Edition
BusinessInside

Frank Reese (Hrsg.)

Website Testing

Conversion Optimierung für Landing Pages und Online-Angebote

BusinessVillage
Update your Knowledge!

Frank Reese (Herausgeber)
Website-Testing
Conversion Optimierung für Landing Pages und Online-Angebote
1. Auflage, Göttingen: BusinessVillage, 2009
ISBN 978-3-938358-58-0
© BusinessVillage GmbH, Göttingen

Bestellnummer
Druckausgabe Bestellnummer PB-806
ISBN 978-3-938358-58-0

Das Buch ist alternativ auch in einer komplett farbig gedruckten Ausgabe erschienen:

Frank Reese (Herausgeber)
Website-Testing
Conversion Optimierung für Landing Pages und Online-Angebote
1. Auflage, Farbdruck, Göttingen: BusinessVillage, 2009
ISBN 978-3-938358-96-2

Bezugs- und Verlagsanschrift
BusinessVillage GmbH
Reinhäuser Landstraße 22
37083 Göttingen
Telefon: +49 (0)551 2099-100
Fax: +49 (0)551 2099-105
E-Mail: info@businessvillage.de
Web: www.businessvillage.de

Layout und Satz
Sabine Kempke

Illustrationen auf dem Umschlag
Titelbild-Collage: Sabine Kempke
Titelbild Hintergrund: Kheng Guan Toh, fotolia
Titelbild Figur: Ioannis Kounadeas, fotolia

Druck und Bindung
Hubert & Co, Göttingen

Copyrightvermerk
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar.
Das gilt insbesondere für Vervielfältigung, Übersetzung, Mikroverfilmung und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Alle in diesem Buch enthaltenen Angaben, Ergebnisse usw. wurden von dem Autor nach bestem Wissen erstellt. Sie erfolgen ohne jegliche Verpflichtung oder Garantie des Verlages. Er übernimmt deshalb keinerlei Verantwortung und Haftung für etwa vorhandene Unrichtigkeiten.
Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürfen.

Einleitung 9

Abschnitt 1: Grundlagen

I. Die Kunst des Experiments – Oder: Wozu eigentlich testen? ***(Frank Reese)*** 13

Regeln des Experiments 24
Regel Nr. 1: Probieren statt Diskutieren 24
Regel Nr. 2: Erst Ziele festlegen, dann Testen 25
Regel Nr. 3: Nach dem Test ist vor dem Test 25

II. Optimierung als Prozess: Usability Engineering ***(Martin Beschnitt)*** 27

1. Definition & Vorteile 29
2. Der Usability-Engineering-Lifecycle 31
3. Wann welche Methode? 33
4. Fazit 34

III. Parameter des Erfolgs: Die Festlegung der richtigen Kennzahlen ***(Frank Reese)*** 35

1. The Name of The Game: Wozu gibt es diese Website? 37
2. Etappen, Brennpunkte und Call-to-Action 37
3. Performance-Messung 39
4. Einzelne Kennzahlen 43

IV. Ansatzpunkte zur Conversion Optimierung ***(Frank Reese)*** 57

1. Optimierung der Startseite 59
2. Landing Pages 64
3. Formularseiten 70
4. Interne Suchmaschinen 74
5. Bestellprozesse 76

Abschnitt 2: Methoden

V. Optimierung in der Konzeptphase: Mit Nutzern gestalten ***(Daniel R. Schmeißer)*** 81

1. Besonderheiten der Konzeptphase 83
2. Verfahren zur Optimierung in der Konzeptphase 87
3. Fazit: Von der Idee zum erfolgreichen Konzept 108
4. Literaturverzeichnis 108

VI. Rapid Prototyping: Optimierung durch iteratives Usability-Testing ***(Daniel R. Schmeißer)*** 111

1. Paper Prototyping: Entwickeln von Interaktionsabläufen 113
2. Usability-Tests mit Prototypen: Optimierung von Interaktionsabläufen 117
3. Besonderheiten bei der Durchführung von Rapid Prototypings 120

4. Anforderungen an die Projektbeteiligten 124
5. Fazit: Iterativ mit Nutzerfeedback optimieren 126
6. Literaturverzeichnis 126

VI.I Fallbeispiel: Rapid Prototyping für die Deutsche Post ***(Daniel R. Schmeißer; Tobias Kauer)*** 127

1. Testdesign: Iteratives Prototyping in zwei Wellen 128
2. Optimierungsresultate aus den beiden Testwellen 130
3. Fazit: Durch iteratives Prototyping zum Erfolg 135

VI.II Fallbeispiel: Rapid Prototyping für netdoktor.de ***(Carsten Rehmann)*** 137

1. Der Untersuchungsgegenstand 138
2. Vorarbeiten 138
3. Ziele des Tests 139
4. Vorgehensweise 139
5. Ergebnisse – Übersicht über die Probleme und Empfehlungen 141
6. Ausgewählte Erkenntnisse 142
7. Learnings 144

VII. Mouse Tracking und Klickanalysen ***(Dirk Schulze)*** 147

1. Usability authentisch testen 148
2. Technik 149
3. Heatmaps von Klicks und Mausbewegungen 150
4. Webformulare optimieren 151
5. Ausblick: Aktivitätsdauer statt Page Impressions? 152

VIII. Einfache Tests mit A/B-Vergleichen ***(Frank Reese)*** 153

1. A/B-Tests: Schlicht aber effektiv 154
2. Praktische Herangehensweise 157

VIII.I Fallbeispiel: A/B-Test für einen Download-Link ***(Frank Reese)*** 161

1. Der „Power Split Tester" 162
2. Fazit 165

IX. Multivariate Testverfahren ***(Frank Reese)*** 167

1. Ansatzpunkt, Elemente und Variationen 168
2. Repräsentativität und Segmentierung 170
3. Auswahl des Testsystems 173
4. Testkombinationen und Teilnehmerzahl 173
5. Wahrscheinlichkeitsverteilungen und Konfidenzintervalle 174
6. Reduzierung der notwendigen Seitenabrufe 180
7. Fazit 183

IX.I Fallbeispiel: AutoScout24 – Evaluierung von Testing Tools ***(Sebastian Wetterauer)*** 185

1. Fachliches Test Setup 186
2. Technisches Test Setup 186
3. „Fahrzeuge Inserieren" Flights 187
4. „Fahrzeuge Suchen und Finden" Flights 188
5. Vorläufige Learnings Testing 190

IX.II Divolution – Optimierung eines KFZ-Versicherungsangebotes ***(Michael Jonas)*** 193

1. Ausgangssituation 194
2. Projektablauf: Phasen der Planung 195
3. Testkonzeption 196
4. Testdurchführung 197
5. Testergebnisse 198
6. Konkrete Empfehlungen und weiterführende Maßnahmen 199
7. Fazit 200

IX.III Fallbeispiel: Arbeit mit dem Google Website Optimizer ***(Birger Friedrichs)*** 201

1. Test-Typ 1: A/B Test 202
2. Test-Typ 2: Multivariater Test 202
3. Was soll ich testen oder die Hypothese 203
4. Das Testszenario und die Hypothese bei SIBO 204
5. Die Durchführung des Experiments 205
6. Das Ergebnis 211
7. Nützliche Tools, Websites und Bücher 212

X. User Experience Tests – Webseiten ganzheitlich optimieren ***(Daniel R. Schmeißer)*** 215

1. Gegenstand des User Experience Tests 218
2. Vorbereitung von UX-Tests 225
3. Durchführung von UX-Tests 231
4. Ableiten von Handlungsempfehlungen 240
5. Fazit: Durch Full User Experience zum Erfolg 243
6. Literaturverzeichnis 244

X.I Fallbeispiel: User Experience Test für neckermann-reisen.de
(Daniel R. Schmeißer; Oliver Rengelshausen) 247

1. Untersuchungsdesign zur Erfassung der UX 249
2. Ergebnisse und Optimierungsansätze 251
3. Fazit: Mit verbesserter User Experience zum Erfolg 261

XI. Eyetracking – mit den Augen der Nutzer sehen ***(Daniel R. Schmeißer)*** 263

1. Eyetracking und Web-Usab: Anwendungsbereiche 265
2. Blickmessung: Kennzahlen und Parameter 271

3. Alternativen zur Blickmessung? 276
4. Fazit: Eyetracking als Bestandteil von UX-Tests 277
5. Literaturverzeichnis 278

XII. Remote User Experience Tests Messen, befragen und beobachten in natürlicher Testumgebung ***(Tim Bosenick)*** 279

1. Definition Remote User Experience Tests 280
2. Das Testverfahren 282
3. Einsatzmöglichkeiten 283

XII.I Fallbeispiel: Remote Testing DER CLUB ***(Tim Bosenick)*** 285

XII.II Fallbeispiel: Ein E-Commerce-Monitor mit Remote Testing ***(Tim Bosenick)*** 289

1. Wettbewerbsvergleichende Studien am Beispiel des E-Commerce-Monitor 290
2. Ausblick 293
3. Literatur 294

XIII. Expertengutachen: User Experience kurz und gut ***(Andreas Selter)*** 295

1. Überblick 296
2. Ausgangssituationen und Einsatzmöglichkeiten 296
3. Auf die richtige Vorgehensweise kommt es an 297
4. Die richtigen Themenschwerpunkte setzen 298
5. Qualitätsmaßstäbe 300
6. Fazit 301

XIII.I Fallbeispiel: Expertengutachen für D&B ***(Andreas Selter)*** 303

1. Die Ausgangssituation 304
2. Die Vorgehensweise 304
3. Erkenntnisse und Lösungen aus dem Expertengutachten 305
4. Fazit 310

XIV. Anbieter, Systeme, Agenturen ***(Frank Reese)*** 311

1. Clickmaps, Mausbewegungsanalysen und ähnliches 312
2. Multivariate Analysen 313
3. Die große Familie der User Experience Tests 314

Die Autoren 316

Vorwort & Danksagung

Die Analyse von Besucherbewegungen auf Online-Angeboten hat sich in den letzten Jahren deutlich weiterentwickelt. Nicht zuletzt das kostenlose Angebot von Google – Google Analytics – hat zu einem starken Vebreitung von Traffic-Analysen beigetragen, aber auch große Online-Angebote werden inzwischen von professionellen Systemen routiniert gemessen und analysiert.

Google hat mit seinem ebenfalls kostenlosen „Google Website Optimizer" auch gezeigt, dass die Messung nur der erste Schritt zu einer erfolgreicheren Website ist. Das wahre Potenzial von automatisierten Auswertungen liegt in einer systematischen Optimierung entlang gemessener Performance-Parameter. Neben dem Google-Tool gibt es im Bereich sogenannter multivariater Analysen bereits seit einigen Jahren eine Reihe ausgereifter Testsysteme aus Deutschland, Großbritannien und den USA, die hier ebenfalls kurz vorgestellt werden.

Aber Website-Optimierung bzw. Conversion Optimierung ist nicht nur über Traffic Analysen möglich. Eine wichtige Idee dieses Buches war es darum, ergänzende und alternative, darunter teilweise lange etablierte Verfahren, wie etwa Mausbewegungsanalysen, Clickmaps, Eyetracking oder User Experience Tests vorzustellen und dem Leser so ein rundes Bild über die verschiedenen Möglichkeiten der Optimierung zu bieten.

Dieses gesammelte Fachwissen fundiert und anschaulich darzustellen, wäre für einen Autor allein kaum machbar gewesen. Darum war ich sehr froh, dass viele Experten aus der Branche spontan ihre Mitarbeit zugesagt haben. Ohne sie wäre dieses Buch nicht möglich gewesen. Ich bedanke mich darum herzlich bei allen, die neben ihrer täglichen Arbeit Inhalte und Fallbeispiele zu diesem Buch beigetragen haben.

Zuletzt noch ein Wort zum verwendeten Vokabular und der Form. Da sich das Buch an ein Fachpublikum richtet, schien es angemessen, die Bedeutung der meisten Begriffe vorauszusetzen. Zudem werden oft englische Bezeichnung und auch Zitate verwendet. Ungewohnt und etwas umständlich mag manchem Leser die Verwendung von Fußnoten erscheinen. Wir haben uns jedoch bemüht, nützliche Zusatzinformationen in den Fußnoten unterzu-bringen und damit den Text noch leserlicher zu gestalten, ohne dem Leser weitergehende Informationen vorzuenthalten. Für das Verständnis des Textes sind die Fußnoten in keinem Fall notwendig.

Frank Reese,
Berlin, März 2009

Einleitung

Vor etwa sechs oder sieben Jahren sollte ich einem Unternehmen dabei helfen, das Online-Angebot persönlicher zu machen. Es gab eine Registrierungsmöglichkeit, die damals üblichen ‚my'-Bereiche und ähnliches und die Frage war: "Wie können wir die Besucher individueller ansprechen und binden. Ein Vorschlag von mir war, die Nutzer – wenn sie denn schon namentlich bekannt sind – auch mit ihrem Namen zu anzusprechen, etwa ‚Schön, dass Sie uns wieder besuchen Herr Müller'. Kein besonders origineller Vorschlag, zugegeben, aber in der Runde brach sofort eine muntere Diskussion zum Für und Wider von solchen Personalisierungsfunktionen aus: „Ich möchte nicht so angesprochen werden", „Die Besucher werden sich ausspioniert fühlen", etc.

Die Frage, „Wie können wir unsere Besucher persönlicher ansprechen?", wurde aus der persönlichen Sicht der Beteiligten diskutiert: „Was finde ich gut?", „Wie würde ich reagieren?", „Wie sind meine Erfahrungen?", Argumente und Meinungen wurden ausgetauscht und gegenübergestellt ohne zu einem neuen Konsens zu finden. Wahrscheinlich kennt jeder Leser diese Art von Meeting, in der die Anwesenden ihre persönliche Meinungen zum Besten geben und am Ende – wie es bei Kaushik heißt – nach HiPPO (*‚Highest Paid Person's Opinion'*) entschieden wird.

Natürlich ist es immer wenig erfreulich, wenn der eigene Vorschlag nicht umgesetzt wird. Wichtiger für mich war aber damals die Erkenntnis, dass die beste Lösung mit solchen Diskussionsrunden nur zufällig gefunden werden kann.

Nicht lange danach habe ich für mich mit den Protokolldateien der Webserver die Möglichkeit entdeckt, die Bewegungen der Website-Besucher direkt nachvollziehen zu können. Diese Daten bieten die Grundlage für gute Entscheidungen jenseits von Trendrhetorik und Meinungen. Die Beschäftigung mit den Traffic-Analysen führte im letzten Jahr zur Veröffentlichung des Buches „Web Analytics – Damit aus Traffic Umsatz wird". Das Buch konzentriert sich auf die effektive Analyse der Datenströme aus den Besucherinteraktionen: Welche Daten sind wichtig? Welche Kennzahlen, Segmentierungen oder Visualisierungen hilfreich? Wie müssen Reports aussehen, die nicht nur für die Galerie abgespeichert werden sondern Entscheidungen effektiv unterstützen?

Web Analytics entfaltet sein volles Potenzial aber erst in einer systematischen, datengestützten Weiterentwicklung von Online-Angeboten. Die Frage, „Ist es sinnvoll die Besucher persönlich anzusprechen?" kann auf einfache, zuverlässige und zeitsparende Weise durch die Besucher selbst entschieden werden. Alle Erfahrung zeigt, dass Experten – wir alle – bei der Einschätzung von sinnvollen Funktionen, Gestaltungen und Formulierungen oft

irren. Dagegen zeigen schon die Ergebnisse von einfachen A/B-Tests oft erstaunliche Optimierungspotenziale (zum Beispiel in Form der Steigerung von Konversionsraten).

Das hier vorliegende Handbuch konzentriert sich auf die Optimierung von Website-Angeboten. Vorgestellt werden eine Reihe von Verfahren, die bei der systematischen Verbesserung von Online-Angeboten helfen. Die Methoden reichen von der Konzeptphase bis zur fertigen Site, von Befragungen bis zu Verhaltensmessungen, von Labortests bis zu Tests im laufenden Betrieb.

Damit werden in diesem Buch das erste Mal – jedenfalls soweit wir wissen – ganz unterschiedliche Ansätze gemeinsam vorgestellt. Traditionell fallen die Vorgehensweisen insbesondere von Befragungen, Usability-Methoden und Analytics-Techniken in unterschiedliche Fachbereiche. Die Struktur der Unternehmen, die Ansprechpartner ja auch die Denkweisen scheinen nicht viel gemeinsam zu haben. Und doch gehören sie alle in den gleichen Werkzeugkasten und haben das gleiche Ziel: Ihnen bei der Optimierung Ihrer Online-Angebote zu helfen. Nicht die Methode gibt das Vorgehen vor, sondern das Ziel bestimmt die Auswahl des Werkzeugs, mit dem es am einfachsten erreichbar ist.

Denn so unterschiedlich die hier vorgestellten Verfahren auch scheinen mögen: Gemeinsam helfen Sie bei einer messbaren Verbesserung im Sinne des Website-Ziels. Sie sollen nicht dazu dienen, irgendein diffus-abstraktes besser-als-vorher zu erreichen. Eine höhere Verweildauer der Besucher, klarere Nutzerführungen, mehr oder bessere Produktinformationen, transparente Bestellprozesse, mehr Social-Web-Elemente … alles das sind Hilfsmittel (keine Ziele) auf dem Weg zu einer effektiveren Website.

Das Handbuch soll dem Praktiker in den Unternehmen eine Toolbox mit verschiedenen Werkzeugen in die Hand geben und dazu noch Hinweise und Anleitungen wann welches Werkzeug am besten einzusetzen ist. Um den realen und konkreten Ablauf der Optimierungsprojekte noch deutlicher zu machen sind allen Verfahren ausführliche Beschreibungen eines praktischen Projekts zugeordnet. Und nicht zuletzt findet der Leser auch Unternehmen, Ansprechpartner und Preise für die einzelnen Verfahren. Dieses Handbuch sollte Praktikern alles liefern, um eine Optimierung ihrer Online-Angebote in Angriff zu nehmen.

I. Die Kunst des Experiments Oder: Wozu eigentlich testen?

von Frank Reese

Regeln des Experiments 24
Regel Nr. 1: Probieren statt Diskutieren 24
Regel Nr. 2: Erst Ziele festlegen, dann Testen 25
Regel Nr. 3: Nach dem Test ist vor dem Test 25

Wir sind alle Experten. Aus unserer eigenen Erfahrung wissen wir, was geht und was nicht geht. Flash-Intros auf der Startseite? „Geht gar nicht!“ oder: „Unsere Zielgruppe erwartet emotionale Visualisierungen!“. Vom „Artikel-in-den-Warenkorb“-Button direkt in den Bestellprozess? „Unsere Besucher wollen eher noch nach weiteren Produkten schauen.“ oder: „Wir wollen es unseren Kunden möglichst leicht machen.“

Seit den professionellen Anfängen des Netzes toben Grundsatzstreits um Navigationsarten, Sitemaps, Cookies, Frames, Personalisierungen, Farbgebungen usw. usf. mit nicht nachlassender Zähigkeit. Moden und Trends, neue Zielgruppen und Technologien, wechselnder Geschmack und veränderte Gewohnheiten vermehren und ergänzen jenen allbekannten Schatz an Weisheiten, dessen Sätze wie Mantras nachgebetet und verteidigt werden. Und: Wer kann schon über Geschmack streiten? Ist dieses Blau nicht doch besser als jenes? Gefällt unseren Besuchern ein dreispaltiges Layout besser als ein zweispaltiges? Meetings, die sich mit solchen Fragestellungen beschäftigen, können zu endlosen Debattierveranstaltungen mutieren, in denen die jeweiligen Befürworter noch die abseitigsten Begründungen für „ihr“ Blau bemühen. Am Ende entscheidet dann doch der Leiter oder Chef oder Kunde nach seinem Geschmack, welches Blau auserwählt wird.

Warum fällt es uns so schwer, zuzugeben, dass wir nicht genau wissen, was besser und was schlechter ist? Offensichtlich ist unser Wissen über die Vorlieben der Website-Besucher doch begrenzt, schließlich geht es dabei nicht selten um Millionen von Menschen. Wir können gar nicht wissen, was diese in der Mehrheit präferieren. Gut, es gibt Mitarbeiter, die haben das richtige Gefühl im Bauch. Die wissen vielleicht, was ankommt, die haben eine Nase für den Geschmack der Besucher. Für diese – und alle anderen natürlich auch – ein kleiner Test zum Bauchgefühl. Die Lösungen zu den fünf Beispielen stehen nach der letzten Frage verkehrt herum am Seitenende – wer mehr als drei Mal ohne zu Raten und ohne das Beispiel bereits zu kennen, richtig liegt, dem sei herzlich gratuliert.

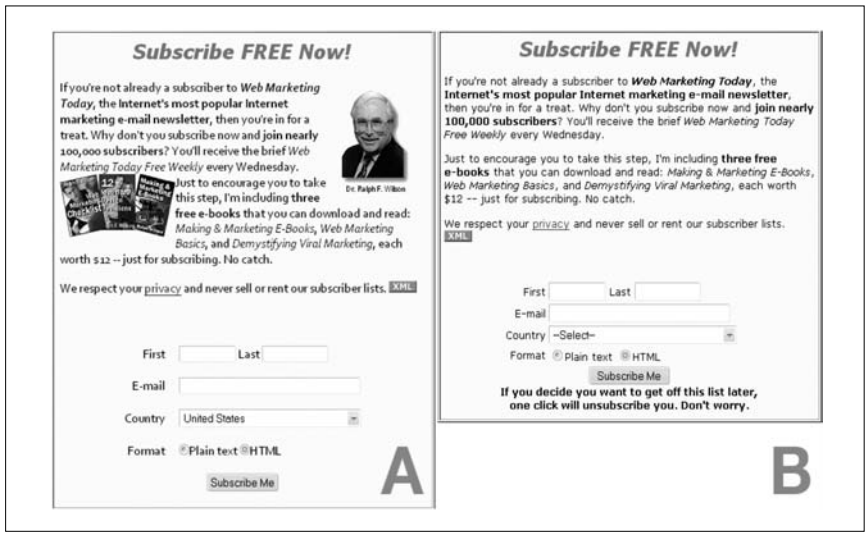
Fünf Testfragen

Frage 1: Welcher Button wird am häufigsten geklickt?

Frage 2: Welche Gestaltung führt zu mehr Registrierungen?

Frage 3: Welche E-Mail generiert mehr Bestellungen?

Frage 4: Mit welchem Formular registrieren sich mehr Besucher?



Frage 5: Welche Betreffzeile führt zu höheren Klickraten bei einer E-Mail?

A: „Computer Security Quarterly – Spring '08"

B: „2007 Record Year For Data Breaches"

Sie haben weniger als drei Mal richtig entschieden? Kein Problem: Erstens wird es wahrscheinlich den allermeisten Lesern so gehen und zweitens – nicht zuletzt – freue ich mich, Ihnen noch ein wenig die Idee des Experiments nahebringen zu dürfen.

Lösung zu den Fragen:

1. Button A
2. Version B erzeugt 15 Prozent mehr Konversionen
3. Die rechts stehende Version generiert eine geringe Klickrate, aber doppelt so viele Konversionen
4. Mit Version B wurde die Konversionsrate um 25 Prozent gesteigert
5. Betreffzeile A

Wir alle kennen den Begriff des Experiments aus den Wissenschaften, genauer den Naturwissenschaften. Bis etwa ins 17. Jahrhundert hinein war es üblich, strittige Fragen zwischen Gelehrten mit Hilfe des Disputs zu entscheiden. In einem festgelegten Verfahren wurden die Argumente der Parteien dargelegt, Prämissen und Prinzipien abgewogen und im besten Fall einige als „wahr" anerkannt". Ein logischer Schluss, der die Regeln beachtet und auf wahren Prämissen beruht, musste dann notwendig auch „wahr" sein. Widerspruchsfreiheit, Präzision in der Analyse, Unterscheidung von notwendigen und hinreichenden Bedingungen, waren die Grundlagen der Erkenntnis und der Unterscheidung von Wahrem und Unwahrem. Mit Empirie (von griechisch empireia: „Erfahrung"), also dem Test in der realen Welt hatte das nicht viel zu tun.

Der englische Wissenschaftler Francis Bacon (1561-1626) war einer der ersten, der die Empirie, das heißt, die Erfahrung, zum entscheidenden Werkzeug der Wissenschaft machte.

„Der beste Beweis liegt in der Erfahrung."
Francis Bacon, englischer Philosoph und Wegbereiter des Empirismus

Mit seinen Zeitgenossen Galilei, Kepler und anderen wurde die Erfahrung und methodische Induktion [1] in das Zentrum des Erkenntnisprozesses gestellt.

Es gibt Dinge, die sind kaum über ein klassisches Experiment entscheidbar und Menschen werden sich noch lange mit Fragen wie „Was geschieht nach dem Tod?" oder „Gibt es einen Weltenschöpfer?" beschäftigen. Aber der naturwissenschaftliche und technische Fortschritt hat sich von philosophischen und religiösen Erkenntnismethoden emanzipiert. Galilei belegte mit sorgsam durchdachten Anordnungen seine Fall- und Wurfgesetze, Isaac Newton klärte in seiner Mechanik die die physikalischen Gesetze von Kraft, Masse und Gewicht, Gregor Mendel zeigte mit seinen Versuchen zur Artenkreuzung die Vererbungsgesetze, nach vielen Diskussionen und Ablehnung belegten die Experimente von Michelson & Gale Einsteins Relativitätstheorie und viele mehr ließen sich anführen. Die zeitgenössischen Gelehrten diskutierten jeweils so lange mit ihren Lieblingsargumenten und Lieblingstheorien ohne zu einem Schluss zu kommen, bis ein Experiment die Entscheidung brachte.

[1] Induktion = von mehreren Beobachtungen auf ein gemeinsames Prinzip schließen, versus Deduktion = vom Allgemeinen (Prinzip) auf das Besondere (zum Beispiel konkrete Beobachtung) schließen

Dabei hat sich im Laufe der Zeit die Idee der Wahrheit, des einzig gültigen Gesetzes und der vollkommenen Richtigkeit verflüchtigt. Mit den menschlichen Methoden ist ‚Wahrheit' nicht zu erlangen. Der österreichisch-britische Philosoph Karl Popper hat dieses Vorgehen nach dem Prinzip Versuch und Irrtum in den Mittelpunkt wissenschaftlichen Handelns gestellt.

„Alles Leben ist Problemlösen."
Karl Popper, österreichischer Philosoph und Begründer des Kritischen Rationalismus

Nach Popper dient eine Hypothese dazu, widerlegt zu werden. Stimmt eine Hypothese mit den Ergebnissen eines Experimentes überein, so wird sie beibehalten. Sie wird so lange gegen andere, neue Hypothesen so getestet, bis es eine bessere oder elegantere Erklärung für das in Frage stehende Phänomen gibt.

Wahrheit spielt keine Rolle mehr. Es ist unmöglich, eine endgültige These aufzustellen, die nicht widerlegt werden kann. Was wir haben, sind Theorien über die Welt, die mehr oder weniger gut die Phänomene darin erklären können. Morgen werden es vielleicht andere sein. Als Kindern wurde uns noch gesagt: „Iss Spinat, der enthält viel Eisen, das ist gut für die Blutbildung". Heute weiß man es besser: Eisen findet sich in vielen Nahrungsmitteln in ausreichenden Mengen, Spinat wird Kindern – jedenfalls aus diesen Gründen – nicht mehr aufgetischt.

Theorien, die Grundlage für eine bestimmte Vorgehensweise sind, beanspruchen keinen Wahrheitsstatus mehr, sondern bieten einen Nutzen: Phänomene zu erklären, Naturereignisse vorherzusagen, eine Maschine funktionieren zu lassen, eine stabile Brücke zu konstruieren. Nutzen ist immer mit Zielen verknüpft: Es geht nicht um absolute Wahrheiten, sondern darum, Dinge zu Laufen zu bringen.

Um nun wieder den Bogen zu unserem Thema zu schlagen: Ob „unsere Zielgruppe emotionale Visualisierungen mag", „Nutzer eine Abneigung gegen Cookies haben", oder blau unterstrichene Hyperlinks besser erkannt werden als grüne, das sind abstrakte Wahrheiten die weder beweisbar noch wichtig sind. Sie liefern nur Grundlagen für lange, ausufernde Diskussionen ohne befriedigendes Ergebnis. Entscheidend ist: Was will ich mit diesem Projekt erreichen? Am Anfang einer *unfruchtbaren Diskussion* über die Frage „Sollen wir unsere Artikel auf der Website mit einem Preis versehen oder nicht?", stehen grundsätzliche Meinungen und Einschätzungen, die ohne konkreten Bezug auf die Wirklichkeit hinsichtlich ihres Wahrheitsgehaltes diskutiert werden. Da es kein hartes Kriterium für Wahrheit

und Priorität gibt, muss man den Teilnehmern eine Neigung zum Rechthabenwollen unterstellen.

Am Anfang einer *fruchtbaren Diskussion* über die Frage: ‚Preise oder keine Preise auf der Website?" steht dagegen zunächst die Festlegung eines konkreten, messbaren Nutzens dieser Entscheidung und ihrer Alternativen: „Was soll am liebsten auf genau dieser Seite des Online-Angebotes passieren?" „Was wollen wir, die verantwortlichen Gestalter dieser Seite, dass der Nutzer auf dieser Seite macht?"

Natürlich gibt es Fragen, die lassen sich nicht experimentell testen: „Gibt es außerirdische Wesen?", „Werden wir wiedergeboren?", oder eben: „Ist die Angabe von konkreten Preisen auf einer Website richtig oder falsch?" – „It depends" muss man auf die letzte Frage mit Jim Sterne[2], einem der heute prominentesten Fürsprecher von Web Analytics und Testmethoden, antworten. Und in diesem „It depends" steckt der Hebel, um aus unfruchtbaren Diskussionen effektive Entscheidungsfindung zu machen. Ob die Preisangabe gut oder schlecht ist, hängt davon ab, worum es im konkreten Einzelfall geht. Sie können sich nicht auf allgemeine Ratschläge verlassen, Ihr Geschäft, Ihre Kunden und Besucher sind zu individuell, um pauschal beurteilt zu werden. Dabei geht die Bewegung immer von allgemeinen, abstrakten Aussagen zu konkreten, messbaren Zusammenhängen: Konversionen, Besucherloyalität, Absprungraten … – was ist unser Ziel? Wo liegt der Nutzen und wie kann ich ihn messen? Das Schöne am Online-Marketing ist ja die unmittelbare Erfassung aller Interaktionen der Besucher mit dem eigenen Angebot. Das sogenannte Performance Marketing hat gezeigt, dass Online-Marketing nicht auf vielfach vermittelte, oft verdeckte Branding-Effekte vertrauen muss, um die eigene Wirksamkeit umständlich zu begründen. Die Kette von der Präsentation der Anzeige (AdView), über den Klick darauf, den Aufruf von Landing Page und Produktseiten, bis zu Bestellvorgang und Abschluss, ist vom Anfang bis zum Ende erfassbar. Sicher ist diese Kette nicht die letzte, vollkommene Darstellung der Konversion eines Anzeigenbetrachters zu einem Kunden. Es gibt viele Hinweise, dass vor dem Klick auf das Adword sowie zwischen Adword-Klick und Konversion mehrere Kontakte mit anderen Inhalten (zum Beispiel Bannern, Produktbewertungen, Preisvergleichen) stehen. Die einfache Performance-Marketing-Kette bietet also noch keine eindeutige Erklärung für das Verhalten der Besucher – weitere Glieder und Zusammenhänge werden aufgedeckt werden. Diese Unvollkommenheit liegt aber in der Natur der Sache, das heißt

[2] Jim Sterne veröffentlichte zusammen mit Matt Cutler bereits 2000 einen wegweisenden Text zum Thema („E-Metrics: Business metrics for the new Economy", pdf, im Internet kostenlos als Download abrufbar), begründete mit anderen die Branchenvereinigung WAA (Web Analytics Association) und veranstaltet mit den weltweit stattfindenden Emetrics-Summits die wichtigste Konferenzplattform der Branche.

im Prozess der Hypothesenformulierung, dem Austesten dieser neuen Annahmen und der Erkenntnisgewinnung – eine endgültige Erklärung wird dabei nicht herauskommen, sehr wohl aber Grundlagen für pragmatische Handlungsweisen. Vor allem aber sollte diese Unvollkommenheit nicht neuerlichen Anlass zu wolkigen Begründungen und unbestimmten Behauptungen geben.

Die Messbarkeit der Interaktionen, die Erfassung, Wiedererkennung und Verfolgung von Besuchern über Cookies, Registrierungen, automatische Abrufe, GPS-Ortungen, RFID, Kundenkarten uvm. gehört sicher zu den revolutionärsten Elementen der neuen digitalisierten Beziehung zwischen Anbietern und Besuchern. Viele Bücher sind bereits über die neuen Verhältnisse zwischen Unternehmen und Kunden geschrieben worden, zu persönlichen 1to1-Beziehungen, zu neuem Informations- und Kaufverhalten, zu Geschäftsmodellen und Gestaltungsmöglichkeiten. Hier soll dies nicht weiter ausgeführt werden, auf einige interessante Bücher sei aber verwiesen [3].

Auch wenn der Online-Wirtschaft die Werkzeuge und Informationen zur Analyse von Besuchern und Bewegungen geradezu in die Wiege gelegt wurden, bleibt doch noch viel von traditionellen Methoden zu lernen.

Zwar ist es inzwischen nicht unüblich, Newsletter und E-Mails automatisch ansatzweise personalisiert zu versenden („Hallo Herr %Vorname %Nachname"), doch oft wird dies eher nachlässig behandelt. Diese Art von Personalisierung scheint vielen als nice-to-have aber nachrangig zu gelten. Wer einmal im klassischen Direkt-Marketing nachfragt, wird dort schnell eines Besseren belehrt: Selbstverständlich gibt es einen klaren Zusammenhang zwischen Response-Raten und personalisierter Ansprache auf Wurfsendungen, Briefen, Werbemitteln. Jeder von uns weiß, dass er viel eher einen an ihn persönlich adressierten Brief öffnet und liest – „Herzlichen Glückwunsch, Herr Reese, Sie haben gerade 7 Millionen Euro gewonnen!" – als eine anonyme Werbenachricht.

Eine bestimmte Art von Usability war immer ein großes Thema im E-Commerce. Dem guten Jakob Nielsen [4] sei das Beste gewünscht, aber ist eigentlich mal jemandem aufgefallen, dass klassische Supermärkte nicht eben übersichtlich bzw. nach den Regeln von ‚Usability'

[3] Akin Arikan: Multichannel Marketing: Metrics and Methods for on and Offline Success (Wiley 08); Chris Anderson: The Long Tail (Hyperion 06); Amy Shuen: Web 2.0: A Strategy Guide (O'Reilly 08); Jeff Howe: Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business

[4] Nielsen ist einer der bekanntesten und einflussreichsten US-Autoren im Usability-Bereich. Sein Online-Angebot www.useit.com und seine Bücher, zum Beispiel Web Usability (2008), wurden vieltausendfach gelesen und diskutiert.

eingerichtet sind? Wie viele Artikel, Leitfäden, „10 Tipps zur Gestaltung Ihrer Homepage", sind geschrieben worden mit Hinweisen wie: „Gestalten Sie die Navigation übersichtlich", „Setzen Sie Flash-, Grafik-, Java-Elemente nur dort ein, wo es nicht anders geht.", „Nutzen Sie dezente Farben, Kontraste und klare Schriftarten" etc.

Tatsächlich geht es in Online-Shops aber nicht um die Erledigung von Aufgaben (s. Software-Ergonomie), sondern um den Verkauf von Produkten, ganz so wie ein Ladengeschäft eben nicht den Gestaltungskriterien etwa einer öffentlichen Bibliothek gehorcht. Supermärkte sind darum nicht primär übersichtlich, vielmehr leiten sie den Kunden gezielt durch die verschiedenen Produktsparten, versuchen das Besucher-Tempo zu beeinflussen (Obst- & Gemüsebereich als Bremszone), verlegen vielbesuchte Sparten an das Ende des Marktes (Frischetheken), positionieren im Regal nicht den günstigsten, am meisten gewünschte Artikel in Augenhöhe (Sichtzone) oder in der am bequemsten zu erreichenden Greifzone, sondern platzieren dort die teureren Artikel, die der Supermarktbetreiber lieber verkaufen will. Und schließlich haben Mütter mit kleinen Kindern eine große Freude im Kassenbereich: Just in Reichweite der kleinen Hände sind jene Artikel eingerichtet, die einen besonderen Reiz auf den Nachwuchs ausüben.

All diese und viele weitere Methoden sind nicht das Ergebnis von persönlichen Vorlieben oder Zufällen. Im konventionellen Marketing ist es seit Jahrzehnten üblich, verschiedene Positionen, Farben oder Ansprachen an kleineren Gruppen von Besuchern zu testen. Große Lebensmittelunternehmen betreiben sogar besondere Testfilialen, in denen neue Produkte, Platzierungen oder Gestaltungen ausprobiert werden. Bevor eine bundesweite Umstrukturierung von Ladengeschäften vorgenommen wird, muss sich eine Neuerung erst im Test beweisen. Das bietet zwei Vorteile: Zum Einen können kostenintensive Fehlschläge vermieden werden, zum Anderen bietet ein dediziertes Testumfeld den Raum für die schnelle Umsetzung von neuen Ideen und Ansätzen. Dabei werden für die Bewertung so viele Wege der Datensammlung und Auswertung herangezogen wie nur möglich. Scannerkassen und Kundenkarten bieten viele Möglichkeiten, Zusammenhänge zwischen Preisbildung, Gestaltung und Zielgruppen herauszufinden. Ein besonders eindrucksvolles Beispiel, wie das Verhalten von Kunden in einem Ladengeschäft analysiert werden kann, beschreibt Paco Underhill in seinem Buch „Warum kaufen wir?"[5].

[5] Paco Underhill: Why We Buy, Simon & Schuster, 2004, deutsche Ausgabe vergriffen

Underhill nutzt anthropologische Beobachtungs- und Analysemethoden nicht um das Verhalten fremder Volksstämme zu erforschen, sondern um die Bewegungen und Handlungen der Menschen während eines Besuches in einem Supermarkt oder einem Kleidungsgeschäft zu beschreiben. Heraus kommen dabei interessante Hinweise auf Unterschiede zwischen Männern und Frauen oder das Phänomen des sogenannten „einarmigen Käufers“ [6].

Die Erforschung des Kundenverhaltens in traditionellen Wirtschaftszweigen hat eine jahrzehntelange Tradition mit eigenen Verfahren, Erkenntnissen und Spezialisten. Im Vergleich dazu stehen kommerzielle Online-Angebote noch ganz am Anfang. Aber die Betreiber von Websites haben einen immensen Vorteil: Sie können nicht nur auf die Vorleistungen ihrer Kollegen aus dem traditionellen Marketing zurückgreifen, für sie läuft die Erfassung und Auswertung des Auswahl- und Kaufverhaltens auch praktisch von alleine und nebenbei. Zudem ist die Einrichtung von Testumfeldern sehr viel kostengünstiger und unkomplizierter. Alles, was es noch braucht ist Offenheit, Flexibilität und Engagement.

Aber Erfahrungen können nicht einfach übernommen werden. Was in dem einen Online-Shop gut funktioniert, kann in einem anderen ganz gegensätzliche Ergebnisse zeigen. Auch der Rat von Online-Experten oder Best Practices-Beispiele sollten eher als Hinweise verstanden werden, denn als Wahrheiten. Es gibt sicher Experten, die wissen, wie eine Website für die Ergebnislisten bei Google optimiert wird oder wie man einen Goldenen Löwen in Cannes gewinnen kann. Aber jenseits dessen wird es leicht unübersichtlich: Da gibt es selbsternannte Gurus, die genau zu wissen vorgeben, wie eine State-of-the-Art-Website auszusehen hat. Da gibt es Fachleute, die die Erstellung von speziellen Landing Pages empfehlen, allein zugeschnitten auf eine bestimmte Kampagne und andere, die eher bestimmte Knotenpunkte im Online-Shop als Landing Page nutzen. Es gibt Experten, die großen Wert auf Sicherheitszertifikate und Verschlüsselungen legen und andere, die sagen, der Besucher würde diese Zertifikate doch nicht mehr ernst nehmen, da es derer zu viele verschiedene gibt, was zu Unübersichtlichkeit führt. Es gibt Fachleute, die glauben, dass ein Bestell- oder Registrierungsformular möglichst auf einer Seite dargestellt werden muss und andere, die eine bestimmte Anzahl von kleinen, übersichtlichen Schritten für empfehlenswert halten. Es gibt Experten, die meinen, vor dem Eintritt in den Bestellprozess sollte der Besucher sich erst einmal registrieren oder anmelden, andere empfehlen diesen

[6] Die meisten Menschen tragen fast immer etwas in ihren Händen: Eine Zeitung, eine Hand- oder Aktentasch oder sonst eine Tüte. Betreten sie nun ein Geschäft in der Absicht, etwas Bestimmtes zu kaufen, nehmen sie keinen Einkaufswagen oder -korb, sondern gehen direkt zum Regal und nehmen den Artikel. Für weitere Produkte – die die Anbieter so prominent als Zugreif-Artikel präsentieren, bleibt keine Hand frei. Die Entscheidung fällt gegen den Spontankauf. Einige Unternehmen, zum Beispiel das Möbel-Haus Ikea, platzieren Einkaufskörbe darum nicht nur am Eingang, sondern überall im Markt.

Schritt am Schluss des Bestellprozesses. Es gibt in Online-Shops rote, gelbe, blaue, grüne Buttons für den Warenkorb, es gibt beschriftete Buttons und einfache Pfeile, es gibt kleine Einkaufswagen-Icons und „JETZT BESTELLEN"-Schaltflächen. Auch wenn Content-Management- und Shop-Systeme für eine deutliche Vereinheitlichung von Websites gesorgt haben, steht doch jeder Seitenbetreiber und Online-Verantwortliche wieder allein vor den vielen Fragen, wie dieses oder jenes am Besten zu machen sei – und er verfolgt dies gewöhnlich mit bester Absicht und mit ganzem Herzen.

Diese Entscheidungen müssen getroffen werden. Jeder Seitenbetreiber muss seine eigenen Erfahrungen machen und die besten Elemente für die eigene Site herausfinden. Gefährlich wird es nur, wenn Sie glauben, Sie wüssten es schon vorher immer besser. Wenn Sie auf einer Website etwas sehen, was in Ihren Augen „gar nicht geht" – vielleicht funktioniert das auf der Site prima. Seien sie vorsichtig, wenn Sie oder Ihre Kollegen scheinbar sehr genau wissen, was richtig und was falsch ist. Ob eine Idee funktioniert oder nicht, zeigt sich nicht daran, ob Experten dazu raten, ob die Meeting-Teilnehmer davon überzeugt werden können oder ob der erfolgreiche Konkurrent X dies auch so macht. „Was zählt ist auf'm Platz", wie man beim Fußball sagt. Und was hier zählt, ist allein das Verhalten Ihrer Besucher auf der Live-Site, gemessen an Ihren Zielen. Misstrauen Sie „Best Practices", Leitfäden und allen, die immer genau wissen, was funktioniert oder nicht. Man kann eine neue Website oder einen Relaunch monatelang vorbereiten, Berater hinzuziehen, Fallstudien berücksichtigen, die beste Agentur einkaufen und serienweise Konzepttests im Vorfeld durchführen lassen. Aber das alles ist nur Vorbereitung, Trainingslager, Schaulauf. Ohne dass Sie die verschiedenen Konzepte und Vorschläge auf Ihrer Live-Site testen und messen, werden Sie bestenfalls zum Trainingsweltmeister. Es gibt vielleicht keine endgültige Wahrheit und keine optimale Website, aber es gibt ein endgültiges Urteil: Das Ihrer Besucher.

Dabei sagen Ihnen die Besucher allerdings nie, warum sie dieses oder jenes Layout bevorzugen. Selbst wenn Sie einzelne Besucher fragen oder Testpersonen in Labors „Gedankenprotokolle" aufzeichnen lassen, kommen Sie der Frage nach dem „Warum" nur unwesentlich näher. Tatsächlich ist eine Begründung für die Tatsache, warum Version A erfolgreicher ist als Version B im Testkontext auch gar nicht wichtig. Da Tests und Anpassungen so schnell und live durchgeführt werden können, ist es viel wichtiger, dass eine neue Gestaltung eine höhere Effektivität erzielt, als die Frage, warum sie das tut. Durch Tests gewinnt die Gestaltung von Websites einen so unmittelbaren praktischen Bezug, dass ein Räsonieren über die grundlegenden Prinzipien des Nutzerverhaltens zu einer Luxustätigkeit wird. Der Pragmatismus des Testens führt direkt und ohne Umwege zu einer kontinuierlichen Steigerung der Effektivität einzelner Webpages und von Online-Angeboten insgesamt.

Regeln des Experiments

Das Experiment stellt Ideen auf die Probe. In einem festgelegten Rahmen werden die Parameter für den Erfolg und das Scheitern festgelegt (mehr zu den entscheidenden Zahlen finden Sie im dritten Kapitel) und dann der Realität mit ihren unbarmherzigen Website-Besuchern zum Ausprobieren bereitgestellt. Wie der amerikanische Web-Analytics-Berater Avanish Kaushik sagt: Tests sind „No-Brainer" – jeder versteht die Qualität von Ideen und deren Ergebnisse, wenn sie in harten Zahlen präsentiert werden: Der neue Button zur Kontaktaufnahme ist zwar nicht hübsch und widerspricht allen Gestaltungsempfehlungen, aber die Rate der Kontaktaufnahmen hat sich deutlich erhöht? Voilá – ein weiterer Schritt auf dem Weg zu einer effektiven Website. Ihr Chef will unbedingt ein schickes Flash-Intro im Zentrum der Homepage haben und glaubt, so kommt die Site besser bei den Besuchern an? Gut, sagen Sie nicht: Das geht aus diesen oder jenen Gründen nicht. Stellen Sie lieber die neue Startseite für einen Teil der Besucher online und messen, wie sich Besucherloyalität und Absprungraten entwickeln. Vielleicht lag er ja richtig, aber wenn nicht, können Sie Daten präsentieren, die zeigen, was wirklich passiert (und was das Flash-Intro hinsichtlich der Website-Effektivität kostet).

Regel Nr. 1: Probieren statt Diskutieren

Niemand weiß besser, was Ihre Website-Besucher wollen, als Ihre Website-Besucher.

Voraussetzung für die systematische Nutzung von Tests ist die Bereitschaft, theoretische Gewissheiten und eigene Meinungen auf die Probe zu stellen und auf Scheuklappen und Vorurteile zu verzichten. Offenheit bei der Formulierung von Gedanken und Neugier auf unkonventionelle Ansätze sind der Nährboden auf dem Ideen für eine Verbesserung der Website gedeihen. Einschränkungen wie „Das geht jetzt aber nicht anders", „Das macht man heute so" oder „Unsere Kunden sind daran gewöhnt" ersticken dagegen jede Initiative im Keim. Erfolgreiche Projekte off- und online leben immer von der Kreativität und Flexibilität der Beteiligten.

Regel Nr. 2: Erst Ziele festlegen, dann Testen

Die Regeln der Entscheidung müssen vor dem Ergebnis festgelegt werden.

Wenn die neue Startseite mit dem gewünschten Flash-Intro online ist und gemessen wird, wie die Besucher mit der Seite umgehen, muss schon vorher klar sein, welche Zahlen als positiv und welche als negativ gelten sollen. Die gemeinsame Vereinbarung dieses Korridors bevor die Ergebnisse vorliegen ist ungemein wichtig. Sämtliche Bedenken, Überlegungen und Faktoren können so bereits bei der Formulierung der entscheidenden Zahlen mit einfließen. Hingegen ist mit äußerst unfruchtbaren Diskussionen zu rechnen, wenn Sie erst nach Durchführung der Tests festlegen, welche der Ergebnisse als Erfolg gelten und welche Seiteneffekte zusätzlich beachtet werden könnten. Man sollte nie das kreative Potenzial von Mitarbeitern und Vorgesetzten bei der Generierung zusätzlicher Faktoren zur Erklärung des Besucherverhaltens unterschätzen, wenn erst einmal Ergebnisse vorliegen, die eine gefestigte Meinung in Frage stellen.

Regel Nr. 3: Nach dem Test ist vor dem Test

Es gibt keine Website, die so perfekt ist, dass sie nicht durch weitere Verbesserungen effektiver werden kann.

Bei der Erstellung und Verbesserung von Websites gibt es kein letztgültiges Wort. Wenn der Nutzen einer Gestaltung oder einer Funktion in das Zentrum der Beurteilung und des Handelns gestellt wird, gibt es keine Grenze für eine weitere Steigerung dieses Nutzens. Jedes Testergebnis und jede Steigerung der Effektivität ist nur ein Schritt in die richtige Richtung. So pragmatisch aber das Vorgehen und die Beurteilung von Gestaltungsalternativen zu handhaben ist, so pragmatisch ist auch die Auswahl des nächsten Optimierungsprojektes zu sehen: Eine Verengung der Bemühungen auf einen Button, eine Seite, ein Formular, der, die oder das auf Biegen und Brechen auf ein bestimmtes Konversionsniveau geführt werden soll, ist nicht empfehlenswert. Der im Zentrum stehende Nutzen eines Tests und einer Verbesserung lässt sich nur im Gesamtbild der Website oder der Online-Aktivitäten sehen. Der Nutzen in Form einer Effektivitätsteigerung entsteht nicht abstrakt als Konversionsrate oder Klickrate, sondern als Nutzensteigerung im Sinne des Website-Zieles – und das im Verhältnis zu den eingesetzten Mitteln. 5.000 Euro, die eben noch dafür genutzt wurden, die Absprungrate einer Startseite um 30 Prozent zu vermindern, können als nächstes vielleicht mit größerer Effizienz bei der Optimierung einer Formularseite helfen.

II. Optimierung als Prozess: Usability Engineering

von Martin Beschnitt

1. Definition & Vorteile 29
2. Der Usability-Engineering-Lifecycle 31
3. Wann welche Methode? 33
4. Fazit 34

Spricht man von Usability-Test oder Usability-Testing, so wird damit meist der sog. klassische Nutzertest im Labor assoziiert, bei dem Probanden eine bestehende Website/-applikation oder einen Prototyp aufgabenbasiert nutzen. Diese Nutzung wird genauestens beobachtet und analysiert, um daraus Handlungsempfehlungen zur Optimierung des Testobjektes ableiten zu können.

Solch ein ganzheitlicher Usability-Test (im Buch nachfolgend als User Experience Test bezeichnet) wird meist als abschließender Projektbaustein beauftragt bzw. durchgeführt. Ziel ist dabei eine finale Qualitätskontrolle (Freigabetest) und auch eine gewisse Beruhigung für die Projektverantwortlichen.

Meist liegt diese abschließende Maßnahme jedoch zeitlich so weit hinten, dass die Empfehlungen häufig erst nach dem feststehendem Launchtermin umgesetzt werden können. Zudem entstehen durch die nötigen Änderungen teilweise hohe Kosten, die nicht im Budgetplan einkalkuliert waren.

Nichtsdestotrotz oder vielleicht gerade deswegen ist Usability heutzutage ein sehr wichtiger Wettbewerbsfaktor, um sich von den inhaltlich homogenen Webangeboten, die das gleiche Zielpublikum anvisieren, positiv abzuheben. Daher ist es wichtig bzw. unumgänglich zu „analysieren" und zu „testen".

Usability ist jedoch weder ein Produkt, dass man einmal kauft und danach nicht mehr betrachten muss, noch etwas, was erst abschließend, kurz vor dem (Re-)Launch einer Website/-applikation getestet werden sollte.

Denn: Zum einen ändern sich in unserem digitalen Zeitalter die Nutzungsgewohnheiten und Erfahrungen der Internetnutzer immer schneller. Zum anderen macht es keinen Sinn, am Ende eines Entwicklungsprozesse ein Produkt zu testen, um schließlich herauszufinden, dass man es für die völlig falsche Zielgruppe entwickelt hat. Soll heißen: Man muss bereits in frühen Phasen der Produktentwicklung die Zusammensetzung und Charakteristika seiner potenziellen bzw. bestehenden Nutzer sowie deren Anforderungen und Erwartungen an das Produkt analysieren. Außerdem sollte man sich tunlichst nicht nach dem Launch einer Website/-applikation auf seinen Lorbeeren ausruhen.

Die Vorgehensweise sollte also weg vom finalen Usability-Test hin zu einem optimalen und kontinuierlichen Methodenmix gehen, der die Benutzerfreundlichkeit eines Produktes in unterschiedlichen Entwicklungsstadien und aus verschiedenen Perspektiven betrach-

tet (zum Beispiel Experten-/Nutzersicht). Nicht umsonst sprechen gestandene Usability-Praktiker wie Dumas & Redish von einer sogenannten Triangulation von Methoden[7]. Beispiel: Eine Online-Befragung im Sinne einer Anforderungs- und Nutzerstrukturanalyse liefert beispielsweise zum einen valide Daten über die tatsächlichen Nutzer einer Website. Zum anderen zeigt sie deren Wünsche & Erwartungen an bestehenden oder neue Inhalte/Funktionen. Die Durchführung einer dem Usability-Test im Labor vorgelagerten Expertenevaluation hilft (als Ergänzung) erste Problembereiche zu identifizieren. Diese können anschließend im Usability-Test mit repräsentativen Nutzern (entsprechend der erhobenen Nutzerdaten rekrutiert) näher betrachtet werden können.

Die persönliche Erfahrung aus Projekten zeigt, dass die meisten Unternehmen in Deutschland die Vorzüge eines solchen systematischen und nutzerzentrierten Entwicklungsprozesses (von der Analyse über die Konzeption und Programmierung bis hin zum Betrieb des Produktes) leider noch nicht vollständig wahrnehmen. Obwohl diese Vorgehensweise, genannt Usability-Engineering, im Bereich der Software Ergonomie bzw. bei der Erstellung klassischer Software bereits seit Jahren praktiziert wird.

1. Definition & Vorteile

Nach Brau und Sarodnick ist *„das Usability-Engineering […] der methodische Weg zur Erzeugung der Eigenschaft Usability. Es ist ein Teilprozess der Entwicklung und Gestaltung technischer Systeme und ergänzt das klassische Engineering, beispielsweise Software-Engineering, um ergonomische Perspektiven. Dabei werden Ansätze, Methoden, Techniken und Aktivitäten für einen benutzerorientierten Entwicklungsprozess bereitgestellt."*[8]

Diese Definition beschreibt die ingenieursmäßige Vorgehensweise, die Benutzungsfreundlichkeit eines Produktes mit bestimmten Maßnahmen und Methoden sicherzustellen, sehr treffend. Ziel ist es, durch Definition von Anwendungsanforderungen, Setzung von messbaren Erfolgskriterien, iteratives Design und Evaluation, Usability als Produkteigenschaft zu objektivieren und messbar zu machen.

[7] Dumas, Joseph S./Redish, Janice C.: A Practical Guide To Usability Testing. Revised Edition, Intellect Books: Exeter, 1999.

[8] Brau, Henning/Sarodnick, Florian: Methoden der Usability Evaluation: wissenschaftliche Grundlagen und praktische Anwendung. Bern: Huber, 2006.

Jakob Nielsen, einer der Pioniere in Sachen Usability, betont dabei, dass Usability-Engineering nicht in Form von Einzelmethoden, sondern in einem übergeordneten Lifecycle durchgeführt werden sollte:

"Usability engineering is not a one-shot affair where the user interface is fixed up before the release of a product. Rather, usability engineering is a set of activities that ideally take place throughout the lifecycle of the product, with significant activities happening at the early stages before the user interface has even been designed." [9]

Sogenanntes Usability-Engineering spart nicht nur Zeit. Es schafft Erfolgssicherheit und Planbarkeit durch die Fokussierung auf den Nutzer. Dies führt zu weniger Diskussionen, schnelleren Entscheidungsprozessen und unter Umständen zu weniger Iterationen innerhalb des Entwicklungsprozesses.

Zudem schützt ein benutzerorientierter Entwicklungsprozess vor hohen Folgekosten. Probleme werden bereits früh aufgedeckt – zu einem Zeitpunkt, an dem sie leichter und billiger zu beheben sind. **Es kann davon ausgegangen werden, dass sich der Zeitaufwand für Änderungen mit jeder Projektphase verdoppelt.** Glaubt man der Aussage von Brinck, Gergle und Wood [Brinck, Gergle, Wood: Usability for the Web: Designing Web Sites that Work, S. 371, 2002, Kaufmann] so steigen die Kosten für eine nötige Änderung pro fortschreitende Projektphase sogar um den Faktor 10 (!)

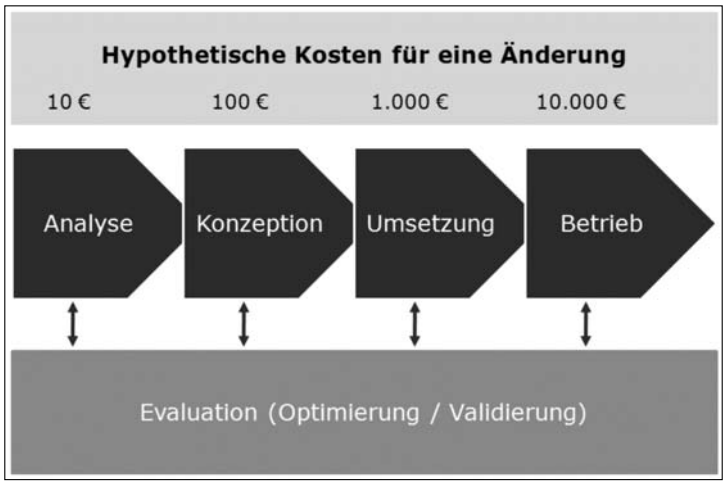


Abbildung 1:
Kosten für nachträgliche Änderungen

[9] Nielsen, Jakob: Usability Engineering. San Diego: Academic Press, 1993.

Zudem werden die Risiken des Imageschadens und des Kunden-/Besucherverlusts minimiert. Infolgedessen wurden sogar allgemeine Charakteristiken, die ein nutzerzentrierter Entwicklungsprozess beinhalten bzw. erfüllen sollte, in der Norm DIN EN ISO 13407 „Benutzer-orientierte Gestaltung interaktiver Systeme“ [10] festgehalten.

2. Der Usability-Engineering-Lifecycle

Grundsätzlich lassen sich folgende vier Entwicklungsphasen eines interaktiven Produktes unterscheiden:

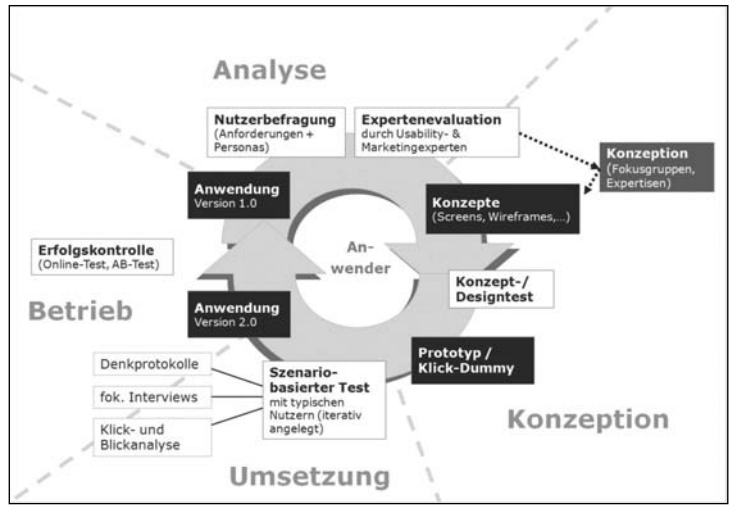


Abbildung 2:
Beispiel für einen typischen, systematischen Re-Design-Prozess

1. (Anforderungs-)Analyse

Die Anforderungsanalyse hilft, die Nutzungsabsichten, Erwartungen und Anforderungen der Zielgruppe(n) zu untersuchen. Anhand der Ergebnisse werden die einzelnen Zielgruppen “greifbarer“. Es lassen sich folglich Schritte zur nutzerzentrierten Konzeption bzw. Optimierung des Produktes ableiten. Beispielsweise eignet sich der KANO-Ansatz [11] hervorragend dazu Kundenbedürfnisse und Anforderungen differenziert zu erforschen und sie gezielt und ressourcengerecht zu erfüllen. Mithilfe des Persona-Ansatzes lassen sich archetypische Nutzer einer Website beschreiben – auf Basis vorhandener Marktforschungsdaten.

[10] DIN EN ISO 13407: Benutzer-orientierte Gestaltung interaktiver Systeme. Berlin: Beuth Verlag GmbH, 1999.

[11] N. Kano: Attractive Quality and Must-be Quality; Journal of the Japanese Society for Quality Control, H. 4, S. 39-48, 1984

So kann sich das Entwicklerteam besser in die (potenziellen) Nutzer hineinzuversetzen und diese Perspektive während des gesamten Designprozesses zu bewahren.

2. Entwicklung/Konzeption

Auf der Basis der erhobenen Anforderungen und Erwartungen erarbeiten Konzepter und Designer erste Entwürfe in Form von Scribbles oder bedingt funktionstüchtiger Prototypen. Noch bevor die endgültigen Spezifikationen für Design und Struktur festgelegt werden, testen repräsentative Nutzer die frühen Konzepte. Somit lassen sich teure Fehlentscheidungen vermeiden. Die Ziele von Usability-Maßnahmen während der Konzeptphase sind klar definierte Interaktionsprozesse, funktionale Strukturen und die stetige Optimierung des Produktes von Beginn an. Hier bieten sich ein sogenanntes Rapid Prototyping an. Beim Rapid Prototyping werden klickbare bzw. funktionsfähige Prototypen mehrfach in kurzen Abständen getestet und kontinuierlich bis hin zur Produktreife optimiert. Das iterative Testen mit nur wenigen Anwendern ermöglicht ein frühzeitiges und effizientes Erkennen und Beheben von Fehlern und Problemen, sodass keine unnötigen Programmierkosten und Projektverzögerungen entstehen. Aber auch Konzept-/Designtest sind ein probates Mittel, um erste Konzepte bzw. Designs abzutesten. Hier geht es jedoch vielmehr um die Gestaltung und Anmutung von Seitenelementen sowie Wording-Fragestellungen als um Interaktionsprozesse wie beim Rapid Prototyping.

3. Umsetzung (Programmierung)

Auch während der Umsetzung sollte kontinuierlich anhand vollständig funktionaler Prototypen getestet werden. Hierbei erfolgt eine ganzheitliche Erfassung von positiven und negativen Erfahrungen beispielsweise in Bezug auf das Interface- und Interaktionsdesign sowie die Navigation und Informationsarchitektur einer Website. Aber auch weiche Faktoren wie Emotionalität, Spaß der Nutzung, Vertrauen und optische Gestaltung werden betrachtet. Lassen sich keine (groben) Probleme im Umgang mit dem Produkt mehr ausmachen, so kann es „produziert" werden. Hierfür eignet sich der klassische szenariobasierte Usability-Test im Labor für die meisten Fragestellungen, da er die Möglichkeit einer „360 Grad-Datenerhebung" bietet. Das heißt, es ist eine Beobachtung und Befragung der Testpersonen – inklusive der Möglichkeit, andere Testmethoden bzw. -verfahren zu integrieren, wie zum Beispiel eine Blickverlaufsmessung (Eyetracking).

4. Betrieb

Der regelmäßige Einsatz von Usability-Maßnahmen im Live-Betrieb dient der Erfolgskontrolle und der kontinuierlichen Qualitätssicherung, um den sich permanent wechselnden Nutzerbedürfnissen Rechnung zu tragen und die Nutzer zufrieden zu stellen. Die Tracking-

daten einer Website liefern beispielsweise wiederum direkte erste Hinweise auf potenzielle Problembereiche einer Website. Identifiziert man während des Check-Out-Prozesses eines Online-Shops eine erhöhte Abbruchrate, kann dies der Anknüpfungspunkt für eine systematische Beleuchtung des Problems durch die entsprechenden Usability- und Marktforschungsdienstleistungen sein, zum Beispiel eine Online-Befragung oder das Aufsetzen eines Erfolgs-Misserfolgsmonitors. Bei letzterem wird der Nutzer beim Aufruf des Online-Shops nach seiner Intention/seinem Ziel befragt. Anschließend kann er wie gewohnt den Shop benutzen. Hat er sein Ziel erreicht bzw. möchte er den Shop aus einem anderen Anlass verlassen, wird er in Bezug auf die Erfüllung/Nicht-Erfüllung seines Ziels befragt.

3. Wann welche Methode?

Nicht jede Usability-Methode ist zu jedem Zeitpunkt bzw. für jede Fragestellung geeignet. Wann welche Methode zum Einsatz kommen sollte, hängt von verschiedenen Faktoren ab. Dies richtet sich zum einen nach der Entwicklungsphase in der man sich aktuell befindet. Zum anderen hängt die Auswahl der Methoden & Verfahren von den folgenden Kriterien bzw. Fragen ab:

- Was für Fragestellungen sollen beantwortet werden? Das heißt sind es eher Fragen in Bezug auf das Design, die Funktionalität, die Struktur (Informationsarchitektur) oder den Inhalt? Geht es eher um das Erheben von Anforderungen & Erwartungen oder um das „Abtesten“ von Interaktionsprozessen?
- In welcher Form liegt die zu testende Applikation bzw. Website vor? Als reines niedergeschriebenes Konzept, in Form schematischer Wireframes, grafisch ausgestaltete Screens oder gar als interaktiver Klick-Dummy?
- Welche Ressourcen und welcher Zeitrahmen stehen zur Verfügung? Welche Vorarbeit kann durch Sie geliefert werden? Wie hoch ist das gesamte Projektbudget? (Anmerkung: In Bezug auf Usability-Maßnahmen ist Durchschnitt mit 10 Prozent des Gesamtprojektvolumens zu planen.)
- Welche Zielgruppe soll erreicht bzw. zum Testen herangezogen werden? Sind es eher Privatpersonen oder vielbeschäftigte Manager, die man nur mit sehr anspruchsvollen Incentivierungskonzepten für eine Studie gewinnen kann. Auch räumliche und zeitliche Aspekte sind hierbei gegebenenfalls zu berücksichtigen.

Je nach Aufgabenstellung und Entwicklungsphase sind die geeigneten Methoden auszuwählen und an die konkreten Projektbedürfnisse und Fragestellungen anzupassen – da keine Website/-applikation der anderen gleicht. Gerade hieran erkennt man die Qualität einer Beratungsdienstleistung durch eine Agentur im Vorfeld an den Usability-Test oder eine Online-Befragung an sich. Denn: Eine Methode/ein Verfahren kann noch so toll sein, wenn man das falsche misst, dann können die Ergebnisse noch so schön aufbereitet sein. Sie sind schlicht und einfach wertlos.

4. Fazit

Im heutigen Internetzeitalter bedarf es einer kontinuierlichen Begleitung bzw. systematischer Unterstützung jedes Internet-Projektes durch Online-Marktforschung bzw. Usability-Maßnahmen in allen Entwicklungsphasen.

Besonders wichtig ist dabei die Analyse bestehender Nutzerstrukturen und/oder der Anforderungen & Erwartungen (potenzieller) Zielgruppen. Dies kann im Rahmen von Online-Befragungen – entweder über ein Panel oder direkt, falls vorhanden, auf der bestehenden Website (onsite) – erfolgen.

Nur so ist eine nachfolgende nutzerorientierte Konzeption neuartiger Interaktionselemente und -prozesse überhaupt möglich.

Durch Konzept-& Designtest bzw. Prototypentests können in der Konzeptionsphase erste Probleme bereits frühzeitig erkannt und behoben werden – noch bevor aufwendige vollfunktionsfähige Beta-Versionen erstellt wurden.

Während der Umsetzung ist immer noch genügend Zeit für abschließende bzw. ganzheitliche Nutzertests bzw. User Experience Tests. Hier gilt es den letzten Feinschliff vorzunehmen.

Im Live-Betrieb gewährleisten multivariate Analyse-Verfahren oder Remote Testing-Tools eine kontinuierliche Qualitätskontrolle und liefern somit Aufschluss über die tatsächliche Nutzung der Website.

III. Parameter des Erfolgs: Die Festlegung der richtigen Kennzahlen

von Frank Reese

1. The Name of The Game: Wozu gibt es diese Website? 37
2. Etappen, Brennpunkte und Call-to-Action 37
3. Performance-Messung 39
4. Einzelne Kennzahlen 43

Sicher kennen Sie das: Sie und ein Bekannter sind jeweils davon überzeugt, dass der andere im Unrecht ist, dass der Fußballspieler XY in der Bundesliga-Saison 1982/83 ganz bestimmt nicht für den Verein Z gespielt hat. Sie wetten um einen Kasten Bier. Anschließend stellt sich heraus, dass der Spieler tatsächlich an keinem Pflichtspiel in der Saison teilgenommen hat, dafür aber an einem Freundschaftspiel. Was für ein ärgerliches Ergebnis. Wer hat nun Recht? Wer kann sich über das Bier freuen?

Bei der Interpretation der Ergebnisse von Tests kann es einem genauso ergehen. Eigentlich erscheint die Fragestellung überaus klar: Generiert die Formulierung „Halber Preis" mehr Klicks als die Formulierung „2 zum Preis von 1"? Aber der Teufel steckt im Detail. Vielleicht liegt ja die Klickrate bei Alternative A niedriger als in B, dafür ist aber die Konversionsrate höher? Oder das Bestellvolumen? Oder die Stornoquote geringer? Die Ziele der Optimierung müssen vor der Durchführung des Tests definiert werden. Wer erst nach dem Test beginnt die Ziele zu diskutieren, hat ein Problem.

Aber so wie Website und Kampagnen bei kontinuierlichen Verbesserungsbemühungen immer effektiver werden, so steigt auch das Know-How der Mitarbeiter hinsichtlich der Durchführung von Tests. Es ist bisher weder eine perfekte Website vom Himmel gefallen, noch ein Meister des Testens. Der Zugewinn an Wissen darüber, wie Tests erfolgreich durchgeführt werden können, wird anfangs fast ebenso groß sein, wie der Effektivitätszuwachs der Website.

Zur Bedeutung und Festlegung von Kennzahlen (auch Kennziffern, KPI für *Key Performance Indicator*) gibt es inzwischen eine Reihe interessanter Publikationen, die eine ausführlichere Darstellung bieten, als im Rahmen dieses Buches sinnvoll ist. Empfehlenswert sind vor allem „The Big Book of KPI" von Eric Peterson und „Web Analytics – Damit aus Traffic Umsatz wird" [12] vom Autor.

Vor dem praktischen Angehen des Tests ist es aber wichtig, möglichst konkret und genau festzulegen, was unter welchen Bedingungen als Erfolg bzw. Misserfolg gelten muss. Als erster Schritt ist zunächst die Ausgangslage zu klären:

[12] Eric T. Peterson: The Big Book of KPI, 2006, beziehbar als pdf über die Homepage des Autors: www.webanalyticsdemystified.com; Frank Reese: Web Analytics – Damit aus Traffic Umsatz wird, Business Village 2008

1. The Name of The Game: Wozu gibt es diese Website?

Mit „The Name of the Game“ wird die Frage nach dem Wozu des ganzen Projektes gestellt. Um was geht es eigentlich? Was ist das oberste Ziel der Website? Sollen Bestellungen in einem Online-Shop erzielt werden? Geht es um Lead-Generierung im BtoB-Bereich, also die klassische Gewinnung von Kundenkontakten über die Website, die dann in direkten Kundengesprächen ausgewertet werden können? Oder liegt die Generierung und Steigerung von AdViews, also der Verkauf und die Darstellung von Werbemitteln im Zentrum der Website? Es gibt tausend Ziele, zu deren Erreichung eine Website im Netz beitragen könnte – jedenfalls ist davon auszugehen, dass jede professionelle Website ein deutlich formulierbares Ziel hat. Websites, für keine solchen Ziele formuliert werden können, sind dann auch nicht die Investition wert, die ihre Erstellung und Pflege darstellt. Vor jedem Ansatz zur Optimierung der Website muss dieses Ziel unter den Verantwortlichen vereinbart und definiert worden sein.

2. Etappen, Brennpunkte und Call-to-Action

Ausgehend von diesem Hauptziel lassen sich Unterziele und Etappen auf dem Weg zum Hauptziel herausfinden. Diese Etappen bilden das Rückgrat der Website. Mit ihrer Struktur lassen sich jene Schritte darstellen, die ein Besucher idealerweise absolviert. Jede einzelne Page der Website ist mehr oder weniger Teil einer der Etappen. Bei einer Page, die nicht einer der Etappen zugeordnet werden kann, ist entschlossen zu fragen, ob sie überhaupt online stehen muss.

Ein Optimierungs-Test kann sich aber nicht auf Abläufe insgesamt oder gar die komplette Website beziehen. Ein erfolgreicher Test kann nur sehr klar voneinander unterscheidbare Alternativen gegeneinanderstellen und helfen, die bessere Version herauszufinden. Das ist der Grund, warum der Vergleich einer ehemaligen und einer neuen Version der Website, zum Beispiel nach einem Relaunch, keinen harten Vorher/Nachher-Vergleich erlaubt: Die Veränderungen sind so zahlreich und grundlegend, dass die Effekte einzelner Elemente nicht herausgezogen werden können.

Der zweite Schritt liegt in der Festlegung einzelner, möglichst entscheidender Wegmarken auf der Website: eine Landing Page, ein Registrierungsformular, eine Startseite, eine Produkt-Detail-Seite. Den Meisten wird eine derartige Festlegung auf die entscheidenden Seiten nicht schwer fallen – viele Business-Modelle und Websites haben naturgemäß sehr

deutliche Brennpunkte, an denen sich zeigt, ob ein Besucher zum Kunden im Sinne des Betreibers wird. Diese einzelnen Seiten tragen auf eine ganze bestimmte Weise zur Erreichung des Website-Zieles bei. Auf diesen Seiten gibt es Elemente, die dafür gedacht sind, vom Besucher genutzt zu werden: Ein Button zum Download, eine Schaltfläche zum In-den-Warenkorb-legen, ein Link auf besondere Neuheiten – was immer. Wenn der Weg vom primären Website-Ziel über die einzelnen Etappen und Bereiche bis hin zu den einzelnen Seiten gemacht ist, fallen die entscheidenden Elemente auf der Seite sofort ins Auge. Diese sind es, die Bryan Eisenberg nicht zuletzt in seinem Buch „Call to Action“[13], in das Zentrum jeder Betrachtung, Evaluierung und Optimierung von Websites stellt.

Eisenberg und seine Mitarbeiter haben damit sehr erfolgreich vielen großen amerikanischen Unternehmen bei der Verbesserung ihrer Online-Angebote geholfen. Die Idee des Call-to-Action kann nicht radikal genug in den Fokus der Gestaltung und Messung einer Webpage gestellt werden. Der Platz auf Websites, in E-Mails oder sonstigen Online-Angeboten mag kostenlos und unbegrenzt erscheinen – er ist es aber nicht. Jeder einzelne Link auf einer Startseite ist eine Option für den Besucher. Und Besucher können nur einmal klicken: Wenn sie sich für jenen Link entscheiden, können sie nicht gleichzeitig dort klicken. Mal abgesehen von Positions- und Gestaltungseffekten: Je mehr Links es auf einer Page gibt, desto geringer die Klick-Wahrscheinlichkeit für jeden einzelnen Link. Betreiber von Websites müssen sich also bei jedem Link auf der Site fragen: Will ich, dass der Nutzer hier klickt? Und was soll passieren, nachdem er dies getan hat? Was ist mein Gewinn dabei? Mit welcher Aktion und den vielen möglichen Optionen erziele ich den höchsten Gewinn? Und wie erreiche ich, dass möglichst viele der Besucher der Site genau diese Aktion ausführen? Die radikale Fokussierung auf diese entscheidenden, wertvollen Punkte sowohl auf einer Website, als auch in einem Ablauf und auf einer einzelnen Webpage ist unerlässlich, um Tests nicht in Beliebigkeit versinken zu lassen, sondern um den Aufwand genau auf den richtigen Punkt zu bringen und den größten Effektivitätszuwachs einzufahren.

Dieser Punkt im Fokus der Optimierung bleibt dabei nicht statisch. Wenn Sie mit einer kleinen Veränderung auf einer Formularseite die Rate der Newsletterbesteller um 30 Prozent erhöht haben, bedeutet dies nicht, dass Sie mit gleichem Aufwand auf der gleichen Seite wieder den gleichen Effektivitäts-zuwachs einfahren können.

[13] Bryan & Jeffrey Eisenberg: Call to Action. Wiley 2004

Denken wir uns also ein dynamisches Verhältnis zwischen einem festgelegten Optimierungsbudget, den Ansatzpunkten der Optimierung und möglichen Effektivitätszuwächsen. Dieses Verhältnis kann auf die verschiedenen, wichtigen Bereiche einer Website angewendet werden. Nachdem mit den ersten 5.000 Euro (als Optimierungsbudget) eine 30-prozentige Steigerung der Newsletterbesteller erreicht wurde, verspricht eine erneute Veränderung der gleichen Page einen geringeren Effekt. Sie bekommen für das Budget also weniger zurück, als beim ersten Mal. Durch die Verbesserung der Formularseite kann es also jetzt sinnvoller sein, erst einmal einen Punkt auf der Startseite zu verbessern. Nach und nach wandert der Punkt mit dem besten Verhältnis aus Einsatz und Gewinn also durch die verschiedenen Bereiche der Site und diese wird nach und nach insgesamt immer effektiver.

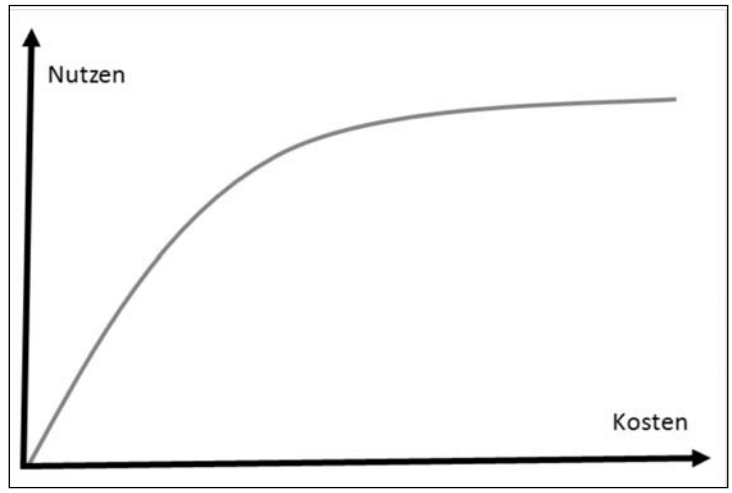


Abbildung 3:
Der Nutzen fortgesetzter Optimierungen der gleichen Page nimmt ab, aber die Kosten nicht.

Die Kurve zwischen Kosten und Effektivitätszuwachs verläuft exponentiell: Je öfter eine Page bereits optimiert wurde, desto geringer ist (bei gleichbleibendem Einsatz) der jeweilige Gewinn der Optimierung. Darum ist es sinnvoll, immer dort anzusetzen, wo aktuell der größte zu erwartende Gewinn wartet.

3. Performance-Messung

Es erscheint offensichtlich: Wer eine Verbesserung erreichen will, muss vorher wissen, wo er steht, um nach den Änderungen einen messbaren Unterschied ausmachen zu können. Wenn die obigen Fragen beantwortet sind, also die zentralen Performance-Punkte einer Website festgelegt wurden, sollte es leicht sein, Kennzahlen zu ermitteln, und somit die

Effektivität dieser Seiten und Seitenelemente zu messen. Den einfachsten Fall bilden wohl Landing Pages: Da diese allein dem Zweck dienen, den Traffic von externen Quelle, zum Beispiel einer Kampagne, auf der Website zur gewünschten Stelle zu lenken, zeigt sich die Qualität einer Landing Page gewöhnlich in der Klickrate zum Beispiel auf die Produktdetailseite. Da dies das einzige Ziel der Landing Page ist, muss jede Aktion, die nicht zum Aufruf der entsprechenden Detailseite führt, als ein Misserfolg gelten. Die Idee zur Berechnung ist also schlicht: Wie groß ist der Anteil der Besucher, die nach dem Aufruf der Landing Page auf den entsprechenden Link klicken? Leider ist die Ermittlung der Performance nicht immer so einfach. Viele Faktoren können dafür sorgen, dass bestimmte Gestaltungen zwar zunächst hohe Klickraten erzeugen, anschließend aber zu enttäuschenden Konversionsraten, Bestellvolumen oder Stornoquoten führen. Darum sollte der Festlegung der Kennzahlen für einen Test eine gründliche Analyse des Website-Traffics und der Besuchergruppen vorausgehen. Welche Umsätze, Registrierungen oder Kontakte werden von welcher Gruppe von Besuchern generiert? Gibt es Unterschiede nach der Art des Zugriffsgerätes (PC versus Mobilgerät), zwischen Stamm- und Erstbesuchern, bestehenden Kunden und Neukunden sowie nach verschiedenen Kampagnen, Produkt- oder Inhaltsinteressen? Ein solides Wissen über die aktuellen Eigenschaften und Bewegungen von Besuchern und Besuchergruppen ist die Grundlage, um für bestimmte Gruppen einzelne Site-Elemente erfolgreich zu verändern. So wird etwa ausgeschlossen, dass eine bestimmte Landing Page zwar besser konvertiert als eine andere, die erstere aber von weniger interessanten Besuchern aufgerufen wird und am Ende nichts zum Ziel der Website beiträgt.

Von den Hauptzielen eines Online-Projektes ausgehend lässt sich eine Hierarchie von Kennzahlen festlegen, die jeweils einen Aspekt hinsichtlich der Zielerreichung abbilden und so insgesamt zu einer Steigerung beitragen können. Einfügen lassen sich zudem Module für die Messung von Kampagnen, für Bestellprozesse, für interne Suchmaschinen oder ähnliches. Wenn notwendig, können Kennzahlen aufgesplittet und für verschiedene Zielgruppen einzeln dargestellt werden.

Eine solche Hierarchie bildet den Einfluss einzelner eher operativer Elemente – in der Abbildung zum Beispiel Kampagnen oder Bestellprozesse – auf die Gesamtperformance der Website ab. Mit einer Zunahme der Effektivität der Kampagnen steigt auch das Gesamtergebnis[14].

[14] Zu Kennzahlen und Hierarchien siehe ausführlich: Frank Reese: „Web Analytics – Damit aus Traffic Umsatz wird", Business Village 2008

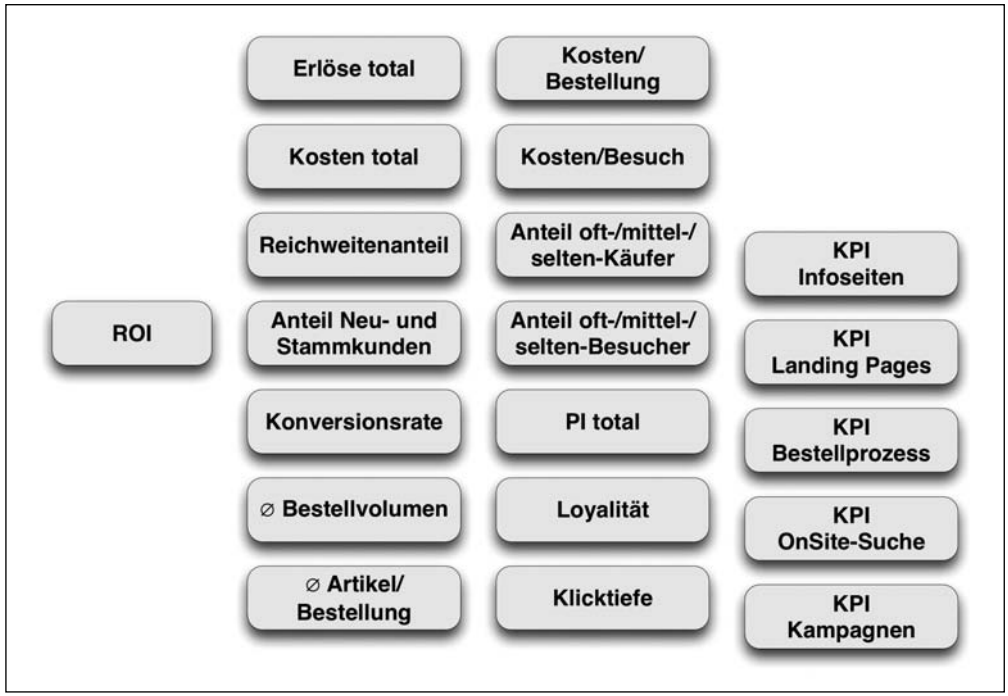


Abbildung 4: Kennzahlen-Hierarchie für einen Online-Shop

Schwierig ist manchmal die Balance zwischen gleichrangigen Zielen auf der obersten Ebene. So kann die Startseite eines BtoB-Unternehmens zwar weitgehend darauf ausgelegt sein, Kontakte zu Neukunden zu generieren, allerdings darf dabei nicht die Vermittlung eines professionellen Images vernachlässigt werden. Wäre dem nicht so, könnte auf der Startseite einfach ein großer „Rufen Sie uns an!" Button stehen. Das würde vielleicht mehr Kontaktaufnahmen erzeugen, aber bei vielen Besuchern ein unseriöses Bild des Unternehmens erzeugen.

Abbildung 5:

Spam-Seiten konzentrieren sich dagegen ganz auf einen bestimmten Zweck. Im abgebildeten Beispiel gibt es keinen Link außer im „Hier eintragen"-Feld

Dazu kommt die notwendige Bedienung unterschiedlicher Zielgruppen: Presse, potenzielle Bewerber, bestehende Kunden und andere Die Ziele gehen hier nicht einfach ineinander auf; im Gegenteil: Wird der Pressebereich auf der Startseite stärker betont, kann dies auf Kosten der Kontaktgewinnung geschehen. Besonders Startseiten sind darum immer Kompromisse zwischen den verschiedenen Interessen, die ein Unternehmen mit seinem Online-Angebot verfolgt.

Eine Diskussion wie die folgende ist also leicht vorstellbar

A: „Durch die Umgestaltung der Startseite konnten wir letzten Monat 20 Prozent mehr Kundenkontakte über die Website verzeichnen."

B: „Allerdings sind dafür die Abrufe im Pressebereich deutlich gesunken. Es ist für unser Unternehmen sehr wichtig, in der Öffentlichkeit präsent zu bleiben."

C: „Auch die Seiten zu den offenen Stellen sind weniger nachgefragt worden."

B: „Und wie sieht es mit der Qualität der neuen Kontakte aus? Haben die Besucher, die wirklich ernsthaft Interesse an unseren Produkten hatten, nicht schon vorher den Weg zum Kontaktformular gefunden?"

Eine Diskussion wie die folgende ist also leicht vorstellbar (Fortsetzung)

A: „Leider sind aus dem Vertrieb bisher keine Rückmeldungen zur Qualität der Kundenkontakte gekommen."

D: „Ich glaube auch, dass unsere Website den Besuchern ein ausgewogenes Informationsangebot bieten muss. Wir sollten zu der alten Version zurückkehren."

Damit wäre der Test mit der neuen Version der Startseite umsonst gewesen. Dieses Ergebnis ist vorhersehbar, wenn es keine klaren Prioritäten zwischen den Zielen gibt. Die Gewichtung lässt sich am besten mit Geld- oder Punktwerten für einzelne Interaktionen herstellen. Wenn vorher gemeinsam vereinbart wurde, dass zum Beispiel der Abruf einer Pressemeldung 5 Punkte wert ist, ein PDF-Download 20 Punkte, eine Bewerbung 50 Punkte, eine Kontaktanfrage 100 Punkte usw., so kann die Performance der gesamten Site in diesen Punktwerten abgebildet werden. Dabei müssen die Punkte nicht realen Werten entsprechen, sondern vor allem im Verhältnis zueinander stimmen. Es mag zunächst eine gewisse Mühe kosten, so ein System zu erstellen, eine effektive Performance-Messung und -Steigerung wird dadurch allerdings ungeheuer erleichtert.

4. Einzelne Kennzahlen

Die Pyramide zur Messung von Kampagnen

Das Performance Marketing – die Aussteuerung und Bezahlung insbesondere von Pay-Per-Click- und Affiliate-Kampagnen entsprechend ihres Erfolges, also ihrer „Performance" – hat einen großen Beitrag zur Messung, Verbreitung und Etablierung harter Erfolgsparameter zur Bewertung von Kampagnen geleistet. Der Erfolg einer Kampagne kann dabei Schritt für Schritt von oben nach unten abgebildet werden (siehe Abbildung 6 auf der folgende Seite).

Ganz vorne steht natürlich die Anzahl der Gelegenheiten, bei denen ein Werbemittel überhaupt die Chance hatte gesehen zu werden. Dies kann die Anzeige in Folge einer Keyword-Eingabe in einer Suchmaschine sein, der Aufruf eines bestimmten Themas oder einer Seite mit entsprechenden Anzeigen (zum Beispiel AdSense), oder auch eine geöffnete E-Mail. Jedes Mal besteht die theoretische Chance, dass ein Besucher die Anzeige sieht und – das ist das Ziel – darauf klickt.

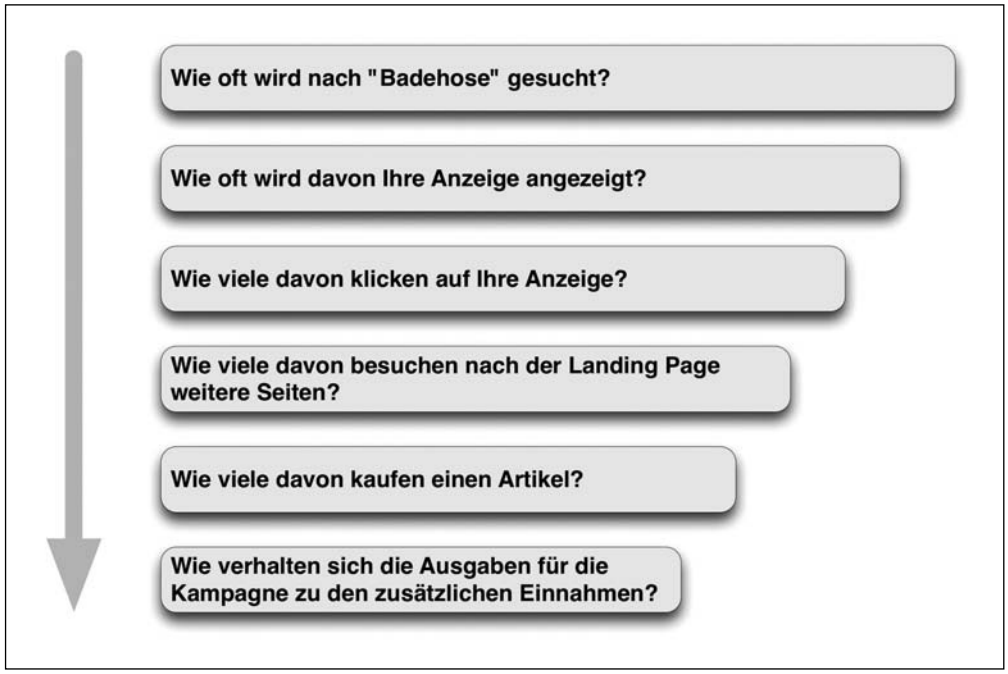


Abbildung 6: Kennzahlen zur Messung von Kampagnen

Dieser Klick ist also nach der reinen Anzahl der AdViews die nächste wichtige Zahl. Oft auch CTR (Click-Through-Rate) genannt. Die CTR errechnet sich aus dem Verhältnis der Anzahl der Anzeigen zu den Klicks:

$$\text{CTR} = \frac{\text{Klicks}}{\text{Anzahl der Anzeigen}} \times 100$$

Die Anzahl der Klicks kann am einfachsten über die Abrufe der Zielseite von dieser Herkunft aus ermittelt werden. Üblicherweise werden die Links in Anzeigen mit einem URL-Parameter ausgestattet (zum Beispiel *www.meine-landingpage.html?meineanzeige*), um die Herkunft eindeutig zuordnen zu können.

Wenn der Besucher die Landing Page aufgerufen hat, ist die erste große Hürde genommen. Jetzt geht es darum, die Erwartung des Besuchers nicht zu enttäuschen und ihn auf die Produktdetailseite, eine Informationsseite oder ein Registrierungsformular weiterzuleiten. Entsprechend kann die Effektivität der Landing Page also aus dem Verhältnis zwischen der Anzahl der Aufrufe und der Anzahl der Klicks auf die nächste Seite errechnet werden. Die

Kalkulation läuft dabei immer nach dem gleichen Schema wie in der zuerst genannten Formel ab: Anzahl der Klicks auf die weiterführende Seite geteilt durch die totale Anzahl der Aufrufe der Landing Page multipliziert mit 100.

Was auch immer der Besucher nun im Einzelnen auf der Website treibt, letztlich soll eine bestimmte Aktion – Bestellung, Kontaktaufnahme oder Ähnliches – erreicht werden. Diese Zuordnung kann innerhalb eines Web-Analytics-Tools über eine Segmentierung bzw. einen Filter erreicht werden: „Zeige nur Bestellungen von Käufern an, die mindestens einmal diese bestimmte Landing Page aufgerufen haben". Allerdings setzt dies voraus, dass der Besucher direkt während seines ersten Besuches den Kaufabschluss vollzieht. Da dies ist aber eher selten der Fall ist, werden Cookies genutzt, um Besucher, die einmal diese Landing Page oder bereits das Werbemittel gesehen haben, wiederzuerkennen.

Mit diesen Mitteln lässt sich nun insgesamt die Effektivität von einzelnen Anzeigen und Kampagnen ermitteln. Die Anzeigen und Werbeträger können nicht nur hinsichtlich eines Branding-Effekts (Anzeige wurde von x Millionen Besuchern gesehen), sondern vor allem auch hinsichtlich Bestellungen und damit dem ROI (Return on Investment) bewertet werden. Damit ergibt sich eine belastbare Grundlage, um für jeden einzelnen Schritt vom Werbemittel bis zur Bestellung Tests und Optimierungsmaßnahmen durchzuführen.

Absprungrate

Die Absprungrate (auch: Bounce Rate, Abprallrate) ist eine sehr einfache Zahl, aber sie gibt Ihnen einen klaren Hinweis auf die Effektivität Ihrer Seite. Stellen Sie sich vor, Sie betreiben ein Spielzeuggeschäft für Kinder. Sie stehen hinter dem Verkaufstresen und sehen, wie die Kinder den Laden betreten. Die Kinder gehen in die Mitte des Geschäfts, schauen sich einmal rundherum um, sehen Brettspiele, Sportgeräte, Modellbausätze, Spielfiguren. Nach diesem Rundblick dreht sich jedes dritte Kind um und verlässt den Laden direkt – ohne einen weiteren Blick, etwas anzufassen oder eine Frage zu stellen. Wenn Sie ein wenig Zeit und Interesse hätten, können Sie einem der Kinder nachgehen. Vielleicht besucht es anschließend ein weiteres Geschäft und sucht wieder, vielleicht sehen Sie, was in Ihrem Sortiment fehlt, vielleicht erkennen Sie, was ein Drittel Ihrer Ladenbesucher vermisst.

Die Absprungrate zeigt Ihnen genau das für Ihre Website. Automatisch und präzise und ohne dass Sie den ganzen Tag Kinder zählen müssen[15].

[15] Sie müssen nicht einmal den Kindern nachgehen, um zu wissen, was diese suchen: In den Traffic-Analysen wird direkt ausgewiesen, was die Besucher der Website bei den Suchmaschinen als Keyword eingegeben haben.

$$\text{Absprungrate} = \frac{\text{Besucher mit der Seite als einzigen Seitenabruf}}{\text{Besuche mit dieser Seite als Einstiegsseite}} \times 100$$

Die Absprungrate verrät Ihnen den Anteil der Website-Besucher, die eine bestimmte Seite aufrufen, aber anschließend keine weitere mehr. Die Seite, für die die Absprungarte ermittelt wird ist also die erste Seite, die ein Besucher des Online-Angebotes aufruft und gleichzeitig die letzte.

Viele Traffic-Analyse-Systeme (auch Google Analytics) berechnen zudem eine durchschnittliche Absprungrate für die gesamte Website. Dies mag gelegentlich ein interessanter Hinweis sein, führt aber weder zu einem brauchbaren Erfolgsparameter noch zu Optimierungshinweisen.

Soll eine Landing Page, Startseite oder eine andere Seite, auf der die Besucher in Ihre Website einsteigen, optimiert werden, so ist die Absprungrate eine der einfachsten und wichtigsten Kennzahlen. Mit einer Variation von Inhalten, Gestaltungselementen und Positionierungen kann die Absprungrate vermindert und damit die Effektivität der Site erhöht werden. Und, um noch einmal das Kinderbeispiel aufzugreifen: Sie können zwar nicht wissen, wohin die Besucher als nächstes gehen, dafür wissen Sie aber genau, woher sie kommen: Der Inhalt der Herkunftsseite, ob Suchmaschine (dazu mit einem bestimmten Suchwort) oder eine Seite mit speziellen Inhalten, verrät Ihnen, was die Besucher eigentlich gesucht haben.

Konversionsraten

Die Konversionsrate ist wohl die beliebteste und – im engeren Rahmen der Trafficanalyse – die wichtigste Kennzahl zur Performance von Websites. Am einfachsten funktioniert diese bei Online-Shops:

$$\text{Konversionsrate} = \frac{\text{Besucher mit mindestens einem Bestellabschluss}}{\text{Anzahl der Besucher}} \times 100$$

Mit dieser Formel wird das Verhältnis zwischen der Anzahl aller Besucher Ihrer Seite zu jenen Besuchern deutlich, die etwas bestellen. Wenn Sie also 1000 Besucher pro Monat auf Ihrer Site haben und davon zehn etwas bestellen, ergibt sich eine Konversionsrate von 1 Prozent. 1 Prozent ist natürlich nicht viel, oft wird eine mittlere Zahl von 2 Prozent für Online-Shops genannt. Viele erreichen aber höhere, manche sogar zweistellige Konversionsraten. Was als gut gelten kann, ist sehr abhängig von den konkreten Produkten.

In diesem Fall wird die Konversionsrate mit Besuchern (nicht Besuche) berechnet. Besucher können einmal oder mehrmals eine Website besuchen, ihre Bestellung kann bereits beim ersten Besuch, aber auch erst beim zehnten Besuch erfolgen.

Besucher sind nur ermittelbar über Cookies

Cookies sorgen dafür, einen Besucher wiederzuerkennen, also Besuch 1 und Besuch 2 einem einzelnen Besucher zuzurechnen. Cookies sind daher unbedingt notwendig zur Ermittlung der Besucherzahl, allerdings sind die Zahlen ein wenig wackelig. Cookie-Blockaden und mehr noch Cookie-Löschungen führen zu verzerrten Besucherzahlen. Wird ein Cookie gelöscht und der Besucher ruft erneut die Website auf, so wird er als ein neuer Besucher gezählt, obwohl hier eigentlich ein Wiederholungsbesuch vorliegt. Pragmatischerweise kann mit einer Cookie-Lebenszeit von vier Wochen gerechnet werden. Innerhalb dieses Zeitraums können die Daten noch als belastbar angesehen werden, darüberhinaus sind sie mit deutlicher Vorsicht zu genießen.

Wegen der Fehleranfälligkeit der Besucherberechnung wird manchmal empfohlen, nicht Besucher, sondern Besuche zur Berechnung der Konversionsrate zu nutzen. Besuche (auch: Visits, Sessions) sind technisch eindeutig erfassbar und damit zuverlässiger. Allerdings wird hierdurch nicht nur die Konversionsrate vermindert (es gibt mehr Besuche als Besucher, aber nicht notwendig mehr Bestellungen) sondern auch ein Teil der Realität ausgeblendet. Einige Websites sind so angelegt, dass der Besucher direkt beim ersten Besuch etwas bestellt, bei den meisten Produkten ist es aber vernünftiger davon auszugehen, dass der Besucher erst beim zweiten, dritten oder vierten Besuch eine Bestellung abschließt.

Die geschilderte Konversionsrate bezieht sich auf die Berechnung für einen Online-Shop, die Kennzahl kann aber für fast alle Website-Ziele berechnet werden: Registrierungen von Newsletter-Empfängern, Kundenkontakte, Downloads – alle wichtigen Ziele einer Website können als Konversionen (Umwandlungen) betrachtet und berechnet werden. Dabei kann eine Website durchaus mehr als ein Ziel haben und für jedes eine Konversionsrate berechnen.

Für konkrete Optimierungen ist die Festlegung und Berechnung eine primären Zieles (siehe oben) ein wichtiger Ausgangspunkt und Rahmen, in dem dann aber kleinere Schritte gegangen und berechnet werden müssen. Mikrokonversionen bezeichnen darum das erfolgreiche Absolvieren einzelner Stufen im Konversionsprozess.

Als **Mikrokonversionen** können zum Beispiel die folgenden Schritte auf dem Weg zur Bestellung gesehen werden:

- Eingabe eines Suchbegriffs in der site-internen Suchmaschine,
- Klick auf eines der Suchergebnisse,
- Aufruf einer Detailseite,
- Ablegen des Produktes im Warenkorb,
- Eingabe von Adress- und Zahldaten,
- Bestätigung der Bestellung.

Jede einzelne dieser Aktionen kann als Ziel beim Aufruf der entsprechenden Seite festgelegt werden. Auf einer Seite mit einem Formular – sei es für Adressdaten, für eine Newsletterbestellung oder eine Kontaktanfrage – besteht das Ziel und die Mikrokonversion immer im Absenden eines vollständig ausgefüllten Formulars. Dies ist das einzige wesentliche Ziel dieser Seite. Alle Aktionen der Besucher jenseits dieses Zieles müssen als potenziell vermeidbar gesehen werden. Die Erhöhung dieser Mikrokonversionsrate ist der Erfolgsparameter für diese Seite.

Nicht immer ist es allerdings so einfach: Eine Produktdetailseite eröffnet dem Besucher aus guten Gründen viele Möglichkeiten zur Interaktion. Er kann das Produkt im Warenkorb ablegen, er könnte sich aber auch weitere Details anschauen, vielleicht den Artikel auf eine Wunschliste setzen oder auf Hinweise zu ähnlichen Artikeln klicken. Alle diese Möglichkeiten konkurrieren um die Aufmerksamkeit des Besuchers, sie müssen für eine Optimierung darum in eine klare und funktionierende Reihenfolge der Prioritäten aus Sicht des Betreibers gebracht werden.

Was tun Nutzer auf einer Produktdetailseite?

Nutzeraktion	Aktuelle Klickhäufigkeit
Produkt in Warenkorb	10 %
Download Produktbroschüre	12 %
Speicherung auf Wunschliste	2 %
Setzung eines Lesezeichens	3 %
Klick auf ‚ähnliches' Produkt	10 %
sonstige Aktionen	25 %
Absprünge	38 %

Das zentrale Ziel der Optimierung ist natürlich die Erhöhung der Produkte im Warenkorb, aber es muss berücksichtigt werden, dass viele Besucher auf keinen Fall das Produkt jetzt in den Warenkorb legen werden. Vielleicht später, aber jetzt sind sie noch nicht entschlossen oder überzeugt genug. Für diese Gruppe von Besuchern stehen Optionen wie Info-Downloads, Wunschlisten und Ähnliches zur Verfügung.

Das Reservoir zur Optimierung der Detailseite liegt damit bei jenen Besuchern, die die Seite ganz verlassen oder Links wählen, die nicht wichtig für die Performance dieser Page sind (Lieferbedingungen, Impressum, Startseite …). Mögliche Ziele von Optimierungen wäre darum zum Beispiel die Verminderung der Absprungrate, die Verlagerung von Klicks auf unwichtige Seiten hin zu Klicks auf der Prioritätenliste oder die Verschiebungen innerhalb der Prioritätenliste nach oben.

Um die konkreten Elemente, die bei der Gestaltung einer Seite im Sinne einer solchen Prioritätenliste in Frage kommen, soll es in Kapitel IV gehen. Hier bleibt festzuhalten, dass bei einer Optimierung der Seitengestaltung die Interaktionen der Effekte mitzudenken sind. Durch die prominentere Darstellung des ‚In-den-Warenkorb'-Buttons werde ich wahrscheinlich eine Erhöhung der entsprechenden Zahlen bewirken, muss aber einerseits im Auge behalten, wie sich die Konversionsrate insgesamt verändert (Häufen sich die Abbruchquoten im Bestellprozess?) und andererseits aus welchen Quellen die zusätzlichen Warenkorb-Klicks kommen, welche anderen Klickhäufigkeiten in der Prioritätenliste also zurückgehen.

Für eine langfristige Effektivitätssteigerung ist ein gut durchdachtes Modell für potenziellen Handlungen der Besucher hilfreich: Was kann durch Optimierungen erreicht werden? Welche Verluste sind unvermeidbar? Welche Erfolge werden durch Wunschlisten, Lesezeichen und Ähnliches erzielt? Was bringen die Klicks auf „ähnliche" Artikel? Wie verhalten sich die erwartbaren Gewinne durch Optimierungen zu den notwendigen Kosten?

Suchbewegungen

Die Nutzung von internen Suchmaschinen auf Online-Shops oder Content-Angeboten wird in ihrer Bedeutung für die Performance der Website von vielen Site-Betreibern noch immer unterschätzt.

Für Seitenbetreiber ist die Gruppe der Suchmaschinennutzer aus drei Gründen sehr interessant: 1. Sie zeigt deutlich ihr Interesse an einem bestimmten Thema oder Produkt. 2. Der Konversionspfad – Startseite, Ergebnisseite, Detailseite – ist sehr klar und darum für

Optimierungen gut geeignet. 3. Die Nutzer interner Suchmaschinen konvertieren gewöhnlich deutlich höher als der Besucherdurchschnitt.

Für die Optimierung der Suchbewegungen stehen vor allem die Suchergebnisseiten im Zentrum der Aufmerksamkeit. Für die Performance-Optimierung ist es wichtig, auf die Null-Treffer-Keywords zu achten (wenn die Suche gar kein Ergebnis liefert) und die Listung der Suchergebnisse zu verbessern. Viele interne Suchmaschinen legen offenbar vor allem darauf Wert, möglichst viel zu finden, doch dies nützt weder dem Besucher noch dem Betreiber. Wichtig ist, dass die relevanten Suchergebnisse auf den ersten Plätzen der Liste dargestellt werden. Alle weiteren werden kaum angeschaut, von Ergebnissen auf weiteren Seiten (zum Beispiel Ergebnisse 21 bis 30 oder Ähnliches) ganz zu schweigen. Da es sich bei Suchmaschinen auf kommerziellen Seiten nicht um neutrale Recherche-Werkzeuge handelt, geht es nicht primär um eine vollständige Abbildung der Resultate, sondern um eine effektivitäts-orientierte Präsentation der Suchergebnisse. Im Sinne des Betreibers sollten also jene Produkte oder Inhalte auf den vorderen Plätzen stehen, die die höchste Gewinnmarge gepaart mit einer großen Bestellwahrscheinlichkeit haben. Auch wenn diese Daten für Nicht-Online-Shops schwerer zu fassen sind, sollte doch trotzdem versucht werden, ein Bild vom „Wert" der Klicks auf die verschiedenen Suchergebnisse zu bekommen.

Um die Performance der Präsentation der Suchergebnisse im engeren Sinn zu messen, bietet sich die Berechnung der durchschnittlichen Klickhäufigkeit der Suchergebnisse an, die Absprungrate der Seite und die Anzahl der Suchdurchläufe pro Besuch. Bestellungen, Registrierungen oder Kontaktanfragen im weiteren Verlauf des Besuches können auf den Klick des Suchergebnisses zurückgerechnet werden.

Abspielraten und Abspieldauer

Für die immer wichtiger werdenden Multimedia-Elemente auf den Websites wie Video- oder Audiostreams, Clips oder Podcasts ist die Erfassung ihrer Beliebtheit besonders wichtig. Je nachdem, in welchem Format die Streams implementiert werden, kann die Messung technisch unterschiedlich realisiert werden (siehe Abbildung 7).

Flash-Inhalte, seien es ganze Bereiche mit Navigationen, einzelne Anwendungen wie ein „Handy-Finder" oder Multimedia-Streams, lassen sich vergleichsweise leicht messen, da in die Inhalte oder die Applikation Code als Trigger für Zählimpulse integriert werden können. Theoretisch ist dies genauso in RealAudio/Video- oder in Microsoft-Mediaplayer-Formaten möglich, oft aber aufwendiger in der Umsetzung. Manchmal kann es sinnvoll

Abbildung 7:
Video-Inhalte
auf *Spiegel Online*

sein, den Code nicht in die Inhalte selbst (da sie eventuell sehr oft wechseln oder sehr zahlreich sind), sondern in die Abspielapplikationen zu integrieren.

Jedenfalls ergeben sich vor allem drei Zahlen, die man im Auge behalten sollte:

1. Wie oft wird der Stream gestartet? Wie viele Besucher zeigen also überhaupt Interesse an den Inhalten?
2. Wie lange ist die Dauer der Nutzung im Durchschnitt? Über verschiedene Parameter kann versucht werden diese Zahl zu errechnen, zum Beispiel über den durchschnittlichen Zeitpunkt der Nutzung der Stop/Pausebuttons, die Größe des übertragenen Datenvolumens (wenn messbar über Serveraktivität). Mindestens verfügbar sein sollte die Anzahl der Nutzungsvorgänge, bei denen die Datei vollständig angesehen/angehört wurde.
3. Wie sieht das Verhalten der Nutzer im weiteren Verlauf aus? Werden weitere Inhalte abgerufen, wird etwas bestellt, registrieren sich die Besucher? Im Kontext der Bewertung der Performance der Inhalte sollte jedoch das Primärziel der Website nicht aus den Augen verloren gehen.

Eine lesenswerte Beschreibung der Möglichkeiten zur Messung von Multimedia-Elementen findet sich zum Beispiel einem Whitepaper des Web-Analytics-Anbieters Nedstat aus dem Februar 2008[16].

[16] „Messung von Multimedia Content in einer Web 2.0 Umgebung", Eric Peterson, Web Analytics Demystified, Inc. Michiel Berger, Nedstat BV Februar 2008, http://www.nedstat.de/web/nedstatde.nsf/pages/messung-von-multimedia-content-publications

Rahmendaten: Bestellvolumen, Stornoquote, CPO, CPC, ROI

Wie bereits erwähnt, ist die Effektivität einer einzelnen Seite nur im Rahmen der Gesamtperformance verstehbar. Ebenso muss man bei der Optimierung einzelner Elemente ein Auge auf mögliche Veränderungen übergeordneter Kennzahlen haben.

Dabei spielen vor allem folgende Größen eine entscheidende Rolle:
Veränderungen einzelner Seiten können nicht nur positive Effekte auf die Gesamtkonversion haben (Berechnung siehe oben), sondern diese auch vermindern. Wenn zum Beispiel beim Nutzer Erwartungen geweckt werden, um ihn eine Seite weiterzuleiten, diese dann aber enttäuscht werden, wird die Absprungrate nur einen Schritt nach hinten verzögert.

Ebenso können sich die **Zusammensetzung von Warenkörben, die Anzahl der Artikel darin und das durchschnittliche Bestellvolumen** unter Umständen auch negativ verändern. Es gehört letztendlich zum Ziel der Optimierung die Anzahl der Artikel im Warenkorb zu erhöhen oder bestimmte Artikel attraktiver zu machen. Es bleibt aber darauf zu achten, dass die positive Veränderung nach der einen Seite keine (stärkeren) negativen Effekte nach der anderen Seite begleitet.

Stornoquoten geben an, wie viele Bestellungen oder ähnliche Transaktionen nach der Bestätigung vom Kunden wieder rückgängig gemacht werden. Wenn beispielsweise bestimmte Artikel stärker in den Vordergrund gestellt werden um den Kaufanreiz zu erhöhen werden, oder wenn die Schwelle zur Bestellung verringert wird, kann es sein, dass die Besteller in größerer Anzahl anschließend doch nicht so glücklich mit ihrem Kauf sind und den Auftrag wiederrufen. Dieser Widerruf zum Beispiel in Form von Warenrücksendungen geschieht meistens jenseits der vom Web-Analytics-System erfassten Vorgänge und findet darum nicht automatisch Eingang in die Berechnung der Kennzahlen. Umso wichtiger ist es, auf die eine oder andere Weise diese Entwicklungen zu beobachten. Teilweise können zusätzliche Daten wie zum Beispiel Stornierungen in die Web-Analytics-Systeme integriert werden. Ist dies nicht möglich, so können umgekehrt die Transaktionsdaten aus dem Traffic-Analyse-System mit internen Daten im Unternehmen selbst verknüpft werden. Wie schwierig dies im Einzelfall auch ist: Stornierungen können eine entscheidende Rolle bei der Beurteilung von Optimierungen im Gesamtkontext spielen.

CPO (Cost per Order) und **CPC (Cost per Customer)** zeigen das Verhältnis des Mitteleinsatzes zum Gewinn auf der Ebene einzelner Transaktionen bzw. Kunden. Diese Kennzahlen sind bisher vor allem bei der Beurteilung von Kampagnen verbreitet. Da Veränderungen in der Seitengestaltung aber Einfluss auf das Bestellverhalten nehmen können (und sollten),

gehören diese Kennzahlen zu einer sorgfältigen Performance-Beobachtung der Website dazu.

Der ROI (Return on Investment) ist die einfachste Kennzahl, um die Kosten und den Nutzen einzelner Maßnahmen sowie der Online-Aktivitäten insgesamt zu messen. Durch die Verrechnung des Einsatzes mit dem daraus resultierenden Gewinn erhält jede Optimierung eine harte und zuverlässige Beurteilung in der Währung, um die es am Ende bei einem professionellen Online-Angebot geht: dem Geld. Zudem ergibt sich der Vorteil, dass verschiedene Optimierungsansätze und -werkzeuge (zum Beispiel Optimierung der Startseite versus Optimierung des Registrierungsprozesses oder Online-Befragungen versus A/B-Test) gegeneinander vergleichbar werden (wie ja im Übrigen auch Kampagnen oder Spezialangebote) und der effizienteste Ansatz weiterverfolgt werden kann.

Scoring & Monetarisierung

Trafficdaten wie Page Impressions, Konversionsraten oder Besucherloyalität haben ihre Relevanz in einer eigenen, virtuellen Realität. Zunächst haben sie nichts mit der „wirklichen" Geschäftsentwicklung zu tun. Ein Abteilungsleiter, dessen Budget für ein Optimierungsprojekt in Anspruch genommen werden soll, kann durchaus mit Recht fragen, was es ihm bzw. dem Unternehmen denn bringt, wenn die Anzahl der durchschnittlich aufgerufenen Seiten pro Besuch ansteigt.

Dieser Argumentation kann am besten begegnet werden, wenn das Geschehen auf der Website in seinem Beitrag zur Unternehmensentwicklung bereits definiert wurde. Es ist noch einfach, eine geringere Cost-per-Customer-Quote bei Aufträgen über das Netz im Vergleich zum Offline-Vertrieb in die Waagschale zu werfen. Aber wie ist es mit Aufrufen von Inhalten im Pressebereich, Download von Produkt-Broschüren oder Steigerung der Page Views insgesamt? Modelle zum Scoring (Zuweisung von Punktwerten) oder zur Monetarisierung (Zuweisung von Geldwerten) sollen dabei helfen, diese etwas luftigen Statistiken auf den Boden der Realität zu holen. Diese Zuweisungen sind immer hochgradig individuell und fußen auf einer im Unternehmen gemeinsam getroffenen Einschätzung, welche Aktion am Ende wie viel wert sein könnte. Weiter oben wurde bereits an einem Beispiel eine mögliche Zuteilung von Punktwerten zu einfachen Seitenabrufen, Downloads oder Kontaktanfragen dargestellt. Hinsichtlich der Entwicklung und Ausgestaltung sei auch noch einmal auf das Buch *Web Analytics – Damit aus Traffic Umsatz wird*" [17] verwiesen, in dem dies ausführlicher

[17] Frank Reese: Web Analytics – Damit aus Traffic Umsatz wird, Business Village 2008

dargestellt wird. Systeme wie Google Analytics bieten bereits einfache Möglichkeiten zur Verbindung einzelner Interaktionen mit Punkt- oder Geldwerten.

Zielgruppen und Segmentierung der Besucher

Mit Segmentierung wird im Analytics-Kontext die Definition von Filtern bezeichnet, um Nutzer mit bestimmten Eigenschaften zu Gruppen zusammenzufassen und miteinander vergleichbar zu machen. So können etwa mit dem Filter „Zeige alle Besuche, die über die Kampagne X gekommen sind" die Merkmale dieser Gruppe von Besuchen genauer betrachtet werden und auch mit den Merkmalen der Besuche insgesamt verglichen werden. Theoretisch kann jede Art von Merkmal, das über Traffic-Analysen erfasst wird – Herkunft, Seitenabrufe, Häufigkeit von Besuchen oder auch technische Eigenschaften – als Filter zur Unterscheidung von Nutzergruppen dienen. Die meisten der heute verfügbaren Systeme bieten recht gute Möglichkeiten zur Segmentierung und machen die Suche nach interessanten Nutzergruppen vergleichsweise leicht und komfortabel. Die Idee und das Wissen, welche Nutzergruppe für eine genauere Analyse interessant sein könnte, liegen aber immer noch beim Anwender.

Klassischerweise zeigen folgende Besucher-Gruppen recht deutliche Verhaltensunterschiede:

Erst- versus Wiederholungsbesucher

Je nachdem, an welchem Punkt Sie in der Entwicklung Ihrer Website stehen, kann es wichtiger sein, zusätzliche neue Besucher auf die Website zu bringen oder die bereits mit Ihrem Angebot bekannten Besucher gezielter anzusprechen. Konkrete Zahlen zu nennen ist immer ein wenig schwierig, denn ein gutes Verhältnis hängt vom Ziel der Site ab. Verhältnisse jenseits von 70/30 (also 80/20, 90/10) sind aber jedenfalls bedenkenswert.

Generell ist misslich, wenn fast immer nur die gleichen Besucher Ihre Website besuchen und kaum neue hinzukommen. Dann sind zwar die durchschnittlichen Zahlen für die Loyalität der Besucher (zum Beispiel Zeit/Besuch, Seitenabrufe/Besuch, weitere Engagement-Zahlen) dann gewöhnlich recht hoch und das Angebot scheint gut zu funktionieren – aber eben nur mit den bereits bekannten. Im Sinne einer Erhöhung der Reichweite des Angebotes (vor allem im Vergleich zu ähnlich gelagerten Konkurrenz-Sites und in Bezug auf die Anzahl der täglichen Suchanfragen zum Thema in externen Suchmaschinen) könnten Werbe- oder PR-Maßnahmen für einen größeren Erfolg der Site sorgen.

Kritischer noch ist es, wenn ein sehr hoher Anteil neuer Besucher verzeichnet wird. Dies deutet auf vordergründig sehr gute Werbe-Kampagnen hin, ohne dass diese aber dauerhaften Erfolg erzeugen zu können.

Eine differenzierende Betrachtung dieser Gruppen liefert jedenfalls sehr gute Hinweise auf die funktionierenden und nicht funktionierenden Teile der Website.

Kunden und Besucher
Ähnlich gelagert wie die vorangegangene Unterscheidung liegt die Segmentierung von bereits bestehenden Kunden und Nur-Besuchern. Auch hier können einfache Mittelwerte Erfolge zeigen, die tatsächlich nicht vorhanden sind. So kann die Site eine gute Konversionsrate vorweisen, aber dies nur mit stets wiederkehrenden Kunden, die genau die Abläufe der Site kennen und sich mit den Umständlichkeiten angefreundet haben. Alle anderen Besucher werden möglicherweise von Unverständlichkeiten oder Fehlleitungen abgeschreckt. Die Betrachtung von Besuchergruppen, die bereits auf der Site sind oder waren, dann aber aus irgendeinem Grund nicht konvertiert sind, kann ein großes Wachstumspotenzial aufdecken.

Interessencluster
Je grösser oder vielfältiger ein Online-Angebot ist, desto unterschiedlicher sind auch die Motivationen der Besucher auf einer Site. Diese inhaltlichen Interessenunterschiede (zum Beispiel Presse, Kundensupport, Stellensuche, Produktgruppen) führen folgerichtig zu einem sehr unterschiedlichen Verhalten auf der Site. Eine Optimierung, die allein auf ein bestimmtes Ziel abstellt, aber die Diversität der Nutzergruppen aus dem Auge verliert, bedeutet wird viel Mühe in ein unerreichbares Ziel zu investieren. Sinnvoller ist es, nach einer gründlichen Analyse der vorhandenen Gruppen, die möglicherweise unterschiedlich wertvollen Gruppen auch bei der Durchführung und Bewertung von Optimierungsprojekten getrennt zu verfolgen.

Inhaltliche Interessenunterschiede der Besucher können vor allem über zwei Wege festgestellt und definiert werden:
Beim **Aufruf bestimmter Bereiche** auf der Seite kann von einem entsprechenden Interesse ausgegangen werden. Wer beispielsweise Informationen im Pressebereich aufruft, ist an allgemeinen Informationen über das Unternehmen interessiert. Bei kleinen Seiten kann dies allerdings in eine Sackgasse führen, da Besucher alle Bereiche auf der Website streifen, eine Differenzierung also keinen Sinn ergibt. Je beliebter und vielfältiger aber das Online-Angebot ist, desto schärfer kann mit solch einfach Regeln getrennt werden. Dies

gilt selbstverständlich mehr noch für Online-Shops, die eine ganze Palette unterschiedlicher Produkte anbieten oder Content-Angebote, die Inhalte aus verschiedenen Interesse-Bereichen vorhalten. Immer sollten die so gewonnenen Ergebnisse aber eher als Hinweise für Optimierungsoptionen genutzt werden, denn als wirklichkeitsgetreue Abbildung der Besucherinteressen.

Besucher, die über Suchmaschinen auf die Seite finden, liefern für die Analyse außerdem das Stichwort, nach dem sie gesucht haben. Ebenso wie die Eingaben in interne Suchmaschinen bieten diese Informationen sehr konkrete Hinweise auf die Interessen der Besucher. Um diese Informationen für eine Segmentierung fruchtbar zu machen, ist es notwendig Cluster (Gruppen) von Stichworten zu bilden, da sonst Begriffe und Schreibweisen wie zum Beispiel ‚Optimierung', ‚Site-Optimierung', ‚Optimirung', ‚optimierung' getrennt gelistet werden.

Je nachdem, welche Aktivitäten auf der Site gewünscht sind oder angeboten werden, ergeben sich vielfältige Ansätze zur sinnvollen Unterscheidung der Besucher: abgegebene Kommentare, Broschürendownloads, RSS-Abonnenten, Besucher von Preisrobotern, In-den-Warenkorbleger, Mobilgerät-Nutzer, geographische Herkunft, Zeitrahmen des Besuches uvm. – letztlich liegt es in der Kompetenz und am Engagement der Website-Verantwortlichen, wie und wo sie ihre verborgenen Schätze entdecken können.

IV. Ansatzpunkte zur Conversion Optimierung

von Frank Reese

1. Optimierung der Startseite 59
2. Landing Pages 64
3. Formularseiten 70
4. Interne Suchmaschinen 74
5. Bestellprozesse 76

Jede Website ist das Ergebnis harter Arbeit. Die Gestaltung von Seitenstrukturen, Startseiten, Navigationen, Übersichten, Layouts, Landing Pages, Bestellprozessen und vielem mehr erfordert eine ganze Reihe von Entscheidungen, die (fast) immer nach bestem Wissen und gründlicher Erwägung getroffen werden. Am Ende steht sie online, die fertige Website, viele Besucher nutzen sie, steigen hier ein und verlassen sie dort, bestellen Artikel oder registrieren sich für einen Newsletter, drucken Inhalte aus oder setzen sich ein Lesezeichen. Zeit um sich zu gratulieren und auf die gemeinsame Leistung stolz zu sein.

Der Launch der Site ist ein Schritt um das Ziel – mehr Besucher, mehr Umsatz, bessere Übersichtlichkeit, höhere Attraktivität – zu erreichen, Aber eine Website initiativ zu erstellen ist eher wie der Entwurf und das Anlegen eines Gartens. Die Arbeit hört nicht auf, sondern fängt danach erst richtig an. Hier muss etwas herausgenommen, dort eine Lücke gefüllt werden. Ein Bäumchen wächst nicht recht und braucht zusätzliche Stützen, anderes Kraut wächst ohne Hemmung und wuchert die ganze Nachbarschaft zu.

Zwar mögen der Entscheidung für zum Beispiel die Anordnung der Inhalte auf einer Produktdetailseite lange Diskussionen vorausgegangen sein, aber jetzt, wo die Arbeiten abgeschlossen sind, glänzt das Resultat mit der Sicherheit des Vorhandenen: „So muss es aussehen und nicht anders". Das ist ein Irrtum. Auch wenn es schwer ist, das Vorhandene gleich wieder in Frage zu stellen, aber so genial war noch kein Entwurf, dass er im Live-Betrieb nicht angepasst, kalibriert, modelliert werden müsste. Das Vorhandene ist die grundlegende Struktur, nicht das endgültige Ergebnis (hart, aber wahr: Sie werden mit der Arbeit an der Seite eigentlich nie fertig). Stellen Sie sich einen See voller Fische vor, aber Sie haben gerade eine erste Version Ihrer Angel entwickelt, mit der Sie nur die am leichtesten zu fangenden Fische angeln. Um dem Reichtum des Sees, dem vollen Potenzial Ihrer Website langsam näherzukommen, bedarf es aufmerksamer und sorgfältiger Detailarbeit, die – langsam aber sicher – zu immer größerem Erfolg führt.

In diesem Kapitel sollen typische Ansatzpunkte für diese Detailarbeit dargestellt werden. Nach dem Prinzip des größten Nutzens sind dies vor allem empfindliche Bereiche der Site – mit einer klaren Zielvorgabe einerseits und andererseits mit einem hohen Risiko, dass die Besucher nicht das machen, was der Betreiber sich vorstellt. Das wären vor allem:

Die Startseite. Das Tor zur Website, häufig die meist-besuchteste Seite, viele Nutzer beginnen hier ihren Besuch und – nach einem schnellen Blick auf die Inhalte – gewinnen einen Eindruck, ob sie hier finden was sie suchen oder nicht.

Landing Pages. Viele kleine Tore zur Website, die für eine bestimmte Gruppe von Besuchern im Kontext einer Online-Kampagne gestaltet werden. Anders als die Startseite, können Landing Pages kompromisslos für diese Gruppe und den Zweck der Kampagne gestaltet werden.

Formulare. Sind die Besucher endlich bereit, sich tiefer oder dauerhafter mit den Inhalten der Site zu beschäftigen (Newsletter, Info-Material, Anfragen), zeigen viele dann doch eine erhebliche Scheu, das Formular korrekt abzusenden und brechen ab.

Interne Suchmaschinen. Die besten Helfer für Besucher, die wissen was sie wollen. Diese auch für die Betreiber sehr interessante Zielgruppe wird leider oft mit nur wenig durchdachten Suchergebnissen beliefert.

Bestellprozesse. Der heilige Gral von Online-Shops und allen Seiten, die einen mehrstufigen Bestell- oder Kontaktablauf benötigen. Jeder investierte Optimierung-Euro kann sich hier vielfach auszahlen.

1. Optimierung der Startseite

Das Eingangstor für die Besucher, das Aushängeschild, Schaufenster, Portal – wie wurden Startseiten nicht noch genannt. Sicher gehören sie zu den wichtigsten Seiten in einem Online-Angebot. Bei den meisten Websites werden Besucher vor allem über diese Page Zugang zum Angebot suchen. Diese faktische Wichtigkeit – die sich an den Zugriffszahlen und der Herkunft der Besucher ablesen lässt, wird auf der Anbieter-Seite ergänzt durch den Willen, auf der Startseite einen Überblick und Eingang in die Inhalte der Site zu bieten. Auf der Website zeigt sich die Wichtigkeit der verschiedenen Inhalte und Bereiche: Was steht für das Unternehmen im Vordergrund? Was präsentiert es auf den zentralen Positionen und dort wo es am sichtbarsten ist?

Die Startseite von IBM Deutschland (siehe Abbildung 8) zeigt die Inhalte des Online-Angebotes im Überblick. Das Unternehmen bietet eine sehr große Menge an Informationen, die für die Besucher in Menus, Navigationsstichworte oder inhaltliche Bereiche zusammengefasst werden müssen und doch den entsprechenden Zielgruppen schnell ins Auge fallen sollen.

Abbildung 8: Startseite IBM Deutschland Februar 2009

Deutlich im Zentrum der Seite steht ein Themenschwerpunkt: „Intelligenz auf der ganzen Welt." Nach einem Klick werden Inhalte rund um das Thema „Smart Planet" präsentiert. Ein weiteres Zukunftsthema und Inhalte zum Messeauftritt sind ebenfalls im Zentrum der Seite abrufbar.

Darunter finden sich die Bereiche „Information", „Kaufen", „Support", „Meine IBM Links" und „Über IBM". In der Navigationsleiste im oberen Bereich könne über ausklappbare Menus sehr viele weitere Inhalte direkt angesteuert werden.

Wenn Sie jetzt von einem IBM-Mitarbeiter gefragt würden: „Was könnten wir auf unserer Startseite besser machen?", was würden Sie sagen?

„It depends" wäre vielleicht eine gute, diplomatische Antwort. Die Seite ist selbstverständlich professionell gestaltet, da finden sich keine Fehler in der Darstellung, der gesamte Screen ist ohne scrollen sichtbar, die Hinweise sind klar und eindeutig formuliert und laden zum weiteren Lesen ein.

Die Frage wäre vielmehr: Was will IBM mit dieser Startseite erreichen? Was wird bereits erreicht? Wo liegen die Erfolgsparameter der Site insgesamt? Welches Verhalten des Besuchers zahlt sich wie und in welcher Höhe aus? Will das Unternehmen Themen setzen? Kunden mit zusätzlichen Informationen und Unterstützung versorgen? Neue Kunden gewinnen? Sich selbst darstellen? Mitarbeiter gewinnen? – Wir wissen es nicht, können aber davon ausgehen, dass alle diese Ziele mehr oder weniger gleichzeitig verfolgt werden. Eine systematische Optimierung der Startseite jenseits allgemeiner Kriterien (zum Beispiel Fehler- und Barrierefreiheit) ist darum unter diesen Bedingungen nicht möglich. Was wäre gewonnen, wenn die Absprungrate durch einen Austausch der zentralen Überschrift gesenkt würde? Oder die Klickhäufigkeit auf einzelne Bereiche „Kaufen“ oder „Support“?

Um einen vernünftigen Ansatz zur systematischen Optimierung zu gewinnen, ist zunächst eine durchgängige Definition der verfolgten Ziele, der Messung dieser Ziele und der Bewertung der verschiedenen Ziele über einen vergleichbaren Maßstab notwendig. Fehlt dies, so kann die Erhöhung der Klickrate auf den Bereich „Kaufen“ auf Kosten der Abrufzahlen für „Informationen“ nicht gewichtet bzw. bewertet werden.

Die gezielte Veränderung einer so komplexen Einheit wie der Startseite verlangt eine sorgfältige Beachtung der vielen nachgelagerten möglichen Konsequenzen einer Anpassung. Anders als zum Beispiel Landing Pages oder Formularseiten müssen Startseiten vielen Ansprüchen gleichzeitig genügen. Sie sind immer das Resultat von Kompromissen zahlreicher verschiedener legitimer Interessen. Die mühevoll definierten und berechneten Punktwerte sind für eine Optimierung die Voraussetzung, weil sie zum einen verhindern, dass durch die positive Veränderung eines kleinen Elements in einem Bereich, eine größere negative Veränderung bei einem anderen Element verursacht wird. Zum anderen liefern die vergleichbaren Werte eine belastbare Diskussionsgrundlage hinsichtlich der Gewichtung und Entwicklung der einzelnen Seitenelemente.

Bevor ein Optimierungsprojekt starten kann, ist aber noch ein zweiter Analyseschritt notwendig. Bisher ist zwar klar, was die Besucher auf der Site machen, was diese ‚wert‘ sind und wie hoch die Anteile der Aussteiger jeweils sind (Absprungraten der einzelnen Seiten), unklar ist aber die primäre Motivation der Besucher. Um dieser näher zu kommen und die Bewegung der verschiedenen Besuchergruppen dann im Vergleich zu sehen, ist es notwendig, die Besucher zu segmentieren, also Zielgruppen zu unterscheiden. Zwar könnte unterstellt werden, dass zum Beispiel die Besucher im Bereich „Support“ sich für Support interessieren, mithin ein Filter „alle Besucher, die mindestens eine Seite im Bereich Support abgerufen haben, zählen zur Gruppe der ‚Supportsucher‘“ sinnvoll wäre, al-

lerdings dreht sich diese Zuordnung um sich selbst: Unsichtbar bleiben dabei ja genau jene Besucher, die zwar an Support interessiert wären, aber die Inhalte nicht gefunden oder abgerufen haben. Diese wären aber das bisher unausgeschöpfte Potenzial und Ziel der Optimierung. Besser ist es deswegen, die Herkunft der Besucher zu analysieren. Hier können vor allem thematisch unterscheidbare Partnerseiten als Herkunft ein Differenzkriterium bilden, als auch die Suchworte, mit denen die Besucher über eine externe Suchmaschine auf die Seite gefunden haben (siehe *Suchcluster* in Kapitel III).

So vorbereitet zeigt ein Blick auf die Analyse der Startseite noch weitere Informationen und damit ausreichend Daten um eine Optimierung in Angriff zu nehmen.

Da der zu erwartende Nutzen immer dort am größten ist, wo eine Veränderung den stärksten positiven Effekt hat, kommen für eine Optimierung der Startseite vor allem die zentralen, häufig nachgefragten und leicht zu verändernden Elemente in Betracht. Die folgende Aufzählung gibt einige Ansatzpunkte zum Testen verschiedener Elemente und Varianten:

Visuals (Hero Shot, Abbildung, Emotional)
Abbildungen sind zweifellos wichtig für die Attraktivität der Seite. Nicht alle Besucher reagieren direkt auf Bilder, aber oft lenken sie die Aufmerksamkeit der Betrachter. IBM hat in diesem Fall ein sachlich-technisches Motiv gewählt, das mit den Farben und dem Thema harmoniert. Neben solchen abstrakten Motiven wären aber auch zum Beispiel möglich:

- Darstellung von Menschen (Mann, Frau, Gruppen, verschiedene Situationen und Haltungen)
- Humorvolle Abbildungen
- Problem-/Nachfrageorientierte Abbildungen

Überschriften
Auch hier bleibt die Startseite nüchtern und technisch: „Gute Ausstattung, zunehmende Vernetzung, Intelligenz auf der ganzen Welt."

Viele Untersuchungen haben gezeigt, wie entscheidend die Headline für den Erfolg einer Website sein kann, sie ist das erste, was der Besucher liest und kann ihn interessieren, langweilen oder sogar abschrecken. Jedenfalls zeigen Ergebnisse von Tests sehr oft, dass die von Fachleuten oder Redakteuren ausgewählte Überschrift nicht die attraktivste für die Besucher ist. Da die zentrale Überschrift zudem das jeweils wichtigste Thema der Startseite

transportieren soll, ist es viel effektiver, nicht eine, sondern drei, vier oder sogar fünf verschiedene Überschriften zu produzieren und das Verhalten des Besuchers über die Auswahl entscheiden zu lassen. Auf unserer Beispiel-Seite wäre zum Beispiel ebenso möglich:

- „Wir laden Sie ein: Willkommen auf dem Smart Planet."
- „IBM gestaltet Zukunft: Planet Smart"
- „Intelligenz und Vernetzung. Leistung durch Ausstattung"

Nicht, dass der Autor der Meinung wäre, diese Überschriften wären nun besser als die tatsächliche. Die Beispiele sollen vor allem den Raum der Möglichkeiten wieder öffnen, der durch die Entscheidung für eine bestimmte Überschrift und deren Umsetzung geschlossen wurde. Seien Sie bereit für das Experiment.

Zudem kann natürlich auch gezielt mit der Länge der Überschriften, der Schriftart, -farbe und –größe experimentiert werden.

Teaser und weitere Angebote

Auf der Beispiel-Startseite werden im unteren Bereich einige Container mit thematisch gruppierten Inhalten präsentiert. Viele Webseiten zeigen den Besuchern auf diese Weise neben dem zentralen Thema weitere – aus ihrer Sicht – interessante Inhalte. Diese Container können – einmal unabhängig von ihrer Positionierung gesehen – verschiedene Angebote oder Themen in den Fokus der Aufmerksamkeit bringen. Im Fall der Beispiel-Seite könnte zunächst variiert werden, welches Thema initiativ hervorgehoben wird. Aktuell ist dies „Informieren", möglicherweise könnte aber die direkte Hervorhebung von „Support" oder „Kaufen" eine stärkere Wirkung auf die Besucher ausüben.

Die farbliche Gestaltung, Formulierung der Headlines, inhaltliche Auswahl können bei diesem zweitwichtigsten Element (nach dem zentralen Thema) ebenso leicht variiert werden.

Navigationsleisten und Suchfelder

Die oben genannten Elemente stehen sicher zuerst im Fokus einer systematischen Optimierung, die die Performance der Site und aktuelle Ziele im Auge hat. Aus dem Verhalten der Besucher ergeben sich möglicherweise aber noch weitere interessante Ansatzpunkte. So kann eine Analyse von Suchbegriffen die zum Aufruf der Seite führen und der Absprungrate andererseits auf Themen hinweisen, die Nutzer zwar erwarten und suchen, aber nicht schnell genug finden. Hier können sich Hinweise auf eine Verlagerung der bevorzugt präsentierten Themen ergeben. Ebenso zeigen die Eingaben im Suchfeld und die

Auswahl von Seitenbereichen über – die sehr reichhaltige – Navigationsleiste im oberen Seitenbereich das Interesse an Inhalten, die entweder gar nicht oder nur schwer erfassbar präsentiert werden.

2. Landing Pages

Landing Pages (Landeseiten) sind ein sehr beliebter Gegenstand der Optimierung. Viele werden für einen ganz bestimmten Zweck gestaltet und sollen so die Besucher von einer externen Seite – zum Beispiel einem Partner, Banner, Adword – empfangen und in die weiteren Angebote der Website leiten. Damit sind sowohl die Ziele der Site als auch die Interessen der Besucher klar definiert. Ob für den einzelnen Klick bezahlt wird oder nicht: Hier kommt eine äußerst interessante Gruppe von Besuchern auf die Seite und die Landing Page soll dafür sorgen, dass diese nicht gleich wieder verschwinden. Die Erfolgskennzahl für eine Landing Page ist also klar: Wie hoch ist der Anteil der Besucher, die auf die weiterführenden Angebote klicken? Und umgekehrt: Wie hoch ist der Anteil der Besucher, die die Landung Page unmittelbar wieder verlassen? Mit der Konzeption und Gestaltung der Landing Page kann nicht die Performance der Banner oder Adwords beeinflusst werden, auch die Preisgestaltung oder Produktauswahl ergeben sich nicht aus der Landing Page-Konzeption. Alles, was vor der Landing Page passiert und alles was später passiert (werden Artikel wirklich bestellt oder nur angesehen?), hat mit der Gestaltung der Landing Page nicht zwingend etwas zu tun. Damit kann sich die Optimierung ganz auf diesen klar bestimmten Zweck konzentrieren.

Aber was ist mit Landing Pages, die nicht für einen bestimmten Zweck entwickelt wurden, sondern einfach Übersichtsseiten im normalen Online-Angebot sind? Leicht zu sehen, dass diese Seiten wegen ihres doppelten Zweckes nicht so leicht optimiert werden können. Es mag für manche Unternehmen sehr viel Aufwand sein, für tausende wechselnde Keywords jeweils eigene Seiten zu erstellen und kleine Unternehmen können schon mit der Gestaltung weniger Landing Pages überlastet sein – allein: Erfolgreicher laufen die exakt zugeschnittenen Pages. Zudem können alle weiteren Seiten, die besonders häufig als Eingangstor von einer bestimmten Quelle, von einer bestimmten Zielgruppe oder zu einem Thema genutzt werden, als Landing Page verstanden und entsprechend optimiert werden. Dies gilt selbstverständlich auch für die in den Suchergebnislisten der Suchmaschinen aufgeführten Adressen, vor allem, wenn für die Listung an guter Stelle viel Mühe in die entsprechende Suchmaschinenoptimierung investiert wurde.

Abbildung 9: Startseite als Landing Page

Auf diese Seite der Firma Rabe gelangt man nach einem Klick auf die Google-Adword-Anzeige nach der Sucheingabe „Mountainbike“. Soweit der Autor als Nicht-Experte zum Thema Mountainbike das beurteilen kann, genießen die Rabe-Räder einen guten Ruf unter Kennern.

Als Landing Page fungiert hier die gewöhnliche Startseite. Man kann wohl annehmen, dass ein Nutzer auf der Suche nach Mountainbikes versteht, dass sich hinter dieser Seite Informationen zum Thema befinden, seine konkrete Situation wird allerdings nicht aufgenommen. Bryan Eisenberg spricht hier von einer ‚scent‘, die der Besucher auf seinem Weg durch das Netz verfolgt. Diese ‚Witterung‘ (so eine freie deutsche Übersetzung von ‚scent‘) auf der Spur nach der gewünschten Information sucht der Besucher auf jeder Seite auf seinem Weg wieder. Wissenschaftliche Untersuchungen zum Surfverhalten von Nutzern haben gezeigt, dass Besucher kleinsten Hinweisen nachgehen, mit der Idee, dort das Gesuchte zu finden,

allerdings brechen die Nutzer diesen Weg verhältnismäßig schnell ab, wenn sich der Weg verliert. Es gibt zu viele verschiedene Spuren zu verfolgen, als dass man Zeit hätte, eine kalte oder doch nur lauwarme Spur zu verfolgen. Die Witterung zu dieser Startseite wird in der Google-Anzeige mit „Mountainbikes Tiefpreise“ angestoßen. Der Betreiber der Seite nimmt dieses Thema aber nicht wieder auf, sondern verlässt sich auf die Hartnäckigkeit der Besucher und die Gesamt-Qualität seines Angebotes.

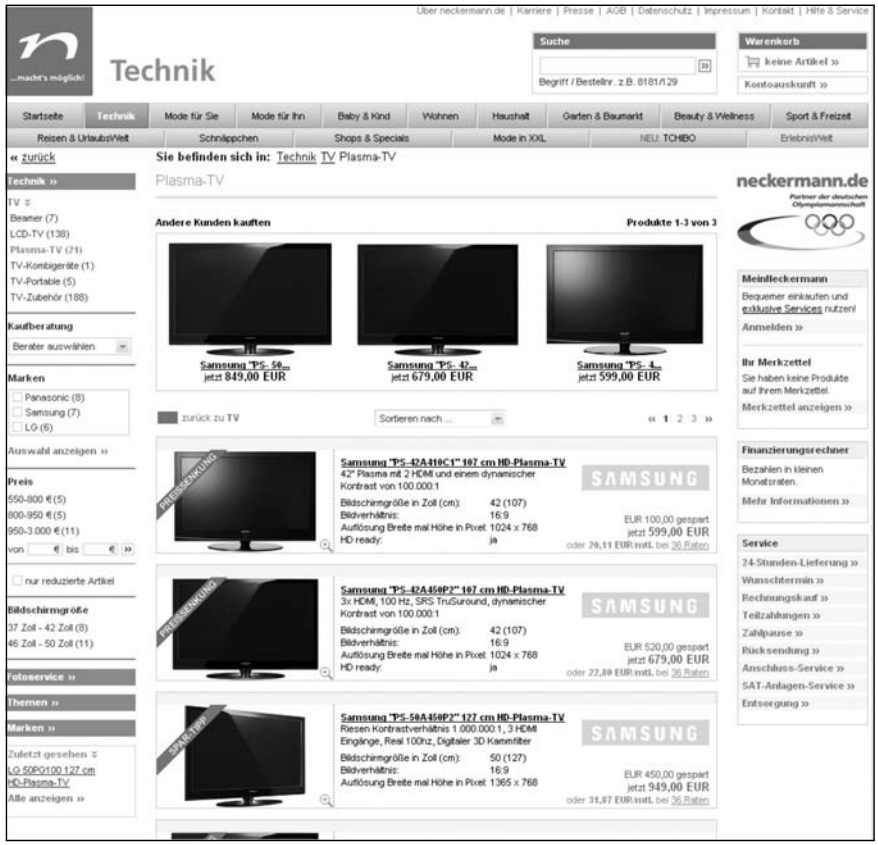


Abbildung 10: Übersichtsseite als Landing Page

Der Versandhändler Neckermann hat die oben abgebildete Seite als Landing Page für das Keyword „Plasma-TV“ eingerichtet. Die Google-Anzeige verspricht „Plasma-TV Angebote“ und die Besucher werden direkt zur normalen Katalogseite weitergeleitet. Damit findet der Besucher sein Thema schnell wieder: Es gibt verschiedene Modelle zur Auswahl, verschiedene Kategorisierungen und sogar eine Kaufberatung.

Als letztes Beispiel für eine Landing Page soll eine Seite von Direct Line dienen:

Abbildung 11: Spezielle Landing Page

Nach einem Klick auf die Adword-Anzeige „Direct Line Versicherung“ bekommt der Nutzer eine Seite präsentiert, die ihm sehr genau zugeschnittene Informationen liefert und nicht viele Möglichkeiten zur Ablenkung bietet. Der inhaltliche Unterschied zur Neckermann-Landing Page besteht vor allem in der Vermeidung einer vollständigen Navigation und der Fokussierung auf die beiden roten Buttons rechts in der Mitte „Mit Kfz-Schein“ und „Ohne Kfz-Schein“.

Welche der Seiten oder Typen auch im individuellen am besten funktioniert (und im Verhältnis zum Aufwand den höchsten Nutzen generiert) kann nur nach einer Analyse der tatsächlichen Trafficströme beurteilt werden. Für eine Messung und Optimierung der Seiten ist aber das letzte Beispiel ganz klar im Vorteil. Sie folgt zudem dem klassischen Aufbau einer Landing Page: Überschrift, Hero Shot, Angebot, Call-to-Action.

Aber noch einmal zurück zu den beiden ersten Beispielen. Eine Optimierung hinsichtlich der Performance als Landing Page muss zwei Momente beachten: Zunächst sollte der Traffic segmentiert werden. Über die URL-Parameter ist der zum Beispiel von dem AdWord stammende Traffic klar unterscheidbar von allen anderen Quellen und kann so als besondere Gruppe gezielt analysiert werden. Als Landing Page verlieren diese Seiten nicht ihre wichtigste Funktion: Eine Brücke zu bilden zwischen der externen Quelle und den intern vorhandenen Inhalten und Angeboten. Nach der Segmentierung kann also die Absprungrate und Weiterklickrate genauer untersucht werden. Ohne diesen Schritt würden die Besucher mit anderen ‚Spuren' die Resultate nicht nur verwässern sondern auch verfälschen. Wenn jetzt optimiert wird, dürfen diese anderen Besucher aber nicht vergessen werden. Die Optimierung im Sinne der Landing Page-Funktion darf nicht auf Kosten der sonstigen Besucher gehen.

Mit den oben erwähnten klassischen Komponenten einer Landing Page liegen auch die bevorzugten Punkte zur Optimierung offen:

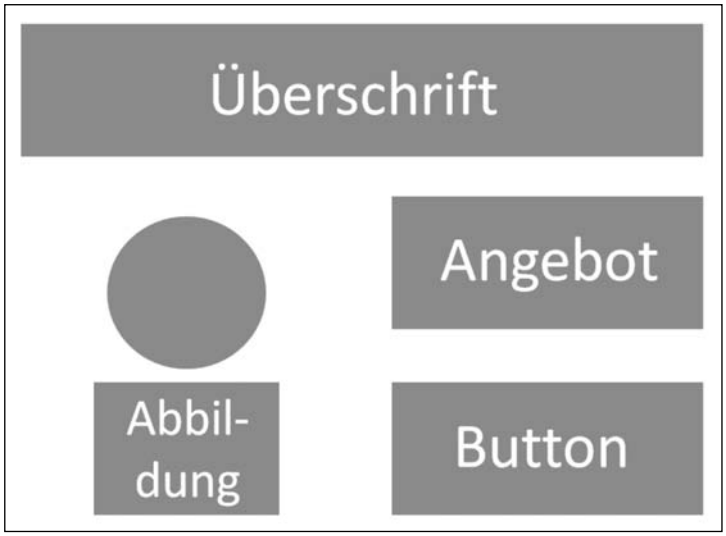


Abbildung 12:
Schema einer spezielle Landing Page

Überschrift

Wie bereits oben gesagt: Die Bedeutung der Überschrift kann gar nicht überschätzt werden. Sie kann die Motivation der Besucher direkt aufnehmen, sie neugierig machen, vor allem ihre Witterung aufnehmen und verstärken, oder – langweilig und wie Zeitverschwendung wirken. Das Verfassen einer guten Headline ist bei Fachleuten gut aufgehoben, allerding sollte man nicht theoretisch die perfekte Formulierung suchen, sondern über Versuche verschiedene Ansätze ausprobieren. Eine Variation der Headline ist ein guter Start in eine systematische Performance-Optimierung.

Abbildung (Hero-Shot)

Dass Bilder eine starke Wirkung haben ist keine Neuigkeit. Die Stimmung, die über ein Bild ausgestrahlt wird, färbt auf alle anderen Inhalte ab. Eine unbeholfene Auswahl kann einen Eindruck von Aufdringlichkeit, Billigkeit oder Unprofessionalität erzeugen, der oder die über die anderen Elemente kaum noch aufzuwiegen ist. Wenigstens unterschieden werden sollte nach Bildern mit Menschen, Frauen oder Männern, direktes Ansehen oder Ansprache des Besuchers, abstrakte Inhalte, humorvolle Inhalte. Was den Besucher im Sinne der Klickraten am stärksten anspricht ist bei Bildern noch viele weniger vorherzusehen als bei den übrigen Inhalten. Ein Ausprobieren ist dringend zu empfehlen.

Angebot

Die Formulierung des Angebotes muss die Erwartungen des Besuchers möglichst genau treffen und eine Bereitschaft zu genauerem Hinsehen erzeugen. Der kurze Text stellt die Brücke zwischen der Headline und dem Weiterklick-Button dar und steht in deutlicher Interaktion mit der Abbildung. Bei einer Variation des Angebotstextes sollte man darum besonders auf Wechselwirkungen achten.

Button

Hier soll der Besucher aktiv werden. Ein Klick auf die Schaltfläche ist das ganze Ziel der Landing Page. Oben im ‚Direct Line'-Beispiel sind die Buttons sehr stark hervorgehoben. Farben, Positionierung und Größe machen die Buttons zum Blickfang der Seite. Automatisch ziehen sie die Aufmerksamkeit auf sich und stellen den Besucher gleich noch vor eine Frage: Mit oder ohne Kfz-Schein? Worin auch immer die genaue inhaltliche Funktion der Unterscheidung an diesem Punkt liegt (eine spätere Unterscheidung beim Ausfüllen des Formulars wäre ja vielleicht auch möglich), zieht diese Frage den Besucher automatisch in den weiteren Prozess hinein, er wird nicht gefragt „Möchten Sie jetzt unverbindlich ein Angebot berechnen?", sondern direkt vor eine Entscheidung gestellt, die im Sinne der Landing Page-Performance nur positiv ausfallen kann. Welche Art von Button den besten

Effekt zeigt, ist im Voraus kaum abzusehen. Weiter oben, im Einführungskapitel ist ein Beispiel von „Enroll Online"-Buttons, bei denen der ursprüngliche Button im Vergleich zu den beiden Varianten eher zurückhaltend wirkt („Online" statt „Now" oder „Now!"), im Test aber trotzdem die beste Performance zeigt. Im Zuge einer multivariaten Analyse sollte ein Test verschiedener Layouts und Beschriftungen der Buttons mitbedacht werden.

Entwirft man nun eine multivariate Analyse dieser Landing Page ergibt sich ein 4×4×4×4-Testfeld, in dem die einzelnen Faktoren (sprich: Seitenelemente) in ihrer Performance untersucht werden (siehe folgende Abbildung [18]).

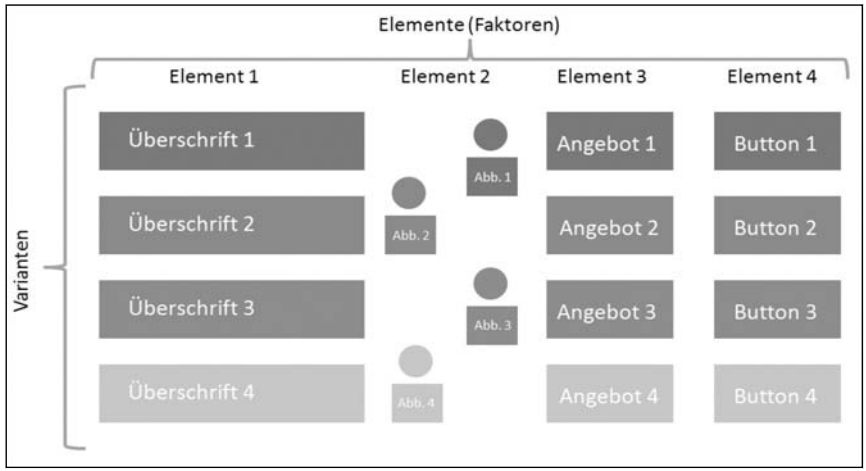


Abbildung 13: Elemente und Varianten einer multivariaten Analyse

Die Einzelheiten zur Vorgehensweise werden im entsprechenden Abschnitt ausführlicher dargestellt. Die Kombination und Variation der verschiedenen Elemente kann aber ebenso gut die Vorlage für andere Verfahren der Website-Optimierung sein.

3. Formularseiten

Ebenso wie Landing Pages haben Formularseiten einen klaren Call-to-Action: Der Besucher soll das Formular möglichst vollständig und im richtigen Format ausfüllen und dann absenden. Allerdings liegen die Abbruchsquoten auf Formularseiten aus naheliegenden Gründen

[18] Abbildungen inspiriert durc hdas lesenswetre Blog http://testingblog.widemile.com/

deutlich höher als bei Landing Pages: Der Besucher muss hier aktiv werden und folgende fünf Voraussetzungen müssen erfüllt sein:

- Grundsätzliche Bereitschaft ein Formular zu bearbeiten
- Bereitschaft, die einzelnen Informationen mitzuteilen
- Kenntnis der richtigen Eingaben („Wie war noch mal meine Mobilnummer?“)
- Eingabe im richtigen Format (zum Beispiel 020209 für den 2. Februar 2009)
- Absenden der Daten

Die folgende Abbildung zeigt eine Seite aus dem Online-Angebot von kicker.de. Um Mitglied der ‚Community‘ zu werden, müssen die Besucher dieses Formular ausfüllen (siehe Abbildung 14 auf der folgenden Seite).

Die *grundsätzliche Bereitschaft, das Formular auszufüllen,* ist bei den Besuchern gegeben, da sie die Services für registrierte Mitglieder in Anspruch nehmen wollen. Je stärker diese Motivation, desto höher die Bereitschaft (und die Geduld) das Formular auszufüllen. Um diese Motivation zu stärken, ist es sinnvoll – wie im Beispiel – noch einmal auf die Vorteile der Registrierung hinzuweisen.

Aus der Motivation, was ein Besucher mit dieser Anmeldung bezweckt und aus seiner Vorstellung, welche Daten dazu wohl notwendig sind, ergibt sich die *Bereitschaft, die einzelnen Felder zu füllen*. Die Relevanz mancher Felder ist jedem sofort einsichtig, andere scheinen eher überflüssig. Diese Zuweisung hängt stark vom Zweck des Formulars ab (bei einer einfachen Newsletter-Bestellung könnte allein die E-Mail-Adresse reichen, bei einer Zahlungsabwicklung sind dagegen Kontodaten unerlässlich), aber auch die Einstellungen der einzelnen Besucher und Zielgruppen variieren stark. So kann es auf technik-affinen Seiten schon problematisch sein, nach einer Postleitzahl zu fragen, während auf Social-Web-Seiten wie Facebook oder StudiVZ die Abfrage von persönlichen Vorlieben nicht wundert. Der Betreiber muss diese Momente im Auge haben: Welche Daten brauche wirklich? Sehen auch die Besucher diese Notwendigkeit? Welche Daten hätte ich gerne zusätzlich? Haben diese Daten einen negativen Effekt auf die Rate der abgesendeten Formulare insgesamt? Die Markierung von Feldern als „Pflichtfelder“ ist sicher eine Möglichkeit, den Besuchern eine Wahlmöglichkeit zu lassen, prinzipiell senkt aber jedes zusätzliche Feld die Motivation des Besuchers, ein Formular auszufüllen. – Im Fall des obigen Kicker-Beispiels könnte also überlegt werden, ob die zusätzlichen Felder für Telefonnummer und Geburtsdatum – und damit die potenziell zusätzlichen Informationen in der Mitgliederdatenbank – eine, wenn auch geringe, Senkung der Registrierungen insgesamt lohnen.

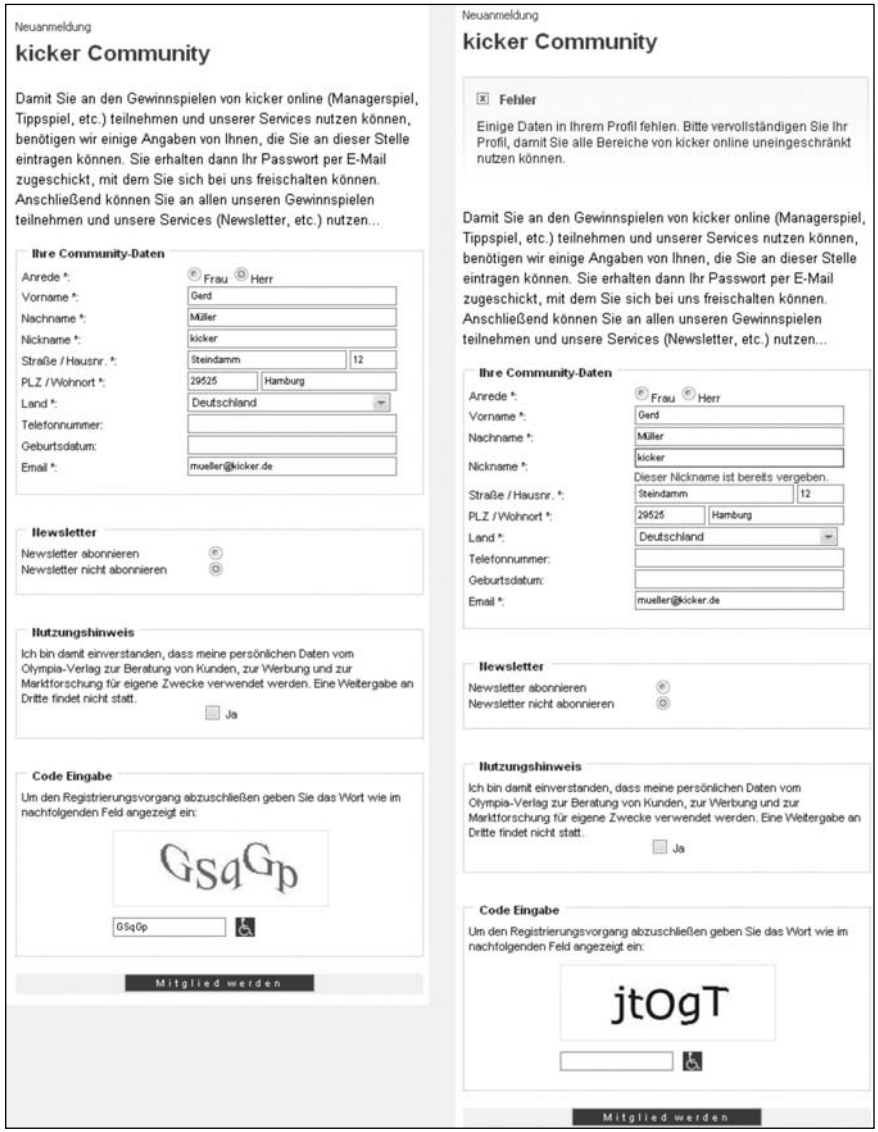


Abbildung 14: Anmeldeformular und Korrekturen bei www.kicker.de

Die Frage, ob die Besucher dieses Formulars überhaupt im *unmittelbaren Besitz der Informationen* sind, scheint im Kicker-Beispiel nicht besonders kritisch. Das Beispiel mit der Landing Page für Autoversicherungen weiter oben (Abbildung 11) ist da schon sensibler. Für die Kalkulation der Versicherungsprämie werden dort entweder Informationen aus dem Fahrzeugschein bzw. der Zulassungsbescheinigung benötigt oder detaillierte Informationen zum Modell des Kfz. Das kann durchaus zu häufigen Abbrüchen führen, da man ja gewöhnlich nicht mit seinem Kfz-Schein in der Tasche am PC sitzt.

Die Verantwortlichen für den Online-Auftritt beschäftigen sich gewöhnlich ausgiebig mit der Frage, welches Feld wie angelegt und abgefragt wird. Trotzdem kann es sein, dass die Besucher Schwierigkeiten mit dem Ausfüllen haben, obwohl sie eigentlich wissen, was dort hineingehört. So erfragen manche Betreiber die Hausnummer im gleichen Feld wie den Straßennamen, andere trennen diese Eingabe, in manchen Formularen werden Vor- und Nachname gemeinsam eingetragen, in anderen getrennt, mal dieser mal jener zuerst. Und allein für das Datumsfeld gibt es sicher mehr als zehn Format-Varianten. All diese Variationen führen unweigerlich zu Fehleingaben – aus Nachlässigkeit, Unlust sich auf das Formular zu konzentrieren oder Unverständnis. Dazu kommen Tippfehler und Unentschlossenheit. All diese Faktoren führen nicht nur zu direkten Abbrüchen, sondern auch zu unvollständig ausgefüllten Formularen, die den Besuchern Fehler zurückmelden. Diese Fehlerschleife kann mit ein wenig Sorgfalt so angenehm und unkompliziert wie möglich gehalten werden. Sehr hilfreich sind dabei direkte Rückmeldungen noch während der Besucher das Formular ausfüllt. In dem obigen Beispiel soll der Besucher sich etwa einen Nickname für die Community geben. Bei großen Communities kann es aber durchaus schwierig sein, einen Namen einzugeben, der noch nicht belegt ist. In diesem Beispiel war der Nickname bereits belegt, aber erst nachdem der Besucher das komplette Formular ausgefüllt und abgesendet hatte, wird dieser Fehler erkannt. Als Resultat muss jetzt noch einmal ein Name eingegeben werden und auch der Code am Ende der Seite muss erneut getippt werden. Zweiter Versuch: Wird dieser Name noch frei sein? (Oder wie oft soll ich dieses Formular noch bearbeiten?).

Aufgrund der Bedeutung der Formularseiten auf dem Weg vom unbekannten Besucher zum registrierten Mitglied, Newsletter-Abonnenten oder Kunden sollte die Konzeption der Seite, die Anzahl und das Format der Felder, die Fehlerrückmeldungen und die Gestaltung von Hinweisen und Buttons mit äußerster Sorgfalt konzipiert und – wenn möglich – mit Usability-Werkzeugen oder Traffic-Analysen analysiert und optimiert werden.

4. Interne Suchmaschinen

Interne Suchmaschinen, also die Möglichkeit für Besucher über ein Freitextfeld nach Inhalten auf der aktuellen Website zu suchen, haben in den letzten Jahren kontinuierlich an Bedeutung gewonnen. Das Freitextfeld ist sozusagen für sich schon ein revolutionäres Moment für den Zugang zu Informationen[19]. Über Jahrhunderte erforderte die nicht-digitale Natur der Inhalte ein raffiniertes System der Kategorisierung und Indizierung. Es brauchte großes Expertenwissen um diese Kategorisierungen zu konzipieren, zu pflegen und letztlich auch zu bedienen. Wer einmal in einer Bibliothek anhand von Karteikästen nach Literatur gesucht hat, wird wissen wovon die Sprache ist. Auch viele Websites pflegen ihre Art von inhaltlichen Hierarchien. Jedem Navigationsbaum liegt letztlich eine Art von Kategorisierung und Ordnung zu Grunde, die versucht, eine möglichst logische, widerspruchsfreie und einfache Abbildung der Inhalte in einer Hierarchie darzustellen. Je zahlreicher und vielfältiger allerdings diese zu verwaltenden Inhalte werden, desto komplizierter und vielschichtiger wird die Hierarchie, so dass es ab einem gewissen Punkt wiederum eine Anleitung braucht, um sich in den Strukturen zurechtzufinden. Freitextfelder sind nun – aus der Sicht des Nutzers gesehen – die Lösung für dieses Zugriffsproblem. Ein Stichwort eingegeben und schwupps – kommt die richtige Information. Um nicht missverstanden zu werden: Navigationsbäume und inhaltliche Hierarchien sind notwendig und haben ihren Nutzen, insbesondere dann, wenn den Besuchern aktiv Vorschläge gemacht werden sollen, wenn inhaltliche Verbindungen bestehen, die sich nicht leicht aus Titelstichwörtern ergeben oder sinnvolle Ergänzungen darstellen[20].

Mit der Entwicklung besserer allgemeiner Suchmaschinen (sprich: Google) hat sich das Freitextfeld als Zugriffswerkzeug allgemein durchgesetzt. Frühe Versuche, Web-Kataloge zu erstellen und zu pflegen um den Nutzern zuverlässig qualitätsvolle Inhalte zu liefern haben sich nicht durchsetzen können. Das Freitextfeld mit seinem „Suche absenden"-Button ist zum zentralen Werkzeug des Informationszuganges geworden.

Dieses Werkzeug findet sich auch auf sehr vielen einzelnen Websites und die Besucher nutzen es gerne und zunehmend. Für die Betreiber der Sites stellen sich mit der Implementierung zuverlässiger und effektiver Suchmaschinen neue Herausforderungen. Zwar ist

[19] lesenswert dazu: *The Search* von John Battelle, 2005

[20] Das „Tagging" von Inhalten im Internet stellt darum auch ein oft übersehenes, sehr wichtiges Moment des vielbeschworenen Web 2.0 dar. Die einzelnen Nutzer legen mit einer Bedeutungszuweisung („Tag", zu Deutsch etwa:) – die gewöhnlich explizit durch eine Begriffszuweisung erfolgt, aber auch durch ein implizit erfasstes Bewegungsmuster ersetzt/ergänzt werden kann – unabhängig von Experten oder abstrakter Logik, ein Netz von Verknüpfungen über die digitalen Inhalte.

der Einbau einer einfachen Stichwortsuche zunächst kein Problem, allerdings sind deren Ergebnisse oft irrelevant, schwer lesbar und unübersichtlich. Suchmaschinen auf kommerziellen Seiten dienen eben nicht primär dem Ziel, möglichst alle Treffer nach irgendwelchen objektiven Relevanzparametern zu präsentieren, sondern die Erwartungen der Nutzer schnell und umstandslos zu erfüllen und dabei die Interessen der Betreiber nicht aus dem Auge zu verlieren.

Zwei Fragen sind zu beachten:

Welche Ergebnisse werden präsentiert?
Die Bewertung der Relevanz von Suchtreffern ist eines der zentralen Merkmale der Qualität von Suchmaschinen. Suchende wollen selten alles finden, was zu dem Thema verfügbar ist, sondern haben meist etwas ganz bestimmtes im Sinn. Was dieses bestimmte ist, kann die Maschine nicht wissen, aber es braucht in der Konzeption eine Idee, was es sein könnte. Für Content-Anbieter käme etwa die Unterscheidung in aktuelle Artikel und ältere in Frage, Texte aus bestimmten Bereichen (Artikel, Schlagzeilen, Produktbewertungen Rezensionen) eher als andere. In Online-Shops ergeben sich Unterscheidungen über Produktkategorien („Handschuhe" aus den Bereichen Winterkleidung, Garten, Haushalt, Arbeit) oder aktuelle Angebote oder regionale Produkte. Ebenso kann in BtoB-Websites nach Produkten, Support oder Pressemitteilungen unterschieden werden.

Dazu geht es bei kommerziellen Websites nicht um eine möglichst objektive Abbildung des Bestandes, sondern um die Führung des Besuchers im Sinne des Betreibers. Welche Suchtreffer sind also für den Betreiber wichtiger als andere? Gute Suchmaschinen bieten ihren Betreibern verschiedene Möglichkeiten, Suchergebnisse zu gewichten, so dass zum Beispiel auf der Suche nach ‚Handschuhen' immer einer der meistverkauften Artikel ganz vorne dabei ist oder Sonderangebote höher bewertet werden.

Wie immer die Treffer zusammengestellt werden: Für den Besucher zählen meist nur die ersten fünf bis sieben. Ist unter diesen kein relevanter Treffer, werden die meisten Nutzer die Suche aufgeben oder ein neues Stichwort eingeben.

Wie werden Ergebnisse präsentiert?
Google hat einen Standard für die Darstellung von Suchergebnissen etabliert: Eine Überschrift, eine kurze Beschreibung und dann die Adresse der Quelle. Diese Dreiteilung wird von Nutzern erwartet und sie bietet den Betreibern genug Raum, um ihre Art von Überschrift und Beschreibung in den Suchtreffern unterzubringen. Die Online-Verantwortlichen

sollten darauf achten, dass nicht einfach zufällige Textauszüge aus dem Indizierungsalgorithmus als Beschreibung genutzt werden, sondern wenn möglich lieber gezielte Informationen (Preise, Beschreibungen, Bewertungen) in den Metadaten ablegen.

Ebenso lässt sich testen, wie viele Suchergebnisse eigentlich geliefert werden sollen und wie viele davon auf einer Seite. Ist es besser alle 1356 Treffer zum Stichwort „Handschuhe" zurückzugeben oder besser nur die zehn meistgeklickten? Will der Besucher lieber 25 Suchtreffer auf einer Seite sehen oder nach jeweils 10 weiterblättern? Probieren sie aus! Nutzen sie die Möglichkeit, den Besuchern auf die Finger zu schauen. Was tun die denn eigentlich auf den Suchergebnisseiten? Bis zu welcher Position zeigen sich nennenswerte Klickzahlen? Wie hoch ist der Anteil der Besucher, die sich für die Ergebnisse 11 bis 20 interessieren? Mit modernen Analytics-Tools lässt sich sehr gut das Verhalten der Besucher auf diesen Seiten beobachten: Wie groß sind die Absprungraten? Wie viele Suchvorgänge werden im Durchschnitt pro suchendem Nutzer durchgeführt?

Der Anteil der Besucher, die nicht über vorgedachte Navigationen zu ihrem Ziel finden wollen, sondern ein Freitextfeld nutzen, steigt deutlich. Und sie sind eine besonders interessante Besuchergruppe, denn diese Besucher wissen (ungefähr) was sie wollen und zeigen allgemein eine höhere Bereitschaft zur Konversion. Ein systematisches Austesten von Gestaltungsalternativen verspricht darum schnellen und deutlichen Erfolg.

5. Bestellprozesse

Bestellprozesse, vom Klick auf den „In den Warenkorb"-Button bis zum Abschluss der Bestellung, gehören zu den meistdiskutiertesten und mysteriösesten Elementen von Online-Shops und allen anderen Websites, die mehrstufige Konversionsprozesse benötigen. Jemand entscheidet sich Artikel A in den Warenkorb zu legen und bricht dann nach dem Aufruf der ersten Bestätigungsseite den Einkauf ab? Wie kann das sein? Was steckt dahinter? Das Bild von tausenden gefüllten Einkaufswagen, die verlassen in einem Supermarkt herumstehen verstärkt den Charakter des Rätsels. Aber wer keine kulturtheoretischen oder netz-anthropologischen Betrachtungen anstellen will, dem kann das Warum? erst mal egal sein. Wichtiger ist, wie die Abbruchraten reduziert werden können.

Es gibt einen reichen Bestand an Leitfäden und Gestaltungshinweisen, aber trotzdem (oder deswegen?) ungezählt viele Erscheinungsweisen des Bestellprozesses. In den meisten Fällen wird für Betreiber eine vollständige Umstellung des Prozesses erst einmal nicht in Frage

kommen, das bedeutet aber nicht, dass nicht Testläufe mit alternativen Schrittabfolgen oder Gestaltungen unmöglich wären.

Da es sich bei einem Bestellprozess eben nicht um eine einzelne zu optimierende Seite handelt, sondern um eine Abfolge von Schritten, ist es wichtig, Optimierungsparameter über die einzelnen Schritte hinweg festzulegen. Teilweise sind auch Nutzungsvorgänge zu berücksichtigen, die über einzelne Besuche hinausgehen (Fortsetzung von Bestellprozessen bei Wiederholungsbesuchen). Als Testverfahren bei größeren Online-Angeboten bieten sich eventuell zunächst einmal Durchläufe in Usability-Labors mit ausgewählten Nutzergruppen und Aufgabenstellungen an. Aber auch professionelle Testplattformen aus dem Bereich multivariater Verfahren bieten Möglichkeiten komplexere Nutzungsvorgänge abzubilden.

Unabhängig von einem konkreten Optimierungsverfahren können Veränderungen unter folgenden Gesichtspunkt erwogen werden:

Erst den Klick auf „Mein Warenkorb" abwarten oder gleich in den Bestellprozess?

Viele Online-Shops möchten ihre Besucher nicht direkt nach der Wahl eines Artikels in den Bestellprozess befördern. Unter der Annahme, dass die Besucher möglicherweise noch weitere Artikel suchen wollen oder es generell nicht schätzen, zu etwas gedrängt zu werden, aktualisieren diese Online-Shops nur im Hintergrund des Warenkorb-Bestand und warten dann auf den Klick auf „Zur Kasse" durch den Besucher. Zu diskutieren, ob dies grundsätzlich besser oder schlechter ist, ist müßig, ein Ausprobieren der jeweils anderen Variante kann aber für den eigenen Online-Shop und die eigene Zielgruppe klare Antworten geben. Vorurteilsloses Testen ist auch in diesem Fall sehr erfolgversprechend.

Registrierung und Bestellung

Besucher, die bereits etwas bestellt haben, müssen in vielen Online-Shops nicht jedesmal wieder ihre Adress- und Zahlungsdaten eingeben. Sind die Eingaben einmal gespeichert, können diese Besucher mit einer einfachen Nutzername/Passwort-Kombination auf ihr Konto zugreifen und die Bestellung so komfortabel und schnell durchführen. Da die Websites aber oft bereits registrierte Kunden nicht automatisch erkenn kann (wegen fehlender Cookies) steht am Beginn des Bestellprozesses in vielen Online-Shops die Frage: Sind Sie bereits Kunde oder wollen Sie sich neu registrieren? Die genaue Formulierung und Gestaltung dieser Abfrage sollte wohlüberlegt sein. So soll hier für bekannte Kunde ein möglichst einfacher und klarer Durchgang verfügbar sein, andererseits sollten neue Kunden

nicht umständlich umgeleitet und durch aufwendige Abfragen abgeschreckt werden. Zu beachten ist, dass der Besucher zu diesem Zeitpunkt entschlossen ist, einen bestimmten Artikel zu bestellen – und sich mehr oder weniger im Klaren über die dafür notwendigen Informationen ist – er aber nicht primär ein Kundenkonto eröffnen, sich registrieren oder ein Profil anlegen will, sondern eben schlicht eine Bestellung aufgeben will. In den folgenden zwei Abbildungen wird sichtbar, wie unterschiedlich diese Klippe auf dem Weg zur Konversion angegangen wird. Im ersten Fall, auf der Website von falconstein.de, bestimmt der Begriff Kundenkonto deutlich den Vorgang. Die Idee des Besuchers einen Artikel zu bestellen tritt in den Hintergrund und die Beziehung zum Anbieter dominiert.

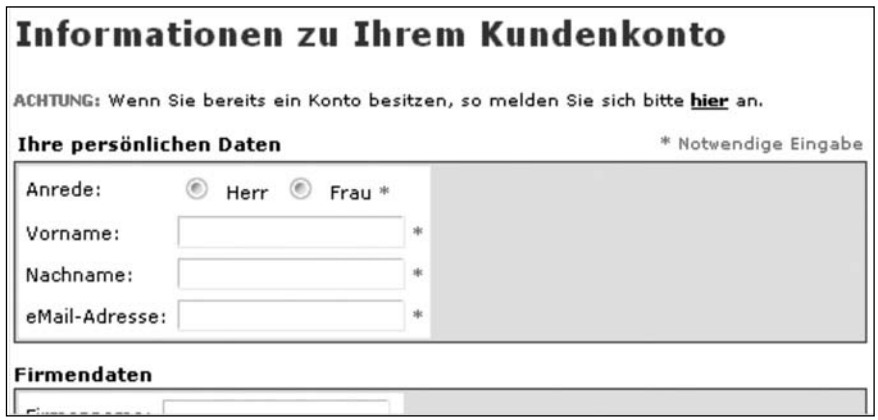


Abbildung 15: Eröffnung eines Kundenkontos bei falconstein.de

Abbildung 16: Neuanmeldung bei quelle.de

Bei quelle.de werden die verschiedenen Anforderungen von Neukunden und Stammkunden über eine Auswahl im oberen Bereich angesprochen, die Durchführung einer Bestellung bleibt aber das bestimmende Thema der Seite. Dazu findet der Besucher hier schon eine klare Vorstellung davon, in wie vielen Schritten er zu seiner Bestellung kommt.

Die Führung des Besuchers durch den Bestellprozess ist eine sensible Angelegenheit, bei die Formulierung, Platzierung und Gestaltung von Überschriften, kleinen Hinweistexten eine entscheidende Rolle spielen können. Genau diese Elemente können mit systematischen Testverfahren gezielt optimiert werden.

Anzahl der Felder und Übersichtlichkeit

Die Formulare in den letzten beiden Abbildungen bei quelle.de und insbesondere bei falconstein.de zeigen nur den oberen Abschnitt der Seiten. Tatsächlich begegnen dem Neukunden hier recht lange Formularseiten mit vielen Feldern. Eine Reihe von Angaben ist unerlässlich für eine Bestellung. Verzichten die Anbieter auf nur sporadisch wichtige Informationen wie Geburtsdaten, Faxnummern und ähnliche ‚freiwillige Angaben', bleiben immer noch 10 bis 15 Felder auszufüllen. Dazu kommen die Zahlungsdaten. Trotz dieser Ähnlichkeit in den Anforderungen ist es doch erstaunlich, wie stark allein der Eindruck von Übersichtlichkeit und Transparenz variiert. Durch kleine Änderungen in der Farbwahl und der Strukturierung der Inhalte kann schon viel gewonnen werden.

Zusätzliche Angebote und Links? Volle Navigation oder Tunnel zum Abschluss?

Eine wichtige Frage stellt die Anzeige von zusätzlichen Artikeln und möglicherweise sogar der vollen Navigation auch im Bestellprozess dar. Durch die Verfügbarkeit der ganzen Produktpalette aber auch durch Hilfsangebote und Beschreibungen zu Liefer- und Zahlungsbedingungen bleiben dem Nutzer alle Optionen offen. Dazu kommt für die Betreiber die Möglichkeit durch ergänzende Artikel eine Art digitaler ‚Mitnehm'-Zone einzurichten und das Volumen der Bestellvorgänge zu steigern. Dem steht aber die Möglichkeit entgegen, dass der Besucher nach weiteren Produkten sucht, Informationen studiert und dann am Ende seinen Kurs verliert – und gar nichts mehr bestellt.

Jeder Betreiber eines Online-Shops muss für sich die richtige Balance zwischen den beiden Extremen finden und durch ausgewogene Angebote ein Optimum für sich herauszuholen. Die Ermittlung dieses optimalen Designs ist aber theoretisch-konzeptionell allein kaum zu schaffen, ein sorgfältiges Austesten der verschiedenen Optionen ist daher sehr empfehlenswert.

Abbildung 17: amazon.de verzichtet vollständig auf zusätzliche Angebot und Links

Gestaltung von Buttons, Formularfeldern, Hinweisen, Überschriften

Viele einzelne Elemente sind bereits angesprochen worden. Und doch noch einmal: Kleine Veränderungen bei Überschriften, Hinweistexten, Button- oder Formulargestaltungen können große Effekte auslösen. Es ist eben keine Frage des Stils, der Ästhetik, des Styleguides oder des persönlichen Geschmacks, wo der „weiter"-Button positioniert ist (oben und/oder unten, links oder rechts), welche Farbe er hat oder wie er beschriftet ist, sondern eine Frage der Page-Performance. Alle Erfahrungen zeigen, dass Experten in der Mehrzahl der Fälle nicht richtig bei der Einschätzung der besten Überschrift, des besten Buttons und der anderen Elemente liegen – auch hier gilt: probieren statt diskutieren.

V. Optimierung in der Konzeptphase: Mit Nutzern gestalten

von Daniel R. Schmeißer

1. Besonderheiten der Konzeptphase 83
2. Verfahren zur Optimierung in der Konzeptphase..... 87
3. Fazit: Von der Idee zum erfolgreichen Konzept 108
4. Literaturverzeichnis 108

Wenn von Website-Optimierung die Rede ist, denkt man automatisch an die nachgelagerte Optimierung vorhandener Funktionalitäten und Prozesse im Anschluss an evaluative Maßnahmen. Dieses Verständnis von Optimierung (und im Speziellen von Usability-Optimierung) äußert sich in der Praxis darin, dass Websites häufig erst in einen Nutzertest gegeben werden, wenn sie bereits (fast) fertig entwickelt sind. Die Nachteile dieser Haltung liegen auf der Hand, denn konzeptuelle Schwachstellen, die zu einem späten Zeitpunkt aufgedeckt werden, verursachen in der Regel viel Zeitaufwand und Kosten, wenn sie geändert werden sollen. Wäre es also nicht hilfreicher, Konzepte so nahe an den Bedürfnissen ihrer potenziellen Nutzer zu entwickeln, dass Schwachstellen im späteren Entwicklungsprozess gar nicht mehr auftreten? Was aber bedeutet Optimierung in der Konzeptphase, wenn konkrete Anwendungen noch gar nicht vorhanden oder ausgearbeitet sind – Nutzer also noch keine Vorstellung davon haben, wie die Website künftig aussehen wird?

In der Konzeptphase legen Sie die wesentlichen Determinanten Ihres Erfolges fest: Wie soll die Informationsarchitektur und Seitenstruktur später aussehen? Mit welcher Tonalität sprechen Sie Ihre Nutzer an? Welche Funktionalitäten und Services stellen Sie zur Verfügung? Alle Entscheidungen, die Sie in dieser Phase mehr oder weniger reflektiert treffen, wirken sich auf die spätere Entwicklung der Website und Akzeptanz in der Zielgruppe aus. Deshalb ist es entscheidend, potenzielle Nutzer bereits in dieser frühen Entwicklungsphase mit einzubeziehen.

Bei der Planung und Entwicklung einer neuen Website gibt es streng genommen keine klaren Markierungen, wann die „Konzeptphase" anfängt und wann sie endet. In den seltensten Fällen wird eine Website zum Launch von Grund auf neu geplant. Weitaus häufiger ist der Fall, dass Veränderungen sukzessive vorgenommen oder neue Produktideen parallel zu bereits bestehenden Anwendungen entwickelt und umgesetzt werden. Dieser Sachverhalt wird meist durch einen kreisförmigen, *iterativen Prozess* mit verschiedenen, aufeinander folgenden Phasen der Websiteentwicklung und -optimierung veranschaulicht (siehe Abbildung 18).

In der **Konzeptphase**, auf die dieses Kapitel eingeht, geht es um die erste Entwicklung von Konzepten, Layout-Entwürfen, Informationsarchitekturen und Prototypen. Sie unterscheidet sich grundlegend von der frühesten Phase der Website-Entwicklung – der Planungsphase. Diese ist überwiegend strategischer Natur und hat weniger mit der konkreten Ausgestaltung einer Website zu tun. In der Planungsphase beschäftigen Sie sich mit der Positionierung Ihres eigenen Angebots im Wettbewerbsumfeld, der Zielgruppenansprache und dem generellen Funktionsumfang. Es gibt eine ganze Reihe sinnvoller Forschungs-

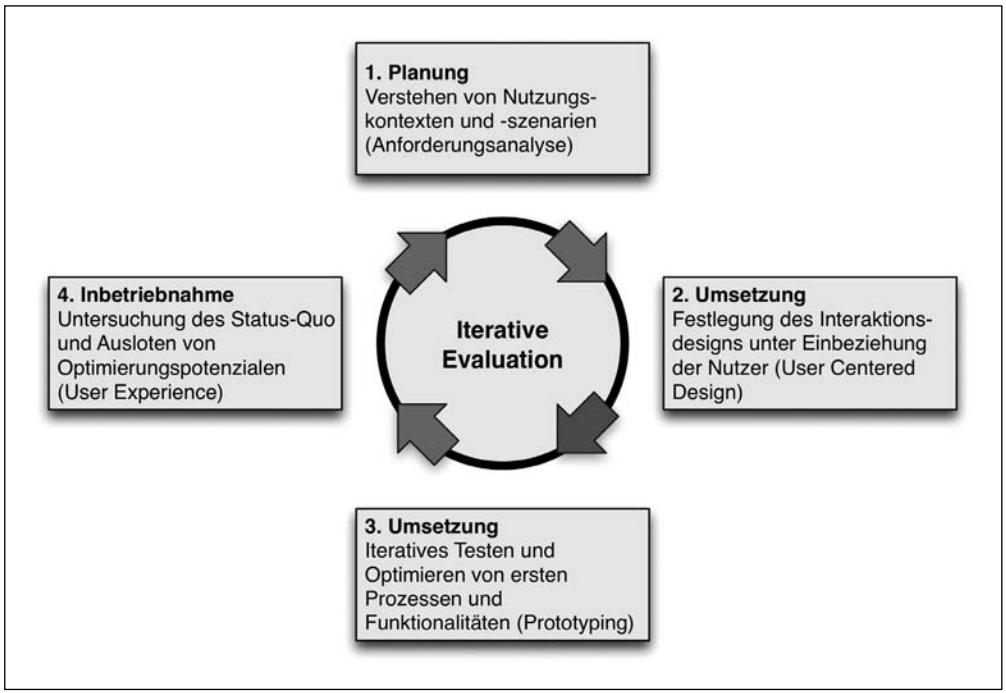


Abbildung 18: Phasenmodell der entwicklungsbegleitenden Website-Optimierung (Quelle: phaydon)

ansätze, die Sie in dieser Phase anwenden können und die Sie dabei unterstützen, die Ausrichtung und Planung Ihrer Website zielgerichtet anzugehen. Hierzu gehören beispielsweise Anforderungs- und Wettbewerbsanalysen, Marktscreenings sowie – unter Berücksichtigung potenzieller Zielgruppen – qualitative Motiv- und Bedarfsanalysen zur Ermittlung relevanter Markteintrittsfaktoren (Analyse von Benefits und Barriers). Da es sich bei den genannten Verfahren aber im engeren Sinne nicht um Optimierungsmaßnahmen handelt, die sich auch auf bereits bestehende Websites anwenden lassen, werden diese nur am Rande berücksichtigt. Fragestellungen, die in der Planungsphase und bei der Positionierung eines Produktes eine Rolle spielen, wirken sich allerdings auch auf die nachfolgende Konzeptphase aus – insofern gehen beide Phasen häufig fließend ineinander über.

1. Besonderheiten der Konzeptphase

Der skizzierte (idealtypische) Ablauf der Website-Entwicklung verdeutlicht, dass man im Entwicklungsprozess nie wirklich an ein Ende kommt: Sind alle Änderungen gelauncht, beschäftigt man sich bereits mit den nächsten Veränderungen und Anpassungen. Nutzer-

ströme verändern sich, die Nutzer selbst ändern ihr Verhalten, neue technische Features erweitern das mögliche Spektrum usw. Das bedeutet, dass auch die Methoden zur Optimierung flexibel und ergebnisoffen sein sollten und den Nutzer nicht „zwangsführen" dürfen. Bezogen auf Optimierungsansätze für die Konzeptphase geht es daher vor allem um die Art und Weise, wie Ergebnisse erzielt und Nutzer in den Entwicklungsprozess einbezogen werden.

Spezielle Anforderungen in der Konzeptphase

Optimierung in der Konzeptphase dreht sich zum Einen um die Frage, wie sich Ideen und Konzepte in frühen Phasen der Produktentwicklung so verbessern lassen, dass die weitere Ausarbeitung zielführend und nutzerzentriert erfolgt. Optimierungspotenzial ergibt sich dabei aus der frühen Einbeziehung der Nutzer, die ihre Erwartungen und Eindrücke äußern, und bezieht sich auf das Konzept selbst. Die Erkenntnisse aus einem derartigen **partizipativen Prozess** können unmittelbar in die Entwicklung einer Website einfließen. Zum Anderen wird in der Konzeptphase auch der Grundstein für den gesamten *Product Life Cycle* gelegt – Optimierung in dieser Phase bezieht sich immer auch auf den gesamten Entwicklungsprozess eines Produkts: Wie muss der Entwicklungsprozess selbst gestaltet sein, damit am Ende ein Produkt entsteht, das gerne und leicht genutzt wird?

In der Konzeptphase ergibt sich das Optimierungspotenzial durch die frühe Einbeziehung der Nutzer in den Entwicklungsprozess. Dieser Prozess wird als User Centered Design oder Participatory Design bezeichnet.

In der Praxis erweist es sich dabei häufig als schwierig, Verfahren zur Optimierung erster Produktideen in der Konzeptphase einzusetzen, da Entwickler, Produktmanager und Agenturen unter erheblichem Zeitdruck agieren und die verschiedenen, an der Konzeption beteiligten Abteilungen häufig mit unterschiedlichen Verantwortlichkeiten ausgestattet sind.

Hier ist eine enge Zusammenarbeit zwischen den Beteiligten erforderlich: Insbesondere der forscherische Part, der entweder In-House oder durch ein unabhängiges Forschungsinstitut erfolgt, sollte frühzeitig in den Entwicklungsprozess einbezogen werden, damit Entwicklungsschritte und Optimierungstools inhaltlich und zeitlich aufeinander abgestimmt werden können. Im Idealfall werden hierfür bei der zeitlichen Planung eines Projekts definierte Zeitpunkte festgelegt, an denen eine Überprüfung und Optimierung erfolgt.

- Typische Zeitpunkte, an denen Tests und Optimierungsphasen ansetzen:
- Definition der intendierten Zielgruppen und Use Cases: Personas und Szenarios
- Festlegung der Informationsarchitektur und Navigationsstruktur: Card Sorting
- Konzeption erster Design- und Layout-Entwürfe zur Zielgruppenansprache: Kreativ-Workshops

Auch wenn im Unternehmen eine grundsätzliche Bereitschaft besteht, Zwischenschritte der Optimierung im Entwicklungsprozess einzuplanen, erweist es sich in der Praxis immer wieder als schwierig, unter Umständen weitreichende Anregungen oder Kritikpunkte, die durch die Nutzer zur Sprache kommen, in der weiteren Entwicklung zu berücksichtigen. Hier spielen natürlich unternehmensinterne Restriktionen eine zentrale Rolle, häufig ist es aber auch die fehlende Zeit oder mangelnde Flexibilität, einmal vorgefasste Meinungen und Strategien neu zu überdenken, wenn sie sich nicht als zielführend erweisen.

Besonders bei neuen Konzepten, die aus der Offline-Welt in die Online-Welt übertragen werden, zeigen sich häufig bereits in frühen Phasen Stolpersteine, deren Nichtbehebung gravierende Auswirkungen auf die spätere Akzeptanz der Website haben kann. Deshalb ist es von entscheidender Bedeutung, dass der Optimierungsprozess in der Konzeptphase so weit wie möglich *ergebnisoffen* abläuft. Die Beteiligten sollten ermutigt werden, kreativ zu denken und auch spontane, abweichende Ideen zur Sprache zu bringen.

Innovative Konzepte und User-Tests: Ein Widerspruch?

Eine weitere Schwierigkeit in der Konzeptphase kommt hinzu, wenn es sich um innovative Konzepte handelt. Hier besteht auf Seiten von Agenturen und Unternehmen häufig die Ansicht, dass der Nutzer echte Innovationen nicht beurteilen könne, weil sie seiner erworbenen und liebgewonnenen Gewohnheit per se entgegenstünden und daher immer kritisch eingeschätzt würden. Empirische UX- oder Marktforschung behindere daher Kreativität und Innovation, weil sie Neuerungen aus dem Blickwinkel des Status Quo heraus beurteile. Dieser Vorbehalt trifft tatsächlich häufig zu, weil ein Großteil angewandter Forschung nicht angemessen – sprich nicht ergebnisoffen – durchgeführt wird. Wirkliche Innovationen bei Funktionalitäten, Content-Darstellung oder Navigationskonzepten setzen sich dann im User-Test nicht durch, wenn von vornherein Vorbehalte geäußert und im Zweifelsfall Empfehlungen für die „bewährte" Machart ausgesprochen werden.

Das muss aber nicht so sein, und der Einsatz geeigneter Verfahren kann dabei helfen, Innovationen von Ihren Nutzern so beurteilen zu lassen, dass eine valide Einschätzung über Akzeptanz und Nutzbarkeit möglich wird. Hierzu gehören die Einbeziehung des **Nut-**

zungskontextes einer Anwendung, sowie die Berücksichtigung non-verbaler, emotionaler und **impliziter Faktoren**, die über den Einsatz assoziativer, kreativer und projektiver Verfahren ermittelt werden können. Aber auch die **Wahl der Zielgruppe** ist entscheidend für die Beurteilung von Innovationen: Befragt man den ‚Otto-Normalverbraucher' oder sogenannte Frühadoptoren, also besonders aufgeschlossene, Technik-affine „First-Mover", bei denen man davon ausgehen kann, dass sie als Vorreiter eine Innovation besser einschätzen können als der Normalverbraucher? Eine schwierige Frage, die nicht pauschal beantwortet werden kann, aber immer wieder neu vor dem Hintergrund des sogenannten „Pro-Veränderungsbias" beurteilt werden sollte.[21]

Der Pro-Veränderungsbias beruht auf der Annahme, dass der Gebrauch innovativer Anwendungen per se den Nutzen für die Nutzer steigert. Dies kann der Fall sein, wenn die Website einen echten Mehrwert bietet, muss aber nicht so sein, zum Beispiel dann, wenn neue Anwendungen schwer erlernbar sind, nicht wirklich als echte Innovation wahrgenommen werden oder sich der Mehrwert nicht auf Anhieb erschließt. Es empfiehlt sich daher häufig, den „Normalverbraucher" und nicht die Frühadoptoren in den frühen Entwicklungsprozess einzubeziehen, zumal beide Gruppen nicht zwangsläufig über dieselben Werte, Erwartungen und Grundeinstellungen bei der Beurteilung neuer Produkte verfügen. Als grundlegende Zielsetzung sollte hier der Abbau der Nutzungsbarrieren durch die Schaffung möglichst intuitiver Konzepte und Nutzungsszenarien sichergestellt werden.

Wenn davon die Rede ist, „normale Nutzer" in den Entwicklungsprozess einzubeziehen, geht es nicht darum, Vorschläge der Nutzer 1 : 1 in den Entwicklungsprozess einzubauen, sondern aus den Urteilen und Einschätzungen der Nutzer die richtigen Hinweise für die Optimierung abzuleiten. Die Nutzer werden also nicht direkt nach ihren Wünschen gefragt („Was würden Sie verbessern?"), stattdessen werden Einstellungen, Urteile, Gedanken und Gefühle während der Nutzung bzw. in der Konfrontation mit einem neuen Konzept exploriert. Das Verfahren, mit dem valide Ergebnisse erzielt werden, ist daher indirekt und erfordert immer eine Interpretationsleistung, die gemeinsam zwischen den Projektbeteiligten erfolgt. Selbst für den Fall, dass Nutzer im Rahmen von Workshops kreativ werden, handelt es sich um ein indirektes Verfahren, das nicht die Arbeit der Agenturen ersetzt: Baut der Nutzer Scribbles oder stellt Collagen zusammen, bringt er damit Erwartungen, Bedürfnisse und Gefühle zum Ausdruck, die er vielleicht in dieser Form nicht verbalisieren könnte

[21] Siehe Königstorfer, Jörg/Gröppel-Klein, Andrea (2007): Projektive Verfahren zur Ermittlung der Akzeptanz technologischer Innovationen. Eine empirische Studie zu Internetanwendungen auf mobilen Endgeräten. In: Buber, Renate/Holzmüller, Hartmut H. (Hrsg.): Qualitative Marktforschung. Konzepte – Methoden – Analysen, Wiesbaden. S. 847-862.

und die dem Entwicklerteam gerade deshalb helfen, die eigenen Konzepte nutzerzentriert weiter zu entwickeln.

Die frühe Einbeziehung „typischer" Nutzer ersetzt nicht die Arbeit der Agenturen, sondern unterstützt Sie dabei, Konzepte nutzerorientiert umzusetzen. Hierbei werden mittels indirekter Verfahren die teils verborgenen und impliziten Bedürfnisse der Nutzer aufgedeckt und für die Optimierung genutzt.

2. Verfahren zur Optimierung in der Konzeptphase

Zunächst werden Verfahren vorgestellt, die eine Optimierung der zielgruppen-gerechten Entwicklung der Website zum Ziel haben (Personas und Szenarios). Danach erfahren Sie, wie Sie mit Hilfe von Kartenlegetechniken (Card-Sorting) die Informations- und Navigationsstruktur Ihrer Website nutzergerecht optimieren können. Abschließend werden offene Gruppenverfahren skizziert, die sowohl für Design- und Konzepttests, als auch für die kreative Entwicklung neuer Produktideen, Bildwelten und Ansprachetonalitäten eingesetzt werden können (Fokusgruppen und Kreativ-Workshops).

Personas – den Nutzern ein Gesicht geben

Thomas ist 39 Jahre, verheiratet, arbeitet als wissenschaftlicher Mitarbeiter an der Universität und spielt in seiner Freizeit gerne Handball – er ist jedoch kein realer Mensch, sondern eine fiktive Person, eine sogenannte Persona. Personas sind typische Nutzer, die die Zielgruppen Ihres Produkts repräsentieren. Sie repräsentieren Typen, aber sind doch konkret mit Namen, Merkmalen und Eigenschaften ausgestattet. Sie dürfen daher nicht mit klassischen Marketing-Zielgruppen, die aus repräsentativen Marktsegmentierungen resultieren, verwechselt werden. Zielgruppen oder Marktsegmente repräsentieren echte Anteile in der Nutzerschaft, die man mit einem Produkt ansprechen möchte. Zielgruppen-Merkmale sind jedoch häufig sehr abstrakt, was konkrete Eigenschaften, Vorlieben und Nutzungsgewohnheiten betrifft. Deshalb eignen sie sich zwar für eine strategische Ausrichtung und Positionierung, aber weniger für die konkrete und nutzerzentrierte Entwicklung einer Website.[22]

[22] Dynamische Marktsegmentierungen können in späteren Phasen der Website-Optimierung gewinnbringend eingesetzt werden, um die Positionierung und Zielgruppenansprache eines Angebots zielgerichtet auf einzelne Segmente auszurichten. Vgl. hierzu Schmeißer, Daniel et al. (2008): Effektive Kundenansprache in komplexen Märkten. Ergebnisse einer Segmentierung im Touristikmarkt. In: planung & analyse, Heft 6/2008, S. 31-37.

Mit Personas lässt sich hingegen ein sehr praktischer Zweck verfolgen. Sie geben dem oft unbekannten Nutzer einer Anwendung von Anfang an ein Gesicht – und dem Entwicklungsteam klare Vorstellungen darüber, was ihre späteren Nutzer auszeichnet. Personas stellen abstrakte Zielgruppeneigenschaften in anschaulichen und lebendigen Nutzertypen dar und helfen dadurch dem Entwicklerteam, ein gemeinsames Verständnis für die (potenzielle) Zielgruppe zu entwickeln und sich während des Entwicklungsprozesses in die Lage der Nutzer hineinzuversetzen („Was würde Thomas dazu sagen?"). Sie sind eine Hilfestellung bei der Priorisierung von Funktionen und Inhalten und ermöglichen die Fokussierung auf wichtige Zielsetzungen, indem sie die Diskussionen über die konkrete Zielgruppe anhand der typischen und anwendungsnahen Eigenschaften der Nutzer erleichtern.

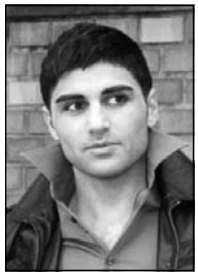
Personas repräsentieren eine große Zahl von Nutzern, gleichzeitig können sich die Entwickler auf eine konkrete Auswahl von Personen konzentrieren. Mit Hilfe von Personas lassen sich darüber hinaus typische Nutzungs- und Bedienungsszenarien sowie mögliche Barrieren in der Nutzung durchspielen.

Personas geben dem unbekannten Nutzer durch konkrete Typenbeschreibung ein Gesicht und dem Entwicklerteam eine Vorstellung davon, welche Wünsche, Ziele und Motive die späteren Nutzer auszeichnet.

Personas können auf Grundlage der bestehenden oder intendierten Nutzergruppe sowohl hypothetisch als auch empirisch mittels teilnehmender Beobachtung oder Tiefeninterviews generiert werden. In der Praxis empfiehlt sich ein kombiniertes Vorgehen, das sowohl vorhandene Zielgruppenbeschreibungen (Soziodemografie, Nutzerstrukturdaten, Marktsegmente), Interviews mit der Zielgruppe (zum Beispiel im Rahmen von Nutzungskontextanalysen In-Home), und hypothetisch-analytische Typenbildung (Anreicherung durch konkrete Eigenschaften und abgeleitete Nutzungsszenarien) mit einschließt (siehe die beispielhafte Darstellung von zwei Personas für den Bereich Consumer Electronics in Abbildung 19).

Neben der Einbeziehung von soziodemografischen und geografischen Daten fließen in die Bildung von Personas auch spezifische Einstellungen und Charaktereigenschaften ein. Die Konzeption von Personas verläuft somit sehr konkret und umfasst neben Namen, Alter, Aussehen meist auch Wünsche, Fähigkeiten, Meinungen und Hobbies. Benutzer lassen sich anhand verschiedener (Charakter-)Eigenschaften und Verhaltensweisen beschreiben:

Erkan, der Technik-affine Vielseher



Alter: 24 Jahre

Familienstand: ledig

Beruf: Student

Freizeit: Freunde treffen, Kino, Internet

„Ohne Fernseher und Festplattenrekorder würde ich mich zuhause langweilen."

Erkan ist ein Fernseh-Fan. Sein Plasma-Flatscreen ist sein ganzer Stolz und hat einen Ehrenplatz an der Wand seiner 40m²-Studentenbude. Immer wenn er für die Universität lernt, kocht oder telefoniert, läuft der Fernseher nebenbei.

Am liebsten schaut Erkan Fußball. Wenn sein Verein Bayern München ein Champions-League-Spiel absolviert, ist er nicht aus dem Haus zu locken. Fußball ist auch der Grund, weshalb Erkan ein Abonnement bei Premiere gebucht hat.

Damit Erkan keine Serie und keinen Film verpasst, hat er sich einen Festplattenrecorder angeschafft, mit dem er seine Lieblingssendungen standardmäßig für die Woche aufnimmt. Aufgenommene Filme brennt Erkan auch häufig für Freunde auf DVD.

Ruth, die Gelegenheits-Rekorderin

Alter: 53 Jahre

Familienstand: verheiratet, 3 Kinder

Beruf: Tierärztin

Freizeit: Backen, Tennis, Gemeinderat

„Zum Fernsehen fehlt mir häufig die Zeit, und die meisten Sendungen sind nichts für mich."

Ruth steht mit beiden Beinen im Leben und ist der Vorstand ihrer 5-köpfigen Familie. Wenn sie nicht gerade in der hauseigenen Praxis Tiere verarztet, macht sie den Haushalt oder engagiert sich in ihrer Gemeinde.

Für Fernsehen fehlt Ruth bisweilen die Zeit und das Interesse.

Fernsehprogramme sucht sie sich gezielt aus der TV-Zeitschrift. Da für sie interessante Sendungen auf ARTE und den Dritten Programmen häufig recht spät am Abend laufen, bittet sie ihren Mann die Sendung auf VHS aufzunehmen. Sie selbst möchte sich mit der Technik nicht weiter auseinandersetzen.

Auch wenn ihre Kinder auf eine moderne TV-Ausstattung mit Festplattenrekorder drängen, ist Ruth gegenüber Innovationen skeptisch und zufrieden mit Geräten, die sie grundlegend bedienen kann.

Abbildung 19: Exemplarisch ausgewählte Personas für die Verwendung von Consumer Electronics (Quelle: phaydon)

- soziodemografisch, zum Beispiel Alter, Geschlecht, Beruf, Einkommen
- psychografisch, zum Beispiel Einstellungen, Motive, Wünsche
- technografisch, zum Beispiel technische Ausstattung, Nutzungshäufigkeit PC, Internet und Consumer Electronics
- geografisch, zum Beispiel Wohnort, Land, Kultur
- verhaltensorientiert, zum Beispiel Nutzung bestimmter Produkte, Freizeit- und Konsumverhalten
- physisch orientiert, zum Beispiel motorische Fähigkeiten

Bei der Gruppierung und Detaillierung von Personas sollten möglichst prägnante und heterogene Typen resultieren. Dabei sollten die Merkmale innerhalb eines Typus möglichst ähnlich sein (*interne Homogenität* auf der Ebene des Typus) und sich die Typen möglichst stark voneinander unterscheiden (*externe Heterogenität* auf der Ebene der Typologie).[23] Für den praktischen Einsatz im Entwicklungsprozess erweist sich daher eine Anzahl von **vier bis maximal acht Personas** als sinnvoll, denn eine geringere Anzahl differenziert nicht mehr ausreichend (die Merkmalseigenschaften überlappen sich), eine zu große Anzahl dagegen verringert die Praxistauglichkeit durch irrelevante Merkmalsdifferenzierung. Zentral ist, dass sich die Personas in den nutzungsspezifischen einstellungs- und verhaltensrelevanten Eigenschaften unterscheiden.

Um Personas gewinnbringend im Entwicklungsprozess einzusetzen, empfiehlt sich darüber hinaus die Verknüpfung mit sogenannten Use Cases oder Nutzungsszenarios. Hierdurch werden Personas nicht nur statisch anhand ihrer Eigenschaften und Vorlieben beschrieben, sondern man legt konkrete Nutzungsziele (Use Cases) fest, die eine Persona typischerweise verfolgt, wenn sie zum Beispiel nach einem bestimmten Produkt sucht. Um im Beispiel zu bleiben: Thomas surft gerne nach Feierabend im Internet und bestellt regelmäßig Bücher und CDs online. Dabei geht er zielgerichtet vor, das heißt er weiß in der Regel genau, wonach er sucht. Eine gute Suchmaschine ist für ihn daher ein absolutes Muss. Der daraus resultierende Use-Case liegt auf der Hand, wenn es um die Konzeption eines Online-Buchshops gehen würde: Nutzer vom Typ „Thomas" erwarten eine effiziente Suchmaschine und strukturierte Suchergebnisseiten, weil sie zielgerichtet suchen.

[23] Vgl. zur Vorgehensweise qualitativer Typologien: Maas, Jörg/Walter, Melanie (2007): Anders auf die Daten schauen. Qualitative Typenbildung am Beispiel einer Reisestudie. In: planung & analyse, Heft 6/2007, S. 69-73.

Szenarios – mit Use Cases Prozesse optimieren

Bevor Funktionsumfang, Nutzungswege und Navigationsstruktur einer Website festgelegt werden, müssen alle am Gestaltungsprozess Beteiligten eine gemeinsame Vorstellung bezüglich der relevanten Anforderungen bekommen. Die Anforderungen, die aus dem verfügbaren Angebotsumfang auf der einen Seite und den Nutzererwartungen auf der anderen Seite resultieren, müssen in ein **konsistentes Produktnutzungskonzept** gegossen werden. Dabei unterscheidet man zwischen Use Cases (Anforderungen und Nutzungsziele), die einen Business-Case beschreiben und solchen, die eine Interaktion zwischen Nutzer und Anwendung beinhalten. Im Rahmen der nutzerzentrierten Entwicklung spielen die letztgenannten Use Cases eine zentrale Rolle, denn ihre systematische Berücksichtigung und die strikte Fokussierung auf die möglichen Ziele der Nutzer trägt maßgeblich zum späteren Erfolg der Website bei und verhindert, dass Diskussionen über technische und gestalterische Details von den eigentlichen Zielen ablenken.

Die übergeordneten Fragen lauten daher immer: *Was will der Nutzer auf Ihrer Seite erreichen?* (Use-Case) bzw. *Wie erreicht er dieses oder jenes Ziel?* (Szenario). Dabei können die Hauptszenarien in einzelne kleinere Szenarien unterteilt werden. Die einzelnen Teilschritte, die von den Nutzern durchlaufen werden, um das mit dem Use-Case verbundene Ziel zu erreichen, können wiederum detailliert und einzelnen Prozessebenen zugeordnet werden. Hauptszenarien können dann einzelnen Personas zugeordnet werden, damit man eine Vorstellung von den Kontext- und Zielgruppenbedingungen bei der Lösung der Aufgaben bekommt. Wie aber lassen sich die zentralen Nutzerziele ermitteln und gegebenenfalls den Zielgruppen und Personas zuordnen? Hier bieten sich mehrere Möglichkeiten an:

Analyse bestehender Nutzungsmuster

In der Regel beginnen Sie bei der Entwicklung oder beim Relaunch Ihrer Website nicht bei Null, sondern setzen auf vorangegangene Entwicklungen auf bzw. entwickeln diese weiter. Daher macht es Sinn, sich die Nutzungsmuster und Szenarien auf den jeweiligen Anwendungen anzuschauen. Hier können Web-Analytics-Tools eine wichtige Hilfestellung bieten, denn sie geben quantitative Hinweise auf typische Nutzungspfade und -muster. Zum Anderen existieren bereits die zentralen Business-Cases, die vor dem Hintergrund bisheriger Kennzahlen beurteilt werden können. Planen Sie einen Relaunch Ihrer Website, empfiehlt sich im ersten Schritt, die typischen Nutzungsmuster und Szenarien im Zustand des Status Quo festzuhalten und kritisch zu hinterfragen.

Hierfür eignen sich folgende Leitfragen[24]:

- Welche Szenarien führen die Nutzer tatsächlich und wie häufig durch?
- Welche Szenarien werden erfolgreich durchlaufen und an welchen Stellen gibt es Abbrüche?
- Welche Zeit und welche Vorkenntnisse benötigen die Nutzer, um bestimmte Ziele zu erreichen?
- Wie reagieren die Nutzer, wenn es zu Fehlern oder Abbrüchen kommt?

Auf dieser Basis können erfolgreiche Szenarien identifiziert und Schwachstellen kritisch hinterfragt werden. Unklar bleibt jedoch, was Nutzer tun, wenn sie ihre Ziele nicht erreichen, bzw. ob sie die Website überhaupt zum Erreichen ihrer Ziele nutzen. Deshalb ist es unumgänglich, auch die Nutzer selbst im Umgang mit der Website zu beobachten bzw. zu befragen.

Befragung und Beobachtung typischer Nutzergruppen

Um ein echtes Anforderungsprofil zu erhalten, müssen die Nutzer selbst beobachtet und befragt werden, insbesondere wenn es um die Neuentwicklung einer Website geht. Bei der Aufgabenanalyse sollte der Nutzer mit seinem Nutzungskontext und allen relevanten direkten und indirekten Einflussvariablen (zum Beispiel Medienkompetenz, Einkommen, Bildungsstand etc.) im Mittelpunkt stehen. Auch wenn die Nutzungskontext- und Anforderungsanalyse eigentlich noch in die vorgelagerte Planungsphase gehören, bieten sich entsprechende Verfahren auch zur Identifikation der relevanten Use Cases an. So ist besonders die Nutzungskontextanalyse von Bedeutung, denn hier werden die Nutzer in ihrem **natürlichen Umfeld** beobachtet, das heißt der reale Nutzungskontext und seine Wirkung auf die Nutzung einer Website finden explizit Eingang in die systematische Analyse und Formulierung der spezifischen Anforderungen.

Der Nutzungskontext wird durch den Benutzer selbst, die Aufgaben, die physikalische Anwendungsumgebung sowie durch soziale Faktoren bestimmt und spielt bei realen Gegenständen und Produkten wie Haushaltsgeräten, Maschinen und Unterhaltungselektronik sicherlich eine noch größere Rolle als bei Websites. Trotzdem können die Gegebenheiten des Nutzungskontextes wichtige Hinweise darauf liefern, auf welche Art und Weise eine Website genutzt wird und welche Ziele die Nutzer im Einzelnen verfolgen. Gerade bei der Entwicklung innovativer Anwendungen ist der Nutzungskontext entscheidend. Neue Anwendungen scheitern häufig daran, dass sie für die falschen Kontexte entwickelt oder in

[24] Vgl. Röse, Kerstin (2003): Task-Analyse. In: Heinsen, Sven/Vogt, Petra (Hrsg.): Usability praktisch umsetzen. Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte. München Wien. S. 98-114.

nicht-intendierten Kontexten genutzt werden. Hinweise auf Nutzungsbedingungen, Rezeptionsverfassungen und soziale Kontexte werden mittels **ethnografischer Verfahren** (In-Home-Interviews, teilnehmende Beobachtung, Tagebuchverfahren) ermittelt und fließen systematisch in die Entwicklung von Use Cases und Szenarien ein. [25]

Ableitung hypothetischer (Problem-)Szenarien

Ein wichtiger Bestandteil der Konzeptionsphase ist die Festlegung der relevanten Problemszenarien, die als Ausgangspunkt für die weitere Entwurfsspezifizierung gelten. Mit Problemszenarien werden keine Probleme bezeichnet, sondern typische Nutzungssituationen, die das Konzept prägen.

Die Erstellung möglichst vollständiger Szenarien dient dazu, einen Überblick über die zu entwickelnden Prozessabläufe und Features zu bekommen. Szenarien können auch in der Diskussion zwischen den Beteiligten im Unternehmen zur systematischen Produktentwicklung genutzt werden.

Schlüsselt man sämtliche Problemszenarien auf Basis vergangener Beobachtungen und Nutzerbefragungen auf und ergänzt diese durch hypothetische Szenarien, erhält man ein relativ vollständiges Set an Use Cases und Szenarien, das im nächsten Schritt priorisiert und mit Stärken und Schwächen versehen werden kann. Dabei stehen die Anforderungen an die User Experience im Vordergrund (vgl. hierzu detailliert Kapitel X).

„User Experience" bedeutet in der Konzeptphase, das Erleben der Nutzer in der Interaktion mit der Anwendung anhand der Hauptszenarien und Nutzerziele frühzeitig zu antizipieren und so zu optimieren, dass die Herstellung möglichst positiver Nutzungserlebnisse die weitere Entwicklung der Prozesse und Funktionalitäten bestimmt. Use Cases bilden aber auch die Grundlage für die Entwicklung einer tragfähigen Informationsarchitektur, die den Nutzer beim Erreichen seiner Ziele unterstützen soll. Hier können sogenannte Kartensortiertechniken helfen, die Informationsarchitektur auf die subjektive Logik und Zieldefinition der Nutzer auszurichten.

[25] Siehe hierzu im Detail: Schmeißer, Daniel/Joisten, Martina (2008): Mehr als das Gewohnte: Innovative Produktentwicklung durch ethnografische Marktforschung. In: planung & analyse, Heft 3/2008, S. 30-35.

Card Sorting – Informationsstrukturen intuitiv gestalten

Die User Experience einer Website „lebt" von einer intuitiven und erwartungs-konformen Informationsarchitektur. Eine Informationsarchitektur ist dann intuitiv, wenn sie nicht aufwändig erlernt, sondern von den Nutzern mühelos antizipiert werden kann. Nur dann ist sichergestellt, dass die Nutzer das finden, wonach sie suchen.

Bei der Entwicklung eines Navigationskonzepts werden häufig zwei Fehler begangen: Im einen Fall werden unternehmensinterne Strukturen (zum Beispiel Abteilungsbezeichnungen, Produkt- oder Katalogkategorien) einfach in Navigationsstrukturen übersetzt – dabei wird vernachlässigt, dass die Organisations- und Unternehmenslogik häufig wenig mit der subjektiven Logik der Nutzer gemein hat (es sei denn es handelt sich um Mitarbeiter des Unternehmens). Im anderen Fall werden auf dem Reißbrett Szenarien entwickelt, die der Businesslogik entsprechen und abbilden, wie Nutzer nach den definierten Business-Cases vorgehen sollten. Beide Herangehensweisen führen in der Praxis dazu, dass Informations- und Navigationsstrukturen nicht intuitiv verständlich sind, da sie nicht die subjektive Logik der Nutzer berücksichtigen. Die Logik der Nutzer folgt aber grundsätzlich anderen Prinzipien: Sie ist alltagsnah, intuitiv, habitualisiert und besteht aus Schemata und Skripten, die in der Praxis erlernt und nicht ohne weiteres verändert werden. Als Website-Betreiber können Sie natürlich dagegen angehen und die Nutzer zur konsistenten Verwendung Ihrer Produktkategorien „erziehen". Aber wollen Sie wirklich riskieren, dass sich der Nutzer nicht zurechtfindet oder gar verständnislos abwendet?

Das Card Sorting stellt ein Set an Techniken dar, mit dem Sie die subjektive Logik der Nutzer entschlüsseln können. Das Grundprinzip des Card-Sortings besteht zunächst darin, alle in Frage stehenden Rubriken und Begriffe auf einzelne Karten zu schreiben (zum Beispiel sämtliche Rubriken der Sekundärnavigation), wobei die Anzahl der Begriffe überschaubar bleiben sollte (maximal 50 bis 60 Begriffe), da sonst der Zeitaufwand und die gedankliche Komplexität für die Teilnehmer zu hoch werden. Diese Begriffe bilden die Ausgangsbasis für die weitere Befragung. Dabei können drei Herangehensweisen unterschieden werden, die einzeln, aber auch kombiniert miteinander durchgeführt werden können: der Verständnis-, Namens- und Klassifikationstest.

Durch das Sortieren, Priorisieren und Umbenennen von Begriffen, Rubriken und Oberkategorien sammeln Sie wichtige Informationen zur Struktur und zum Verständnis des künftigen Navigationskonzepts und können diese gezielt zur Optimierung von Menüstrukturen einsetzen.

Verständnistest

Im ersten Schritt können Sie überprüfen, wie die einzelnen Rubriken oder Begriffe, die Sie gewählt haben, von den Nutzern verstanden werden bzw. welche Assoziationen sie auslösen. Die Verständnisabfrage sollte dabei möglichst spontan erfolgen. Hierzu werden die Karten jeweils einzeln vorgelegt und der Befragte wird aufgefordert, möglichst spontan seine Assoziationen zu dem Begriff zu äußern. Die Äußerungen werden pro Rubrik notiert oder auf die Rückseite der Karten geschrieben, anschließend werden die Äußerungen gesammelt und hinsichtlich ihrer Verständlichkeit analysiert. Die Analyse erfolgt in der Regel nicht computergestützt, da die semantische Bedeutung in allen Nuancen erfasst und von den Beteiligten kommentiert werden sollte. Ziel dieses Vorgehens ist es, möglichst nur diejenigen Begriffe bzw. Rubrikbezeichnungen auszuwählen, die ein eindeutiges Verständnis nahelegen. Bei Begriffen, die mehrdeutig oder missverständlich sind, können bereits im Erhebungsprozess von den Befragten neue Begriffe entwickelt werden, die sukzessive in das Kartenset aufgenommen werden und von den nachfolgenden Befragten überprüft werden. Das Vorgehen entspricht daher einer gegenstandsbezogenen Theoriebildung mit offener Stichprobe: Man geht so lange vor und befragt so viele Teilnehmer, bis die Ergebnisse gesättigt sind, das heißt keine neuen Erkenntnisse mehr zutage gefördert werden und ein vollständiges Set an eindeutigen und verständlichen Begriffen vorliegt.

Namingtest (offenes Card-Sorting)

Liegt bereits ein fertiges Set an Rubriken bzw. Begriffen vor, geht es im offenen Card Sorting darum, dass jede Testperson alle Karten gruppiert und den so entstandenen Gruppen selbst sinnvolle Namen gibt. Geht es im Verständlichkeitstest also um die verständliche Benennung der Rubriken selbst, soll beim offenen Card Sorting auch die zugehörige Oberkategorie gefunden und benannt werden. Eine präzise Definition der Begriffe (die gegebenenfalls durch einen Verständlichkeitstest abgesichert wird), ist daher für diesen Schritt absolut erforderlich, denn die Ergebnisse des Card Sortings sind nur so gut wie das Material, auf das es aufbaut. Im Idealfall entsteht so eine vollständige Gruppierung aller Karten pro Teilnehmer. Diese werden (einschließlich der Karten, die gegebenenfalls nicht zugeordnet werden konnten) anschließend ausgezählt und inhaltlich erfasst.

Das skizzierte Verfahren ist induktiv, da die Teilnehmer auf Basis der Unterrubriken die Oberkategorien entwickeln. Man geht also davon aus, dass die Nutzer diejenigen Begriffe und Rubriken gruppieren, die sie als zusammengehörig erleben. Darin liegt aber auch die Schwierigkeit der Übertragbarkeit des Verfahrens in der Praxis, denn (a) orientieren sich Nutzer während sie navigieren immer im Kontext der gesamten Seite und erschließen sich die Bedeutung einzelner Begriffe aus dem Zusammenspiel mit anderen Begriffen und (b)

schauen sie häufig zunächst auf die Oberkategorien und leiten von ihnen aus ab, welche Rubriken sich darunter befinden müssten. Daher liefert das offene Card Sorting allein noch keine validen Ergebnisse. Die Begriffe sollten daher (a) zusätzlich in ihrem Kontext und im Zusammenspiel mit den anderen Begriffen überprüft werden (zum Beispiel durch einen abschließenden Vergleich aller Oberkategorien im Kontext der Website) und (b) auch umgekehrt die gefundenen Oberkategorien nochmals in einer Kontrollgruppe kreuzvalidiert werden, das heißt ähnlich wie im ersten Schritt auf Verständlichkeit überprüft werden.

Abbildung 20: Card Sorting zur Ermittlung einer intuitiven Informationsarchitektur (Quelle: phaydon).

Klassifikationstest (geschlossenes Card Sorting)

Beim geschlossenen Card Sorting sind die verschiedenen Oberkategorien bereits vorgegeben. Die Testpersonen werden also lediglich aufgefordert, die Karten auf die Kategorien zu verteilen, nach Möglichkeit vollständig. Dieses Verfahren empfiehlt sich, wenn die Oberkategorien aus unternehmensinternen Gründen nicht mehr geändert werden können oder wenn bereits ein offenes Card Sorting im Vorfeld durchgeführt wurde und man nun im zweiten Schritt die Verständlichkeit dieser Kategorien in Bezug auf definierte Unterkategorien überprüfen möchte. Bei bestimmten Fragestellungen hat es sich auch bewährt, die durch die Nutzer ermittelten und benannten Oberkategorien mit den durch das Unternehmen gesetzten abzugleichen. Der Abgleich erfolgt jeweils am Ende des Interviews und hat zum Ziel, das subjektive Verständnis mit den gegebenenfalls vorgegebenen Kategorien in Einklang zu bringen bzw. vorhandene Diskrepanzen festzuhalten.

Grundsätzlich sollte bei der Durchführung von Card Sorting Tests darauf geachtet werden, dass die Befragten eine klare Instruktion bekommen und die Inhalte auf den Karten gut lesbar (keine Handschrift) und möglichst prägnant benannt sind (keine Sätze, sondern klare Begriffe). Um überhaupt ein Set an denkbaren Begriffen zu entwickeln, empfehlen sich vorgelagerte Workshops, um die zu testenden Begriffe und Strukturen auszuarbeiten und festzulegen.

Die Erhebung erfolgt beim Card Sorting entweder mittels persönlicher Interviews im Studio oder computerunterstützt als Onlinebefragung. Je mehr Fragen und Begrifflichkeiten noch ungeklärt sind, desto eher bieten sich persönliche Interviews an, in denen die Teilnehmer vorab und im Anschluss an das Kartenlegen vertiefend zu ihrem subjektiven Verständnis, zu spontanen Assoziationen und alternativen Wordings befragt werden. Hier geht es auch bei der Auswahl der Stichprobe nicht um ein repräsentatives Abbild der Nutzerschaft, sondern um funktionale Repräsentativität. Funktional repräsentativ sind Ergebnisse dann, wenn alle in Frage stehenden Facetten eines zu testenden Phänomens in der Analyse berücksichtigt werden. Es genügen daher kleinere Stichproben von 20 bis 40 Personen.[26]

Liegen bereits prägnante Begriffe vor, die lediglich vorhandenen Kategorien zugeordnet oder gruppiert werden sollen, ist auch eine Durchführung zum Beispiel über flashbasierte Onlinetools denkbar. Die computergestützte Durchführung eines Card Sortings hat neben der schnellen und kosteneffektiven Durchführbarkeit den Vorteil, dass größere Stichproben berücksichtigt werden können, so dass die Ergebnisse im Idealfall repräsentativ für die Nutzerschaft sind.[27] Dieses Vorgehen sollte daher gewählt werden, wenn Stichprobenfehler vermieden werden sollen, zum Beispiel wenn durch einen Produktbereich sehr unterschiedliche Zielgruppen angesprochen werden oder wenn es für die Zuordnung der Rubriken keine inhaltlichen Begründungen gibt. Hier existiert eine ganze Reihe von Tools, mit denen sich das Card Sorting realisieren lässt. Allerdings ist zu beachten, dass der Aufwand im Vorfeld der Erhebungsphase nicht unerheblich ist. Lassen sich beim qualitativen Card Sorting im Interview Begriffe flexibel gegeneinander austauschen (iterativ-explora-

[26] In einem offenen Card Sorting mit 168 Teilnehmern gingen Tullis und Wood mittels Reliabilitätstests der Frage nach, ab welcher Stichprobengröße stabile Ergebnisse erzielt werden können. Über die Bildung von Subsamples ermittelten die Autoren, dass die Reliabilität der Ergebnisse des Card Sortings bereits mit 15-20 Teilnehmern bei über 90 Prozent liegt. Auch wenn die Ergebnisse nicht generalisierbar sind, leiten sich Anhaltspunkte für die geeignete Stichprobengröße ab. Diese sollte nicht unter 20 Teilnehmern liegen, ab 40 Teilnehmer ist allerdings keine signifikante Verbesserung der Ergebnisqualität mehr zu erwarten. Vgl. Tullis, Tim/Albert, Bill (2008): Measuring The User Experience. Collecting, Analyzing, and Presenting Usability Metrics. Elsevier Inc. S. 223.

[27] Vgl.: Lorenzen-Schmidt, Olde et al. (2008): Nutzerzentrierte Strukturierung des Produktbereichs beim Webauftritt der comdirect bank. In: Zeitschrift für interaktive und kooperative Medien (i-com), Heft 2/2008, S. 52-54.

tives Vorgehen), muss bei der computergestützten Methode vorab genau überlegt werden, welche Begrifflichkeiten man vorgibt.

Menüstrukturen können durch eine Kombination aus Verständlichkeits- und Kategorisierungstest auf Basis der subjektiven Nutzerlogik optimiert werden. Sowohl bei der Erhebung, als auch in der Auswertung können computergestützte Verfahren zur Unterstützung eingesetzt werden.

Eine weitere Herausforderung stellt auch die Auswertung von Card Sorting Experimenten dar. Ist der Umfang an Begriffen sehr groß, wird eine manuelle Auswertung der Daten äußerst komplex und unüberschaubar. Hier existieren mittlerweise gut einsetzbare Tools zur Auswertung. CASOLYSIS ist beispielsweise eine frei verfügbare, plattformunabhängige Java-Anwendung, mit der beliebige Datensätze aus Card Sorting Tests analysiert werden können.[28] Wichtig bei der Auswertung ist die Transparenz des Verfahrens. Im Idealfall ermöglicht die Auswertungssoftware das Experimentieren mit den Datensätzen, so dass unterschiedliche Ergebnisvarianten und Zuordnungen ausprobiert und gerechnet werden können. Als statistisches Auswertungsverfahren bei offenen Card Sorting Tests kommt in der Regel die Cluster-Analyse zum Einsatz (zum Beispiel hierarchische Ward-Analyse). Diese bestimmt die Nähe der Kartengruppen zueinander (Cluster) durch eine Bewertungsfunktion und ermittelt so, welche Begriffe zusammengehören und welche nicht. Clusteranalysen erlauben aber nur bedingt Ableitungen darüber, welcher Kategorie sie zugeordnet werden müssen und wie die Kategorie benannt werden sollte. Die Kategorienbezeichnungen müssen daher auf Basis der vorhandenen Struktur und intendierter Ziele von den Projektbeteiligten und Experten bestimmt werden.

Es existieren eine Reihe weiterer Ansätze (Distanzmatrix, DBSCAN-Verfahren), die eine Auswertung von Datensätzen erlauben, diese sollen aber an dieser Stelle nicht im Detail vorgestellt werden. Entscheidend ist, dass Sie die Ergebnisse automatisierter Analysen immer mit den Projektbeteiligten besprechen und auf ihre Sinnhaftigkeit hin diskutieren. Hier bieten sich begleitende Workshops an, in denen verschiedene Naming-Alternativen oder Clusterlösungen auch mit den inhaltlich-strategischen Vorhaben abgeglichen werden. Qualitative Lösungen, die offen-explorativ angelegt sind, haben hier den Vorteil, dass die Projektbeteiligten und Nutzer von Anfang an stärker in die Analyse einbezogen sind, mithin der gesamte Analyseprozess iterativ verläuft. Ein offen-flexibles Vorgehen erweist sich

[28] Emmrich, Sebastian/Szwillus, Gerd (2008): Auswertung von Card-Sorting-Experimenten mit CASOLYSIS. In: Brau, Henning et al. (Hrsg.): Usability Professionals 2008, S. 149-152.

häufig als zielführend, denn in der Praxis zeigt sich immer wieder, dass auf exaktem Weg gewonnene Lösungen nicht zwangsläufig den Anforderungen und Vorgaben der Unternehmen entsprechen. Im gesamten Prozess müssen daher die Unternehmens- und Nutzerperspektive in gleichem Maße berücksichtigt werden.

Workshops – Konzepte und Layouts überprüfen

Qualitative Gruppenverfahren werden im Rahmen der Website-Entwicklung und Optimierung leider noch vergleichsweise selten eingesetzt, obwohl sie bei vielen Fragestellungen klare Vorteile gegenüber Einzelerhebungen (Interviews) aufweisen. Das liegt zum Einen daran, dass Fokusgruppen, Gruppendiskussionen oder Kreativ-Workshops zu den klassischen Methoden der Marktforschung gehören und eine Berücksichtigung der besonderen Anforderungen für die User Experience-Forschung noch nicht systematisch ausgearbeitet wurde. Zum Anderen werden Fokusgruppen aber häufig auch als „quick-and-dirty" Verfahren missverstanden, von denen man sich erhofft, in kurzer Zeit preisgünstig Ergebnisse zu bekommen. Sie werden aber im weiteren Verlauf sehen, dass Fokusgruppen eine sehr komplexe Methodik beinhalten und hohe Anforderungen an die Vorbereitung, Durchführung (Moderation) und den Analyseprozess stellen.

Zunächst lassen sich verschiedene Arten von Gruppenverfahren unterscheiden, die in der Konzeptphase sinnvoll eingesetzt werden können.[29] Eine einheitliche Definition existiert nicht, aber es lassen sich je nach Vorgehensweise, Inhalt und Moderationsstil differierende Gruppenverfahren für den Einsatz in der Konzeptphase benennen:

Gruppendiskussion: Dabei handelt es sich um die am häufigsten angewendete Form einer offenen und explorativen Gruppenerhebung, in der die Teilnehmer (maximal acht bis zehn) vorgegebene Themen relativ frei angehen und sich untereinander austauschen (Dauer circa zwei Stunden). Dabei werden die unterschiedlichen, subjektiven Wahrnehmungs- und Erlebnisformen aus verschiedenen Perspektiven betrachtet, unter starker Einbeziehung der Fantasie, von Symbolen, bildhaften Materialien und projektiven Aussagen.

Fokusgruppen: Die sogenannte „Focus Group" ist amerikanischer Herkunft und im Vergleich zur klassischen Gruppendiskussion eher output-orientiert (zehn bis zwölf Teilnehmer, Dauer circa 1,5 bis 2 Stunden). Die Reihenfolge, in der die Themen diskutiert werden, ist genau festgelegt, ebenso wie die Verweildauer bei einem Thema. Es geht in erster Linie

[29] Ein allgemeiner Überblick findet sich bei Blank, R. (2007): Gruppendiskussionsverfahren. In: Naderer, G./Balzer, E. (Hrsg.): Qualitative Marktforschung in Theorie und Praxis, Gabler. S. 279-299.

Abbildung 21: Gruppendiskussion mit einem Moderator und acht Teilnehmern. (Quelle: phaydon).

um das Einholen eines schnellen Meinungsbildes, zum Beispiel durch Handheben bei der Abfrage bestimmter Design- und Layout-Konzeptvarianten.

Kreativ-Workshops stellen eine erweiterte Form der Gruppendiskussion dar, die in der Regel drei bis fünf Stunden dauert und den Einsatz verschiedener Kreativtechniken umfasst. Durch die aktive Einbindung von Nutzern sollen Flops in der Produkt- und Konzeptentwicklung vermieden werden, weshalb im Kreativ-Workshop alles erlaubt ist, unabhängig davon, ob die generierten Ideen zukünftig durchführbar oder generell sinnvoll sind. Die intensive Auseinandersetzung mit dem Untersuchungsgegenstand fördert die „Mitmach"-Motivation der Befragten, wodurch im Vergleich zu reinen Befragungen validere Erkenntnisse über Bewertungsdimensionen und Bedürfnisstrukturen gewonnen werden. Zudem können Projektbeteiligte und Designer selbst am Workshop teilnehmen und im Rahmen von Kleingruppen die kreative Arbeit der Teilnehmer anleiten bzw. unterstützen.

Mini-Gruppen oder **Workshops** mit sechs bis acht Teilnehmern erlauben einen intensiven Austausch der Teilnehmer untereinander und eignen sich daher vor allem für den B2B-Bereich, bei Experten-Befragungen oder für unternehmensinterne Veranstaltungen, in denen verschiedene Leitungsebenen oder Abteilungen miteinander diskutieren sollen. Workshops werden im Rahmen von Nutzertests häufig eingesetzt, wenn sich die Teilnehmer intensiver untereinander austauschen sollen (zum Beispiel durch Kleingruppenarbeit) oder wenn digitale Präsentationstechniken zum Einsatz kommen (zum Beispiel Bildschirmgebrauch mit Remote-Steuerung durch den Moderator).

Dyaden: In dyadischen Interviews werden zwei Personen zu einem gemeinsamen Gespräch eingeladen, das von einem Interviewer moderiert wird. Diese Gesprächskonstellation bietet eine ausgeprägte Intensität der interpersonalen Beziehung und Interaktion, zum Beispiel zwischen Paaren, Freunden oder Experten. Die Befragten werden motiviert, während des Interviews einen Perspektivenwechsel vorzunehmen und die eigene Sicht kritisch zu hinterfragen, was in der Regel besser gelingt als im Einzelinterview. Dyaden bieten so die Chance, Themengebiete aus verschiedenen Perspektiven zu beleuchten und soziale Meinungsbildungsprozesse auf Mikro-Ebene abzubilden.

Abbildung 22: Gruppendiskussion mit einem Moderator und acht Teilnehmern. (Quelle: phaydon).

Das generelle Ziel von Gruppenverfahren besteht darin, unter Vorgabe eines bestimmten „Grundreizes" (zum Beispiel ein Verbalkonzept, Layout-Entwürfe) eine Diskussion anzustoßen, um möglichst weitreichende Informationen zum Thema zu sammeln, ein breites Set an Meinungen und Einstellungen zu erkunden sowie gruppenspezifische Verhaltensweisen zu beobachten, um Rückschlüsse auf die soziale Relevanz eines Produkts zu erhalten. Der Vorteil der Gruppensituation besteht darin, dass über das bewusste Zulassen der Gruppendynamik „tiefer liegende" Strukturen und Emotionen stimuliert werden, die Dynamik selbst zum Abbau psychischer Kontrollen beiträgt und spontane Reaktionen provoziert. Sie eignen sich daher bei allen Fragestellungen, in denen es um psychologisch-relevante Nutzungsfaktoren, die Markenwahrnehmung oder das kommunikative und visuelle Konzept einer Website geht.

Natürlich gibt es auch Nachteile, weshalb Sie Gruppenverfahren nicht bei allen Fragestellungen einsetzen sollten. Geht es zum Beispiel um Verständnisfragen, können in der Gruppensituation Verzerrungen aufgrund sozial erwünschten Antwortverhaltens oder durch Konformitätsdenken auftreten. Gruppenverfahren eignen sich auch nicht, wenn es um die Abbildung und Überprüfung erster Nutzungsprozesse geht, die von jedem Nutzer einzeln durchlaufen werden.

Um die Vorteile von Gruppenverfahren in der Konzeptphase optimal ausnutzen zu können, empfiehlt sich häufig ein kombiniertes oder sequentielles Vorgehen von zum Beispiel vorgelagerten Workshops und vertiefenden Einzelinterviews.

Durchführung von Gruppenverfahren

Bei der Durchführung von Gruppenverfahren ist die Festlegung der Ziele gemeinsam mit dem (internen oder externen) Auftraggeber im Vorfeld entscheidend. Geht es um die Überprüfung eines Konzepts, liegen verschiedene Entwürfe bereits vor oder sollen vertiefende „Insights" in der Nutzergruppe ermittelt werden, die in eine erste Konzeptentwicklung münden? Danach werden die geeigneten Verfahren zusammengestellt und gegebenenfalls weitere Moderationstechniken ausgewählt. Je weniger konkret das Material, das vorliegt und in die Gruppe hineingegeben werden kann, desto offener und kreativer sollte auch das gewählte Verfahren sein (Gruppendiskussionen, Kreativworkshops). Liegen bereits konkrete Konzepte oder Entwürfe einer Website vor, desto strukturierter kann das Verfahren ablaufen (Fokusgruppen, Workshops).

Grundsätzlich empfiehlt sich eine Anzahl von mindestens zwei bis drei Gruppen, idealerweise vier Gruppen, um zu validen Erkenntnissen zu kommen. Ist die Zielgruppe sehr heterogen, können die Gruppen auch nach Nutzersegmenten aufgeteilt werden (zum Beispiel zwei Gruppen mit Kunden und zwei mit Nicht-Kunden, jeweils unterteilt in zwei Altersgruppen). Die Gruppen selbst können *homogen* (ähnliche Zielgruppenmerkmale) oder heterogen zusammengesetzt sein. Heterogene Gruppen eignen sich vor allem, wenn Sie kontroverse Diskussionen anstoßen und gegebenenfalls festgefahrene Denkrichtungen verlassen möchten (zum Beispiel Verwender und Nicht-Verwender eines Angebots).

Gruppendiskussionen und Workshops können gewinnbringend für die Optimierung in frühen Phasen eingesetzt werden. Die Art der Durchführung sowie Anzahl und Zusammensetzung der Gruppen müssen aber individuell und abhängig von den Untersuchungszielen entschieden werden.

Die Qualität der Ergebnisse von Gruppenverfahren steht und fällt mit der **Qualifikation des Moderators** und der Güte des Diskussionsleitfadens. Dem Moderator fällt die Aufgabe zu, eine lockere und zugleich wertschätzende Gesprächssituation zu schaffen, in der die grundlegende Akzeptanz aller Beteiligten sowie methodische Neutralität gewährleistet sind. Er muss sämtliche Teilnehmer in die Gruppe integrieren und den dynamischen Prozess fördern, ohne die Rolle eines „Lehrers" einzunehmen. Er gibt die Struktur der Diskussion vor (alle für den Auftraggeber relevanten Informationen sollten angesprochen werden), hakt nach und vertieft einzelne Themen. Hierfür benötigt er einen sorgfältig ausgearbeiteten Leitfaden als groben Rahmen für die anzusprechenden Themen und den Ablauf der Diskussion. Darüber hinaus muss er ein gutes Gefühl für Timing und Sinn für Dramaturgie haben – alles Voraussetzungen, die nur mit einer entsprechenden Ausbildung, Supervision und Erfahrung erworben werden können.

Die Analyse von Gruppenverfahren ist in der Regel der aufwändigste Schritt und komplexer als die Auswertung zum Beispiel halbstandardisierter Einzelinterviews. Grundlage sind selektive oder vollständige wörtliche Protokolle der Diskussionsrunden. Auf dieser Basis werden gegenstandsbezogen typische Aussagen und Phänomene gesammelt und schrittweise und im Austausch mit dem Projektteam verdichtet. Bei offen angelegten Diskussionen (zum Beispiel Motiv- und Bedarfsanalysen) eignen sich Ansätze der Grounded Theory, die ein axiales Kodieren der Aussagen vorsehen.[30] Hierbei werden die Aussagen der Teilnehmer nach vier Dimensionen analysiert: Ausgehend vom beobachteten Phänomen werden Aussagen zu den ursächlichen Bedingungen, dem Kontext bzw. zu intervenierenden Bedingungen, zu den subjektiven Handlungsstrategien und deren Konsequenzen gesammelt. Ziel der Analyse ist die Ausformulierung erklärender Theorien bzw. Modelle und deren Überprüfung an den Daten. Der Interpretationsvorgang wird an dem Punkt abgebrochen, an dem die theoretische Sättigung erreicht ist, das heißt eine weitere Kodierung und Anreicherung der Kategorien keine neuen Erkenntnisse mehr liefert.

[30] Vgl. Glaser, Barney G./Strauss, Anselm (2008): Grounded Theory. Strategien qualitativer Forschung. (Übersetzung der Originalausgabe von 1967 unter dem Titel The Discovery of Grounded Theory: Strategies for Qualitative Research). Bern. Aufschlussreich auch die Weiterentwicklung der Grounded Theory in: Strauss, Anselm/Corbin, Juliet (1996): Grounded Theory: Grundlagen Qualitativer Sozialforschung. Weinheim.

Die Entscheidung, eine Fokusgruppe oder einen Workshop durchzuführen, muss inhaltlich begründet sein und die Vorteile der Gruppensituation angemessen berücksichtigen. Grundsätzlich lässt sich sagen, dass Gruppenverfahren immer dann eingesetzt werden sollten, wenn Sie sich durch den Austausch und die kontroverse Diskussion der Teilnehmer untereinander einen Mehrwert gegenüber der Einzelgesprächssituation erhoffen und wenn das vorliegende Material sich in einem frühen Entwicklungsstadium befindet, so dass funktionale Aspekte eher im Hintergrund stehen.

Im Folgenden werden drei typische Studiendesigns mit Gruppenverfahren vorgestellt, die häufig in der Konzeptphase Anwendung finden: die Motiv- und Bedarfsanalyse, die Nutzungskontextanalyse und der Konzept- und Designtest.

Motiv- und Bedarfsanalyse: Treiber und Barrieren der Nutzung identifizieren

Die Motiv- und Bedarfsanalyse wird häufig entwicklungsbegleitend oder zur Überprüfung des Status Quo durchgeführt. Das Ziel der Motiv- und Bedarfsanalyse besteht darin, die Nutzungsmotive der Nutzer, ihre (expliziten und impliziten) Wünsche und Bedürfnisse sowie Hemmnisse und Barrieren in der Nutzung bestimmter Angebote zu explorieren, um die künftige Website-Gestaltung optimal auszurichten.[31] Wenn Sie zum Beispiel einen Shop mit Markenkleidung konzipieren, könnte es im Rahmen einer vorgelagerten Motiv- und Bedarfsanalyse zunächst sehr offen und explorativ um die Frage gehen, wie die Nutzer normalerweise in ihrem Alltag vorgehen, wenn sie Markenartikel kaufen. Welche Gedanken und Gefühle gehen ihnen dabei durch den Kopf? Worauf achten sie? Welche Shops werden in der Offline- und Online-Welt aufgesucht? Was spricht die Nutzer an und wovon fühlen sie sich abgeschreckt? Die nachgelagerte Analyse unterscheidet systematisch zwischen positiven Anreizen (Benefits) und negativen Nutzungsbarrieren (Barriers).

Da sowohl Nutzungstreiber als auch Barrieren häufig unbewusst sind oder zumindest nicht auf Anhieb verbalisiert werden (können), ist es wichtig, auch assoziative, projektive und kreative Techniken bei der Erhebung einzusetzen, um an die tieferliegenden, vor- und unbewussten Bedürfnisse und Ängste heranzukommen. Da die Analyse vor dem Hintergrund psychologischer Theorien erfolgt, ist in der Regel eine professionelle Auswertung durch einen psychologisch geschulten Projektleiter erforderlich. Als Verfahren eignen sich hier Gruppendiskussionen und Kreativ-Workshops. Kreative Techniken werden eingesetzt, um zum Beispiel im Rahmen von Rollenspielen gewohnte Sichtweisen auf ein Phänomen auf-

[31] Siehe hierzu exemplarisch die Ergebnisse einer qualitativen Motiv- und Bedarfsanalyse in Rengelshausen, Oliver/Schmeißer, Daniel (2007): Touristik 2.0 – Chancen und Risiken von User Generated Content für den Online-Reisevertrieb. In: planung & analyse, Heft 4/2007, S. 58-61.

zubrechen und neue Denkrichtungen und Ideen anzustoßen. Die Technik „Disneys Denkstühle“ teilt zum Beispiel die Gruppe in drei Denkhaltungen auf (der Träumer, der Realist, der Kritiker). Jede Denkhaltung soll für das in Frage stehende Thema (zum Beispiel Entwicklung eines Markenshops) Konzepte erarbeiten und mit den anderen diskutieren. Der Träumer darf waghalsige Verbindungen aufstellen und sich das Unmögliche möglich machen; der Realist greift die Themen des Träumers auf und versucht sie in ein realistisches Konzept zu überführen; der Kritiker nimmt das Konzept des Realisten kritisch unter die Lupe. Anschließend werden auf Basis der Diskussion tragfähige Konzepte weiter erarbeitet.

Nutzungskontextanalyse: Verwendungszusammenhänge optimal nutzen

Eine Erweiterung der Motiv- und Bedarfsanalyse stellt die Nutzungskontextanalyse dar. Hier wird versucht, alle relevanten Informationen zum Verwendungskontext zusammenzutragen und anschließend in der Gruppe zu diskutieren. Da der natürliche Kontext entfällt, eignen sich Gruppenverfahren vor allem in Kombination mit anderen Verfahren. Häufig bietet sich zum Beispiel der ergänzende Einsatz von Tagebüchern an, die im Vorfeld der Gruppendiskussion von den Teilnehmern über einen bestimmten Zeitraum hinweg (in der Regel ein bis zwei Wochen) ausgefüllt werden und den Umgang mit dem Angebot oder die Internetnutzung im Hinblick auf ein eingegrenztes Thema festhalten (siehe Abbildung 23). In paper & pencil-Tagebüchern dokumentieren die Teilnehmer zum Beispiel ihren Alltag oder Situationen der Produktnutzung über mehrere Tage hinweg. Je nach Fragestellung wird das Tagebuchdesign sehr offen gehalten oder auf den Umgang bzw. die Erfahrung mit einem Produkt fokussiert. Es empfiehlt sich ein Tagebuchaufbau, der sowohl Freiraum für Kreativität lässt, als auch Strukturvorgaben (zum Beispiel Tagesablauf nach Zeitschema) macht. Bei Online-Tagebüchern schreiben die Teilnehmer PC-gestützt ihre Eindrücke und Erlebnisse nieder – diese Methode bietet sich nicht nur bei besonders online-affinen Zielgruppen an, sondern auch dann, wenn die Nutzung bestimmter Online-Angebote im Fokus steht.

Neben den ethnografischen Ansätzen, die vor allem auch die teilnehmende Beobachtung im natürlichen Kontext mit einschließt, auf die an dieser Stelle aber nicht näher eingegangen werden kann, werden Nutzungskontextanalysen in der Konzeptphase so durchgeführt, dass die Befragten ihre Erfahrungen und typischen Erlebnisse bei der Verwendung bestimmter Websites schildern und sich miteinander austauschen. Die Praxis zeigt, dass der Einsatz von Monitoren in der Gruppenerhebung sinnvoll ist; teils können hier durch den Moderator Websites gezeigt und von den Teilnehmern genutzt werden. Der Vorteil der Gruppenerhebung liegt darin, dass verschiedene Nutzungsszenarien auch anhand von Wettbewerberseiten durchgespielt werden können und sich die Teilnehmer zwischendurch

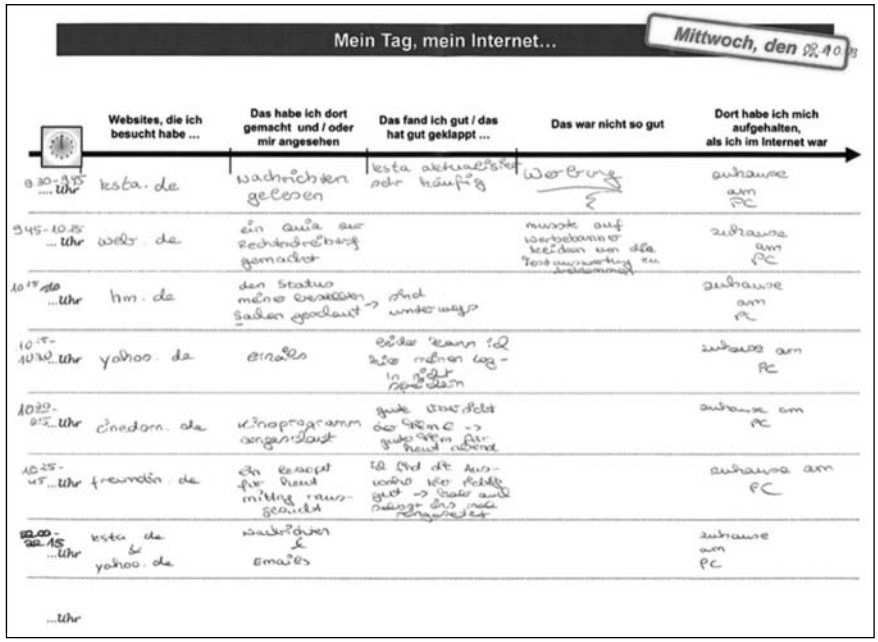


Abbildung 23: Ausschnitt aus einem Tagebuch mit tageweisen Einträgen zum Nutzungsverhalten des Internets bei jungen Frauen (Quelle: phaydon).

austauschen können. Dieses Vorgehen erfordert eine zugleich offene, aber doch kontrollierte Planung durch einen geschulten Moderator, der situativ in die Prozesse eingreift und versucht, möglichst den gesamten Nutzungskontext zu explorieren.

Konzept- und Designtest: Layout-Entwürfe testen und optimieren

Konzept- und Designtests werden eingesetzt, um die Akzeptanz eines oder mehrerer Konzeptentwürfe sowie die emotionale Wirkung von Layout-Entwürfen in der fokussierten Zielgruppe zu überprüfen. Konzepte liegen entweder als Verbalkonzepte vor, oder beinhalten idealerweise bereits erste visualisierte Entwürfe, Skizzen oder Dummies, an denen man die spätere Umsetzung zumindest in Ansätzen erkennt. Ziel eines Konzept- und Designtests ist die erste Überprüfung funktionaler Aspekte, im Vordergrund stehen aber häufig Fragen zur (emotionalen) Wirkung und Akzeptanz eines (Design-)Konzepts in der Zielgruppe, beispielsweise im Hinblick auf zentrale Dimensionen wie:

Erwartungskongruenz: Gemeint ist die Übereinstimmung der angedachten Services und Funktionalitäten mit den Erwartungen und Bedürfnissen der Nutzer. Um diese zu beurteilen, ist eine vorgelagerte Motiv- und Bedarfsanalyse erforderlich, die zunächst ohne

Vorlage eines Konzepts (ungestützt) in der ersten Phase einer Gruppendiskussion erfolgen kann.

Funktionaler Mehrwert: Hier steht die Frage im Vordergrund, inwieweit die geplante Website gegenüber der früheren Seite (Relaunch) bzw. anderen Angeboten des Wettbewerbs eine Erleichterung in funktionaler Hinsicht darstellt. Werden die Konzepte als echter Mehrwert wahrgenommen? Bieten sie künftig Leistungen, die als echter Nutzen erlebt werden?

Emotionaler Mehrwert: Es geht aber auch darum zu überprüfen, ob die Gestaltung der Seite die Nutzer emotional anspricht und auf impliziter Ebene die Bedürfnisse und Ängste der Nutzer ernst nimmt. Hier spielt die Verwendung künftiger Bildwelten eine zentrale Rolle, die bei Vorlage konkreter Designentwürfe bereits getestet und optimiert werden können.

Glaubwürdigkeit & Markenpassung: Die Website sollte bezüglich ihrer Inhalte, Angebote und Ansprachetonalitäten zur Marke passen. Mögliche Diskrepanzen im Markenerleben können bereits in der frühen Phase auf Basis von Designentwürfen erkannt und zielführend in der weiteren Entwicklung berücksichtigt werden.

Differenzierung & Faszinationspotenzial: Eine Website kann in sich noch so gut gestaltet sein – sie muss sich auch vom Wettbewerb abheben und über ein ausreichendes Faszinationspotenzial verfügen, um sich zu behaupten. Daher ist der Abgleich des eigenen Konzepts mit den Konzepten und Angeboten der Wettbewerber unumgänglich.

Gerade bei Vorliegen erster Designentwürfe bieten sich Gruppendiskussionen und Expertenworkshops für einen kreativen Austausch der Teilnehmer an. So können zum Beispiel verschiedene Layout-Varianten einer Homepage inklusive Bildwelten visualisiert und daraufhin getestet werden, ob die intendierte Tonalität erzielt wird. Gegebenenfalls können über Kreativtechniken (Collagen) neue Entwürfe und Bildwelten durch die Teilnehmer entwickelt werden und in die weitere Detaillierung des Konzepts einfließen. Über die Auswahl der Gruppenzusammensetzung lassen sich die Entwürfe dabei gezielt in den relevanten Nutzergruppen testen.

3. Fazit: Von der Idee zum erfolgreichen Konzept

In diesem Kapitel haben Sie eine Reihe qualitativer Verfahren kennengelernt, die sich in der Konzeptphase eignen, Potenziale für die Optimierung Ihrer Website zu ermitteln. Zusammengenommen bieten die vorgestellten Verfahren ein wirkungsvolles Instrumentarium, die richtigen Ideen in erfolgsversprechende Konzepte zu gießen. Entscheidend dabei ist, dass Sie sich im Vorfeld genau über die Fragestellung und das Ziel der Untersuchung klar werden, um das richtige Vorgehen zu wählen. Denn: Verfahren in der Konzeptphase sind schwerer zu planen als in späteren Phasen, weil häufig noch kein konkretes Material vorliegt, die Zeit knapp ist und die Auswertung aufwändig erscheint. Dies sollte Sie aber nicht davon abhalten, Ihr Augenmerk bei der Optimierung schon auf diese Phase zu richten. Denn in der frühen Phase werden alle entscheidenden Weichen für Ihren späteren Erfolg gelegt – Verbesserungen, die Sie hier vornehmen, wirken sich zeitsparend und kostengünstig auf die weiteren Entwicklungsschritte aus und sollten daher nicht vernachlässigt werden.

4. Literaturverzeichnis

Blank, R.: Gruppendiskussionsverfahren. In: Naderer, G./Balzer, E. (Hrsg.): Qualitative Marktforschung in Theorie und Praxis, Gabler, Wiesbaden 2007. S. 279-299.

Emmrich, Sebastian/Szwillus, Gerd: Auswertung von Card-Sorting-Experimenten mit CASOLYSIS. In: Brau, Henning et al. (Hrsg.): Usability Professionals 2008, S. 149-152.

Glaser, Barney G./Strauss, Anselm (2008): Grounded Theory. Strategien qualitativer Forschung. (Übersetzung der Originalausgabe von 1967 unter dem Titel The Discovery of Grounded Theory: Strategies for Qualitative Research). Bern.

Königstorfer, Jörg/Gröppel-Klein, Andrea: Projektive Verfahren zur Ermittlung der Akzeptanz technologischer Innovationen. Eine empirische Studie zu Internetanwendungen auf mobilen Endgeräten. In: Buber, Renate/Holzmüller, Hartmut H. (Hrsg.): Qualitative Marktforschung. Konzepte – Methoden – Analysen, Wiesbaden 2007. S. 847-862.

Lorenzen-Schmidt, Olde/Wörmann, Michael/Oettinger, Bernd: Nutzerzentrierte Strukturierung des Produktbereichs beim Webauftritt der comdirect bank. In: Zeitschrift für interaktive und kooperative Medien (i-com), Heft 2/2008, S. 52-54.

Maas, Jörg/Walter, Melanie: Anders auf die Daten schauen. Qualitative Typenbildung am Beispiel einer Reisestudie. In: planung & analyse, Heft 6/2007, S. 69-73.

Schmeißer, Daniel et al.: Effektive Kundenansprache in komplexen Märkten. Ergebnisse einer Segmentierung im Touristikmarkt. In: planung & analyse, Heft 6/2008, S. 31-37.

Schmeißer, Daniel/Joisten, Martina: Mehr als das Gewohnte: Innovative Produktentwicklung durch ethnografische Marktforschung. In: planung & analyse, Heft 3/2008, S. 30-35.

Strauss, Anselm/Corbin, Juliet: Grounded Theory: Grundlagen Qualitativer Sozialforschung. Weinheim 1996.

Rengelshausen, Oliver/ Schmeißer, Daniel: Touristik 2.0 – Chancen und Risiken von User Generated Content für den Online-Reisevertrieb. In: planung & analyse, Heft 4/2007, S. 58-61.

Röse, Kerstin: Task-Analyse. In: Heinsen, Sven/Vogt, Petra (Hrsg.): Usability praktisch umsetzen. Handbuch für Software, Web, Mobile Devices und andere interaktive Produkte. München Wien 2003. S. 98-114.

Tullis, Tim/Albert, Bill: Measuring The User Experience. Collecting, Analyzing, and Presenting Usability Metrics. Elsevier Inc., London 2008

VI. Rapid Prototyping: Optimierung durch iteratives Usability-Testing

von Daniel R. Schmeißer

1. Paper Prototyping: Entwickeln von Interaktionsabläufen 113
2. Usability-Tests mit Prototypen: Optimierung von Interaktionsabläufen 117
3. Besonderheiten bei der Durchführung von Rapid Prototypings 120
4. Anforderungen an die Projektbeteiligten 124
5. Fazit: Iterativ mit Nutzerfeedback optimieren 126
6. Literaturverzeichnis 126

Prototypen sind das klassische Feld für Usability-Testing. Dabei gilt die Devise, je früher desto besser. Es ist also nicht immer erforderlich, mit möglichst weit entwickelten Prototypen zu testen. Auch in frühen Phasen können bereits wichtige Abläufe abgebildet und genutzt werden. Voraussetzung ist das Vorliegen erster Prototypen, anhand derer die Interaktion mit einer Website untersucht werden kann, ohne das die hierfür erforderlichen Funktionalitäten bereits alle implementiert sind. Prototypen stellen Funktionalitäten, Interfaces und Bedienungskonzepte mehr oder weniger detailliert dar, so dass ausgewählte Aspekte der Website vorab ausprobiert, überprüft und optimiert werden können. Geht es um die Optimierung der Usability sollte also nicht erst bei der fertigen Website angesetzt werden, sondern bereits in frühen Entwicklungsstufen, denn in späteren Phasen steigt der Optimierungsaufwand exponentiell an. Je früher Sie testen, desto flexibler sind Sie in der Umsetzung von Änderungen und der Entwicklung neuer Ideen.

Allgemein wird zwischen horizontalen und vertikalen Prototypen unterschieden. Mit vertikalen Prototypen werden ausgewählte Funktionalitäten umgesetzt, die bereits realistisch und in vollem Umfang genutzt werden können. Horizontale Prototypen bilden die komplette Breite einer Benutzeroberfläche ab, ohne dass diese mit Funktionalitäten ausgestattet ist. In der Praxis hat man es gewöhnlich mit Mischformen zu tun, das heißt einzelne Teile sind bereits mehr oder weniger funktionsfähig, andere sind wiederum nur in der Oberfläche dargestellt. Für das Testen mit Prototypen von Websites ist daher die Unterscheidung zwischen dem Grad der Funktionalität zielführender. Hier lassen sich Prototypen mit Low-, Medium- und High-Fidelity unterscheiden.[32]

Arten von Prototypen

Scribbles: handgezeichnete Skizzen, die erste Interaktionsabläufe verdeutlichen

Storyboards: Zeichnungen oder Screenshots, die Interfaces in einer Handlungskette zeigen

Paper Mock-Up (Papier-Prototyp): einfache Handskizzen, komplette Papiermodelle oder Abfolgen von Powerpoint-Folien der Websites

Wireframes: schematische Darstellungen einer Website, mit denen in einem Ablaufmodell die Interaktionsabläufe festgehalten und auf Funktionalität geprüft werden können

Interaktiver Prototyp: Flash- oder Html-basierte Prototypen, in denen ausgewählte Funktionalitäten bereits vollständig nutzbar sind

[32] Siehe Nielsen, Jakob (1993): Usability Engineering. Academic Press, San Diego. Und: Sarodnick, Florian/Brau, Henning (2006): Methoden der Usability-Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung. Bern.

Es gibt einige generelle Aspekte, die Sie beim Testen mit Prototypen bzw. bei der Wahl des geeigneten Funktionsumfangs bei Prototypen beachten sollten: Bei einfachen Prototypen fällt es den Probanden in der Regel leichter, auch grundsätzliche Kritik an den Abläufen und Funktionen anzubringen, da der Blick nicht auf Design- und Layout-Elemente gerichtet ist. Umgekehrt verhindert ein unfertiger und skizzenhafter Prototyp häufig auch die realistische Einschätzung der späteren Anwendung. Insbesondere bei Websites spielt auch der visuelle Eindruck eine entscheidende Rolle bei der User Experience; grafische Elemente können die Nutzerführung mehr oder weniger intuitiv unterstützen oder erschweren.

Bei besonders ausgearbeiteten Prototypen besteht hingegen die Gefahr, dass die Prozesse und zugrundeliegenden Funktionalitäten nicht mehr grundlegend in Frage gestellt werden, das heißt je fertiger die Website auf die Befragten wirkt, desto weniger fühlen sich die Nutzer animiert, über die grundsätzlichen Eigenschaften des Systems zu diskutieren. Die Wahl eines geeigneten Prototypen sollte daher nicht nur in Abhängigkeit vom Entwicklungsstadium getroffen werden, sondern auch die Zielrichtung der Fragestellungen berücksichtigen. Hier lassen sich grob zwei grundlegende Untersuchungsziele unterscheiden: die Entwicklung und die Optimierung von Interaktionsabläufen.

1. Paper Prototyping: Entwickeln von Interaktionsabläufen

Liegen noch keine Konzepte oder Ablaufskizzen für eine Anwendung vor, können bereits mit einfachen horizontalen Prototypen (Scribbles, Storyboards und Paper Mock-Ups) erste Ideen zu den Interaktionsabläufen entwickelt werden. Mit dem Einsatz von Kärtchen, Papiervorlagen und Post-its lassen sich im Rahmen von Workshops gemeinsam mit den Projektbeteiligten und späteren Nutzern auf einfachstem Niveau erste Interaktionsabläufe simulieren. Aufgrund des frühen Projektstatus empfiehlt es sich, die Anwendung zunächst als Papier-Prototyp zu konzipieren, mit einem kleinen Nutzerkreis zu testen und anschließend zu optimieren. Durch dieses Vorgehen ist gewährleistet, dass schon früh im Entwicklungsprozess Nutzerfeedback eingeholt werden kann, ohne dass bereits aufwändige Programmierarbeiten durchgeführt werden müssen. Außerdem lassen sich zu diesem Zeitpunkt flexibel und mit geringem Aufwand Änderungen einbringen, zum Beispiel auch dadurch, dass der Nutzer selbst einfach seine Vorschläge „zu Papier bringen“ und kreativ werden kann.

Abbildung 24: Erstellung eines Papier-Prototypen mit Hilfe von Schere, Kleber und Print-its (Quelle: phaydon).

Ein Papier-Prototyp ist eine konzeptionelle Darstellung einer Anwendung, der das Verhalten der fertigen Anwendung simuliert. Auf Basis von Papier-Skizzen werden typische Prozesssequenzen in der Anwendung Schritt für Schritt dargestellt und damit das Verhalten der Anwendung anschaulich visualisiert und beschrieben. Diese Prozessschritte werden dann in einem Nutzertest anhand eines szenariobasierten Ansatzes überprüft. Der Nutzer bekommt eine bestimmte Aufgabe, die er mit der Anwendung lösen kann. Dazu interagiert er mit dem Papier-Prototypen und simuliert die Nutzung der echten Anwendung. Ein Mausklick kann hier beispielsweise durch einen Druck mit dem Finger auf eine bestimmte Stelle simuliert werden. Die Reaktionen des Systems werden also von einem Menschen simuliert (in der Regel durch den Versuchsleiter oder ein Mitglied des Projektteams). Da es sich im Rahmen von Workshops um eine offene Testsituation handelt, können verschiedene Reaktionen des Systems durchgespielt und gleichzeitig die Kommentare und Reaktionen der Nutzer festgehalten werden.[33]

Am Ende eines solchen Paper Prototyping Tests wird der Papier-Prototyp basierend auf dem Feedback der Nutzer optimiert, wobei die wesentlichen Funktionen und Strukturen anschaulich schematisch dargestellt und erneut getestet werden können (zum Beispiel in einem „Wireframe"). In der Regel genügt hier eine geringe Fallzahl von Testteilnehmern (acht bis zwölf Teilnehmer), da es ausschließlich um die Optimierung der Interaktionsabläufe, nicht jedoch um die Erhebung subjektiver Design- oder Akzeptanzurteile geht.

[33] Man spricht auch von Wizard-of-Oz-Verfahren, wenn ein versteckter Operator – der Wizard – die Reaktionen des Systems erzeugt, die durch eine Aktion des Nutzers hervorgerufen wird.

Fallbeispiel Bayer CropScience – Entwicklung eines Registrierungstools

Die Bayer CropScience AG zählt mit einem Jahresumsatz von ca. 5,8 Milliarden Euro zu den weltweit führenden, innovativen Crop-Science-Unternehmen mit den Tätigkeitsbereichen Pflanzenschutz (Crop Protection), Schädlingsbekämpfungsmittel für den nicht-landwirtschaftlichen Gebrauch (Environmental Science) sowie Saatgut- und Pflanzenbiotechnologie (BioScience). In einem Unternehmen, das weltweit über 17.800 Mitarbeiter beschäftigt, kommt der Informationsvermittlung und der Kommunikation von Strategien und Entscheidungen an die Mitarbeiter eine hohe Bedeutung zu. Bayer CropScience führt zu diesem Zweck im Headquarter regelmäßig sogenannte Business Briefings zu verschiedenen unternehmensrelevanten Themen durch. Die Mitarbeiter sollen so ein tieferes Verständnis für die Unternehmensorganisation und ihre Ziele und Märkte erlangen. Der Anmeldeprozess zu solchen Business Briefings stellt eine organisatorische Herausforderung dar.

Um den Anmeldeprozess für diese Informationsveranstaltungen möglichst effizient zu gestalten, sollte im Rahmen des „Registrierungstool"-Projekts eine Intranet-Anwendung entwickelt werden, die einen reibungslosen Anmeldeprozess für die Business Briefings mit minimalem Aufwand für Teilnehmer und Organisatoren ermöglicht. Diese Anwendung sollte zunächst nur für die Business Briefings entwickelt und bei erfolgreicher Etablierung auch für weitere Unternehmensveranstaltungen angewendet werden.

Bei der Erstellung der Anwendung sollte insbesondere auf Benutzerfreundlichkeit (Usability) Wert gelegt werden, um eine hohe Nutzerakzeptanz zu erreichen. In diesem frühen Planungsstadium bot sich der Einsatz eines Papier-Prototypen an, der zeit- und kostengünstig erstellt werden konnte und trotzdem die wichtigsten Funktionen des Tools simulierte. In der Interaktion von Nutzern mit dem Papier-Prototypen ließen sich somit zahlreiche Usability-Probleme aufdecken, ehe auch nur eine einzige Zeile der Intranet-Anwendung programmiert werden musste. Basierend auf dem Feedback der Nutzer wurde der Papier-Prototyp anschließend optimiert und konkrete Design-Vorschläge („Wireframes") erarbeitet, die die wesentlichen Funktionen und Strukturen anschaulich schematisch darstellen.

Es wurden zehn Interviews im Headquarters durchgeführt. Die Dauer der Interviews betrug 30 Minuten, in denen zunächst die generellen Bedürfnisse und Wünsche der Nutzer bezüglich einer Onlineregistrierung für Business Briefings abgefragt wurden. Im zweiten Teil des Interviews wurden dann anhand eines Papier-Prototypen die Abläufe mittels Kärtchen, Scribbels und Klebezetteln Schritt für Schritt simuliert. Dabei ging es um den gesamten Registrierungsprozess von der Begrüßung, über die Veranstaltungsauswahl und den persönlichen Daten bis zur Bestätigung der Anmeldung. Am Anfang standen dabei vor allem der Ablauf

selbst sowie das Wording der einzelnen Prozessschritte im Mittelpunkt. Abbildung 25 zeigt den Papier-Prototypen für den ersten von drei Schritten zur erfolgreichen Anmeldung: die Terminauswahl.

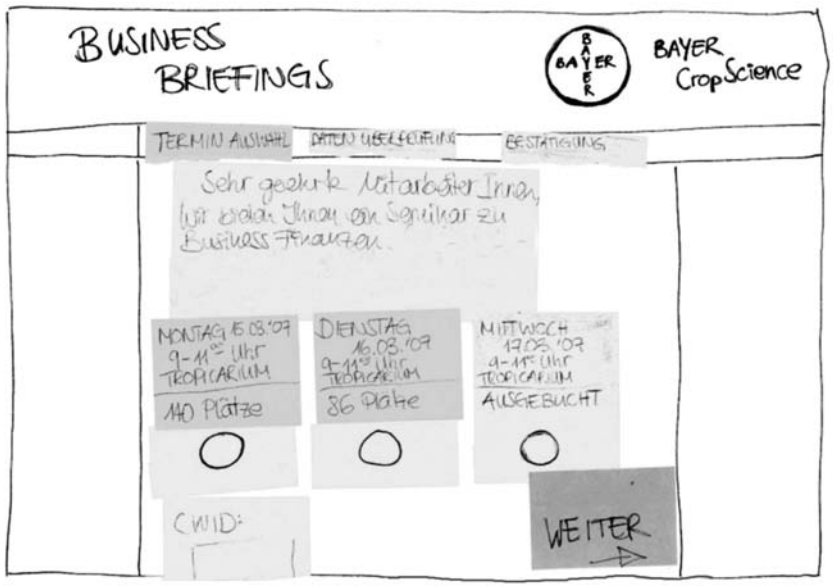


Abbildung 25: Papier-Prototyp Business Briefings: Schritt 1 „Terminauswahl" (Quelle: Bayer CropScience AG).

Die Detaillierung der Papier-Prototypen berücksichtigte dabei weder das künftige Design, noch die technische Umsetzbarkeit der Prozessabfolge, sondern hielt lediglich die grundsätzlichen Anforderungen an den Prozess fest. Die Papier-Prototypen wurden dabei systematisch exploriert, wobei auch Schwachstellen und Usability-Probleme deutlich wurden, zum Beispiel bei der Benennung einzelner Ablaufschritte oder bei der Logik der Prozessschritte.

Bereits nach wenigen Interviews kristallisierten sich die Stärken und Schwächen dieser ersten Version heraus: Die klare Ausweisung der notwendigen Schritte in Form eines Fortschrittsbalkens machten den Nutzern den Prozessablauf transparent. Ebenso überzeugten die großzügig gestalteten Navigationselemente (Radio-Buttons und „Weiter"-Button). Zu den Schwachpunkten zählte zum Beispiel die Angabe von Plätzen, weil für die Nutzer nicht klar war, ob es sich bei um bereits belegte oder freie Plätze handelt. Im optimierten Wireframe wurde dieses Problem durch die genauere Bezeichnung „freie Plätze" behoben.

Die so entwickelten Prozessabläufe für die Online-Anmeldung zu den Business Briefings wurden im nächsten Schritt optimiert und in Wireframes (siehe Abbildung 28) übersetzt. Hierbei wurden die von den Nutzern genannten Mängel berücksichtigt und auf Basis eines detaillierten Konzepts ausgearbeitet.

Der Prozessschritt „Bestätigung" wurde aufgrund des Nutzerfeedbacks hinsichtlich einer deutlicheren Erfolgsmeldung optimiert. Neben der Registrierungsbestätigung und dem Hinweis auf die Bestätigungsmail wurde auch ein prominent gestalteter „Fenster schließen"-Button eingeführt, der den Prozess für die Nutzer psychologisch endgültig beendet.

Mit Hilfe des Paper Prototypings konnte so ein erster Entwurf der Anwendung entwickelt werden, der die Bedürfnisse der Anwender ernst nimmt und bei der Umsetzung der Interaktionsabläufe systematisch die Nutzersicht berücksichtigt. Der frühe Einsatz in der Entwicklungsphase vermied eine kostenaufwändige Programmierung, die im Live-Betrieb unter Umständen zu Problemen geführt hätte. Das Registrierungstool ist seit seiner Entwicklung erfolgreich im Einsatz und genießt eine hohe Akzeptanz bei den Anwendern.

2. Usability-Tests mit Prototypen: Optimierung von Interaktionsabläufen

Je weiter der Entwicklungsprozess einer Website voranschreitet, desto klarer werden Nutzungskonzept und mögliche Nutzungsszenarien. Der Prototyp selbst wird dabei zunehmend interaktiver und mit Funktionalitäten ausgestattet. Die Entwicklung interaktiver Prototypen ist dabei Voraussetzung für die Durchführung von Usability-Tests, bei denen die Nutzer möglichst frei durchspielen sollen, wie sie unter normalen Bedingungen mit der Website oder Anwendung interagieren würden. Der Prototyp muss daher zumindest in ausgewählten Bereichen alle Funktionalitäten umfassen, damit ausgewählte Interaktionsabläufe abgebildet werden können. Prototypen in Usability-Tests liegen daher als Wireframes oder als in Flash- oder Html programmierte Klickstrecken bzw. in einer voll funktionsfähigen Testumgebung vor. Diese können nun im Rahmen von Nutzungsaufgaben detailliert im Hinblick auf Usability-Schwachstellen untersucht werden.

Der Usability-Test mit Prototypen zählt dabei zu den formativen Evaluationsverfahren (siehe detailliert zu Usability-Kriterien Kapitel X). „Formativ" ist das Usability-Testing deshalb, weil (im Gegensatz zur summativen Evaluation, bei der nachträglich Websites überprüft werden) die Weiterentwicklung und Optimierung des Prototypen unter aktiver

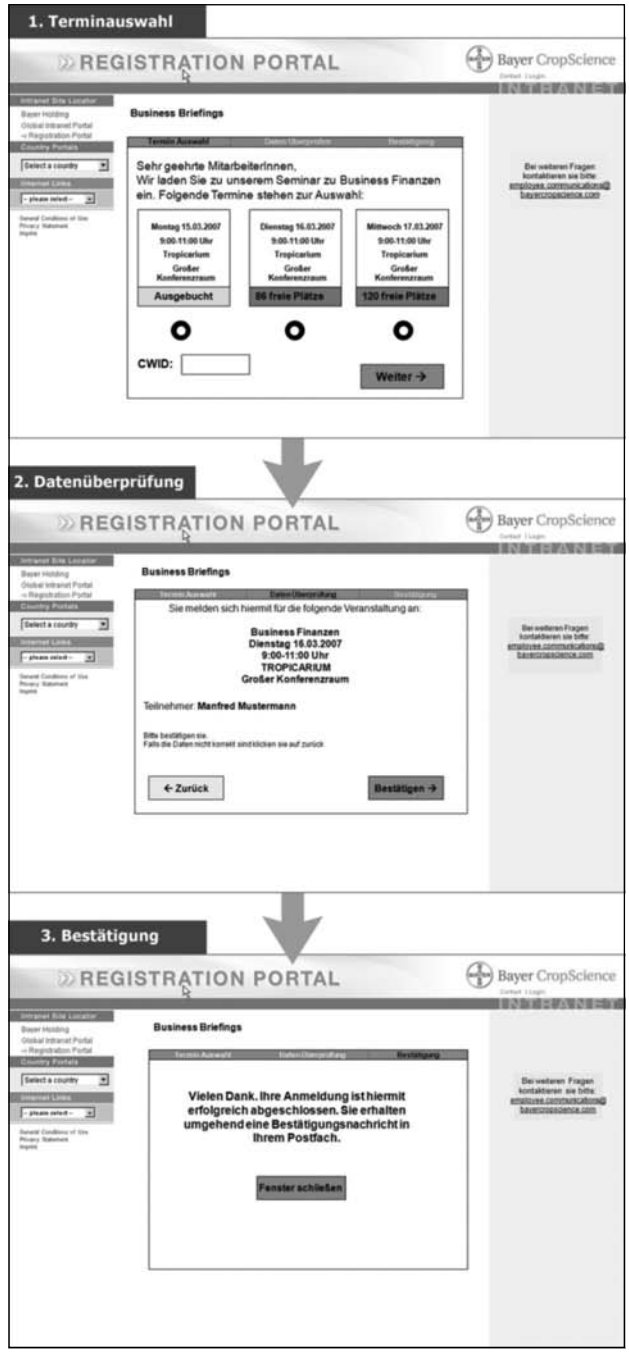


Abbildung 26: Umsetzung der Papierprototypen in Wireframes: Registrierungsprozess in drei Schritten – Terminauswahl, Datenüberprüfung und Bestätigung (Quelle: Bayer CropScience AG).

Einbeziehung des Nutzerfeedbacks erfolgt. Beim Usability-Testing mit Prototypen steht also weniger das Prüfen, sondern vielmehr das Auffinden von Schwachstellen und Optimierungspotenzialen im Vordergrund, wobei der Fokus ganz klar auf der Usability bzw. im Auffinden von Usability-Problemen liegt. Hierfür bearbeiten die Nutzer möglichst typische Aufgaben, die das System vorsieht; während sie interagieren, verbalisieren sie laut ihre Gedanken und Gefühle (siehe zu Stichprobengröße, Interviewdurchführung und Verhaltensbeobachtung Kapitel X).

Während der Nutzung oder im Anschluss daran fragt der Interviewer gezielt nach und exploriert, welche Erwartungen und Eindrücke der Nutzer hat. Die gesamte Interaktion und das Interview werden vollständig aufgezeichnet und können nachträglich in der Auswertung herangezogen werden. Datengestützte Explorationen oder Blickaufzeichnung (Eyetracking) empfehlen sich in den frühen Phasen des Usability-Testings mit Prototypen allerdings nicht, da der Interviewer hin und wieder in den Prozess eingreifen muss, um die richtigen Handlungsabläufe sicherzustellen bzw. um Fehler, die aufgrund der unvollständigen Programmierung entstehen können, zu vermeiden. Zum Anderen geht es nicht um die Beurteilung von Design- und Layout-Aspekten, die erst in späteren Phasen an Relevanz gewinnen und zum Beispiel im Rahmen eines User Experience-Tests überprüft werden können (vgl. Kapitel X).

Welche Nutzungsaufgaben in den Test einbezogen werden sollten, hängt von der Ausarbeitung des Prototypen sowie vom geplanten Interaktionsumfang ab. Bei der Auswahl der Nutzungsaufgaben anhand von Prototypen hat sich die Wahl möglichst realistischer und kurzer Use Cases bewährt, die von den Nutzern soweit wie möglich ohne fremde Hilfestellung bearbeitet werden können. Sollen zum Beispiel Interaktionselemente und -abläufe in einem Anmeldeprozess getestet werden, empfiehlt sich das einfache Durchspielen der Anmeldung Schritt für Schritt, Screen für Screen. Liegen die Prototypen in unfertiger Form vor (das heißt, es können nur die für den Use-Case relevanten Bedienelemente genutzt werden), muss der Interviewer den Nutzer entsprechend anleiten und immer wieder auf den „richtigen“ Weg bringen.

Das beim Test von Prototypen angewendete Verfahren entspricht dem sogenannten „Walkthrough“ – also dem schrittweisen Durchgehen einer Aufgabe. Zentral bei dem Vorgehen ist, dass der Interviewer mit dem Nutzer die Aufgaben (in der Regel zwei bis sechs Use Cases) Schritt für Schritt durchspielt und dabei Fragen nach dem Umgang und dem Erleben während der Interaktion stellt. Zur Aufdeckung von Usability-Problemen wird danach gefragt, ob der Nutzer weiß, was zu tun ist, um ein Ziel zu erreichen, ob er geeignete

Aktionen erkennen kann und wie die Reaktionen des Systems beurteilt werden, bzw. ob der Nutzer erkennt, dass er sein Ziel erreicht hat.

Der Interviewer hält dabei anwendungsbegleitend fest, an welchen Stellen der Nutzer Schwierigkeiten hatte, sein Ziel zu erreichen, bzw. an welchen Stellen Fehler aufgetreten sind, die einer Zielerreichung im Wege standen. Hier (und auch nachträglich in der Analyse der Interaktionsprotokolle und Aufzeichnungen) werden also die „critical incidents" festgehalten, nach Nutzergruppen und Aussagetypen gruppiert, gewichtet und anschließend mit den Projektbeteiligten im Hinblick auf konkrete Möglichkeiten der Optimierung diskutiert. Da es sich um echte Interaktionen mit der künftigen Website handelt, ist es wichtig, dass im Test auch die tatsächlichen bzw. künftigen Nutzer der Website einbezogen werden. Diese müssen in der Zielgruppenbestimmung definiert und in einem Stichprobenplan festgelegt werden (siehe Kapitel X).

3. Besonderheiten bei der Durchführung von Rapid Prototypings

Das Rapid Prototyping ist eine besondere Form des Usability-Testings, bei dem die Vorzüge von Prototypen-Tests mit dem iterativen und partizipativen Ansatz des User Centered Design verknüpft werden. Grundsätzlich gelten daher alle eingangs genannten Verfahrensweisen auch für das Rapid Prototyping. Im Unterschied zum Paper Prototyping kommen allerdings interaktive Prototypen oder Wireframes zum Einsatz. Der wesentliche Unterschied zu den klassischen Prototyping Tests besteht darin, dass sich beim Rapid Prototyping Phasen der Evaluation und der Optimierung abwechseln. Durch die iterative Durchführung von Test- und Optimierungsphasen ermöglicht das Rapid Prototyping die effiziente Optimierung eines Prototyps innerhalb eines kurzen Zeitraums von sechs bis maximal zwölf Tagen. Das Verfahren eignet sich daher besonders entwicklungsbegleitend zur Konzeption und Programmierung einzelner Website-Bereiche oder isolierter Anwendungen (zum Beispiel Buchungs- und Anmeldeprozesse).

Folgende Einsatzbereiche und Prototypen haben sich in der Praxis bewährt

- Test von Navigationskonzepten anhand rudimentärer Klickstrecken, die lediglich das Navigationsprinzip veranschaulichen
- Test von Anmelde- und Bestellprozessen, bei dem der komplette Prozess anhand von Wireframes durchgespielt werden kann
- Test von interaktiven und visuell gestalteten Prototypen, die ein Gefühl für die spätere Nutzung der Website vermitteln, aber auf bestimmte Bereiche beschränkt bleiben (zum Beispiel Homepage oder ausgewählte Produktseiten)
- Test von Landing Pages, um den optimalen Workflow vom Werbemittel über die Landing Page bis in den Shop abzubilden

Erfahrungsgemäß genügt eine Anzahl von drei Explorationsphasen, um zu einem tragfähigen und im Detail optimierten Prototyp zu gelangen. Nur bei sehr komplexen Anwendungen sind gegebenenfalls mehrere Explorationsphasen erforderlich. In der Praxis hat sich herausgestellt, dass in der Regel drei Testrunden im Abstand von jeweils zwei bis drei Tagen mit jeweils darauf folgenden Optimierungsphasen ausreichen.

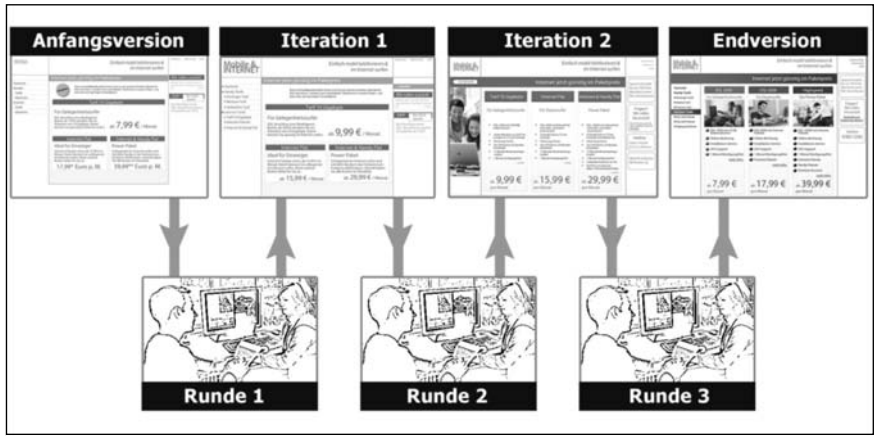


Abbildung 27: Vereinfachter, typischer Ablauf eines Rapid Prototypings mit drei aufeinander folgenden Interviewrunden (Quelle: phaydon).

Die Explorationsphasen werden in Form von Einzelinterviews mit einer Länge von 45 bis 120 Minuten durchgeführt. Bei der Stichprobengröße hat sich eine Anzahl von mindestens vier, alternativ sechs Personen pro Testrunde als ausreichend erwiesen. Bei drei Explorationsrunden ergibt sich somit eine Stichprobengröße von insgesamt 12 bis 18 Personen. Je nach Fragestellung können anstelle von Einzelinterviews auch Workshops bzw. sogenannte „Mini-Groups“ mit vier bis sechs Teilnehmern zum Einsatz kommen. Diese bieten sich an,

wenn es bei der Nutzung der Prototypen auch um den interaktiven Austausch der Teilnehmer untereinander geht.

Im Rahmen der Explorationsphasen werden zunächst die allgemeinen Wünsche und Erwartungen der User an die Interaktion mit der Website herausgefunden. Anschließend werden die Screens sukzessive durchgespielt und anhand von Use Cases mit Blick auf Verständlichkeit, Wording und Handhabbarkeit exploriert. Da der Ablauf der Interviews hohe Anforderungen an die Testsituation stellt, sollten die Interviews ausschließlich von psychologisch geschulten Interviewern durchgeführt werden, die über entsprechende Usability-Expertise verfügen. Dabei ist es wichtig, dass der Interviewer bzw. Usability-Experte ein besonderes Gespür dafür entwickelt, wie der einzelne Nutzer mit dem Prototyp umgeht.

Gerade aufgrund des wechselnden Detaillierungsgrades reagieren die Nutzer sehr unterschiedlich: Einige fühlen sich durch ihre frühe Einbeziehung „privilegiert" und dazu animiert, die vermeintliche „Expertenbrille" aufzusetzen, anderen fehlt es dagegen an Vorstellungskraft und Abstraktionsfähigkeit, wieder andere verlieren sich unnötig in Details. Der Interviewer sollte daher immer wieder auf die relevanten Use Cases und die interaktionsbegleitenden Gedanken der Nutzer fokussieren. Durch die Vermittlung des Entwicklungsstandes und Zweck der Untersuchung kann bei den Teilnehmern ein besseres Verständnis der Fragestellungen erzielt werden.

Beim Rapid Prototyping haben sich folgende Fragetypen bewährt, um die Interaktion mit dem Prototyp vertiefend zu erfassen und mögliche Usability-Probleme aufzudecken

- „Was sehen Sie hier?" (Zustandsbeschreibung)
- „Was können Sie auf dieser Seite alles machen?" (Erlebter Handlungsspielraum)
- „Wie würden Sie als nächstes vorgehen, um Ziel x oder y zu erreichen?" (Kognitionen zur Zielerreichung)
- „Was erwarten Sie, was passiert, wenn sie auf diesen Button klicken?" (Handlungsvorbereitende Kognitionen)
- „Haben Sie sich das, was passiert ist, auch so vorgestellt oder inwiefern weicht es ab?" (Erwartungsabgleich)

Nach jeder Explorationsphase werden die Ergebnisse mit allen Projektbeteiligten diskutiert und Lösungsvorschläge für die Überarbeitung bzw. Optimierung der Screens im Rahmen einer moderierten Debriefing-Sitzung entwickelt. Hierbei empfiehlt es sich, dass die Projektbeteiligten bzw. Produktmanager der Fachabteilung an den Interviews beobachtend

teilnehmen, um im Anschluss mit den Usability-Experten und Interviewern die Lösungsschritte bei der Optimierung der Prototypen zu planen. Eine besondere Herausforderung besteht darin, die unterschiedlichen Perspektiven auf das Produkt (Nutzersicht, Expertensicht, Unternehmenssicht) so miteinander zu integrieren, dass daraus eine optimale Anwendung resultiert. Das ist nicht immer einfach und stellt hohe Anforderungen an die Kommunikationsfähigkeit und die konstruktive Zusammenarbeit der Projektbeteiligten (siehe weiter unten im Detail). Gelingt eine differenzierte Auseinandersetzung mit den unterschiedlichen Perspektiven und Interessenschwerpunkten, ist das Rapid Prototyping ein sehr effizientes Verfahren, wenn Sie die Zielgerichtetheit Ihrer Entwicklung und Optimierung sicherstellen wollen.

Im Anschluss an das jeweilige Debriefing werden die Prototypen und Workflows im Rahmen der „Optimierungsphase" entsprechend angepasst. Die Anpassung erfolgt in der Regel innerhalb des vorgegebenen Layouts, das heißt das Design wird nicht konzeptuell verändert, da es nicht von den eigentlichen Interaktionsabläufen ablenken soll. Die Optimierung erfolgt auf Basis der identifizierten Usability-Schwachstellen und berücksichtigt vor allem den Seitenaufbau, einzelne Funktionen und das Wording. Screens und Klickdummy werden gemäß den Vorgaben entweder vom durchführenden Institut oder den beteiligten Agenturen umgesetzt.

Ein Rapid Prototyping verläuft typischerweise wie folgt:

1. Festlegen der zu berücksichtigenden Use Cases und Lieferung des Prototypen durch den Auftraggeber
2. Programmierung bzw. Zusammenstellen der Klickstrecken, damit die relevanten Interaktionsabläufe durchlaufen werden können
3. Erstellung und Abstimmung eines detaillierten Interviewleitfadens, der die Use Cases, Instruktionen und Fragen an die Benutzer in strukturierter Form festhält
4. Durchführung der ersten Interviewrunde, in der Regel mit jeweils zwei parallel ablaufenden Interviews, die von den Projektbeteiligten durch eine Einwegscheibe im Use-Lab beobachtet werden können
5. Anschließendes Debriefing mit allen Projektbeteiligten: Analyse der critical incidents und Ableitung von Handlungsempfehlungen zur Optimierung des Workflows
6. Umsetzung der erarbeiteten Empfehlungen durch einen Interaktionsdesigner
7. Anpassen des Interviewleitfadens und der Klickstrecke gemäß den vorgenommenen Änderungen
8. Wiederholtes Durchlaufen von Interview- und Optimierungsrunden (je nach Bedarf und Optimierungspotenzial)

9. Abschließendes Debriefing: Festlegen der noch offenen To-dos inklusive Pflichtenheft und Anforderungsanalyse für die nächsten Schritte
10. Erstellung des finalen Prototypen auf Basis der letzten Ergebnisrunde
11. Erstellung eines Abschlussberichts, in dem die Prozessschritte und Veränderungen dokumentiert sind

Es ist unschwer erkennbar, dass ein Rapid Prototyping erhebliche Anforderungen an alle Projektbeteiligten hinsichtlich Planung, Durchführung und Umsetzung der Optimierung stellt. Ein herkömmlicher Usability-Test, in dem einmalig Nutzer befragt werden und der Auftraggeber zwei Wochen später den Abschlussbericht erhält, ist ungleich weniger zeitintensiv für die Beteiligten. Es passiert daher immer wieder, dass die Durchführung eines Rapid Prototypings unternehmensintern zunächst auf Widerstand stößt, da primär der kurzfristig hohe Aufwand, aber nicht der langfristige Nutzen gesehen wird. Die Investition in ein Rapid Prototyping lohnt sich aber in jedem Fall und verfügt über ein hohes Return on Invest, weil eine bereits fertige Anwendung später nur sehr viel aufwändiger und kostenintensiver zu optimieren ist und gegebenenfalls bis zu diesem Zeitpunkt bereits Kunden durch eine mangelhafte Nutzerführung vergrault wurden. Ein sauber durchgeführtes Rapid Prototyping ermöglicht im Idealfall eine nutzerzentrierte und trotzdem pragmatische Entwicklung einer Website, deren Interaktionsabläufe später problemlos handhabbar sind. Für den reibungslosen Ablauf eines Rapid Prototypings sind aber eine Reihe von Anforderungen an die Projektbeteiligten zu beachten, die nicht zuletzt über seinen Erfolg entscheiden.

4. Anforderungen an die Projektbeteiligten

Neben dem bereits erwähnten Vorbehalt bezüglich des Aufwands für Rapid Prototypings betreffen die Anforderungen an die Projektbeteiligten vor allem zwei Aspekte – zum Einen die Erstellung des Prototyps im Vorfeld der Durchführung, zum Anderen die Beteiligung am Optimierungsprozess selbst.

Erstellung der Prototypen

Neben der termingerechten Lieferung der Ausgangsprototypen, die für die Durchführung eines Rapid Prototypings entscheidend ist, ist es vor allem der Prototyp selbst, der zu Problemen führen kann. Zum Einen besteht die Gefahr, dass durch einen ungeeigneten Detaillierungsgrad der Fokus der Testteilnehmer von der eigentlichen Fragestellung abgelenkt wird. Dies ist zum Beispiel der Fall, wenn ein Prototyp bereits durch die Kreativagentur visuell sehr differenziert ausgestaltet wird und dadurch im Test viele Diskussionen über

die visuelle Gestaltung provoziert, obwohl das Untersuchungsziel in der Optimierung des Interaktionsdesigns liegt. Aber auch umgekehrt kann der Fall auftreten, dass der Prototyp zu wenig detailliert ist und die Diskussion über Interaktionsabläufe in der Folge zu abstrakt bleibt. In solchen Fällen ist schwierig festzustellen, wie der Workflow im Falle einer detaillierten visuellen Umsetzung ablaufen würde.

Bereits bei der Erstellung der Prototypen ist daher darauf zu achten, den Detaillierungsgrad in Abhängigkeit vom Untersuchungsziel festzulegen. Der Prototyp sollte der Fragestellung angemessen sein – dies setzt voraus, dass diese vor der Erstellung des Prototyps festgelegt und für alle Projektbeteiligten transparent ist. Ist die Fragestellung unklar, kann es passieren, dass der Prototyp unter Umständen mit zu großem Aufwand entwickelt wird (beispielsweise auch deshalb, weil sich die Entwickler oder die Agentur dazu verpflichtet fühlen, einen möglichst detaillierten Prototypen in das „Rennen" zu schicken, um „vorläufige" Ergebnisse zu vermeiden.) Aber gerade wegen des hohen Detaillierungsgrades des Prototyps ist das Risiko hoch, dass die Kritik am Prototyp ähnlich detailliert und umfangreich ausfällt, was letztlich zu einer Verschwendung von Ressourcen führt.

Auch ein weiterer Aspekt scheint zentral, der beim Testen mit Prototypen häufig zu beobachten ist.[34] Sowohl für die Testteilnehmer als auch für die unternehmensinternen Projektbeteiligten kann es schwierig sein, den Stellenwert eines Prototyps richtig einzuschätzen, zum Beispiel wenn durch die visuelle Gestaltung der Eindruck entsteht, dass die Entwicklung schon sehr weit fortgeschritten ist. Dies kann in der Interviewsituation zu Komplikationen führen, wenn die Nutzer nicht mehr von auftretenden Problemen abstrahieren können und diese als schwerwiegender einschätzen, als sie aufgrund des frühen Entwicklungsstadiums tatsächlich sind. Aber auch bei Inhouse-Entwicklungen kann hierdurch der Druck auf die Projektverantwortlichen erhöht werden, da unter Umständen die Akzeptanz für Modifikationen und Verzögerungen der Entwicklung reduziert wird.

Beteiligung am Optimierungsprozess

Das direkte Aufeinanderfolgen von Evaluations- und Optimierungsphasen macht eine enge Zusammenarbeit zwischen den verschiedenen Projektbeteiligten erforderlich. Dies betrifft nicht nur die Vorbereitung und Festlegung des Prototyps im Vorfeld des Tests, sondern vor allem die testbegleitende Anwesenheit aller Projektbeteiligten während der Interviews und den Abschlussbesprechungen zur Festlegung der jeweils erforderlichen Optimierungsschritte. Alternativ zu den Debriefing-Sitzungen vor Ort eignen sich auch Video- oder

[34] Vgl. Weber, Markus/Immich, Thomas (2008): User Interface Prototypen in Usability Engineering und Software Entwicklung: Probleme und Synergien. In: Brau, Henning et al. (Hrsg.): Usability Professionals 2008, S. 40-44.

Telefonkonferenzen; diese können aber den direkten Austausch unter dem unmittelbaren Eindruck der live verfolgten Interviews mit „echten" Nutzern nicht ersetzen. Findet das Rapid Prototyping unter Beteiligung externer Dienstleister statt (Agenturen, Forschungsinstitute), ist eine klare Aufgabenteilung zur Optimierung der Kommunikationswege zwingend erforderlich. So erfolgt die Planung gemeinsam mit dem beauftragenden Unternehmen, die Durchführung der Interviews sowie die Auswertung und Berichtslegung erfolgen in der Regel durch das Forschungsinstitut. Viele UX-Institute verfügen inzwischen auch über Personal mit entsprechender Ausbildung in *User Centered Design* (UCD) bzw. UI-Design. Deshalb ist bei Beteiligung der ausführenden Agentur eine Aufgabenverteilung bezüglich der Umsetzung bzw. Optimierung im Interaktionsdesign und bei der Weiterentwicklung des Prototyps im Vorfeld notwendig. Unabhängig davon, ob die Agentur oder das Forschungsinstitut die Umsetzung durchführt, muss gewährleistet sein, dass diese auf Grundlage der Ergebnisse aus den einzelnen Optimierungsrunden erfolgt.

5. Fazit: Iterativ mit Nutzerfeedback optimieren

Prototyping Tests gehören zu den effizientesten Verfahren im Rahmen des User Centered Design Prozesses. Durch den flexiblen Einsatz unterschiedlich detaillierter und interaktiver Prototypen können sie während des gesamten Entwicklungsprozesses eingesetzt und zur Optimierung der Website genutzt werden – von einfachen Scribbles bis hin zu interaktiven Klickstrecken, die mit weitgehend vollständiger Funktionalität ausgestattet sind. Das Rapid Prototyping stellt unter diesen Verfahren das effektivste dar, da es nutzerzentriert und iterativ zur Optimierung des Prototypen in kürzester Zeit beiträgt. Es erfordert von allen Beteiligten ein hohes Maß an Professionalität, Usability-Expertise, Ausdauer und Leidenschaft; der Nutzen eines iterativen Optimierungsprozesses erscheint dafür aber evident. Richtig durchgeführt helfen Rapid Prototypings, Ihre Website zeit- und kosteneffizient bereits zu einem frühen Zeitpunkt nutzerorientiert umzusetzen und zu optimieren.

6. Literaturverzeichnis

Nielsen, Jakob: Usability Engineering. Academic Press, San Diego 1993.

Sarodnick, Florian/Brau, Henning: Methoden der Usability-Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung. Bern 2006.

Weber, Markus/Immich, Thomas (2008): User Interface Prototypen in Usability Engineering und Software Entwicklung: Probleme und Synergien. In: Brau, Henning et al. (Hrsg.): Usability Professionals 2008, S. 40-44.

VI.I Fallbeispiel: Rapid Prototyping für die Deutsche Post

von Daniel R. Schmeißer und Tobias Kauer

1. Testdesign: Iteratives Prototyping in zwei Wellen 128
2. Optimierungsresultate aus den beiden Testwellen 130
3. Fazit: Durch iteratives Prototyping zum Erfolg..... 135

Seit einigen Jahren bietet die Deutsche Post AG sowohl privaten wie auch geschäftlichen Nutzern mit dem Produkt STAMPIT die Möglichkeit, von zu Hause bzw. dem Arbeitsplatz aus über PC, Drucker und Internet Brief- oder Paketsendungen selbst zu frankieren. Dieses Angebot trägt dem Wunsch einer immer größer werdenden Zahl von Kunden Rechnung, die eigenen Postgeschäfte unabhängig von den Filial-Öffnungszeiten erledigen zu können. Regelmäßige Nutzerbefragungen zeigten, dass das Angebot auf eine hohe Wertschätzung und Weiterempfehlungsbereitschaft traf; allerdings wurden auch relativ schnell Optimierungspotenziale deutlich. Diese betrafen zum Einen die Funktionalität der IT-Lösung: Eine Basisversion von STAMPIT ist webbasiert, die vollen Funktionalitäten lassen sich aber nur nutzen, wenn man die entsprechende Frankier-Software auf dem eigenen PC installiert, was sich als generelle Einstiegsbarriere für die Nutzung erwies.

Daneben äußerten zahlreiche Nutzer den Wunsch nach zusätzlichen Anwendungen und Optionen, zum Beispiel hinsichtlich des Layouts der verwendeten Briefmarken. Dies alles führte letztendlich zu Überlegungen, das Alternativprodukt „Internetmarke" zu entwickeln, das sich durch folgende Kernmerkmale auszeichnen sollte:

- eine rein webbasierte Lösung, um die Einstiegshürden zur Nutzung so niedrig wie möglich zu halten;
- eine einfache, extrem nutzerfreundliche Funktionalität („mit 3 Klicks zur Briefmarke");
- die Möglichkeit, das Layout der Briefmarke weitestgehend individuell zu gestalten.

Um vor diesem Hintergrund die tatsächlichen Nutzerbedürfnisse optimal zu berücksichtigen, wurde bereits die Konzeptionsphase forscherisch begleitet.

1. Testdesign: Iteratives Prototyping in zwei Wellen

Die Methode des „Rapid Prototyping" erschien allen Beteiligten dabei optimal geeignet, die ersten Entwürfe auf Akzeptanz, Bedienungsfreundlichkeit und (generelle) Nutzungsbereitschaft in allen relevanten Zielgruppen zu überprüfen und die Vorschläge für die weitere Entwicklung zu optimieren. Durch die frühe Einbeziehung der Nutzer in der Entwicklungsphase sollte sichergestellt werden, dass die Anwendung von Beginn an nutzerzentriert entwickelt wird und keine unnötigen Entwicklungskosten durch eine fehlerhafte Umsetzung entstehen. Bereits im Frühjahr 2007 wurden daher die ersten Entwürfe in Form eines Klick-Dummy, der den Bestell- und Druckprozesses der „Internetmarke" prototypisch

abbildete, einem Test unterzogen. Zielsetzung des ersten (der insgesamt zwei) „Rapids" waren die

- Überprüfung des Klick-Dummys hinsichtlich seiner Usability
- Anpassung der Informationsarchitektur an die Bedürfnisse und Wünsche der Nutzer
- Erarbeitung einer optimierten Version für die Anwendung „Internetmarke" im Rahmen des Prototypings

Das Rapid bestand aus drei Explorationsphasen mit je vier Interviews pro Runde und jeweils darauf folgenden Optimierungsphasen, in denen die Screenentwürfe nach den Ergebnissen der Nutzerbefragung weiterentwickelt wurden. Die Änderungen des Screendesigns wurden durch phaydon interne Interaktionsdesigner umgesetzt, die Untersuchung fand in den phaydon Usability-Labs in Köln statt.

Im Rahmen der Einzelinterviews wurden die Schwachstellen der geplanten Anwendung identifiziert und an die Bedürfnisse der Nutzer angepasst. Hierzu wurden sowohl die Eignung der grafischen, textlichen und technischen Gestaltung als auch Inhalt, Aufbau, Navigation und Layout bei der Überprüfung berücksichtigt. Ferner wurden Informationsgehalt und Informationsaufarbeitung, Anmutung und Zufriedenheit der Nutzer exploriert.

Die Interviews hatten eine Länge von 60 Minuten und beinhalteten eine Warm-Up-Phase zur Erhebung der nutzerseitigen Erwartungen an eine Anwendung wie die „Internetmarke" und einen szenario-orientierten Hauptteil, in dem anhand kleinerer Use Cases der Prozess der Briefmarkenerstellung gemeinsam mit den Nutzern durchgespielt wurde. Neben der Identifikation von Schwachstellen und Optimierungshinweisen ging es auch um die Frage, wie das Produkt als solches von den Nutzern aufgenommen wird bzw. welche Wünsche bezüglich der persönlichen Briefmarkenerstellung geäußert werden.

Ergebnis des ersten Rapids waren optimierte Screenentwürfe, die im nächsten Schritt unternehmensintern sowohl konzeptuell als auch technisch und grafisch weiterentwickelt werden konnten. Der optimierte Prototyp wurde dann erneut im Sommer 2008 in einem weiteren Rapid Prototyping einer detaillierten Überprüfung unterzogen. Neben den ursprünglich angedachten Prozessschritten handelte es sich in der erweiterten Version um einen weiter ausgearbeiteten Klickdummy, in dem bereits kommunikative und grafische Elemente umgesetzt waren. Dieser wurde analog zum ersten Rapid in drei Explorationsrunden optimiert.

Befragt wurden in beiden Testwellen je zwölf Internetnutzer im Alter zwischen 20 und 59 Jahren, gleichmäßig verteilt auf zwei Altersgruppen (20 bis 39 Jahre; 40 bis 59 Jahre). Alle Testteilnehmer bestellten zumindest gelegentlich im Internet, hatten aber noch keine Briefsendung online frankiert. Die Stichprobe bestand je zur Hälfte aus Frauen und Männern sowie zu gleichen Teilen aus Privatkunden und gewerblichen Kunden. Darüber hinaus wurde die Häufigkeit der Internetnutzung quotiert (Light-, Medium- und Heavy-Nutzer).

Die Ergebnisse und das Vorgehen des Rapid Prototyping werden im Folgenden für beide Testwellen exemplarisch anhand einzelner Prozessschritte erläutert.

2. Optimierungsresultate aus den beiden Testwellen

Ausgewählte Ergebnisse der ersten Welle (2007)

Der Prototyp in der ersten Welle bestand aus vier Prozessschritten (siehe Abbildung 28): „Briefmarke auswählen", „Motiv auswählen", „Bezahlart auswählen" und „Briefmarke drucken". Obwohl nur ein sehr rudimentärer Prototyp als Testbasis zur Verfügung stand, äußerten sich die Befragten spontan sehr positiv zu dem Angebot. Die Möglichkeit, individuell gestaltete Briefmarken ausdrucken und versenden zu können, stieß unmittelbar auf großes Interesse. Positiv hervorgehoben wurde auch die Kürze und Prägnanz des Prozesses, der auf den ersten Blick unkompliziert und einfach zu durchlaufen wirkte.

Der Haupt-Benefit wurde sowohl von Privat- als auch Gewerbekunden (vor allem Freiberufler und kleinere Unternehmen) darin gesehen, Briefmarken einfach und bequem von zu Hause aus drucken und diese individuell gestalten zu können – eine Erwartung, die hohe Ansprüche an Gestaltung und Usability der Seiten stellt.

Abbildung 28: Überblick über die vier Haupt-Prozessschritte in der ersten Testwelle des Bestell- und Druckprozesses der „Internetmarke" (Quelle: Deutsche Post World Net/phaydon).

Die Ergebnisse sollen im Folgenden am Beispiel von zwei Seiten aufgezeigt werden: „Briefmarke auswählen“ und „Briefmarken drucken“.

Prozessschritt „Briefmarke auswählen“: Schwachstellen und iterative Optimierung
Abbildung 29 verdeutlicht, dass die Seite über drei Explorationsrunden bis zum finalen Screen sukzessive optimiert wurde (die finale Optimierung erfolgte im Anschluss an die dritte Explorationsrunde). An dieser Stelle ließe sich kritisch die Frage stellen, warum nicht ein gründlicher Usability-Test direkt alle Schwachstellen aufgedeckt hat? Dafür gibt es zwei Gründe: Zum Einen verändert sich aus gestaltpsychologischer Sicht die Gesamtwahrnehmung einer Seite, wenn einzelne Teile verändert werden. So treten bestimmte Schwachstellen erst dann zutage, weil bzw. nachdem andere Details geändert werden. Der andere Grund liegt darin, dass das Vorgehen in einem Rapid Prototyping prozessorientiert erfolgt, also nicht nur der Status Quo nach bestem Wissen evaluiert, sondern das *Optimum* im Ablauf des Prozesses realisiert werden soll. Dies ist nur mit einem iterativen Vorgehen möglich, in dem gegebenenfalls einmal gemachte Veränderungen weiterentwickelt, revidiert oder gegebenenfalls wieder ganz zurückgenommen werden können.

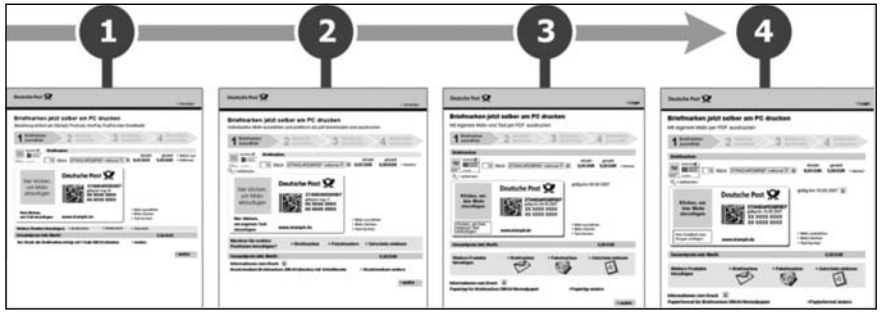


Abbildung 29: Iterative Optimierung der Seite „Briefmarke auswählen“ in vier Schritten (Quelle: Deutsche Post World Net/phaydon).

Zusammenfassend traten folgende Schwachstellen zutage, die in der Endversion berücksichtigt und umgesetzt wurden (siehe Abbildung 30):

1. Zur besseren Orientierung wurde die Überschrift so geändert, dass die Zielsetzung „Mit eigenem Motiv per PDF ausdrucken“ direkt beim Einstieg deutlich wird.
2. Hauptproblem war die fehlende Angabe des Briefmarkenpreises im Pull-Down-Menü, die in der Endversion ergänzt wurde.

3. Der Gültigkeitszeitraum der Internetmarke wurde in der Endversion prägnanter hervorgehoben, da er häufig übersehen wurde; außerdem wurde ein Info-Icon hinzugefügt, das dem Nutzer weitere Informationen zur begrenzten Gültigkeit bei Bedarf zur Verfügung stellt.
4. Das Wording für die Eingabe des Textfeldes („Hier klicken, um Text einzufügen“) wurde häufig missverstanden („welcher Text?“) und in der Endversion ersetzt durch „Hier Grußbotschaft oder Slogan einfügen“.
5. Der Hinweis auf weitere Produkte, die hinzugefügt werden können, ging unter und wurde in der Endversion durch intuitiv verständliche Icons ergänzt.
6. Da die Nutzer häufig Fragen zum Druckvorgang und zu erforderlichen Papierarten stellten, wurde in der Endversion ein Info-Icon zu „Informationen zum Druck“ hinzugefügt.

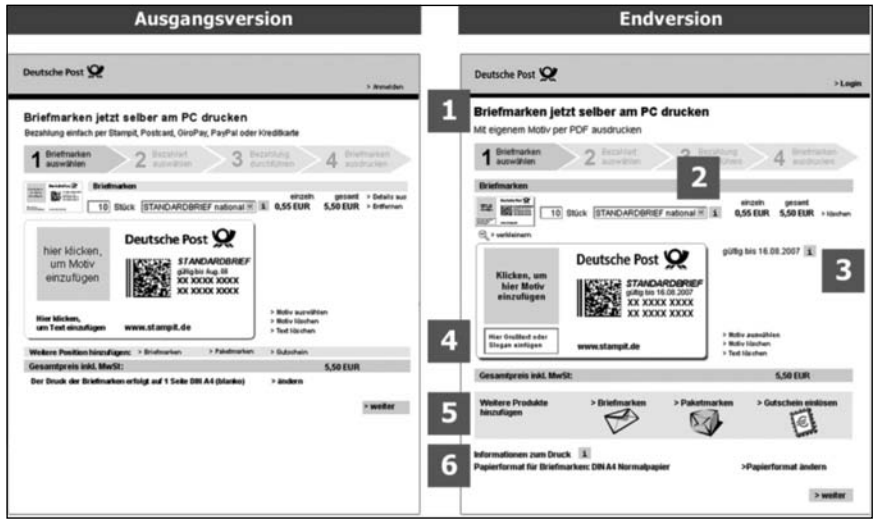


Abbildung 30: „Briefmarke auswählen“ – Ausgangs- und Endversion (Testwelle 1) (Quelle: Deutsche Post World Net/phaydon).

Prozessschritt „Briefmarke drucken“: Schwachstellen und iterative Optimierung

In diesem Prozessschritt werden das Ergebnis der Auswahl angezeigt und die druckfähigen PDF-Dateien zur Verfügung gestellt. Hier wurden folgende Schwachstellen aufgedeckt und in der Endversion optimiert (siehe Abbildung 31):

1. Die Handlungsaufforderung zum Ausdrucken der PDF-Dokumente trat nicht klar genug hervor („Ihre Unterlagen zum Druck“). Deshalb wurde in der Endversion der Text im oberen Bereich reduziert und eine klare Handlungsaufforderung formuliert: „Drucken Sie hier Ihre Briefmarken aus“.
2. Fachtermini zur Papiergröße („mit Schnittkante“ oder Ähnliches) wurden durch die praktische Anwendung ersetzt („zum Ausdruck auf A4 Normalpapier“).
3. Die umständlichen und textlastigen Ausführungen zum Umgang mit dem PDF-Dokument („Unser Tipp für den Druck“) wurden durch einen kurzen und gängigen Hinweis auf den Download der Adobe Software ersetzt und weniger auffällig unter die eigentlichen Hauptdokumente platziert.

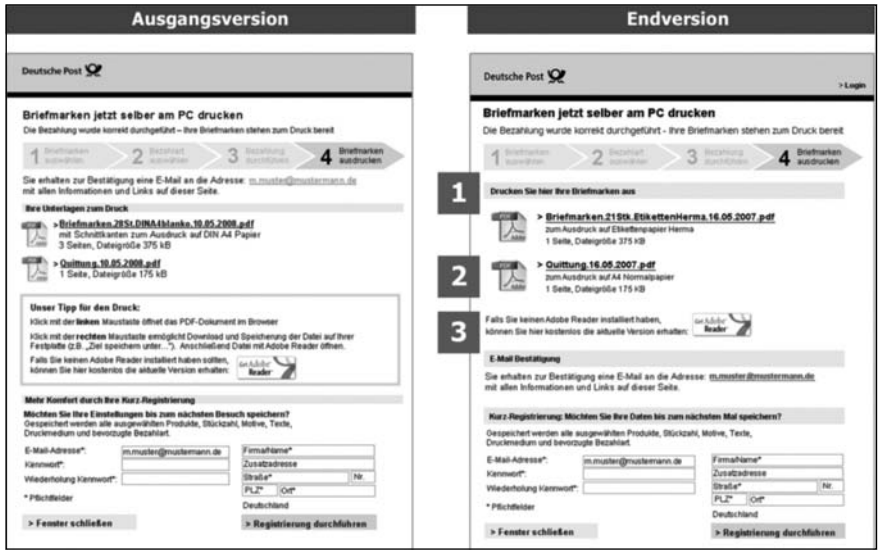


Abbildung 31: „Briefmarke drucken“ – Ausgangs- und Endversion (Testwelle 1) (Quelle: Deutsche Post World Net/phaydon).

Ausgewählte Ergebnisse der zweiten Welle (2008)

Der auf Basis des ersten Rapid Prototypings optimierte und weiterentwickelte Prototyp wurde zwischenzeitlich um weitere Prozessschritte ergänzt, darüber hinaus ging es neben der Optimierung der Prozesslogik auch um die Ansprache-Tonalität und Kommunikation des Angebots selbst.

Optimierung des Bezahlprozesses

In der zweiten Welle wurde im Detail auch die Prozessabfolge in den einzelnen Bezahlschritten überprüft. Hier standen den Nutzern grundsätzlich verschiedene Zahlarten zur Verfügung. Neben den gängigen Zahlarten wie Lastschrift, Kreditüberweisung und Zahlung mit PayPal wurde den Nutzern auch eine Onlineüberweisung mittels giropay angeboten.

Das Online-Zahlverfahren giropay wurde innerhalb der deutschen Kreditwirtschaft entwickelt und garantiert mit der Postbank, den Sparkassen, den Volksbanken und Raiffeisenbanken nach eigenen Angaben die höchsten Sicherheitsstandards. Wenn der Nutzer sich in einem Online-Shop für die Bezahlung mit giropay entscheidet, wird er aus dem Shop heraus über eine sichere Verbindung direkt zum Online-Banking-Konto seiner Bank oder Sparkasse geleitet, wo die Überweisung für den Einkauf veranlasst wird. Das Besondere dabei ist, dass der Händler eine sofortige Zahlungsbestätigung bekommt, so dass er die Ware umgehend an den Kunden versenden kann. Im Rahmen des Rapid Prototypings wurde nun die grundsätzliche Fragestellung untersucht, wie der Übergang von der Seite der Internetmarke auf die durch giropay initiierte Bankseite gestaltet sein muss, damit die Nutzer nicht abgeschreckt werden. Im Detail ging es um folgende Fragen: Welche Informationen benötigt der Nutzer beim Übergang? Wird die Weiterleitung als vertrauenswürdig erlebt? Wie erlebt der Nutzer die Weiterleitung auf die Zahlungsseite seiner Bank?

In der Ausgangsversion (siehe Abb. 33) führte der Klick auf den Button „weiter zur Bezahlung“ bei der Auswahl der Bezahlart giropay direkt auf die externe Bezahlseite von giropay (hier: eine spezielle Einstiegsseite der Stadtsparkasse). Da viele Nutzer auf den abrupten und unvorbereiteten Wechsel irritiert und verunsichert reagierten (es war kein Hinweis auf die Weiterleitung platziert), wurde in der optimierten Variante eine vorgeschaltete Zwischenseite im Look & Feel der Internetmarke entworfen, auf der die Weiterleitung angekündigt wird. Eine animierte Fortschrittsanzeige zeigte darüber hinaus die Zeit bis zur Weiterleitung an, so dass der Wechsel für die Nutzer nachvollziehbar wurde. In den weiteren Runden wurde die Zwischenseite so optimiert, dass der Übergang bzw. die Bezahlung mit giropay von den Nutzern als problemlos und sicher erlebt wurde – ein wichtiges Kriterium auch für die Gesamtakzeptanz der Internetmarke.

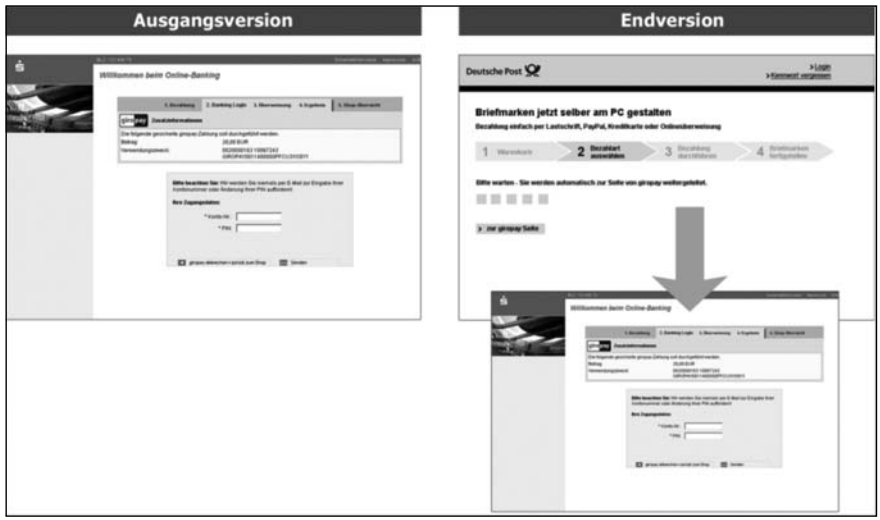


Abbildung 32: „Bezahlung durchführen" am Beispiel von giropay – Ausgangs- und Endversion (Testwelle 2) (Quelle: Deutsche Post World Net/phaydon).

3. Fazit: Durch iteratives Prototyping zum Erfolg

Mit Hilfe der beiden aufeinander folgenden Rapid Prototypings konnte der Entwicklungsprozess der Internetmarke von Beginn an so optimiert werden, dass Prozessablauf und Kommunikation des Angebots genau auf die Bedürfnisse der späteren Nutzer zugeschnitten wurden. Da in den Rapids immer wieder neue Umsetzungsvarianten ausprobiert und getestet wurden, konnten auch innovative Features auf Akzeptanz hin überprüft werden. Bereits im ersten Rapid wurde deutlich, dass sich die Nutzer die Auswahl eigener Motive – zum Beispiel private Fotos oder Bilder – wünschen, so dass auch die Möglichkeit des Uploads bzw. der Auswahl eigener Motive testweise implementiert und über das Rapid optimiert wurde.

Dies erforderte eine kooperative Zusammenarbeit zwischen allen Projektbeteiligten, die von Anfang an gegeben war. An allen Debriefing-Sitzungen nahmen neben den UX-Experten von phaydon auch Vertreter der Fachabteilung und ein Vertreter der unternehmensinternen Marktforschung (MRSC) teil. Der Rolle des Marktforschers kam in diesem Prozess eine wichtige Bedeutung zu, da dieser die zum Teil unterschiedlichen Sichtweisen zwischen UX-Perspektive und Unternehmenssicht konstruktiv vermitteln konnte.

Der Erfolg spricht für sich: Neben den kontinuierlich steigenden Nutzungszahlen als quantitatives Ergebnis wurde die Internetmarke bereits kurz nach der Implementierung in einer Kooperation mit dem Burda-Verlag getestet. 500 Nutzer der LISA Online Community erhielten Gutscheine für die einmalige Nutzung der Internetmarke und wurden gebeten, ihr Feedback im Online-Forum zu posten. Dieses war ausgesprochen positiv, sowohl im Hinblick auf Gesamtakzeptanz (93 Prozent bewerteten die Internetmarke mit gut oder sehr gut), Wiederverwendungsbereitschaft (90 Prozent) und Benutzerfreundlichkeit (80 Prozentgeben an, keinerlei Probleme gehabt zu haben). Die Ergebnisse stimmen positiv, zeigen aber auch, dass die Weiterentwicklung der Internetmarke kontinuierlich einer Überprüfung durch die Nutzer unterzogen werden muss. Nur so kann gewährleistet werden, dass das Angebot auch weiterhin auf die hohe Akzeptanz bei den Kunden trifft.

VI.II Fallbeispiel: Rapid Prototyping für netdoktor.de

von Carsten Rehmann

1. Der Untersuchungsgegenstand 138
2. Vorarbeiten 138
3. Ziele des Tests 139
4. Vorgehensweise 139
5. Ergebnisse – Übersicht über die Probleme und Empfehlungen 141
6. Ausgewählte Erkenntnisse 142
7. Learnings 144

Im Rahmen eines iterativen Nutzertests (Test in mehreren Schleifen) wurde der Prototyp des neuen Internet-Angebots von NetDoktor.de, der Ende 2008 online ging, getestet. Neue Anforderungen und Bedürfnisse der Nutzer und Redakteure führten zu einer grundlegenden Umgestaltung des Internet-Angebots – sowohl im Frontend als auch im Backend. Der gesamte Umstellungsprozess nahm etwa ein Jahr in Anspruch.

1. Der Untersuchungsgegenstand

NetDoktor.de ist das älteste unabhängige Gesundheitsportal in Deutschland, welches seit 1999 online ist. Seit ca. einem Jahr gehört NetDoktor zum Holzbrink-Verlag. Die Seite hat circa 1,2 Millionen Nutzer pro Monat.

Eine unabhängige Redaktion von Journalisten, Medizinern und Biologen erstellt dabei die Inhalte zu den Themen „Medikamente", „Krankheiten" und „Symptome". Nutzer können sich darüber informieren, wie Eingriffe und verschiedene Untersuchungen ablaufen. Es sind etwa 600 Krankheiten in unterschiedlicher Tiefe und Breite dargestellt. Neben der Informationsrecherche können Nutzer auch ihr Wissen testen.

Im Rahmen der Umstrukturierung wurde über eine Online-Befragung die Nutzerstruktur und deren Intentionen erfasst. Mit den Ergebnissen war es möglich, das Angebot Zielgruppengerecht zu entwerfen. Viele Nutzer gelangen über Google auf das Internet-Angebot von NetDoktor.de. Sie kommen oft kurz vor oder kurz nach einem Arztbesuch auf die Seite (70 Prozent). Häufig mit Diagnosen, die chronische und/oder schwerwiegende Krankheiten beinhalten. Dem Seitenbetreiber ist es wichtig, den betroffenen Nutzer auf einfache und angenehme Art und Weise an die wichtigen Informationen für sein Krankheitsbild heranzuführen.

2. Vorarbeiten

Über eine Onsite-Umfrage auf der alten Seite von netdoktor.de wurden die typischen Nutzer auf NetDoktor ermittelt und Personas erstellt. Zu den Zielgruppen gehören die „leicht" Kranken (Grippekranken), medizinisches Personal, besorgte Väter und auch die chronisch und schwer Kranken.

Im Vorfeld wurden rudimentäre Klickdummys sowie Screenshots des neuen Angebots durch Experten evaluiert. Die im Rahmen der Expertenevaluation festgestellten Probleme und Empfehlungen dienten als Basis, den Prototyp für den Test anzupassen. In den Test ging dann ein Prototyp, mit dem die zentralen Test-Cases, wie z. B. Recherche über das Krankheitsbild „Asthma“ oder eine Suche nach Forenbeiträgen, durchgeführt werden konnte.

3. Ziele des Tests

Ziel des Tests war es zu ermitteln, ob zentrale Interaktionsprozesse der Kernzielgruppen und das neue Navigationskonzept verstanden wurden.

4. Vorgehensweise

Im Juli 2008 wurden zwei Testschleifen angesetzt, in der jeweils 6 Nutzer aus den Zielgruppen den Prototyp testeten. Während des Tests sollten die Nutzer verschiedene Aufgaben bzw. Szenarien durchlaufen. Dabei wurde beobachtet, wie sie vorgegangen sind. Zudem wurden fokussierte Interviews durchgeführt, um Ursachen für Missverständnisse und Fehlbedienungen zu ermitteln.

Zu Auswertungszwecken wurde der Test mit einer speziellen Screencapture Software aufgezeichnet und die Aussagen der Testpersonen mitprotokolliert.

Jeder Test dauerte circa 75 Minuten. Die Erstellung des Studienkonzeptes nahm insgesamt drei Werktage in Anspruch. Die Zeitdauer für die Datenerhebung und -Auswertung belief sich auf sechs Werktage.

Am ersten Tag fand die Datenerhebung in Göttingen statt. Dabei waren die beteiligten Agenturen (Webdesign und Usability) und die Projektverantwortlichen seitens NetDoktor zugegen. Am Abend des ersten Tages diskutierten alle Projektbeteiligten gemeinsam die beobachteten Probleme und entwickelten erste Lösungsansätze. Am zweiten Tag erfolgte eine detaillierte Auswertung seitens der Usability-Experten und Herleitung von konkreten Handlungsempfehlungen. Nach Lieferung des Ergebnisberichtes zur ersten Testschleife hatten die Design-Agentur und die Programmierer zwei Tage Zeit für die Umsetzung, bevor am fünften Projekttag die zweite Erhebungsschleife mit weiteren sechs Nutzern aus den Zielgruppen stattfand. Am fünften Tag wurde geprüft, ob die Qualität und die Umsetzung

der Empfehlungen zu Verbesserungen im Umgang mit dem Prototypen geführt haben. Auch wurden neue Probleme entdeckt. Hierbei ist interessant, dass die Probanden auf einige der neu entdeckten Probleme sicherlich nicht gestoßen waren, da sie bereits an den in der ersten Testschleife entdeckten Problemen scheiterten (Stichwort „Überlagerung von Problemen“). Am sechsten Tag konnte dann die abschließende Auswertung erfolgen und ein detaillierter Endergebnisbericht fertig gestellt werden.

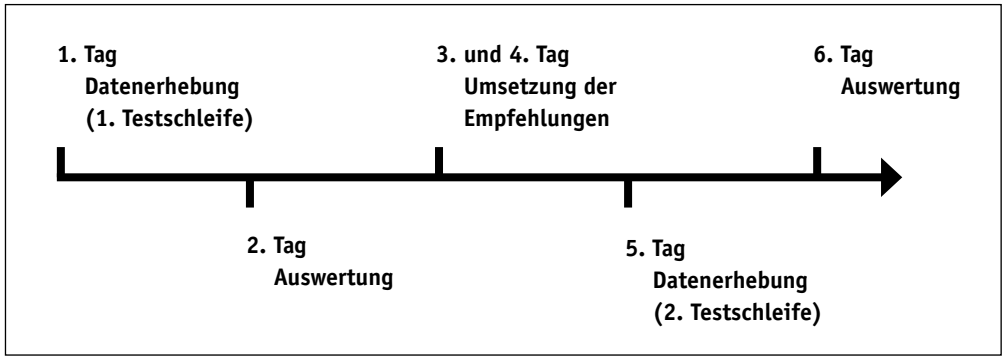


Abbildung 33: Zeitschiene der Datenerhebung und Auswertung.

5. Ergebnisse – Übersicht über die Probleme und Empfehlungen

In der ersten Testschleife wurden durch die sechs Probanden 15 leichte, mittlere und größere Probleme entdeckt, für die Handlungsempfehlungen gegeben wurden. Im zweiten Test zeigte sich bei der Kontrolle der umgesetzten Empfehlungen, dass die meisten Empfehlungen angenommen wurden und für Nutzer diesbezüglich keine Probleme mehr darstellten. Bei vier Aspekten, traten auch im zweiten Test Probleme auf. Hier mussten die Empfehlungen überarbeitet werden. Des Weiteren wurden neue Probleme festgestellt, die im ersten Test kein Nutzer aufdeckte.

Problembereiche	**1. Testschleife** gefundene Probleme	**2. Testschleife** Kontrolle der Empfehlungen		neu gefundene Probleme
Startseite	2	✓✓	–	–
Navigation	1	✓	–	1
Suche	5	✓✓✓	××	2
Artikelseite	4	✓✓✓	×	1
Katalog	2	✓	×	0
Forum	0	–	–	1
Footer	1	✓	–	0
			Empfehlung optimiert	neue Empfehlung

Abbildung 34: Übersicht über Probleme und Empfehlungen, Häkchen stehen für behobene, Kreuze für nicht behobene Probleme

6. Ausgewählte Erkenntnisse

Auf der Einstiegsseite des Internet-Angebotes wurde im Einstiegsteaser ein Bild dargestellt, durch das sich nicht jeder Nutzer angesprochen fühlte. Da zudem der ‚Weiter'-Button, mit dem sich andere Themen anzeigen ließen, nicht nutzergerecht gestaltet wurde, unterschätzten einige Nutzer die Vielfalt des Angebots. Da aber oft die erste Seite darüber entscheidet, ob ein Nutzer auf einer Seite verweilt oder über Google nach einem anderen Angebot recherchiert, ist es wichtig die Angebotsvielfalt schnell zu kommunizieren.

Abbildung 35: Problem

Als Lösung wurde eine seitliche Navigation in Form von Reitern bzw. Registerkarten (Tabs) vorgeschlagen, damit Nutzer in der Lage sind zwischen verschiedenen Einstiegsthemen zu wechseln. Zudem sollte das Einstiegsthema alle 5 Sekunden automatisch wechseln (dezenter Übergang).

Die These für die zweite Testschleife war, dass der animierte Wechsel die Nutzer nicht stört und die Inhalte der Webseite besser kommuniziert.

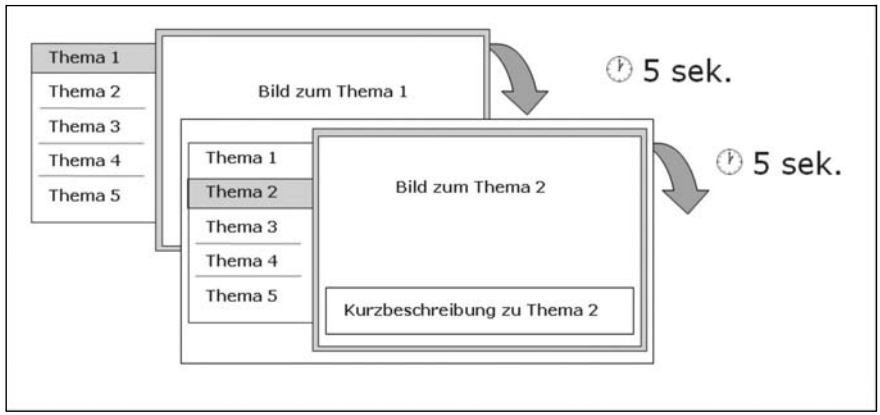


Abbildung 36: Lösung

Abbildung 37: Umsetzung für den zweiten Test

Ein weiteres Problem in der ersten Testschleife stellte die Benutzung von interaktiven Flächen bei der Aufgabe: „Suche nach Informationen zum Krankheitsbild Asthma“ dar.

Im Lexikon ist die Krankheit in verschiedene Punkte wie zum Beispiel Wissen, Ursachen etc. gegliedert (linke Spalte in der Abbildung). Die Nutzer wollten einen Unterpunkt des Themas „Wissen“ auswählen (diese sind in der rechten Spalte). Dabei fuhren sie mit der Maus über eine andere Kategorie in der linken Spalte. Da die linke Spalte mit den Kategorien interaktiv ist, wurde die rechte Liste beim „drüberfahren“ automatisch aktualisiert.

Dadurch wurde die Auswahl einer Unterrubrik erschwert, weil Nutzer zuerst den Mauszeiger waagerecht bewegen mussten.

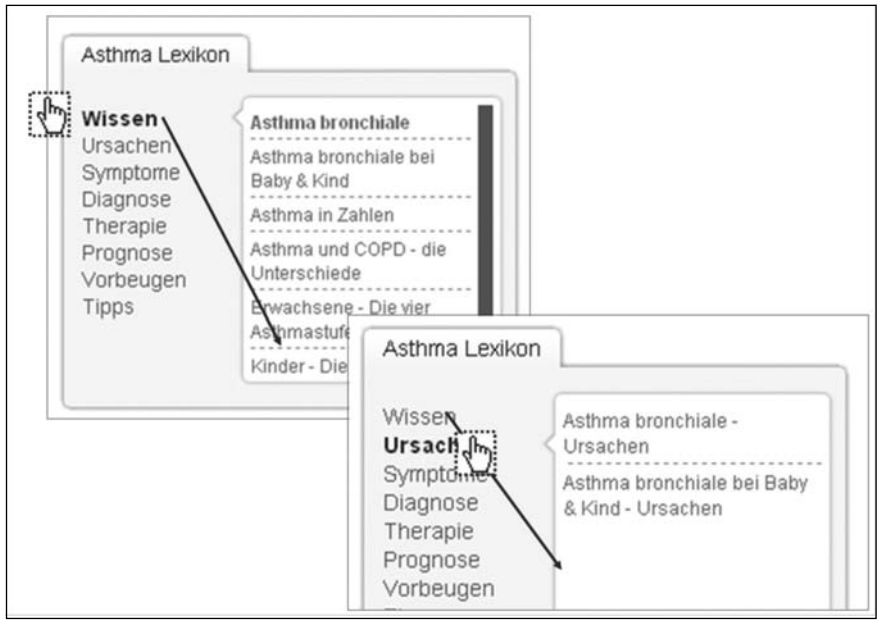


Abbildung 38: Problem bei interaktiven Flächen

Als Lösung wurde ein Verzicht auf die interaktive Fläche vorgeschlagen. Das Untermenü im Lexikon sollte sich zukünftig erst per Mausklick ändern.

Die zweite Testschleife zeigte, dass beide Empfehlungen akzeptiert bzw. als leichter zu bedienen beurteilt wurden.

7. Learnings

Es hat sich gezeigt, dass die Kommunikation mit den beteiligten Projektpartnern (Seitenbetreiber, Programmierer, Designer, Usability-Agentur) sehr wichtig ist. Insbesondere vor dem Hintergrund, dass die Empfehlungen zur Behebung der Probleme, die sich während einer Testschleife gezeigt haben, für die nächste Testschleife umgesetzt werden müssen. Dies erfordert genaue zeitliche Abstimmungen und Umsetzungen der Empfehlungen, da

nur dann der Erfolg getestet werden kann, wenn die Empfehlungen (genau) umgesetzt wurden.

Da bei den Tests alle Projektpartner zugegen waren, konnten „vor Ort" gemeinsam Handlungsempfehlungen für die aufgedeckten Probleme entwickelt und sofort mit der Umsetzung begonnen werden. Programmierer und Entwickler kennen die plattformspezifischen Restriktionen, Designer wissen wie Änderungen im Corporate Design umsetzbar sind und Usability-Experten bringen Branchenwissen (?)sowie das Wissen über benutzerfreundliche Gestaltungsrichtlinien mit ein.

Die Gruppengröße in den einzelnen Testschleifen könnte verändert werden. Zum Beispiel könnten in der ersten Testschleife mehr Probanden eingeladen werden, damit anfangs mehr unterschiedliche Probleme aufgedeckt werden. In weiteren Testschleifen sind dann weniger Probanden nötig, um die Handhabbarkeit der umgesetzten Empfehlungen zu überprüfen. Auch könnte die Stichprobe in der zweiten bzw. dritten Testschleife auch zum Teil (circa 50 Pozent) aus Probanden aus vorherigen Testschleifen bestehen, da diese am Besten beurteilen können, ob eine umgesetzte Empfehlung das aufgedeckte Problem löst. Die andere Hälfte der Testpersonen wäre für „neue" Probleme zuständig.

In der zweiten Testschleife wurden einige neue Probleme aufgedeckt und einige Empfehlungen führten nicht vollständig zum gewünschten Erfolg und mussten angepasst werden. Daher ist es ratsam eher drei Testschleifen anzusetzen.

Abschließend bleibt anzumerken, dass ein iterativer Test nicht unbedingt teurer sein muss als ein „normaler" Nutzertest, der nur einen Testdurchgang vorsieht. Der Test wird mit etwa der gleichen Anzahl an Testpersonen durchgeführt – nur eben in verschiedenen Erhebungsschleifen. Zwar muss etwas mehr Zeit für die Koordination eingeplant werden, aber im Idealfall reduziert sich der Aufwand in späteren Testschleifen bzw. Verfeinerungsstufen, da nur noch für kleinere Probleme Handlungsempfehlungen gegeben werden müssen.

„Durch den iterativen Usability-Test wurde der Prototyp der neuen Seite schrittweise und zügig in vielen Details optimiert. Die reibungslose Zusammenarbeit aller Beteiligten (NetDoktor.de/eResult GmbH/it doesn´t look like a sculpture (idllas.com)) führte schnell zu guten Ergebnissen."

Philipp Montgelas, CEO von NetDoktor.de:

VII. Mouse Tracking und Klickanalysen

von Dirk Schulze

1. Usability authentisch testen 148
2. Technik 149
3. Heatmaps von Klicks und Mausbewegungen 150
4. Webformulare optimieren 151
5. Ausblick: Aktivitätsdauer statt Page Impressions? 152

Bekannte Webanalyse-Lösungen geben bisher einen Überblick über den Traffic auf einer Website. Über welche Links und Suchbegriffe die Besucher kommen, wie viele Seiten sie besuchen, wie lange sie bleiben, ob sie schon mal da waren. Entsteht eine Diskrepanz zwischen diesen Zahlen und den angestrebten Zielen der Website, kann diese zwar festgestellt werden – die Ursachen für das Problem aber bleiben oft verborgen.

Es ist relativ leicht feststellbar, wie viele Besucher eine Website an einem bestimmten Punkt verlassen haben. Herauszufinden warum sie das getan haben, ist schon schwieriger. Hier setzen Mouse-Tracking-Tools das Vergrößerungsglas an und machen detailliert sichtbar wie Nutzer mit einer Website interagieren, welche Bereiche das größte Interesse wecken und auf welche Probleme sie stoßen.

1. Usability authentisch testen

Seitenbetreiber können ihre Besucher mit Hilfe von solchen Tools beim Surfen beobachten. Auf der technischen Seite muss dazu nur eine Zeile Javascript-Code in den Quelltext der Seite eingefügt werden. Ist der aktiv, lassen sich sämtliche Mausbewegungen, Klicks und Tastatureingaben jedes Besuchers einer Website aufzeichnen. Anschließend können sie als Video bequem im Browser abgespielt werden. Ein Klick auf den Play-Button genügt und man sieht den Mauszeiger eines Besuchers über die eigene Website wandern. Probleme in der Bedienbarkeit offenbaren sich durch beobachten des Nutzerverhaltens, ganz so als würde man tatsächlich hinter dem Nutzer stehen und ihm über die Schulter schauen. Einblicke wie diese waren bisher nur in aufwendigen Labortests mit Probanden möglich, und blieben dementsprechend umsatzstarken Websites vorbehalten. Doch selbst bei der Investition in einen solchen – ohne Frage aufschlussreichen – Usability-Test bleibt das Problem der experimentellen Effekte: Die Versuchspersonen befinden sich in einer hochgradig künstlichen Situation und wissen, dass sie beobachtet werden. Außerdem besteht immer die Gefahr, dass von einer sehr kleinen Gruppe auf die Masse geschlossen wird. Mouse-Tracking bietet sich hier an, gefundene Usability-Probleme hinsichtlich ihrer Schwere und Verbreitung zu überprüfen. Das ist die große Stärke der Methode: Sie eröffnet Einblicke in das Verhalten echter Besucher unter realen Bedingungen.

2. Technik

Mouse-Tracking wird meist als webbasierter Dienst angeboten, der ohne Software-Installation auskommt. Um den Service zu nutzen, muss lediglich ein Stück Javascript-Code in den Quelltext der zu trackenden Website eingefügt werden. Dieser Code sammelt Informationen über die Mausoperationen auf einer Website und sendet die Daten an den Anbieter. Der Betreiber der Website kann sich dann dort einloggen und die Aufnahmen abspielen, sowie Heatmaps oder Clickmaps abrufen. Die Anzahl der aufzunehmenden Besuche kann zwischen einem und 100 Prozent beliebig variiert werden. Passwörter werden nicht aufgezeichnet. Für die Nutzer der Website entstehen keine spürbaren Veränderungen.

Praxisbeispiel

In einen Onlineshop wurden ungewöhnlich oft volle Warenkörbe nicht abgeschickt. Mit Hilfe von Webanalyse-Daten ließ sich das Problem zwar feststellen, die Ursache aber blieb rätselhaft. Ein Mausbewegungs-Film zeigte: Der „Bestellung abschicken"-Button war so ungünstig platziert, dass er sich bei einer niedrigen Auflösung außerhalb des sichtbaren Bildschirms befand. Unerfahrene Nutzer kamen nicht auf die Idee, nach unten zu scrollen und konnten ihn deshalb nicht sehen. Sie brachen ab, obwohl sie eigentlich kaufen wollten.

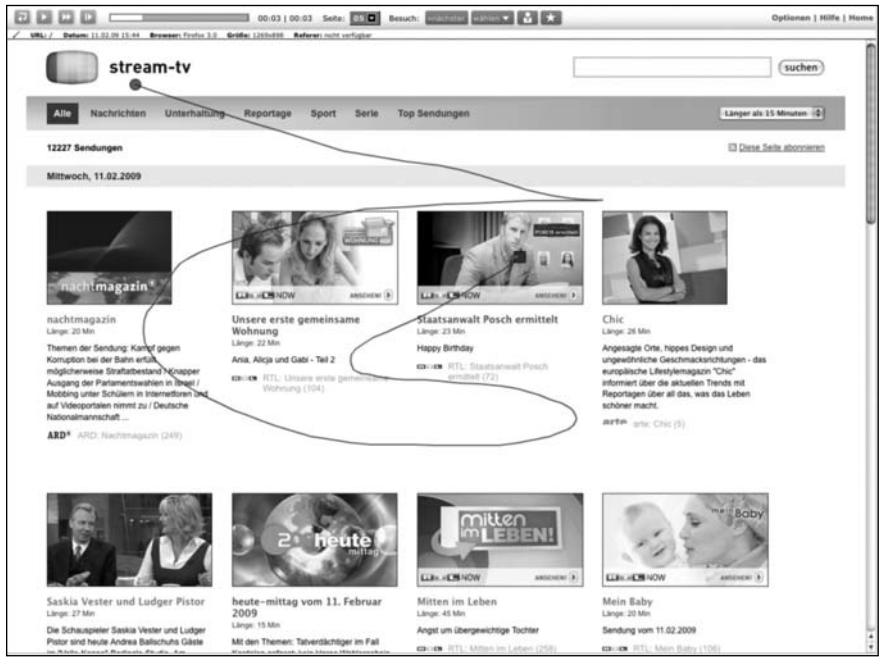


Abbildung 39: Darstellung einer Mausspur (m-pathy)

3. Heatmaps von Klicks und Mausbewegungen

Natürlich kann sich niemand alle Besuche auf einer Website einzeln anschauen. Das ist auch nicht nötig. Die aufgezeichneten Mausdaten lassen sich zu Heatmaps aggregieren. Diese gibt es sowohl für Klicks (Clickmaps) als auch für Mausbewegungen (Movementmaps). Auf den Wärmebildern, die über die Website gelegt werden, ist dann auf einen Blick zu sehen, welche Bereiche einer einzelnen Seite wie viel Interesse auslösen. Rot bedeutet viele Klicks oder Mausbewegungen an dieser Stelle, blau wenig. Die Clickmaps zeigen auf einen Blick, welche Buttons und Navigationselemente die Besucher wie häufig anklicken und sind damit besonders für die Analyse von Startseiten und Landing Pages sehr aufschlussreich. Die Anordnung der verschiedenen Elemente auf einer einzelnen Seite kann so optimiert werden. Außerdem zeigen Clickmaps, an welchen Stellen die Nutzer intuitiv Links vermuten, die aber nicht vorhanden sind. Häufig werden beispielsweise kleinere Bilder angeklickt, in der Hoffnung, es gäbe noch eine vergrößerte Ansicht. Movementmaps zeigen hingegen wie häufig Besucher mit dem Mauszeiger über einen Bereich gefahren sind. Sie können Hinweise auf die Verteilung des Aufmerksamkeitsfokus innerhalb einer Seite geben. Denn der Mauszeiger dient als Orientierungshilfe und bewegt sich oft parallel zum Auge, rund ein Drittel der Nutzer verwendet ihn als Lesehilfe. Es wird also sichtbar, welche Bereiche einer Seite überhaupt wahrgenommen werden sowie welche Textabschnitte gelesen werden und welche nicht. Studien haben beispielsweise gezeigt, dass sich die Mausnutzung bei schlecht lesbarem Text deutlich sichtbar verändert.

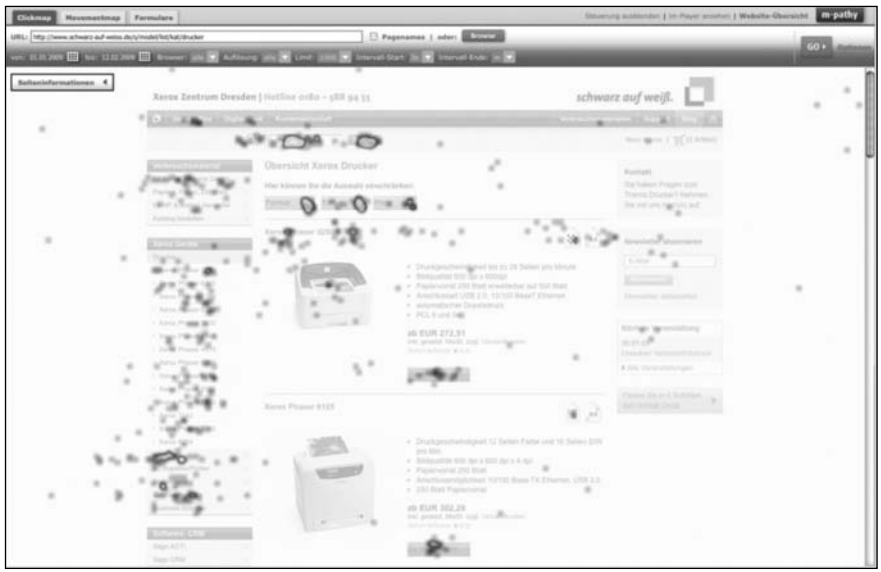


Abbildung 40: Darstellung einer Mausspur (m-pathy)

Praxisbeispiel

Von der Startseite eines Portals gab es zwei Wege zur Seite „Kontakt", einmal über einen kleinen Textlink am unteren Ende der Seite und einmal über einen Menüpunkt in der Hauptnavigation. Interessanterweise wurde der untere Link wesentlich häufiger angeklickt als der obere. Die Besucher haben den Button „Kontakt" aus Erfahrung eher im Fußbereich der Seite vermutet. Im Hauptmenü der Seite konnte so auf eine Schaltfläche verzichtet werden, was die Übersichtlichkeit erhöhte.

4. Webformulare optimieren

Dass Mouse-Tracking viel feiner granulierte Daten liefert als herkömmliche Webanalyse-Software wurde bereits erwähnt. Besonders deutlich wird dies bei der Analyse von Webformularen. Formulare stehen an entscheidenden Punkten der Interaktion zwischen Website und Nutzer wie der Registrierung oder dem Abschluss eines Einkaufs. Trotzdem sind die Abbruchraten hier noch immer besonders hoch.

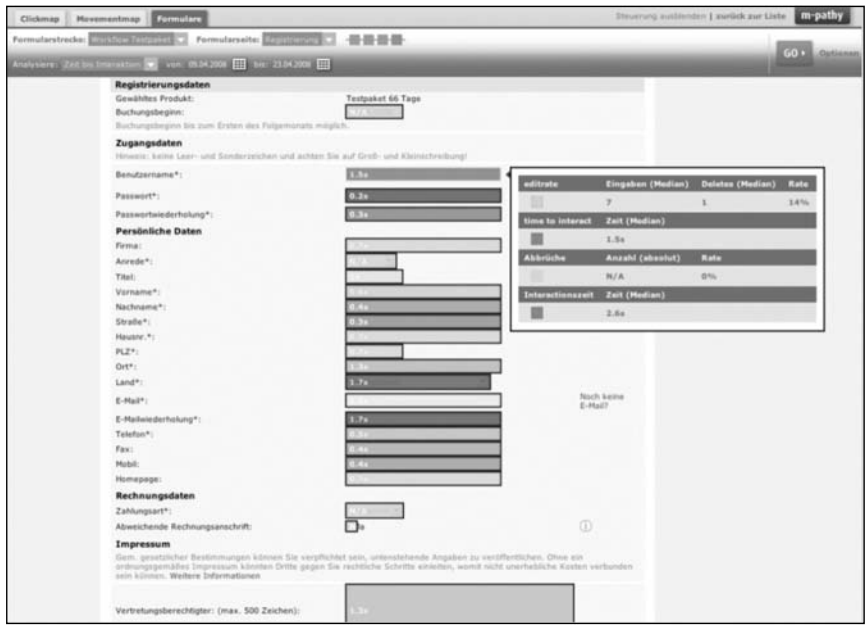


Abbildung 41: Formularanalyse

Durch die Mouse-Tracking-Technologie lassen sich Aussagen nun nicht mehr nur auf der Ebene von Seiten treffen, sondern für jedes einzelne Feld eines Formulars. Für welche Angabe wird überdurchschnittlich viel Zeit benötigt? Wie hoch sind die Abbruchraten für die einzelnen Formularfelder? Wie oft wurde eine Eingabe korrigiert? An welcher Stelle wurde die Hilfe aufgerufen? All diese Fragen lassen sich beantworten.

Praxisbeispiel

Auf der Website einer Versicherung konnten Kunden einen erlittenen Schaden per Webformular melden. Eine Analyse der Daten konnte zunächst ein Feld mit auffälligen Werten identifiziert werden. Gefragt wurde hier nach der geschätzten Schadenhöhe. Die genaue Ursache für das Problem zeigte sich dann bei der Sichtung der aufgezeichneten Besuche. In den Filmen war zu sehen, dass ein Teil der Besucher den Betrag inklusive eines Euro-Symbols eingegeben hatten (zum Beispiel 500 Euro). Das Formular akzeptierte aber nur reine Zahlenwerte und diese auch nur ohne Dezimalstellen. Erschwerend kam hinzu, dass die ausgegebene Fehlermeldung nur unspezifisch dazu aufforderte, die Angabe im betreffenden Feld zu korrigieren. Dass das Eurosymbol falsch war und nicht ihre Schätzung, mussten die Nutzer in mehreren Versuchen selbst herausfinden.

5. Ausblick: Aktivitätsdauer statt Page Impressions?

Als eine Art attraktives Nebenprodukt verrät die Aufnahme der Maus- und Tastaturaktivitäten auch die Zeitdauer, in der sich ein Nutzer wirklich aktiv mit einer Website beschäftigt. In Zeiten von dutzenden im Hintergrund geöffneten Browser-Tabs und der Diskussion um die Aussagekraft von Page Impressions als Onlinewährung dürfte das noch spannende Perspektiven eröffnen.

VIII. Einfache Tests mit A/B-Vergleichen

von Frank Reese

1. A/B-Tests: Schlicht aber effektiv 154
2. Praktische Herangehensweise 157

Was wäre, wenn man einfach mal etwas anderes ausprobieren würde? Wenn Unternehmen beginnen, die Bewegungen ihrer Besucher zu messen, stellt sich oft eine Mischung aus Stolz, Verwunderung, aber auch Enttäuschung ein: Zuerst ist es ja ganz beeindruckend, wenn in den Auswertungen von Google Analytics, Etracker, Econda oder wie die einschlägigen Systeme alle heißen, hunderte oder sogar tausende Besucher die Inhalte abrufen und sich über Services, Produkte und das Unternehmen informieren. Der zweite Blick ist dann aber oft hilflos: Viele Besucher sind es ja schon, aber warum bleiben die denn nicht länger? Warum ruft denn keiner an oder schreibt eine E-Mail?

Dabei werden Möglichkeiten zur Umgestaltung oft erst einmal nicht gesehen. Zu umständlich scheint die Umstellung und zu unsicher der Erfolg. Lieber vertraut man darauf, dass die „wirklich interessierten Besucher" sich schon zurechtfinden. Es ist doch alles da: Produkt- und Unternehmensinformationen, Neuigkeiten und Kontaktadressen – es muss also an den Besucher liegen, wenn diese dann doch nur kurz bleiben. – Vielleicht ist das so. Aber könnte es nicht doch sein, dass die Besucher die Website ganz anders als die Verantwortlichen sehen und andere Interessen und Ansprüche haben?

1. A/B-Tests: Schlicht aber effektiv

Um diese Ansprüche der Nutzer kurz, schmerzlos und ohne große Umstände herauszufinden, können A/B-Tests helfen. Sie sollen die Option eröffnen, einfach mal einen alternativen Vorschlag auszuprobieren. Die Idee dabei ist – in der einfachsten Version – denkbar schlicht: Welche von zwei Versionen, die sich in einem Element unterscheiden, zeigt die bessere Performance? Ein testendes, vergleichendes Vorgehen ist ja gar nichts Ungewöhnliches. Überlegt man, in welches Restaurant man mittags zum Essen geht, wählt man einfach die nächstbeste Möglichkeit. Am nächsten Tag wechselt man in ein anderes. Irgendwann sind alle Restaurants in der Gegend einmal ausprobiert worden, so dass sich offenbar sagen lässt: In diesem Restaurant ist das Essen am besten. Leider ist das meist nicht so einfach. Zu viele Variablen spielen eine Rolle bei der Einschätzung: Das konkrete Menü, vielleicht der Wochentag, die Stimmung und Begleitung, eventuell auch noch die Laune des Kochs oder der Bedienung. Der Vergleich der Restaurants leidet an der Komplexität der beinflussenden Faktoren. Zu viele Dinge spielen gleichzeitig eine Rolle, als dass ein allgemein geltendes Resultat erzielbar wäre. Auch die Performance von Webpages wird von einer Vielzahl von Faktoren bestimmt. Auch hier spielen Tageszeit und Wochentag eine Rolle, die Art der Motivation oder der soziodemographische Hintergrund des Besuchers uvm. Darum ist es für effektive Tests so wichtig, die Komplexität so gering wie möglich zu halten. A/B-

Tests heißen so, weil nur die beiden Varianten A und B gegeneinander getestet werden. Das ist das eiserne Gesetz des A/B-Tests: Eine Page mit zwei Variationen. Im Unterschied zu den in Kapitel IX vorgestellten multivariaten Verfahren wird der Test auf die einfachste mögliche Form heruntergefahren.

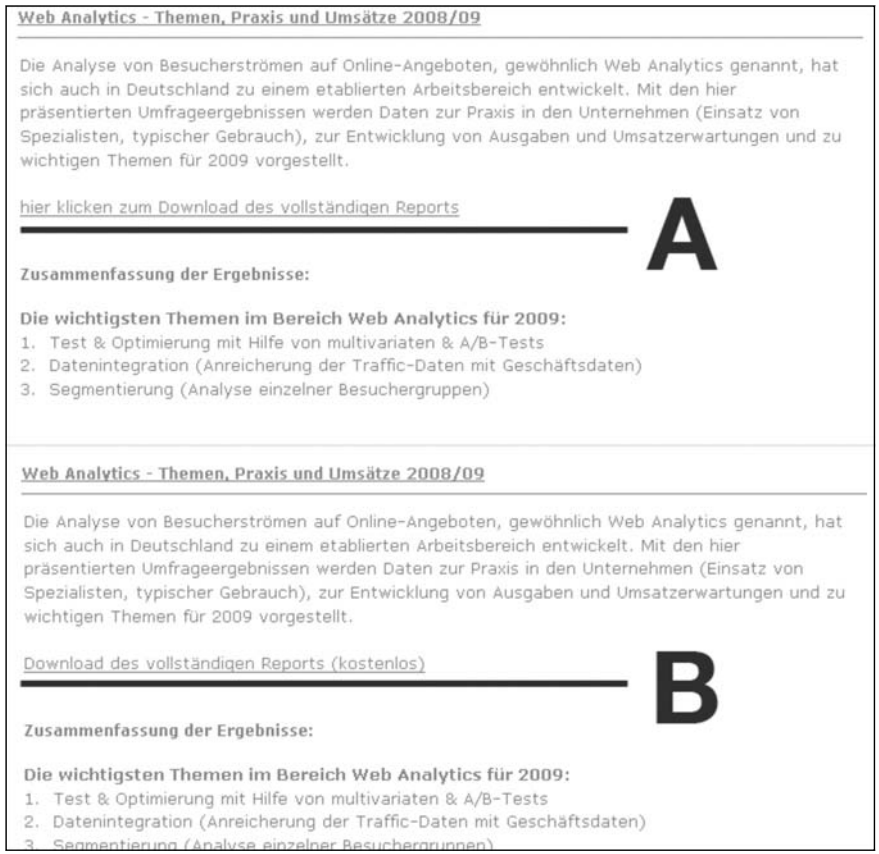


Abbildung 42: A/B-Test mit zwei Varianten (siehe auch das nachfolgende Fallbeispiel)

Die Abbildung 42 zeigt zwei Versionen einer Seite, die sich in ihrer Überschrift unterscheiden. Würden gleichzeitig noch die Farben, Button-Beschriftungen oder gar die Navigation variiert, würde der Test wiederum nur diffuse Ergebnisse liefern. Darum ist es so schwierig, die Performance einer Website vor ihrem Relaunch mit der Performance nach dem Relaunch zu vergleichen: Zuviel hat sich gleichzeitig verändert, als dass sich die Veränderungen auf einzelne Elemente zurückführen ließen. Selbst eine gleichzeitige Variation von nur zwei Elementen (zum Beispiel Überschrift und Button) macht die Analyse von insgesamt schon

vier Seitenvariationen notwendig und darüberhinaus auch der Interaktionen der Effekte untereinander (zum Beispiel Überschrift A funktioniert mit Button B besser als Überschrift B mit Button B, aber schlechter als Überschrift B mit Button A). Wegen dieser zusätzlichen Komplexität werden A/B-Tests auf den Vergleich von nur zwei Versionen beschränkt.

Wie bei jedem Verfahren der Optimierung ist es auch hier wichtig, genau zu wissen, was man mit der einzelnen Seite erreichen will. Ohne Zielformulierung vor dem Test gibt es keine Zielerreichung nach dem Test. Landing Pages oder auch andere Seiten mit einer sehr klaren, engen Zielstellung eignen sich darum besonders für A/B-Tests. Im einfachsten Fall wird eine Hälfte der Besucher auf Version A gelenkt und die andere Hälfte auf Version B. Dies kann auf sehr verschiedene Weise geschehen, aber am Ende erhält man immer eine Auswertung, die klare Ergebnisse ergibt, zum Beispiel: Version A hatte eine Absprungrate von 30 Prozent, Version B eine Absprungrate von 20 Prozent. Der Unterschied zwischen diesen beiden Ergebnissen kann dabei nur an der unterschiedlichen Überschrift liegen.

Für A/B-Tests gelten prinzipiell die gleichen Voraussetzungen hinsichtlich der Belastbarkeit der Ergebnisse wie für multivariate Verfahren: Je größer der zu erwartende Performance-Unterschied sein muss, um die Umgestaltung auch unter Kosten/Nutzen-Aspekten sinnvoll zu machen, desto proportional geringer sind die notwendigen Seitenabrufzahlen. Wenn eine Umgestaltung erst dann lohnenswert ist, wenn eine 20-prozentige Steigerung der Konversionsrate erfolgt, dann sind für belastbare Testergebnisse weitaus geringere Abrufzahlen notwendig, als wenn eine 2-prozentige Steigerung schon als vielversprechendes Ergebnis angesehen wird (zur Begründung dieses Zusammenhanges siehe Kapitel IX).

Wer einen A/B-Test angeht, sollte also schon vorher überlegen, was genau am Ende als Erfolg angesehen werden kann. Ebenso sind Effekte mangelnder Repräsentativität zu bedenken: Ein Test an Wochentagen kann zu anderen Ergebnissen führen als einer am Wochenende; Einflüsse von temporären Kampagnen, die die Traffic-Zusammensetzung ändern oder auch Effekte von aktuellen Pressemeldungen und Sonderangeboten können das Verhalten der Besucher verändern.

Gleichwohl liegt der Charme des A/B-Tests in seiner Schlichtheit. Er ist der kleine ungeschliffene Bruder des großen multivariaten Tests, das „Quick & Dirty"-Modell der Testfamilie. Darum entfaltet der A/B-Test seine größte Wirkung auch bei kleinen schnellen Fragestellungen, die präzise auf den Punkt kommen, aber methodologische Reinheit gegen pragmatisches Handeln eintauschen.

Optimierung mit Testreihen: Shoot-out der Alternativen

A/B-Tests können singulär eingesetzt werden, sie werden aber meistens als Testreihe angelegt, um nach und nach zu einer durchoptimierten Seite zu gelangen. Da der Test verhältnismäßig wenige Seitenabrufe verlangt und nach kurzer Zeit abgeschlossen ist, kann der nächste Test gleich im Anschluss folgen. Schritt für Schritt wird so die Klickrate zum Beispiel einer Landing Page immer besser. Sie kann nicht sinken, nur steigen. Der Grad der Steigerung wird zwar nach und nach immer flacher (bis der Aufwand weiterer Tests nicht mehr lohnt), aber nach dem einfachen Prinzip, dass nur die bessere Version überlebt, klettert die Performance immer weiter. Bei einem solchen sukzessiven Shoot-out ist allerdings zu beachten, dass der Gewinner späterer Shoot-Outs nicht notwendig den Verlierern früherer Tests überlegen ist. Z (besser als Y) muss nicht notwendigerweise besser als A sein (A war B unterlegen). Interaktive Effekte zwischen Variablen (wenn erst in mehreren Schritten die beste Überschrift gesucht wird und anschließend die beste Button-Gestaltung) rücken ebenso wenig in das Blickfeld.

A/B-Tests mit zurückliegenden Daten

Während der Vorbereitung für eine Testdurchführung taucht oft der Gedanke auf: Wenn wir doch nur vor der Veränderung dieser oder jener Seite auch schon Tests durchgeführt hätten! Aber Hilfe naht: Moderne Analytics-Tools bieten viele Möglichkeiten zur Segmentierung des Traffics, also der Filterung der Besucherbewegungen nach bestimmten Kriterien. Wenn die Veränderungen über Zeiträume eingegrenzt werden können – Montag, den 12.1., gab es noch Version A und Montag, den 19.1., gab es Version B – und die Veränderungen zwischen Version A und B nicht zu vielfältig sind, kann ein A/B-Test auch rückblickend gemacht werden. Natürlich ist die Aussagekraft nicht so hoch, wie bei einem sorgfältig geplanten Test, **aber es ist immer besser die Daten sprechen zu lassen als Vorurteile zu pflegen.**

2. Praktische Herangehensweise

Technisch gesehen gibt es drei Hebel, um einen A/B-Test in Angriff zu nehmen. Da die Auswertung des Tests keine besonderen mathematischen Analysen verlangt, liegt die Unterscheidung vor allem in der gleichmäßigen und steuerbaren Verteilung des Traffics auf die verschiedenen Versionen.

A/B-Tests mit dem Google Website Optimizer

Google hat mit seinem Website Optimizer allen Online-Verantwortlichen ein großes Geschenk gemacht. Was auch immer die Motivation dafür war: Mit dem Google Website Optimizer (GWO) eröffnen sich auch für Unternehmen, die keine fünfstelligen Beträge in Testsysteme investieren können, reichhaltige Testmöglichkeiten. Das Tool wird in Kapitel IX.III genauer vorgestellt, hier soll darum nur kurz auf die Funktionsweise eingegangen werden: In die eigene Website wird ein Javascript-Code integriert, der die Steuerung und Zählung der zu testenden Variationen übernimmt. Die Variationen selbst, also zum Beispiel Überschriftenformulierungen oder HTML-Anweisungen für den Abruf von Grafikdateien, werden in den Online-Formularen bei Google hinterlegt. Jedesmal, wenn ein Besucher die Seite abruft, bekommt er entsprechend des Testdesigns eine Version zugespielt, deren Abruf und Effekt (Wird die nächste Seite im Sinne der Konversion abgerufen?) bei Google registriert wird. Der komplette Test ist sehr einfach anzulegen und auch die Auswertung der Ergebnisse ist recht anschaulich. Allen Unternehmen, die die Kosten für ein kostenpflichtiges Testsystem (noch) scheuen und die bereit und in der Lage sind, den zusätzlichen Google-Javascript-Code in die Website zu integrieren, kann unbedingt und ohne Zweifel zu einem Ausprobieren des GWO geraten werden. Mit dem Tool kann nicht nur die Seite selbst optimiert werden, es ist zudem eine gute Möglichkeit, um den Umgang mit den neuen Verfahren und Herangehensweisen zu erlernen.

Vergleich von Zeiträumen

Im einfachsten Fall können die unterschiedlichen Seitenversionen auch einfach nacheinander getestet werden. Wenn es leichter ist, komplette Seitenversionen bzw. Grafiken oder Überschriften auszutauschen als ein Javascript zur Steuerung einzubauen, kann ein Test auch einfach an zwei verschiedenen Wochentagen (je nachdem, wie viele Seitenabrufe gebraucht werden und wie hoch das Trafficaufkommen ist) durchgeführt werden. Dass die Verantwortlichen dabei auf die Vermeidung unberechenbarer Seiteneffekte achten müssen, versteht sich von selbst. Der Test *kann* in dieser Weise durchgeführt werden, aber die Ergebnisse sind anfällig für Einwände hinsichtlich der Zuverlässigkeit. Mehr noch als bei den besser kontrollierbaren Testdesigns ist es hier überaus wichtig, dass sich alle Beteiligten bereits vorher darüber im Klaren sind, was sie von einem solchen Test erwarten können und was nicht. Wenn dies nicht der Fall ist, werden die Resultate zu keiner Entscheidung führen, da sich immer wieder Gegenargumente finden lassen.

Serverseitige Ansätze

Wird die Anzeige von Elementen über Javascript gesteuert, spricht man von clientseitiger Steuerung, da der Javascript-Code erst im Browser (client) des Besuchers ausgeführt wird. Im Unterschied dazu liefert bei einer serverseitigen Steuerung der Webserver selbst die Varianten für den Test. Wenn die Web-Administratoren eine zufällige Verteilung des Trafficstroms für die betreffende Seite über Weichen oder ähnliches installieren können, müssen sie dabei nur dafür sorgen, dass die unterschiedlichen Seiten in einem Traffic-Analyse-System wiedererkannt werden. In diesem Analyse-System kann dann über Segmentierungen eine Auswertung des Tests vorgenommen werden. Das Ganze hört sich vielleicht kompliziert an, aber bei einem eingespielten Team aus Marketing- und Technik-Experten kann diese eine gangbare Lösung sein.

Einen wiederum anderen Weg ermöglichen die Anbieter von Testsystemen auf PHP-Basis. Hierbei handelt es sich um kleine, recht günstige Lösungen, die von einem technisch erfahrenen Seitenbetreiber selbst installiert werden können und dann praktisch eine eigene Testplattform darstellen. Da Webserver und Testplattform auf dem gleichen System laufen, werden eventuelle Datenschutzbedenken (wie bei der Nutzung externe Plattformen) ausgeschaltet.

Adressen für kleine Testsysteme

Conversion Booster (www.l-space-design.com/Products/Conversion_Booster.aspx)
Split Test Accelerator (www.splittestaccelerator.com)
Power Split Tester (www.powersplittester.com)

VIII.I Fallbeispiel: A/B-Test für einen Download-Link

von Frank Reese

1. Der „Power Split Tester“ 162
2. Fazit 165

Im Herbst 2008 wurde eine Umfrage zum Thema „Web Analytics in der Praxis“ von Ideal Observer erstellt. Knapp über 60 Teilnehmer gaben Auskunft zu ihren Erfahrungen und ihren Zukunftseinschätzungen. Die Ergebnisse wurden zu einem Report zusammengefasst und dann kostenlos im Internet zur Verfügung gestellt[35].

Unter anderem in einem Blogbeitrag wurde auf die Veröffentlichung der Ergebnisse aufmerksam gemacht und den Lesern sollte auf möglichst einfache Weise ein Download des Reports geboten werden. Schließlich fiel die Entscheidung für einen Text-Link im oberen Drittel des Beitrages, der bei einem Aufruf der Seite im zentralen Blickfeld liegen sollte.

Da bereits zum Zeitpunkt der Veröffentlichung des Blogbeitrages an der Konzeption des hier vorliegenden Buches gearbeitet wurde, lag die Idee nahe, einmal statt des allseits bekannten Google Website Optimizers eines der kleineren Testsysteme zu nutzen. Diese machen mit teils sehr plakativer Sprache („WANT MORE SALES?“) auf sich und ihr Produkt aufmerksam und nun sollte ausprobiert werden, was denn dahintersteckt.

1. Der „Power Split Tester“

Der „Power Split Tester“ (http://www.powersplittester.com) für aktuell 47US$ ist eine kleine Bibliothek mit php-Dateien, die auf dem Web-Server eingerichtet werden müssen. Das geht schnell und unproblematisch. Das System kommt sogar ohne eine Datenbank aus, da die Testdaten in reinen Textdateien gespeichert werden.

Nach absolvierter Einrichtung wird das System über eine Web-Oberfläche gesteuert. Die Navigation ist etwas umständlich, aber am Ende doch einfach.

In der Weboberfläche werden die Anzahl der zu variierenden Elemente und die Anzahl der Variationen festgelegt. Schließlich werden auch die verschiedenen Textversionen eingegeben.

[35] „Web Analytics – Themen, Praxis und Umsätze 2008/09“, noch immer abrufbar unter http://www.idealobserver.de/news/?p=459

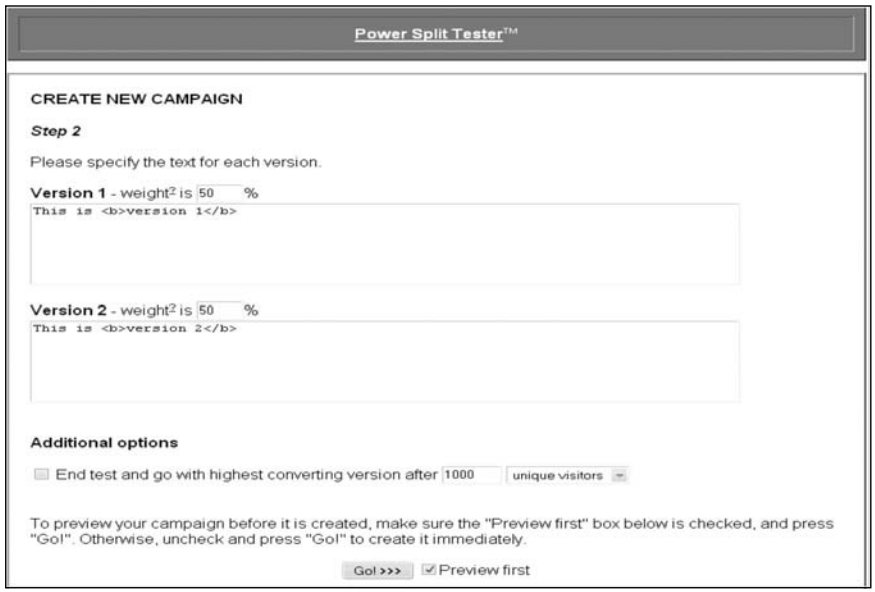


Abbildung 43: Einrichtung eines Tests mit dem Power Split Tester

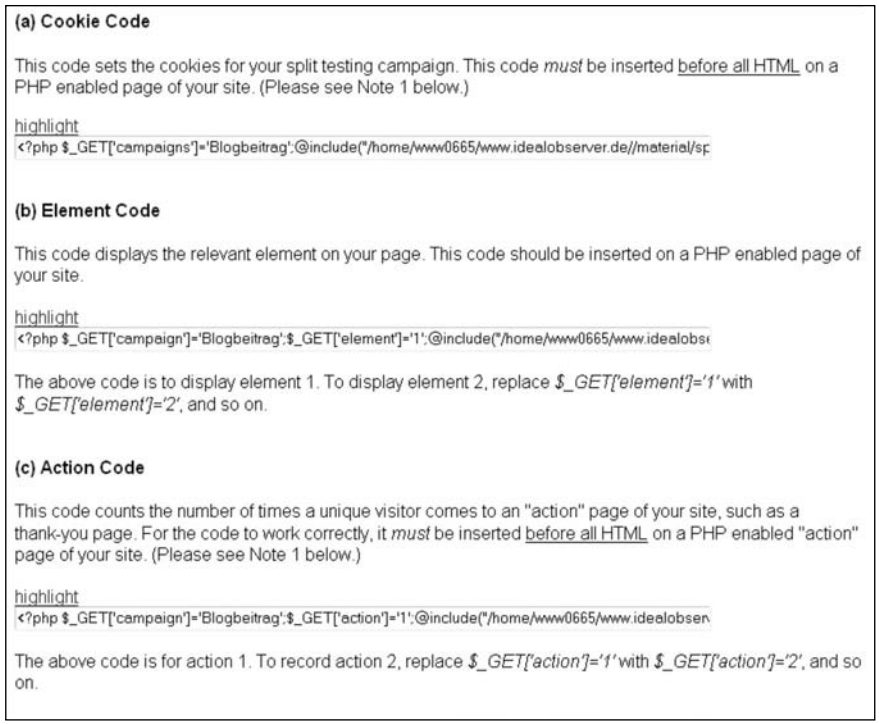


Abbildung 44: Einrichtung eines Tests mit dem Power Split Tester

Das System liefert dann die fertigen Code-Schnipsel, die in die Website eingebaut werden sollen. Ist das geschafft, wird die Darstellung dieses Containers durch den Split Tester gesteuert. Von hier aus werden den Besuchern die beiden Varianten zugespielt, der Abruf gemessen und ausgewertet. Der Cookie Code sorgt dafür, dass ein Besucher bei späteren, erneuten Seitenaufrufen die gleiche Version erhält wie beim ersten Aufruf. Der Action Code meldet an das Testsystem einen erfolgreichen Klick. Hier werden also die Konversionen gezählt.

Nach in diesem Fall 975 Seitenabrufen sieht das Ergebnis wie folgt aus: Version 1 („Download des vollständigen Reports (kostenlos)“) erreicht eine Konversionsrate von fast 13 Prozent, während Version 2 („hier klicken zum Download des vollständigen Reports“) nur auf gut 7 Prozent kommt. Damit steht der Gewinner des Tests eindeutig fest. Der kleine Unterschied in der Formulierung sorgt für einen fast 50-prozentigen Anstieg der Konversion.

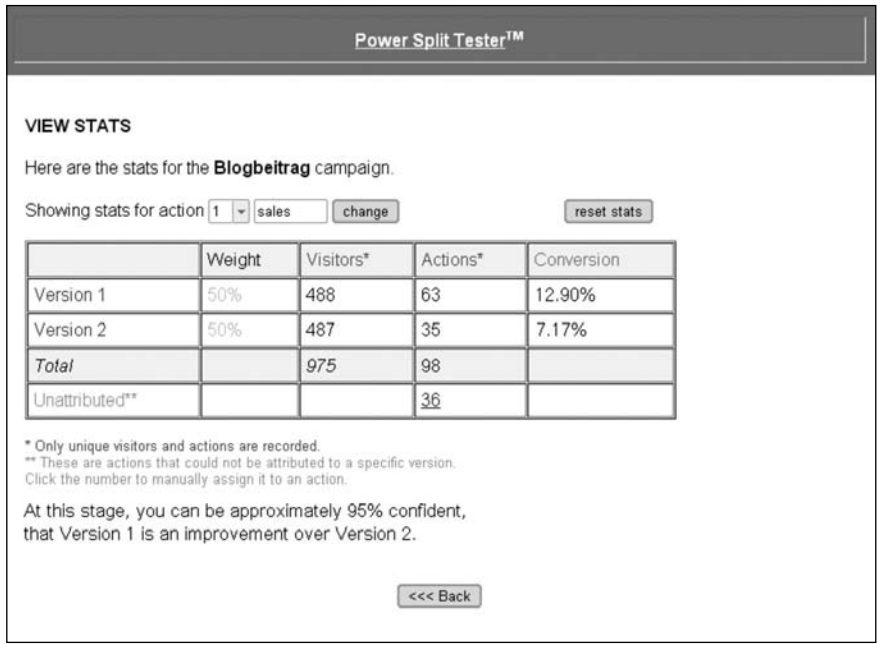


Abbildung 45: Die Testergebnisse nach 975 Seitenabrufen.

Nach in diesem Fall 975 Seitenabrufen sieht das Ergebnis wie folgt aus: Version 1 („Download des vollständigen Reports (kostenlos)“) erreicht eine Konversionsrate von fast 13 Prozent, während Version 2 („hier klicken zum Download des vollständigen Reports“) nur auf gut 7 Prozent kommt. Damit steht der Gewinner des Tests eindeutig fest. Der kleine Unterschied in der Formulierung sorgt für einen fast 50-prozentigen Anstieg der Konversion.

2. Fazit

Wo auch immer die Ursache für dieses stark variierende Verhalten seitens der Besucher liegt, muss im Unklaren bleiben. Eine vorher stattfindende Diskussion, welche Version warum besser sein würde, hatte im Gegensatz dazu keine klare Entscheidung gebracht. Durch den kleinen Test wurde auf diese Weise eine klare Entscheidungsgrundlage und eine deutliche Steigerung der Abrufrate erreicht. Das Testsystem kann Online-Verantwortlichen von kleinen Websites, die keine Scheu vor Quell-Code haben, durchaus empfohlen werden.

IX. Multivariate Testverfahren

von Frank Reese

1. Ansatzpunkt, Elemente und Variationen 168
2. Repräsentativität und Segmentierung 170
3. Auswahl des Testsystems 173
4. Testkombinationen und Teilnehmerzahl 173
5. Wahrscheinlichkeitsverteilungen und Konfidenzintervalle 174
6. Reduzierung der notwendigen Seitenabrufe 180
7. Fazit 183

Multivariate Testverfahren – die Bezeichnung klingt vielleicht kompliziert, dafür ist das Versprechen ist auch nicht klein: Statt mühevoll Schritt für Schritt einzelne Elemente auszutesten, können hier viele – theoretisch unbeschränkt viele – Variationen von Elementen einer Webpage gleichzeitig online gestellt und gemessen werden. Für die Verantwortlichen ist dies ein riesiger Vorteil. Schon bei der Gestaltung von Landing Pages, Startseiten oder Formularen muss nicht endgültig zwischen mehr oder weniger gleich überzeugend wirkenden Alternativen gewählt werden, sondern für jedes fragliche Element, zum Beispiel eine Überschrift, ein Button oder ein Formularfeld, können mehrere Alternativen in den Test einbezogen werden.

Aber im Unterschied zum einfachen A/B-Test kommen multivariate Verfahren nicht ohne ein wenig raffiniertere Mathematik aus. Da nicht einfach nur zwei (oder vielleicht noch drei) Versionen direkt miteinander verglichen werden, sondern sich leicht mehrere hundert Kombinationen ergeben und dazu mögliche Beeinflussungen der Elemente untereinander zu berücksichtigen sind, kommen bei multivariaten Tests einige statistische Verfahren zum Einsatz. Die Verfahren rechnen die Einflussstärke einzelner Varianten, zum Beispiel einer bestimmten Überschrift, aus den vielen Kombinationen heraus und zeigen, ob sich jenes Element besonders mit diesem Button oder jenem Visual verträgt.

Aber von vorne: In diesem Kapitel sollen zunächst die Merkmale und notwendigen Entscheidungen für einen multivariaten Test an Hand eines Beispiels dargestellt werden und dann einige Testplattformen und deren Unterschiede kurz vorgestellt werden.

1. Ansatzpunkt, Elemente und Variationen

Multivariate Tests brauchen zunächst eine sorgfältige Vorbereitung und mehrere alternative Versionen von Website-Elementen müssen produziert werden. Dazu kommen die Durchführung des Tests (meist mit Hilfe eines externen Dienstleisters) und die Umsetzung der Ergebnisse. Das Ganze ist also nicht mal eben vor der Frühstückspause zu erledigen, entsprechend werden vor allem sehr wichtige Bereiche einer Website mit diesen Verfahren optimiert. In Kapitel IV wurden bereits die typischen Ansatzpunkte für eine Optimierung vorgestellt, die auch für dieses Verfahren vor allem in Betracht kommen.

Als Beispiel soll hier ein Formular auf den Seiten des Fußballverbandes FIFA dienen (Abbildung 46). Die Testbeschreibung ist dabei ganz fiktiv, zu den Betreibern der Site besteht keinerlei Verbindung.

Der Betreiber lädt seine Website-Besucher zu einem Club-Beitritt ein. Auf der vorangehenden Seite werden einige Vorteile genannt und mit dem Ausfüllen und Absenden dieses Formulars (das hier verkürzt dargestellt wird) ist die Mitgliedschaft vollzogen.

Das Ziel des Formulars ist eindeutig: Der „Call-to-Action" liegt allein bei der Anmeldung zur Mitgliedschaft. Damit kann die Konversionsrate: „Wie hoch liegt der Anteil der Aufrufe dieser Page, der zu einer gültigen Absendung des Formulars führt?" als Kennzahl für die Performance der Testvarianten genutzt werden. Diese Rate soll in unserem Beispiel der Einfachheit halber bei 10 Prozent liegen. In 10 von 100 Fällen wird also das Formular nach dem Aufruf der Seite vollständig abgesendet. Für den Test wird eine Erhöhung der Konversionsrate auf 15 Prozent als Ziel definiert, also eine Steigerung um 50 Prozent. Die Festlegung der Differenz hat später Einfluss auf die notwendige Anzahl von Seitenaufrufen im Test bzw. auf die Dauer des Tests.

Durch die eindeutige Festlegung der entscheidenden Kennzahl ist die Variation bzw. Kombinationen mit der höchsten Konversionsrate der „Gewinner" des Tests, wobei der Test insgesamt als abgeschlossen gilt, wenn eine der Kombination zuverlässig eine Konversionsrate von 15 Prozent erreicht.

Im Formular sind vier Bereiche markiert, deren Inhalt mit Hilfe eines Tests optimiert werden sollen:

- Die Überschrift (rot)
- Die Darstellung, Positionierung und Ausgestaltung von zwei Formularfelder (grün)
- Die Darstellung, Gestaltung und Reihenfolge der Kontrolleingabe und der „Nutzungsbedingungen" (gelb)
- Die Gestaltung und Positionierung der Buttons am Ende der Seite (blau)

Abbildung 46: Anmeldeformular mit Testbereichen

Für den Test werden jetzt für jedes der einzelnen Elemente (im Folgenden auch „Faktoren" genannt) Alternativen formuliert. Für die Überschrift könnte dies zum Beispiel „Werden Sie Mitglied im Club" oder „Werden Sie FIFA-Mitglied" sein. Die Felder für Geburtsdatum und Geschlecht könnten a) ganz entfernt werden, b) „Alter" und „Geschlecht" lauten oder „Alter" und „Anrede" lauten. Ebenso können für das gelbe und das blaue Element Varianten festgelegt werden.

2. Repräsentativität und Segmentierung

Neben den inhaltlichen Variationen ist es wichtig zu überlegen, ob sämtliche Besucher potenziell am Test teilnehmen sollen oder nur Besucher mit bestimmten Eigenschaften. Dabei geht es zunächst nicht um die absolute Anzahl der Test-Teilnehmer (diese wird an anderer Stelle festgelegt), sondern um die Repräsentativität der Testgruppe einerseits und die Mög-

lichkeit, sich auf bestimmte Zielgruppen zu konzentrieren andererseits. Multivariate Tests basieren auf statistischen Methoden, die Aussagen über Wahrscheinlichkeiten, nicht über Ursache-Wirkung-Zusammenhänge machen. Unabhängig von den konkreten Umständen liegt die Wahrscheinlichkeit, dass ein Aufruf zu einer Formularabsendung führt, zum gegenwärtigen Zeitpunkt bei 1:10. In dieses Verhältnis fließt die Gesamtheit der bisherigen Besuche ein, gewöhnlich über einen langen Zeitraum gemessen. Die Ergebnisse des Tests basieren aber nicht auf der verfügbaren Gesamtzahl der Vorgänge, sondern stellen eine angemessen große Menge von Seitenabrufen zum Test, um von diesen auf das wahrscheinliche Gesamtergebnis schließen zu können. Damit rückt die Frage der Repräsentativität in den Blickpunkt. Nur wenn die Eigenschaften der Seitenabrufe im Test genauso strukturiert sind, wie die Gesamtheit aller Seitenabrufe, kann von der kleineren Menge (Sample) zuverlässig auf die größere Menge (Gesamtheit) geschlossen werden. Damit ergeben sich zwei Fragen:

Repräsentativität

Unterschiede können auftauchen, wenn beispielsweise das Verhalten der Besucher durch saisonale Effekte das Verhalten der Besucher im Testzeitraum anders als im Jahresdurchschnitt ist. So kann im vorliegenden Fall davon ausgegangen werden, dass mit dem Näherrücken einer Fußballweltmeisterschaft der Traffic auf den Seiten und sicher auch die Bereitschaft zur Registrierung zunimmt. Ebenso kann es Unterschiede im Wochenverlauf (unterschiedliches Zugriffsverhalten an Wochentagen und Wochenenden) und im Tagesverlauf (tagsüber, abends, nachts) geben. Das Beispiel ist zudem eine sprachlich regionalisierte Version eines internationalen Online-Angebotes. So könnte es starke Unterschiede zwischen verschiedenen Ländern aus dem deutschsprachigen Raum einerseits geben und dazu zwischen deutschsprachigen Besuchern aus anderen, nicht-deutschsprachigen Ländern. Für den Test ist also darauf zu achten, dass der Testzeitraum diese Schwankungen möglichst– soweit praktikabel – berücksichtigt. Der Zeitraum sollte zum Beispiel nicht nur nachts oder nur an Wochentagen laufen. Auch ein Test im direkten Vorfeld einer Fußball-WM kann keine Voraussagen zum Verhalten nach einer WM liefern.

Neben diesen Überlegungen, die bei der Anlage des Tests berücksichtigt werden sollten, beinhaltet der Test einige Bedingungen, die definitiv nicht von der Gesamtheit der Besucher erfüllt werden: Fast alle Testsysteme nutzen Javascript, eine im Internet sehr verbreitete Programmiersprache, die zum Beispiel für dynamische Elemente (Ajax) oder auch zur Traffic-Analyse genutzt wird. Allerdings gibt es Nutzer, die diese Funktion in ihrem Browser ausschalten. Sie können die Website normal abrufen, aber sie werden nicht am Test teilnehmen können. Das gleiche gilt für Cookies. Auch hier gibt es einen Anteil von

Besuchern, die deren Verarbeitung blockieren, so dass auch sie (teilweise) vom Test ausgeschlossen sind. Der Anteil dieser Nutzer ist allerdings gering. Er wird auf technik-affinen Seiten höher sein als im Durchschnitt, insgesamt für Javascript-Blockaden aber nicht über 1 Prozent liegen, bei Cookie-Blockaden nicht über 5 bis 10 Prozent. Diese Abweichungen können im Sinne der Gültigkeit der Testergebnisse gewöhnlich akzeptiert werden.

Segmentierung

Segmentierung, also die Unterscheidung von Besuchergruppen nach ihrem Verhalten, ist eine wichtige Methode für effektive Traffic-Analysen. Mit der Differenzierung zum Beispiel nach Herkunft (Kommt der Besucher über eine Kampagne, eine Partnerseite oder über eine Suchmaschine? Wenn Letzteres: mit welchem Stichwort?), nach geographischer Herkunft (Analyse der Sprachversion des Browsers oder des nationalen oder regionalen Standorts der IP-Adresse), nach der Art des Zugriffsgerätes (Mobilgerät oder PC) oder auch dem bisherigen Verhalten des Besuchers (Erst- oder Wiederholungsbesucher), können sich deutlich unterschiedliche Verhaltensweisen der Nutzer ergeben – und damit unterschiedliche Wahrscheinlichkeiten, dass sie das Formular absenden.

Wenn solche Segmente bzw. Zielgruppen vorher bekannt sind und unterschiedliche Potenziale im Sinne des Website-Zieles erkannt werden (nach oben oder unten abweichende Konversionsraten), so ist es effektiver, bestimmte Besuchergruppen vom Test auszuschließen oder exklusiv zuzulassen.

Zurück zum konkreten Ablauf des multivariaten Tests: Nach der **Festlegung der zu testenden Seite (Anmeldeformular)**, der Kennzahl **(Konversionsrate)**, des **Zielwertes** (15 Prozent), der **Testvarianten** (Überschrift, Buttons etc.), legen wir jetzt als letztes konzeptionelles Moment den **Testeilnehmerkreis** fest: Zunächst sollen nur Besucher mit einer geographischen Herkunft aus einem deutschsprachigen Land oder einer deutschsprachigen Browserversion zugelassen werden. Für alle nicht-deutschsprachigen Besucher kann dieses Formular kaum sinnvoll optimiert werden. Ein irrtümlicher Ausschluss ist natürlich möglich, fällt aber nicht ins Gewicht.

Außerdem werden alle Besucher ausgeschlossen, für die das Formular die erste aufgerufene Seite (Einstiegsseite) im Online-Angebot der FIFA ist. Sie haben nicht vorher einen Link oder einen Button gewählt, der sie zu einer Mitgliedschaft einlädt, sondern sind auf irgendeine Weise eher zufällig auf der Seite gelandet. Eine Konversion dieses Segments sollte eher unwahrscheinlich sein, eine Optimierung für diese Gruppe jedenfalls wenig lohnenswert.

3. Auswahl des Testsystems

Für die weitere Konkretisierung des Testdesigns und die Durchführung des Tests wird jetzt ein zusätzliches Tool gebraucht, das die Varianten entsprechend des Designs einspielt sowie die Performance der Faktoren und der Kombinationen misst und auswertet. Diese Werkzeuge sind nur selten im Unternehmen vorhanden, also wird in der Regel eine externe Lösung zur Durchführung gebraucht. Da die Konfiguration ein gewisses Maß an Fachwissen voraussetzt, ist zudem die Unterstützung durch externe Expertensinnvoll. Es gibt eine Reihe amerikanischer und britischer Unternehmen, die Tests anbieten. In Deutschland gehören aktuell Omnitures „Test & Target", der kostenlose „Web Site Optimizer" von Google und die Testplattform von Divolution zu den weitverbreitetsten. Es gibt große Unterschiede zwischen diesen Systemen (für eine Übersicht s. Kapitel XIV), gemeinsam ist ihnen jedoch das technische Prinzip: Wie bei Website-extern gesteuerten Werbemitteln werden die einzelnen Test-Varianten über einen Aufruf aus der eigenen Webpage initiert. Für jeden Faktor wird also eine Art Container definiert, in dem dann die Varianten vom externen Test-Tool abwechselnd eingespielt werden.

Auf der eigenen Website werden nur die Container zur Darstellung mit Javascript definiert, die einzelnen Varianten (Überschriften Buttons usw.) bzw. deren Code werden im System des Test-Anbieters abgelegt. Bei jedem Aufruf der Webpage wird neben dem Code für die eigentlichen Inhalte auch der Code für den Abruf des Container-Inhaltes beim Test-Anbieter verarbeitet. Für den Besucher bleibt dieser Vorgang unsichtbar, ihm wird eine vollständige, in dieser Hinsicht in keiner Weise auffällige Website präsentiert. Um langfristige Wirkungen nachverfolgen zu können (Konversion bei späteren Besuchen) und um eine Konsistenz der Seite selbst bei späteren Besuchen oder Seiten-Aktualisierungen zu gewährleisten, wird vom Test-Anbieter ein (oder mehrere) Cookie(s) im Browser des Besuchers abgespeichert. Über diese Technik können die Besucher wiedererkannt und eine Page mit den gleichen Elementen und Kombinationen wie beim ersten Besuch angezeigt werden.

4. Testkombinationen und Teilnehmerzahl

Neben der konkreten Testplattform, der Zusammenarbeit mit den Beratern und den anfallenden Kosten wird mit der Wahl der Testplattform auch teilweise (Sind datenreduzierende Methoden möglich?, siehe unten) entschieden, wie hoch die Zahl der Testteilnehmer sein muss, um zu einem zuverlässigen Ergebnis zu gelangen.

Diese Zahl definiert sich über fünf Momente:

- Wie hoch ist die gegenwärtige Konversionsrate?
- Wie viele Kombinationen gibt es?
- Wie groß soll die Steigerung sein?
- Wie zuverlässig sollen die Ergebnisse sein?
- Sollen sämtliche Kombinationen und ihre Interaktionen getestet werden?

Die gegenwärtige Konversionsrate ist gewöhnlich gegeben (in diesem Beispiel 10 Prozent). Die Anzahl der Kombinationen ergibt sich aus der Anzahl der zu testenden Elemente (Überschrift, Felder, Nutzungshinweise, Buttons) (auch „Level" genannt) und der Anzahl der Varianten (Faktoren) pro Element. Wird die Zahl der Varianten pro Element miteinander multipliziert, ergibt sich die Anzahl der möglichen Kombinationen. In unserem Beispiel gibt es drei Überschriftenvarianten, drei Feldvarianten, zwei Nutzungshinweisvarianten und drei Buttonvarianten. Das ergibt:

Anzahl der Kombinationen = $3 \times 3 \times 2 \times 3 = 54$ Kombinationen

Dieser Umfang ist bei multivariaten Analysen nicht ungewöhnlich, oft wird sogar eine deutlich größere Anzahl von Kombinationen getestet.

Den gewünschten Anstieg haben wir auf 50 Prozent festgelegt. Zu bedenken ist bei dieser Festlegung: Je höher sie veranschlagt wird, desto geringer ist die Anzahl der notwendigen Testdurchläufe unter der Bedingung eines zuverlässigen Ergebnisses (siehe unten zu Konfidenzintervallen).

5. Wahrscheinlichkeitsverteilungen und Konfidenzintervalle

An diesem Punkt ist es notwendig, ein wenig auszuholen und die theoretischen Hintergründe kurz zu beschreiben. Für die tägliche Praxis ist dieses Hintergrundwissen im Detail kaum wichtig, aber für ein Verständnis darüber, wie die Vorbedingungen das Ergebnis beeinflussen, ist es durchaus hilfreich.

Die Ergebnisse von multivariaten Tests basieren auf Aussagen über das wahrscheinliche Eintreffen von Ereignissen. Bei einem Roulette-Spiel beispielsweise liegt die Wahrscheinlichkeit, dass eine bestimmte Zahl eintrifft, bei 1:37. Von den 36 Zahlen plus der Zero kommen in 37 Durchgängen alle einmal dran, könnte man denken, aber natürlich stimmt

das nicht. Vielmehr würde eine Verteilung nach 37 Würfen große Ungleichgewichte zeigen. Ebenso bei unserer Konversionsrate: Theoretisch würde einer von zehn Seitenaufrufen zu einer Konversion führen bzw. bei 100 Seitenaufrufen hätten wir zehn Konversionen. In Wirklichkeit ergibt sich in beiden Fällen diese Verteilung erst nach sehr vielen Würfen bzw. Seitenaufrufen. In der Theorie nähert sich die beobachtete Häufigkeit der Ereignisse mit der Anzahl der Durchläufe der theoretischen Wahrscheinlichkeit an. Je öfter also die Kugel rollt, desto stärker wird sich die Verteilung der Zahlen der theoretisch bestimmten annähern – in 37.000 Durchgängen sollte jede Zahl etwa 1.000 Mal vorkommen. Eine so hohe Zahl liegt gewöhnlich auch der Kalkulation einer Konversionsrate, die über einen längeren Zeitraum ermittelt wurde, zugrunde.

Wenn wir jetzt das Formular verändern, wollen wir eine höhere Wahrscheinlichkeit für das Eintreffen der Konversion erreichen. Von 1:10 soll die Wahrscheinlichkeit auf 1,5:10 steigen. Es wäre also so, als würden wir anfangen, den Roulette-Kessel zu manipulieren, um öfter eine bestimmte Zahl gewinnen zu lassen. Da wir allerdings nicht so lange testen können, bis das neue Formular (und jede seiner neuen Kombinationen) ebenso hohe Fallzahlen hat, wie jene für die Berechnung der bisherigen Konversionsrate, stellt sich die die Frage: Ab wann können wir einigermaßen sicher sein, dass eine bestimmte neue Konstellation wirklich besser ist als das Original (das bestehende Formular). Die Zuverlässigkeit der Aussagen von multivariaten Tests basieren auf sogenannten Konfidenzintervallen oder auch Vertrauensintervallen. Diese beschreiben die Wahrscheinlichkeit, dass ein Ergebnis tatsächlich die (stets unbekannte) Wirklichkeit abbildet.

Am höchsten Punkt einer Normalverteilung (die sehr bekannte Glockenkurve) findet sich die von uns berechnete Konversionsrate. Sie ist ein Durchschnittswert, um den mögliche andere Durchschnittswerte innerhalb der Kurvenfläche liegen. Da wir die tatsächliche Verteilung nie kennen können (das würde unendlich viele Durchläufe beanspruchen), zeigt die Kurve die Wahrscheinlichkeit, dass ein bestimmter Mittelwert (unsere Konversionsrate) tatsächlich die Realität trifft. Je weiter ein Punkt von der Mitte entfernt liegt, desto unwahrscheinlicher ist sein tatsächliches Eintreffen. Diese Verteilungen sind gewöhnlich in Gültigkeitsbereiche beschrieben.

Nach der Berechnung für Abbildung 48 liegt die Wahrscheinlichkeit, dass die Konversionsrate zwischen 35 und 64 liegt, bei 95 Prozent. Auf jeder Seite der Kurve bleiben 2,5 Prozent abgeschnitten. Hier liegen die Mittelwerte geringer als 35 oder höher als 64 und werden also nur mit jeweils 2,5 prozentiger Wahrscheinlichkeit eintreffen.

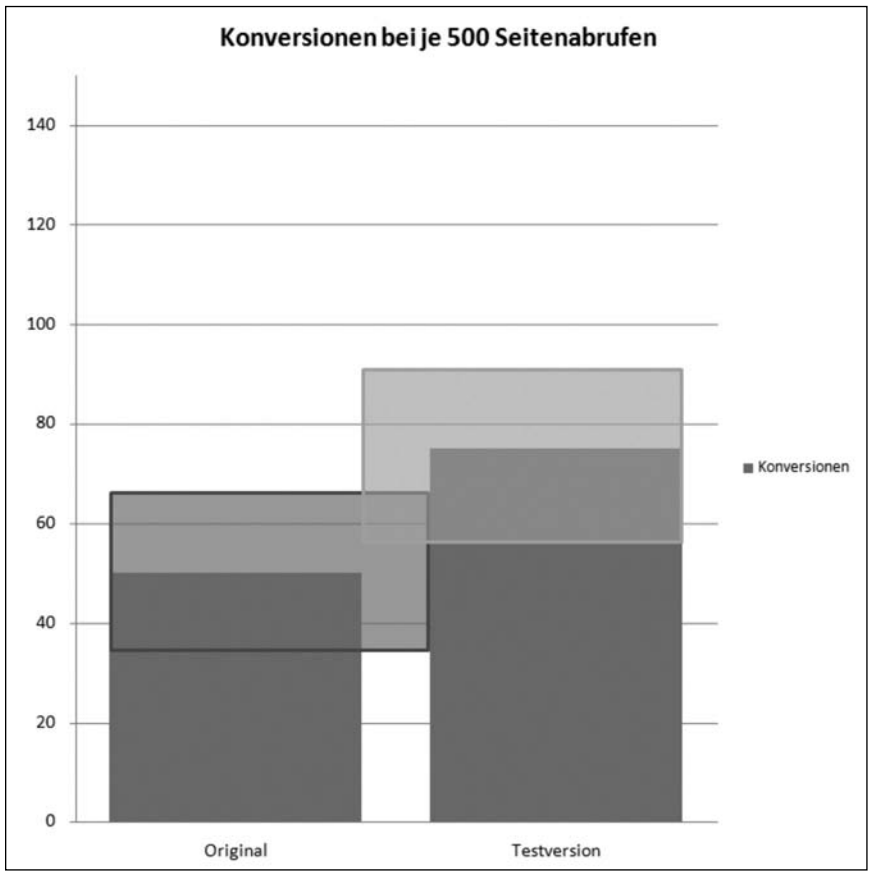


Abbildung 47: Die blauen Balken zeigen die Anzahl der Konversionen der beiden Versionen (50 bzw. 75) entsprechend der Konversionsrate (10 Prozent) der Originalversion und der anvisierten Konversionsrate(15 Prozent). Die transparenten Rechtecke um diese Mittelwerte zeigen die Bereiche, in denen die Anzahl der Konversionen in der Zukunft auch liegen könnten (laut Konfidenzintervall). Diese Bereiche überschneiden sich.

onfidenzintervalle sind Konventionen. Übliche Konfidenzintervalle sind auf 90 Prozent, 95 Prozent oder 99 Prozent angelegt. Ein Intervall von 50 Prozent würde gerademal die Qualität einer Zufallsentscheidung besitzen: Die Wahrscheinlichkeit, dass ein Mittelwert (eine Konversionsrate) in diesem Bereich liegt, ist ebenso groß, wie dass er außerhalb liegt. Für wirtschaftsnahe Untersuchungen, Marktforschungen und Umfragen werden eher 90 Prozent oder 95 Prozent akzeptiert, für wissenschaftliche oder medizinische Untersuchungen eher 99 Prozent gefordert. Ein Konfidenzintervall von 95 Prozent besagt also, dass

die Verteilung der beobachteten Werte (im Test) mit 95-prozentiger Wahrscheinlichkeit die tatsächliche Verteilung abbildet. Die Nutzung eines 95-prozentiger Konfidenzintervalls ist auch bei multivariaten Analysen verbreitet. Ein Fehler von 5 Prozent wird also akzeptiert.

Ein Konfidenzintervall von 95 Prozent wird allgemein akzeptiert und gibt Sicherheit, dass die Konversionsrate nur mit geringer Wahrscheinlichkeit nicht im vorausgesagten Bereich liegt.

Dieser Fehler bedeutet, dass die tatsächliche Verteilung der Werte und damit der daraus berechnet Mittelwert mit 2,5-prozentiger Wahrscheinlichkeit höher oder niedriger liegt. Wie groß diese potenzielle Abweichung nach oben oder unten nominell ist, also im Falle unserer anvisierten 15-prozentigen Konversionsrate, ergibt sich aus der Anzahl der Test-Durchläufe: Je höher die Anzahl der Testdurchläufe, desto zahlenmäßiger geringer wird der Bereich, in dem der tatsächliche Mittelwert mit entsprechender Wahrscheinlichkeit vermutet werden kann. Das Testergebnis kann als zuverlässig gelten (mit dem oben genannten Konfidenzintervall), wenn die Abweichungen der beiden Werten sich nicht mehr überschneiden:

Wenn sich die Bereiche, in der der Mittelwert der Originalversion einerseits und der Mittelwert der alternativen Version andererseits wahrscheinlich liegen, nicht mehr überschneiden, hat man mit großer Zuverlässigkeit einen Gewinner: Der Test war erfolgreich.

Der unterste mögliche Wert der neuen Kombination (15 Prozent abzüglich Abweichung) liegt über dem obersten möglichen Wert der Originalversion (10 Prozent plus Abweichung) [36].

Berechnet man das Beispiel mit je 500 Seitenabrufen und einem Konfidenzintervall von 95 Prozent, so zeigen sich folgende Verhältnisse (Abbildung 47).

[36] Eigentlich liegen hier zwei verschiedene Vorgänge vor: Der Bereich, in dem der tatsächliche Mittelwert liegt, wird mit einem 95-prozentigen Konfidenzintervall festgelegt. Ob der Unterschied zwischen zwei Testversionen und ihren Intervallen zuverlässig ist, kann wiederum mit einem Signifikanz-Test berechnet werden und als Ergebnis zum Beispiel ein 90 Prozent, 95 Prozent oder 99-prozentiges Signifikanzniveau. In der Praxis werden diese beiden Vorgänge gewöhnlich zusammengefasst und ein Zuverlässigkeitsniveau für beide Vorgänge gewählt.

Laut Abbildung 48 ergeben sich im Original bei einer Konversionsrate von 10 Prozent und 500 Seitenabrufen 50 Konversionen. Bei der Testversion mit anvisierten 15 Prozent ergeben sich bei der gleichen Anzahl von Seitenabrufen 75 Konversionen. Auf den ersten Blick ist also die Testversion der Originalversion deutlich überlegen.

Die transparenten Rechtecke zeigen dagegen die mögliche Spanne der tatsächlichen Verteilung. Mit 95-prozentiger Sicherheit kann danach gesagt werden, dass die Konversionen in der Originalversion bei den nächsten (zukünftigen) 500 Seitenabrufen zwischen 35 und 64 liegen werden, bei der Testversion dagegen zwischen 57 und 92. Es ergibt sich eine Überschneidung der möglichen Anzahl der Konversion. Es lässt sich also *nicht mit 95-prozentiger Sicherheit sagen*, ob die Testversion mehr Konversionen generiert.

Wird die Anzahl der Seitenabrufe erhöht, ergibt sich folgendes Bild (Abbildung 48):

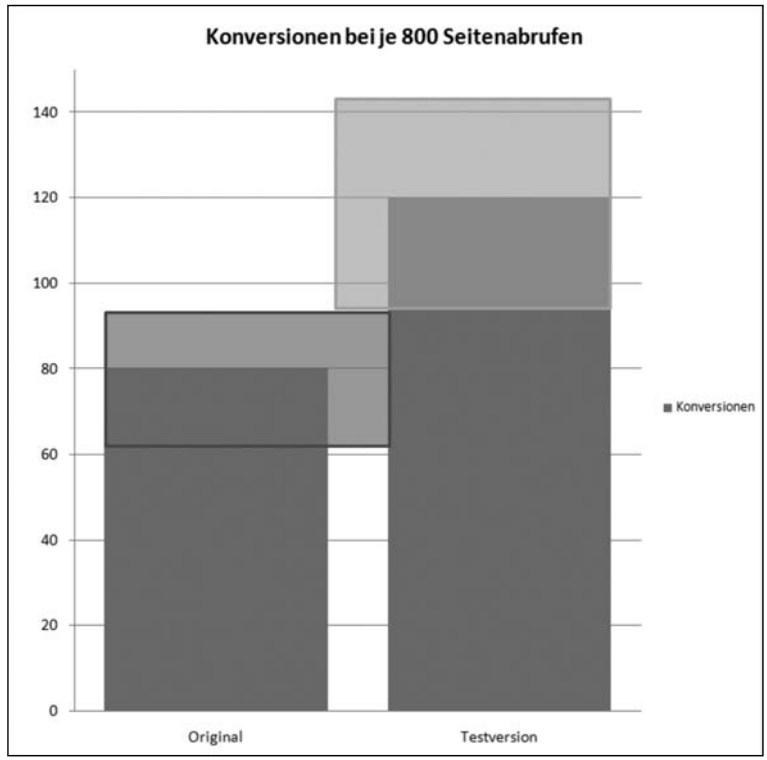


Abbildung 48:
Die Balken zeigen die Anzahl der Konversionen für die beiden Versionen (80 bzw. 95). Die transparenten Rechtecke um diese Mittelwerte zeigen die Bereiche, in denen diese Werte in Zukunft auch liegen könnten (laut Konfidenzintervall). Diese Bereiche überschneiden sich nicht mehr. Die Testversion ist mit 95-prozentiger Sicherheit besser.

Durch die Erhöhung der Seitenabrufe steigern sich die Konversionen auf 80 bzw. 120. Dazu verringern sich die Abweichungen (die Messung wird zuverlässiger), die Werte in den Intervallen überschneiden sich nicht mehr und jetzt kann mit 95 Prozent Sicherheit gesagt werden, dass die Testversion eine höhere Konversionsrate erzielt als die Originalversion. Die hier verwendete Anzahl von 800 Seitenaufrufen pro Version ist damit die untere Grenze für einen Test, der in diesem Fall zuverlässig eine Steigerung der Konversion rate ergeben soll. Dabei sollte beachtet werden, dass mit dem Verfahren nicht die genaue Konversionsrate von 15 Prozent mit 95-prozentiger Wahrscheinlichkeit vorhergesagt wird, sondern Konversionen im Bereich zwischen 140,91 und 98,09 pro 800 Seitenabrufe, also eine Konversionsrate zwischen 17,75 Prozent und 12,25 Prozent.

Für eine detaillierte Beschreibung dieser Berechnungen möchte ich den Leser auf die einschlägige Literatur verweisen[37]. Die Testplattformen berechnen diese Werte gewöhnlich automatisch und nutzen teilweise abweichende Modelle.

Die Berechnung von Konfidenzintervallen, Mindestanzahl der Seitenabrufe und ob das Ergebnis zuverlässig ist oder nicht, führen die Testsysteme automatisch aus.

Wichtig ist hier allerdings folgender Zusammenhang: Da die Abweichungen mit zunehmender Zahl der Durchläufe geringer werden, ist für die zuverlässige Erreichung einer kleinen Steigerung (zum Beispiel von 10 Prozent auf 11 Prozent) eine größere Anzahl von Durchläufen notwendig, als für die zuverlässige Erreichung einer großen Steigerung (zum Beispiel von 10 Prozent auf 15 Prozent). Die Differenz zwischen dem Ausgangswert (10 Prozent) und dem Zielwert (15 Prozent) ist so groß, dass schon bei vergleichsweise kleinen Fallzahlen keine Überschneidungen mehr auftreten. Würden wir für die Testversion eine Konversionsrate von 20 Prozent als Ziel vorgeben (100-prozentige Steigerung), genügten bereits 250 Seitenabrufe für ein zuverlässiges Ergebnis.

Je grösser die zu erwartende Steigerung der Konversionsrate, desto geringer ist die Anzahl der notwendigen Seitenabrufe um die Steigerung zuverlässig zu belegen.

[37] Zum Beispiel Fahrmeir, Künstler & Pigeot: Statistik (Springer 2007); auch Tim Ash gibt einen guten Einblick in „Landing Page Optimization“ (Wiley 2008), für März 2009 ist eine deutsche Übersetzung des Buches angekündigt

Wir halten also fest: In aller Regel sollte für einen multivariaten Test ein Konfidenzintervall von 95 Prozent gewählt werden[38], und eine höhere zu erwartende Steigerung erlaubt eine geringere Anzahl von Testdurchläufen. Da die Anzahl der täglichen Seitenabrufe für Online-Abrufe nicht leicht zu beeinflussen ist (jedenfalls nicht ohne zusätzliche Verzerrungen zum Beispiel zusätzliche durch Kampagnen einzuführen), bedeutet ein höheres Ziel eine kürzere Laufzeit des Tests. Für eine einfache, beispielhafte Kalkulation der Zusammenhänge für den Google Website Optimizer kann zum Beispiel ein Online-Kalkulator unter folgenden Web-Adressen genutzt werden:
https://www.google.com/analytics/siteopt/siteopt/help/calculator.html

6. Reduzierung der notwendigen Seitenabrufe

Jeweils 800 Seitenabrufe sind in unserem Beispiel also notwendig, um eine Testversion mit der Originalversion zu vergleichen. In unserem Beispiel gibt es allerdings nicht nur eine Testversion, sondern 54 (3 × 3 × 2 × 3 Kombinationen).

Für jede Testversion werden 800 Seitenabrufe benötigt, um ein belastbares Ergebnis zu bekommen. Das sind zusammen 43.200 Seitenaufrufe. Dazu wird gewöhnlich nicht der gesamte Traffic auf das Testfeld gelenkt, sondern ein wesentlicher Anteil auf der Originalversion gelassen. Gehen wir jetzt für unser Beispiel davon aus, dass das Formular täglich 15.000-mal aufgerufen wird, so ergibt sich folgende Rechnung:

15.000	tägliche Seitenabrufe des Formulars
– 3.000	(20 Prozent des Traffics bleibt ungetestet wegen der Segmentierungen siehe oben)
– 6.000	(50 Prozent des verbleibenden Traffics bleibt auf der Originalseite)
= 6.000	Seitenabrufe pro Tag für die Testversionen

[38] Für Nutzer des Google Website Optimizers erübrigt sich diese Überlegung – das Konfidenzintervall ist auf 80 Prozent festgelegt. Warum Google hier von weltweit etablierten Standards aus Wissenschaft und Wirtschaft abweicht, bleibt ihr Geheimnis. Der Spalteneintrag „Chance to beat the original" zeigt allerdings das Ergebnis einer zusätzlichen Signifikanzberechnung. Ein Wert hier von zum Beispiel 90 Prozent ergibt eine 90-prozentige Sicherheit, dass der im Konfidenzintervall liegende Mittelwert der neuen Version über dem der Originalversion liegt. Dazu etwas kryptisch bei Google: „The 80 % confidence level is what determines the width of the range. It is what allows us to say, with 95 % confidence, that a range that exceeds and does not overlap another range will outperform the other range." (https://www.google.com/support/websiteoptimizer/bin/answer.py?answer=61146)

Damit sollte der Test gut sieben Tage lang laufen, um aussagekräftig zu sein. Das Testsystem steuert in diesem Zeitraum die Zuweisung der Teilnehmer auf die verschiedenen Kombinationen. Dies hat zudem den Vorteil, dass sowohl Wochentage wie auch ein Wochenende im Testzeitraum liegen.

In dem bisherigen Modell wird davon ausgegangen, dass alle 54 Versionen getestet werden müssen, da potenziell jede Kombination der Gewinner des Tests sein könnte. Dieser Ansatz wird **voll-faktorielles Design** genannt. Unter der Voraussetzung, dass jede einzelne Variante eine wesentliche Interaktion mit einer anderen Variante haben könnte, müssen alle Kombinationen durchgespielt werden, um diese Effekte berechnen zu können. Die Idee ist also: Überschrift1 könnte mit Feld2 und Button3 das beste Ergebnis zeigen, mit Button1 oder Button4 aber wiederum deutlich schlechter, als Überschrift2 mit Feld2 und Button4, usw.

Die Komplexität dieser Interaktionsebenen kann sehr weit getrieben werden. Die mögliche Auswertung der Performance liegt zwischen zwei Polen:

Nur einzelne Varianten werden in ihrem Effekt gemessen

So würde etwa eine Variante der Überschriften, eine der Felder, der Nutzungshinweise und der Buttons als Test-Gewinner ausgewiesen. Dabei wird außer Acht gelassen, ob der Gewinner des einen Elements auch tatsächlich in Kombination mit dem Gewinner eines anderen Elementes effektiv ist.

Nur Kombinationen werden in ihrem Effekt gemessen

In diesem Fall werden alle 54 Kombinationen direkt miteinander verglichen. Das hat den Vorteil, dass kontraproduktive Interaktionen zwischen Elementen ausgeschaltet werden, allerdings bleibt unklar, welche einzelnen Elemente und Elementkombinationen auf niedrigerer Ebene (hier Kombinationen mit zwei oder drei Elementen) für den Effekt verantwortlich sind.

Vorteilhafterweise gibt es in den meisten Testsystemen die Möglichkeit, im Vorhinein bestimmte, vermutlich interagierende Elemente festzulegen, andere dagegen als unabhängig davon zu betrachten. So könnte etwa von einer Interaktion zwischen einem Text und einem daneben stehenden Bild ausgegangen werden, die Gestaltung oder Positionierung des Buttons wäre dagegen unabhängig davon.

Solche Systeme werden **teil-faktorielle Designs** genannt. Eines der bekanntesten Modelle zur Reduzierung der zum Test notwendigen Kombinationen ist nach dem japanischen Mathematiker Taguchi benannt. Dieser hatte für Produkttests im industriellen Umfeld ein Verfahren entwickelt, um bestimmte Kombinationen in einem Test auszuschalten und trotzdem zu einem aussagekräftigen Ergebnis zu gelangen. Mit der Taguchi-Methode (und anderen teil-faktoriellen Verfahren) wird es also möglich, große Tests mit vielen Elementen gleichzeitig durchzuführen, ohne die dafür notwendigen Testdurchläufe ins Unermessliche steigen zu lassen. Für diese Verfahren wird zum Beispiel davon ausgegangen, dass Interaktionen auf niedriger Ebene (nur zwischen zwei Elementen) wichtiger sind, als Interaktionen auf höherer Ebene (zwischen drei, vier oder mehr Elementen) und dass Elemente, die für sich allein keinen Effekt zeigen, auch in ihrer Interaktion mit anderen Elementen wahrscheinlich keine bedeutenden Effekte verursachen werden.

Das Interesse an einer Reduktion der Testdurchläufe, zum Beispiel durch die Taguchi-Methode ist von ihrem Ursprung her den Kosten für einen Test mit tatsächlich produzierten Gütern geschuldet. Diese Aufwände treten für Online-Angebote nicht notwendig auf. Auch in unserem Beispiel ist das Prinzip der Sparsamkeit in der Datenerhebung nicht ausschlaggebend, da die Anzahl der Testkombinationen, die Differenz zur Zielkennzahl und vor allem das tägliche Trafficaufkommen groß genug sind, um den Test in angemessener Zeit zum Ende zu bringen. Aber auch in der Online-Welt verursacht der Traffic direkt Aufwand. So berechnen die Anbieter der Testplattformen die Kosten der Testdurchführungen in Abhängigkeit von der Anzahl der Seitenabrufe (Events). Je mehr Kombinationen also getestet werden müssen, desto teurer die Durchführung.

Unabhängig von automatischen Verfahren wie der Taguchi-Methode gibt es in den Testsystemen teilweise Möglichkeiten, während des Tests die Gewichte der einzelnen Kombinationen zu verändern. So werden etwa auf Kombinationen, die eine deutlich positive Entwicklung zeigen, stärkere Traffic-Anteile gelenkt, während Kombinationen mit schwachem Effekt schon während des Tests zurückgefahren, teilweise auch komplett ausgeschaltet werden. Ein solch dynamisches Vorgehen kann ein guter Kompromiss zwischen einem Start mit maximaler Analysetiefe und Sparsamkeitsprinzip sein.

Die Frage, in wieweit und in welcher Tiefe Interaktionen gemessen und analysiert werden müssen, ist in der Praxis umstritten. Einige (wie Google) plädieren für vollständige Analysen ohne Rücksicht auf zusätzliches Traffic-aufkommen bzw. -notwendigkeit. Professionelle Systeme wie Divolution, Widemile oder Maxymiser setzen dagegen auf eine differenzierte Reduzierung des Testvolumens.

Welches Verfahren genau genutzt werden kann und wie es sinnvoll gesteuert wird, ist hochgradig vom individuellen Test-Design und dem genutzten System abhängig.

7. Fazit

Der Einsatz und die Grundlagen multivariater Verfahren sind nicht ganz unkompliziert und es erfordert gut fundiertes Expertenwissen um sie erfolgreich umzusetzen. Dieses Kapitel sollte einen Eindruck von den konzeptionellen Überlegungen für einen Projektablauf geben und die statistischen Methoden, die ihm zu Grunde liegen erläutern. In der Praxis können Unternehmen heute auf ausgereifte Testsysteme (s. Kapitel XIV) zurückgreifen und mit der Unterstützung erfahrener Praktiker sollte einer erfolgreichen Anwendung nichts im Wege stehen.

Multivariate Verfahren bieten für die Conversion Optimierung von Landing Pages, Bestellprozessen oder Formularseiten unschätzbare Vorteile. Ohne Zweifel werden diese Verfahren die Gestaltung und Produktion von wichtigen Bereichen der Website stark verändern bzw. beeinflussen. Professionell arbeitende Unternehmen werden mittelfristig nicht darauf verzichten können, Online-Angebote in dieser Weise zu optimieren. Die bisherige Arbeitsweise mit einer zunächst ausführlichen Konzeptphase und einem anschließenden Stillstand bis zum nächsten Komplett-Relaunch wird abgelöst werden durch kleine, gezielte und fortgesetzte Anpassungen und Optimierungen der entscheidenden Seitenbereiche. Mit Avinash Kaushik, dem amerikanischen Berater für Web Analytics, lässt sich sagen: „Experiment or Die“.

IX.I Fallbeispiel: AutoScout24 – Evaluierung von Testing Tools

von Sebastian Wetterauer

1. Fachliches Test Setup 186
2. Technisches Test Setup 186
3. „Fahrzeuge Inserieren“ Flights 187
4. „Fahrzeuge Suchen und Finden“ Flights 188
5. Vorläufige Learnings Testing 190

Im April 2008 haben wir bei AutoScout24 vier verschiedene A/B/N und multivariate Testing Tools evaluiert. Ziel war es, eine Lösung zu finden, mit der wir die Webseite optimieren und für Onlinemarketingkampagnen konversionsoptimierte Landing Pages im Hinblick auf spezifische Benutzergruppen schalten können. Für die Evaluierung wurde anhand der internen Anforderungsanalyse eine Bewertungsmatrix mit verschiedenen Dimensionen erstellt, wie zum Beispiel Account- und Supportmanagement, Funktionsumfang, technische Leistungsfähigkeit und Preisstaffelungen. Zwei Lösungen wurden anschließend einem Proof of Concept unterzogen und die Ergebnisse in der Gesamtevaluierung entsprechend berücksichtigt.

1. Fachliches Test Setup

Im Rahmen des Proof of Concept haben wir uns für folgendes fachliches Setup entschieden. Beide Hauptprozesse von AutoScout24 (Suchen & Finden von Fahrzeugen und Inserieren von Fahrzeugen) waren Bestandteil dieses Tests mit der jeweiligen Hauptkonversionsstufe Kontaktaufnahme mit dem Anbieter beziehungsweise erfolgreich eingestelltes Inserat). Um die Nachhaltigkeit der Varianten zu analysieren, wurde jede Konversionsstufe in zwei zeitlichen Dimensionen aufgesetzt, zum einen in einer visitbasierten Konversionsrate und zum anderen in einer sieben-Tage-Konversionsrate. Beide Tests bestanden aus je zwei Flights (Test-Durchgängen) mit je sieben bis acht Rezepten, einer Dauer von zwei bis drei Wochen pro Flight und ungefähr 5.000 Testbesuchern pro Tag. Die Tests wurden der AutoScout24 Applikation durch eine vorgelagerte Subdomain vorangeschaltet, die ausschließlich über Google Adwords angesteuert wurde. Dieses Setup war notwendig, um das Trafficvolumen kontrolliert zu steuern, Keywordsegmentierung und -targeting zu testen und um flexibel und unabhängig von unseren Releasezyklen zu sein. Das Testing Team wurde intern aus vier Spezialisten (Onlinemarketing, Product Management, Usability und Web Analytics) zusammengestellt. So konnten thematische Interdisziplinarität beziehungsweise die unterschiedlichen Abteilungsziele angemessen berücksichtigt werden.

2. Technisches Test Setup

Im technischen Setup haben wir unser Web Analytics Tool zusätzlich in die Landing Pages implementiert. So konnten die Ergebnisse des Testingtools validiert werden und weitere Methoden wie Pfadanalysen und Bounce Rates standen zu Analysezwecken zur Verfügung. Desweiteren haben wir unser Web Analytics Tool zu einer Testcookiesetzung auf

den Landing Pages und der Auslieferung der Konversionsmessung – nur unter der Voraussetzung, dass der User an dem Landing Page-Test teilnimmt – „missbraucht". Das Trafficvolumen von AutoScout24 macht diese Lösung zwingend notwendig, um die Anzahl der Servercalls und damit die Kosten eines Testingtools geringer zu halten. Im ersten Flight wurden zusätzliche User-Segmentierungen implementiert, um dann die User des zweiten Flights nach den relevanten Kriterien differenziert anzusprechen (targeting).

3. „Fahrzeuge Inserieren" Flights

Ziel des Tests war die Entwicklung einer neuen Inserieren-Startseite, die eine Steigerung an erfolgreich eingestellten Inseraten generiert.

Abbildung 49: Alte Inserieren-Startseite

Abbildung 50: Neue Inserieren-Startseite

Während der Testphasen wurden folgende unterschiedliche Testelemente analysiert:

- 1 Button versus 2 Buttons
- Darstellung der Prozessschritte
- Darstellung/Positionierung der Vorteile
- Hervorhebung Pricing
- Varianten ohne Navigation

Die optimierte Inserieren-Startseite hat die Konversionsrate (in der sieben-Tage-Betrachtung) um 50 Prozent gesteigert. Die Hauptfaktoren für diesen Lift waren vor allem die 1-Button-Variante und die deutliche Hervorhebung der einzelnen Prozessschritte. Als zusätzliche Learnings konnten ebenfalls ermittelt werden, dass Landing Pages ohne Navigation visitbasiert sehr gut konvertieren und sich bei AutoScout24 Pricing-Hervorhebungen im „Discounterstil“ negativ auswirken.

4. „Fahrzeuge Suchen und Finden“ Flights

Ziel dieser Flights war es, unterschiedliche Landing Pages für Marketingkampagnen zu entwickeln, die Besucher anhand von Keywordgruppen zu targeten und eine Steigerung erfolgreicher Nachfragen zu generieren.

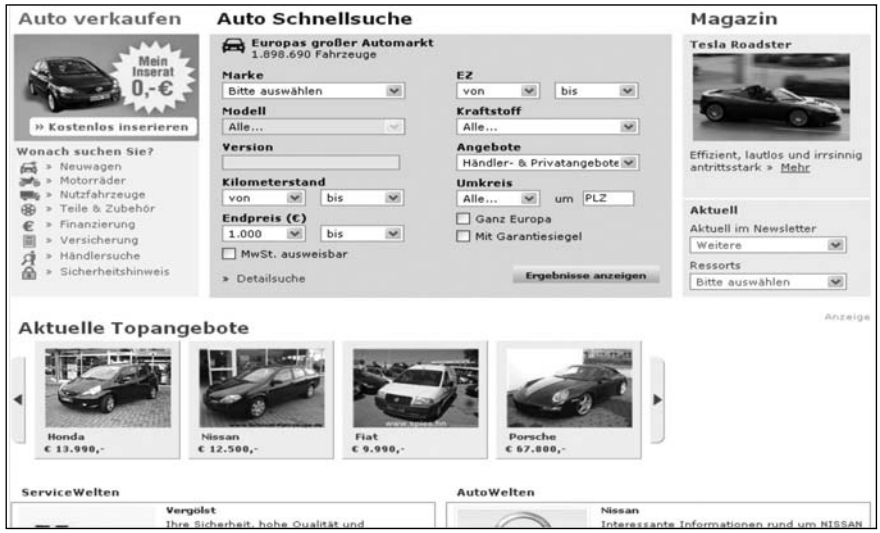


Abbildung 51: Suchen und Finden-Startseite

Die klassische AutoScout24-Startseite ist durch ein dreispaltiges Layout mit der hervorgehobenen, zentral platzierten Suchbox gekennzeichnet. Als Landing Pages für Suchen und Finden, sollten die neuen Seiten das Thema Suchen stärker in den Vordergrund stellen und dem Besucher komfortable Schnelleinstiege in den Suchprozess ermöglichen, ohne dabei vom Prozess abzulenken.

Folgende zwei Varianten wurden daraufhin mit unterschiedlichen Testelementen variiert.

Abbildung 52: Suchen und Finden Variante – Branded Keywords

Abbildung 53: Suchen und Finden Variante – Unbranded Keywords

Zum einen wurde das klassische 3-Spalten-Layout gegenüber einem 2-Spalten-Layout, und zum anderen verschiedene Suchschnelleinstiege getestet (Tagclouds, Icons). Desweiteren variierten wir neben den autoaffinen auch andere Arten von Emotionsbildern.

Das Hauptergebnis des ersten Flights zeigte, dass Besucher mit branded Keywords das bereits gewohnte 3-Spalten-Layout bevorzugen, während Besucher mit unbranded Keywords die 2-Spalten-Varianten vorziehen. Zusätzlich wurden von beiden Besuchergruppen die Suchschnelleinstiege sehr gut angenommen. Bei den Emotionsbildern haben wir festgestellt, dass autoaffine Bilder der Optimierung deutlich stärker dienlich sind und humorvolle Bilder nur wir selbst gut fanden.

Mit diesen Erkenntnissen haben wir im – nach Keywordgruppen differenzierten – zweiten Flight gelernt, dass beide Testsieger zwar im Durchschnitt am besten abgeschlossen haben (Konversionssteigerungen zwischen 10 und 20 Prozent), jedoch erhebliche Unterschiede zwischen Werktagen und Wochenenden existieren, und wir hierfür weitere Targetingkriterien einführen müssen.

5. Vorläufige Learnings Testing

Die Interdisziplinarität eines Testteams aus unterschiedlichen Abteilungen ist zwingend notwendig, da verschiedene Ansätze und Herangehensweisen sinnvoll sind.

Konversionsbetrachtungen mit verschiedenen Cookielaufzeiten sind hilfreich, um die Nachhaltigkeit einzelner Varianten zu messen.

Die zusätzliche Integration einer Web Analytics Lösung hilft, tiefere Einsichten zu bekommen und weiterführende Analysemöglichkeiten zu erhalten, die neue Varianten anregen können.

Eine Loslösung von Styleguides für Layouts ist möglich und sinnvoll.

Besucher-Segmentierung und -Targeting kann der Schlüssel zum Erfolg sein.

Entscheidung: Zu Beginn dieses Jahres wurde Vertrag mit dem Anbieter der besten Lösung und die weitere Testplanungsroadmap sieht einen größeren Test pro Quartal vor. Die Roadmap wird aktuell intern von unserem interdisziplinären Testteam diskutiert und abgestimmt.

IX.II Divolution – Optimierung eines KFZ-Versicherungsangebotes

von Michael Jonas

1. Ausgangssituation 194
2. Projektablauf: Phasen der Planung 195
3. Testkonzeption 196
4. Testdurchführung 197
5. Testergebnisse 198
6. Konkrete Empfehlungen und weiterführende Maßnahmen 199
7. Fazit 200

Nachdem viele Unternehmen ihre Online-Budgets ausgeschöpft haben, geraten jetzt mehr und mehr Maßnahmen zur Effizienzsteigerung ins Blickfeld der Marketingverantwortlichen: die Erhöhung des Return on Investment (RoI) steht im Vordergrund. Aktuelle Online-Projekte zeigen, dass dieser durch eine Optimierung mit Testing und Targeting signifikant gesteigert werden kann.

Anhand eines konkreten Beispiels soll aufgezeigt werden, welche Potenziale in Website-Optimierungsmethoden wie dem multivariaten Testing liegen. Beim multivariaten Testing werden mehrere Contentelemente wie zum Beispiel ein Bild, eine Überschrift, ein Text oder ein Formular in unterschiedlichen Ausprägungen parallel getestet. Dabei geht es darum herauszufinden, welche Zusammenstellung aus den verschiedenen Elementen am meisten Erfolg bringt.

Ein weiterer Pluspunkt: Das Testing erfolgt nicht in der gestellten Umgebung eines Usability-Testlabors, sondern unter Live-Bedingungen mit realen Online-Nutzern, ohne dass diese wahrnehmen, dass sie sich in einem Test befinden. Die Prozessoptimierung findet also unter den konkreten Nutzungsbedingungen statt.

1. Ausgangssituation

Im beschrieben Fall entschied sich ein Anbieter aus der Versicherungsbranche zum Einsatz der Optimierungslösungen von DIVOLUTION für multivariates Website- Testing und Onsite-Targeting. Bei dem Unternehmen handelt es sich um die deutsche Sektion eines internationalen Konzerns, der seine Produkte vorwiegend über das Internet anbietet. Der Schwerpunkt liegt dabei auf KfZ-Versicherungsprodukten.

Der Online-Reifegrad des Unternehmens kann als hoch eingestuft werden: Sowohl das digitale Marketing als auch der Online-Vertrieb des Versicherers sind fest in die Unternehmensstrategie integriert, die entsprechenden Budgets werden bereitgestellt. Zudem ist das Unternehmen offen für neue Entwicklungen beim Online-Marketing (early adopter).

Die Aufgabenstellung beinhaltete, vor dem Hintergrund der Erhöhung der Zufriedenheit der Website-Nutzer, zwei Ziele: Zum einen sollte der Prozess, der den Nutzer von den Werbemitteln auf externen Websites über die Landing Page hinein in den Transaktionsprozess führt, optimiert werden. Zum anderen sollte die Content-Kombination mit der höchsten Conversion Rate bis zum Vertragsabschluss ermittelt werden.

2. Projektablauf: Phasen der Planung

In der Regel durchläuft ein Testing- und Optimierungsprojekt vier Hauptschritte, bis die Ausgangsbasis für eine kontinuierliche Optimierung erreicht ist:

1. Kick-off mit Testzieldefinition
2. Testkonzeption und Produktion der Testelemente
3. Abnahme
4. Livestellung

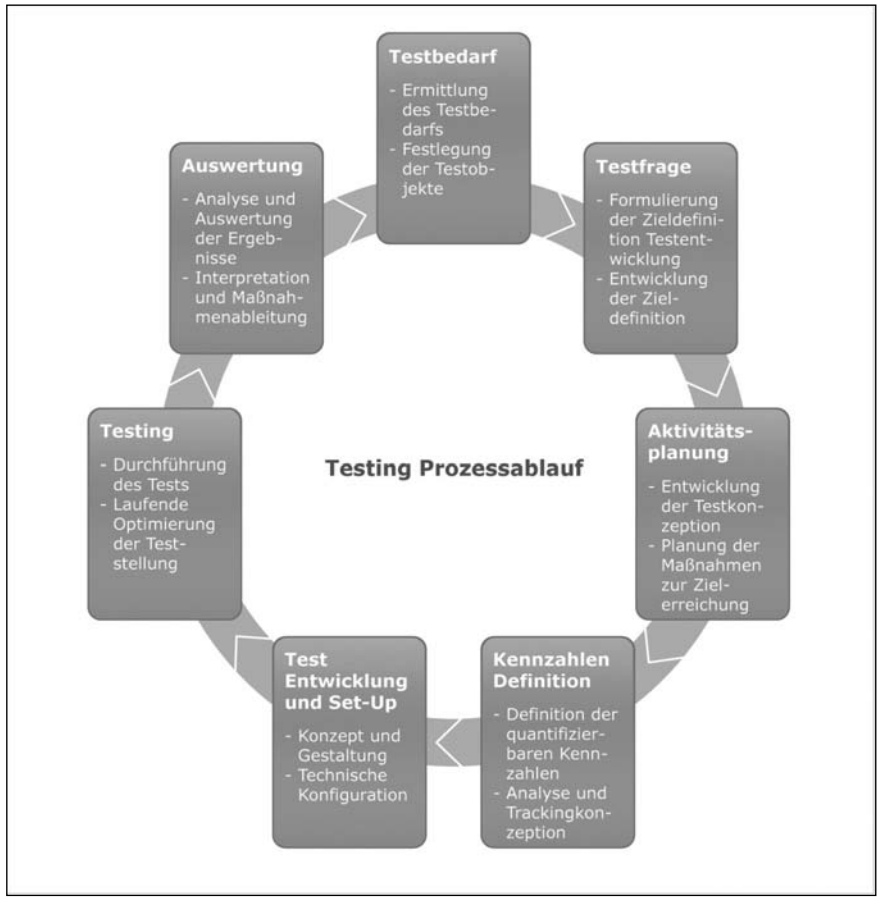


Abbildung 54: Testing Prozessablauf

Am Beginn jedes Testingprojekts steht ein Kick-off. Eine Aufgabe dabei besteht in der Vereinbarung einer gemeinsamen Terminologie, also der Begrifflichkeiten, die bei der Zieldefinition Verwendung finden: Die Projektpartner einigen sich darauf, was gemessen werden soll und welche Messeinheiten verwendet werden (zum Beispiel PageImpressions, Visits, Conversions und Aktionen). Nachdem die Begriffsdefinition erfolgt ist, werden die Optimierungsziele festgelegt.

Im konkreten Fall war das die Optimierung der Teststrecke „Einstieg – Angebot – Vertrag' mit folgender Zielhierarchie:

1. Primärziel: Erhöhung der Sales Conversion durch Prozessoptimierung vom Werbemittel bis zum Vertragsabschluss.
2. Sekundärziel: Erhöhung der Klick-Conversion auf der gesamten Strecke.
3. Ein zusätzlicher Kundenwunsch bestand in der detaillierten Klick-Messung einzelner Teaser.

Auf Basis der Optimierungsziele erfolgt im nächsten Schritt die Erstellung der Testkonzeption, deren zentraler Bestandteil die Testmatrix ist. Dabei handelt es sich in den meisten Fällen um eine einfache Kreuztabelle, welche die verschiedenen Testelemente in Beziehung zueinander setzt und aus der sich ein grober Überblick über die Testvarianten gewinnen lässt.

Parallel dazu kann mit der Kreation der Testelemente (Bilder, Grafiken, Buttons) in den verschiedenen Varianten begonnen werden.

Ein weiterer wichtiger Schritt ist die Abnahme durch den Kunden. Zwingend erforderlich ist dabei die Qualitätssicherung der gesamten technischen Konfiguration. Im Rahmen der Abnahme werden alle inhaltlichen und technischen Parameter geprüft und durch den Kunden frei gegeben. Sind all diese Projektschritte durchgeführt und ist die Freigabe durch den Kunden erfolgt, kann die Livestellung der Testanordnung stattfinden. Im besprochenen Fall betrug der Zeitraum vom Kick-off bis zur Livestellung lediglich vier Wochen.

3. Testkonzeption

Die Erstellung der Testkonzeption erfolgte mit klarer Aufgabentrennung zwischen der Agentur des Kunden und DIVOLUTION: die Technologie und das Frameset kamen von DIVOLUTION, die Inhalte wurden durch die Agentur erstellt.

Im vorliegenden Fall ergaben sich aus der Zieldefinition drei Optimierungsprojekte, die parallel umgesetzt wurden:

1. Optimierung des großen Teasers auf der Startseite inklusive Hintergrundbildern, Buttons und Texten
2. Optimierung der Randteaser auf der Startseite inklusive Hintergrundbildern, Buttons und Texten
3. Optimierung der Landing Page, die als Tarifrechner fungiert: Durch multivariates Testing mit integriertem Targeting sollte die Besucherführung in den Transaktionsprozess optimiert werden.

Alle Testelemente werden mit Hilfe der Testmatrix nachgehalten. Die Anzahl der maximalen Varianten stellt den Gesamtumfang einer Testanordnung dar. Je nachdem, wie viele einzelne Elemente getestet werden sollen, kann die Kombination zu einer riesigen Variantenvielfalt führen. Mit Hilfe einer Variantenreduzierungsmethode können jedoch ähnliche Varianten ausgeschlossen werden.

Die automatische Variantenreduzierung bewirkt, dass Varianten, die unter eine definierte Konversionsrate sinken, abgeschaltet werden und somit nicht teuer eingekaufter Traffic auf diese sogenannten „Looservarianten" gelenkt wird. Der von DIVOLUTION eingesetzte Algorithmus zur automatischen Variantenreduzierung wurde in Zusammenarbeit mit der Uni Bonn entwickelt. Mit dieser Methode wird sichergestellt, dass die Testergebnisse ein Validitätsniveau von 99 Prozent erreichen. Zum Vergleich: In der Werbung, insbesondere im Werbemitteltest, gilt bereits ein Niveau von 90 Prozent als statistisch valide.

4. Testdurchführung

Für das multivariate Testing kam die Testing- und Optimierungssoftware siteOptimizer von DIVOLUTION zum Einsatz. Das Tool wurde über eine Schnittstelle an das Content-Management-System (CMS) des Kunden angedockt. An Stelle des Teasers im CMS wurde dort ein Javascript eingebunden, das sich die entsprechenden Testelemente vom DIVOLUTION-Server holte. Beim Testing der Landing Page wurden dynamische Inhalte vom Werbemittel automatisch übernommen. Dabei wurde der Nutzerstrom segmentiert: Die Besucher fanden auf der Landing Page die Inhalte jenes Werbemittels wieder, das sie auf der externen Website angeklickt hatten. Auf diese Weise kam die Targeting-Technologie von DIVOLUTION zum Einsatz.

Beim multivariaten Testing der Website wurden Hintergrundbilder und Layouts, Buttongrößen und Buttonlayouts, Headlines und Argumentationen, Teaser, Angebote und Inhaltsboxen automatisch und parallel getestet. Konkrete Fragestellungen im Detail waren dabei:

1. Hat das Design Einfluss auf das Nutzerverhalten: Was verursacht Aufmerksamkeit, was nicht? Was wird gut wahrgenommen? Konkrete Testobjekte hierzu: „1 großer Sidebar-Teaser versus 2 Sidebar-Teaser“ oder „1 langer und 1 kleiner Bottom-Teaser versus 3 Bottom-Teaser”.
2. Wie beeinflussen sich Themengruppen/Themencluster gegenseitig? Konkrete Testvarianten hierzu: “Nur Produktfeatures” oder “Produktfeatures mit Servicethemen” oder “Produktfeatures mit Reputation” oder “ein bunter Mix aus allem”.

Auf der gesamten Teststrecke ‚Einstieg – Angebot – Vertrag’ wurde eine Erfolgsmessung durchgeführt: Hierzu erfolgte sowohl eine Analyse der Klickraten auf die einzelnen Inhaltskomponenten der getesteten Elemente für sich, als auch eine Analyse des Erfolgs der einzelnen Varianten im Kontext des Bestell- und Buchungsprozesses.

Um bei den Besuchern keine Verunsicherung zu erzeugen, stellte DIVOLUTION durch die patentierte Client-DNA zur cookielosen Nutzererkennung sicher, dass ein wiederkehrender Besucher stets die gleiche Variante zu sehen bekam, bei der er das letzte Mal die Site verlassen hatte.

Beim vorliegenden Case wurden über einen Zeitraum von drei Monaten insgesamt etwa 940.000 Auslieferungen der getesteten Teaser vorgenommen, wobei der Top-Teaser 570.000 Mal und der Bottom-Teaser 370.000 Mal ausgeliefert wurde.

5. Testergebnisse

Das Testing und Targeting ermöglichte es, potenzielle Versicherungskunden dynamisch und kanalbezogen anzusprechen. Durch das dynamische Targeting funktionierte die Landing Page als Tarifrechner-Startseite und der gesamte Prozess konnte verschlankt werden. Dies führte dazu, dass die Abbrecherquote im Vergleich zur Ausgangskonversion um 25 Prozent gesenkt werden konnte.

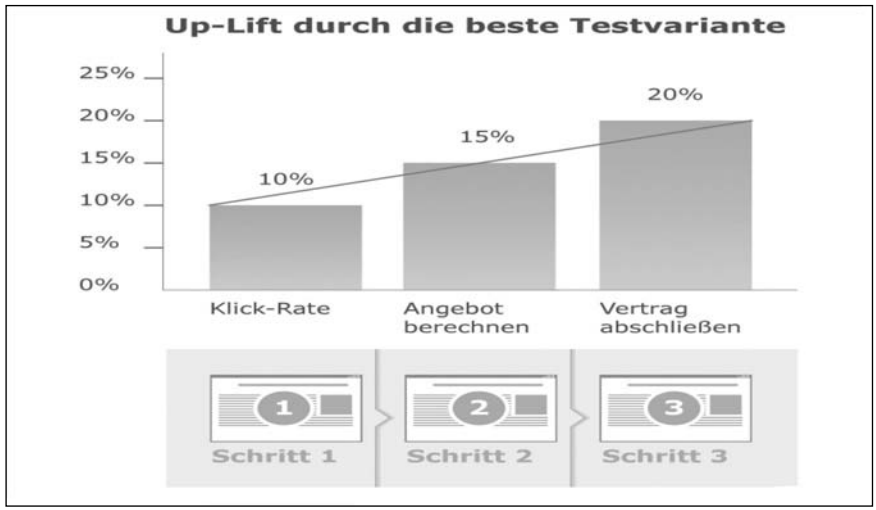


Abbildung 55: Steigerung der Erfolgsraten

Die beste Testvariante führte zu folgendem Up-Lift (Erhöhung) der Ergebnisse im Vergleich zur Ausgangsvariante: 10 Prozent bei der Klickrate, 15 Prozent bei der Berechnung von Angeboten und 20 Prozent bei Vertragsabschlüssen. Somit konnte die Zahl der Anträge und Abschlüsse bei gleichen Kosten signifikant erhöht werden.

Bei den Detailfragen konnten folgende Einzelergebnisse gewonnen werden:

1. Große Sidebar-Teaser erwiesen sich als eher ungeeignet. Offensichtlich kannibalisieren sich ein großer Top- und ein Sidebar-Teaser in der Aufmerksamkeit der Nutzer.
2. Hingegen funktionierten längere Bottom-Teaser.
3. Es konnte kein Zusammenhang bei der Kombination von Themengruppen gefunden werden.

6. Konkrete Empfehlungen und weiterführende Maßnahmen

Aus den Erkenntnissen des Testings wurden Empfehlungen für weiterführende Maßnahmen abgeleitet:
Im nächsten Schritt werden die gut funktionierenden Themen bei der Anordnung der Bottom- und Sidebar-Teaser berücksichtigt. Zudem werden die Learnings aus dem Testverfahren auch auf tiefere Websitebereiche übertragen. Ebenso können die Erkenntnisse auf die SEM-Landing Pages angewendet werden.

Aus den Ergebnissen des Testings konnten Kriterien abgeleitet werden, die in ein Regelwerk für Geo- und Behavioral Targeting eingehen: So wird ein optimiertes Targeting auf Produktebene, auf Customer LifeCycle Ebene oder Tarifebene ermöglicht. Mit Hilfe des Testings in Verbindung mit Targeting konnten die Produktangebote individuell an die Bedürfnisse der Interessenten angepasst werden.

Eine besondere Herausforderung stellte die Optimierung des Tarifrechners dar: Da er durch die Berechnungsalgorithmen und die Programmierung eng mit der bestehenden Website verzahnt war, konnte er nicht isoliert auf einer getrennten Strecke getestet werden. Die Lösung besteht in einem parallelen Testing: Ein Dummy in Form einer parallelen Teststrecke wird aufgesetzt, auf die die Nutzer über eine Traffic-Weiche geleitet werden.

7. Fazit

Die Vorteile des multivarianten Testing-Verfahrens liegen, wie das Projekt deutlich zeigte, insbesondere darin, dass viele Kommunikationsideen nebeneinander getestet werden können. So können sehr schnell aussagekräftige Ergebnisse erzielt werden. Das gesamte Verfahren ist hoch flexibel, die Optimierung kann durch neue Erkenntnisse aus den Tests fortwährend angepasst und die laufenden Testvarianten kontinuierlich ausgewertet werden. Funktionierende Gewinner-Varianten können mit weiteren Elementen neu kombiniert und auf Basis der gewonnen Ergebnisse sofort weiter optimiert werden.

Ein Grund für den positiven und reibungslosen Projektablauf bestand unter anderem darin, dass das Verständnis für die Online-Optimierungsthematik bei allen Beteiligten (Kunde, Agentur und DIVOLUTION) stark ausgeprägt war und die notwendige Unterstützung von Entscheiderebene vorhanden war. Somit konnten Entscheidungsprozesse bei den Online-Marketing-Verantwortlichen verkürzt werden und die Umsetzung schneller erfolgen. Das Projekt kombinierte zudem das multivariate Testing-Verfahren mit einem einfachen Targeting. Dieses Zusammenspiel stellt eine der zukunftsweisenden Entwicklungen im Online-Marketing zur effektiveren Nutzeransprache dar.

Das Beispiel zeigt auch eindrucksvoll, dass der geforderte RoI durch eine Optimierung mit multivariatem Testing signifikant gesteigert werden kann. In der Kombination mit Targeting auf der Basis generierter Nutzerprofile wird diese Effizienzverbesserung zukünftig noch einmal einen deutlichen Schub erfahren. Testing ist somit nur der Einstieg in die 360-Grad-Optimierung der idealen Kundenansprache im Online-Vertriebskanal.

IX.III Fallbeispiel: Arbeit mit dem Google Website Optimizer

von Birger Friedrichs

1. Test-Typ 1: A/B Test 202
2. Test-Typ 2: Multivariater Test 202
3. Was soll ich testen oder die Hypothese 203
4. Das Testszenario und die Hypothese bei SIBO 204
5. Die Durchführung des Experiments 205
6. Das Ergebnis 211
7. Nützliche Tools, Websites und Bücher 212

Das kostenlose Tool Google Website Optimizer hilft Ihnen, Ihre Website für Conversions zu optimieren. Dabei sind Conversions als allgemein anzusehen und Sie können frei definieren, was eine Conversion auf Ihrer Website darstellt. Dies kann zum Beispiel eine Bestellung in Ihrem Online-Shop, eine Kontakt-Anfrage, eine Newsletter-Anmeldung, ein Download einer Broschüre oder das Ablegen eines Produktes in den Warenkorb sein.

Google Website Optimizer

http://www.google.com/websiteoptimizer

Google Website Optimizer stellt Ihnen zwei verschiedene Typen von Tests zur Verfügung. Damit erlaubt er Ihnen, unterschiedliche Varianten von Webseiten miteinander zu vergleichen (A/B Test) oder auch einzelne grafische oder inhaltliche Bereiche auf einer Webseite zu testen und die beste Kombination bzw. Variante zu ermitteln (Multivariater Test). Langfristig können Sie durch den Einsatz des Google Website Optimizers die Conversion-Rate und somit den Ertrag Ihrer Website steigern, da Sie mit der gleichen Anzahl an Besuchern mehr Verkäufe, Anfragen, etc. generieren können.

1. Test-Typ 1: A/B Test

Der erste Typ ist der A/B Test. Mit ihm lassen sich mehrere Webseiten-Varianten gegeneinander testen und die beste Variante ermitteln. Wenn Sie zum Beispiel das Layout Ihrer Website ändern möchten, können Sie hier einen oder mehrere Layout-Vorschläge gegen Ihr ursprüngliches Layout testen. Am Ende des Tests wissen Sie dann, welcher Layout-Vorschlag die meisten Conversions erzielen wird. Welche Änderungen genau dafür verantwortlich sind, dazu sagt dieser Test-Typ nichts aus. Google empfiehlt diesen Typ für Webseiten mit geringen Zugriffszahlen (weniger als 1.000 Zugriffe pro Woche).

2. Test-Typ 2: Multivariater Test

Der zweite Test-Typ nennt sich Multivariater Test. Mit diesem Typ haben Sie die Möglichkeit auf der gleichen Seite mehrere inhaltliche Änderungen durchzuführen und diese in Kombination gegeneinander zu testen. Sie definieren für die Seite einen Bereich, den Sie testen möchten, und erstellen für diesen Bereich eine oder mehrere Varianten. So können Sie zum Bespiel für Ihre Überschrift mehrere Textvarianten entwerfen sowie bei einem Bild

Abbildung 56: Auswahl des Testtyps

unterschiedliche Bildvarianten gegeneinander testen. Google Website Optimizer wird dann alle verfügbaren Kombinationen durchprobieren und Ihnen die beste Kombination(en) bzw. den Gewinner präsentieren. Bei einem Test von drei Überschriften und drei Bildern ergeben sich somit insgesamt neun verschiedene Kombinationen. Da hier sehr schnell viele Kombinationen zustande kommen können, empfiehlt Google diese Variante bei mehr als 1.000 Zugriffen pro Woche. Sie sollten klein anfangen und es vermeiden, mit einer großen Anzahl von Kombinationen zu starten. Je größer die Anzahl an Kombinationen, desto mehr Traffic bzw. Besucher benötigen Sie für ein statistisch aussagekräftiges Ergebnis.

3. Was soll ich testen oder die Hypothese

Bei der Arbeit mit dem Website Optimizer steht die Frage „Was soll ich testen?“ im Vordergrund. Das ist eine Frage, die Sie sich ganz am Anfang stellen sollten. Meistens werden Sie aus der eigenen Neugier heraus selbst Fragen haben:

- *Warum legen so wenige Besucher ein Produkt in den Warenkorb?*
- *Warum erhalte ich so wenige Anfragen über meine Website?*

Jetzt gilt es aus der Frage eine Hypothese zu bilden, die Sie dann beweisen oder widerlegen können. Erstellen Sie aus Ihren Fragen eine Hypothese, indem Sie sich mögliche Antworten auf Ihre Fragen überlegen:

→ *Ein größerer Warenkorb-Button erhöht die Conversion Rate.*
→ *Ein veränderter Anfrage-Text erhöht die Anzahl der Anfragen.*

Durch eine simple Antwort auf Ihre Frage, haben Sie nun eine Hypothese geschaffen, die Sie mit dem Google Website Optimizer in einem Experiment testen und beantworten können. Als Beispiel könnten Sie ein Experiment zur Größe des Warenkorb-Buttons erstellen. Sie erstellen mehrere Varianten des Buttons und messen über eine bestimmte Zeit den Erfolg oder Misserfolg gegenüber der Original-Version.

Mehr Anregungen zu Test-Möglichkeiten finden Sie auch in der Google Hilfe (http://www.google.com/support/websiteoptimizer/bin/answer.py?answer=61205&topic=14314&hl=de).

4. Das Testszenario und die Hypothese bei SIBO

Die SIBO Website www.sibo-netzwerkschraenke.de ermöglicht es Benutzern, sich Netzwerkschränke, Serverschränke, Wandgehäuse oder Racks zu konfigurieren und dann eine Angebotsanfrage über die Website zu stellen. Eine Preisinformation zu den Produkten war vom Anbieter nicht vorgesehen, da diese abhängig ist von der individuellen Konfiguration und der Bestellmenge. Das Ziel der Website ist die Anzahl der Anfragen über die Website zu steigern. Eine Frage, die SIBO bewegt ist demnach:

- *Wie kann ich die Anfragen über die Website erhöhen?*

Nach Absprache mit dem Anbieter ergibt sich, dass er die Möglichkeit hat, Netzwerkschränke zu einem günstigen Preis im Vergleich zur Konkurrenz anzubieten. Die Erfahrung zeigt auch, dass im Internet viel über den Preis angefragt oder entschieden wird. Die gesuchten Produkte werden meistens auf mehreren Websites angeboten und sind über die Suchmaschinen auch nur wenige Klicks entfernt. Mit dieser Überlegung wird folgende Hypothese aufgestellt:

- *Die Darstellung von Preisinformationen steigert die Anfragerate.*

Abbildung 57: Auswahl des Testtyps

Diese Hypothese gilt es nun mit dem Google Website Optimizer in einem Experiment zu testen und zu beobachten, ob die Anzahl der Anfragen bei der Darstellung von Preisinformationen wirklich höher ist als beim Weglassen von Preisinformationen.

5. Die Durchführung des Experiments

Da die Hypothese feststeht, geht es nun um die Einrichtung und die Durchführung des Experiments mit dem Google Website Optimizer. Da sich nicht für jeden Konfigurationstyp Preise darstellen lassen, werden nur Ab-Preise für jede Produktkategorie dargestellt. Die Preise sollen für jede Kategorie (Netzwerkschrank, Serverschrank, etc.) individuell sein und auf der Beschreibungsseite der Kategorie mittels eines Preisschildes deutlich dargestellt werden. Da es sich bei der Zielgruppe von SIBO um Firmen handelt, soll der Netto-Preis auf dem Preisschild angegeben werden. Zusätzlich wird unterhalb des Beschreibungstextes

noch einmal der Netto- als auch der Brutto-Preis angegeben, um mögliche Abmahnungen auszuschließen (siehe auch Abbildung 61).

Auswahl Test-Typ: Multivariater Test

Die Einrichtung des Experiments beginnt mit der Entscheidung für den Test-Typ (siehe auch Abbildung 56). Beim SIBO Experiment fällt die Entscheidung auf den multivariaten Test, da in diesem Fall keine vollkommen neue Seite getestet, sondern ein bestimmter Inhalt bzw. Bereich auf mehreren Seiten variiert werden soll. Letzteres ist mit dem Test-Typ A/B-Test so nicht möglich.

Test- und Zielseiten bestimmen

Mit dem nächsten Schritt wird dem Experiment ein Name gegeben. Damit kann das Experiment in der Liste auf der Übersichtsseite des Google Website Optimizers leichter identifiziert werden. Weiterhin muss die Testseite und die Ziel- bzw. Conversion-Seite angegeben werden (siehe Abbildung 58).

Abbildung 58: Testseite und Zielseite bestimmen

Die Testseite ist die Seite, auf der später die Inhalte bzw. Preisinformationen variiert und getestet werden sollen. Auf der SIBO-Seite sind das mehrere Seiten. Da aber nur eine Testseite angegeben werden kann, wird auch nur eine beispielhaft eingetragen.

Die zweite Seitenangabe bezieht sich auf die Ziel-Seite bzw. die Conversion-Seite. Das ist die Seite, anhand derer Google Website Optimizer erkennt, dass eine Anfrage abgeschickt oder eine Conversion erzielt wurde. In diesem Fall ist es das Skript, welches die Mail mit den Konfigurationswünschen verschickt.

Installation des Google Website Optimizer Codes

Nachdem Testseite und Zielseite definiert sind, kann der JavaScript-Code des Google Website Optimizers auf den Seiten eingebaut werden. Google bietet hier zwei Möglichkeiten an (siehe Abbildung 59): Ein Dienstleister übernimmt die Implementierung des Codes für Sie oder Sie installieren den Code selbst.

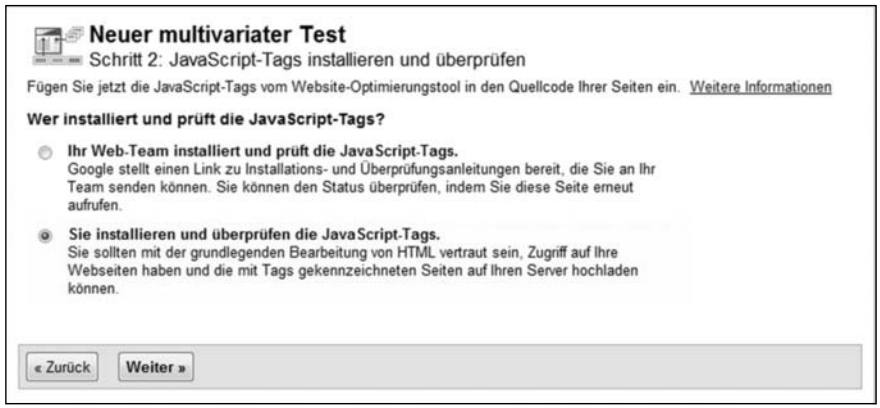


Abbildung 59: Wer übernimmt die Installation des JavaScript-Codes?

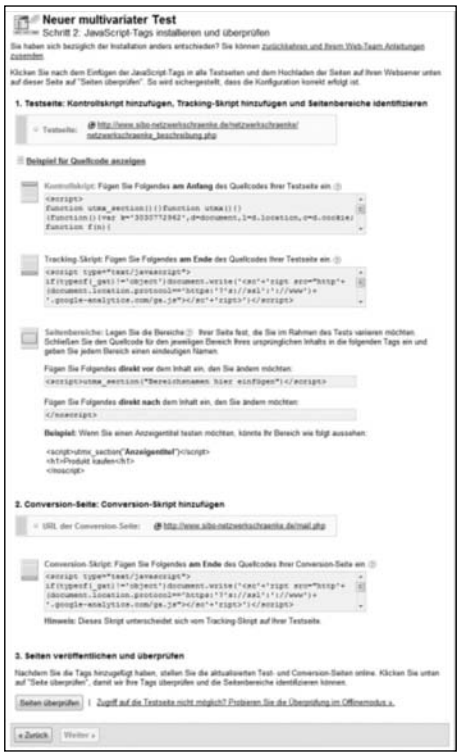


Abbildung 60:
Der Javascript-Code für den Google Website Optimizer

Wählen Sie einfach die gewünschte Option aus. Wenn Sie die Implementierung einem Fachmann überlassen möchten, dann wählen Sie die erste Option. Google bietet Ihnen dann einen speziellen Link an, den Sie Ihrem Fachmann zum Beispiel per E-Mail schicken können. Der Link führt zu einer Seite, auf welcher eine Anleitung sowie der JavaScript-Code für dieses Experiment dargestellt sind. Wenn Sie sich für die zweite Option entscheiden, kommen Sie auf der nächsten Seite direkt zu der Anleitung und den JavaScript-Codes (siehe Abbildung 60). Diese müssen Sie dann auf der Testseite und auf der Zielseite wie angegeben einbauen.

Auf den Testseiten müssen drei verschieden Skripte eingefügt werden: Das Kontrollskript steuert die anzuzeigenden Varianten. Bei jedem Aufruf der Webseite wird einem neuen Besucher eine bestimmte Kombination von Varianten zugewiesen. Ein wiederkehrender Besucher wird anhand eines vom Kontrollskript zuvor gesetzten Cookies erkannt und ihm wird die gleiche Kombination wie beim letzten Besuch angezeigt.

Das Tracking-Skript ist für die spätere Auswertung wichtig. Es zählt die Besucher und ordnet sie den angezeigten Varianten bzw. der Kombination zu.

Die Seitenbereiche, die den zu variierenden Inhalt enthalten, werden ebenfalls mit einem speziellen JavaScript-Code definiert. Jeder Bereich muss mit einem Namen gekennzeichnet werden. In diesem Fall gibt es zwei Bereiche, die definiert werden:

„Preistag" – Bereich für das Preisschild

„Preistag-Text" – Bereich für die Preisinformationen unterhalb des Beschreibungstextes
Der HTML-Inhalt innerhalb der Bereiche kann später über die Oberfläche des Google Website Optimizer angepasst oder ersetzt werden. Das ist wichtig: Die HTML-Änderungen für die verschiedenen Varianten werden nicht mehr direkt im Code der Seite, sondern im Google Website Optimizer durchgeführt.

Auf der Zielseite muss nur ein JavaScript-Code eingebaut werden: Das sogenannte Conversion-Skript. Es ist ähnlich wie das Tracking-Skript auf der Testseite, zählt aber die Besucher, die in diesem Fall erfolgreich eine Anfrage erzeugt haben.

Der Website Optimizer ermöglicht eine Prüfung der eingebundenen JavaScript-Codes, um sicherzustellen, dass alles korrekt eingebaut wurde.

Hürde: Dynamischer Webseiten-Inhalt

Bei der Nutzung von dynamischen Webseiten-Inhalten können sich Probleme ergeben. Auf der SIBO-Website werden die Preisinformationen für jede Kategorie dynamisch durch Skripte auf dem Server ausgelesen und dann wird die Webseite mit dem Preisschild und den Preisinformationen an den Browser des Benutzers gesendet. Die inhaltlichen Variationen des Google Website Optimizers basieren allerdings auf JavaScript-Code, welcher erst auf dem Computer des Benutzers ausgeführt wird. Somit lassen sich keine Varianten erzeugen, die auf Server-Skripte zurückgreifen.

Auf der SIBO-Website wird das Problem so gelöst, dass die Website im Original die Preisschilder und die Preisinformationen anzeigt. Im Google Website Optimizer wird dann für die beiden Bereiche jeweils eine zusätzliche Variante angelegt, bei welcher die HTML-Bereiche leer sind und somit keine Preisinformationen ausgegeben werden (siehe Abbildung 62 und 63). Diese Variante stellt die ursprüngliche Version dar (keine Preisinformationen). Im Google Website Optimizer wird es dann genau umgekehrt dargestellt: Die ursprüngliche Version ist in diesem Fall die Version mit Preisinformationen.

Abbildung 61: Ursprüngliche SIBO-Website mit Preisschild und Preisinformation

Abbildung 62: SIBO-Website Variante ohne Preisschild und ohne Preisinformation

Test-Kombinationen deaktivieren

Der Website Optimizer ermöglicht das Deaktivieren von bestimmten Kombinationen. Dies ist nützlich für Kombinationen, die keine wirkliche Verbesserung zeigen oder für das Experiment generell nicht relevant sind. Damit stehen diese Kombinationen den Benutzern nicht mehr zur Verfügung und der Website-Traffic wird besser für die noch bestehenden Kombinationen genutzt. Denn die Anzahl der Kombinationen und der Website-Traffic haben einen entscheidenden Einfluss auf die Dauer des Experiments.

Durch die Verwendung von zwei Bereichen auf der SIBO-Webseite (Preisschild und Preisinformation) mit jeweils zwei Varianten (Preis sichtbar und Preis nicht sichtbar) ergeben sich insgesamt vier verschiedene Kombinationen, die Google Website Optimizer gegeneinander testet. Von diesen vier Kombinationen sind aber nur zwei wirklich relevant. Die Kombinationen, die entweder nur das Preisschild oder nur die Preisinformation im Text anzeigen, sind sinnlos und werden deaktiviert (siehe Abbildung 63).

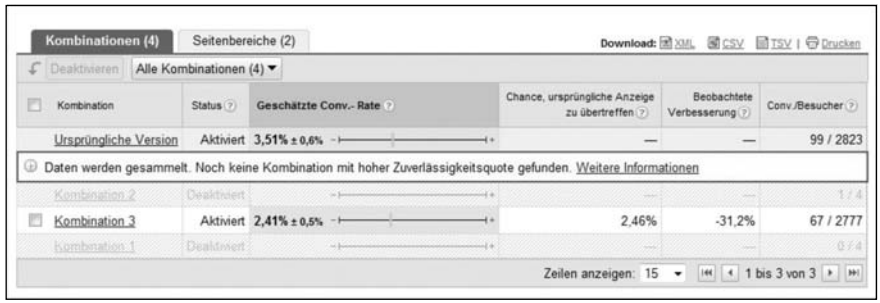
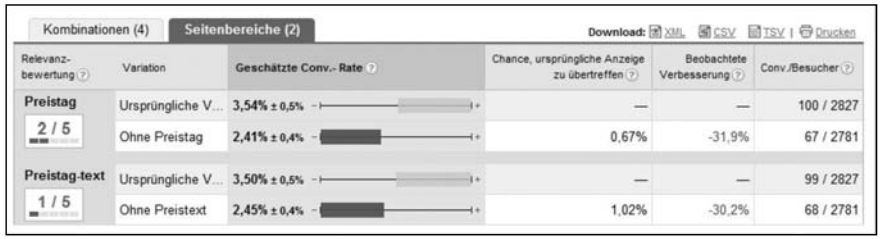


Abbildung 63: Deaktivierte Kombinationen

6. Das Ergebnis

Während das Experiment läuft, zählt Google Website Optimizer mittels der in den Seiten implementierten Tracking-Skripte die Besucher und Conversions zu den einzelnen Varianten und Kombinationen. Sobald ein klarer Gewinner feststeht, wird Ihnen Google dies Bericht zu Ihrem Google Website Optimizer Experiment mit dem Satz „Kombinationen mit hoher Zuverlässigkeitsquote gefunden“ mitteilen. Dies hängt je nach Experiment von der Anzahl der Kombinationen und des zur Verfügung stehenden Website-Traffics ab. Es kann auch vorkommen, dass keine Kombination mit hoher Zuverlässigkeit angezeigt wird (oder es werden einfach noch mehr Daten benötigt).

Beim SIBO-Experiment ist in den Berichten eine Besonderheit zu beachten: Da die als ursprüngliche Version angezeigte Variante aufgrund der Implementierung, die eigentliche Testvariante ist (Webseite mit Preisinformationen), müssen die Ergebnisse umgekehrt bewertet werden. Eine angezeigte Verschlechterung gegenüber der ursprünglichen Version ist in diesem Fall eine Verbesserung durch die Darstellung der Preisinformationen und eine angezeigte Verbesserung wäre eine Verschlechterung. Dies klingt kompliziert, lässt sich aber sehr leicht anhand Abbildung 64 nachvollziehen. Im SIBO-Beispiel erzielt die Variation ohne Preisinformationen durchschnittlich eine Verschlechterung von 31 Prozent gegenüber der ursprünglichen Version. Dies bedeutet im Umkehrschluss, dass die Darstellung von Preisinformationen auf der Website eine Verbesserung der Conversion-Rate bzw. der Anfrage-Rate von 45 Prozent zur Folge hat: Die Conversion-Rate erhöht sich von 2,41 Prozent auf 3,5 Prozent.



Kombinationen (4) | Seitenbereiche (2) | Download: XML CSV TSV | Drucken

Relevanz-bewertung	Variation	Geschätzte Conv.-Rate	Chance, ursprüngliche Anzeige zu übertreffen	Beobachtete Verbesserung	Conv./Besucher
Preistag 2 / 5	Ursprüngliche V...	3,54% ± 0,5%	—	—	100 / 2827
	Ohne Preistag	2,41% ± 0,4%	0,67%	-31,9%	67 / 2781
Preistag-text 1 / 5	Ursprüngliche V...	3,50% ± 0,5%	—	—	99 / 2827
	Ohne Preistext	2,45% ± 0,4%	1,02%	-30,2%	68 / 2781

Abbildung 64: Ergebnis: Die Version ohne Preisinformation schneidet circa 31 Prozent schlechter ab

Sie sehen also, welche positiven Folgen kleinere Änderungen an Ihrer Website haben können. Das Tolle daran ist, dass die Tools zum großen Teil kostenlos zur Verfügung stehen. Sie müssen sich nur mit den Tools auseinandersetzen oder mit Ihrem Dienstleister zusammensetzen, um die Performance Ihrer Website zu steigern.

7. Nützliche Tools, Websites und Bücher

In diesem Abschnitt möchte ich Ihnen noch eine kleine **Übersicht von nützlichen Websites, Tools und einem Buch** zur Verfügung stellen, die das Arbeiten mit dem Website Optimizer erleichtern oder zusätzlichen Mehrwert schaffen. Diese Liste soll keinen Anspruch auf Vollständigkeit erheben und kann durchaus schnell wieder veralten. Gerne nehme ich auch Tipps von Ihnen entgegen.

Website Optimizer Google Hilfe Seiten

Google stellt zu seinen Produkten eine umfangreiche Online-Hilfe zur Verfügung. So ist es auch beim Google Website Optimizer Produkt. Bei Problemen und Fragen ist dies Ihre erste Anlaufstelle. Viele Themen sind hier schon aufgegriffen und beschrieben.
Siehe: *http://www.google.com/support/websiteoptimizer/*

Dauer eines Experiments berechnen

Wie lange wird ein Experiment laufen müssen, um genügend Daten für ein Ergebnis zu sammeln? Mit dem Website Optimizer Duration Calculator können Sie durch Eingabe einiger Variablen die voraussichtliche Dauer Ihres Experimentes berechnen bzw. schätzen lassen. Geben Sie einfach die Anzahl der Kombinationen Ihres Tests, Ihre Seitenzugriffe pro Tag, die momentane Conversion-Rate sowie die erwartete prozentuale Verbesserung ein. Der Rechner wird Ihnen dann ausrechnen, wie viele Tage das Experiment voraussichtlich laufen wird.
https://www.google.com/analytics/siteopt/siteopt/help/calculator.html

Tool zur Abschätzung der voraussichtlichen Stichprobengröße

Zum Einsatz des Google Website Optimizer brauchen Sie selbstverständlich Besucher, mit denen Sie Ihre Website-Experimente durchführen können. Mit dem folgenden Tool **(Sample Size Estimation Tool)** können Sie ungefähr ermitteln, wie groß Ihre Stichprobengröße pro Kombination (Besucheranzahl) sein muss, wenn Sie bei einer bestimmten Conversion-Rate der Original-Seite, einer erwarteten Verbesserung um x-Prozent, eine Sicherheit von y-Prozent erzielen möchten.
http://www.websharedesign.com/conversion-tools/sample-size-estimator-tool/

Conversion Confidence Calculator

Mit dem Conversion Confidence Calculator können Sie zusätzlich überprüfen, wie sicher (statistisch gesehen) die bisher gesammelten Daten für zwei Test-Varianten sind. Geben Sie einfach die Anzahl der Besucher und Conversions für Variante A und B ein. Das Tool berechnet dann, wie wahrscheinlich es ist, dass Variante B auch in Zukunft besser oder schlechter konvertiert als Variante A.
http://www.prusak.com/tools/conversion-confidence-calculator/

Google Website Optimizer und Google Analytics verbinden

Der Google Website Optimizer hat einen großen Nachteil: Mit diesem Tool können Sie nur Besucher und Conversions messen – mehr nicht. Sie wissen also nicht, was auf Ihrer Website genau passiert bzw. welchen Einfluss Ihre Experimente und Kombinationen auf

das Besuchsverhalten Ihrer Website-Besucher haben. Mit einem kleinen Trick und dem Einsatz von Google Analytics ist dies aber dennoch (auch beim Einsatz eines multivariaten Tests) möglich. Den notwendigen JavaScript-Code und eine Erklärung finden Sie auf dem englischsprachigen Blog von ROI Revolution. Durch die Verknüpfung der Daten haben Sie nun Zugriff auf weitere Berichte wie zum Beispiel Absprungrate, Seitenaufrufe, eindeutige Seitenaufrufe, durchschnittliche Besuchszeit und mehr.
http://www.roirevolution.com/blog/2008/05/using-website-optimizer-with-google-analytics-new.html

Google Website Optimizer Guide und Hilfe

Eine private Sammlung von Hinweisen, Tipps und Tricks zur Nutzung und zur Einrichtung des Google Website Optimizers steht mit dem Google Website Optimizer Guide (in Englisch) zur Verfügung. Hier erfahren Sie wertvolle Tipps rund um die Planung und Durchführung von Google Website Optimizer Experimenten. *(http://www.websiteoptimizerguide.com/)*

Buch: Always Be Testing

Wer sich nicht scheut, auch ein englisches Buch zu lesen, dem lege ich das folgende Buch ans Herz: Always Be Testing – The Complete Guide to Google Website Optimizer. Geschrieben wurde es von Bryan Eisenberg, John Quarto-vonTivadar und Lisa T. Davis. Bryan Eisenberg ist Gründer der Firma FutureNow, Autor und hat langjährige Erfahrungen im Bereich der Website und Conversion Optimierung. Sein Buch ist nicht nur für Nutzer des Google Website Optimizer interessant, da er in vielen Kapiteln und mit realen Beispielen Anregungen und Ideen liefert, was sich alles testen lässt.

X. User Experience Tests – Webseiten ganzheitlich optimieren

von Daniel R. Schmeißer

1. Gegenstand des User Experience Tests 218
2. Vorbereitung von UX-Tests 225
3. Durchführung von UX-Tests 231
4. Ableiten von Handlungsempfehlungen 240
5. Fazit: Durch Full User Experience zum Erfolg 243
6. Literaturverzeichnis 244

Wenn Sie an dieser Stelle ein Kapitel über Usability-Tests erwartet haben, sollten Sie nicht abbrechen zu lesen. Es hat sich nämlich in den letzten Jahren zunehmend durchgesetzt, den engeren Begriff der Usability durch den weiteren und spannenderen Begriff der User Experience zu ersetzen. Insbesondere bei der Optimierung von Webseiten hat sich dieser Ansatz als zielführend erwiesen, denn im Zuge der Etablierung verbindlicher und allgemein zugänglicher Usability-Standards zeigt sich, dass sich Webseiten nicht allein durch eine gute Nutzungsfreundlichkeit (Usability) voneinander differenzieren lassen. Webseiten müssen nutzerfreundlich sein, damit sie erfolgreich sind. Eine gute Usability allein macht eine Website aber weder begehrlich, noch ist sie eine Garantie für den Erfolg am Markt.

Man kann daher auch von **Usability als Hygienefaktor** sprechen (in Unterscheidung zu sog. Motivationsfaktoren). Die Unterscheidung zwischen „Hygienefaktoren" und „Motivatoren" geht auf die Zwei-Faktoren-Theorie (auch Motivator-Hygiene-Theorie) von Frederik Herzberg zurück. Grundthese ist, dass Zufriedenheit und Unzufriedenheit nicht die beiden äußersten Ausprägungen einer Eigenschaft darstellen, sondern als zwei unabhängige Eigenschaften zu betrachten sind: Unter Hygienefaktoren versteht Herzberg Faktoren, welche bei positiver Ausprägung die Entstehung von Unzufriedenheit verhindern, aber nicht zur Zufriedenheit beitragen. Häufig werden diese Faktoren gar nicht bemerkt oder als selbstverständlich betrachtet. Sind sie aber nicht vorhanden, empfindet man dies als Mangel.

Übertragen auf Usability bedeutet dies: Ihr Vorhandensein ist notwendige Voraussetzung für ein positives Nutzungserleben (User Experience), ihre mangelhafte Umsetzung macht unzufrieden; umgekehrt motiviert Usability allein aber noch nicht zur Nutzung einer Website. Usability ist vielmehr als **Teil des umfassenderen Konzepts der User Experience** zu sehen. Eine positive User Experience umfasst sämtliche Facetten des menschlichen Erlebens und Verhaltens, von der anfänglichen Orientierung auf einer Website, über das Markenerleben und Kaufverhalten, bis hin zur Wahrnehmung emotionaler und hedonischer Qualitäten. Neben der „technischen" Sicht, die auf Funktionalität und Prozesslogik fokussiert, ist daher eine professionelle **psychologische Perspektive** zwingend erforderlich, um bei der Optimierung einer Website ein uniques Nutzungserlebnis zu schaffen.

Wenden wir uns aber zunächst noch einmal dem Begriff der Usability zu. Dieser stellt nach wie vor eine zentrale Kategorie dar und steht in der Alltagspraxis bei User Experience Tests häufig im Mittelpunkt. Es gibt mit der DIN EN ISO 9241-11 eine seit Jahren etablierte Norm, die Usability als das Ausmaß definiert, „in dem ein technisches System durch seine Benutzer in einem bestimmten Nutzungskontext verwendet werden kann, um bestimmte

Ziele **effektiv, effizient** und **zufriedenstellend** zu erreichen".[39] Aus dieser Definition resultieren eine Reihe anerkannter Kriterien, die leistungsorientiert sind und vor allem Systemeigenschaften, weniger die subjektive Erlebensqualität der Nutzer beschreiben. Hierzu gehören die (stärker aus den Ingenieurswissenschaften und der Software-Ergonomie abgeleitete) Kriterien „Aufgabenangemessenheit", „Selbstbeschreibungsfähigkeit", „Steuerbarkeit", „Erwartungskonformität", „Fehlerrobustheit", „Individualisierbarkeit" und „Lernförderlichkeit". Andere Definitionen fokussieren auf die Interaktionsebene, so zum Beispiel die Definition von Nielsen[40], der Usability ganz allgemein als die Qualität des Nutzungserlebens während der Interaktion mit einem System beschreibt.

In seinem bekannten Buch *Don't Make Me Think* schlägt Steve Krug[41] ein einfaches und zugleich alltagsnahes Verständnis von Usability vor: Usability heißt eigentlich nur, dass etwas gut funktioniert – dass eine Person mit durchschnittlicher Begabung und Erfahrung ein Produkt für ihre Ziele nutzen kann, ohne davon frustriert zu werden. Wir wollen daher im Folgenden die Unterscheidung von Usability und User Experience so verwenden, dass Usability die Fähigkeit einer Website beschreibt, die Nutzer erfolgreich bei der Erreichung ihrer Ziele zu unterstützen; User Experience hingegen beinhaltet die weitere Perspektive und umfasst die **gesamte Interaktion** der Nutzer mit der Website, einschließlich ihrer Wahrnehmungen, Emotionen und Gedanken im Hinblick auf Design, Inhalt und Marke[42]. Diese Perspektive hat in den letzten Jahren an besonderer Bedeutung gewonnen, denn im Zuge aller gemeinhin als „Web 2.0" und „Social Commerce" benannten Phänomene ist vor allem die soziale Komponente als Teil des ganzheitlichen Mediennutzungserlebnisses in den Vordergrund gerückt. Web 2.0 steht für eine Reihe interaktiver und kollaborativer Elemente des Internets, die dem Nutzer eine neue Art und Weise ermöglichen mit dem Internet umzugehen bzw. das Internet aktiv für ihre Belange nutzen (Stichwort User Generated Content). Social Commerce ist eine Spielart des elektronischen Handels (E-Commerce), bei der die aktive Beteiligung der Kunden sowie die persönliche Kommunikation und Interaktion der Kunden untereinander im Vordergrund stehen. Es liegt auf daher auf der Hand, auch die sozialen Interaktionen der Nutzer bei der Optimierung von Webseiten systematisch zu berücksichtigen.

---

[39] DIN EN ISO 9241 (1996): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten, Berlin.

[40] Nielsen, Jacob (1993): Usability Engineering. London: AP Professional Ltd.

[41] Krug, Steve (2000): Don't make me think! A Common Sense Approach to Web Usability, Indianapolis. Seit 2006 auch als deutsche Übersetzung erhältlich: Krug, Steve (2006): Don't make me think! Web Usability – das intuitive Web, 2. Aufl., Heidelberg.

[42] Siehe Schmeißer, Daniel/Oberg, Nicole (2009): User Experience als Erfolgsfaktor im E-Commerce. In: planung & analyse, Heft 1/1009, S. 26-31.

1. Gegenstand des User Experience Tests

In einem User Experience Test steht das gesamte Erleben der Nutzer auf dem Prüfstand. Das Vorgehen lässt sich daher als ganzheitlich beschreiben, denn im Gegensatz zur Usability, die sich primär auf die Erfüllung vorab definierte Aufgaben (Use Cases) und die Untersuchung funktional-technischer Faktoren konzentriert, geht es bei der UX neben den funktionalen auch um die nicht-funktionalen (emotionalen) Faktoren. Daher kann es gelegentlich vorkommen, dass sich UX und Usability widersprechen, beispielsweise wenn bestimmte Designaspekte über ein so hohes Faszinationspotenzial verfügen, dass ihre erschwerte Nutzbarkeit in Kauf genommen wird.

User Experience ist weniger technisch orientiert und an den Systemeigenschaften interessiert, sondern fragt nach dem subjektiven Erleben der Nutzer im Umgang mit der Website.

Der User Experience Test ist daher ein Prozess, in dem unvoreingenommen und wertfrei die Meinungen, Gefühle, Gedanken und Wahrnehmungen der Nutzer zum Vorschein kommen sollten. Diesem Zugang liegt die Grundannahme zugrunde, dass Websites nur dann erfolgreich sein können, wenn sie an die expliziten und impliziten Bedürfnisse, Wünsche und Kenntnisse der Nutzer anknüpfen und Lösungen bereitstellen. Daher ist es zunächst unerheblich, ob Verbesserungsvorschläge, die aus UX-Tests resultieren, technisch umsetzbar sind oder nicht. Natürlich sollte die Machbarkeit rein aus pragmatischen Gründen nicht vernachlässigt werden – im ersten Schritt geht es aber darum, die Prozesse so zu konzipieren bzw. zu optimieren, dass sie aus Nutzersicht zu einem befriedigenden Erlebnis führen. Erst im nächsten Schritt wird geprüft, wie sich die so definierten Workflows technisch und visuell umsetzen lassen.

Und noch ein Aspekt scheint zentral: Geht es bei Usability-Optimierung vor allem um das Reduzieren von Stressfaktoren und Nutzungsbarrieren, steht bei der Optimierung der User Experience die **Herstellung möglichst positiver und uniquer Erlebensqualitäten** im Vordergrund. Diese beinhalten aber nicht alleine das Erreichen der individuellen Nutzungsziele, sondern beispielsweise auch das Erleben im Erstkontakt auf einer Landing Page, das spontane Stöbern im Shop oder die Wahrnehmung einer Marke, die im Web auf spezielle Weise inszeniert wird. Es liegt daher auf der Hand, dass der UX-Test nicht allein über das Lösen von Use Cases zu Ergebnissen kommen kann. Solche szenario-basierten Testansätze

wurden in Kapitel V vorgestellt und es wurde betont, dass funktionale Aspekte vor allem in frühen Phasen der Entwicklung durch Prototyping-Verfahren optimiert werden sollten. Voraussetzung für das Testen der vollständigen User Experience ist dagegen die volle Funktions- und Interaktionsfähigkeit der Website. Der UX-Test findet daher im Live-Betrieb einer Seite statt, entweder als Status-Quo Analyse eines bestehenden Angebots, oder unmittelbar vor bzw. nach einem Relaunch zur Feinoptimierung in der Betriebsphase.

Kriterien und Prinzipien bei UX-Tests

Bei Usability-Tests unterscheidet man gemeinhin zwischen **formativen** und **summativen Evaluationen**. Die formative Evaluation setzt entwicklungsbegleitend an und hat zum Ziel, die Usability einer Website iterativ und sukzessive zu optimieren. Die summative Evaluation beschäftigt sich dagegen mit dem „fertigen" System und hat zum Ziel, die Übereinstimmung der Nutzererfahrungen mit den intendierten Funktionalitäten zu überprüfen. Der UX-Test kann daher als Form der summativen Evaluation bezeichnet werden; Varianten des Usability-Prototypings (vgl. Kapitel VI) sind dagegen der formativen (entwicklungsbegleitenden) Evaluation zuzuordnen.

Beim Aufsetzen eines UX-Tests sollte zwischen **Leistungs- und Zufriedenheitskriterien** unterschieden werden.[43] Beide resultieren aus den Anforderungen der Nutzer an eine Website, unterscheiden sich aber hinsichtlich ihrer Erfassung und Interpretation. In der Regel hängen beide Bereiche eng miteinander zusammen, aber nicht zwangsläufig. So kann es vorkommen, dass die Performance der Nutzer bei der Zielerreichung auf einer Website unterdurchschnittlich ist (viele finden zum Beispiel ein bestimmtes Produkt erst nach umständlicher Suche), dies führt aber nicht automatisch zu einer geringeren Zufriedenheit (zum Beispiel wenn die User gar nicht bemerken, dass sie das Produkt nicht auf die „richtige" Weise gefunden haben). Es ist daher wichtig, beide Kriteriumsbereiche parallel zu erfassen.

Abbildung 65 gibt einen Überblick über die verschiedenen Leistungs- und Zufriedenheitsfaktoren der User Experience:

[43] Sowohl Leistungs- als auch Zufriedenheitskriterien sind bereits in älteren Usability-Definitionen enthalten. Im Web-Bereich ist in den letzten Jahren vermehrt der Versuch unternommen worden, Leistungskriterien medienadäquat zu operationalisieren. Dasselbe steht für die Zufriedenheitskriterien noch aus. Vgl. Zur Metrisierung von UX- Leistungskriterien: Tullis, Tim/Albert, Bill (2008): Measuring The User Experience. Collecting, Analyzing, and Presenting Usability Metrics. Elsevier Inc.

Leistungskriterien (Performance)	Zufriedenheitskriterien (Satisfaction)	
Aufgabenlösung	Ease of Use	Markenpassung
Zeit zur Aufgabenlösung	Erwartungskongruenz	Innovation
Fehlerhäufigkeit	Vertrauen und Glaubwürdigkeit	Faszinationspotenzial
Effizienz	Nützlichkeit	Emotionalität und Ästhetik
Erlernbarkeit	Joy of Use	Differenzierungsfähigkeit

Abbildung 65: Überblick über die zentralen UX-Kriterien, unterteilt nach Leistungs- und Zufriedenheitsfaktoren (Quelle: phaydon).

Leistungskriterien (Performance)

Typische Leistungskriterien leiten sich aus den Anforderungen an die Gebrauchstauglichkeit einer Website ab. Die Kriterien operationalisieren dabei die Aspekte Effizienz und Effektivität und beschreiben die Ergebnisse aus der Beobachtung von Verhaltensdaten während des Lösens vorab definierter Aufgaben (zum Beispiel Bestellung eines bestimmten Produkts oder Auffinden eines inhaltlichen Artikels; siehe zur Definition und Festlegung von Use Cases Kapitel V). Leistungskriterien haben den Vorteil, dass sie weitgehend unbeeinflusst und standardisiert sowohl in qualitativen UX-Tests als auch in quantitativen Remote-Tests zur Identifikation von Usability-Schwachstellen erhoben und quantifiziert werden können. Inzwischen existieren auch Versuche, die so gewonnenen Leistungskriterien zu metrisieren, um auf dieser Basis Wiederholungsmessungen und Benchmarks im Sinne eines UX-Controllings zu ermöglichen.[44] Folgende Leistungsparameter lassen sich unterschieden:

Aufgabenlösung. Beschreibt die **Anzahl** der Personen, denen es gelingt, die Aufgabe innerhalb einer bestimmten Zeit zu lösen. Dieses Maß wird verwendet, sobald verschiedene Nutzungsaufgaben („Use Cases" oder „Tasks") Bestandteil des Tests sind, wobei für die Interpretation auch die Aufgabenschwierigkeit mit berücksichtigt werden sollte. Zusätzlich können **unterschiedliche Erfolgslevel** unterschieden werden, zum Beispiel zwischen vollständiger Lösung, Teillösung, fehlerhafter Lösung und Abbruch. Die Abbruchquote gibt zusätzlich Auskunft über den Anteil der Personen, die die Aufgabe zu einem bestimmten Zeitpunkt von sich aus beendet haben.

[44] Vgl. Tullis, Tim/Albert, Bill (2008): Measuring The User Experience. Collecting, Analyzing, and Presenting Usability Metrics. Elsevier Inc.

Zeit zur Aufgabenlösung. Ermittelt die durchschnittliche Zeit, die Nutzer vom Start bis zur Lösung einer Aufgabe benötigen. Die Schnelligkeit, mit der bestimmte Aufgaben gelöst bzw. Ziele erreicht werden können, ist ein verlässliches Maß für die Usability einer Website. Sie variiert aber erfahrungsgemäß in hohem Maße mit den Fähigkeiten und Intentionen der Nutzer, so dass die Zeit allein als Effizienzmaß wenig tauglich ist. Darüber hinaus ist hinlänglich bekannt, dass objektive Dauer und subjektives Zeitempfinden (zum Beispiel beim Durchlaufen eines Bestellprozesses) erheblich voneinander abweichen können. So zeigt sich in Usability-Studien, dass die Zufriedenheit beim Durchlaufen eines Bestellprozesses nicht zwangsläufig von der tatsächlichen Länge des Prozesses oder der objektiven Dauer der Bestellung abhängt, sondern von der **subjektiv empfundenen Länge**. Diese ist umso kürzer, je übersichtlicher die Nutzerführung und verständlicher die Handlungshinweise sind.[45]

Fehlerhäufigkeit. Erfasst die Anzahl von Fehlern bzw. fehlerhaften Aktionen, die für die Aufgabenlösung nicht zielführend sind. Was genau als Fehler angesehen werden kann, muss dabei von Fall zu Fall entschieden werden. Hierzu kann die fehlerhafte Bedienung einer Navigation ebenso wie das fehlerhafte Eingeben in die Bestellmaske zählen. Die **Häufigkeit** der Fehler ist dabei weniger aussagekräftig, als die **Qualität** der Fehler: Handelt es sich um Fehler, die das System nahelegt oder um vermeidbare Fehler, die zum Beispiel infolge einer missverständlichen Kommunikation durch irritierende Wordings entstehen?

Effizienz. Neben der Zeit zur Aufgabenlösung lassen sich weitere Effizienzparameter bestimmen. Relevant sind in diesem Zusammenhang die Verhaltensdaten, die mittels Logfile-Aufzeichnung und Pfadanalyse gewonnen werden. Hier lassen sich die für eine Aufgabenlösung genutzte Anzahl von besuchten Seiten, Kontakthäufigkeiten, Verweildauer auf Seiten oder die Häufigkeit von Browser-Back-Bewegungen („Lostness“) quantifizieren und qualitativ ausweisen.

Erlernbarkeit. Beschreibt die Fähigkeit, ähnliche Aufgaben unter **vergleichbaren** Bedingungen in Wiederholungstests effizienter und schneller durchzuführen. Wiederholungsmessungen können innerhalb desselben Tests oder zeitlich aufeinanderfolgend vor und nach der Optimierung einer Website zur Erfolgskontrolle erfolgen.

[45] Siehe hierzu: Schmeißer, Daniel/Sauer, Michael (2005): Usability-Forschung im E-Commerce. Methoden und Ergebnisse am Beispiel des Online-Versandhandels. In: planung & analyse, Heft 1/2005.

Zufriedenheitskriterien (Satisfaction)

Bereits in älteren Usability-Definitionen geht es um den Faktor Zufriedenheit. Dieser lässt sich ausweiten auf das gesamte Erleben während der Nutzung einer Website. Zufriedenheitskriterien umfassen alles, was die User während bzw. im Anschluss an die Interaktion mit der Website sagen, denken und empfinden. Dabei muss auch berücksichtigt werden, dass Nutzer in Befragungen häufig nicht das sagen, was sie denken oder fühlen. Der Einsatz indirekter, impliziter Verfahren und Messmethoden spielt daher ebenfalls eine wichtige Rolle. Zufriedenheitsfaktoren lassen sich im Rahmen von UX-Tests in folgenden Dimensionen abbilden und für den Einsatz in Benchmarks metrisieren:

Ease of Use. Zur Erfassung des Ease of Use kommen in der Regel Post-Task Ratings zur Anwendung. Hierzu wird im Anschluss an die Lösung der Nutzungsaufgaben mittels skalierter Abfrage erfasst, wie gut die Nutzer mit der Aufgabenlösung zurechtgekommen sind, wie zufrieden sie mit der Aufgabenlösung sind und wie einfach sie die Aufgabenlösung empfunden haben. Die Abfrage erfolgt über 4-, 5- oder 6-stufige Skalen (zum Beispiel 1 = starke Zustimmung, 5 = Ablehnung). Alternativ erfolgt die Abfrage über semantische Differenziale (zum Beispiel Gegensatzpaar „leicht versus schwierig"). Die subjektiven Aussagen über die Leichtigkeit und Erlernbarkeit der Aufgabenlösung können in metrisierter Form dazu genutzt werden, zwischen (a) Nutzergruppen, (b) Aufgaben oder (c) Messzeitpunkten zu vergleichen.

Erwartungskongruenz. Alternativ zur Erfassung der Selbsteinschätzung bezüglich einer Aufgabenlösung kann auch danach gefragt werden, wie die tatsächliche Nutzung der Website gegenüber der erwarteten Nutzung erlebt wird. Auch übergreifend lässt sich mittels offener Fragen ermitteln, wie ein Angebot die Erwartungen der Nutzer erfüllt und an welchen Stellen der Nutzer enttäuscht oder frustriert wird. Für eine vertiefende Analyse der Erwartungskongruenz empfiehlt es sich, vor Durchführung von Nutzungsaufgaben die Testteilnehmer ungestützt zu ihren Erwartungen an das Angebot, die Website oder die Marke zu befragen. Die so ermittelten Aussagen können später mit den tatsächlichen Eindrücken abgeglichen werden und geben Aufschluss über Diskrepanzen im Erleben einer Website.

Vertrauen & Glaubwürdigkeit. Ein häufig vernachlässigtes Kriterium ist das Vertrauen, das der Nutzer einer Website schenkt. Vertrauen ist aber essenziell, denn ohne Vertrauen fühlen sich Nutzer nicht sicher und zögern unter Umständen, erforderliche Angaben für die Bestellung oder Transaktion zu tätigen. Dabei geht es im Wesentlichen um das Ausmaß an wahrgenommener Glaubwürdigkeit, über die eine Website verfügt. Zu Glaubwürdigkeit

und Vertrauen tragen der erste visuelle Eindruck, die allgemeine Reputation des Anbieters, aber auch das zuverlässige Feedback während der Interaktions- und Transaktionsprozesse bei.

Nützlichkeit. Das Ausmaß der subjektiv empfundenen Nützlichkeit beschreibt den Grad, in dem eine Anwendung oder eine Website für die persönlichen Belange hilfreich ist. Das betrifft sowohl die Website als Ganzes, als auch einzelne Funktionalitäten und Services. Dieser subjektiv empfundene Mehrwert muss nicht deckungsgleich sein mit den intendierten Businesszielen oder den explizit geäußerten Nutzerwünschen, sondern beschreibt eine psychologische Komponente: Produkte sind nützlich, wenn sie die für den Markt oder die Produktgruppe typischen (expliziten und impliziten) Konsumentenbedürfnisse und Konflikte aufgreifen und relevante Lösungen anbieten.

Joy of Use. Dabei handelt es sich um ein viel und kontrovers diskutiertes Konzept, das häufig missverstanden wird.[46] Es bedeutet nicht, dass eine Website unterhaltsam oder „spaßig" sein muss, sondern meint das persönliche Empfinden von Freude bei der Nutzung. Freude ist sehr subjektiv und ihr Ausmaß bei der Nutzung von Websites variiert natürlich von Person zu Person und ist schwer messbar. Es liegt aber auch auf der Hand, dass Freude bei der Nutzung wichtig ist und dazu führen kann, Fehler in der Bedienung leichter zu verzeihen. Joy of Use kann aber weder angemessen durch „hedonische Qualitäten" oder „Freude am Klicken" übersetzt werden. Vielmehr sind verschiedene Faktoren für die Freude an der Nutzung entscheidend. Hierzu gehört nicht nur eine angenehme, visuelle Erscheinung (Ästhetik), klare Inhalte und Strukturen (Prägnanz), sondern auch eine intuitive Bedienung (Flow-Erleben), Abwechslungsreichtum (Tiefenstruktur) und eine nutzerinitiierte Interaktion (Kontrolle). Freude an der Nutzung bedeutet, dass der Nutzer zu jedem Zeitpunkt das Gefühl hat, die Kontrolle zu haben, seine Ziele erreichen zu können und relevante Inhalte zu bekommen. Freude an der Nutzung heißt aber auch, dass eine Website ihre Nutzer immer wieder mit Features, Inhalten und positiven „Nutzungserlebnissen" überraschen muss.

Markenpassung. Entscheidend für das Nutzererleben ist auch die Wahrnehmung bzw. Inszenierung der Marke. Es ist bekannt, dass das Wissen über eine Marke Wahrnehmung und Verhalten maßgeblich steuert, und das gilt natürlich genauso für den Auftritt einer Marke

[46] Siehe für einen allgemeinen Überblick Reeps, Inga E. (2006): Joy-of-Use. Ästhetik, Emotion und User Experience für interaktive Produkte. VDM Verlag Dr. Müller, Saarbrücken.

im Web.[47] Passt der Auftritt nicht zur Marke, entstehen Diskrepanzen, die bewusst oder unbewusst zu unterschwelligen Zweifeln führen. Diese haben beispielsweise über die erlebte Glaubwürdigkeit eines Auftritts Auswirkungen auf die Bereitschaft von Neukunden, den Bestellprozess abzuschließen oder überhaupt Produkte im Internet zu bestellen. Ein markenkongruenter Auftritt erleichtert dagegen die Übertragung markentypischer Attribute auch auf die Angebote im Internet. Daher ist eine offen-verbale Abfrage der Markenwahrnehmung zentral für die Evaluation der User Experience. Diese sollte im ersten Schritt über eine offene und ungestützte Abfrage des Markenbildes (Markenkern und -image) sowie im zweiten Schritt über die skalierte Abfrage der Markenpassung exploriert werden.

Innovation & Faszinationspotenzial. Es gibt Websites, die über eine sehr gute Usability verfügen und technisch auf dem neuesten Stand sind, aber trotzdem nicht zur Nutzung animieren. Die User Experience umfasst daher auch das unmittelbare Faszinationspotenzial, das von einer Website bei ihrer ersten Nutzung ausgeht. Eine Website verfügt über Faszinationspotenzial, wenn sie spontan begeistert, ein besonderes Erlebnis in Aussicht stellt, gerne weiterempfohlen und wieder genutzt wird. Die Wahrnehmung von Innovation kann dabei eine wichtige Rolle spielen – sie steht für Neuartigkeit und einen spezifischen Mehrwert, der in dieser Form noch nicht umgesetzt worden ist. Faszinationspotenzial wird (qualitativ) über spontane Assoziationen oder (skaliert) über Adjektivlisten und (projektiv) über die Nutzungs- und Empfehlungsbereitschaft erfasst.

Emotionalität & Ästhetik. Die emotionale und damit implizite Wirkung einer Website wird im Rahmen klassischer Usability-Tests häufig vernachlässigt oder nebenbei mittels standardisierter Adjektivlisten erfasst. Die emotionale Wirkung einer Seite lässt sich aber weder durch eine skalierte Abfrage, noch durch Zustandsbeschreibungen angemessen beschreiben. Es erfordert unterschiedliche qualitative Ansätze, um die emotionale und ästhetische Wirkung einer Website zu ermitteln, denn es geht nicht allein um die anmutungsmäßige Beschreibung der Seite, sondern um die Ermittlung der Gefühle und Eindrücke, die hierdurch beim Nutzer ausgelöst werden. Diese ergeben sich zum einen aus den Bedürfnissen und Konflikten, die Nutzer im Umgang mit einem Angebot entwickeln, zum anderen aus der direkten Wahrnehmung impliziter Codes und Symbole, die über die Wahl der Bildwelten, Ikonografie und Produktdarstellungen beeinflusst werden.

[47] Siehe hierzu Esch, Franz-Rudolf et al. (2005): Strategische Empfehlungen für den Aufbau von Markenbekanntheit und für die Markenpositionierung im Internet. In: Esch, Franz-Rudolf: Moderne Markenführung. Grundlagen – Innovative Ansätze – Praktische Umsetzungen. Wiesbaden, S. 676-705.

Gerade die Wirkung von Key-Visuals und Teasern, aber auch die Inszenierung von Produktwelten auf Angebotsseiten wird im Hinblick auf ihre emotionale Wirkung häufig unterschätzt. Dabei handelt es sich zumeist um subjektive Einschätzungen im Sinne der visuellen Schönheit, die zumeist eine hohe emotionale – und damit implizite – Wirkung auf die Wahrnehmung einer Seite haben. Hier empfiehlt sich der Einsatz assoziativer („Bitte erzählen Sie spontan, was Ihnen beim Anblick von… durch den Kopf und Bauch ging"), projektiver (*„Wie würden Sie die Markenwelt von xy-Shop beschreiben, wenn xy ein Tier wäre?"*) und kreativer Verfahren (*„Bitte stellen Sie aus dem vorliegenden Bildmaterial zusammen, wie Sie gerne auf der Website angesprochen werden möchten"*). Diese unterlaufen die rationale Ebene und animieren die Teilnehmer dazu ihre Gefühle und Eindrücke auf indirekte Art zum Ausdruck zu bringen.

Differenzierungsfähigkeit. Eine Website kann noch so gut gestaltet und optimiert sein, sie muss sich auch ausreichend vom Wettbewerb abheben, um aufzufallen und einen echten Mehrwert zu bieten. Differenzierung kann auf unterschiedlichen Ebenen ansetzen, entscheidend ist aber die Erfassung der relevanten und **markttypischen** Differenzierungskriterien für Ihr Angebot. Das Differenzierungspotenzial einer Seite lässt sich qualitativ durch eine Kombination von Pre-Befragung zum allgemeinen Nutzungsverhalten auf Wettbewerberseiten und einer Post-Befragung zu relevanten Differenzierungsmerkmalen des eigenen Angebots in Relation zum Wettbewerb explorieren.

Fazit

Im Rahmen von UX-Tests ist es wichtig, Leistungs- und Zufriedenheitskriterien parallel zu betrachten. Denn erst die gemeinsame Analyse objektiver und subjektiver Experience-Kennwerte lässt Aussagen über das gesamte Nutzererleben – die User Experience – einer Website zu.

2. Vorbereitung von UX-Tests

Stichprobenauswahl

Bei qualitativen UX-Tests kommt der Auswahl der Stichprobe eine zentrale Rolle zu – mit ihr steht und fällt die Qualität und Aussagekraft der Ergebnisse. Bei quantitativen Stichproben ist die Grundgesamtheit in der Regel bekannt: Das Sample ist dann repräsentativ für die Grundgesamtheit, wenn alle Fälle dieselbe Chance hatten in die Stichprobe aufgenommen zu werden (Zufallsauswahl). Bei qualitativen Stichproben ist die Grundgesamtheit häufig zu Beginn der Untersuchung nicht bekannt und lässt sich erst am Ende beschrei-

ben. Qualitative UX-Studien sind daher nicht durch eine statistische, sondern durch eine **funktionale Repräsentativität** bzw. analytische Verallgemeinerbarkeit charakterisiert.[48]

Funktionale Repräsentativität ist dann gewährleistet, wenn die Stichprobengröße geeignet ist, die in der Grundgesamtheit der betroffenen Nutzer anzutreffenden Varianten von Merkmalen abzubilden. Hierfür muss die Stichprobe die Zielgruppe inhaltlich repräsentieren, das heißt die typischen Merkmale aufweisen. Das bedeutet: Es muss sichergestellt werden, dass die Fragestellung möglichst facettenreich erfasst wird (zum Beispiel durch Wahl extremer, typischer oder kritischer Fälle).

Zielgruppenbestimmung

Für die Auswahl der Stichprobe bedeutet dies, dass sich Auftraggeber und -nehmer im Vorfeld darüber Gedanken machen müssen, wer die typischen Nutzer der Website sind. Diese werden gemäß ihrer Bedeutung im Rahmen eines Quotenplans gewichtet, das heißt, der Quotenplan legt die Verteilung der Stichprobe fest. So kann zum Beispiel die Festlegung lauten, dass Bestandskunden des Online-Angebots (die bereits etwas bestellt haben) und potenzielle Neukunden (zum Beispiel wechselbereite Wettbewerbskunden) in die Studie einbezogen werden. Auch wenn beispielsweise Bestandskunden den größten Anteil in der Nutzerschaft stellten, können die potenziellen Neukunden gleichgewichtet werden, um das Optimierungs- und Nutzungspotenzial der Website angemessen einschätzen zu können. Daneben kann zum Beispiel die Häufigkeit der Internetnutzung (Heavy- versus Light-Nutzer) oder verschiedene Altersgruppen paritätisch berücksichtigt werden. Aus diesen Vorgaben ergeben sich also vollständige Quotenpläne, welche die Verteilung über die Stichprobenmerkmale im Vorhinein festlegen (vgl. Abbildung 66).

N=32										Gesamt
Geschlecht		männlich				weiblich				
Kenner der Website		Nutzer		Nichtnutzer		Nutzer		Nichtnutzer		
Nationalität		D	GB	D	GB	D	GB	D	GB	
Alter	20-29	2	2	2	2	2	2	2	2	16
	30-39	2	2	2	2	2	2	2	2	16
		4	4	4	4	4	4	4	4	
		8		8		8		8		
Gesamt		16				16				32

Abbildung 66: Beispiel für einen qualitativen Quotenplan für zwei Länder mit insgesamt 32 Untersuchungsteilnehmern, unterteilt nach Alter, Geschlecht und Websitenutzung. Pro Land ergeben sich zusammengefasste Quoten (Zellbelegungen) für das Geschlecht, Nutzer versus Nicht-Nutzer und zwei Altersgruppen mit jeweils 8 Personen pro Split.

[48] Vgl. für einen Überblick über Stichprobenverfahren: Schreier, M. (2007): Qualitative Stichprobenkonzepte. In: Naderer, G./Balzer, E. (Hrsg.): Qualitative Marktforschung in Theorie und Praxis, Gabler. S. 232-245.

Die Rekrutierung der Testteilnehmer erfolgt dann auf Grundlage eines Rekrutierungsfragebogens, der neben den Quotenmerkmalen auch allgemeine soziodemografische Merkmale (Einkommen, Haushaltsgröße etc.) und medienrelevante Variablen (Internetzugang, Mediennutzung etc.) erhebt.

Stichprobenziehung

Bei der Stichprobenziehung können Sie zum einen auf die unmittelbar verfügbaren Nutzer oder Befragungswillige zugreifen (man spricht von einer „anfallenden Stichprobe" oder einer „Stichprobe aufs Geratewohl", zum Beispiel wenn sie in der Abteilung oder unter Studenten fragen, wer gerade Zeit und Lust hat am Test teilzunehmen), oder Sie können eine sog. „absichtsvolle" Stichprobe ziehen, was in der Regel über ein professionelles Feldinstitut erfolgt. Aus methodischer Sicht ist eine absichtsvolle qualitative Stichprobenziehung immer einer anfallenden Stichprobe vorzuziehen, denn nur so können Sie die Merkmale der Personen kontrollieren, die in die Untersuchung eingehen. Gerade bei Tests mit geringer Anzahl an Befragungspersonen sollten sie vermeiden, Extremmerkmale zu stark zu gewichten, da dies natürlich Auswirkungen auf die Ergebnisse haben wird (zum Beispiel nur Heavy-User oder Kenner eines Angebots).

Bei den qualitativen Stichprobenplänen geht man in der Regel theoriegeleitet vor, das heißt man verfügt vor Beginn der Rekrutierung über Vorannahmen bezüglich der relevanten Merkmale in der Zielgruppe. Eine erste Festlegung erfolgt im nächsten Schritt darüber, ob es sich um einen homogene oder heterogene Stichprobe handeln soll. Homogene Stichproben bestehen aus einander ähnlichen Fällen (zum Beispiel early adopter); heterogene Stichproben decken hingegen die gesamte Bandbreite eines Phänomens ab (zum Beispiel zufriedene Kunden, wechselwillige Kunden, Nicht-Kunden).

Grundsätzlich sind fünf Schritte bei der Erstellung eines qualitativen Stichprobenplans zu beachten[49]:

1. Geltungsbereich der Untersuchung festlegen (zum Beispiel alle Internetnutzer, die online Reisen buchen)
2. Relevante Merkmale im Gegenstandsbereich identifizieren (Altersrange; Internet-Affinität etc.)
3. Festlegung der Merkmalskombinationen„Zellen" des Plans (Erstellung sogenannter Stichprobenpläne)

[49] Schreier, M. (2007): Qualitative Stichprobenkonzepte. In: Naderer, G./Balzer, E. (Hrsg.): Qualitative Marktforschung in Theorie und Praxis, Gabler, S. 239.

4. Bestimmung der Anzahl Fälle pro Merkmalskombination (in der Regel mindestens 5-6 Fälle pro Zelle; Erstellung eines Quotenplans)
5. Fallauswahl (Stichprobenziehung)

Stichprobengröße

In der Usability-Literatur finden sich auch heute noch Hinweise, dass fünf bis acht Personen ausreichten, um 80 Prozent der Usability-Fehler aufzudecken.[50] In der Praxis zeigt sich aber immer wieder, dass der real gefundene Anteil der Benutzungsprobleme mitunter deutlich davon abweichen kann und auch die Bedeutung einzelner Schwachstellen erst im Nachhinein angemessen eingeschätzt werden kann. Wenn Sie trotzdem mit 20 Prozent Fehlern leben können (von denen Sie nicht wissen, wie gravierend sich diese in der gesamten Nutzerschaft auswirken) und Ihr Budget nicht mehr Probanden zulässt, ist dies sicherlich ein zulässiges Vorgehen. Was Sie damit nicht erreichen, sind verlässliche Aussagen über die Akzeptanz, Wirkung, Wahrnehmung und das Nutzungserleben Ihrer Website.

Bei qualitativen UX-Tests, in denen es neben der Usability auch um Brand-Fit, Tonalität und andere Erlebensdimensionen geht, empfiehlt es sich daher, Tests an den Vorgaben der qualitativen Methodologie auszurichten und **mindestens 16 bis 20 Personen** einzubeziehen. Eine genaue Festlegung der Stichprobengröße lässt sich aber nicht allgemeingültig ad-hoc bestimmen und hängt von der Heterogenität der Zielgruppe, der Methodik, der Komplexität der Fragestellung und nicht zuletzt von Ihrem Budget ab:

- Ist die Stichprobe sehr homogen und die Fragestellung klar zugespitzt auf die Ermittlung funktionaler Schwachstellen und Optimierungspotenziale in der Benutzerführung, genügen in der Regel 10 bis 12 Testteilnehmer.
- Werden Verhaltensbeobachtungen (zum Beispiel Pfaddaten oder Blickbewegungen) erfasst und aggregiert ausgewertet, sollte die Stichprobengröße mind. 16 bis 20 Personen umfassen, da die Variabilität von Verhaltens- und Messdaten in der Regel sehr hoch ist und sich daher kleine Stichproben verzerrend auf die Ergebnisse auswirken können.
- Ist die Stichprobe sehr heterogen, werden also im Quotenplan sehr unterschiedliche Merkmalskombinationen berücksichtigt und im Ergebnisbericht gegebenenfalls miteinander verglichen, empfiehlt sich als Faustregel mindestens 5 bis 6 Personen pro Untersuchungszelle. In der Regel vermeidet man in qualitativen Designs Stichprobenpläne mit mehr als vier Hauptzellen (zum Beispiel Kunde versus Nicht-Kunde; Frauen versus

[50] Ursprünglich geht diese Doktrin auf J. Nielsen zurück, der in diesem Punkt aufgrund seines Expertenstatus mit einiger Vehemenz die Erkenntnisse der qualitativen Methodenlehre ignoriert. Vgl. Nielsen, Jacob (1993): Usability Engineering. London: AP Professional Ltd.

Männer), diese können innerhalb der Zellen aber über weitere Merkmale (zum Beispiel Alter und Internetaffinität) verteilt werden. Im Falle eines Vier-Zellen-Plans würde man also mit circa 20 bis 24 Personen testen.

Auswahl von Testszenarien

Die Auswahl der geeigneten Testszenarien ist ebenso wichtig wie die Wahl der Stichprobe. Testszenarien müssen nicht identisch mit Use Cases (Nutzungszielen) sein, werden aber in der Praxis häufig mit solchen gleichgesetzt (vgl. Kapitel V). In die Planung der Testszenarien sollten grundsätzlich drei Faktoren einfließen:

Nutzungsziele (Use Cases). UX-Tests mit Fokus auf Usability operationalisieren Testszenarien durch die Formulierung von Aufgaben, die der Nutzer in der Testsituation lösen muss. Durch Verhaltensbeobachtung und die Erfassung der Leistungsparameter wird sichtbar, wie der Nutzer mit der Website umgeht, um sein Ziel zu erreichen (zum Beispiel Bestellung eines bestimmten Produkts) und welche Stolpersteine dabei auftreten. Bei der Formulierung der Use Cases sollten sowohl möglichst typische Aufgaben gewählt werden, aber auch solche Aufgaben, die besonders kritische Aspekte abdecken und unter Umständen nur selten auftreten. Diese können auch innerhalb einer Stichprobe variieren (zum Beispiel ändert der Kunde auf der Website bestehende Vertragsbestandteile, während der Neukunde einen Vertrag abschließt). In der Regel werden auf diese Weise 4 bis maximal 8 Use Cases definiert und – um Reihenfolge- und Lerneffekte zu vermeiden – in rotierter Abfolge von den Befragten durchgeführt. Zur Auswahl typischer Use Cases ist eine Anforderungsanalyse zu Beginn der Untersuchung unumgänglich.

Nutzungskontext. Die Wahl der Testszenarien resultiert auch aus den typischen Verwendungskontexten, in denen eine Website genutzt wird. Dies hängt weniger damit zusammen, ob ein UX-Test im Studio, bei den Befragten zu Hause oder per Remote-Zugang am eigenen PC durchgeführt wird. Entscheidend sind die Annahmen darüber, wie eine Website von den Usern genutzt wird.[51] So kann es sinnvoll sein, Testszenarien einzubauen, die den Zugang über eine Suchmaschine (zum Beispiel Google) oder über ein Werbemittel (vermittelt über eine Landing Page) abbilden. Alternative Testszenarien berücksichtigen die verschiedenen Verwendungsanlässe und -zusammenhänge systematisch im Rahmen des Studiendesigns.

[51] Letztlich sind sich alle Befragten im Rahmen eines UX-Tests – sei es im Studio oder per Remote-Test am heimischen Computer – bewusst, dass sie für bestimmte Untersuchungszwecke befragt werden. Dieses Wissen steht einer „natürlichen" Nutzung häufig per se im Weg, weshalb es im Wesentlichen auf eine gute Instruktion und gelungene Interviewführung ankommt, um mögliche Ergebnisverzerrungen zu vermeiden.

Nutzereigenschaften. Bei der Festlegung der Testszenarien sollten auch die spezifischen Eigenschaften der Nutzer bzw. Zielgruppen Berücksichtigung finden. So sollten zum Beispiel Wettbewerbskunden nicht unmittelbar dazu veranlasst werden, sich mit der aus ihrer Sicht fremden Konkurrenzseite zu beschäftigen. Stattdessen kann über eine Warm-Up-Phase sichergestellt werden, dass mögliche Nutzungshemmnisse bezüglich der zu untersuchenden Website bereits im Vorfeld angesprochen werden und sich nicht aus Gründen der Reaktanz auf die nachfolgende Nutzung der Seite niederschlagen.

Die Definition der Testszenarien erfolgt im Rahmen der Leitfadenerstellung. Hier wird der grobe Ablauf eines UX-Tests festgelegt, die Reihenfolge der Use Cases bestimmt sowie offene und geschlossene Fragetypen in teil-standardisierter Art formuliert.

Arten von Studiendesigns und -zielen

Bevor wir im nächsten Teil auf die Methoden zur Durchführung von UX-Tests eingehen, ist es zunächst wichtig zu unterscheiden, welche Ziele ein Test verfolgt. Denn je nach Zielsetzung und Fragestellung variieren auch Studiendesign und zum Einsatz kommende Verfahren. Es lassen sich folgende Arten von UX-Tests mit jeweils unterschiedlichem Fokus für die Optimierung unterscheiden:

Innovationstest (Produktneueinführung). Wird eine Website erstmalig gelauncht (Markteintritt), sollten möglichst sämtliche Facetten der User Experience berücksichtigt werden. Für einen Erfolg am Markt sind vor allem die Markenwahrnehmung und die Differenzierung vom Wettbewerb entscheidend. Daher müssen auch die relevanten und intendierten Zielgruppen ausreichend in der Stichprobe abgebildet werden, das heißt es empfiehlt sich, zur Abschätzung der Akzeptanz und des Nutzererlebens mit größeren Stichproben arbeiten und einen möglichst breiten Methoden-Mix bestehend aus qualitativen Interviews, Fragebögen, Verhaltens- und Interaktionsbeobachtung sowie Blickregistrierung einzusetzen.

Vergleichstest (Zwei-Varianten-Test). Beim Vergleichstest geht es in der Praxis häufig um die Frage, ob eine neue Version der Website eine echte Verbesserung gegenüber der alten bzw. aktuellen Seite darstellt. Zu diesem Zweck müssen beide Seiten unter vergleichbaren Bedingungen, das heißt mit denselben Testszenarien und mit jeweils strukturell ähnlichen Stichproben getestet werden. Ein sogenanntes „monadisches" Studiendesign (jede Testperson wird nur einer Untersuchungsbedingung zugeordnet) stellt die gleichwertige Berücksichtigung beider Seiten-Varianten her, indem die Stichprobe in zwei Teile geteilt wird: Die eine Hälfte exploriert die eine Variante, die zweite Hälfte die andere. Aus Kostengründen werden häufig semi-monadische Designs angewendet: Hier werden alle

Testteilnehmer zu beiden Seiten befragt, allerdings in rotierter Reihenfolge (während die eine Hälfte mit Variante A beginnt, startet die andere mit Variante B; abschließend bietet sich ein direkter Vergleich der User Experience auf beiden Varianten an).

Wiederholungstest (Pre-Post-Messung). Veränderungen in der User Experience lassen sich durch Wiederholungstest im zeitlichen Verlauf darstellen. Da sich nicht nur die Websites selbst, sondern auch Kontextbedingungen und Nutzerverhalten mit der Zeit ändern, bieten sich Wiederholungsmessungen in regelmäßigen Abständen (mind. einmal pro Jahr) an, um Veränderungen im Nutzererleben und Effekte vorgenommener Optimierung abschätzen zu können. Wichtig ist auch hier, dass ähnlich wie im Vergleichstest entsprechende Testszenarien und strukturgleiche Stichproben zur Anwendung kommen.

Benchmarktest (Wettbewerbsvergleich). Beim Benchmarktest geht es um den Vergleich der User Experience auf der eigenen Website mit derjenigen auf ausgewählten Wettbewerberseiten. Im Rahmen qualitativer UX-Tests werden aus praktischen Gründen häufig die zwei bis drei wichtigsten Wettbewerber in den Test einbezogen und mit denselben Testszenarien untersucht. Der Testansatz ist entweder monadisch (jede Testperson wird mit einer Website konfrontiert) oder semi-monadisch (jede Testperson wird mit ausgewählten oder allen Webseiten in rotierter Reihenfolge konfrontiert). Über die standardisierte Abfrage und Erfassung vorab definierter Metriken (Leistungs- und Zufriedenheitskriterien) können Sie ihr Angebot direkt mit dem Wettbewerb vergleichen und gezielt eine marktnahe Optimierung vornehmen.

Status Quo-Potenzialanalyse. Natürlich können Sie einen UX-Test auch zur Überprüfung und Optimierung des Status Quos Ihrer Website durchführen. Hier geht es um eine möglichst umfassende Bestandsaufnahme der Website aus Nutzersicht, um aktuelle Optimierungspotenziale zu identifizieren oder eine Potenzialabschätzung für einen Rebrush oder Relaunch abzuleiten.

3. Durchführung von UX-Tests

UX-Tests lassen sich mit mehr oder weniger großem Aufwand betreiben. Die Vorbereitung ist aber zeit- und kostenintensiv, allein wenn man an die externe Rekrutierung und Incentivierung geeigneter Testteilnehmer denkt. Die Durchführung selbst erfordert wiederum personelle Ressourcen, sowohl was die Datenerhebung als auch die Auswertung und Analyse der Daten betrifft. Gerade die Analyse der Daten muss durch erfahrenes Personal

erfolgen, so dass UX-Tests in der Praxis entweder von speziell für diese Zwecke eingerichteten UX-Abteilungen im Unternehmen oder von professionellen UX-Forschungsinstituten durchgeführt werden.

Die im Rahmen von UX-Tests zur Anwendung kommenden Methoden und Instrumentarien sind sehr vielfältig, wobei sich grundsätzlich alle Forschungsmethoden aus der psychologischen Markt- und Medienforschung einsetzen lassen. Es gibt dennoch besondere Anforderungen an die Erhebung von UX-Daten, und es haben sich in der Praxis eine Reihe typischer Verfahren zur Durchführung durchgesetzt, die im Folgenden skizziert werden. Dabei lassen sich die verbalen Erhebungsmethoden (Interviews, Workshops etc.) von den Beobachtungsverfahren unterscheiden. Zu den letztgenannten gehören sowohl die Aufzeichnung des Nutzungsverhaltens (zum Beispiel Logfiles) als auch die Verhaltensbeobachtung via Videoaufzeichnung sowie die Erfassung der Blickbewegungen mittels Eyetracking (siehe Kapitel XI).

Formen von Interviews

Die Grundlage professionell durchgeführter UX-Tests sind qualitative (fokussierte) Interviews, in denen die Testteilnehmer anhand eines teilstandardisierten Leitfadens von einem erfahrenen und psychologisch geschulten Interviewer befragt werden. Qualitative Interviews zeichnen sich durch eine weiche Gesprächsführung aus, die den Befragten nach Möglichkeit nicht in seinem Antwortverhalten beeinflusst, die Spezifität der Sichtweise und Situationsdefinition aus deren Sicht berücksichtigt und ein möglichst breites Spektrum der subjektiven Eindrücke in Bezug auf die Nutzung einer Website erfasst. ‚Fokussiert' sind solche Interviews insofern, als das Gespräch mittels eines strukturierten Leitfadens durchgeführt wird. Der Leitfaden legt die Reihenfolge der Themengebiete und Use Cases fest, sieht bestimmte Fragetypen vor, lässt aber ausreichend Raum für situationsabhängige Abweichungen und eigene Formulierung von Nachfragen. So wechseln sich in einem UX-Interview Phasen offen-explorativer Fragen (in denen es um Markenwahrnehmung und subjektive Erlebensbeschreibungen geht) mit Phasen strukturierter Item-Abfragen (zum Beispiel Erfassung zur Aufgabenlösung) ab. UX-Interviews haben eine Länge von 60 bis maximal 120 Minuten.

Beispielhafter Ablauf eines UX-Interviews

1. Warm-Up: Einleitende Fragen zu Erfahrungen und Erwartungen an die Website
2. Anwendung von drei bis vier typischen Use Cases mit Verhaltensbeobachtung
3. Exploration jedes Use Cases mittels datengestützter Exploration
4. Nachexploration hinsichtlich Gesamtanmutung, Joy of Use, Likeability etc.

Ein häufig eingesetztes Verfahren zur Aufdeckung von Usability-Problemen ist die **Methode des lauten Denkens**. Diese kann zum Beispiel auch als Teil eines längeren qualitativen Interviews zur Anwendung kommen, wenn es darum geht, die Funktionalität einer Website tiefer zu beleuchten bzw. wenn es sich um den Test eines noch nicht voll funktionsfähigen Prototyps handelt (vgl. Kapitel VI). Dabei werden die Testteilnehmer während sie die Use Cases lösen dazu aufgefordert, möglichst alle Gefühle und Gedanken während der Aufgabenbearbeitung zu äußern (*„Was geht Ihnen gerade durch den Kopf?“, „Was denken Sie, was passieren wird?“*). Da sich auf diese Weise sehr schnell und ohne viel Aufwand vielfältige qualitative Informationen sammeln lassen, wird die Methode des lauten Denkens häufig eingesetzt. Sie hat aber eine Reihe entscheidender Nachteile.

Neben dem Effekt, dass das laute Denken alle Erhebungsdimensionen unterläuft, bei denen eine kritische Distanz gerade nicht erwünscht ist, also Spontanität, Intuitivität, Flow-Erleben, Immersion etc., besteht der zentrale Nachteil dieser Methode darin, dass sie sich nicht mit der prozessorientierten Performance-Messung verträgt. Durch die Doppelbelastung der Nutzer, die zeitgleich die Aufgabe lösen und ihre Gedanken in Worte fassen müssen, wird nicht nur die Bearbeitungsgeschwindigkeit gesenkt. Das laute Denken hat auch einen „sokratischen Effekt“, denn der Befragte kann sich bereits während der Nutzung infolge wiederholten Nachfragens und Verbalisierens über Fehler und Inkonsistenzen bewusst werden, die entweder eine weitere Belastung darstellen oder dazu führen, dass weitere oder andere Fehler gemacht werden, die in der natürlichen Situation nicht in dieser Form oder Häufigkeit auftreten würden.

Der Nutzer gerät darüber hinaus in eine Rechtfertigungssituation, wenn die Aufgabenbewältigung sich als schwierig erweist und neigt zu rationalen, defensiven Begründungen. Ein weiterer Nachteil besteht – neben der generell ungewohnten Situation laut zu denken – in der unterschiedlichen Fähigkeit der Probanden, ihre Gedanken und Gefühle surfbegleitend in Worte zu fassen. Das liegt zum einen an personenbezogenen Fähigkeiten (Frauen fällt das laute Denken in der Regel leichter als Männern), aber auch generell an der

Tatsache, dass Gedanken und Gefühle häufig komplexer sind, als sie sich in dem Moment spontan ausdrücken lassen. Durch Anwendung verbaler Verstärker bzw. durch wiederholte Aufforderung, seine Gedanken zu verbalisieren, erhofft man sich ein spontaneres und valideres Feedback der Probanden.[52]

Eine Alternative zum lauten Denken sind datengestützte Interviews.[53] Dabei lösen die Testteilnehmer die Nutzungsaufgaben jeweils im Anschluss an eine detaillierte Instruktion durch den Interviewer alleine am PC. Der Interviewer beobachtet auf einem Kontrollmonitor das Surfverhalten des Probanden und macht sich Notizen zu einzelnen Surfsequenzen. Es wird auf jede Hilfestellung verzichtet, solange nicht technische oder das Verständnis betreffende Schwierigkeiten auftreten, die eine Bearbeitung der Aufgabe grundsätzlich verhindern. Während der Aufgabenlösung wird der gesamte Nutzungsprozess mittels Screencapture-Software aufgezeichnet. Die in Echtzeit aufgezeichneten Navigationsprotokolle werden als Videodatei abgespielt und dienen als Grundlage für die sich daran anschließenden Explorationen, in denen die Testpersonen anhand der Aufzeichnungen ihres eigenen Surfverhaltens gezielt und mit Blick auf Orientierungs- und Verständnisprobleme nachbefragt werden (Videokonfrontation). Parallel hierzu hat der Auftraggeber bzw. die Projektbeteiligten in professionell eingerichteten Use-Labs die Möglichkeit, hinter einem Einwegspiegel beobachtend an den Interviews teilzunehmen (siehe Abbildung 67).

Bei den datengestützten Einzelinterviews handelt es sich um eine Weiterentwicklung des lauten Denkens, die durch die Aufzeichnung von Videoprotokollen und die höheren Anforderungen an die Interviewführung aufwändiger ist. Sie weist aber neben der natürlichen und daher extern valideren Untersuchungssituation (wer von Ihnen denkt laut beim Surfen?) klare Vorteile gegenüber dem lauten Denken auf: Datengestützte Interviews sind nicht auf die rasche Verbalisierung während der Aufgabendurchführung angewiesen, vermeiden Blockaden und Rationalisierungen bzw. können diese gezielt im Rahmen eines offenen Gesprächs hinterfragen. Der entscheidende Vorteil liegt aber darin, dass die Messung von Verhaltens- und Prozessdaten unabhängig von ihrer verbalen Bewertung vorgenommen werden kann. Auf diese Weise erhalten Sie Beobachtungsdaten, die nicht durch den Zwang zur permanenten Reflexion beeinflusst sind und sich mit Blick auf Performance-Metriken valide auswerten lassen.

---

[52] Siehe zum Einsatz des lauten Denkens: Yom, Miriam/Wilhelm, Thorsten H./Gauert, Stefanie (2007): Protokolle lauten Denkens und Site Covering. Eine Erweiterung der Methode zur detaillierten Bewertung des Screendesigns von Webangeboten. In: Buber, Renate/Holzmüller, Hartmut H. (Hrsg.): Qualitative Marktforschung. Konzepte – Methoden – Analysen, Wiesbaden. S. 637-652.

[53] Vgl. Schmeißer, Daniel/Sauer, Michael (2005): Usability-Forschung im E-Commerce. Methoden und Ergebnisse am Beispiel des Online-Versandhandels. In: planung & analyse, Heft 1/2005.

Abbildung 67: Datengestützte Einzelexploration im Usability-Lab. Im Hintergrund der Einwegspiegel, durch den Kunden und Entwickler die Interviews ungestört beobachten können (Quelle: phaydon).

Teilnehmer im Test neigen natürlich immer dazu, sich als Experten eines Problems zu sehen. Dies führt in der Praxis zu Rationalisierungen und Pseudo-Attributionen, teils auch zu unrealistischen Wünschen und Vorschlägen der Nutzer bezüglich der möglichen Optimierung einer Website. Es ist daher wichtig, die Interviews nur als eine mögliche Datenquelle zu betrachten und diese im Zusammenspiel mit indirekten Verfahren zu interpretieren.

Analyse von Verhaltensdaten

Verhaltens- und Beobachtungsdaten geben indirekt Aufschluss über die Erlebnisse und Schwierigkeiten der Nutzer im Umgang mit einer Website. Sie liefern aber auch wichtige Hinweise auf **habitualisierte** Verhaltensweisen, die den Nutzern möglicherweise nicht bewusst sind, das Nutzungsverhalten aber entscheidend prägen. Ihre Interpretation sollte dabei in Zusammenspiel mit Befragungsdaten erfolgen, damit Prozess- und Erlebensebene miteinander abgeglichen werden können. Dabei können durchaus Diskrepanzen auftauchen, zum Beispiel wenn sich Nutzer überschwänglich zufrieden äußern, ihre Surfprotokolle und die Beobachtungen des Interviewers aber belegen, dass die Aufgabenziele tatsächlich nicht erreicht wurden oder umgekehrt die Aufgaben gut gelöst werden, die Befragten aber aus anderen Gründen kritisch reagieren.

Surfpfadanalyse

Nur Produkte und Inhalte, die leicht auffindbar sind, werden auch genutzt – klare Navigationspfade und eine gute Navigationsstruktur sind daher wichtige Erfolgsfaktoren für eine positive User Experience. Welche Wege die Nutzer wählen, von wo sie kommen und wo sie gegebenenfalls abbrechen, wird mittels Interaktionsprotokollierung festgehalten. Eine gängige Form der Verhaltensbeobachtung im Rahmen qualitativer UX-Tests ist daher die Aufzeichnung der **Nutzungsprotokolle** und Logfiles während der Aufgabenbearbeitung bzw. Nutzung der Website mittels „Webrecording". Entsprechende Software-Lösungen (zum Beispiel Morae der Firma TechSmith) zeichnen dabei den gesamten Bildschirminhalt inklusive Eingabeprotokolle und dazugehöriger Zeitwerte auf, so dass im Anschluss mit entsprechenden Analysetools nachvollzogen und visualisiert werden kann, wie die Nutzer bei der Lösung ihrer Nutzungsziele (Use Cases) vorgegangen sind. Werden die Blickbewegungen während der Aufgabenlösung mittels Eyetracker erfasst, liegen die Nutzungsprotokolle inklusive Zeitangaben, Klicks und Scrolling-Aktivitäten automatisch vor und können in entsprechende Analyseprogramme exportiert werden.

Inzwischen haben sich für die qualitative UX-Forschung Verfahren etabliert, die solche Verlaufsdaten in Form von dynamisch visualisierten Navigationspfaden anschaulich darstellen. Diese Surfpfadanalysen dürfen nicht mit herkömmlichen Darstellungen aus Web-Analytics-Tools verwechselt werden, obwohl die Funktionsprinzipien ähnlich sind. Der entscheidende Unterschied: Mittels Surfpfadanalyse sollen typische Navigationspfade dargestellt und analysiert werden, die Nutzer bei der Verfolgung ihrer Nutzungsziele unter kontrollierten Bedingungen wählen. Die Analyse erfolgt daher über die Darstellung hierarchischer Navigationsebenen, das heißt zur besseren Nachvollziehbarkeit der Verlaufsdaten werden die aufgesuchten Seiten nach Navigationsebenen (1. Ebene: Startseite; 2. Ebene: Übersichtsseiten; 3. Ebene: Produktdetailseiten oder Ähnliches) angeordnet. Auf aggregierter Ebene werden die typischen Surfpfade, die Verweildauer auf den Seiten sowie Kontakthäufigkeiten und Browser-Back-Bewegungen dargestellt. Je nach Fragestellung kann auch der durch den Nutzer induzierte Hierarchiebaum mit dem hierarchisch intendierten gegenübergestellt und abgeglichen werden, um auf mögliche strukturelle Schwachstellen aufmerksam zu machen.

Abbildung 69 zeigt die aggregierten Surfpfade von 20 Nutzern, die im Rahmen einer Nutzungsaufgabe auf neckermann-reisen.de nach einem für sie passenden Hotelangebot gesucht haben.

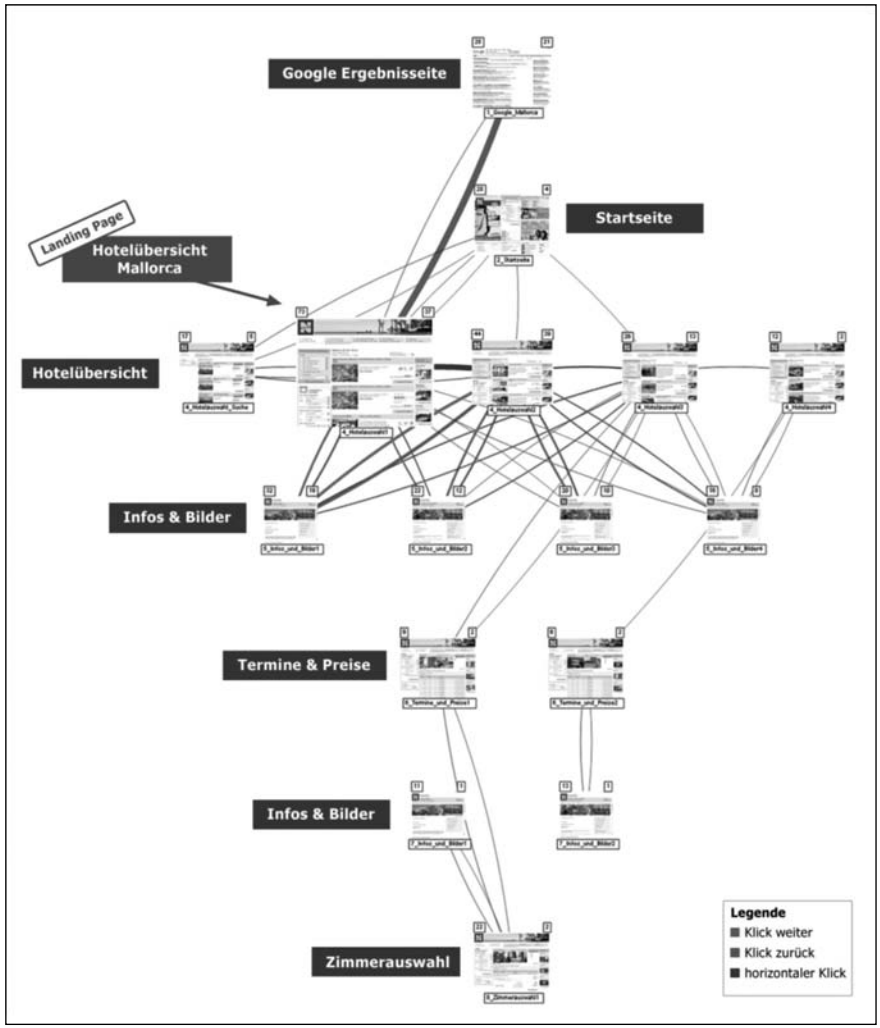


Abbildung 68: Surfpfad auf der Neckermann-Reisen Website, Erläuterung im Text. Vgl. auch die Case-Study Thomas-Cook AG, S. 319

Die Darstellung erfolgt von oben nach unten über die verschiedenen Navigationsebenen hinweg. Die Größe der abgebildeten Screenshots stellt die durchschnittliche **Verweildauer** auf der jeweiligen Seite dar, zusätzlich können Angaben über die **Kontakthäufigkeit** eingeblendet werden. Die grünen Linien stehen für Abwärtsbewegungen in der Hierarchieebene, die roten Linien für Aufwärtsbewegungen, die blauen für horizontale Klicks innerhalb einer Ebene. Die Dicke der Linien stellt die aggregierte Kontaktzahl dar, das heißt je breiter der Navigationspfad, desto mehr Nutzer haben diesen Weg gewählt. Mit Surfpfadanalysen können „Autobahnen", Sackgassen und diejenigen Stellen identifiziert werden, an denen die Nutzer „hängenbleiben", kehrt machen oder mehr Zeit als nötig verbringen.

Die meisten Analyseprogramme erlauben darüber hinaus einen Split nach unterschiedlichen Nutzergruppen und Seitenebenen. So können gezielt (strukturelle) Schwachstellen in der Nutzerführung aufgedeckt und für die Optimierung ganzer Websites ebenso wie für einzelne Produktauswahl- und Bestellprozesse genutzt werden. Durch die Visualisierung, Quantifizierung und problemzentrierte Analyse der Surfpfade anhand wichtiger Kennzahlen wie durchschnittlicher Verweildauer pro Seite oder Anzahl der Seitenaufrufe, eignet sich die Surfpfadanalyse ideal zum Vergleich verschiedener Zielgruppen (zum Beispiel Bestandskunden verus Neukunden) sowie zur Erfolgskontrolle bei Wiederholungsmessungen.

Videoaufzeichnung

Ergänzend zur Surfpfadanalyse werden häufig **Video-Interaktions-Protokolle** angefertigt. Dabei wird die gesamte Interaktion des Nutzers mit der Website während der Aufgabenlösung und des Interviews mit einer oder mehreren Kameras aufgezeichnet. Zur Erfassung der Gesprächssituation (Mimik, Gestik etc.) werden frontale Aufnahmen gewählt, zur Aufzeichnung der Interaktion mit der Website kommt die „over-the-shoulder" Perspektive zum Einsatz, die zeitgleich auch die Mausaktionen und den Bildschirminhalt erfasst. In der Regel dienen Videoprotokolle vor allem der Dokumentation, in seltenen Fällen werden diese aber nach festgelegten Regeln (zum Beispiel mittels Video-Interaktionsanalyse)[54] systematisch ausgewertet.

Für eine systematische Auswertung lohnt sich der sehr hohe Aufwand meist nicht, darüber hinaus werden wichtige Beobachtungen bereits durch den Interviewer oder Beobachter hinter der Einwegscheibe (im Studio) festgehalten. Die Videoprotokolle können aber nachträglich vor allem mit Blick auf besonders kritische Stellen analysiert werden – Critical-Incidants-Analyse – oder aber sie dienen als Ausgangsmaterial für Best-of-Videos, in denen

[54] Knoblauch, Hubert (2004): Die Video-Interaktions-Analyse. In: sozialersinn, Heft 1/2004, S. 123-138.

ausgewählte und besonders aussagekräftige Sequenzen zur Illustration der Berichtsergebnisse in einem anschaulichen und kommentierten Kurzfilm zusammengeschnitten werden. Best-of-Videos haben vor allem eine didaktische Funktion, indem sie bei Präsentationen vor Publikum, das keine Gelegenheit hatte einen „echten" Nutzertest zu beobachten, die Nutzungssituation durch O-Töne der Befragten veranschaulichen. Vor allem bei Entscheidern, die häufig einen geringeren Bezug zu ihrer Zielgruppe haben, fördert dies die Akzeptanz und das Verständnis der Ergebnisse eines UX-Tests.

Ein weiteres und zunehmend häufiger eingesetztes Verfahren ist die Aufmerksamkeitsmessung mittels Blickregistrierung. Da es sich um ein wichtiges und zugleich sehr komplexes Verfahren handelt, wird die Blickregistrierung in einem eigenen Kapitel behandelt (siehe Kapitel XI).

Nutzerbefragungen

Die Befragung der Nutzer im Rahmen eines UX-Tests erfolgt in der Praxis sehr anwendungsbezogen und auf die jeweilige Fragestellung zugeschnitten. Neben den teilstandardisierten und ad-hoc entworfenen Fragebögen empfiehlt es sich jedoch – auch zum Zweck späterer Vergleichbarkeit bei Wiederholungstests – zusätzlich mit standardisierten (und im Idealfall normierten) Fragebogenteilen zu arbeiten. Diese erfassen bestimmte Leistungskriterien und Einstellungsdimensionen in standardisierter Form und gewährleisten so eine hohe Validität und Reliabilität (Zuverlässigkeit) der Ergebnisse. Häufig werden Fragebögen eingesetzt, die sich aus der DIN EN ISO 9241-11 ableiten. Problematisch ist allerdings, dass nicht alle Dialogprinzipien (Erlernbarkeit, Aufgabenangemessenheit etc.) auf die Nutzung von Websites anwendbar sind bzw. nur in abgewandelter Form für die standardisierte Beurteilung ausgewählter Funktionsbereiche genutzt werden können (zum Beispiel für die Evaluation von Bestell- und Check-Out-Prozessen bei E-Commerce-Angeboten). Auch andere Verfahren stammen aus dem Bereich der Software-Entwicklung und eignen sich nur bedingt für die Anwendung auf Websites (QUIS: Questionnaire for User Interface Satisfaction; SUMI: Software Usability Measurement Inventory; IsoMetrics; EU-CON II; IfADo-Fragebogen zur Bewertung von Software).

Während für den Bereich Software-Ergonomie eine ganze Reihe validierter Instrumente vorliegt, existieren für die Website-Evaluation nur wenige und meist adaptierte Verfahren.

Ein Befragungsinstrument, das sich explizit mit dem Konzept der User Experience auseinandersetzt, ist der AttrakDiff2-Fragebogen.[55] Er besteht aus 28 bipolaren, siebenstufigen Items, die zu jeweils vier Skalen zusammengefasst werden. Die im Fragebogen operationalisierten Dimensionen fokussieren weniger auf die funktionale Qualität einer Website, sondern auf die Qualitätswahrnehmung der Nutzer. Diese ist durch die pragmatische und „hedonische" Qualität beschrieben.[56]

In der Regel verfügt jede UX-Abteilung und jedes größere Forschungsinstitut über eigene Instrumente, mit denen die User Experience einer Website standardisiert beurteilt werden kann. Wichtig bei der Verwendung von Fragebögen bzw. Itemlisten ist dabei in jedem Fall die angemessene Skalierung und Validierung der Items über entsprechende statistische Verfahren (zum Beispiel Faktorenanalyse zur Identifikation übergreifender Einstellungsdimensionen). Es gibt aber nicht das richtige Instrument, sondern jedes Verfahren ist nur so gut, wie die Empfehlungen, die daraus resultieren. Deshalb sollten Sie abwägen und je nach Fragestellung darüber entscheiden, welches Verfahren sinnvoll eingesetzt oder für die eigenen Belange entwickelt werden sollte.

4. Ableiten von Handlungsempfehlungen

Die Analyse und Aufbereitung von UX-Daten erfolgt auf verschiedenen Ebenen. Ziel ist es, die verbalen Daten aus Interview und Fragebogen mit den Verhaltensdaten aus Navigations- und Interaktionsprotokollierung (sowie Blickbewegungen) zu integrieren und daraus die relevanten Handlungsempfehlungen abzuleiten. Die Handlungsempfehlungen sind dabei nicht zwangsläufig deckungsgleich mit den Wünschen und Verbesserungsvorschlägen, die von den Nutzern in den Interviews geäußert werden. Im Gegenteil: Erst die Einschätzung und Gewichtung durch die UX-Experten entscheidet, welche Empfehlungen zur Optimierung einer Website wirklich handlungsrelevant sind und welche nicht. Für den Analyse- und Auswertungsprozess gibt es daher keine allgemeingültigen Regeln, sondern die Aggregation der Daten erfolgt in der Praxis gegenstandsbezogen.

---

[55] Hassenzahl, Marc/Beu, Andreas/Burmester, Michael (2001): Engineering Joy. In: IEEE Software, Heft 1/2001. S.70 bis 76.

[56] Die vier Skalen umfassen (a) die pragmatische Qualität, die in der wahrgenommenen Fähigkeit eines Produkts besteht, Handlungsziele zu erreichen; (b) hedonische Qualität – Stimulation, als Befriedigung der Bedürfnisse nach Verbesserung der eigenen Kenntnisse; (c) hedonische Qualität – Identität, als Kommunikation selbstwertdienlicher Botschaften; und (d) Attraktivität als globale Bewertung des Produkts im Sinne von gut, attraktiv und angenehm. Vgl. Hassenzahl, Marc et al. (2008): Der User Experience (UX) auf der Spur: Zum Einsatz von www.attrakdiff.de. In: Brau, Henning et al. (Hrsg.): Usability Professionals 2008, S. 78-82.

Gegenstandsbezogene Ableitung von Empfehlungen

Gegenstandsbezogen heißt, dass man schon im Vorfeld der Untersuchung mit bestimmten Vorannahmen beginnt, diese während des gesamten Erhebungsprozesses revidiert, modifiziert und anreichert und im Analyseprozess systematisch auf das Ausgangsmaterial anwendet.[57] Anhand des Datenmaterials werden neue Hypothesen gebildet, die wiederum in einem iterativen Prozess sukzessive an den Daten überprüft werden. Da es sich um einen iterativen Prozess handelt, ist der kommunikative Aspekt zentral: UX-Studien sollten im Austausch mit unterschiedlichen Projektbeteiligten und nicht ‚im stillen Kämmerlein' erfolgen, denn verschiedene Perspektiven auf ein und dasselbe Problem lassen oft neuartige Erkenntnisse und Lösungen zu.

Ziel des Analyseprozesses ist die Ableitung von Handlungsempfehlungen. Diese resultieren zum Einen aus den subjektiven Befragungsdaten, die darüber Auskunft geben, welche Eindrücke, Erwartungen und Vorlieben die Nutzer haben und wichtige Rückschlüsse auf die Optimierung der Angebote, Services, der Tonalität und kommunikativen Ansprache zulassen. Bei der Optimierung von Funktionsgesichtspunkten (Usability und Navigation) sollten zum Anderen auch die „objektiven" Verhaltensdaten mit einbezogen werden. Diese geben oftmals genauer als Befragungsdaten darüber Auskunft, ob und wie die Nutzer tatsächlich mit einem System umgehen. Die Gewichtung der Handlungsempfehlungen erfolgt daher unter Berücksichtigung sämtlicher Datenquellen.

Eine **inhaltsanalytische** Auswertung nach Häufigkeit der auftretenden Probleme oder genannten Schwachstellen empfiehlt sich nur als Ergänzung, denn auch das häufige Auftreten eines bestimmten Problems korrespondiert nicht zwangsläufig mit der tatsächlichen oder auch subjektiv erlebten Bedeutung des Problems. Denkbar sind zum Beispiel gravierende Schwachstellen, die von den Probanden aber kaum bemerkt werden (zum Beispiel weil sie anderweitige Lösungsstrategien entwickeln). Handlungsempfehlungen sollten vielmehr danach gewichtet werden, wie sehr eine Schwachstelle das Nutzerempfinden beeinträchtigt (von „sehr stark" bis „überhaupt nicht"). Entsprechend können dann auch die Handlungsempfehlungen nach ihrer Bedeutsamkeit priorisiert werden, zum Beispiel von „must do" (müssen unbedingt umgesetzt werden, wenn die Website erfolgreich sein soll) bis „nice to do" (Umsetzung nicht zwingend erforderlich, aber bei geringem Aufwand zu empfehlen). Neben der inhaltlich-ausführlichen Darstellung sämtlicher Detailergebnisse

[57] Vgl. Glaser, Barney G./Strauss, Anselm (2008): Grounded Theory. Strategien qualitativer Forschung. (Übersetzung der Originalausgabe von 1967 unter dem Titel The Discovery of Grounded Theory: Strategies for Qualitative Research). Bern.

können auf diese Weise Tabellen mit Handlungsempfehlungen erstellt werden, die den Entwicklern und Designern eine wichtige Hilfestellung bei der Optimierung liefern.

Für eine gute Lesbarkeit von Handlungsempfehlungen bei UX-Tests eignen sich Tabellen, die neben Kategorie und Seitentyp, auf die sich die Empfehlung bezieht, auch eine Kurzbeschreibung der Schwachstelle, die konkrete Empfehlung und deren Priorisierung nebeneinander auflisten. Neben den schriftlichen Reports, die den allgemeinen Leitlinien für die Darstellung von Studienergebnissen genügen sollten, empfehlen sich für eine Diskussion von UX-Ergebnissen vor allem die Durchführung von Workshops. Im Rahmen von UX-Workshops, in denen die unterschiedlichen Projektbeteiligten sowie interne und externe Auftraggeber teilnehmen, werden die zentralen Aspekte der User Experience vorgestellt und speziell mit Blick auf die daraus resultierenden Optimierungsansätze diskutiert. Dabei sollte ein geschulter Moderator durch den Workshop führen. Zu den Aufgaben des Moderators gehört im Vorfeld auch die Abstimmung der Zielsetzungen und die Vorbereitung zentraler Leitfragen, die pointierte Darstellung und moderierte Diskussion der Ergebnisse im Rahmen des Workshops sowie die zusammenfassende und kommentierte Nachbereitung der Ergebnisse inklusive Planung weiterer Schritte. Workshops dienen daher vor allem einem vertiefenden Verständnis der Ergebnisse über alle Projektebenen hinweg und führen im Idealfall zu einer mit den Unternehmenszielen abgeglichenen Priorisierung und Erarbeitung weiterer Lösungsschritte für die Optimierung Ihrer Website.

Gewichtung und Identifikation von „Erfolgsfaktoren"

Angesicht der Vielzahl von UX-relevanten Leistungs- und Zufriedenheitsfaktoren stellt sich berechtigterweise die Frage, wie deren Bedeutung bzw. Relevanz für die praktische Anwendung und Optimierung festgestellt werden soll: An welchen Stellschrauben sollte bei der Optimierung einer Website angesetzt werden, um den größten Effekt auf Gesamtzufriedenheit, Conversion und Nutzungsbereitschaft zu erzielen?

Hier helfen regressionsanalytische Modelle, die für jeden Faktor bzw. für die erhobenen Eindrucksdimensionen den Beitrag zur Erklärung der Gesamtzufriedenheit oder Nutzungsbereitschaft einer Website aufklären. Regressionsanalysen gehören zu den multivariaten Statistikverfahren und dienen der Analyse von Beziehungen zwischen einer abhängigen Variable (zum Beispiel Gesamtzufriedenheit) und einer oder mehreren unabhängigen Variablen (zum Beispiel den Leistungsfaktoren).[58] Sie werden eingesetzt, um Zusammenhänge zwischen den Variablen zu erkennen und abzuleiten, welche Vorhersagekraft eine unab-

[58] Backhaus, Klaus et al. (1996): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Springer, Berlin Heidelberg, S. 1-55.

hängige Variable für die Ausprägung der abhängigen hat. Diese sogenannten „Treiber" oder „Erfolgsfaktoren" können dann entsprechend ihrer Bedeutung gewichtet werden. Da sowohl abhängige als auch unabhängige Variablen metrisch skaliert sein müssen, eignet sich die Regressionsanalyse aber eher für quantitativ ausgerichtete Remote-Tests oder Onlinebefragungen mit Stichprobengrößen ab circa hundert Personen.

Eine Alternative bei qualitativen UX-Tests stellt die klassische **Stärken-Schwächen-Analyse** dar. Hier werden die Faktoren in einem vierdimensionalen Raum zwischen den Achsen „Ausprägung des Faktors" und „Bedeutung des Faktors für die Gesamtzufriedenheit" angeordnet. Die Bedeutung eines Faktors für die Gesamtzufriedenheit oder Nutzungsbereitschaft wird über eine einfache Korrelation hergestellt und bedarf nicht unbedingt eines metrischen Skalenniveaus. So weist beispielsweise ein Faktor, der negativ bewertet wird, aber eine hohe Bedeutung für die Gesamtzufriedenheit der Website hat, hohes Verbesserungspotenzial auf („must do"). Ein Faktor dagegen, der negativ bewertet wird, aber nur eine geringe Bedeutung für die Gesamtzufriedenheit hat, wird von den Nutzern in Kauf genommen, bedarf aber nicht zwangsläufig der Optimierung („nice to have"). Faktoren, die positiv bewertet werden, aber eine geringe Bedeutung haben, werden von den Nutzern als selbstverständlich wahrgenommen („must have"); Faktoren, die positiv bewertet werden und eine hohe Bedeutung haben, gelten als Erfolgsfaktoren und sollten unbedingt beibehalten werden („must hold").

Es liegt auf der Hand, dass neben den übergreifenden Erfolgstreibern auch individuelle Faktoren für den Erfolg einer Website verantwortlich sein können. Soweit es die Stichprobengröße oder quantitative Begleitforschung (zum Beispiel über Onlinebefragungen) erlauben, sollten daher zur Abstützung der qualitativen Ergebnisse auch Verfahren zur Ermittlung von Erfolgsfaktoren zum Einsatz kommen.

5. Fazit: Durch Full User Experience zum Erfolg

In dem Kapitel wurden die wichtigsten Aspekte zur Durchführung von UX-Tests sowie Vor- und Nachteile einzelner Verfahren beschrieben. Wie eingangs erwähnt, müssen die verschiedenen Facetten der User Experience operationalisiert und gemessen werden, was die Komplexität von UX-Tests häufig beträchtlich steigert. Hat man es beim Rapid Prototyping (vgl. Kapitel VI) noch mit überschaubaren Prozessen und Entwicklungsschritten zu tun, müssen beim UX-Test verschiedene Datenebenen und Facetten des menschlichen Erlebens miteinander in Einklang gebracht werden.

Eine weitere Herausforderung liegt im qualitativen Ansatz selbst. Zwar können über die Wahl größerer Stichproben einzelne Kennwerte metrisiert werden, diese fördern aber nicht zwangsläufig auch die relevanten und bedeutsamen Facetten der User Experience zutage. Entscheidend ist letztlich eine psychologisch fundierte Interpretation der Ergebnisse, die sich (auch) als Sprachrohr der Nutzer versteht – was viel Erfahrung, Einfühlungsvermögen und Usability-Expertise erfordert. Im letzten Schritt messen sich die Ergebnisse eines UX-Tests aber an den Empfehlungen, die aus ihnen resultieren und die in Zusammenarbeit mit den Projektbeteiligten in konkrete und umsetzbare Änderungsanforderungen münden müssen. Gelingt dies, liefern qualitative UX-Tests nicht nur die entscheidenden Optimierungshinweise für Ihre Website, sondern auch ein vertieftes Verständnis für die zugrunde liegenden Ursachen und Wirkfaktoren einer ganzheitlich verstandenen User Experience.

6. Literaturverzeichnis

Backhaus, Klaus et al.: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Springer, Berlin Heidelberg 1996.

DIN EN ISO 9241: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten, Berlin 1996.

Esch, Franz-Rudolf et al.: Strategische Empfehlungen für den Aufbau von Markenbekanntheit und für die Markenpositionierung im Internet. In: Esch, Franz-Rudolf: Moderne Markenführung. Grundlagen – Innovative Ansätze – Praktische Umsetzungen. Wiesbaden 2005, S. 676–705.

Glaser, Barney G./Strauss, Anselm: Grounded Theory. Strategien qualitativer Forschung. (Übersetzung der Originalausgabe von 1967 unter dem Titel The Discovery of Grounded Theory: Strategies for Qualitative Research). Bern 2008.

Hassenzahl, Marc/Beu, Andreas/Burmester, Michael: Engineering Joy. In: IEEE Software, Heft 1/2001. S.70–76.

Hassenzahl, Marc et al.: Der User Experience (UX) auf der Spur: Zum Einsatz von www.attrakdiff.de. In: Brau, Henning et al. (Hrsg.): Usability Professionals 2008, S. 78–82.

Herzberg, Frederick; Mausner, Bernard; Snyderman, Barbara Bloch: The Motivation to Work. 2. Aufl., Wiley, New York 1959.

Herzberg, Frederick: One more time: how do you motivate employees? In: Harvard Business Review 46 (1968) 1, S. 53–62.

Knoblauch, Hubert (2004): Die Video-Interaktions-Analyse. In: sozialersinn, Heft 1/2004, S. 123–138.

Krug, Steve: Don't make me think! A Common Sense Approach to Web Usability, Indianapolis. Deutsche Übersetzung (2006): Don't make me think! Web Usability – das intuitive Web, 2. Aufl., Heidelberg 2006.

Nielsen, Jacob: Usability Engineering. AP Professional Ltd., London 1993.

Reeps, Inga E.: Joy-of-Use. Ästhetik, Emotion und User Experience für interaktive Produkte. VDM Verlag Dr. Müller, Saarbrücken 2006.

Schmeißer, Daniel/Oberg, Nicole: User Experience als Erfolgsfaktor im E-Commerce. In: planung & analyse, Heft 1/1009, S. 26–31.

Schmeißer, Daniel et al.: Wie aus Nutzern Kunden werden. Schlüssel zur Optimierung von Online-Kampagnen durch systematische Erfolgskontrolle. In: planung & analyse, Heft 1/2007, S.58–63.

Schmeißer, Daniel/Oberg, Nicole: Der Onlineshop – Revier von Jägern und Sammlern oder von Stöberern? Ergebnisse einer qualitativen Studie zum Online-Versandhandel. In: planung & analyse, Heft 5/2006, S. 59–65.

Schmeißer, Daniel/Sauer, Michael: Usability-Forschung im E-Commerce. Methoden und Ergebnisse am Beispiel des Online-Versandhandels. In: planung & analyse, Heft 1/2005, S. 75–81.

Schreier, M.: Qualitative Stichprobenkonzepte. In: Naderer, G./Balzer, E. (Hrsg.): Qualitative Marktforschung in Theorie und Praxis, Gabler, Wiesbaden 2007. S. 232–245.

Tullis, Tim/Albert, Bill: Measuring The User Experience. Collecting, Analyzing, and Presenting Usability Metrics. Elsevier Inc., Oxford 2008.

Yom, Miriam/Wilhelm, Thorsten H./Gauert, Stefanie: Protokolle lauten Denkens und Site Covering. Eine Erweiterung der Methode zur detaillierten Bewertung des Screendesigns von Webangeboten. In: Buber, Renate/Holzmüller, Hartmut H. (Hrsg.): Qualitative Marktforschung. Konzepte – Methoden – Analysen, Wiesbaden 2007. S. 637–52.

X.I Fallbeispiel: User Experience Test für neckermann-reisen.de

von Daniel R. Schmeißer und Oliver Rengelshausen

1. Untersuchungsdesign zur Erfassung der UX 249
2. Ergebnisse und Optimierungsansätze 251
3. Fazit: Mit verbesserter User Experience zum Erfolg 261

Der Vertrieb von Reisen über das Internet wächst seit Jahren beständig. Reiseinteressierte nutzen immer häufiger das Internet, um sich über Reiseziele und Hotels zu informieren, Angebote zu vergleichen und online zu buchen. Die technologischen Möglichkeiten und die Erfahrungen und damit auch die Erwartungen der User steigen stetig. Unternehmen, die im Online Vertrieb erfolgreich sein wollen, müssen daher ihre Websites ständig optimieren und weiterentwickeln, um sich den kontinuierlich verändernden Rahmenbedingungen anzupassen und den steigenden Ansprüchen der Nutzer an die Online-Auftritte gerecht zu werden.

Die Thomas Cook AG deckt in Deutschland den gesamten Bereich touristischer Leistungen für verschiedene Zielgruppen ab und ist seit Jahren auch im Online-Vertrieb aktiv. Zu den Veranstaltermarken der Thomas Cook AG gehört unter anderem **Neckermann Reisen**, der **zweitgrößte Reiseveranstalter Deutschlands**. Neckermann Reisen hatte im April 2007 einen Relaunch seiner Webpräsenz neckermann-reisen.de durchgeführt. Die neue Seite diente als „Online-Schaufenster" für den Veranstalter, das mit einem weit gefassten Sortiment und Angebotsportfolio eine breite Zielgruppe ansprechen sollte. Neue Features und Funktionen wie eine blätterbare Navigationsleiste und verschiedene Filtersetzungen und -optionen sollten zudem die Attraktivität der Seite erhöhen und die Suche nach Informationen erleichtern.

Nach dem Relaunch zeigte sich relativ bald eine **Verschlechterung der Conversion Rate**, darüber hinaus stagnierte die Umsatzentwicklung der Website. Die Auswertungen des Nutzerverhaltens mithilfe der eingesetzten Webanalytics Software zeigten eine Zunahme an Abbrüchen der User auf der Website auf. Die Auswertungen lieferten jedoch keine Hinweise darauf, welches die konkreten **Gründe für die Abbrüche** waren. Vor diesem Hintergrund fiel die Entscheidung, den bestehenden Problemen auf der Website mithilfe eines User-Experience-Tests auf den Grund zu gehen.

Im Rahmen dieses Tests sollte der neue Internetauftritt von Neckermann Reisen hinsichtlich Akzeptanz, Verständlichkeit und Benutzerfreundlichkeit in der Zielgruppe überprüft werden. Im Vordergrund standen die Beurteilung der neuen Navigationsstruktur sowie das Verständnis der neuen Filteroptionen bei der Informationssuche und Reisebuchung. Darüber hinaus sollten Fragen nach möglichen Unterschieden in Wahrnehmung und Akzeptanz in Abhängigkeit von den verschiedenen Einstiegsmöglichkeiten auf die Neckermann Reisen Seite beantwortet werden. Typische Einstiege in das Angebot sind neben dem direkten Weg über die Startseite die Nutzung einer Suchmaschine (Suche nach einem Urlaubsland bzw. einem konkreten Hotel) oder der Einstieg über ein Online-Werbemittel.

Ziel der Untersuchung war es, mögliche **Schwachstellen** zu identifizieren und konkrete Empfehlungen an die Hand zu geben, um den Informations- und Buchungsprozess optimal an den Bedürfnissen der Nutzer auszurichten. Auf Grundlage der Analyse sollten konkrete **Empfehlungen für eine kunden- und abverkaufsorientierte Gestaltung** der Landing Pages sowie des gesamten Shops und der Buchungsstrecke abgeleitet werden.

1. Untersuchungsdesign zur Erfassung der UX

Zur Beantwortung der Untersuchungsfragen wurde ein breiter **Methoden-Mix** zur Erfassung der vollständigen User Experience gewählt. Grundlage der Studie waren 32 ausführliche Tiefeninterviews mit einer Dauer von circa zwei Stunden. Zusätzlich wurden 48 teilstandardisierte Interviews von 60 Minuten Länge durchgeführt, um das Post-Klick-Verhalten in Abhängigkeit unterschiedlicher Einstiege quantitativ zu erfassen. Die Stichprobe von **insgesamt 80 Nutzern** war für eine qualitative Studie also bewusst breit angelegt, um die Ergebnisse mit Blick auf unterschiedliche Zielgruppen und Nutzungsszenarien auf eine robuste Basis zu stellen.

Befragt wurden Internetnutzer im Alter zwischen 20 und 59 Jahren, gleichmäßig verteilt auf drei Altersgruppen. Frauen wurden mit 60 Prozent etwas höher gewichtet als Männer. Alle Teilnehmer hatten in den letzten zwölf Monaten eine Reise bzw. einzelne Reisebestandteile (mindestens Flug plus Hotel) online gebucht und beabsichtigten erneut eine Reisebuchung-Online in den kommenden zwölf Monaten. Alle Teilnehmer waren Pauschalreise-affin, darüber hinaus äußerte jeweils die Hälfte Interesse an Städtereisen und Last-Minute-Angeboten. Im Quotenplan wurden neben den Urlaubsvorlieben außerdem die Haushaltsgröße (Singles, Paare, Familien mit/ohne Kleinkinder) und das Haushaltsnettoeinkommen kontrolliert.

In den Interviews wurden unterschiedliche Einstiegsmöglichkeiten untersucht:
1. Gruppe: Einstieg über ein Werbebanner mit zugehöriger Landing Page
2. Gruppe: Einstieg über die Ländersuche bei Google (Stichwort zum Beispiel „Mallorca“)
3. Gruppe: Einstieg über konkrete Hotelsuche bei Google (Stichwort zum Beispiel „Villa Cati“)
4. Gruppe: Direkteinstieg durch Eingabe der URL (www.neckermann-reisen.de)

Ziel der unterschiedlichen Einstiegsoptionen war es im ersten Schritt herauszufinden, wie sich die Nutzer spontan verhalten, wenn sie auf die Neckermann Reisen Website gelangen. Der Einstiegs-Use-Case war sehr allgemein formuliert (Suche nach relevanten Reiseinformationen) und stellte sicher, dass sich die Probanden frei bewegen konnten. Im Fokus des darauf folgenden Interviews stand zunächst das Erleben der Nutzer im Erstkontakt mit der Startseite bzw. der jeweiligen Google- oder Kampagnen-Landing Page. Mittels Surfpfadanalyse sollte ermittelt werden, wie sich die Art des Einstiegs auf das **Post-Klick-Verhalten** und die gesamte User Experience auswirkt, um Landing Pages und nachfolgende Prozesse zielgerichtet optimieren zu können.

Im zweiten Teil der Interviews lösten die Teilnehmer **typische Nutzungsaufgaben** (Use Cases), die im Vorfeld der Anforderungsanalyse von den Projektbeteiligten festgelegt wurden:

- Suche nach einer Pauschalreise inklusive vollständiger Buchung
- Recherche von Last-Minute- und Städteangeboten
- Buchung einer Flugreise

Die Reihenfolge und Anzahl der Aufgaben wurde über die Stichprobe rotiert, so dass Reihenfolge- und Lerneffekte vermieden wurden. Zeitsynchron zur Lösung der Use Cases wurden Interaktionsprotokolle angefertigt sowie die Blickbewegungen der Teilnehmer mittels kontaktloser Eyetracker aufgezeichnet, parallel hierzu fand eine Verhaltensbeobachtung mittels Videoaufzeichnung statt. Jeweils im Anschluss an jede Nutzungsaufgabe wurden die Teilnehmer auf Basis eines teilstandardisierten Leitfadens datengestützt (also auf Basis ihrer individuellen Surfprotokolle) hinsichtlich ihrer subjektiven Eindrücke vertiefend befragt.

Als Ergebnis der Studie konnte eine **Liste mit insgesamt über 30 Optimierungsempfehlungen** erstellt werden. Sie umfasste kleinere Details wie das Wording einzelner Buttons bis hin zu weitreichenden Empfehlungen bezüglich der konzeptuellen Neugestaltung der Navigationsstruktur und des Aufbaus von Landing Pages. Ausgewählte Ergebnisse werden im Folgenden am Beispiel der Startseite und Hauptnavigation, der Hotelauswahl und typischer Landing Pages vorgestellt.

2. Ergebnisse und Optimierungsansätze

Optimierung von Startseite und Navigationskonzept

Das übergreifende Erleben der Nutzer war zunächst positiv: Die Website wurde aufgrund ihrer Tonalität und kommunikativen Ansprache als sympathisch, ansprechend und einladend erlebt. Dem neuen Design war es gelungen durch modernes Layout und Design ein Urlaubsgefühl bei den Nutzern zu wecken und in hohem Maße zu animieren, sich mit den Angeboten zu beschäftigen. Das Look & Feel der Seite verfügte darüber hinaus über eine hohe Markenpassung und transportierte die für Neckermann Reisen intendierten Imagefacetten. In Übereinstimmung mit der Ausgangssituation rückläufiger Buchungsraten konnte aber auch im Test beobachtet werden, dass die Befragten an verschiedenen Stellen ihre Nutzung abbrachen bzw. Stolpersteine und Missverständnisse die Nutzung des Angebots erschwerten. Welche Gründe gab es hierfür?

Als ein zentrales Problem stellte sich die Primärnavigation heraus. Diese war als Karussell-Navigation so konzipiert, dass sie auf der Startseite nicht vollständig sichtbar war, sondern rechts oben in einem abgegrenzten Bereich per Mausklick rotiert werden konnte. Neben der generellen Problematik, dass nicht alle Rubriken auf Anhieb sichtbar waren, machte bereits die Blickverlaufsanalyse sichtbar, dass die Navigation von vielen Nutzern spät oder gar nicht gesehen wurde, was zur Folge hatte, dass alternative Einstiege genutzt wurden. Der Scanpfad auf der Neckermann Reisen Startseite (siehe Abbildung 69 auf der folgenden Seite) bildet die typische Blickreihenfolge in den ersten Sekunden ab (= Orientierungsphase) und belegt anschaulich, dass die Nutzer die mittig platzierte Buchungsmaske für den (visuellen) Einstieg nutzen. Die ersten vier Blicksprünge entfallen auf dieses Seitenelement, anschließend werden die Urlaubs-Teaser rechts und links hiervon betrachtet, bevor der Blick vergleichsweise spät (an 14. Stelle) auf die Navigation fällt. Die Parameter der aggregierten Blickverweildauer im Orientierungs- versus Intensivkontakt (vgl. Abbildung 70 auf Seite 253), bestätigen dieses Bild. Nicht nur im Erstkontakt entfällt das Gros der Aufmerksamkeit auf die mittlere Buchungsmaske, sondern auch nach intensiver Beschäftigung mit der Seite (definiert als maximale Beschäftigungsdauer zwischen 10 und 30 Sekunden) werden weite Teile der Seite nur sehr flüchtig wahrgenommen. Die Hauptnavigation findet durchgängig kaum Beachtung. Es fällt aber auch auf, dass die Angebote und Services im unteren Seitenbereich, der bei einer Standardauflösung von 1024 × 768 nicht mehr im initial sichtbaren Monitorbereich liegt, auch nach längerer Betrachtungsdauer auf der Seite weder beachtet noch genutzt wird; unter anderem auch deshalb, weil der sichtbare, obere Seitenbereich einen in sich geschlossenen Eindruck macht und daher nicht zum Scrollen der Seite animiert.

Abbildung 69: Scanpfad für die Orientierungsphase auf der alten Neckermann Reisen-Strartseite: Die Zahlen in den Kreisen geben die typische Blickreihenfolge wieder. Die Größe des Kreises indiziert die durchschnittlichen Verweildauer im Erstkontakt mit dem jeweiligen Areal. Die Position des Kreises zeigt den Schwerpunkt der Fixation innerhalb des Areals an (Quelle: Thomas Cook AG, phaydon).

Schwachstellen und Konsequenzen für die User Experience

Zieht man sämtliche Befunde aus Blickmessung, Interview und Interaktionsanalyse zusammen, ergeben sich bereits für die Startseite eine Reihe zentraler Schwachstellen, die sich kritisch auf die Erfüllung der relevanten Use Cases und damit auf die Leistungskriterien (Zeit zur Aufgabenlösung, Fehlerhäufigkeit, Effizienz etc.) auswirkten:

- Wie erwähnt, beachten die Nutzer die Hauptnavigation spät oder gar nicht; erschwerend kam hinzu, dass die Navigation in den unteren Ebenen gar nicht mehr auftaucht. Den Befragten mangelte es daher durchgängig an einem feststehenden Orientierungselement, was innerhalb des Suchprozesses immer wieder zu Verunsicherung führte und die Aufgabenerfüllung unnötig verlängerte.

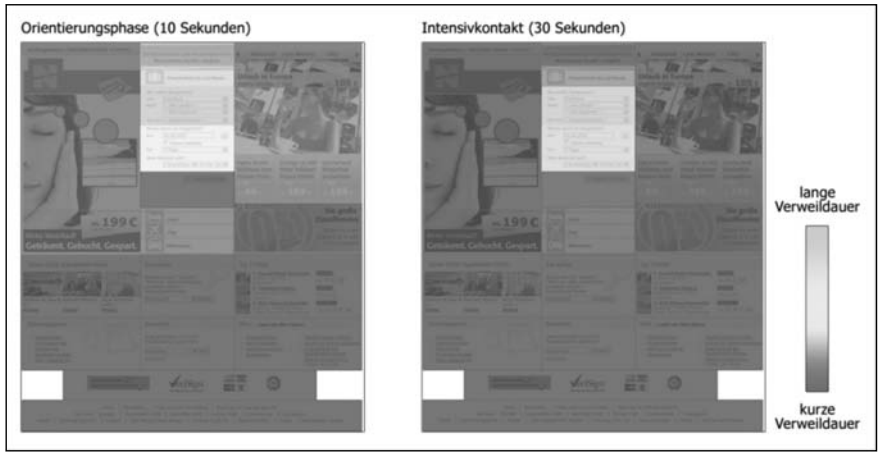


Abbildung 70: Verweildauer auf der Neckermann Reisen Startseite in der Orientierungsphase und während des Intensivkontakts; Erläuterung im Text (Quelle: Thomas Cook AG/phaydon).

- Die kleinteilige Darstellung der Angebote im unteren Seitenbereich wird im Vergleich zur großflächig-ansprechenden Bebilderung der oberen Seitenhälfte als „Bruch" erlebt. Die Folge: Nutzer blenden diesen Seitenbereich einfach aus; subjektiv bleibt der Eindruck insgesamt wenig attraktiver Angebote und einer nicht-kohärenten Produktinszenierung.
- Unabhängig von der geringen Auffälligkeit der Navigation erschwerten einzelne Benennungen (zum Beispiel „City" für Städtereisen) das Verständnis.
- Die Nutzer bewerteten die mittig platzierte Buchungsmaske grundsätzlich als hilfreich. Dennoch wurde die Funktionalität nicht vollständig genutzt, da die vertikal angeordnete Reiterstruktur als solche nicht verstanden wurde. So entstand der Eindruck, man könne lediglich nach Pauschalreise & Last Minute Angeboten suchen.
- Die Zusammenfassung von Pauschalreise- und Last-Minute-Angeboten in der Suchmaske sorgte für Verwirrung – die Nutzer beurteilten die geringe Differenzierung an Zugängen als unzureichend.

Empfehlungen für die Optimierung der Benutzeroberfläche

Auf Grundlage der Ergebnisse und ausformulierten Handlungsempfehlungen nahm Neckermann Reisen eine Priorisierung vor und erstellte detaillierte Änderungsanforderungen. Dabei entschloss man sich

- die Top-Navigation inklusive dem Button „Startseite" und der Freitextsuche freigestellt im Header zu platzieren und als festes Orientierungselement auf allen Seitenebenen anzubieten (siehe Abbildung 71).

- die IBEs im sichtbaren Bereich als Radio-Buttons anzuordnen und zusätzlich einen Last-Minute-Einstieg zu integrieren (siehe Abbildung 72)
- das Layout (Look and Feel) im nicht-sichtbaren Bereich an das im sichtbaren Bereich anzupassen (zum Beispiel durch Hinterlegung von Flächen ähnlich wie die Teasergestaltung), und bei einer Auflösung von 1024 × 768 visuell anzuzeigen, dass auch im nicht-sichtbaren Bereich attraktive Inhalte folgen
- die große Teaserfläche im oberen Bereich anzupassen, zum Beispiel durch eine prägnante Abgrenzung von Text und Bild

Durch die Veränderungen sollte gewährleistet werden, dass der Nutzer bereits auf der Startseite die möglichen Buchungseinstiege (Pauschal, Last Minute, Flug etc.) auf den ersten Blick als Optionen wahrnimmt und nutzt.

Abbildung 71: Die Neckermann Reisen Startseite vor und nach der Optimierung (Quelle: Thomas Cook AG/phaydon).

Außerdem sollten das Look & Feel und damit die Attraktivität der Seite weiter optimiert werden, um Verweildauer und Klickrate zu erhöhen.

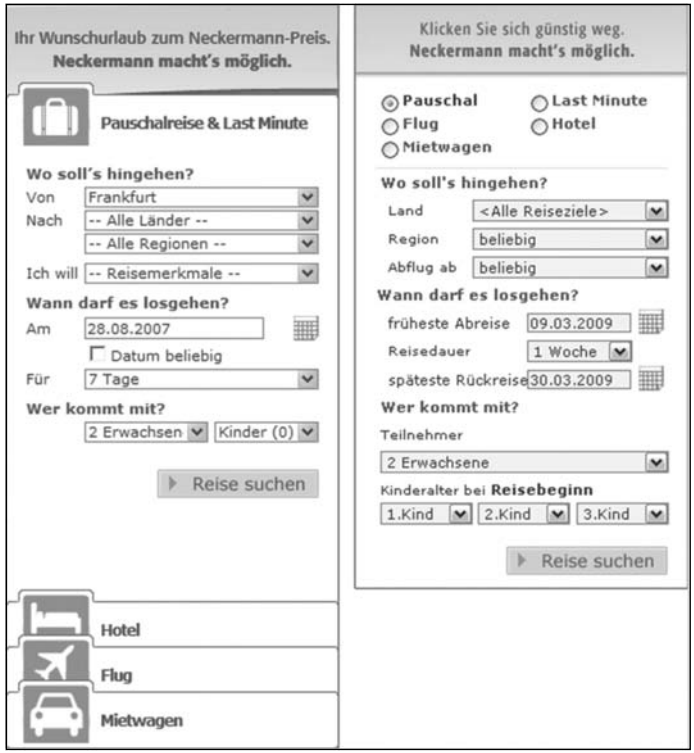


Abbildung 72: IBE-Einstiegsmaske auf der Startseite – vor und nach der Optimierung (Quelle: Thomas Cook AG/phaydon).

Optimierung der Hotelübersichtsseiten

Bei der Reisebuchung sind die Hotelübersichtsseiten zentraler Dreh- und Angelpunkt der Suche. Hier verschaffen sich die Nutzer einen ersten Überblick über die verfügbaren Angebote, Hotelkategorien und Empfehlungen für einen Reiseort. Die Anforderungen an die Hotelübersichten sind daher hoch: Sie müssen für den Auswahl- und Buchungsprozess alle Optionen bereitstellen und gleichzeitig einen anschaulichen und komprimierten Überblick über die verschiedenen Hotelangebote bieten.

Abbildung 73 zeigt den Surfpfad der Nutzer, die über die Ländersuche bei Google einsteigen und auf eine Hotelübersichtsseite gelangen. Die Nutzer verbringen insgesamt über 40 Prozent ihrer Zeit auf den Übersichtsseiten, im Durchschnitt werden fünf bis sechs Übersichtsseiten auf der Suche nach einem geeigneten Hotel angeklickt. Die grünen und roten Pfade, die von ihnen hin und weg führen, zeigen das sprunghafte Verhalten der Nutzer, die immer wieder einzelne Hotels anklicken, aber häufig bereits nach kurzer Zeit wieder auf die Übersichtsseiten zurückkehren, um sich weiter zu orientieren.

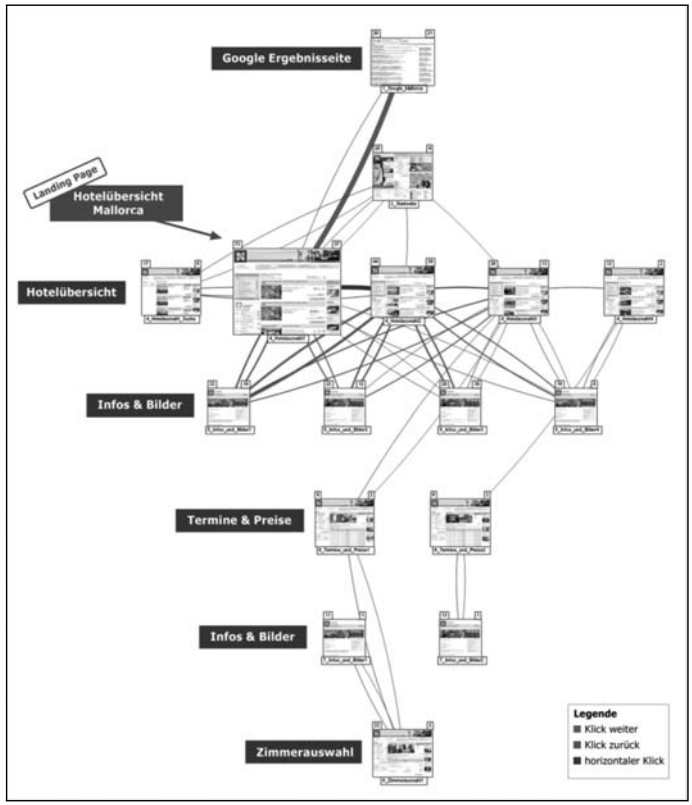


Abbildung 73: Surfpfade im Anschluss an den Einstieg über Google – im Mittelpunkt der Nutzung stehen die Hotelübersichten, die mehrfach angesteuert werden (Quelle: Thomas Cook AG/phaydon).

Die Hotelübersichtsseiten wurden von den Nutzern auf den ersten Blick als übersichtlich und vollständig erlebt. Im Nutzungsverlauf zeigten sich aber bei der konkreten Auswahl einzelner Reiseangebote Probleme, die häufig zu Abbrüchen führten.

Schwachstellen und Konsequenzen für die User Experience

- Hauptgrund für die Abbrüche war die links platzierte Suchauswahl. Diese war in drei Bereiche aufgeteilt: Filtern der Ergebnisse, Reiseauswahl und Freie Suche. Bereits die Komplexität der Such- und Filtermöglichkeiten überforderte viele Nutzer.
- Die Unterscheidung zwischen Ergebnisfilter und Gesamtsuche erschloss sich nicht auf Anhieb: Die Nutzer stellten die gewünschten Spezifikationen einfach von oben nach unten ein, übersahen dabei aber, dass es sich um unterschiedliche Suchmasken handelte.

- Die Freie Suche berücksichtigte nicht die voreingestellten Suchbegriffe und lieferte somit unpassende bzw. ungenaue Ergebnisse.

Empfehlungen für die Optimierung der Benutzeroberfläche

Bei den abgeleiteten Änderungsanforderungen entschloss man sich

- die Filtermöglichkeiten komplett zu überarbeiten und neu zu platzieren, indem die allgemeine Suchauswahl klarer von den Sortierfunktionen der Suchergebnisse getrennt wird
- die Filtermöglichkeiten deutlich zu erweitern, zum Beispiel durch Preis, Top-Reiseziele und durchschnittlichen Kundenbewertungen
- die Hoteltypanzeigen zu reduzieren, um die Nutzer bei ihrem Überblick nicht unnötig zu überfordern
- umfangreiche Sortiermöglichkeiten auch für den nächsten Schritt „Termine & Preise" anzubieten, damit ein Rücknavigieren vermieden wird.

Ziel der Optimierungsmaßnahmen war es, den Nutzern auf jeder Seite der Hotelauswahl (Übersichtsseite und Wahl der Termine & Preise) die erneute Suche bzw. Modifikation von Reiseangaben zur Verfügung zu stellen. Gleichzeitig sollten die Hotelübersichten so gestaltet sein, dass eine problemlose Sortierung der Ergebnisse ermöglicht wurde.

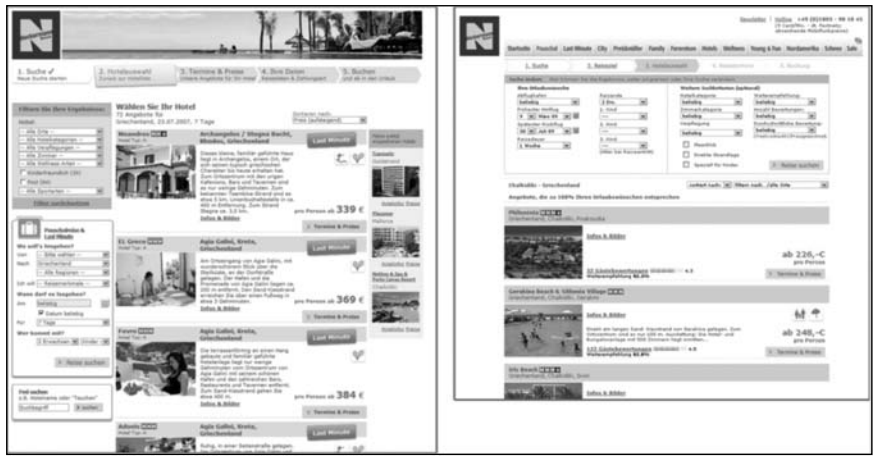


Abbildung 74: Hotelübersichtsseite vor und nach der Optimierung (Quelle: Thomas Cook AG/phaydon).

Das Resultat dieser Anpassung war, dass die durchschnittliche Verweildauer auf den Hotelübersichtsseiten stieg und die Nutzer sich intensiver mit den präsentierten Hotelangeboten beschäftigten.

Optimierung der Landing Pages

Die Studie bestätigte in eindrucksvoller Weise, dass die Art des Einstiegs (Google, Startseite, Werbebanner etc.) nicht nur das Post-Klick-Verhalten auf der Seite, sondern die gesamte User Experience im Sinne des subjektiven Erlebens entscheidend beeinflusst. Der erste Eindruck, der auf einer Startseite oder Landing Page entsteht, bestimmt die Richtung, in der nach weiteren Informationen gesucht wird, hinterlässt ein bestimmtes Bild der Marke, und formt die Erwartungen des Nutzers an die Website.

Der erste Eindruck wirkt aber auch prägend im Sinne eines Priming-Effekts: Ist er positiv, werden spätere Stolpersteine eher verziehen als im Falle eines holprigen Einstiegs. Es verwundert daher nicht, dass der Direkteinstieg über die Startseite trotz aller nachgewiesenen Schwachstellen rückblickend am angenehmsten beurteilt wurde, denn was im ersten Moment zählt ist der positive Gesamteindruck. Ganz anders auf den Landing Pages, auf die man nach dem Klick auf Google-Adwords gelangt. Zwar waren die Landing Pages zugeschnitten auf die jeweilige Art der Suchanfrage (zum Beispiel Ländernamen oder Hotels), die gesamte Gestaltung der Landing Pages entsprach aber nicht den Erwartungen der Nutzer: Wie bei vielen Google-AdWord-Kampagnen zu beobachten, verlinkten die Adwords als Deeplinks direkt auf die Angebote der herkömmlichen Hotelübersichten, was häufige Abbrüche und Irritationen zur Folge hatte.

Einstieg über Google Adwords

Problematisch an der Hotelübersichtsseite als Landing Page war, dass eine Top-Navigation zur weiteren Orientierung im Angebot fehlte und die Nutzer den Eindruck bekamen, sie befänden sich bereits im Buchungsprozess. Durch die oben skizzierte Problematik bezüglich der verschiedenen Such- und Filterfunktionen gestaltete sich die initiale Suche auf der Landing Page schwierig, viele brachen die Suche im weiteren Verlauf ab. Vermisst wurden darüber hinaus speziell ausgezeichnete Top-Angebote und Empfehlungen, die auf einen Mehrwert hinweisen. Kurz gesagt: Nutzer, die über die Ländersuche bei Google auf eine Hotelübersicht gelangten, fühlten sich nicht ausreichend „abgeholt“ und verließen die Seite häufig nach kurzer Verweildauer wieder.

Die daraus resultierenden Empfehlungen berücksichtigten (siehe Abbildung 75):

- eine stärker emotionale und auf die Adwords zugeschnittene Tonalität in der Kundenansprache (Kommunikation impliziter Markenbenefits)
- die Hervorhebung und prominente Ausweisung von besonderen Hotelangeboten, die preislich attraktiv sind (Kommunikation relevanter Reason Why)
- die Integration sämtlicher IBEs in der Suchmaske der linken Spalte (Anbieten verschiedener Handlungsoptionen)
- die knappere Listendarstellung der verfügbaren Hotels zur schnelleren Orientierung
- die Kennzeichnung der Listendarstellung als „die beliebtesten Hotels“
- die Verlinkung auf „ähnliche Themen“ und Auswahl weiterer Regionen pro Land

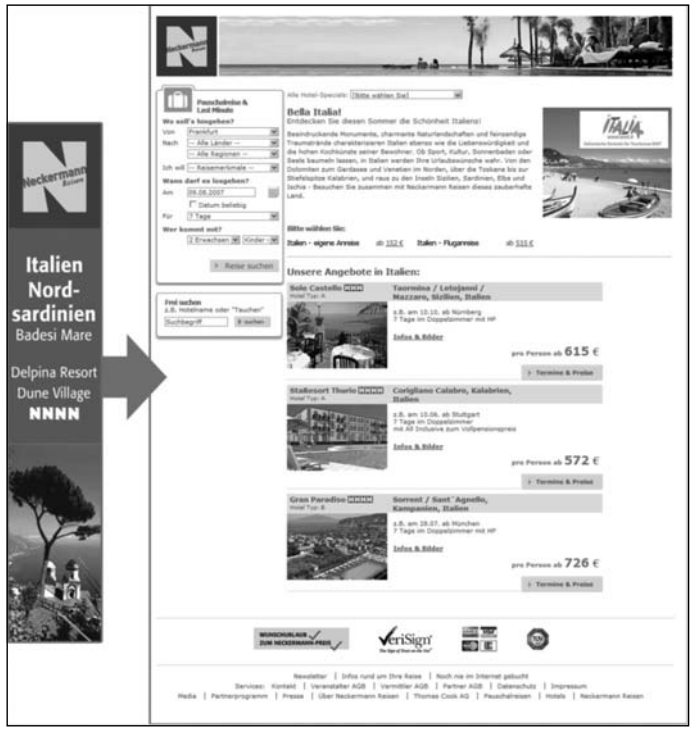


Abbildung 75: Werbebanner und zugehörige Landing Page aus dem Test (Quelle: Thomas Cook AG/phaydon).

Die Optimierung erfolgt daher auf einer grundlegenden konzeptuellen Ebene. Der Direkteinstieg über ein konkretes Hotel oder eine Region hätte den Aufwand nach sich gezogen, jede Landing Page sowohl emotional als auch funktional auf die jeweilige Kampagne abzustimmen. Darüber hinaus zeigte sich in der Kampagnenauswertung, dass Bannerformate mit konkreten Angeboten (siehe Abbildung 76) besser funktionierten als die unspezifischen Formate. Daraufhin wurde die Kampagnenlogik mit großem Erfolg umgestellt.

Abbildung 76: Aktuelle Kampagnenlogik – konkrete Angebote verlinken auf spezifische Hotelauswahlseiten (Quelle: Thomas Cook AG/phaydon).

Die neue Kampagnenlogik verfolgte den Ansatz, anstelle einzelner eher generischer Kampagnenthemen (zum Beispiel Last Minute Urlaub), den Kunden über eine relativ breite Auswahl konkreter Reiseangebote aus verschiedenen Zielgebieten anzusprechen. Die Verlinkung aus dem Banner führte den Kunden auf eine spezifische Hotelauswahlseite, auf der er sich über das Angebot und mögliche Alternativen informieren kann. Die zu diesem Zweck neu konzipierten dynamischen Multilink-Werbemittel ermöglichen die Verwendung von tagesaktuellen Angeboten und Preisen sowie eine fortwährende Optimierung der einzelnen Werbemittellinien.

Als Folge der Umstellung der Werbemittel stieg die Klickrate um 60 Prozent gegenüber der herkömmlichen Kampagne, damit einher ging eine deutliche Verbesserung der Conversion Rate. Dies führte zu einer höheren Kampagneneffizienz und einer Verringerung der Akquisitionskosten pro Neukunde.

3. Fazit: Mit verbesserter User Experience zum Erfolg

Zusammenfassend lässt sich sagen, dass sich die in den UX-Test gesetzten Erwartungen voll und ganz erfüllt haben. Auf Basis der Testergebnisse wurden klare Optimierungsvorschläge erarbeitet. Die Startseite und das Navigationskonzept wurden überarbeitet. Die Filter- und Sortiermöglichkeiten auf der Hotelübersichtsseite wurden neu strukturiert und vereinfacht. Die Darstellung der Hotels wurde angepasst. Die Studie brachte darüber hinaus wichtige Erkenntnisse über das **Zusammenspiel zwischen Werbemittel** bzw. Google Adwords und der dazugehörigen **Landing Page** auf der Website. Auf Basis dieser Studienergebnisse und weiterer gezielter Analysen wurden verschiedene Varianten entwickelt, getestet und erfolgreich eingeführt.

Die vorgenommen Optimierungen der Website von Neckermann Reisen haben dazu geführt, dass sich die **Umsatzentwicklung** der Seite stabilisiert hat und zum Geschäftsjahresende wieder ein Zuwachs gegenüber dem Vorjahr verbucht werden konnte.

Darüber hinaus hat die Analyse der Website im praktischen Gebrauch und die Beschäftigung mit den Ergebnissen im Rahmen der Optimierung den Blick für die tatsächlichen Kundenbedürfnisse geschärft und damit einen wichtigen Beitrag zu Verbesserung der User Experience auf der Website geleistet.

Usability Tests sind bei der Thomas Cook AG zu einem festen Baustein bei der Entwicklung neuer Websites und Funktionalitäten geworden. Das projektbegleitende Usability-Testing erhöht zwar zunächst die Entwicklungskosten der Website – langfristig trägt es in der Regel aber zur Senkung der Kosten bei, da aufwändige Nachbesserungen der Vergangenheit angehören.

XI. Eyetracking – mit den Augen der Nutzer sehen

von Daniel R. Schmeißer

1. Eyetracking und Web-Usab: Anwendungsbereiche 265
2. Blickmessung: Kennzahlen und Parameter 271
3. Alternativen zur Blickmessung? 276
4. Fazit: Eyetracking als Bestandteil von UX-Tests 277
5. Literaturverzeichnis 278

Es gibt eine grundlegende Diskrepanz zwischen der Zielgerichtetheit, mit der das Design und Layout einer Website konzipiert wird und der Flüchtigkeit, mit der die Website später durch den Nutzer rezipiert wird. Verständlicherweise gehen Designer bei der Konzeption davon aus, dass sämtliche Bereiche einer Seite für den Nutzer relevant sind, mithin bestehen auch genaue Vorstellungen darüber, in welcher Reihenfolge die Seite im Idealfall angeschaut werden sollte. Die Realität sieht dagegen häufig anders aus: Webseiten werden flüchtig wahrgenommen, weite Teil einer Seite finden gar nicht den Fokus der Aufmerksamkeit, Texte werden nur im Scan-Modus überflogen – kurz: die Nutzer geben einer Seite nur wenige Sekunden Zeit, um sich zu orientieren und ihre Information zu finden. Gelingt es der Website nicht, die relevanten Informationen möglichst rasch zu transportieren, wendet sich der Nutzer ab, klickt weiter oder verlässt die Seite im schlimmsten Fall ganz. Die Analyse der Orientierungsmuster, Blickverläufe und Aufmerksamkeitsschwerpunkte ist daher unumgänglich, wenn Sie Ihre Website hinsichtlich Design und Layout vollständig optimieren wollen.

Eyetracking Studien können heute – sieht man einmal von den hohen Anschaffungskosten der Geräte ab – mit geringem Zeitaufwand schnell und kostengünstig im Rahmen von UX-Tests durchgeführt werden. Werden Eyetracking Studien von spezialisierten Forschungsinstituten durchgeführt, betragen die **Mehrkosten** für die Durchführung und Analyse von Blickverlaufsdaten in der Regel nicht mehr als **10 bis 20 Prozent des Gesamtpreises einer UX-Studie** (je nach Umfang der aggregierten Analysen). Die Durchführung von Blickverlaufsanalysen (Eyetracking) erlaubt dafür aber eine präzisere Ableitung von Handlungsempfehlungen auf Basis objektiver Aufmerksamkeitsdaten. Da diese nur sinnvoll in Kombination mit Befragungs- und Interaktionsdaten interpretiert werden können (mehr dazu weiter unten), ist dieses Kapitel als ergänzendes Unterkapitel zu UX-Tests konzipiert.

Der Einsatz **berührungsloser** Eyetracker (siehe Abbildung 77), bei denen die Augenkamera in den Monitorrahmen eingebaut ist sowie automatisierte Auswertungstools erlauben einen hohen Grad an Forschungsökonomie und machen das Verfahren für die Anwendung in UX-Tests interessant. Zu Recht wird von den professionell durchführenden Eyetracking-Experten aber immer wieder darauf hingewiesen, dass gerade in Folge des zeitlichen Drucks in der kommerziellen UX-Forschung die Auswertung und Interpretation von Blickdaten häufig simplifiziert und wenig inspiriert erfolgt. Gerade die hohe Augenschein-Validität der Blickdaten kann zu fehlerhaften Schlussfolgerungen führen bzw. verhindern, dass die Aussagekraft der erfassten Rezeptionsprozesse voll ausgeschöpft und adäquat auf die Fragestellung angewendet wird. Im Folgenden sollen daher die relevanten Anwendungs-

bereiche sowie sinnvolle Kennzahlen und Indikatoren für Eyetracking-Studien in der UX-Forschung vorgestellt werden.

Abbildung 77: Kontaktloser Eyetracker der Firma Tobii. Links: Untersuchungssituation beim Surfen, zeitsynchron werden alle Klicks und Blickbewegungen der Nutzer erfasst (Quelle: phaydon).

1. Eyetracking und Web-Usab: Anwendungsbereiche

In der (medien-)psychologischen Grundlagenforschung ist die Aussagekraft des Eyetrackings im Hinblick auf menschliche Prozesse der Aufmerksamkeit und Informationsverarbeitung unumstritten.[59] In der Usability-Forschung wird die Nützlichkeit des Verfahrens trotzdem noch gelegentlich in Frage gestellt, weil die Ansicht vertreten wird, dass sich die Güte und Vollständigkeit der Handlungsempfehlungen durch den Einsatz von Eyetracking nicht signifikant verbessern lasse.[60] Damit verhält es sich ähnlich wie mit der Fünf-bis-acht-Probanden-Regel: Wenn Sie sich mit 80 Prozent Optimierungspotenzial zufrieden

[59] Siehe den allgemeinen Überblick bei Bente, Gary: Erfassung und Analyse des Blickverhaltens. In: Mangold, Roland/Vorderer, Peter/Bente, Gary (Hrsg.): Lehrbuch der Medienpsychologie, Göttingen 2004 u. a. S. 297–324.

[60] Vgl. Burmester, Michael/Weinhold, Thomas (2007): Liefert Eye-Tracking einen Mehrwert im Rahmen formative Usability-Tests? In: i-com, Zeitschrift für interaktive und kooperative Medien, Heft 2/2007, S. 39-43.

geben, dann sollten Sie darauf verzichten. Wenn Sie aber genau wissen wollen, wie die Nutzer Ihre Website wahrnehmen und rezipieren, empfiehlt sich ihr Einsatz, denn es gibt kein adäquateres Verfahren, dass Ihnen mit einer solchen Genauigkeit Aussagen über die Treffsicherheit Ihres Designs und Layouts liefern kann. Aber was genau hat das Eyetracking mit Web-Usability zu tun?

Zunächst einmal ist Aufmerksamkeit grundsätzlich ein selektives Phänomen, das angesichts der Komplexität visueller Stimuli darüber bestimmt, was wir bewusst wahrnehmen. Dieser Filterungsprozess hat eine ordnende und Orientierung stiftende Funktion, denn das Auge wählt aus sämtlichen Informationen nur diejenigen aus, die am vielversprechendsten scheinen und gleichzeitig handlungsrelevant sind. Von diesen sogenannten **Top-Down-** oder schemagesteuerten Aufmerksamkeitsprozessen werden die **Bottom-Up-Prozesse**, also Stimulus-gesteuerte Prozesse der Informationsverarbeitung unterschieden.[61] Bei letztgenannten sind es vor allem die Eigenschaften der Website (Farben, Formen, Kontraste etc.), die das Blickverhalten der Nutzer unwillkürlich steuern. Die beiden Arten der Informationsaufnahme und -verarbeitung verlaufen aber nicht isoliert, sondern parallel zueinander und beeinflussen sich wechselseitig. Die mit Aufmerksamkeit belegten Informationen werden dabei intern repräsentiert und bilden ein Bild von dem, was wir sehen; gleichzeitig wird damit festgelegt, welche Informationen weiterverarbeitet werden und welche Folgehandlungen eingeleitet werden – insofern bleibt alles, was nicht in den Fokus der Aufmerksamkeit gerät, außen vor und hat subjektiv keinerlei Relevanz.

Im Zusammenhang mit der Untersuchung von User Experience ist entscheidend, dass die visuelle Verarbeitung von Website-Informationen (egal ob Text oder Bild) schnell und unwillkürlich abläuft, aber weitere Handlungen vorbereitet, das heißt bevor irgendwo geklickt wird (Handlung), wird der Link bereits im peripheren Gesichtsfeld bemerkt und unser Gehirn beginnt unwillkürlich mit der Vorverarbeitung dieser Information. Die Anteile der Aufmerksamkeit, die unbewusst, unwillkürlich und spontan (implizit gesteuert) ablaufen, sind also deutlicher höher als die bewussten Anteile kognitiver Verarbeitung wie sie zum Beispiel beim Lesen von Texten auftritt. Die Aufmerksamkeit und Analyse von Blickdaten ist daneben aber auch von der Intention des Nutzers bzw. von den Web-spezifischen Aufgabenstellungen (Use Cases) und Phasen der Informationsverarbeitung abhängig. Dabei lassen sich grob drei Bereiche unterscheiden, die für die UX-Forschung relevant sind und aus denen jeweils unterschiedliche Ergebnisse und Optimierungshinweise resultieren. Sie

[61] Siehe Bucher, Hans-Jürgen/Schumacher, Peter (2006): The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. In: Communications 31, S. 347-368.

beschreiben zugleich die drei zentralen Anwendungsbereiche für das Eyetracking bei der Optimierung von Websites innerhalb des UX-Testings.

Orientierung (Spontaneous Looking): Optimierung von Layout-Entwürfen

Wer hat es nicht schon einmal erlebt: Sie kommen auf eine Website, versuchen sich einen Überblick zu verschaffen und verlassen die Seite, enttäuscht nicht das richtige gefunden zu haben, wieder nach wenigen Sekunden. Was ist passiert? Beim Erstkontakt mit einer Seite (häufig die Startseite oder eine Landing Page) verschaffen sich die Nutzer in einer sogenannten Orientierungsphase einen Überblick und selektieren bereits Inhalte hinsichtlich ihrer Bedeutung für die weiteren Nutzungsziele. Die Orientierungsphase dauert in der Regel nur wenige Sekunden, in denen das Auge in schnellen Sprüngen die verschiedenen Seitenbereiche scannt. Die Blickbewegungen in der Orientierungsphase verlaufen spontan und unwillkürlich und werden sowohl über gelernte Schemata (top-down) als auch durch die Gestaltungselemente des Stimulus (bottom-up) gesteuert. (Schemata äußern sich beispielsweise in der typischen Orientierung an der linken Navigationsleiste, das heißt es handelt sich um relativ starre Wahrnehmungsmuster über gelernte Positionen und Gestaltungsmerkmale typischer Seitenelemente.) Auf diese Weise legen die Nutzer in den ersten Sekunden auf einer Website einen individuellen Scanpfad der Seite an, der aus einer festen Abfolge von Fixationen besteht. Dieser dient als eine Art „Landkarte" im Hinblick auf die zentralen Orientierungs- und Steuerelemente und wird bei erneutem Kontakt mit der Seite wieder durchlaufen.[62]

Abbildung 78 zeigt einen typischen (aggregierten) Scanpfad auf der Amazon-Startseite. Die Zahlen in den roten Kreisen geben die typische Blickreihenfolge wieder. Sie bedeuten aber nicht, dass alle Personen genau in dieser Reihenfolge geschaut haben. Die Größe des Kreises indiziert die **durchschnittliche Verweildauer** in dem jeweiligen Areal im Erstkontakt, sagt also etwas aus über die Attraktionsfähigkeit eines inhaltlichen Elements auf der Seite (Area of Interest). Die Lage der Kreise gibt den **Schwerpunkt der Aufmerksamkeit** innerhalb des Areals an. Typischerweise zeigt sich, dass sich die Nutzer häufig von der Seitenmitte aus einen Überblick verschaffen. Auch bei der Amazon-Startseite ist dies der Fall; von da aus wandert der Blick aber schnell auf das Sucheingabefeld, dessen Positionierung häufig gelernt ist und das bei Amazon die am meisten genutzte Funktion darstellt. Erst nach kurzem Blick auf die rechte Spalte orientieren sich die Nutzer an der linken Hauptnavigation. Von da aus werden dann weitere Seitenelemente überwiegend flüchtig und im unteren und rechten Bereich angeschaut.

[62] Noton, David (1970): A Theory of Visual Pattern Perception. In: IEEE transactions on systems science and cybernetics, Heft 6/1970, S. 349-357.

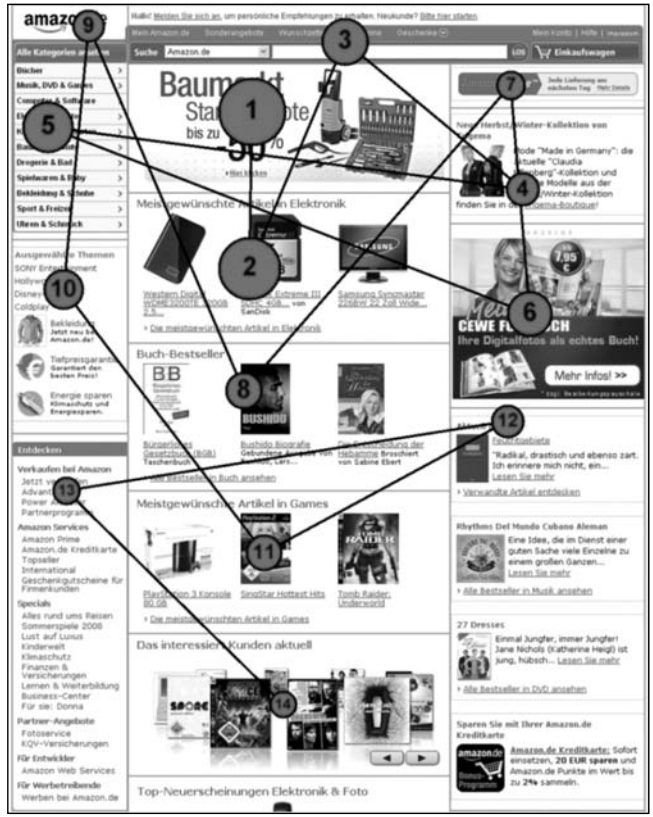


Abbildung 78: Scanpfad der Amazon-Startseite; kumulierte Auswertung von N=40 Teilnehmern. Erläuterung siehe Text (Quelle: phaydon)

Der Scanpfad dient im weiteren Verlauf als Schema, das die Orientierung der Nutzer auf den nachfolgenden Seiten erleichtert; daher ist es zum Beispiel wichtig, dass sich die zentralen Navigationselemente über die Seitenebenen hinweg immer an derselben Position befinden (Stichwort „Erlernbarkeit"). Da der Scanpfad spontan und unwillkürlich entsteht (ohne bewusste kognitive Steuerung), haben visuelle Gestaltungsmerkmale (Form, Farbe, Konturenstärke, Bewegung, Nähe und Distanz von Elementen) einen erheblichen Einfluss auf seinen Verlauf und können gezielt für eine intuitive Blickführung optimiert werden. Scanpfadanalysen eignen sich daher besonders für die Optimierung von Einstiegsseiten, aber auch für den Vergleich unterschiedlicher Layout-Entwürfe in der Konzeptphase, indem sie dem Designer darüber Feedback geben, ob und wie die visuelle Gestaltung der Seite für einen möglichst idealen Blickverlauf optimiert werden kann.

Informationssuche (Task-Oriented Looking): Optimierung funktionaler Visualisierungen

Bei komplexeren UX-Studien geht es in der Regel um das vollständige Interaktionsgeschehen, das im Rahmen von Use Cases abgebildet wird. Mittels kontaktloser Eyetracker lassen sich während des Lösens von Nutzungsaufgaben die Blickbewegungen, Scrolling-Aktivitäten und Klicks zeitsynchron zu den Interaktionsprotokollen (Seitenaufrufe, Verweildauer, Seitenkontakte) aufzeichnen. Die Blickverläufe geben so ein anschauliches Bild von dynamischen Aufmerksamkeits- und Suchprozessen während der Nutzung, indem nachvollziehbar wird, wie sich die Nutzer auf den Seiten orientieren, wo sie relevante Informationen vermuten und an welchen Stellen sie gegebenenfalls scheitern, weil Links nicht zielführend sind oder relevante Informationen schlichtweg übersehen werden.

Die Fragestellungen, die mit Hilfe des Eyetrackings im Rahmen von UX-Tests bearbeitet werden können, sind daher entsprechend vielfältig. Da die Aufzeichnung aber in der Regel während des gesamten Surfprozesses erfolgt, können einzelne Seiten, die sich im Rahmen der Surfpfadanalyse als besonders interessant herausstellen (zum Beispiel weil sie häufig aufgesucht werden oder besondere Schwierigkeiten bereiten), auch nachträglich gezielt ausgewählt und aggregiert ausgewertet werden. Typische Fragestellungen sind beispielsweise:

- Werden die relevanten Links und Hinweise schnell genug wahrgenommen?
- Ermöglicht die Navigation eine intuitive Nutzerführung über verschiedene Seitenebenen hinweg?
- Wie viele Menüpunkte darf eine Navigation oder ein Drop-Down-Menü umfassen, damit sie noch gelesen werden?
- Ziehen speziell platzierte Hinweise, Produkt-Teaser und Werbemittel ausreichend Aufmerksamkeit auf sich?
- Werden relevante Hinweise zur Bestellung (Warenkorb-Button oder Ähnliches) gesehen?
- Sind die relevanten Suchergebnisse so gestaltet und positioniert, dass sie leicht und unkompliziert gefunden werden können?

Entscheidend bei der Beantwortung der Fragestellungen ist, dass die Analyse und Optimierung prozessbezogen erfolgt, das heißt unter Berücksichtigung der typischen Kontextbedingungen, der Navigationspfade und der Aufmerksamkeitsmuster. Erst das Zusammenspiel dieser Komponenten erlaubt eine fundierte Ableitung von Handlungsempfehlungen für die Optimierung, denn die Ergebnisse aus Blickverlaufsanalysen allein sind zunächst wertneutral, das heißt weder positiv noch negativ. Erst vor dem Hintergrund der spezifischen

Aufgabenstellungen, die der Nutzer bewältigen möchte, gewinnen Aufmerksamkeitsdaten an Bedeutung.

Zentral ist daher auch die Berücksichtigung des spezifischen Surf-Modus. Blickbewegungen verlaufen anders, wenn Nutzer gezielt ein Produkt suchen („Goal-Mode"), unspezifisch nach Themen recherchieren oder einfach nur im Web stöbern („Activity-Mode"). Kurzum: Aggregierung, Auswertung und Interpretation von surfbegleitenden Blickdaten sind äußerst komplex und sollten in jedem Fall mit Unterstützung spezialisierter Forschungsinstitute durchgeführt werden.

Informationsverarbeitung (Information-Oriented Looking): Optimierung der Content-Usability

Ein weiteres, häufig vernachlässigtes Anwendungsfeld der Blickverlaufsforschung ist die visuelle und redaktionelle Gestaltung der Inhalte. Website-Inhalte verfügen über eine gute Content-Usability, wenn sie leicht lesbar, intuitiv erfassbar, ästhetisch aufbereitet und anwenderfreundlich gestaltet sind. Hier geht es also um die Frage, wie sich der Content auf einer Seite so optimieren lässt, dass Nutzer ihn als möglichst angenehm erleben. Die Aufmerksamkeitsanalyse gibt hier zum Beispiel Aufschluss darüber,

- wie intensiv die Inhalte und Texte gelesen werden
- in welcher Reihenfolge einzelne Textbestandteile gelesen werden
- welche Bedeutung Headlines, Bilder und Teasertexte für den Einstieg haben
- mit welchen Bildern gearbeitet werden sollte, damit die Seite über Faszinations- und Aufmerksamkeitspotenzial verfügt
- wie ausgewogen die Relation zwischen Text und Bild sein sollte.

Entscheidend für die Analyse Content-bezogener Fragestellungen ist das Ineinandergreifen spontaner und zielgerichteter Aufmerksamkeitsprozesse. Vieles von dem, was wir lesen, nehmen wir flüchtig wahr; anderes lesen wir aufmerksam, weil es uns interessiert oder fasziniert. Ein globales Maß, mit dem sich Interesse messen lässt, ist daher die Verweildauer auf einem Element. Die Verweildauer allein genügt aber nicht, denn eine lange Verweildauer auf einem Areal kann nicht nur Interesse, sondern gegebenenfalls auch Irritation oder Unverständnis signalisieren. Man benötigt also weitere Kennzahlen über das Blickverhalten, um zuverlässige Aussagen über das tatsächliche Rezeptionsverhalten treffen zu können. Mit diesen beschäftigt sich daher das nächste Kapitel etwas ausführlicher.

2. Blickmessung: Kennzahlen und Parameter

Das Thema Blickmessung ist äußerst komplex, deshalb stehen im Folgenden nur die UX-relevanten Fakten im Vordergrund. Wichtig ist zunächst das foveale Sehen vom peripheren Sehen zu unterscheiden, denn zu Recht wird immer wieder der Einwand erhoben, dass wir irgendwie ‚mehr' sehen als das, was uns bewusst in den Fokus kommt. Das foveale Sehen ist für die detaillierte Wahrnehmung unserer Umwelt zuständig – nur im Bereich der Sehgrube (fovea centralis) befinden sich ausreichend Zapfen, um ein hohes farbiges Auflösungsvermögen sicherzustellen. Das periphere Sehen findet außerhalb der Fovea statt und verfügt über eine sehr viel geringere Sehschärfte und ist weniger farbintensiv. Trotzdem ist das periphere Sehen wichtig, denn es dient dazu, die Ausrichtung unserer Aufmerksamkeit zu steuern, indem bereits Konturen oder Teile des visuellen Stimulus erkannt und die nächsten Blicksprünge (Sakkaden) vorbereitet werden.

Die eigentliche Verarbeitung visueller Informationen findet aber erst dann statt, wenn das Auge für eine kurze Zeit in Ruhe verharrt und den Reiz fixiert. Solche Fixationen sind der eigentliche Gegenstand der Blickmessung. Eine **Fixation** dauert **mindestens 100 Millisekunden**, schwankt zwischen Person und Reizsituation und wird im Mittel je nach Aufgabe zwischen 100 und 600 Millisekunden angegeben. Es gibt keine allgemeinverbindliche Definition für die Dauer einer Fixation, andererseits gibt aber bereits die unterschiedliche Länge einer Fixation Auskunft über die Art der zugrundeliegenden Informationsverarbeitung. Kurze Fixationslängen deuten beispielsweise auf die flüchtige Rezeption von bildhaftem Material hin (sogenannte „Express-Fixationen"), während längere Fixationen von mindestens 200 Millisekunden für das Lesen von Texten erforderlich sind. Zwischen den Fixationen richtet sich das Auge immer wieder erneut durch schnelle Blicksprünge aus. Mit diesen als Sakkaden bezeichneten Blickbewegungen springt das Auge von einem Ort zum anderen, zwischendurch finden Fixationen statt, um einen Ausschnitt genauer zu betrachten. Obwohl die Schnelligkeit und Häufigkeit von Sakkaden ebenso informativ ist (zum Beispiel für den Grad der Komplexität eines Stimulus), richtet sich die Analyse von Blickdaten in der Regel auf die Fixationen, denn nur hier findet Informationsverarbeitung statt. Während der Sakkaden sind wir „blind".

Die Messung der Fixationen erfordert auf Seiten der Messtechnik ein hohes Maß an Auflösung und Genauigkeit. Diese werden durch moderne Eyetracking-Systeme weitgehend sichergestellt. Zum Anderen ist die Erhebung der Blickdaten in einer möglichst natürlichen Nutzungssituation erforderlich, woraus sich die Anforderung an berührungslose Messgeräte mit einfachen Kalibrierungsprozeduren ergibt. Trotzdem treten immer wieder eine

Reihe von Messfehlern auf, die nur mit entsprechender Kenntnis und den richtigen Vorkehrungen vermieden werden können. Da es keine etablierten Best Practices für die Berechnung von Fixationen, für die Messauflösung der eingesetzten Eyetracker oder für einzelne Berechnungsalgorithmen gibt, diese aber die Kennwerte und die Berechnung durchschnittlicher Verweildauern massiv beeinflussen, sollten die Parameter nach Möglichkeit (und spätestens auf Nachfrage) offen gelegt und transparent gemacht werden.

Grundsätzlich können Blickdaten bei quantitativen Analysen über drei Ebenen hinweg aggregiert werden (a) über räumliche Areale, (b) über Nutzergruppen und (c) über Zeitphasen.[63] Für alle drei Ebenen der Datenaggregierung können anschließend Aussagen hinsichtlich verschiedener Parameter getroffen werden. Grundlegend ist zunächst die **Aggregierung nach räumlichen Arealen** (Areas of Interest = AoI). Hierbei wird die jeweilige Seite in einzelne Elemente aufgeteilt, für die die Aufmerksamkeitsparameter ausgewiesen werden können. Die Erstellung der AoIs ist flexibel und orientiert sich gewöhnlich an den inhaltlich bedeutsamen Elementen, wobei eine flächenmäßige Aufteilung oder Unterteilung zusammengehöriger Elemente je nach Fragestellung auch sinnvoll sein kann. Die **Aggregierung nach Nutzergruppen** erlaubt eine Analyse der Blickdaten nach verschiedenen Personengruppen. So kann es sinnvoll sein, die Scanpfade von Kennern (Kunden) einer Seite mit denen der Neueinsteiger zu vergleichen, um Rückschlüsse auf die initialen bzw. gelernten Orientierungsmuster zu erhalten: So lässt sich zum Beispiel im Vorfeld eines Relaunches erfassen, ob die neuen Strukturelemente auf der Seite von den bisherigen Nutzer intuitiv verstanden werden oder inwieweit bisher gelernte Schemata einer erneuten Aneignung entgegenstehen; bei Neueinsteigern ginge es hingegen um die intuitive visuelle Orientierung im Erstkontakt.

Die **Aggregierung nach Zeitphasen** wird den unterschiedlichen Fragerichtungen und Interessensschwerpunkten gerecht: Interessieren Sie sich für die frühen Orientierungsmuster in den ersten fünf Sekunden, oder möchten Sie etwas über das Leseverhalten im Intensivkontakt mit einer Seite erfahren? Moderne Auswertungstools gewährleisten problemlos eine dynamische und kombinierte Aggregierung aller drei Ebenen.

Es existiert eine ganze Reihe von Kennwerten, von denen aber nur wenige in der angewandten UX-Forschung zum Einsatz kommen. Die wesentlichen Maße beziehen sich vor allem auf das globale Aufmerksamkeitsverhalten auf einer Website. Diese betreffen die durchschnittliche Verweildauer, die Kontakthäufigkeit, Übergangswahrscheinlichkeiten

[63] Vgl. Schmeißer, Daniel/Sauer, Michael (2003): Mit den Augen der User sehen. In: planung & analyse, Heft 4/2003.

von einem in das andere Areal sowie Anzahl und Dauer von Fixationen. In der Praxis und für die Belange der Website-Optimierung haben sich folgende Kennzahlen und Visualisierungsmöglichkeiten durchgesetzt:

- Scanpfade
- Individuelle Blickspuren („Alphamaps“)
- Heatmaps
- Farbflächenkodierungen (Verweildauer, Kontakthäufigkeit und andere)
- Verlaufs-Presence-Kurven

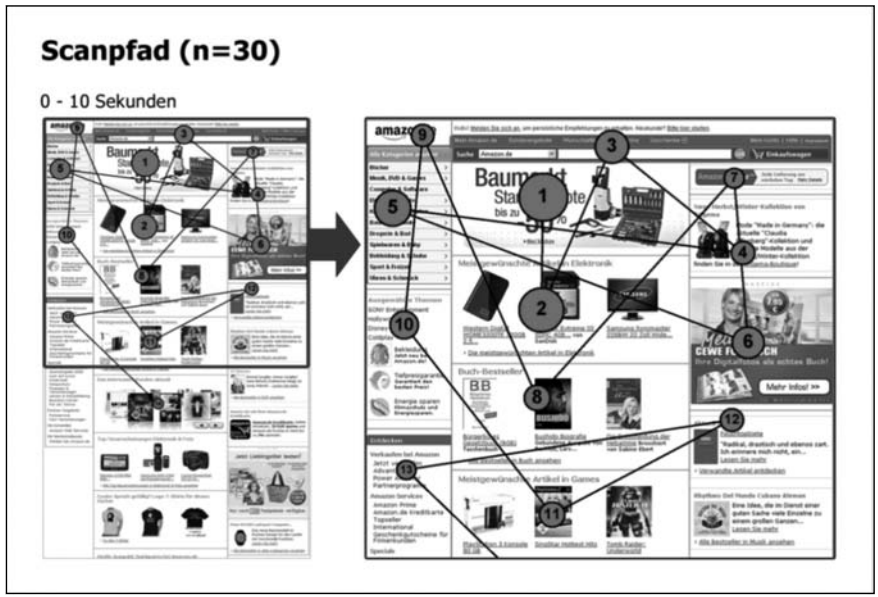


Abbildung 79: Die Zahlen in den roten Kreisen geben die typische Blick-Reihenfolge wieder (dies bedeutet also nicht, dass alle Betrachter in exakt dieser Reihenfolge geschaut haben). Die Größe des Kreises indiziert die durchschnittliche Verweildauer im Erstkontakt mit dem jeweiligen Areal. Die Position des Kreises bedeutet nicht, dass nur auf diesen Punkt geschaut wurde, sondern zeigt den Schwerpunkt der Fixationen innerhalb des Areals an (Quelle: phaydon).

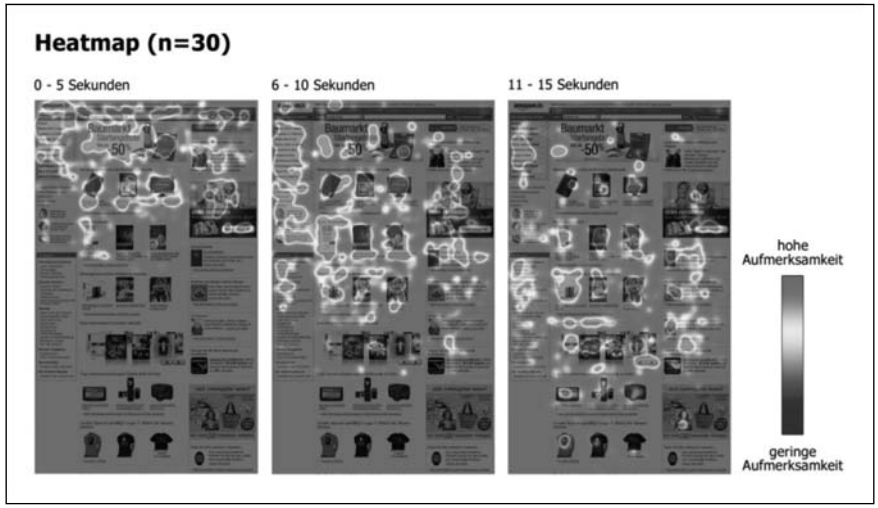


Abbildung 80: Individuelle Blickspuren beim ersten Betrachten einer Website liefern Hinweise auf das Orientierungsverhalten der Nutzer (Quelle: phaydon).

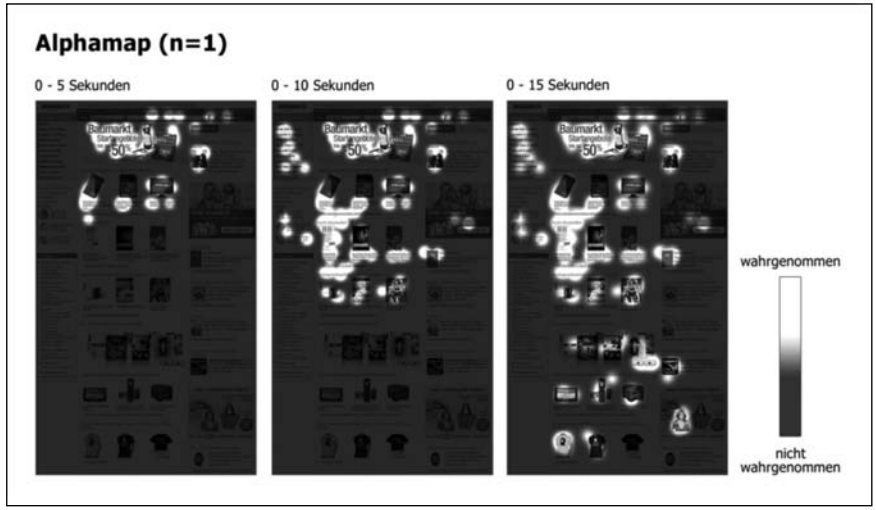


Abbildung 81: Die aggregierte Blickspur mehrerer Nutzer wird mittels Heatmap-Darstellung (hell = intensive Beachtung; dunkler = geringe Beachtung) dargestellt. So wird sichtbar, welche Aufmerksamkeit die verschiedenen Seitenareale im zeitlichen Verlauf erhalten (Quelle: phaydon).

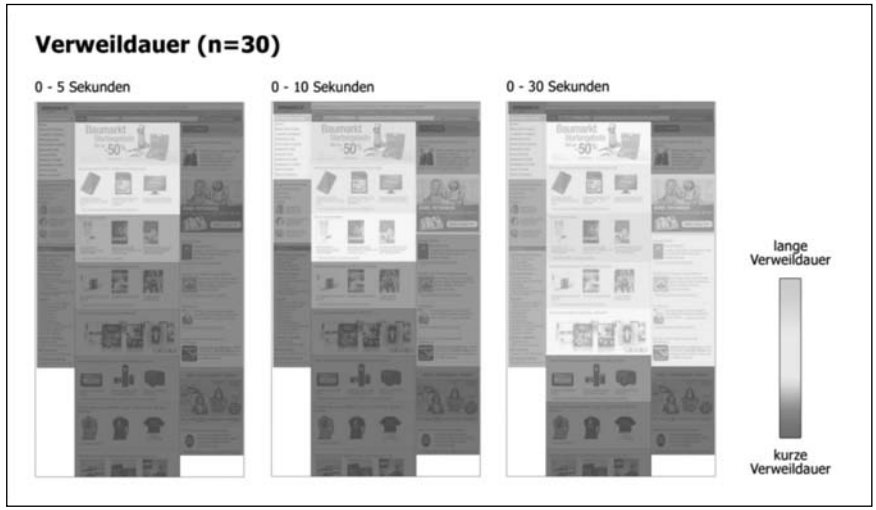


Abbildung 82: Mittels Farbflächenkodierung können unterschiedliche Blickparameter (hier: durchschnittliche Verweildauer in Seitenarealen) visualisiert und nach Zeitverläufen oder Nutzergruppen ausgewertet werden (Quelle: phaydon).

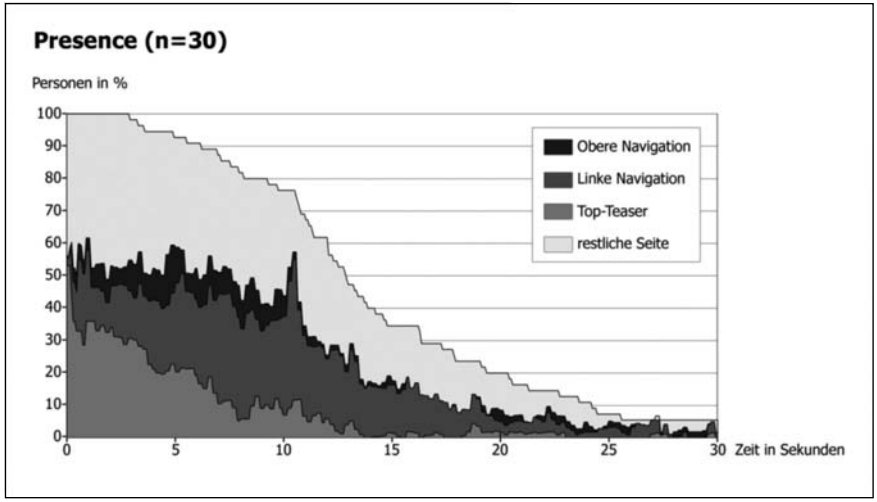


Abbildung 83: Stapeldiagramme (sogenannte „Presence-Kurven“) stellen die Aufmerksamkeitsverläufe auf einer Seite im zeitlichen Verlauf dar. So wird sichtbar, wie sich die Nutzer orientieren bzw. in welchen Seitenbereichen sie zum Beispiel im Rahmen eines Use Cases nach Produkten oder weiterführenden Informationen suchen (Quelle: phaydon).

3. Alternativen zur Blickmessung?

Um es kurz zu machen: Es gibt keine echte Alternative zur klassischen Blickbewegungsmessung. Es gibt lediglich Kompromisse, die es im Idealfall erlauben, schneller und günstiger die Aufmerksamkeit auf Webseiten zu erfassen. Der Vollständigkeit halber sollen drei alternative Verfahren vorgestellt werden.

Site-Covering

Beim Site-Covering wird die einzelne Seite in inhaltliche Felder unterteilt (parallel zu den AoIs), die jeweils mit leicht transparenten Kärtchen (oder online mittels Verwischen der Felder) abgedeckt werden.[64] Die Teilnehmer erkennen also nur schemenhaft, was sich hinter den Kästchen verbirgt, können die Inhalte aber durch Anklicken aufdecken. Beim Anklicken einer neuen Fläche wird die jeweils zuvor gezeigte wieder verdeckt. Das Verfahren zielt darauf ab, die Erwartungen von Usern an eine Benutzeroberfläche zu erfassen und wird häufig als Ergänzung zum lauten Denken eingesetzt. Da es sich aber im Hinblick auf die tatsächlichen Wahrnehmungsprozesse artifiziell verhält, können mit diesem Verfahren lediglich erlernte Schemata aufgedeckt, aber keine spontanen Aufmerksamkeitsverläufe erfasst werden.

Attention-Tracking

Beim Attention-Tracking dient die Maus als Anzeigegerät für die Aufmerksamkeit: Die Probanden geben über schnelle Mausklicks an, welche Elemente einer Seite ihre Aufmerksamkeit erregen, wobei Mausklicks den Fixationen der Blickmessung entsprechen.[65] Durch die Gleichsetzung von Mausklicks mit Fixationen ist eine automatisierte Analyse der Daten möglich, zugleich kann das Verfahren auch online schnell und kostengünstig bei größeren Stichproben angewendet werden.

Der entscheidende Nachteil für die UX-Forschung liegt (neben der generellen Voraussetzung einer höheren Internet-Affinität) darin, dass bei Websites nur einzelne Screens gezeigt werden können und nicht das natürliche Nutzungsverhalten abgebildet werden kann. Eine weitere Schwierigkeit liegt in der Erfassung der Leseaktivität, die durch das schnelle Klicken auf aufmerksamkeitsstarke Areale verhindert wird. Studien belegen vor allem bei

---

[64] Vgl. Wilhelm, Thorsten H./Berger, Dorit/Yom, Miriam (2002): Site Covering: Eine innovative Methode zur Erfassung der Informationsaufnahme und des Entscheidungsverhaltens auf Webseiten. In: planung & analyse, Heft 2/2002. S. 46-50.

[65] Vgl. Scheier, Christian/Egner, Steffen (2005): Beobachten statt Fragen! In: planung & analyse, Heft 1/2005, S. 53-58.

Messungen in der frühen Orientierungsphase eine deutliche Abweichung der Ergebnisse zwischen herkömmlicher Blickmessung und Attention-Tracking.[66] Bei Klickdaten gehen Informationen über ein anfängliches Such- und Orientierungsverhalten der Nutzer verloren, die in den Blickdaten durchaus vorhanden sind.

Restricted-Focus-Viewer

Auch bei diesem Verfahren handelt es sich um einen Ansatz, der vor allem für die Belange der Online-Marktforschung praxisnah entwickelt wurde.[67] Ähnlich wie beim Site-Covering ist der Stimulus (zum Beispiel eine Anzeige) zunächst stark verwischt, analog zu den Eindrücken im peripheren Gesichtsfeld. Es sind also keine Details erkennbar, aber man kann sich soweit orientieren, dass entschieden werden kann, welche Elemente es wert sind, genauer betrachtet zu werden. Der Mauszeiger fungiert dabei als der Scheinwerfer, der die Bildauflösung durch Bewegungen des Mauszeigers steuert, das heißt nur am Ort des Mauszeigers bzw. im kleinen Umkreis des Mauszeigers erscheint die Seite scharf. Durch die Aufzeichnung der Mausbewegungen werden die Bereiche des scharfen Sehens (analog zum fovealen Sehen) erfasst und in entsprechende Kennwerte umgerechnet.

Der Vorteil dieses Verfahrens ist, dass es dem tatsächlichen Sehen durch Nachbildung des peripheren und fovealen Sehens nahe kommt und online auch auf größere Stichproben angewendet werden kann. Aufgrund der Tatsache, dass sich die Betrachter aber bereits auf der verschwommenen Vorlage orientieren müssen, werden die spontanen Orientierungssprünge und unbewussten Aufmerksamkeitsprozesse nicht valide erfasst. Insgesamt eignet sich dieses Verfahren mehr für statische Vorlagen (zum Beispiel Anzeigen) und weniger für UX-Studien, da im dynamischen Nutzungsprozess die Verwendung des Mauszeigers als Scheinwerfer keinen Sinn macht (denn die Nutzer klicken und schauen gewöhnlich nicht mit der Maus).

4. Fazit: Eyetracking als Bestandteil von UX-Tests

Die dargestellten Verfahren machen deutlich, dass es ökonomische Alternativen zur Erfassung von Aufmerksamkeitsverläufen gibt. Diese gelangen aber schnell an ihre Grenzen, insbesondere wenn es um die Erfassung des natürlichen Nutzungsverhaltens bei dyna-

[66] Bente, Gary et al. (2007): Im Auge des Nutzers. Eye-Tracking in der Web-Usability-Forschung. In: Welker, Martin/Wenzel, Olaf (Hrsg.): Online-Forschung 2007. Grundlagen und Fallstudien. Köln, S. 210-215.

[67] Vgl. zum Beispiel Zellner, Martina/Kaufmann, Gwen/Hartmann, Adriane (2007): Versicherungsanzeige mit Humor. Bleibt noch genügend (Wahrnehmungs-) Zeit fürs Branding? In: planung & analyse, Heft 2/2007. S. 32-36.

mischen Seiteninhalten oder um die Erfassung früher und spontaner Such- und Orientierungsverläufe geht. Sobald Prozesse der Aufmerksamkeit und Informationsverarbeitung im Vordergrund stehen, wird die UX-Forschung daher auch in Zukunft nicht ohne das aufwändigere, aber validere Verfahren der Blickmessung mittels Eyetracking auskommen können. Dass sich Eyetracking-Daten alleine nicht zur Identifikation von Usability-Problemen eignen, sondern ihre Aussagekraft hinsichtlich der User Experience erst durch den professionellen Einsatz und eine fundierte psychologisch-rezeptionsorientierte Analyse erhalten, dürfte ebenso deutlich geworden sein. Entscheidend ist, Blickverlaufsanalysen (Eyetracking) nicht losgelöst von den allgemeinen Fragestellungen und Verfahren eines User Experience Tests durchzuführen, denn erst in der Kombination mit qualitativen Befragungs- und Beobachtungsdaten können Blickverlaufsanalysen einen sinnvollen Beitrag zur Ableitung tragfähiger Handlungsempfehlungen leisten.

5. Literaturverzeichnis

Bente, Gary: Erfassung und Analyse des Blickverhaltens. In: Mangold, Roland; Vorderer, Peter; Bente, Gary (Hrsg.): Lehrbuch der Medienpsychologie, Göttingen 2004, u.a. S. 297-324.

Bente, Gary et al.: Im Auge des Nutzers. Eyetracking in der Web-Usability-Forschung. In: Welker, Martin; Wenzel, Olaf (Hrsg.): Online-Forschung 2007. Grundlagen und Fallstudien. Köln 2007, S. 185-219.

Bucher, Hans-Jürgen; Schumacher, Peter: The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. In: Communications 31 (2006), S. 347-368.

Burmester, Michael; Weinhold, Thomas: Liefert Eye-Tracking einen Mehrwert im Rahmen formative Usability-Tests? In: i-com, Zeitschrift für interaktive und kooperative Medien, Heft 2/2007, S. 39-43.

Noton, David: A Theory of Visual Pattern Perception. In: IEEE transactions on systems science and cybernetics, Heft 6/1970, S. 349-357.

Scheier, Christian; Egner, Steffen: Beobachten statt Fragen! In: planung & analyse, Heft 1/2005, S. 53-58.

Schmeißer, Daniel; Sauer, Michael: Mit den Augen der User sehen. In: planung & analyse, Heft 4/2003.

Wilhelm, Thorsten H.; Berger, Dorit; Yom, Miriam: Site Covering: Eine innovative Methode zur Erfassung der Informationsaufnahme und des Entscheidungsverhaltens auf Webseiten. In: planung & analyse, Heft 2/2002. S. 46-50.

Zellner, Martina; Kaufmann, Gwen; Hartmann, Adriane: Versicherungsanzeige mit Humor. Bleibt noch genügend (Wahrnehmungs-) Zeit fürs Branding? In: planung & analyse, Heft 2/2007. S. 32-36.

XII. Remote User Experience Tests Messen, befragen und beobachten in natürlicher Testumgebung

von Tim Bosenick

1. Definition Remote User Experience Tests 280
2. Das Testverfahren 282
3. Einsatzmöglichkeiten 283

Remote Usability-Tests von Internet-Anwendungen gewinnen immer mehr an Akzeptanz. In mehreren Studien konnte nachgewiesen werden, dass der reale und sehr breite Nutzungskontext sowie die größere Zahl der Testpersonen bei Remote-Tests zu mehr und anderen aufgedeckten Usability-Problemen führen als Labor-Tests (Kühn 2005, Nufer 2006). Vor allem das hohe Involvement der Testpersonen gilt als großer Vorteil. Es entsteht zum einen durch die reale Nutzungssituation im natürlichen Kontext der User und wird zum anderen dadurch gefördert, dass die Probanden auf der zu testenden Website das tun, was sie ohnehin vorhatten.

1. Definition Remote User Experience Tests

Remote Testing ist heute eine nicht mehr wegzudenkende Methode in der Toolbox der Usability-Tests. Grundsätzlich lassen sich zwei Ansätze unterscheiden: Das synchrone und asynchrone Remote Testing.

Synchrones Remote Testing

Die ersten Remote Tests wurden, ähnlich Usability-Tests im Labor, synchron durchgeführt. Die Testpersonen bestimmen hierbei den Ort für die Durchführung des Usability-Tests selbst, werden jedoch zeitgleich von einem Usability-Experten beobachtet. Der Testleiter führt am Telefon das Interview und verfolgt die Bildschirmeingaben des Probanden über ein spezielles Tool. So kann der Usability-Experte direkt nachfragen, wenn bestimmte Probleme auftreten und die Gründe dafür näher untersuchen. Die Telefongespräche und Screens der Testpersonen werden zur späteren Auswertung aufgezeichnet. Hat die Testperson eine Webcam installiert, ist auch der Gesichtsausdruck während der Nutzung des Testobjekts zu sehen. Dabei entspricht die Ergebnisqualität synchroner Remote-Tests der von Nutzertests im Usability-Labor. Mit beiden Testmethoden lassen sich die gleichen Usability-Probleme bezüglich Anzahl, Kategorien und Schweregrad nachweisen (vgl. Brush et al. 2004).

Asynchrones Remote-Testing

Auch bei einem asynchronen Remote Usability-Test bearbeiten oder nutzen die Probanden eine internetbasierte Anwendung zu Hause oder am Arbeitsplatz. Die Beobachtung des Tests erfolgt jedoch ebenso wie die Auswertung zeitlich versetzt nach der Test-Session.

Die Idee für diesen Ansatz entstand gemeinsam mit einem Kunden, für dessen E-Commerce-Webseite eine Reihe klassischer Usability-Tests im Labor durchgeführt wurde. Hierbei wurden einzelne Bereiche des Online-Shops, wie der Bestellprozess, die Artikeldetailseiten oder das Navigationsschema optimiert. Die Frage, wie gut im Shop tatsächlich verkauft wurde, konnte aber nie gänzlich zufrieden stellend beantwortet werden, da die Rahmenbedingungen im Labor gewisse Einschränkungen bieten. Zum einen lassen sich aufgrund der in der Regel eher kleinen Stichprobe keine repräsentativen Aussagen treffen bzw. wäre es aufgrund der relativ hohen Incentives sehr teuer, eine entsprechend große Stichprobe zu wählen. Zum anderen beabsichtigen die Testpersonen in einer Laborsituation nur selten einen realen Online-Kauf – selbst wenn sie einen entsprechenden Zuschuss erhalten oder die Absicht als Rekrutierungskriterium vorausgesetzt wird. Der Grund hierfür liegt in der für die Testpersonen künstlichen Atmosphäre – sie sitzen nicht zu Hause an ihrem Rechner und hinter dem Spiegel werden sie zudem vom Testleiter beobachtet. Dies kann unter Umständen dazu führen, dass das Involvement der Probanden eher gering ist und nicht der tatsächlichen Nutzung entspricht.

Im Gegensatz zu Labor-Tests zeichnen sich ethnografische Ansätze in der Markt- und Usabilityforschung dadurch aus, dass durch Beobachtung und Befragung im natürlichen Kontext der Testpersonen, die Daten eine höhere ökologische Validität enthalten. Das Verhalten der Nutzer in ihrer gewohnten Umgebung ist authentischer als im Labor. Folglich geben die Daten, die mit ethnografischen Ansätzen erhoben werden, das wirkliche Verhalten besser wieder als eine Online-Befragung oder ein Usability-Test im Labor. Bekannte server- und nutzerzentrierte Verfahren zur Leistungsmessung und Befragungen im Internet stoßen jedoch schnell an ihre Grenzen, wenn das natürliche Verhalten der Nutzer dokumentiert werden soll. Sie basieren in der Regel auf der Logfile-Analyse und liefern umfangreiche Kennzahlen zur Online-Nutzung bis hin zur Dokumentation von Nutzungspfaden. Der Nutzer bleibt dabei wie bei allen technischen Messverfahren zur Internetnutzung weitgehend unbekannt, da diese Verfahren keine Daten zu Nutzergruppen oder Nutzungszielen liefern können. Userbasierte Ansätze wie zum Beispiel Online-Panels füllen diese Lücke, haben aber den Nachteil, dass die Motivation der Nutzer offen bleibt. Sie beantworten zum Beispiel nicht, warum die Nutzer gerade auf einer bestimmten Internet-Seite surfen oder einen Vorgang abgebrochen haben. Hierfür eignen sich in besonderem Maße Online-Befragungen und Usability-Tests.

2. Das Testverfahren

Remote-Testing-Tools wie LEOtrace® verknüpfen die oben genannten unterschiedlichen Ansätze miteinander: Die technische Messung, die Befragung und die Beobachtung. Somit erfassen sie das gesamte Nutzerverhalten, die sogenannten Customer Experience: Nutzungsmotive, -zufriedenheit und -häufigkeiten eigener und fremder Angebote werden somit vergleichbar.

Wie bereits weiter oben ausgeführt, gewährleistet das Testverfahren, dass die Probanden in einer Alltagssituation das Internet an ihrem eigenen PC wie gewohnt nutzen und damit die Voraussetzung für qualitativ hochwertige und valide Ergebnisse geschaffen wird. Technisch lassen sich zwei Varianten unterscheiden: Der erste Ansatz ist server-basiert. Die Rekrutierung der Probanden erfolgt direkt online über die zu testende Website über ein Pop-Up oder ein Layer. Der Datenstrom wird über eine Art Proxy-Server umgeleitet, der die gesammelten Informationen der Nutzung speichert. Die User surfen nachdem sie die Bereitschaft zur Teilnahme erklärt haben, wie gewohnt auf der Website weiter und müssen keine weitere Installation vornehmen. Verlässt der Nutzer die Webseite, ist die Testsession beendet. Beim zweiten Ansatz erfolgt die Rekrutierung über ein Online-Panel. Hier installieren die Probanden ein Add-On für die Browser Internet Explorer oder Firefox. Wie beim server-basierten Ansatz zeichnet ein Server alle Nutzungsdaten der Probanden auf — dies ist über einen längeren Zeitraum hinweg und für mehrere Webseiten gleichzeitig möglich. So lässt sich zum Beispiel nachvollziehen, wie viele Webseiten ein Proband besucht, bevor er sich entschließt, ein spezielles Produkt online oder offline zu kaufen.

Die Online-Rekrutierung ist ein leichter und gut steuerbarer Zugang zu einem breiten Spektrum an Testpersonen sowie eine kostengünstige Variante der Rekrutierung. Die Auswahl der Testpersonen über ein Panel stellt sicher, dass die Testpersonen der realen Nutzerschaft oder, je nach Fragestellung, einem für die Studie interessanten Profil entsprechen. Je nach Aufgabenstellung der Studie bewegen sich die Testpersonen zeitlich und inhaltlich völlig frei auf der zu testenden Website oder bearbeiten vorgegebene Aufgaben. An jeder beliebigen Stelle einer zuvor definierten Website lassen sich über das Remote-Tool während des freien Surfens Online-Fragebögen einbinden. Bei beiden Ansätzen kann die Befragung vor oder nach sowie ausgelöst durch bestimmte Ereignisse, auch während einer Testsession stattfinden. Verlässt ein Proband zum Beispiel eine Produktdetailseite oder bricht den Bestellprozess ab, lässt sich direkt nachfragen, aus welchem Grund der Abbruch stattfand. Schließlich können die Testpersonen während des freien Surfens jederzeit über eine Leiste am unteren Bildschirmrand ihr positives und negatives Feedback abgeben.

Während der gesamten Nutzung zeichnet das Remote-Testing-Tool neben der URL der aufgerufenen Seiten, Screenshots, Mausbewegungen und Eingaben der User in die Website auf. Erfolgt die Rekrutierung der Testpersonen über ein Online-Panel, lassen sich die Nutzer, anders als bei Logfile-Analysen, die mit Session-Cookies arbeiten oder User-Agents, die Brower-Version, IP-Adresse und Bildschirmauflösung heranziehen, über das Add-On identifizieren. So lassen sich die Logfiles der Testpersonen mit den Daten der Befragungen wie Internetnutzung, Soziodemographie, Erwartungen an die Website, Besuchsgründe, Erwartungserfüllung, Abbruchgründe oder Bewertungen verknüpfen und aus den soziodemografischen Merkmalen, Einstellungen und Lebensstil der Teilnehmer exakte Nutzertypologien und Zielgruppen bestimmen. Über beide Ansätze lässt sich die Ausgabe bestimmter Inhalte oder Werbung gezielt steuern, um zum Beispiel in einem kontrollierten Umfeld zu überprüfen, wie welche Werbung auf welche Nutzergruppen wirkt.

Hieraus ergeben sich Einsatzmöglichkeiten für unterschiedliche Fragestellungen, die im folgenden Kapitel näher erläutert werden.

3. Einsatzmöglichkeiten

Wie die vorangegangenen Ausführungen zeigen, eignet sich ein Remote-Test besonders dann, wenn das authentische Verhalten der Nutzer einer Website abgefragt und gleichzeitig die Frage beantwortet werden soll, wie gut die Website den beabsichtigten Business-Case erfüllt. Der Auftraggeber einer Studie erfährt beispielsweise über die Kombination aus Messen und Beobachten mit welchen Erwartungen und Intentionen die Nutzer auf seine Seite kommen. Die Ergebnisse bieten in der Regel Optimierungspotenziale, wenn die Erwartungen noch nicht bekannt waren oder wenn sie nicht befriedigt wurden. In beiden Fällen lässt sich anhand der Daten nachvollziehen, was die Nutzer für ein Ziel verfolgten, wie sie es verfolgten und warum sie es beispielsweise nicht erreicht haben.

Weiter lässt sich mit dem sogenannten Behaviour Tracking nachvollziehen wie die eigene Website im Vergleich zum Wettbewerb verkauft und/oder informiert. Eine breit angelegte Studie beantwortet Fragen nach den Best Practices im unmittelbaren Wettbewerberumfeld, welche Websites für welche Prozessschritte genutzt werden oder wie die eigene Seite im Vergleich zum Wettbewerb von den Usern bewertet wird.

Die Daten, die über die Beobachtung gewonnen werden, stellen einen hohen Mehrwert zu Online-Befragungen dar: Viele Aspekte des Verhaltens im Internet können nicht sauber über eine Selbstauskunft erhoben werden, da die User oft schwer reflektieren können, wie sie normalerweise vorgehen – einige Verhaltensweisen werden über-, andere unterschätzt. Werden Nutzer zum Beispiel befragt, ob sie die Suchmaschine Google nutzen, wenn sie eine Reise buchen, bejahen dies 99 Prozent. Suchen sie dann tatsächlich online nach einer Reise sind es weitaus weniger, weil sie dabei anders vorgehen als sie es selbst dachten. Remote-Test schließen diese Lücke, da sie einerseits das Verhalten messen und andererseits zum Zeitpunkt der Nutzung nachfragen und nicht erst Tage oder Wochen danach.

Neben den umfangreichen qualitativen Daten aus der Befragung und Beobachtung, liefert ein Remote-Test aufgrund der hohen Stichprobe Kennzahlen, die über reine Zufriedenheits- oder Nutzerzahlenrankings hinausgehen.

Ein Remote-Test stößt jedoch auch an seine Grenzen: Per Gesetz ist es untersagt, in geschlossenen Bereichen wie zum Beispiel Online-Banking zu testen, wo persönliche und/oder sensible Daten abgefragt werden. Ein Labor-Test ist immer dann angezeigt, wenn das Testmaterial nur in Teilen zur Verfügung steht, schnelles Feedback von den Nutzern gefordert ist oder spezifische Fragestellungen auftauchen, die direktes Nachfragen oder Beobachten erfordern. Soll zum Beispiel ein Bestellprozess optimiert werden, steht nicht das persönliche Involvement der Testpersonen im Vordergrund, sondern die zu bewältigende Aufgabe. Allgemein lässt sich festhalten, dass Fragestellungen, die den Fokus auf die Usability von Webanwendungen legen, für einen Labor-Test besser geeignet sind als für einen Remote-Test. Während die Frage nach der Customer Experience, die den Nutzwert, das Design und den Joy of Use untersucht, besser von einem Remote-Test beantwortet wird. Beide Testverfahren gemeinsam bieten sich folglich immer dann an, wenn entsprechend kombinierte Fragestellungen beantwortet werden sollen.

XII.I Fallbeispiel: Remote Testing DER CLUB

von Tim Bosenick

Die folgenden beiden Anwendungsbeispiele zeigen das Einsatzspektrum des Remote-Testing. Beide Studien kombinierten Messung, Befragung und Beobachtung und zielten darauf ab, das authentische Verhalten der User zu erheben, um daraus qualitative (Optimierungspotenzial) und quantitative (Kennzahlen) Rückschlüsse zu ziehen. Während bei der Studie für DER CLUB Bertelsmann die server-basierte Technologie zum Einsatz kam, bei der jeweils ein User über eine Testsession auf einer Website begleitet wird, wurde beim E-Commerce-Monitor das Add-On verwendet. Dies ermöglichte eine Beobachtung von 440 Nutzern über sieben Wochen auf mehreren Webseiten.

Der Erfolg einer Usability-Maßnahme lässt sich je nach Produkt im Gegensatz zur geleisteten Investition nicht immer monetär beziffern, da es in erster Linie um Qualität nicht um Quantität geht. Gute Usability drückt sich in zunehmendem Gebrauch, verbesserter Nutzung oder höherer Kundenzufriedenheit aus. Mit den entsprechenden Kennzahlen, die ein Remote-Testing-System liefern kann, lässt sich jedoch belegen, ob die Investition – zum Beispiel in einen Relaunch – lohnend war.

DER CLUB Bertelsmann verfolgte mit einer Remote-Studie das oben genannte Ziel: Die Website www.derclub.de sollte hinsichtlich ihrer Usability optimiert werden. Nach dem Relaunch sollte mithilfe von Usability-Kennzahlen gezeigt werden, was die Überarbeitung hinsichtlich Konversionsrate, Nutzungszufriedenheit oder Bedienbarkeit bewirkt hatte. Da es sich bei der Studie um die Betrachtung einer einzelnen Website handelte, wurde der server-basierte Ansatz gewählt bei dem pro Testsession ein User beobachtet wird und keine Installation eines Add-Ons nötig ist. Die Studie wurde in drei Schritten umgesetzt: Zu Beginn wurde ein Remote-Test durchgeführt, um mit aktuellen Nutzungsproblemen und Kennzahlen den Status Quo der Website zu erfassen. Daraus ließen sich die ersten Optimierungsideen ableiten. Um den neuen Prototypen zu entwickeln und zu verbessern, wurde im nächsten Schritt ein Rapid Prototyping [68] eingesetzt. Schließlich wurde die neue Website mit einem Remote-Test auf noch vorhandene Nutzungsprobleme untersucht. Gleichzeitig konnte mit dem zweiten Test die Verbesserung gegenüber dem Status Quo gemessen werden.

Am ersten Remote-Test nahmen über einen Zeitraum von zehn Tagen 520 Probanden an der Studie teil, die online direkt über die Website rekrutiert wurden. Cookies stellten sicher, dass jeder User nur eine Einladung zur Teilnahme erhielt.

[68] Rapid Prototyping ermöglicht bereits während der Entwicklungsphase, einen Prototypen innerhalb eines kurzen Zeitraums effizient zu testen. Das Produkt wird während der Testphase iterativ optimiert, so dass schnelle Ergebnisse und frühzeitiges Feedback zum jeweiligen Entwicklungsstand vorliegen und der Zeitraum bis zur Fertigstellung des Produkts verkürzt wird.

Abbildung 84: Jeder zehnte Website-Besucher wurde über einen Layer (links) eingeladen, an der Studie teilzunehmen.

Nachdem sich die Nutzer zur Teilnahme bereit erklärt hatten, wurden sie gebeten, ihr aktuelles Vorhaben auf www.derclub.de umzusetzen. Während der Nutzung zeichnete der zwischengeschaltete Server sämtliche Schritte der Probanden über Logfiles, Screenshots und Mauspositionen auf. Über eine eingeblendete Leiste am unteren Bildschirmrand konnten die Teilnehmer jederzeit positives oder negatives Feedback zu Dingen abgeben, die ihnen im Laufe ihrer Nutzung aufgefallen sind. Zusätzlich beantworteten die Probanden Fragen zum Nutzerprofil, zur Website-Nutzung, zu Usability-Aspekten und Wettbewerbern, zudem bewerteten sie verschiedene Aspekte der Website. Das Feedback, die Screenshots inklusive Mauswege sowie die Nennungen aus den offenen Fragen wurden qualitativ ausgewertet, um daraus Empfehlungen zur weiteren Verbesserung der Webseite herzuleiten. Aus den positiven Rückmeldungen wurden Rückschlüsse auf die Stärken des Bestellprozesses gezogen.

Nach dem Relaunch startete ein weiterer Remote-Test, der unter den gleichen Voraussetzungen stattfand wie die Studie vor dem Relaunch. Innerhalb von zehn Tagen nahmen 510 Probanden teil.

Nach der Auswertung der gesammelten Daten zeigt sich, dass die Potenziale, die im ersten Test identifiziert und umgesetzt wurden, Wirkung zeigen. So wurde das Redesign der Club-Website im Vergleich zur Vorgängerversion in nahezu allen Dimensionen der Customer

Abbildung 85: Beispiel für die Testumgebung nach dem Redesign. Unten ist die Leiste eingeblendet, über die die Teilnehmer positives und negatives Feedback abgeben können.

Experience besser bewertet. Insbesondere das verbesserte Design, die neuen Navigationskonzepte und das vereinfachte Handling wirken sich nun positiv aus. Wurde in der ersten Studie die Produktdarstellung als teilweise unzureichend empfunden, stößt sie nun auf positive Resonanz. Auch die Suchergebnisse, die als nicht nachvollziehbar und irrelevant bewertet wurden, werden nun als stark verbessert wahrgenommen. Ebenso konnte DER CLUB seine Position gegenüber dem Wettbewerb signifikant steigern. Dennoch wird DER CLUB seine Webseite weiter kontinuierlich hinsichtlich einzelner Parameter weiter optimieren.

Das vorliegende Beispiel zeigt, dass mit geringem Aufwand, eine hohe Fallzahl erreicht werden konnte, die die Optimierungspotenziale quantitativ absicherten. Das Rapid Prototyping im Labor half, die Ergebnisse effizient umzusetzen, da die Testpersonen unmittelbar Feedback zu ersten Entwürfen geben konnten.

XII.II Fallbeispiel: Ein E-Commerce-Monitor mit Remote Testing

von Tim Bosenick

1. Wettbewerbsvergleichende Studien am Beispiel des E-Commerce-Monitor 290
2. Ausblick 293
3. Literatur 294

Während bei der Studie für DER CLUB Bertelsmann die server-basierte Technologie zum Einsatz kam, bei der jeweils ein User über eine Testsession auf einer Website begleitet wird, wurde beim E-Commerce-Monitor das Add-On verwendet. Dies ermöglichte eine Beobachtung von 440 Nutzern über sieben Wochen auf mehreren Webseiten.

1. Wettbewerbsvergleichende Studien am Beispiel des E-Commerce-Monitor

Regelmäßig veröffentlichte Studien und Nutzungszahlen über das Onlineshopping belegen, dass die Deutschen gerne und viel im Internet einkaufen. Wie sich die Nutzer beim Onlineshopping verhalten, welche Seiten sie bevorzugen, wie sie ihre Kaufentscheidungen treffen und welche Käufe sie schließlich offline tätigen – darauf gab es bislang jedoch keine stichhaltigen Antworten. Ziel des Branchenmonitor E-Commerce war es daher, das Verhalten von E-Commerce-Nutzern im Internet detailliert zu erfassen und unter anderem anhand von Personas[69] Käufertypen zu beschreiben. Hierzu wurde der gesamte Shopping-Prozess von 440 Personen vom Beginn der Recherche bis zum abschließenden Kauf und der After Sales Phase über einen Zeitraum von sieben Wochen beobachtet. Es wurde auch erhoben, wie häufig die Nutzer zu anderen Shops wechseln und welche Rolle externe Seiten wie Suchmaschinen und Preisvergleichsseiten beim Onlineshopping spielen. Außerdem wurde die Beobachtung der Teilnehmer durch eine Reihe von Befragungen zum Thema Onlineshopping ergänzt.

Voraussetzung für die Rekrutierung über ein Online-Panel war zum einen die Absicht, innerhalb der Feldzeit von sieben Wochen mindestens ein Produkt online kaufen zu wollen und zum anderen die Nutzung von Internet Explorer oder Firefox-Browser zur Installation des Add-Ons. Auf weitere Screening-Kriterien wurde verzichtet, um das Verhalten beim Onlineshopping auf einer möglichst breiten Basis messen zu können. Die Studie dauerte insgesamt zehn Wochen und war in drei Phasen aufgeteilt: Erstbefragung, Beobachtung und Abschlussbefragung (siehe Abbildung 86). In der ersten Phase wurden soziodemografische Daten und Informationen der Teilnehmer bezüglich ihrer bisherigen E-Commerce-Nutzung gesammelt. Während der Beobachtungsphase zeichnete das Remote-System das Surfverhalten der Teilnehmer auf. Die Teilnehmer führten als zentrales Element eine Einkaufsliste (siehe Abbildung 87) und gaben Auskünfte und Feedback zu den von ihnen besuchten Seiten.

[69] Personas beschreiben potenzielle Nutzergruppen mit ihren Eigenschaften, Arbeitsweisen, Bedürfnissen, Zielen und Präferenzen. Sie helfen, interaktive Angebote bedarfsgerecht auf die Zielgruppen auszurichten.

Abbildung 86: Ablauf der Studie „E-Commerce Monitor 2008“

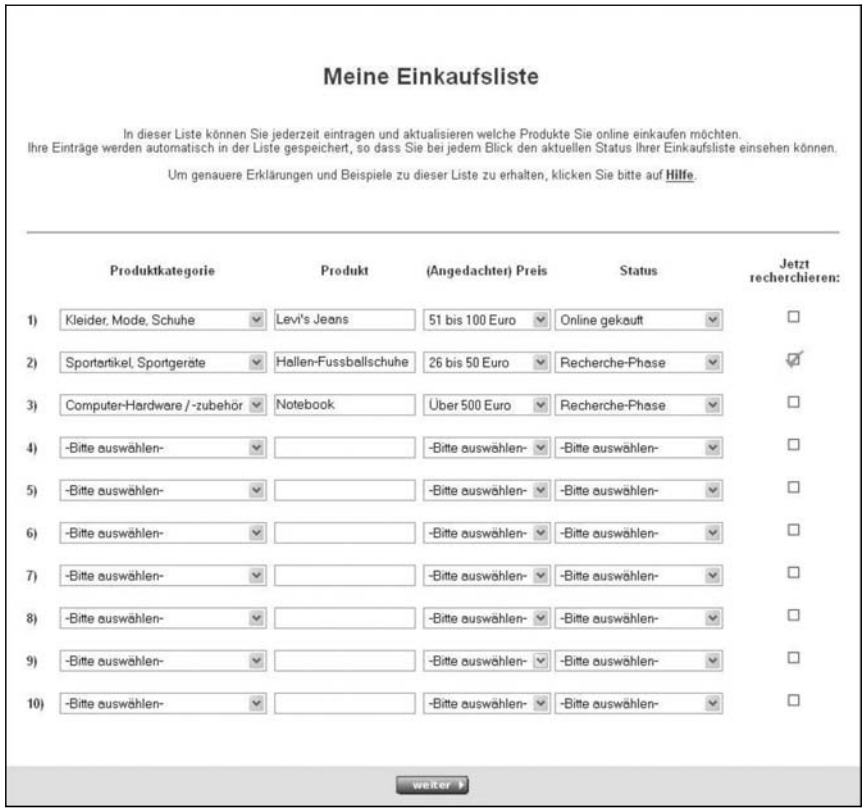


Abbildung 87: Zentrales Element der Beobachtungsphase war die Einkaufsliste. Hier trugen die Teilnehmer die Produkte ein, die sie zu kaufen planten und informierten darüber, wenn Käufe abgeschlossen waren.

Zum Abschluss der Studie wurden alle Teilnehmer zu einer Abschlussbefragung eingeladen, um die Recherchephase, das Markenimage von Online-Shops, personalisierte Shop-Bereiche und das Verhältnis zwischen Online- und Offline-Recherche zu evaluieren. Nach Ende der Feldzeit verknüpften die Customer Experience-Experten die gesammelten Daten miteinander und werteten sie aus.

Die Ergebnisse des Branchenmonitors ergeben einen umfassenden Überblick über das Internet-Kaufverhalten deutscher Online-User. So recherchierten die Teilnehmer über 2.000 Produkte und kauften davon schließlich 51 Prozent online. Im Durchschnitt wurden vier Recherche-Sessions durchgeführt, bis ein Produkt tatsächlich gekauft wurde.

Die Ergebnisse zeigen beispielsweise, dass die Internetkäufer unterschiedliche Recherche-Strategien nutzen, um an das gewünschte Produkt zu gelangen. Insgesamt konnten anhand einer Clusteranalyse drei Produktrecherche-Strategien im Internet differenziert werden: Low Involvement-, High-Involvement- und Offline-Käufe (ROPO – Research online, Purchase offline). Die Strategien unterscheiden sich hinsichtlich Rechercheintensität und des Preises. Bei der letzt genannten Strategie diente das Internet lediglich als Informationsmedium.

Durch die Befragung der Teilnehmer konnte weiter herausgefunden werden, bei welchen Produkten offline recherchiert wird, welche offline gekauft werden und aus welchen Gründen. So handelte es sich zum Beispiel bei den offline getätigten Käufen um hochpreisige Produkte, bei denen die Konsumenten ein hohes Involvement besitzen. Als Gründe hierfür führten die Probanden unter anderem zu hohe Versandkosten an.

Die Analyse des Bestellprozesses zeigte, dass ein Teil der Teilnehmer immer noch starke Sicherheitsbedenken bezüglich des Bezahlens im Internet haben. Hier müssen Onlineshops offenbar weitere mehrheitlich akzeptierte Zahlungsoptionen wie Kauf per Rechnung anbieten.

Anhand des Surf- und Rechercheverhaltens der Studienteilnehmer konnten die User Experience-Experten mithilfe einer Cluster-Analyse sieben Käufertypen identifizieren und beschreiben. Kennt ein Onlineshop-Anbieter die unterschiedlichen Käufertypen, kann er mit den verfügbaren Informationen und Funktionen seines Angebots auf die verschiedenen Charakter eingehen und die jeweiligen Käufertypen zufriedenstellen.

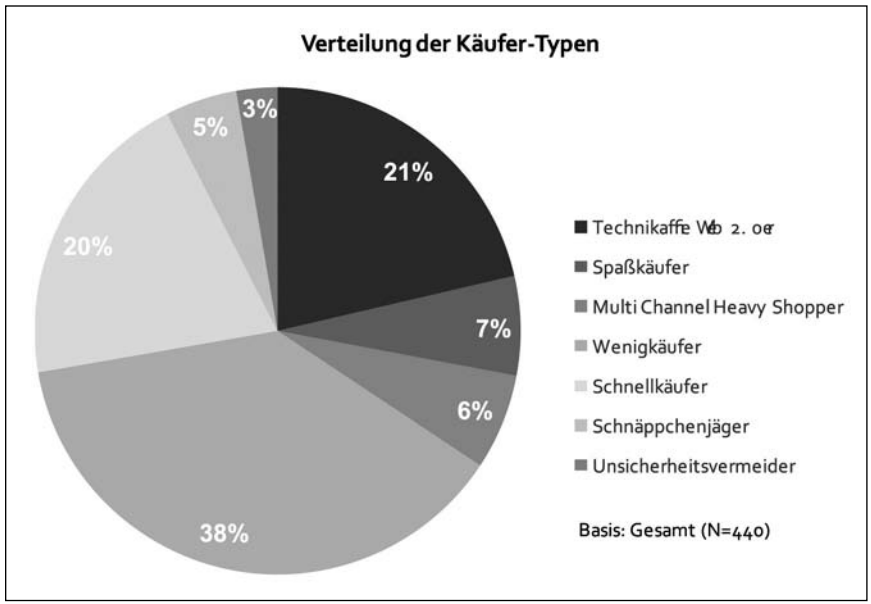


Abbildung 88: Verteilung der Käufer-Typen.

Zusätzlich zu den Ergebnissen ist das Wissen von E-Commerce-Experten in die Auswertung mit eingeflossen: Die von ihnen benannten Best Practices geben einen Überblick über Onlineshops, die sich in bestimmten Aspekten vom Wettbewerb abheben.

Die vorliegende Studie zeigt, dass der Branchenmonitor ein umfassendes Gesamtbild über das Nutzungsverhalten von Onlinekäufern im Internet gibt. Damit können Onlineshopbetreiber potenzielle Kunden besser verstehen und den Nutzern somit ein auf ihre Bedürfnisse zugeschnittenes Shopping-Erlebnis bieten.

2. Ausblick

Die Ausführungen und Fallstudien zeigen, dass die mit einem Remote-Testing mögliche Methodenkombination aus technischer Messung, Befragung und Beobachtung im natürlichen Nutzungskontext eine äußerst effektive Untersuchungsmethode der Usability-Toolbox ist. Von der Betrachtung der User Experience einzelner Webseiten bis hin zum Tracking des Nutzerverhaltens ganzer Branchen ermöglicht das Verfahren eine kontinuierliche Messung des Nutzungsverhaltens und erlaubt somit bisher nicht ausgeschöpfte

Optimierungsmöglichkeiten hinsichtlich der gesamten Customer Experience. Der nächste konsequente Schritt wird es sein, das Verfahren in größeren Panels zu installieren, die das Nutzerverhalten online und offline über einen unbestimmten Zeitraum verfolgen. So ließe sich zum Beispiel ein kontinuierliches Conversion-Rate-Benchmarking etablieren oder der Einfluss von Online-Werbung bzw. dem Internet allgemein auf das Kaufverhalten offline nachvollziehen.

3. Literatur

Brush, A.; et al.: A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface. In: CHI 2004: Late Breaking Results Paper, S.1179-1182, ACM Press, Pittsburgh 2004, PA.

Kühn, Martina: Entwicklung und Bewertung einer Remote Usability Evaluationsmethode für Webseiten, Diplomarbeit, Universität Lüneburg 2005

Nufer, Stephan: Weiterentwicklung einer asynchronen Remote Usability Testmethode mit Bewertung und Evaluation der Ergebnisse, Diplomarbeit, Universität Hamburg 2006

XIII. Expertengutachen: User Experience kurz und gut

von Andreas Selter

1. Überblick 296
2. Ausgangssituationen und Einsatzmöglichkeiten 296
3. Auf die richtige Vorgehensweise kommt es an 297
4. Die richtigen Themenschwerpunkte setzen 298
5. Qualitätsmaßstäbe 300
6. Fazit 301

Gelten Usability Testing und andere nutzerbasierte Evaluationsmethoden oft per se als Königsweg zu einer optimalen User Experience, sind Expertengutachten gerade in Zeiten kurzer Entwicklungszyklen und knapper Projektbudgets extrem effektiv. Professionell durchgeführt decken sie mit vergleichsweise geringem Aufwand sowohl grundlegende Schwachstellen als auch oftmals versteckte Details auf. Expertengutachten sind schon für sich alleine sehr wirkungsvoll, können aber auch in Kombination mit anderen Maßnahmen von großem Nutzen sein.

1. Überblick

Expertengutachten, Expert Review, Expertenevaluation, Usability Expertise, Website Check. Die Methode hat unterschiedliche Namen, die eines gemeinsam haben. Nicht mehr oder weniger repräsentative Nutzer, sondern ein oder mehrere Experten bewerten die Qualität der User Experience einer Website. Expertengutachten beinhalten einerseits eine Heuristische Evaluation, die aufzeigt, in wie weit anerkannte Usability- und Designrichtlinien eingehalten oder vernachlässigt werden. Andererseits versetzen sich die Experten bei einem Cognitive Walkthrough in die Rolle bestimmter Nutzertypen und durchlaufen typische Anwendungsfälle. Dabei decken sie Ungereimtheiten und Schwachstellen im Kontext der individuellen Nutzungssituation auf. Als Ergebnis zeigen die Experten Lösungsmöglichkeiten für die gefundenen Schwachstellen auf und geben Handlungsempfehlungen für weitere Optimierungsmöglichkeiten.

2. Ausgangssituationen und Einsatzmöglichkeiten

Expertengutachten sind für unterschiedlichen Situationen und Fragestellungen geeignet. Im Rahmen einer Konzeptentwicklung helfen sie, frühzeitig Schwachstellen zu erkennen oder diese zu vermeiden. Sind zum Beispiel die Interaktionsprinzipien intuitiv und effizient, werden sie konsistent angewendet? Bei bestehenden Websites können Expertengutachten akut auftretende Probleme, wie zum Beispiel ungewöhnlich hohe Abbruchraten in bestimmten Bereichen beleuchten, dienen jedoch ebenso der kontinuierlichen Qualitätssicherung. Expertengutachten lassen sich außerdem äußerst wirkungsvoll mit anderen Evaluations- und Analysemethoden kombinieren. Bei Auffälligkeiten in Web Analytics Daten liefern Expertengutachten das nötige Verständnis für mögliche Ursachen. Dass zum Beispiel so viele Nutzer den Registrierungsprozess abbrechen, kann daran liegen, dass das Eingabeformular unnötig umfangreich, unübersichtlich und unverständlich ist. Auch in

der Vorbereitung von Usability Tests sind Expertengutachten ein wertvoller erster Schritt. Wozu Zeit und Budget mit aufwändigen Nutzertests für Auffälligkeiten vergeuden, die ein Experte bereits auf den ersten Blick erkannt hätte?

3. Auf die richtige Vorgehensweise kommt es an

Um sicher zu stellen, dass ein Expertengutachten relevante und verwertbare Ergebnisse liefert, sollten Fragestellung und Untersuchungsthemen im Vorhinein geklärt werden. Dem Auftraggeber ist wenig geholfen, wenn der Experte mit einem pauschalen Usability Expertengutachten hauptsächlich Schwachstellen findet, die schon längst bekannt sind. In einem ersten Gespräch zwischen Auftraggeber und Experten sollten die genauen Ziele des Gutachtens definiert werden. Geht es eher um kurzfristig umsetzbare Optimierungen, besteht dringender Handlungsbedarf für akute Schwachstellen, sollen grundlegende Anforderungen für einen umfangreicheren Relaunch gesammelt werden? In welchem Umfang soll die Website untersucht werden, möglichst vollständig oder sind zunächst zentrale Bereiche wie zum Beispiel Startseite, Registrierungsprozess, Produktdetailseiten, Warenkorb oder Buchungspfad ausreichend? In welchem Umfang und Detaillierungsgrad sollen Lösungswege erarbeitet werden, in Form konkreter Konzept- und Designvorschläge oder genügen allgemeine Hinweise auf Best-Practice-Lösungen? Nur wenn die Fragestellung klar definiert ist, erfüllt ein Expertengutachten mit gezielter Vorgehensweise auch den vollen Nutzen.

Um möglichst objektive Ergebnisse zu erzielen, untersuchen in der Regel zwei bis drei Experten unabhängig von einander die fragliche Website. Im Rahmen einer Heuristischen Evaluation überprüfen sie die Website zunächst auf Konformität zu anerkannten Standards.[70] Bei einem anschließenden Cognitive Walkthrough versetzen sich die Experten in die Situation realer Nutzer und durchlaufen typische Nutzungssituationen und -szenarien. Wie reibungslos verläuft zum Beispiel die Buchung eines zweiwöchigen Urlaubs auf den Kanaren für den Vater einer vierköpfigen Familie?

Anschließend fassen die Experten ihre Beobachtungen und Erkenntnisse zusammen, wägen Bedeutung und Auswirkungen der einzelnen Punkte ab und erarbeiten möglichst konkrete Lösungsmöglichkeiten und Handlungsempfehlungen. Bei der Beschreibung von möglichen Schwachstellen und entsprechenden Lösungsmöglichkeiten ist die Betrachtung

[70] Siehe zum Beispiel die „10 Usability Heuristiken“ von Jakob Nielsen oder die Dialogprinzipien aus DIN ISO 9210, Teil 11

Aufwand gegen Nutzen von großer Bedeutung. Wie kritisch für die User Experience ist eine erkannte Problemstelle tatsächlich, welcher Aufwand ist für deren Behebung nötig und wie stark ist der zu erwartende Optimierungseffekt? Undurchsichtige oder durch die Art der Darstellung sich scheinbar widersprechende Preisangaben in einem Online-Shop sorgen mit Sicherheit für Verunsicherung und Argwohn bei den Nutzern, ließen sich aber mit verhältnismäßig geringem Aufwand beheben. Bei missverständlichen, mehrdeutigen Navigationsbegriffen oder unverständlichen Hilfetexte helfen meist einfache sprachliche Optimierungen. Dagegen kann die dringend notwendige Bereinigung einer unnötig aufgeblähten Informationsarchitektur oder Produktkategorisierung zunächst abschrecken, der dazu benötigte Aufwand sich aber hinterher in Form einer deutlich besseren Conversion Rate mehrfach auszahlen. Letztlich empfiehlt es sich, das Angebot einer persönlichen Ergebnispräsentation in Anspruch zu nehmen, bei der die Experten ihre Erkenntnisse und Empfehlungen ausführlich erläutern und mit dem Auftraggeber diskutieren können.

Den Aufwand für ein Expertengutachten bestimmten unterschiedliche Faktoren wie Komplexität und Umfang der Website, Art der Fragestellung und Untersuchungsthemen, Detaillierungsgrad der Optimierungsempfehlungen und anderes. Dementsprechend individuell fällt die Zeit- und Kostenkalkulation aus. Im Durchschnitt bewegt sich diese jedoch zwischen 2.000 und 7.000 Euro bei einer Bearbeitungszeit innerhalb von drei bis zehn Arbeitstagen.

4. Die richtigen Themenschwerpunkte setzen

Am meisten verbreitet sind Expertengutachten, die vor allem den Grad der Usability einer Website überprüfen. Dies umfasst weitgehend pragmatische Gesichtspunkte wie die Übersichtlichkeit der Informationsarchitektur und der einzelnen Seiten, Verständlichkeit von Texten, Links und Navigationsbegriffen, Vollständigkeit der relevanten Inhalte und Informationen und anderes.

Dass Nutzer eine Website aufgrund ausreichender Usability bedienen können, bedeutet aber noch nicht, dass sie es auch tun und die Angebote entsprechend nutzen. Einen Kaufabschluss tatsächlich zu tätigen, persönliche Daten preiszugeben oder in sonstiger Weise auf einer Website aktiv zu werden setzt vor allem die Vermittlung von Glaubwürdigkeit, Vertrauenswürdigkeit und Motivation voraus. Unter dem Begriff Persuasiveness (auf deutsch etwa: „Überzeugungskraft“) sind bestimmte Gestaltungsregeln definiert, welche die wahrgenommene Glaubwürdigkeit und Vertrauenswürdigkeit von Websites fördern und

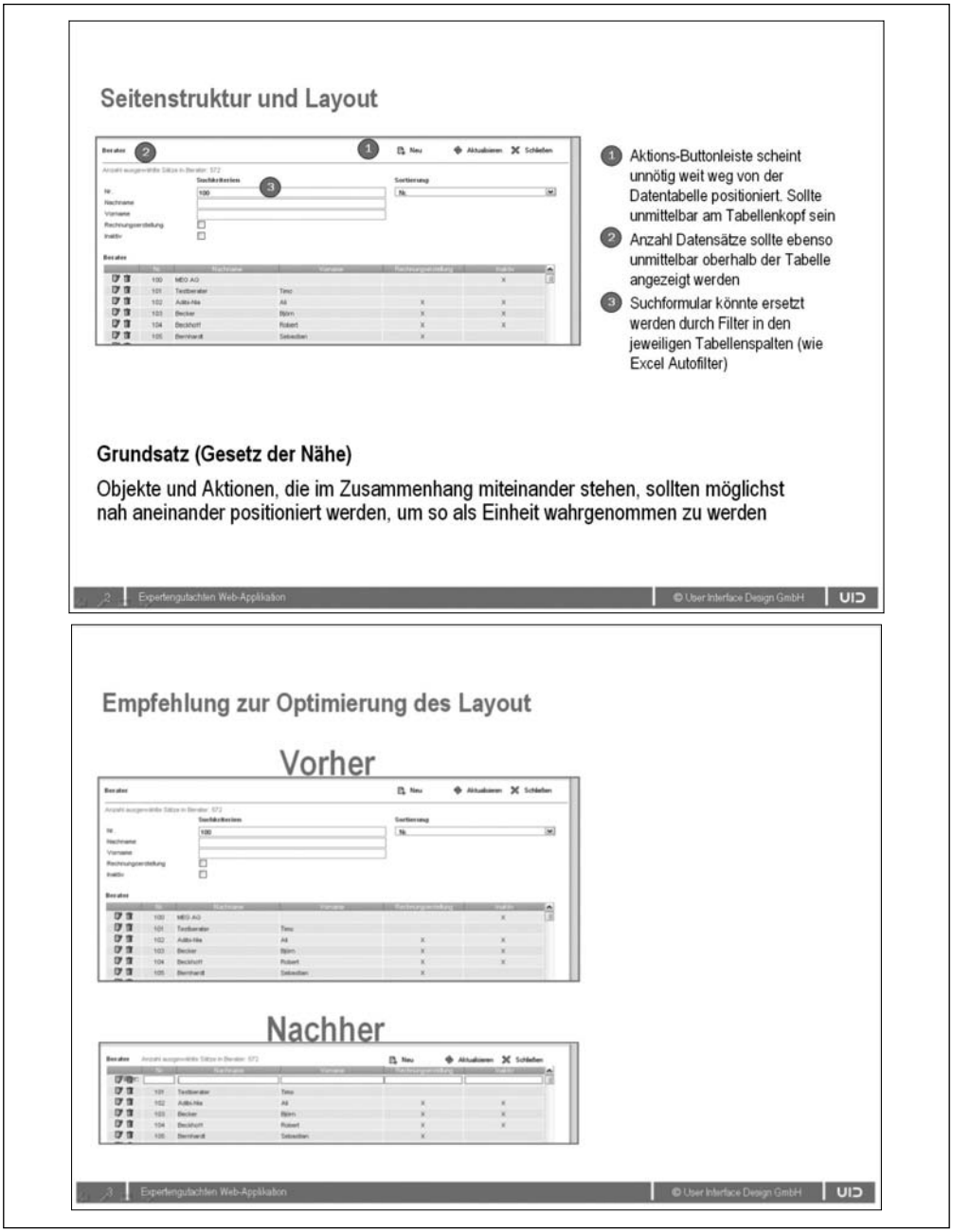


Abbildung 89: Beispiele aus einem Ergebnisbericht eines Expertengutachten mit konkreten Lösungsvorschlägen

Nutzer zu bestimmten Handlungen motivieren. Diese lassen sich im Rahmen eines Expertengutachtens mit Hilfe des Captology Evaluation Tool[71] objektiv überprüfen.

Auf dem stark umkämpften Online-Markt wird es immer wichtiger, sich mit dem eigenen Angebot von den Wettbewerbern abzuheben und herauszustechen. Dies ist nur möglich, wenn man deren Stärken und Schwächen genau kennt und einen objektiven Blick für die eigene Position hat. Mit einer expertenbasierten Benchmark-Analyse lässt sich diese anhand eines klar definierten Kriterienkatalogs gut feststellen. Das Expertengutachten liefert hierbei dann zentrale Erkenntnisse, wo versteckte Potenziale liegen und welche anderswo vorhandenen Defizite in eigene Stärken umgewandelt werden können. Wird zum Beispiel klar, dass ein Online-Tarifrechner für Versicherungen bei den meisten Anbietern unverständlich, umständlich und eher abschreckend wirkt, ist hier die Chance erkannt, mit einer entsprechend guten Lösung auf der eigenen Website zu punkten.

Ein weiterer Aspekt kann die Barrierefreiheit einer Website sein. Dabei ermöglicht Barrierefreiheit nicht nur behinderten und älteren Nutzern die uneingeschränkte Nutzung von Websites, sondern letztlich allen. Unter anderem geht es dabei um ausreichende Farbkontraste, Lesbarkeit von Schriften, Verständlichkeit von Texten, uneingeschränkte Darstellung der Website unter verschiedenen Browsern und individuellen Browsereinstellungen, eindeutige Links und anderes. Die Bundesregierung hat in ihrer „Barrierefreie Informationstechnik Verordnung“ (BITV) eindeutig überprüfbare Leitlinien zur Sicherstellung der Barrierefreiheit erlassen.

5. Qualitätsmaßstäbe

Über die Qualität und damit die Verwertbarkeit der Ergebnisse aus einem Expertengutachten entscheidet vor allem die Professionalität der beteiligten Experten. Unbestritten genügt manchmal bereits eine gute Portion gesunder Menschenverstand, um offensichtliche Unstimmigkeiten aufzudecken. Für das Erkennen und folgerichtige Interpretieren komplexerer Schwachpunkte sind das umfangreiche Wissen und die Erfahrung echter Experten im Bezug auf Nutzerverhalten und -bedürfnisse sowie entsprechend erfolgreichen Gestaltungslösungen unverzichtbar. Neben der Sorgfalt und Gründlichkeit in der Vorgehensweise

[71] Das Captology Evaluation Tool entstand im Rahmen von Forschungsarbeiten an der Hochschule der Medien in Stuttgart unter Leitung von Prof. Michael Burmester. Die Ergebnisse wurden im Rahmen einer Diplomarbeit zusammengefasst (Kaspar Szymanski, 2004). Weitere Informationen unter http://www.evaluationstechniken.de/pdf/12.pdf

sind jedoch auch die differenzierte Betrachtung und genaue Gewichtung der erkannten Schwachstellen mit Blick auf deren tatsächliche Beeinträchtigung bei der Website-Nutzung wichtig. Die Experten müssen hierbei die Interessen der Nutzer bei der Bedienung mit den Möglichkeiten und Absichten des Anbieters in Einklang bringen können.

6. Fazit

Ob für sich alleine oder in Kombination mit anderen Methoden, Expertengutachten sind wertvolle Instrumente, um Optimierungsmöglichkeiten für Websites zu identifizieren. Je fokussierter und professioneller sie durchgeführt werden, desto höher ist auch ihr Nutzen. Dabei sind sie in den unterschiedlichsten Situationen einsetzbar. Ob für die Suche nach akuten und kurzfristig behebbaren Schwachstellen, der Identifikation von Anforderungen und Innovationspotenzialen für einen umfangreichen Relaunch oder als Maßnahme zur Qualitätssicherung innerhalb von Konzept- und Designprojekten. Echte User Experience Profis erkannt man daran, dass sie die genaue Vorgehensweise und Untersuchungsschwerpunkte individuell mit dem Auftraggeber abstimmen und als Ergebnis individuelle und auch konkret umsetzbare Lösungsmöglichkeiten aufzeigen.

sind jedoch auch die differenzierte Betrachtung und genaue Gewichtung der erkannten Schwachstellen mit Blick auf deren tatsächliche Beeinträchtigung bei der Website-Nutzung wichtig. Die Experten müssen hierbei die Interessen der Nutzer bei der Bedienung mit den Möglichkeiten und Absichten des Anbieters in Einklang bringen können.

6. Fazit

Ob für sich alleine oder in Kombination mit anderen Methoden, Expertengutachten sind wertvolle Instrumente, um Optimierungsmöglichkeiten für Websites zu identifizieren. Je fokussierter und professioneller sie durchgeführt werden, desto höher ist auch ihr Nutzen. Dabei sind sie in den unterschiedlichsten Situationen einsetzbar. Ob für die Suche nach akuten und kurzfristig behebbaren Schwachstellen, der Identifikation von Anforderungen und Innovationspotenzialen für einen umfangreichen Relaunch oder als Maßnahme zur Qualitätssicherung innerhalb von Konzept- und Designprojekten. Echte User Experience Profis erkannt man daran, dass sie die genaue Vorgehensweise und Untersuchungsschwerpunkte individuell mit dem Auftraggeber abstimmen und als Ergebnis individuelle und auch konkret umsetzbare Lösungsmöglichkeiten aufzeigen.

XIII.I Fallbeispiel: Expertengutachen für D&B (www.dnbgermany.de)

von Andreas Selter

1. Die Ausgangssituation 304
2. Die Vorgehensweise 304
3. Erkenntnisse und Lösungen aus dem Expertengutachten 305
4. Fazit 310

Die D&B Deutschland GmbH ist ein führender Anbieter von Unternehmensinformationen. Die D&B Datenbank enthält laufend aktualisierte Daten zu mehr als 4,3 Millionen deutschen Firmen und rund 140 Millionen Unternehmen weltweit. Die Website www.dnbgermany.de dient der Darstellung des Unternehmens mit seinen Kerngeschäftsbereichen mit deren Produkten und Lösungen. Die Website wurde zuletzt 2007 einem Relaunch unterzogen, der im Januar 2008 online ging. Der mittelfristige Erfolg hinsichtlich Besucherzahlen, inhaltlicher Nutzung und schließlich Leadgenerierung erfüllte jedoch nicht die Erwartungen. Web Analytics Daten lieferten erste Hinweise auf potenzielle Schwachstellen. Ein Usability & User Experience Expertengutachten ging den Anhaltspunkten schließlich auf den Grund und zeigte entsprechende Optimierungsmöglichkeiten auf.

1. Die Ausgangssituation

Trotz des noch nicht lange zurückliegenden Relaunch der Website war man bei D&B mit der bisherigen Performance nicht zufrieden. Intern war man sich bereits einiger potenzieller Optimierungsmöglichkeiten bewusst, hauptsächlich in Bezug auf Navigation, Orientierung, Textlayout und -gestaltung und anderes. Die Frage war jedoch, welche Usability-Schwächen hat die Website tatsächlich und wie könnten die richtigen Themen angegangen und effektive Lösungen gefunden werden? D&B setzte seit einiger Zeit ein Web Analytics Tool ein und beauftragte die User Interface Design GmbH (UID) damit, ausgehend von den Web Analytics Daten ein Usability & User Experience Expertengutachten zu erstellen, das die wichtigsten Optimierungsmöglichkeiten aufzeigt. Ziel war von vorneherein, mögliche Optimierungen evolutionär in die bisherige Website einfließen zu lassen, anstatt revolutionäre Änderungen vorzunehmen, die übermäßige Zeit- und Kostenaufwände bedeuten könnten.

2. Die Vorgehensweise

In einem gemeinsamen Workshop zwischen D&B und UID wurden zunächst die Ziele und die konkrete Vorgehensweise für das Expertengutachten besprochen. Für die spätere Durchführung des Cognitive Walkthrough war es vor allem wichtig, ein möglichst genaues Verständnis für die eigentliche Zielgruppe der D&B Website mit deren Bedürfnissen und Nutzungszielen zu gewinnen. Zu diesem Zweck wurden Personas (stereotypische Nutzerprofile) und Anwendungsszenarien definiert. Wichtig war zudem, das Selbstverständnis und die Website-Ziele aus der Sicht von D&B zu beleuchten. Nur so war es später möglich,

die richtigen Rückschlüsse aus dem Expertengutachten zu ziehen und entsprechend wirkungsvolle Optimierungen zu empfehlen.

Als Auftakt zum Expertengutachten analysierten die User Experience Experten von UID zunächst die vorliegenden Web Analytics Daten, um hieraus bereits erste Anhaltspunkte für mögliche Schwachstellen aufzudecken. Dabei wurden vor allem folgende Fragen verfolgt: Welche Seiten werden am stärksten oder am wenigsten genutzt? Wie ist die durchschnittliche Verweildauer auf den jeweiligen Seiten, bzw. in wie fern entspricht sie den Erwartungen (zum Beispiel längere Verweildauer auf inhaltsreichen und wichtigen Seiten, bzw. kurze Verweildauer auf Übergangsseiten wie zum Beispiel Überblicksseiten)? Welches sind die häufigsten Ausstiegsseiten? Wie sehen typische Klickverläufe und Navigationspfade aus? Hieraus ergaben sich bereits erste Anhaltspunkte für zentrale Schwachstellen.

Ausgehend von diesen Hinweisen unterzogen die User Experience Experten die Website einer Heuristischen Evaluation anhand einschlägiger Usability und Persuasiveness Richtlinien[72]. In einem weiteren Schritt wurde dann ein Cognitive Walkthrough anhand der zuvor definierten Personas und Anwendungsszenarien durchgeführt. Dabei versetzten sich die User Experience Experten unter anderem zum Beispiel in die Rolle eines Business Analysten, der auf der Suche nach einem Anbieter von Wirtschaftsdaten, Firmeninformationen und entsprechenden Tools ist.

Die User Experience Experten stellten schließlich ihre Erkenntnisse und Thesen den faktischen Web Analytics Daten gegenüber. Dabei konnten sie ihre Thesen über Schwachstellen plausibel mit den faktischen Nutzerdaten unterstreichen. Die Gesamtergebnisse einschließlich der Empfehlungen zur Optimierung Website mit konkreten Lösungswegen wurden D&B im Rahmen eines Ergebnisworkshops vorgestellt. Hierbei wurden dann auch konkrete Maßnahmen und Folgeschritte diskutiert und verabschiedet.

3. Erkenntnisse und Lösungen aus dem Expertengutachten

Die zugrunde gelegten Web Analytics Daten zeigten vor allem zwei Auffälligkeiten. Zum einen war die Abbruchrate auf der Startseite ungewöhnlich hoch, das heißt ein Großteil der Nutzer kam nicht über die Startseite hinaus. Dazu kam, dass die auf der Startsei-

[72] Stanford Guidelines for Web Credibility (http://credibility.stanford.edu/guidelines/index.html), Grundsätze der Dialoggestaltung nach DIN EN ISO 9241 – 10 (http://www.iso.org/iso/home.htm), die 10 Usability Heuristiken nach Jakob Nielsen (http://www.useit.com/papers/heuristic/heuristic_list.html).

te vorhandenen Teaser der drei Kerngeschäftsbereiche verhältnismäßig wenig angeklickt wurden. Verfolgte man andererseits die Klickpfade innerhalb der Website, zeigte sich, dass viele Nutzer innerhalb der Navigation herumirrten und sich offensichtlich schwer taten, die gewünschten Inhalte auf direktem Wege zu finden.

Als mögliche Ursachen für diese Auffälligkeiten wurden aus dem Expertengutachten im Wesentlichen vier Kernthemen identifiziert. Die Layoutgestaltung der Startseite, die allgemeine Informationsarchitektur, die Darstellung der Navigationsmöglichkeiten sowie das visuelle Design.

Die Startseite

Die Inhalte auf der Startseite waren zwar gut strukturiert aber auch sehr gleichwertig dargestellt. Bis auf zwei kleinere Inhaltsbereiche war der Hintergrund durchgehend grau, die Schriftformate sehr einheitlich. Dies machte es dem Auge schwer, visuelle und inhaltliche Schwerpunkte zu erkennen, vor allem die drei Kerngeschäftsbereiche. Insgesamt wirkte die Seite inhaltlich sehr voll und wenig einladend. Darüber hinaus schien die generelle Anmutung des visuellen Designs den Premium-Anspruch von D&B nicht gerecht zu werden. Die Empfehlung lautete hier, die drei Kerngeschäftsbereiche visuell deutlich herauszustellen, die verschiedenen Inhalte nach ihrer Bedeutung visuell besser zu unterscheiden, das Layout insgesamt aufzulockern sowie das Screen Design im Rahmen des vorhandenen CI aufzufrischen.

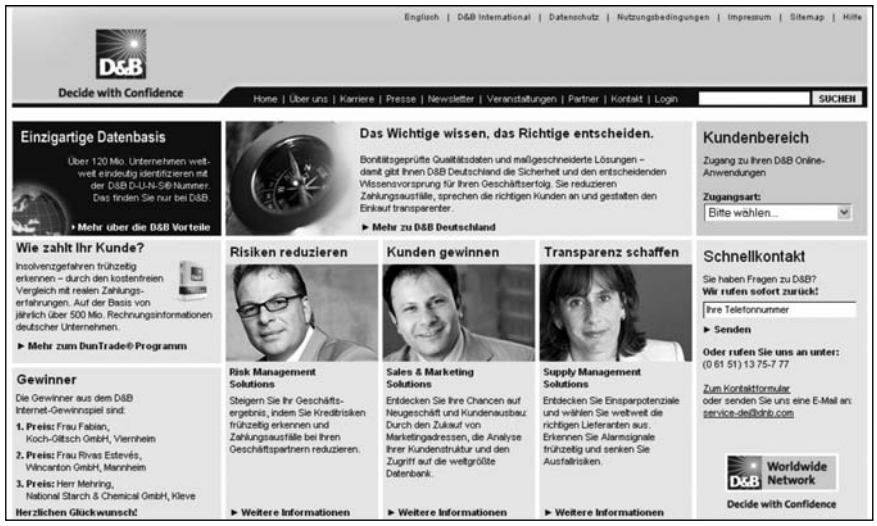


Abbildung 90: Die Startseite von D&B zum Zeitpunkt des Expertengutachten. Das Layout ist inhaltlich gut strukturiert, die visuelle Gestaltung lässt aber jeden Unterschied verwischen, es fehlt an Schwerpunkten.

Abbildung 91: Die optimierte Startseite von D&B wirkt nun deutlich übersichtlicher und hochwertiger.

Informationsarchitektur und Navigationsvisualisierung

Die Informationsarchitektur erschien sehr verzweigt, die Kategorisierung der einzelnen Bereiche für Nutzer nicht immer ganz nachvollziehbar, die Bezeichnungen teilweise nicht eindeutig. Einige Seiten, zu denen man über Textlinks im Inhaltsbereich gelangte, waren nicht in der Gesamtnavigation verankert, das heißt die Orientierung ging hier völlig verloren. Außerdem gab es verschiedene Seiten, die redundant in unterschiedlichen Bereichen verankert waren. Auf Basis der Expertenempfehlungen überarbeite UID die Navigationsarchitektur. Es wurden unnötige Ebenen entfernt, Begriffe verständlicher und eindeutiger formuliert und Seiteninhalte teilweise zusammengefasst. Haupt- und Metanavigation wurden neu strukturiert und so wesentlich übersichtlicher und intuitiver gestaltet.

Abbildung 92: Die alte Haupt- und Subnavigation: Der Bereich „Über uns" enthält 8 Unterpunkte, die teilweise nicht eindeutig unterscheidbar sind (zum Beispiel Über uns, Daten & Fakten, Unternehmensprofil). Die D&B Kerngeschäftsbereiche sind hier nicht über die Hauptnavigation, sondern nur in der Subnavigation unter „Produkte & Lösungen" innerhalb des Bereichs „Über uns" zu erreichen. Die Subnavigation ist insgesamt visuell und räumlich getrennt von der Hauptnavigation, die Orientierung wird damit erschwert.

Abbildung 93: Die neue Haupt- und Subnavigation: Das Unternehmensprofil ist nun ein eigenständiger Bereich, der nur noch 4 anstatt bisher 8 Unterpunkte beinhaltet. Die D&B Kerngeschäftsbereiche sind nun direkt in der Hauptnavigation verankert. Die Subnavigation wurde visuell und räumlich stark an die Hauptnavigation angeglichen, um die inhaltlichen Zusammenhänge eindeutig zu machen.

Layout & Screendesign

Das bisherige Screendesign folgte dem Corporate Style Guide von D&B, der allerdings bisher nicht gesondert auf die Anforderungen des Mediums Website einging. Die starre Umsetzung dieses Styleguides ließ das Design der Website wenig zeitgemäß und attraktiv erscheinen.

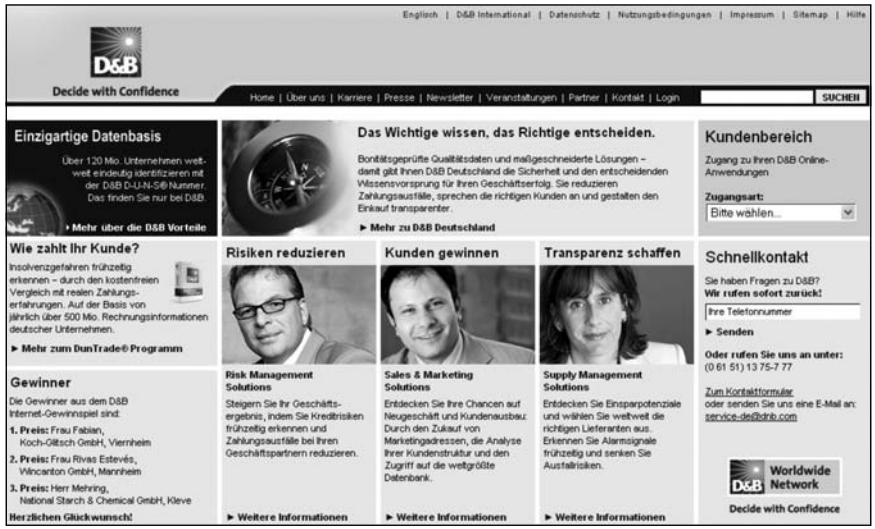


Abbildung 94: Das bisherige Layout und Screendesign. Die beiden CI Farben Gelb und Blau wurden sehr volltonig eingesetzt, was in Kombination mit dem flächigen Hintergrund-Grau wenig zeitgemäß und attraktiv wirkte. Die unterschiedlichen Inhalte waren jeweils in einzelnen Kästchen dargestellt, was die Struktur zwar unterstrich, andererseits aber auch sehr streng und starr erschien. Durch die gleichartige Gestaltung der Inhaltelemente traten die drei Kerngeschäftsbereiche nicht genug hervor.

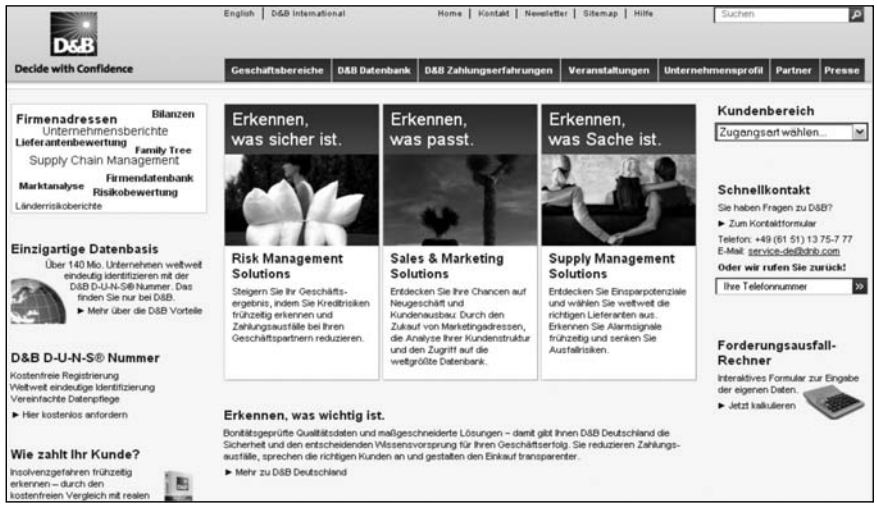


Abbildung 95: Das neue Layout und Screendesign: Die CI Farben Gelb und Blau wurden für die Website optimiert, in dem sie anstatt im Vollton mit einem dezenten Verlauf dargestellt werden. Die separaten Kästchen wurden entfernt zugunsten eines dezenten Hintergrund-Grau, auf dem die Inhalte frei stehen. Die drei Kerngeschäftsbereiche stehen nun deutlich im Vordergrund. D&B setzt außerdem nun eine komplett neue Bilderwelt ein.

4. Fazit

Die Kombination aus Web Analytics und qualitativem Expertengutachten erwies sich bei der Optimierung der D&B Website als optimale Kombination. Die Nutzerdaten dienten sowohl als Vorlage als auch als Nachweis für die Erkenntnisse der User Experience Experten. Die sehr eindeutigen Ergebnisse im Hinblick auf Seitenlayouts, Navigationsstruktur und Informationsdesign waren eine ideale Vorlage für die gezielten Optimierungen. Deren Resultat stieß nicht nur bei D&B auf große Zustimmung, sondern nachweißlich auch bei den Nutzern. Dies belegen unter anderem die Web Analytics Daten in Punkto Seitenaufrufe, Verweildauern sowie durchgängiges Klickverhalten auf den einzelnen Seiten.

XIV. Anbieter, Systeme, Agenturen

von Frank Reese

1. Clickmaps, Mausbewegungsanalysen und ähnliches 312
2. Multivariate Analysen 313
3. Die große Familie der User Experience Tests 314

In diesem Abschnitt soll dem interessierten Leser eine Reihe von Adressen und Ansatzpunkten auf der Suche nach Dienstleistern und Systemen für die in diesem Buch beschriebenen Verfahren geboten werden. Die Listen beanspruchen selbstverständlich weder Vollständigkeit noch eine Bewertung der genannten Anbieter.

1. Clickmaps, Mausbewegungsanalysen und ähnliches

Die Lösung **m-pathy** wird von der Dresdener Firma seto angeboten. Wie hier im Beitrag des seto-Mitarbeiters Dirk Schulze beschrieben (Kapitel VII), liefert m-pathy Mausbewegungsanalysen, Heatmaps und ähnliches zur Analyse des Verhaltens der Besucher auf einer Website. **www.m-pathy.com**

Auch die US-Anbieter Clicktale und Crazyegg bieten ähnliche Services zur Visualisierung der Besucherbewegungen. **www.clicktale.com; www.crazyegg.com**

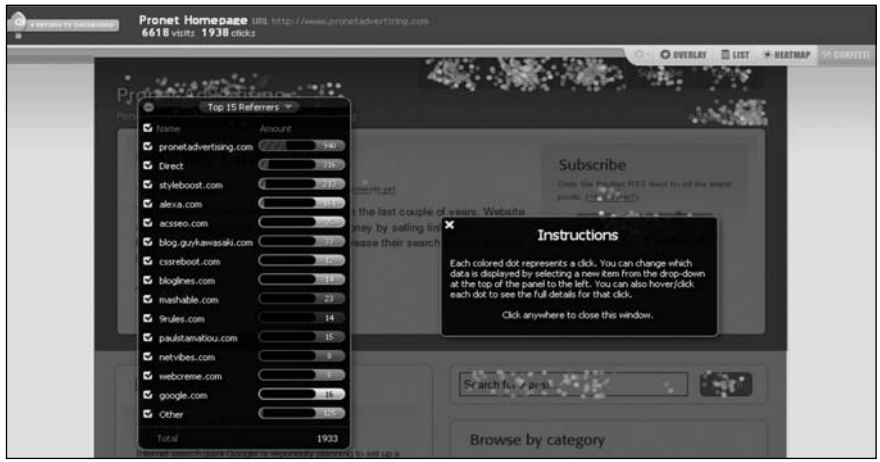


Abbildung 96: Confetti-Visualisierung bei Crazyegg.

2. Multivariate Analysen

Zum Google Website Optimizer muss hier nicht mehr viel gesagt werden. Das sehr nützliche und dabei kostenlose Tool wird in Kapitel IX.III ausführlich beschrieben.
www.google.de/websiteoptimizer

Sitespect

Sitespect ist ein amerikanischer Anbieter einer Testplattform. Als kostenpflichtiger Service bietet Sitespect weit mehr Flexibilität und Analysemöglichkeiten als kostenlose Varianten. Sitespect kann als ASP (oder auch „SaaS") genutzt werden, die Lösung kann im Unterscheid zu vielen Mitbewerbern ebenfalls inHouse eingesetzt werden.
www.sitespect.com

Widemile

Widemile ist ein weiterer amerikanischer Anbieter einer multivariaten Testplattform. Dieses System wird wie die meisten allein im ASP-Modus angeboten. Wie alle professionellen Systeme bietet es weitgehende Möglichkeiten zur Datenreduktion (im Gegensatz zu Google) und zur individuellen Konfiguration von Konfidenzintervallen und Ähnlichem.
www.widemile.com

Maxymiser

Maxymiser ist eine britische Testplattform. Neben den üblichen Funktionen bietet Maxymiser auch Systeme zum Behavioral Targeting. **www.maxymiser.com**

Divolution

Das Detmolder Unternehmen bietet das derzeit einzige ausgereifte Testsystem aus Deutschland. Der Beitrag (Kapitel IX.II) des Mitarbeiters Michael Jonas beschreibt einige der Funktionen. **www.divolution.de**

3. Die große Familie der User Experience Tests

In diesem Buch wurde versucht zu beschreiben, welche Methoden aus dem Gebiet der User Experience Tests bei welchen Projekten weiterhelfen können. Diese Verfahren werden selten singulär eingesetzt, und noch seltener gibt es in den anwendenden Unternehmen selbst die Kompetenz zur Durchführung der Verfahren. Meist wird für diese Dienstleistungen darum eine spezialisierte Agentur engagiert, die sowohl die Beratung, Planung wie auch die Durchführung übernimmt. Die Kosten für ein Optimierungsprojekt variieren entsprechend den Anforderungen und den konkret eingesetzten Methoden stark. Für ein durchschnittliches Vorhaben mit circa zehn repräsentativen Versuchspersonen ist mit Kosten von etwa 10.000 Euro zu rechnen.

In Deutschland gibt es eine ganze Reihe professionell arbeitender und erfahrener Agenturen, die Unternehmen weiterhelfen können. Einige sollen im Folgenden kurz genannt werden:

eResult, Göttingen

Das Göttinger Unternehmen existiert seit 2000 und beschäftigt knapp 20 Mitarbeiter. Das volle Methodenspektrum wird abgedeckt. Zu den Kunden gehören: AIDA, aperto, Baur, Bertelsmann Stiftung, Bosch Rexroth, comdirect Bank, HRS. **www.eresult.de**

phaydon, Köln

Das 2004 gegründete Unternehmen sieht sich als „Full-Service Forschungs- und Beratungsunternehmen" und beschäftigt knapp 30 Mitarbeiter. Zu den Kunden gehören: Quelle, Baur, Vente Privée, Bayer, Schering oder die Deutsche Bank. **www.phaydon.de**

SirValUse, Hamburg

Die Agentur SirValUse aus Hamburg wurde im Jahr 2000 gegründet und beschäftigt inzwischen mehr als 75 Mitarbeiter. Die Agentur sieht sich als Experten für den gesamten Bereich der User Experience Tests. Zu den Kunden in 2008 und 2009 zählen unter anderem: Audi, Beiersdorf, ZDF, Telekom, Swisscom, Microsoft oder Samsung. **www.sirvaluse.de**

User Interface Design, München

User Interface Design (UID) ist ein User Experience und Usability Spezialist aus München mit 65 Mitarbeitern und Standorten in München, Ludwigsburg, Mannheim und Dortmund. Zu den Kunden gehören Expedia, Deutsche Telekom, 1&1 Internet, Ebay, Western Union und Vaillant.

Remote Tests

Da die sogenannten Remote Tests (hier in Kapitel XII vorgestellt) eine gewisse Sonderposition zwischen Traffic- und Usability-Analysen einnehmen, werden die – aktuell nur zwei – Systeme hier noch einmal gesondert aufgeführt.

LEOtrace©

Das System wurde entwickelt und wird eingesetzt von der oben bereits erwähnten Agentur SirValUse aus Hamburg. LEOtrace© zeichnet das Verhalten der Nutzer auf, während diese die zu testende Website in ihrem natürlichen Umfeld (zum Beispiel zu Hause) besuchen. Im Unterschied zu Mitbewerbern kommt das System dabei ohne Browser-PlugIn aus, was die Bereitschaft zur Teilnahme deutlich erhöht.

Ein Projekt mit LEOtrace© wird mit ungefähr 2.000 Probanden durchgeführt, die on-site rekrutiert werden. Die Kosten für ein solches Vorhaben, das circa sechs Wochen Laufzeit beansprucht, werden ungefähr bei 25.000 Euro liegen. SirValUse hat mit diesem Verfahren unter anderem folgende Kunden beraten: Otto, Ebay, Spiegel Online, T-Com, Yahoo Deutschland, tagesschau.de

Ansprechpartner: Torsten Müller, SirValUse, mueller@sirvaluse.de, weitere Informationen unter www.sirvaluse.de

WebEffective

Keynote Systems ist ein amerikanisches, weltweit operierendes Unternehmen, das mit WebEffective ein ausgereiftes System für Remote Tests anbietet. Das Unternehmen selbst ist Deutschland präsent, das System wird allerdings auch von einer Reihe deutscher Agenturen in Partnerschaft mit Keynote angeboten. Unter anderem: scoreberlin (www.scoreberlin.de) und phaydon (siehe oben).

Die Autoren

Martin Beschnitt ist User Experience Consultant bei der eResult GmbH in Göttingen und verantwortet dort als Mitglied der Geschäftsleitung den Bereich Konzeption & Design. Das Unternehmen bietet seit 2000 als Full-Service Usability-Agentur Produkte & Services für unterschiedliche Anwendungen (Websites/-applikationen, Intranets, mobile Portale & Endgeräte, Kiosksysteme, etc.) und Fragestellungen – von der Anforderungsanalyse, über die Konzeptentwicklung bis hin zur Evaluation & Optimierung. Kontakt: martin.beschnitt@eresult.de, www.eresult.de

Tim Bosenick ist Geschäftsführer der SirValUse Consulting GmbH in Hamburg. Er studierte Biochemie, Physik, Philosophie, Psychologie und Soziologie mit Abschluss als Diplom-Soziologe. Direkt nach dem Studium gründete er 2000 SirValUse, das inzwischen eines der größten Institute für Usability-Testing in Europa und Gründungsmitglied des internationalen Netzwerkes „User eXperience Alliance (UXA)" ist.
Kontakt: bosenick@sirvaluse.de, www.sirvaluse.de

Birger Friedrichs ist seit mehreren Jahren im Online Marketing beratend tätig. Seit Oktober 2007 ist er bei der Morgana Castle – Internet Solutions GmbH in Stuttgart für den Bereich Internet Marketing zuständig. Hier betreut er auf Agenturseite die Kunden in den Bereichen Suchmaschinenwerbung und Website Optimierung. Davor hat er für den internationalen Telekommunikationsdienstleister Avaya mehrere Jahre im Internet Marketing gearbeitet und war dort für die Optimierung der Websites in Europa und Asien zuständig.
Kontakt: friedrichs@morgana.de, www.morgana.de

Michael Jonas ist Manager Client Services bei DIVOLUTION. DIVOLUTION® ist ein Software-as-a-Service Anbieter von Optimierungslösungen für E-Commerce, Onlinevertrieb und Onlinewerbung. Seit 2005 bietet der Testing & Targeting Spezialist eine Plattform, die eine Echtzeitanalyse des Kundenverhaltens und die automatische Optimierung der Kundenansprache ermöglicht. Damit wird eine Erhebung relevanter Zielgruppendaten, eine Steigerung der Kundenzufriedenheit, ein Umsatzwachstum bei gleichzeitiger Kostensenkung und eine Effizienzoptimierung erreicht. Kontakt: m.jonas@divolution.com, www.divolution.com

Tobias Kauer. Team Manager Brand & Communication Research bei Deutsche Post World Net. Studium der Diplom-Soziologie, Sozialpsychologie und Politische Wissenschaften an der Universität Mannheim. Seit 2004 als Senior Research Consultant im „Market Research Service Center" der Deutsche Post World Net und andere verantwortlich für die Durchführung von internen und externen Befragungen schwerpunktmäßig zu folgenden Themen: Kundenzufriedenheit, Mitarbeiter- und Führungskräftebefragungen, Konzept- und Usability-Test; derzeit verantwortlich für das Team „Brand and Communication Research". Vor 2004 verschiedene Stationen in der betrieblichen und Instituts-Marktforschung.
Kontakt: tobias.kauer@dpwn-mrsc.com, www.dp-dhl.de

Frank Reese betreibt mit Ideal Observer, Berlin, das wahrscheinlich bekannteste vollständig anbieterunabhängige Beratungsunternehmen zum Thema Web Analytics und Site-Optimierung im deutschsprachigen Raum. Er hat bereits viele große deutsche Unternehmen bei der Auswahl, Einführung und Nutzung von Web-Analytics-Systemen und kann als einer der profiliertesten Experten zum Thema in Deutschland gesehen werden. 2008 veröffentlichte er das erste deutschsprachige Buch zum Thema „Web Analytics – damit aus Traffic Umsatz wird".
Kontakt: reese@idealobserver.de, www.idealobserver.de

Carsten Rehmann ist seit Ende 2003 bei der eResult GmbH als Research Consultant tätig. Er studierte Betriebswirtschaftslehre mit dem Schwerpunkt Marketing und Statistik an der Georg August Universität Göttingen. Im Rahmen seiner Diplomarbeit, beschäftigte er sich mit der Blickverläufen auf Websites bei mehrmaligen Seitenkontakten. In Kundenprojekten ist Herr Rehmann als Studienleiter tätig. Das heißt, er betreut die Studienkonzeption, die Durchführung sowie die Analyse und Interpretation der Daten. Neben dieser Tätigkeit betreut Herr Rehmann als Produktmanager das Produkt „Nutzerbasierter Usability-Test". Hier liegt sein Augenmerk auf der kontinuierlichen Verbesserung des Projektablaufs und der Qualitätssicherung von nutzerbasierten Studien. Des Weiteren obliegt ihm die technische Verantwortung für das stationäre und mobile Usability-Labor der eResult GmbH.
Kontakt: carsten.rehmann@eresult.de, www.eresult.de

Dr. Oliver Rengelshausen ist Leiter E-Commerce der Thomas Cook AG für den deutschen Markt. In dieser Funktion ist er verantwortlich für den Online-Vertrieb der deutschen Veranstaltermarken. Rengelshausen war bereits von 2001 bis 2005 als Leiter E-Commerce und Leiter Non Traditional Sales bei Thomas Cook. Seit April 2005 wirkte er als Head of Alternative & Electronic Sales Channel bei der O2 GmbH & Co OHG in München. Begonnen hatte der gebürtige Hildesheimer seine berufliche Karriere nach dem Studium der Betriebswirtschaftslehre in Göttingen in der Werbeabteilung der Blaupunkt-Werke GmbH, Hildesheim. Er war dann als freiberuflicher Trainer und Berater für neue Medien, E-Commerce und Telekommunikation tätig. Im Jahr 2000 übernahm Rengelshausen die Leitung des Produktmanagements Touristik der Lufthansa E-Commerce GmbH.
Kontakt: oliver.rengelshausen@thomascook.com, www.thomascook.com

Daniel R. Schmeißer, Diplom-Psychologe, ist geschäftsführender Gesellschafter und Mitgründer von phaydon | research+consulting, einem auf Usability und User Experience spezialisierten Forschungs- und Beratungsunternehmen mit Sitz in Köln. Seine Forschungs- und Beratungsschwerpunkte liegen in den Bereichen User Experience (Web und Mobile), Kommunikations- und Medienforschung. Er ist Autor zahlreicher Studien und Beiträge zu den Themen E-Commerce, Usability und Werbewirkungsforschung. Seit 2008 ist er Lehrbeauftragter für die Masterstudiengänge Media Management und Wirtschaftspsychologie an der Hochschule Fresenius, Köln.
Kontakt: daniel.schmeisser@phaydon.de, www.phaydon.de

Dirk Schulze arbeitet seit deren Gründung 2005 bei der Dresdner Agentur Seto GmbH und beschäftigt sich dort mit Usability, Web-Analyse und Online-Marketing. Das Unternehmen brachte 2007 das innovative Mousetracking-Tool www.m-pathy. com auf den Markt. Dirk Schulze studierte Kommunikationswissenschaft und schreibt gerade seine Magisterarbeit zum Thema Mousetracking.
Kontakt: dirk.schulze@m-pathy.com, www.m-pathy.com

Andreas Selter arbeitet seit seinem Studienabschluss in Informationsmanagement im Jahr 2001 bei der User Interface Design GmbH (UID) als Usability Consultant. UID ist eines der führenden User Experience Dienstleistungs- und Beratungsunternehmen in Europa und Gründungsmitglied des International Usability Partners Netzwerk (IUP). Er ist Mitglied der Geschäftsstellenleitung in München und betreut als Manager Web Solutions namhafte Kunden wie Expedia.de, HUK24, SOS Kinderdörfer weltweit, Quelle, Immobilien Scout und andere. Andreas Selter hat fundierte Erfahrung in der Leitung von umfangreichen User Experience Projekten im In- und Ausland. Seine Schwerpunkte umfassen dabei User Research und benutzerzentrierte Gestaltung für E-Business Websites als auch Rich-Internet-Applications im B2B und B2C Umfeld. Sein Web User Experience Wissen hat er bereits in der Vergangenheit weitergegeben im Rahmen von Lehraufträgen an der FH Ulm im Fachbereich Medienergonomie sowie an der FH Joanneum in Graz (Österreich) im Fachbereich Informationsdesign.
Kontakt: andreas.selter@uid.com, www.uid.com

Sebastian Wetterauer, (Teamleiter Data & Web Analytics, AutoScout24 GmbH) ist bei AutoScout24 verantwortlich für den Bereich Data & Web Analytics. Nach erfolgreichem Abschluss seines Markt- und Werbepsychologie, Soziologie und Theaterwissenschafts Studiums an der LMU München, startete er 2005 seine berufliche Karriere als Assistenz der Geschäftsführung einer E-Commerce Plattform in der Sportartikelbranche. Seit Januar 2007 ist er bei AutoScout24 zuständig für den Bereich Web Analytics und der europaweiten Einführung eines Trackingtools.
Kontakt: swetterauer@autoscout24.com, www.autoscout24.de

Hier ist Raum für Ihre Notizen