

IA von Websites: asynchrone Remote-Tests und Laborstudien im Vergleich

Florian Meier¹, Christian Wolff²

Lehrstuhl für Informationswissenschaft¹ bzw. Medieninformatik², Universität Regensburg

Zusammenfassung

Der Beitrag zeigt auf, wie Crowdsourcing-Verfahren für *asynchronous remote usability testing* eingesetzt werden können. Das konkrete Szenario hierfür ist die Untersuchung der Informationsarchitektur (IA) von Websites mit Hilfe des Tree-Test und des Navigations-Stress-Test. Für beide Methoden wurden Crowdsourcing-kompatible Online-Verfahren entwickelt bzw. angepasst. In einer Vergleichsstudie werden jeweils gleiche Aufgaben sowohl über Crowdsourcing-Plattformen als auch in einem Laborsetting getestet. Diese empirische Studie zeigt, dass sich vergleichbare Ergebnisse erzielen lassen, wobei sich für die Crowdsourcing-basierte Untersuchung mit vergleichsweise geringem Aufwand hinreichend viele Testpersonen rekrutieren lassen. Der Beitrag ist insofern auch im Sinne des *network as the extension of the usability laboratory* zu verstehen.

1 Einleitung

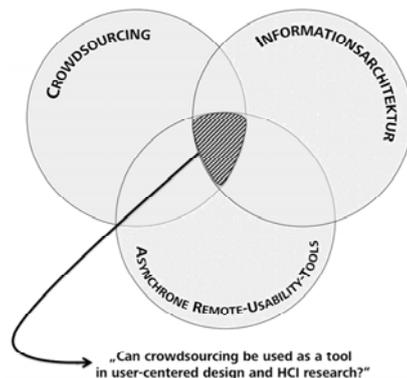


Abbildung 1: Verortung der Studie

Der mit Usability-Studien verbundene erhebliche Zeit- und Kostenaufwand ist nach wie vor ein relevanter Faktor der weiteren Etablierung von benutzerzentrierten Entwicklungs- und Bewertungsverfahren. Neben dem bekannten Konzept der *discount usability*, wie es seit langem von Jacob Nielsen und anderen vorgeschlagen wird (Nielsen 2009), konnten sich zuletzt eine ganze Reihe von Verfahren etablieren, die über die engen Beschränkungen von Laborstudien hinausgehen (Albert et al. 2010). Damit eröffnen sich erweiterte Perspektiven für die Durchführung empirischer Studien: Über netzbasierte Rekrutierungstechniken lassen sich andere Benutzergruppen erreichen, die Durchführungskosten sind gerade bei Nutzung von

Crowdsourcing-Plattformen vergleichsweise niedrig und auf diese Weise können auch deutlich größere Teilnehmerzahlen für Usability-Studien erreicht werden. Im Mittelpunkt des vorliegenden Beitrags steht daher die Frage, inwiefern sich Crowdsourcing-Plattformen in Verbindung mit *asynchronous remote usability testing*-Diensten für die Bewertung der Informationsarchitektur von Websites eignen (Abbildung 1). Im nachfolgenden Beitrag gehen wir zunächst auf *asynchronous remote usability testing* und Crowdsourcing ein (Kap. 2). Kap. 3 stellt den Anwendungsbereich Informationsarchitektur und die hierfür bekannten Bewertungsverfahren vor. Die Vergleichsstudie zur Bewertung der Informationsarchitektur von Websites mit Hilfe von Tree-Test und Navigations-Stress-Test erläutert Kap. 4, wesentliche Ergebnisse werden in Kap. 5 diskutiert. Ein Fazit schließt sich an (Kap. 6).

2 *Asynchronous remote usability testing* und Crowdsourcing

Die jüngst gewachsene Popularität netzbasierter Werkzeuge, die die Durchführung von Usability-Studien erlauben, die Verfügbarkeit von Crowdsourcing-Plattformen und die Defizite bisheriger Bewertungsverfahren für die Informationsarchitektur von Websites haben die vorliegende Studie angeregt. *Usability as a service*-Dienste bzw. das *online outsourcing* von Usability-Untersuchungen haben sich mittlerweile etabliert (Kaushik 2010). So liegen auch eine Reihe von Vergleichsstudien vor, die *remote usability testing*-Verfahren mit Laborstudien vergleichen (Meier 2012, 10f, West & Lehman 2006). Gegenüber dem lange etablierten Methodeninventar im Bereich von Usability und User Experience (Sarodnick & Brau 2011) ergeben sich die im Begriff des *asynchronous remote usability testing* schon erkennbaren Unterschiede:

1. Testperson und Testleiter sind räumlich getrennt, die Versuchsdurchführung erfolgt toolbasiert mit Hilfe von bzw. über Netzwerke, also *remote* (Bolt & Tulathimutte 2010). Solange der Test weiterhin synchron unter der unmittelbaren Kontrolle des Testleiters erfolgt (und ggf. auch weitere *remote* Beobachter teilnehmen können), ist der Ressourcenbedarf dem Labortest vergleichbar oder sogar höher.
2. Als zweite wesentliche Änderung kommt die zeitliche (und letztlich auch organisatorische) Entkoppelung hinzu: Ein asynchroner Test findet zu unbestimmter Zeit und ohne unmittelbare Kontrolle durch den Testleiter statt (*unmoderated, automated, selfguided* etc.). Für diesen Bereich liegen bereits zahlreiche Methoden und Serviceangebote vor (Hassler 2010), die z. B. die Mausinteraktion des Benutzers mit einer Website oder einer Software auf seinem Client protokollieren und an den Testleiter übermitteln.

Bolt & Tulathimutte (Bolt & Tulathimutte 2010, 128) klassifizieren die verschiedenen Methoden des *remote usability testing* wesentlich nach den Dimensionen qualitativ/moderiert vs. quantitativ/automatisiert einerseits sowie konkret-testend vs. konzeptbewertend andererseits. Die Bewertung von Aufbaustrukturen durch einen Tree-Test wie in der vorliegenden Studie fällt demnach in die Kategorien quantitativ/automatisiert und konzeptbewertend.

Die nachfolgende Abbildung 2 klassifiziert verschiedene Werkzeuge für das *asynchronous remote usability testing* nach den Kriterien Art erhobener Daten sowie inhaltlicher Bezugsbereich und eingesetzte Tools (ausführliche Erläuterungen zu den einzelnen Tools bei Albert 2010).

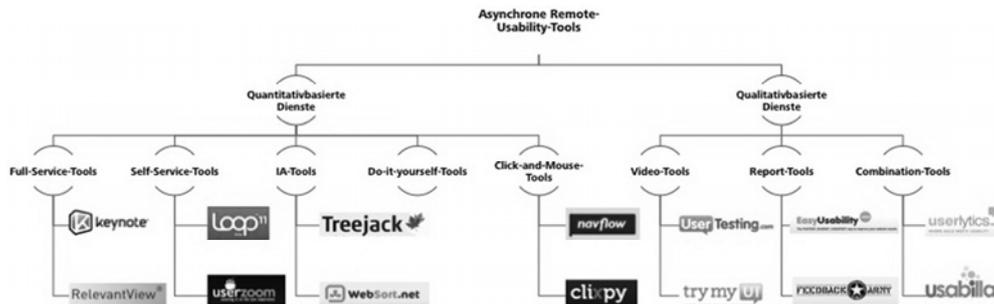


Abbildung 2: Klassifikation von asynchronen Remote-Usability-Tools nach Albert (2010)

Die Koppelung solcher Werkzeuge mit *Crowdsourcing* (Howe 2006) – vor allem als Rekrutierungsstrategie – ist naheliegend, wobei unterschiedliche Koppelungsformen zwischen *remote usability testing*-Plattform und Crowdsourcing-Dienst existieren (Meier 2012, S. 37ff.). Der von Jeff Howe geprägte Begriff des *Crowdsourcing* bezeichnet die Auslagerung von Unternehmensprozessen oder Tätigkeiten an eine breite Masse von Freizeitarbeitern im Web (Howe 2006). Die Anwendungsfälle sind dabei von sehr starker Heterogenität geprägt und reichen von einfachen repetitiven Aufgaben wie dem Verschlagworten von Bildern, über das Designen von Logos, bis hin zur Erarbeitung von Lösungen für komplexe Forschungsprobleme. Erste Studien zur Nutzung von Crowdsourcing-Plattformen im Kontext der Usability liegen bereits vor (z. B. Tidball & Stappers 2011), allerdings kaum für die unmittelbare Bewertung von konkreten Systemparametern wie in der vorliegenden Arbeit (Kittur, Chi & Shu 2008). Im Unterschied zu stärker spiel- und motivationsorientierten Ansätzen wie dem der *Human Computation* (von Ahn 2006) erfolgt beim Crowdsourcing die Motivation der Teilnehmer (auch) über einen unmittelbaren monetären Anreiz, d. h. die über das Netz für die Aufgabebearbeitung rekrutierten Teilnehmer bekommen für erledigte Aufgaben eine (kleine) Entlohnung.

3 Informationsarchitektur und ihre Bewertung

Morville & Rosenfeld (Morville & Rosenfeld 2006, 4) definieren Informationsarchitektur bezüglich vier Teilaspekte. Einer dieser Aspekte lautet wie folgt:

*“in-for-ma-tion ar-chi-tec-ture n.
[...].3. The art and science of shaping information products and experiences to support usability and findability [...].”*

Dabei wird deutlich, dass die Informationsarchitektur eng mit Fragen der Usability verbunden ist. Insofern kann es kaum überraschen, dass auch gängige Verfahren der Bewertung von Informationsarchitektur aus dem Usability-Bereich bekannt sind, wie etwa das Card-Sorting (Toub 2000). Die nachfolgende Tabelle gibt eine Übersicht der für die Bewertung unterschiedlicher Aspekte der Informationsarchitektur bisher vorgeschlagenen Methoden.

Methode	Bezugsbereich
Card sorting	Organisation, Benennung
Checklist review	Organisation, Benennung, Navigation
Goodness rating	Organisation
Navigations-Stress-Test (nach Instone)	Benennung, Navigation
Paper prototype testing (Usability-Test)	Organisation, Benennung, Navigation
Speeded sentence verification	Organisation
Testing a simple navigational interface (Navigations-Stress-Test)	Benennung, Navigation
Testing with pencil and paper (Tree-Test)	Organisation, Benennung
Tracking traversal patterns	Organisation, Benennung
Tree testing	Organisation, Benennung
Trunk testing (Navigations-Stress-Test)	Benennung, Navigation
Usability test of IA (Tree-Test)	Organisation, Benennung

Tabelle 1: Bewertungsmethoden für die Informationsarchitektur (Meier 2012, 106)

Für die vorliegende Arbeit wurden der Tree-Test und der von Keith Instone entwickelte Navigations-Stress-Test (Instone 2000) ausgewählt. Der Tree-Test (auch als *reversed* Card-Sorting bezeichnet) ist eine Testmethode, bei der der Nutzer eine bereits bestehende Informationsarchitektur erhält und dazu aufgefordert wird, eine Suchaufgabe zu erfüllen, indem er eine Kategorie bzw. Unterkategorien wählt, die die zur Suchaufgabe passende Information enthalten könnte. Der Navigations-Stress-Test ist ein einfaches und ursprünglich papiergebundenes Bewertungsverfahren, bei dem der Benutzer wesentliche Fragen zu Struktur und den Navigationsmöglichkeiten einer Website („Wo bin ich?“, „Wie komme ich eine Ebene nach oben?“, „Wie komme ich zur Homepage?“) auf einem Ausdruck der Webseite kennzeichnet. Während der dem Card-Sorting verwandte Tree-Test eher die *top-down*-Struktur einer Website in den Blick nimmt, wird mit dem Navigations-Stress-Test die *bottom-up*-Perspektive untersucht, ausgehend von Unterseiten einer Website. Beide Verfahren eignen sich grundsätzlich für unterschiedliche Durchführungsmodi. Während für den Tree-Test (und verwandte Verfahren wie Card-Sorting) nicht nur unterschiedliche Tools für die Durchführung am Rechner bzw. Servicedienste wie *Tree Jack* existieren, die sich für das Crowdsourcing-Verfahren eignen, ist eine vergleichbare Online- bzw. toolbasierte Fassung des Navigations-Stress-Test bisher nicht bekannt und wurde daher für diese Studie neu konzipiert.

4 Vergleichsstudie: Crowdsourcing vs. Labor

Um die Leistungsfähigkeit des asynchronen Remote-Usability-Tests mit Hilfe von Crowdsourcing zu untersuchen, wurde eine Vergleichsstudie durchgeführt. Am Beispiel der Website einer Hochschule wurden jeweils das Tree-Test- und das Navigations-Stress-Test-Verfahren sowohl mit Hilfe von per Crowdsourcing gewonnenen Testpersonen als auch in einem Labortest untersucht. Neben demographischen Daten wurden Kostenfaktoren (geringerer zeitlicher und finanzieller Aufwand bei Crowdsourcing) und Effektivitätsaspekte (Qualität der Bewertung, Struktur der gewonnenen Daten) erhoben (Meier 2012, 112ff).

Für die Durchführung des Crowdsourcing-Szenarios wurde ein Studiendesign gewählt, bei dem Nutzer die Tests in Form von *micro tasks* über eine Crowdsourcing-Plattform durchführen (Kittur, Chi & Suh 2008). Ausgewählt wurde hierfür der weit verbreitete *Amazon Mechanical Turk*-Service (MTurk). Die *micro tasks* (bei MTurk auch *Human Intelligence Tasks* (HITs) genannt) werden dabei mit Hilfe der Serviceplattform CrowdFlower¹ aufbereitet und bei MTurk eingestellt. Der Benutzer wird strukturiert durch den Test geführt: Nach einem Start-Screen und einer Einführung in die Funktionsweise des Tests erfolgt die Aufgabenbearbeitung. Am Ende wird ein Abschlusscreen mit Kontaktdaten präsentiert. Die Laborstudie erfolgte für beide Verfahren mit je 10 Versuchspersonen sowie denselben Aufgabenstellungen bei Tree-Test und Navigations-Stress-Test (siehe unten). Hierfür wurden bewusst Erstsemesterstudenten des Fachbereichs Medieninformatik rekrutiert, da hier die Annahme zugrunde lag, dass diese über eine ähnlich hohe Web- respektive Technikaffinität verfügen wie Crowdsourcing-Nutzer, beide aber nur geringe bis gar keine Erfahrung mit der Informationsarchitektur des Untersuchungsgegenstandes aufweisen. Beide Labortests erfolgten papierbasiert, der Test wurde per Video aufgezeichnet und vom Testleiter und zwei Assistenten begleitet. Für den Tree-Test wurde dabei eine Mappe mit vorbereiteten Hierarchieausdrucken zu den jeweiligen Oberpunkten der Navigationsstruktur vorbereitet, für den Navigations-Stress-Test lagen den Versuchspersonen schwarz-weiße Ausdrücke der zu bewertenden Webseite sowie die Aufgabenstellung vor.

4.1 Tree-Test: Bewertung der Informationsarchitektur top-down

Der Tree-Test wurde mit Hilfe des *Tree Jack*-Tools aufbereitet (O'Brien 2009). Die Benutzer bekamen typische Aufgabenstellungen für die Interaktion mit der Website wie:

„[...] 3. Um die Sauna des Sportzentrums besuchen zu können, benötigen Sie einen Sportausweis. Finden Sie die Seite, auf der Sie sich Online für einen Sportausweis registrieren können.

4. Ihre nächste Vorlesung findet in ZH 5 statt. Leider wissen Sie nicht, wo sich dieser Hörsaal befindet. Sie benötigen einen Lageplan der Universität, der Rauminformationen enthält.[...]“ (Meier 2012, 135)

¹ Erreichbar unter folgender Webadresse: <http://www.crowdflower.com>

Mit Hilfe des über *MTurk* eingebundenen *Tree Jack*-Tools konnten die Versuchspersonen Ober-Unterbegriffszuordnungen zu den einzelnen Aufgaben durchführen, die dann mit der tatsächlichen Informationsarchitektur abgeglichen werden.

4.2 Navigations-Stress-Test

Für den Navigations-Stress-Test, der an sich papiergebunden konzipiert ist, wurde eine Onlinevariante entwickelt, bei der Benutzer die einzelnen Fragen durch Einzeichnen der Antworten mit der Maus auf einem Screenshot der jeweiligen Webseite durchführen können. Auch hier wurde die Crowdsourcing-Studie über *MTurk* durchgeführt.

Frage	Aufgabe
1. Um welchen Webauftritt handelt es sich?	Kreisen Sie den Seitennamen ein und markieren Sie ihn mit einem 'C' !
2. Wie lautet der Titel der Unterseite auf der Sie sich befinden?	Kreisen Sie den Titel ein und markieren Sie ihn mit einem 'T' !
3. Wo befindet sich die Suchfunktion der Seite?	Kreisen Sie die Suchfunktion ein und markieren Sie diese mit 'S' !
4. Wie kommt man zur Startseite dieses Webauftritts?	Markieren Sie diesen Link mit einem 'H' !
5. Welcher Link entspricht der Unterseite auf der Sie sich befinden?	Kreisen Sie diesen Link ein und markieren Sie ihn mit einem 'X' !
[...]	

Tabelle 2: Fragen und Annotationsanweisungen für den Navigations-Stress-Test (Ausschnitt, Meier 2012, 141)

4.3 Auswertung

Erfasst werden jeweils Aufgabenerfolg und Bearbeitungsdauer. Die Online-Version des Tree-Tests konnte teilweise unmittelbar in *Tree Jack* ausgewertet werden. Die Mausmarkierungen des Online-Navigations-Stress-Tests wurden mit einer Musterlösung verglichen und intellektuell ausgewertet. Die Ergebnisse der Laborstudien wurden unmittelbar durch den Testleiter erfasst. Die Datenauswertung der Versuche sowie der Vor- und Nachbefragungen (demographische Angaben, Einschätzung der IT-Kompetenz, Nachbewertung der Verfahren) erfolgte in *SPSS* und *R*. Abbildung 3 zeigt links die Musterlösung als Bewertungsvorgabe und rechts die aggregierten Testergebnisse des Navigations-Stress-Tests (Mausannotationen aller Testpersonen, jeweils Ausschnitte).



Abbildung 3: Musterlösung (links) und aggregierte Mausinteraktion der Crowdsourcing-Teilnehmer für den Navigations-Stress-Test (Ausschnitte)

5 Ergebnisse

Für die Crowdsourcing-Tests konnten mit Hilfe von *MTurk* 58 Versuchspersonen (30 Tree-Test, 28 Navigations-Stress-Test) rekrutiert werden. Da bei den Laborstudien Studenten als Versuchspersonen gewonnen wurden (vgl. oben), überrascht kaum, dass beim Crowdsourcing Alters- und Bildungsstruktur der Teilnehmer deutlich heterogener ausfallen. Dennoch lässt sich feststellen, dass auch Crowdsourcing-Nutzer überwiegend einen hohen Bildungsabschluss besitzen (bei den Laborteilnehmern zu 100% das Abitur). 70% der Tree-Test- und 85% der Navigations-Stress-Test-Teilnehmer verfügen über das Abitur oder einen höheren Abschluss. Auch in Bezug auf die Webaffinität lassen sich Parallelen zwischen den Versuchsgruppen ausmachen. So geben beim Tree-Test 75% der Crowdsourcing-Probanden und 80% der Laborprobanden an, das Web mehr als 15 Stunden pro Woche zu nutzen. Nur im Falle der Laborprobanden des Navigations-Stress-Tests liegt der Wert in diesem Bereich mit nur 60% etwas niedriger. Die Bearbeitungsdauern beider Szenarien sind, wie Tabelle 3 zeigt, ebenfalls vergleichbar:

	Bearbeitungsdauer Mittelwert	SD	T-Test
Tree-Test Crowdsourcing	10,7 min	4,6 min	p=0,709
Tree-Test Labor	10,2 min	3,1 min	
Navigations-Stress-Test Crowdsourcing	4,2 min	2,0 min	p=0,296
Navigations-Stress-Test Labor	4,8 min	0,8 min	

Tabelle 3: Bearbeitungsdauern, deren Standardabweichung sowie die Ergebnisse der T-Tests im Vergleich. Vor der Anwendung des T-Tests wurden die Daten mit Hilfe des Kolmogorov-Smirnov-Tests auf Normalverteilung überprüft.

Die Erfolgsquoten für beide Testverfahren und Szenarien werden in nachfolgenden Diagrammen illustriert (Abbildung 4):

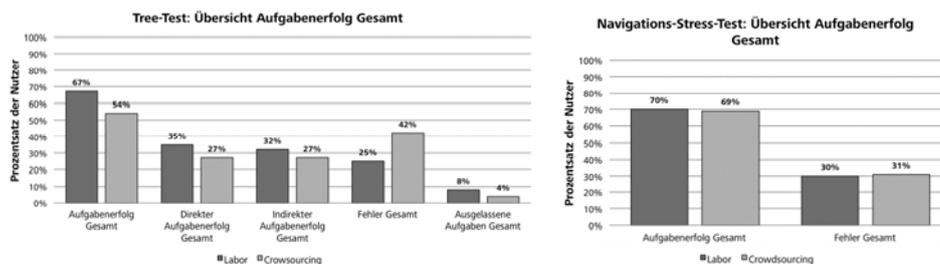


Abbildung 4: Aufgabenerfolg Tree-Test gesamt (links), Navigations-Stress-Test gesamt (rechts)

Deutlich wird, dass die Erfolgsquoten bei beiden Methoden für den Crowdsourcing-Ansatz wie für die Laborstudie in vergleichbaren Bereichen liegen. Angesichts der Tatsache, dass sowohl bei Art und Ergebnis der Rekrutierung als auch bei der konkreten Versuchsdurchführung (Toolbenutzung am Rechner vs. vom Testleiter begleitete Ausführung auf Papier) erhebliche Unterschiede bestanden, ein durchaus überraschendes Ergebnis. Für beide Erhebungsszenarien (Crowdsourcing und Labor) wurden die Rohdaten mit Hilfe des Vier-Felder-

Tests untersucht, um zu zeigen, dass in Bezug auf die jeweilige abhängige Variable (z.B. Aufgabenerfolg) kein Zusammenhang zwischen ihnen besteht. In Abhängigkeit von den Häufigkeiten kam dabei entweder der Exakte Test nach Fisher (geringe Häufigkeiten) oder der Qui-Quadrat-Test zum Einsatz (Du Prel et al. 2010, 343ff.). Wie die nachfolgende Tabelle am Beispiel des Tree-Tests zeigt, ergeben sich keine wesentlichen Unterschiede in den Ergebnissen:

Abhängige Variable	p-Wert
Aufgabenerfolg Gesamt	p=0,444
Direkter Aufgabenerfolg Gesamt	p=0,454
Indirekter Aufgabenerfolg Gesamt	p=0,740
Fehler Gesamt	p=0,156
Ausgelassene Aufgaben Gesamt	p=0,564

Table 4: Vier-Felder-Test für den Vergleich Laborstudie vs. Crowdsourcing-Studie am Beispiel des Tree-Tests. Indirekter Aufgabenerfolg bezeichnet die Tatsache, dass Versuchsteilnehmer Rückschritte in der Hierarchie machen mussten, bevor sie die richtige Kategorie finden konnten. Ein Fehler bezeichnet die Wahl einer falschen Kategorie, in der die Information nicht zu finden war.

Für den in der Bewertung des Aufgabenerfolgs einfacher strukturierten Navigations-Stress-Test ergibt sich für den Vergleich des Gesamtaufgabenerfolgs zwischen Crowdsourcing- und Laborstudie ein p-Wert von 0,953. Die Auswertung mit Bezug zu den einzelnen Aufgaben kann hier aus Platzgründen nicht wiedergegeben werden. Dabei werden allerdings Schwächen dort offensichtlich, wo der Aufgabenerfolg besonders gering ausfällt (z. B. bei der oben zitierten Aufgabe zum Finden eines Lageplans).

Auffällig ist, dass die Crowdsourcing-Teilnehmer bei den Bewertungsfragen zur Informationsarchitektur deutlich bessere Noten vergeben als die „internen“ Laborstudienteilnehmer. Allerdings geben die Crowdsourcing-Teilnehmer bei der Frage nach Erfahrung im Umgang mit Websites von Hochschulen einen höheren Wert an als die Laborteilnehmer (2,9 (Labor) vs. 3,2 (Crowdsourcing) beim Tree-Test; 2,5 (Labor) vs. 2,6 (Crowdsourcing) beim Navigations-Stress-Test (auf einer Likertskala von 1 – 5)).

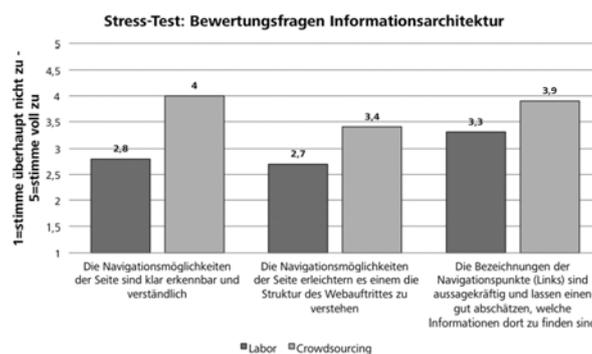


Abbildung 5: Navigations-Stress-Test: Bewertungsfragen zur Informationsarchitektur

Die Auswertung der Freitextkommentare zu den Tests liefert weitere Hinweise – sowohl zu Problemen der Informationsarchitektur als auch zur Verbesserung des Online-Navigations-Stress-Tests (z. B. Vereinfachung der Interaktion durch Bereitstellen einfacher graphischer Primitive für die Markierung).

6 Fazit

Die vorliegende Vergleichsstudie fokussiert auf eine sehr spezifische Methodenauswahl für die IA-Bewertung und hat vergleichsweise überschaubaren Umfang. Ihre wesentlichen Ergebnisse lassen sich durch die Bewertung der folgenden Annahmen zusammenfassen:

1. Für das Crowdsourcing-Szenario ließen sich mit weniger Mitteln in kürzerer Zeit mehr Versuchspersonen rekrutieren: Kosten von etwa 1 € für Crowdsourcing-Gebühren stehen deutlich höhere Aufwandsentschädigungen gegenüber, außerdem konnten in vergleichbarer Zeit erheblich mehr Probanden gewonnen werden.
2. Es stellte sich heraus, dass die Rate an Absprüngen bzw. Spam (unverwertbare Daten aus nicht ernsthaft betriebener Aufgabenbearbeitung beim Crowdsourcing) deutlich niedriger liegt als befürchtet, nämlich bei ca. 10%.
3. Unterschiede im Aufgabenerfolg lassen sich kaum ausmachen, die Qualität der Ergebnisse ist vergleichbar. Bei beiden Verfahren konnten wesentliche Probleme der Informationsarchitektur identifiziert werden. Allerdings erfolgen in der Online-Variante deutlich weniger Freitextkommentare, die als qualitative Zusatzdaten sehr wertvoll sein können.
4. Anders als bei bisherigen Studien zum *asynchronous remote usability testing* zeigt sich hier keine kürzere Bearbeitungsdauer durch die Probanden, sie sind vielmehr bei Fokussierung auf die eigentliche Bearbeitungsdauer kaum unterschiedlich.
5. Auch die Beobachtung, dass Crowdsourcing-Probanden generell besser bewerten, ließ sich bestätigen.

Im Ergebnis eröffnet die hier verglichene Crowdsourcing-Variante neue Perspektiven für die Bewertung von Informationsarchitekturen, da sie auch unabhängig von der kritischen Resource Laborinfrastruktur durchgeführt werden kann und vergleichbare Testergebnisse liefert.

Literaturverzeichnis

- Albert, B. (2010). *How to Shop for Umoderated Usability Testing Tools*. Article Johnny Holland Magazine August 25th 2010. <http://johnnyholland.org/2010/08/25/how-to-shop-for-unmoderated-usability-testing-tools/> - Zugriff am 29.11.2011
- Albert, B. & Tullis, T. & Tedesco, D. (2010). *Beyond the usability lab. Conducting large-scale user experience studies*. Amsterdam [u.a.]: Elsevier/Morgan Kaufmann.
- Bolt, N. & Tulathimutte, T. (2010). *Remote Research*. New York: Rosenfeld Media.
- Bruun, A. et al. (2009). Let your users do the testing: a comparison of three remote asynchronous usability testing methods. *Proceedings of the 27th international conference on Human factors in com-*

- puting systems (CHI '09). New York, NY, USA: ACM, S.1619-1628. Online: <http://doi.acm.org/10.1145/1518701.1518948> – Zugriff am 1.4.2012
- Du Prel, J.-B. et al. (2010). Auswahl statistischer Testverfahren. In: *Dtsch Arztebl Int* 2010; 107(19), S. 343–8. Online: <http://www.aerzteblatt.de/pdf.asp?id=74880> – Zugriff am 1.4.2012
- Hassler, M. (2010). *Web Analytics. Metriken auswerten; Besucherverhalten verstehen; Website optimieren*. Heidelberg; München [u.a.]: mitp.
- Howe, J. (2006). *The Rise of Crowdsourcing*. Online: <http://www.wired.com/wired/archive/14.06/crowds.html> – Zugriff am 1.4.2012
- Instone, K. (2000). *Navigations-Stress-Test*. Online: <http://instone.org/navstress>. Zugriff am 1.4.2012
- Kaushik, A. (2010). *Web analytics 2.0. the art of online accountability and science of customer centricity*. Indianapolis, Ind: Wiley.
- Kittur, A., Chi, E. H. & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI '08)*. New York, NY, USA: ACM, S. 453-456. Online: <http://doi.acm.org/10.1145/1357054.1357127> – Zugriff am 1.4.2012.
- Meier, Florian (2012). *Crowdsourced Information Architecture*. Evaluation der Informationsarchitektur von Websites mithilfe von Crowdsourcing und asynchronen Remote-Usability-Tests. Masterarbeit, Universität Regensburg.
- Morville, P. & Rosenfeld, L. (2006). *Information Architecture for the World Wide Web. Designing Large-Scale Web Sites*. Beijing, Cambridge, Köln [u.a.]: O'Reilly.
- Nielsen, J. (2009). *Discount Usability: 20 Years*. Alertbox, September 14. Online: <http://www.useit.com/alertbox/discount-usability.html> – Zugriff am 1.4.2012
- O'Brien, D.e (2009). *Tree-Testing. A quick way to evaluate your IA*. Online: <http://www.bboxesandarrows.com/view/tree-testing> – Zugriff am 1.4.2012
- Sarodnick, F. & Brau, H. (2011). *Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung. 2. Auflage*. Bern: Verlag Hans Huber.
- Tidball, B. E. & Stappers, P. J. (2011). Crowdsourcing Contextual User Insights for UCD. Position Paper for the SIGCHI 2011 Workshop on Crowdsourcing and Human Computation. *Proceedings of the 2011 annual conference on Human Factors in Computing Systems. (CHI 2011)*. New York: ACM. Online: <http://crowdresearch.org/chi2011-workshop/papers/tidball.pdf> – Zugriff am 1.4.2012
- Toub, S. (2000). *Evaluating Information Architecture. A Practical Guide to Assessing Web Site Organization*. Argus Center for Information Architecture Whitepaper. Online: http://argus-acia.com/white_papers/evaluating_ia.pdf – Zugriff am 1.4.2012
- West, R. & Lehman, K. (2006). Automated summative usability studies: an empirical evaluation. *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*. New York, NY, USA: ACM, S. 631-639. Online: <http://doi.acm.org/10.1145/1124772.1124867> – Zugriff am 1.4.2012