

Die mittlere Güte von Navigationssystemen

Ein Kennwert für komparative Analysen von Websites bei Usability-Nutzertests

Monika Röttger und Wolfgang G. Stock, Düsseldorf

Mit der mittleren Güte des Navigationssystems einer Website stellen wir einen quantitativen Kennwert für die Erfassung der Usability von Webauftritten vor. Der Kennwert arbeitet mit Klickraten und Abbruchhäufigkeiten bei aufgabenbezogenen Nutzertests. Er ist bei vergleichenden Analysen von Websites einsetzbar.

The Mean Quality of Navigation Systems. An Indicator for Comparative Analyses of Web Sites in Usability Tests
We introduce the „mean quality of the navigation system of a Web site“ as a quantitative indicator for the analysis of the usability of the site. The indicator is based upon click rates and break-off rates in task-based user tests. The indicator is useful to comparative analyses of Web sites.

Problemstellung

Seit dem Klassiker von Jakob Nielsen (2000) ist die Usability von Websites zu einem der Standardkriterien für die Analyse von Auftritten im World Wide Web geworden. Ziel entsprechender evaluierender Studien ist in aller Regel ein praktisches: die Verbesserung der Website, gilt doch im E-Commerce die Reihe „Usability vor Marke vor System“ (Hess 2002). Eine eher wissenschaftliche Durchdringung der Usability ohne direkte Praxisfixierung ist demgegenüber weitgehend Desiderat. Vergleichende Untersuchungen unterschiedlicher Websites sind für mindestens folgende drei Aufgaben notwendig:

- (1) quantitative Erfassung des „Fortschritts“ (oder „Rückschritts“) von Versionen derselben Website (Nielsen 2001)
- (2) quantitative Erfassung der Differenz der Usability verschiedener konkreter

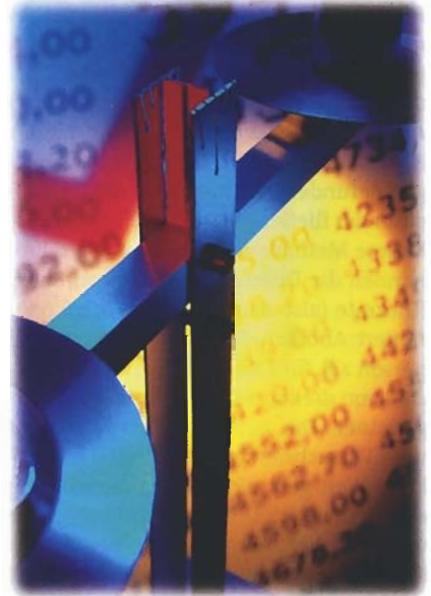
Websites oder ein Benchmarking an einem „Best Case“

(3) quantitative Erfassung der Unterschiede verschiedener Typen von Websites (etwa: werbende versus informative), verschiedener Branchen oder verschiedener Länder

Bei der Methode der heuristischen Evaluation, nach der Experten die Usability von Websites nach einem Kriterienkatalog abschätzen, liegen bereits Ansätze zu quantitativen Vergleichen vor. Erinnerung sei an den „Web Usability Index“ (Harms/Schweibenz 2002) oder an die komparative Analyse „durchschnittlicher“ Websites ausgewählter Industrie- und Dienstleistungsbranchen (Kim/Shaw/Schneider 2003). Ein zentral wichtiges Verfahren ist das Task Based Testing, bei dem Versuchspersonen konkrete Aufgaben innerhalb einer Website bewältigen müssen. Erst diese Nutzertests geben – im Kontext mit heuristischer Evaluation (und ggf. weiteren Methoden) – eine zufriedenstellende Erfassung der Usability einer Website. Erfasst werden nämlich u.a. die Navigationsmöglichkeiten (mit klaren Wegen, Umwegen oder Sackgassen) innerhalb einer Website durch Nutzer. Für einen solchen Usability-Nutzertest mit dem Schwerpunkt auf der quantitativen Analyse des Navigationssystems stellen wir einen komparativ einsetzbaren Kennwert vor.

Klickraten und Abbruchhäufigkeiten als Indikatoren der Güte des Navigationssystems

Aus der Gesamtmenge der Analysemethoden der Website-Usability (darunter Thinking-aloud Nutzertest, Teaching Back, heuristische Evaluation, Cognitive Walkthrough sowie dem Einsatz eines Usability Labors; vgl. Roßmann 2002a, 2002b) geben Nutzertests die wohl am



weitesten verbreitete Methode ab. Gemäß Fichter (2001) und Roßmann werden vom Versuchsleiter für die Website typische Aufgaben entworfen und Probanden zur Bearbeitung vorgelegt. Von einer Startseite ausgehend, werden die Pfade sowie die Klickzahl der Probanden bis zur Zielseite protokolliert. Der Versuch wird nach einer Maximalzeit abgebrochen. Wir erfassen also folgende Werte:

- die günstigste Anzahl der Klicks vom Start zur jeweiligen Zielseite (vom Versuchsleiter zu eruieren)
- die Zahl an Abbrüchen (wenn die Probanden innerhalb der vorgegebenen Zeit die Zielseite gar nicht gefunden haben)
- die Zahl an Klicks, die die Probanden bis zum Ziel benötigten.

Die durchschnittliche Klickzahl und die Zahl der Abbrüche stehen in der bisherigen Methodenliteratur zur Usability in keinem direkten Verhältnis zueinander. Roßmann zählt die Klickraten sowie die Zahl der Abbrüche getrennt. (Sie braucht natürlich auch keinen vergleichenden Indikator, da sie nur eine einzige Website untersucht hat.) Dazu ein Beispiel (entnommen: Roßmann 2002a, 25): Die optimale Klickzahl sei vier. Von den 40 Probanden haben drei (7,5 Prozent) die Aufgabe nicht bewältigt. Die restlichen 37 (92,5 Prozent) benötigen im arithmetischen Mittel 6,2 Klicks, der Median liegt bei fünf Klicks. Wir haben drei Zahlen

(Abbruchquote, arithmetisches Mittel, Median) vor uns, die – obwohl dasselbe messend – nicht miteinander verbunden sind. Unseres Erachtens ist es einfacher, mit einem Kennwert zu arbeiten, der eine Verbindung dieser Aspekte darstellt. Abbrüche repräsentieren einen negativen Aspekt: Der Nutzer hat es nicht geschafft, die Aufgabe zu lösen. Die positive Seite ist die Anzahl der Nutzer, die die Aufgabe mit der optimalen Klickzahl gelöst haben. Hinzu wird noch der Anteil der Nutzer genommen, die die Aufgabe gelöst haben, jedoch mit einer höheren Klickzahl als unbedingt nötig. So erhält man drei Aspekte (Abbrüche, Lösung mit kürzester Klickzahl gefunden, Lösung auf Umwegen gefunden), die in einen Kennwert ineinander fließen.

Unsere Methode errechnet für jeden Probanden die Differenz von der optimalen Klickrate (also dem kürzesten Weg) oder notiert Abbruch. Braucht eine Versuchsperson z.B. fünf Klicks bei einer Optimalzahl von drei, so wird der Aufgabe „2“ zugeordnet. Die optimale Klickzahl wird vom Versuchsleiter festgestellt. Der konkrete Wert geht nicht in die Berechnung ein, sondern stets nur die Differenz der von den Versuchspersonen benötigten Klickzahl und der Optimalzahl. Notiert wird jeweils die Anzahl der Probanden, die einen gleichen Klickwert aufweisen, also alle die mit der Optimalzahl, mit Optimalzahl + 1, Optimalzahl + 2 usw. bis Optimalzahl + 6, danach diejenigen, die mehr als sechs Klicks über dem Optimum liegen und letztlich die Anzahl der Probanden, die abbrechen mussten. Anschließend wird die Anzahl der Nutzer als relative Häufigkeit (in Prozent) ausgerechnet, um Ergebnisse mit unterschiedlichen

Anzahlen von Testpersonen miteinander vergleichen zu können.

Aus der so entstehenden Tabelle lassen sich Problemstellungen bei einzelnen Aufgaben direkt erkennen, zum Beispiel, ob eine bestimmte Aufgabe besser bewerkstelligt wird als andere. Um jedoch eine allgemeinere Aussage über die zu testende Website geben zu können und einen Vergleich mit anderen Homepages zu ermöglichen, sind wir noch einen Schritt weiter gegangen. Wir haben den relativen Häufigkeiten der erreichten Klickraten Gewichtungswerte zugeordnet. So gibt es für den Nutzeranteil, der genau die optimale Anzahl an Klicks gebraucht hat, die Höchstpunktzahl zehn. Der Nutzeranteil, der einen zusätzlichen Klick benötigt hat, wird mit neun Punkten bewertet. So werden absteigend die gebrauchten Klickzahlen bewertet, für mehr als sechs zusätzliche Klicks gibt es letztendlich drei Punkte. Abbrüche wurden mit null Punkten bewertet, da ein Abbruch bedeutet, dass die Aufgabe nicht gelöst wurde, d.h. die Website ist hier für den Nutzer sozusagen „durchgefallen“. Diese Punkte werden anschließend mit den entsprechenden Prozentzahlen multipliziert und addiert.

Eine solche Punktzahl wird für jede der Aufgaben, aus denen ein Test besteht, errechnet. Diese Punktzahlen werden dann addiert und anschließend durch die Anzahl der Aufgaben geteilt, um einen Mittelwert zu erhalten. Dieser Wert steht für die **mittlere Güte des Navigationssystems einer Website**. Maximal erreichbar sind 1.000 Punkte.

Mit dieser „mittleren Güte“ lassen sich problemlos Vergleiche anstellen. Die mittlere Güte „unserer“ Testwebsite kann im Sinne

des Benchmarking mit der Site mit der Höchstpunktzahl verglichen werden. Hat die Website A beispielsweise 750 Punkte, und Website B hat 470, so lässt sich sagen, dass die Nutzer mit B größere Navigationsprobleme haben als mit A. Testet man mehrere Websites eines gegebenen Typus, so werden die einzelnen Messwerte mittels des arithmetischen Mittels verdichtet und geben Auskunft über die mittlere Güte der Navigationssysteme der Stichprobe.

Ein Beispiel

Das Forschungsprojekt E.V.I. Heine (2003) hat acht Websites getestet. Das Projekt arbeitete mit Testgruppen, so dass einige Gruppenmitglieder Zeit hatten, die Klicks und allfällige Bemerkungen (analog dem Thinking-aloud-Test) zu notieren. Gearbeitet wurde mit acht Gruppen, denen jeweils zehn Aufgaben vorgelegt worden sind. Bearbeitungszeit waren maximal drei Minuten. Wir sehen in der Tabelle beispielhaft die Ergebnisse für die Shop-Website von Quelle.

Schauen wir uns in der Tabelle Aufgabe 6 an. Die Aufgabenstellung lautet: „Welche Prämien gibt es bei 5000 Digits?“. Der kürzeste Weg zur Information läuft über insgesamt vier Klicks (also Optimalzahl = 4). Vier der acht Testgruppen brauchten diese vier Klicks, drei Gruppen benötigten einen Klick mehr und eine Gruppe musste abbrechen. Diese Rohdaten stehen im oberen Teil der Tabelle. Im mittleren Teil errechnen wir die relativen Häufigkeiten. In der Zeile „Optimalzahl“ steht nunmehr 50, sind doch vier von acht Gruppen 50 Prozent. Im unteren Teil werden die relativen Häufigkeiten gewichtet und die gewichteten Werte addiert.

So ergibt sich bei dieser Aufgabe eine Navigationsgüte von 837,5.

Wie viele Probanden?

Und welche?

Wie viele Aufgaben?

Wie viele Testpersonen werden benötigt? Wie viele Nutzertypen müssen in den Tests berücksichtigt werden? Gemäß Caulton (2001) haben alle Nutzergruppen die gleiche Wahrscheinlichkeit, über die Usability-Probleme einer Website zu stolpern. Hiernach wäre es irrelevant, dass das Testdesign ausdrücklich

	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7	Frage 8	Frage 9	Frage 10
Optimum	6	3	1	3	2	4	1	0	3	6
Op+1	1	0	2	1	6	3	3	1	2	2
Op+2	0	3	2	0	0	0	1	2	1	0
Op+3	0	1	1	0	0	0	0	5	1	0
Op+4	0	1	0	0	0	0	0	0	0	0
Op+5	0	0	0	2	0	0	0	0	0	0
Op+6	0	0	1	1	0	0	0	0	1	0
Op>6	0	0	1	1	0	0	0	0	0	0
Abbruch	1	0	0	0	0	1	3	0	0	0
Optimalwert	3	1	1	1	2	4	8	1	3	10
Op (in %)	75	37,5	12,5	37,5	25	50	12,5	0	37,5	75
Op+1	12,5	0	25	12,5	75	37,5	37,5	12,5	25	25
Op+2	0	37,5	25	0	0	0	12,5	25	12,5	0
Op+3	0	12,5	12,5	0	0	0	0	62,5	12,5	0
Op+4	0	12,5	0	0	0	0	0	0	0	0
Op+5	0	0	0	25	0	0	0	0	0	0
Op+6	0	0	12,5	12,5	0	0	0	0	12,5	0
Op>6	0	0	12,5	12,5	0	0	0	0	0	0
Abbruch	12,5	0	0	0	0	12,5	37,5	0	0	0
Summe	100	100	100	100	100	100	100	100	100	100
Gew = 10	750	375	125	375	250	500	125	0	375	750
Gew = 9	112,5	0	225	112,5	675	337,5	337,5	112,5	225	225
Gew = 8	0	300	200	0	0	0	100	200	100	0
Gew = 7	0	87,5	87,5	0	0	0	0	437,5	87,5	0
Gew = 6	0	75	0	0	0	0	0	0	0	0
Gew = 5	0	0	0	125	0	0	0	0	0	0
Gew = 4	0	0	50	50	0	0	0	0	50	0
Gew = 3	0	0	37,5	37,5	0	0	0	0	0	0
Gew = 0	0	0	0	0	0	0	0	0	0	0
gew. Summe	862,5	837,5	725	700	925	837,5	562,5	750	837,5	975
Mittelwert										801,25

Tabelle: Klickraten und Abbrüche beim Test des Navigationssystems einer Website. (Acht Testgruppen, zehn Aufgaben). Oberer Teil: Rohwerte, mittlerer Teil: relative Häufigkeiten, unterer Teil: gewichtete relative Häufigkeiten. Der Gütewert des Navigationssystems ist der Mittelwert der Ergebnisse für die einzelnen Aufgaben (hier: 801,25).

unterschiedliche Gruppen von Nutzern ausweist. Nielsen (2001) hingegen unterscheidet zwischen „Neulingen“ und „erfahrenen Nutzern“. Eine Variante wäre auch, nur solche Probanden zu wählen, die der Zielgruppe der zu testenden Websites entsprechen. Für unsere Variante der Nutzertests ist die Caulton-Hypothese allerdings noch nicht geprüft worden.

Bei der Anzahl der Versuchspersonen gibt Jakob Nielsen (2001) „mehr als fünf Nutzer“ an; er selbst arbeitet jedoch in der Regel mit 20 Probanden. Nicola Roßmann (2002, 22) arbeitete sogar mit 40 Versuchspersonen, 20 Information Professionals und 20 „Internet-Laien“. Intuitiv scheint die These zuzutreffen: je mehr desto besser.

Letztlich stellt sich noch die Frage nach der richtigen Anzahl der Aufgaben. Roßmann (2002, 21) wählt sechs Aufgaben; wir schlagen zehn vor. Hier gilt offensichtlich ebenfalls, dass mit der Anzahl der Aufgaben die Aussagekraft der Ergebnisse ansteigt. Bei der Anzahl der Versuchspersonen sowie bei der Zahl der Aufgaben dürfte der Forschungsetat eine Grenze darstellen.

Die Probleme um Zusammensetzung der Testgruppe, deren Anzahl sowie der Anzahl der Fragen können nicht abschließend gelöst werden. Hier steht Forschungsbedarf an.

Literatur

Caulton, David A. (2001): Relaxing the homogeneity assumption in usability testing. – In: Behaviour & Information Technology 20 (2001), 1-7.
 Fahrenkrog, Gabriele; Marahrens Oliver; Bittner, Ewald (2002): Des Surfers Leid, des Surfers Freud. Web Usability und wie man sie testet. – In: NfD. In-

formation: Wissenschaft und Praxis 53 (2002), 73-81.
 Fichter, Darlene (2001): Testing the Web site usability waters. – In: Online 25 (2001) Nr. 2, 78-80.
 Hackos, J.T.; Redish, J.C. (1998): User and Task Analysis for Interface Design. – New York: John Wiley & Sons, 1998.
 Harms, Ilse; Schweibenz, Werner (2002): Usability Evaluation von Web-Angeboten mit dem Web Usability Index. – In: Ralph Schmidt (Hrsg.): Content in Context. Perspektiven der Informationsdienstleistung. 24. Online-Tagung der DGI. – Frankfurt: DGI, 2002, 283-292.
 Heine, E.V.I. (2003): Usability von Navigationssystemen im E-Commerce und bei informativen Websites – des Nutzers Odyssee. – In: Information – Wissenschaft und Praxis 54 (2003), S. 405-414.
 Hess, Dominique (2002): Usability vor Marke vor System. – In: Organisator Nr. 6 (2002), 16-18.
 Kim, Sung-Eon; Shaw, Thomas; Schneider, Helmut (2003): Web site design benchmarking within industry groups. – In: Internet Research. Electronic Networking Applications and Policy 13 (2003), N. 1, 17-26.
 Nielsen, Jakob (2000): Designing Web Usability. – Indianapolis, Ind.: New Riders Publ., 2000. – (Dt. Erfolg des Einfachen. München: Markt+Technik, 2000).
 Nielsen, Jakob (2001): Usability metrics. – In: Jakob Nielsen's Alertbox, January 21, 2001. – Online: <http://useit.com/alertbox/20010121.html>.
 Roßmann, Nicola (2002a): Website-Usability. Landtag NRW. – Köln: FH Köln, 2002. – (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 34).
 Roßmann, Nicola (2002b): Die Usability der Homepage des Landtags NRW. – In: Information – Wissenschaft und Praxis 53 (2002), 461-467.

Usability; Benutzerforschung; Informationswissenschaft; Internet; Benutzerfreundlichkeit; Kennwert

DIE AUTOREN

Monika Röttger

Jahrgang 1979. Seit 1999 Studium der Romanistischen Philologie, der Neueren und Neuesten Geschichte und der Informationswissenschaft an der Heinrich-Heine-Universität Düsseldorf. 2002 Auslandssemester in Nantes. Seit 1997 freie Mitarbeiterin beim „Remscheider Generalanzeiger“.



Wolfgang G. Stock

Jahrgang 1953. Inhaber des Lehrstuhls für Informationswissenschaft der Heinrich-Heine-Universität Düsseldorf. Heinrich-Heine-Universität Düsseldorf Institut für Sprache und Information Abteilung für Informationswissenschaft Universitätsstraße 1 40225 Düsseldorf www.phil-fak.uni-duesseldorf.de/infowiss E-Mail: stock@phil-fak.uni-duesseldorf.de

www.in-netman.de



SOFTWARE

HAN

Hidden Automatic Navigator

- Management von eJournals
- Navigation ohne Passwörter (Single-Sign-On)
- Verschmelzung von CD-ROM- und Internetangebote
- Anbindung an die Elektronische Zeitschriften Bibliothek - Regensburg

NetMan XP

Das Applikationsportal

- Management von Bibliotheksanwendungen
- Zentrale Administration aller Medien & Einstellungen
- Aufbau digitaler und virtueller Bibliotheken
- Aufbau von Multimediazentren und Selbstlernzentren

Inetbib-Tagung
 10.11. - 12.11.2003
 Frankfurt / Main in der
 "Deutschen Bibliothek"

solutions for your application and information management

H+H Software GmbH · www.hh-software.com · frh@hh-software.com
 Maschmühlenweg 8-10 · D-37073 Göttingen · fon +49 (0) 551-52208-0 · fax +49 (0) 551-52208-25