 DOI 10.1515/iwp-2012-0063 — Information. Wissenschaft & Praxis 2012; 63(5):308–313

Florian Meier, Regensburg

Crowdsourcing als Rekrutierungsstrategie im asynchronen Remote-Usability-Test

Der Beitrag stellt vor, wie Crowdsourcing-Dienste (hier *Amazon Mechanical Turk* (MTurk) und Crowdflower) in Kombination mit asynchronen Remote-Usability-Tests zur Evaluation von Websitecharakteristika genutzt werden können. Konkret wird die Informationsarchitektur von Websites mit Hilfe des Navigations-Stress-Tests bewertet. Diese Methode wird durch eine eigene Implementierung so aufbereitet, dass sowohl die Durchführung eines klassischen Labortests als auch eines Crowdsourcing-basierten Online-Tests ermöglicht wird. Der Rückgriff auf Crowdsourcing-Plattformen in Kombination mit netzbasierten Werkzeugen zielt darauf ab, den hohen finanziellen, zeitlichen und infrastrukturellen Aufwand von Laborstudien zu mindern, ohne dabei wesentliche Einbußen in der Qualität und Aussagekraft der erhobenen Daten hinnehmen zu müssen.

Deskriptoren: Benutzerfreundlichkeit, Test, Methode, Netz

Using crowdsourcing services for recruting test persons for asynchronous remote usability tests
This article investigates how crowdsourcing services (specifically Amazon Mechanical Turk and Crowdflower) combined with web based asynchronous remote usability tools can be utilized for the evaluation of website characteristics. In practice, the information architecture of websites is evaluated employing the navigation stress test. This evaluation method was adapted in a way so that not only a classical laboratory based test could be conducted, but through a digital implementation with JavaScript, the realization of a crowdsourcing based online test was made possible. The combination of crowdsourcing services with online usability tools aims at reducing the time, cost and facilities efforts of laboratory studies, without having to cope with data lacking from quality and validity.

Keywords: usability, test method, web

Le crowdsourcing comme stratégie de recrutement dans les tests d'utilisabilité asynchrone à distance
Cet article montre comment des services de crowdsourcing – également appelé externalisation ouverte – à valeur ajoutée (tels que Amazon Mechanical Turk (MTurk) et Crowdflower) peuvent être utilisés en combinaison avec des tests d'utilisabilité asynchrones à distance pour l'évaluation des caractéristiques de sites web. Plus précisément, l'architecture de l'information de sites Web est évaluée à l'aide de tests de résistance de navigation. Cette méthode est préparée par sa propre implémentation de manière à permettre la conduite d'un test de laboratoire classique de même qu'un test basé sur le crowdsourcing en ligne. Le recours à des plates-formes de crowdsourcing en combinaison avec des outils basés réseau vise à réduire les coûts financiers élevés, mais aussi le temps et l'infrastructure des tests de laboratoire, sans pour autant devoir accepter des pertes significatives au niveau de la qualité et de la fiabilité des données collectées.

Mots-clés: facilité d'utilisation, test, méthode, réseau

1 Einleitung

Neben dem Konzept der *discount usability*, wie es seit langem von Jacob Nielsen vorgeschlagen wird, konnten sich in den letzten Jahren Methoden etablieren, die darauf abzielen, den hohen finanziellen und zeitlichen Ressourcenaufwand von Usability-Studien zu reduzieren. Speziell die Rekrutierung von geeigneten Testpersonen stellt oft einen kritischen Faktor bei der Durchführung von Laboruntersuchungen dar. Durch die Kombination von Crowdsourcing als Rekrutierungsstrategie mit asynchronen Remote-Usability-Diensten kann nicht nur eine höhere Teilnehmerzahl erreicht werden, sondern es können auch schwer erreichbare Benutzergruppen zur Teilnahme an Nutzertests gewonnen werden. Somit ergeben sich neue Optionen für empirische Studien, die im Kontext des *network as the extension of the usability laboratory* (Hartson, Castillo, Kelso & Neale, 1996) zu verstehen

sind und Untersuchungen *beyond the usability lab* (Albert, Tullis, & Tedesco, 2010) betreiben. Der Beitrag gliedert sich dabei wie folgt: Nach der Darstellung wesentlicher Aspekte des asychronen Remote-Usability-Tests und dessen Unterschiede zum klassischen Labortest werden Möglichkeiten vorgestellt, wie durch Crowdsourcing eine erleichterte Testnutzerrekrutierung erfolgen kann. Anschließend erfolgt die Darstellung der Vergleichsstudie. Die Präsentation der Ergebnisse sowie ein kurzes Fazit schließen diesen Beitrag.

2 Asynchroner Remote-Usability-Test und Crowdsourcing als Rekrutierungsstrategie

Motivation dieses Beitrags ist die seit einigen Jahren stetig wachsende Popularität netzbasierter Werkzeuge zur Durchführung von Usability-Studien. Immer mehr Usability-Praktiker nehmen im Berufsalltag die Angebote von *Usability as a Service*-Applikationen wahr, bzw. nehmen ein *Online Outsourcing* von Nutzertests vor, wie das Ergebnis der Zweijahresumfrage der *Usability Professionals Association* (UPA) zeigt (Sauro 2011).

2.1 Asynchroner Remote-Usability-Test

Im Vergleich zum klassischen Methodeninventar im Bereich von Usability und User Experience, welches schon seit langem etabliert ist (Sarodnick & Brau, 2011), ergeben sich einige wesentliche Unterschiede zum asynchronen Remote-Test, die sich anhand der Adjektive *remote* und asynchron erklären lassen.

1. *Remote:* Testleiter und Testperson sind räumlich voneinander getrennt. Der Ablauf und die Untersuchungsmöglichkeiten eines Remote-Tests unterscheiden sich aber sonst nicht wesentlich von klassischen Usability-Tests im Labor. Die Aufgabenerfüllung respektive die Kommunikation erfolgt über das Telefon (z. B. *Voice over IP*), mit Hilfe von spezieller Software (also toolbasiert) und mit einer Webcam. Die unmittelbare Kontrolle des Testleiters bleibt somit erhalten, der Ressourcenbedarf ist mit dem Labortest vergleichbar oder höher.
2. *Asynchron:* Der asynchrone Remote-Test fügt zur räumlichen Trennung zwischen Testleiter und Testperson noch eine zeitliche hinzu. Er findet zu unbestimmter Zeit und ohne unmittelbare Kontrolle durch den Testleiter statt (*unmoderated, automated, self-guided etc.*), so dass es auch zu einer organisatorischen Entkopplung kommt.

Wesentlicher Vorteil von (asynchronen) Remote-Usability-Tests ist die Möglichkeit der sog. *Time-Aware-Research* (TAR) (Bolt & Tulathimutte 2010, S. 10 Testteilnehmer können so in ihrer realen Umgebung mit ihrer eigenen Hard- und Softwarekonfiguration zu einem alltäglichen Moment getestet werden (*Live Recruiting* durch HTML-Popup).

Methoden mit denen im weitesten Sinne asynchrone Nutzertests durchgeführt werden können, existieren schon länger. Hierzu zählen Report-Methoden (*user-reported cirtical incident*-Methode, Tagebuchmethode) oder verschiedene Verfahren der *Web Analytics* (Protokoll der Mausklicks). Asynchrone Remote-Usability-Tools führen eine zusätzliche Kontextualisierung durch, indem konkrete Aufgaben vorgegeben werden (wie das Erstellen eines Nutzeraccounts) und bieten somit einen Vorteil gegenüber herkömmlichen asynchronen Methoden. Bolt & Tulathimutte klassifizieren die verschiedenen Methoden des *remote usability testing* wesentlich nach den Dimensionen qualitativ/moderiert vs. quantitativ/automatisiert einerseits sowie konkret-testend vs. konzeptbewertend andererseits. Eine ähnliche Klassifikation findet sich bei Albert 2010, der eine Einteilung hinsichtlich der Art der erhobenen Daten sowie dem inhaltlichen Bezugsbereich und exemplarischen Beispielwerkzeugen vornimmt (vgl. Abbildung 1).

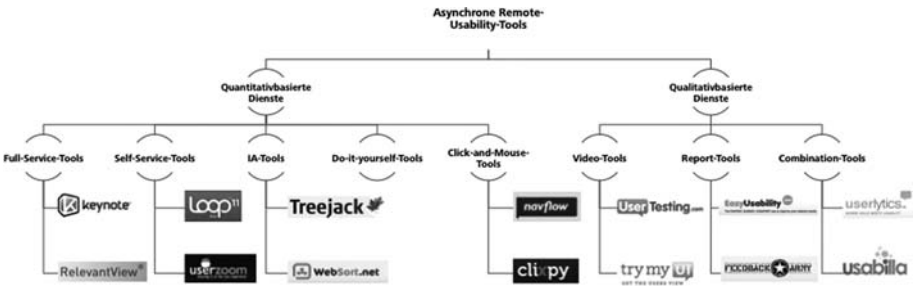

Abb. 1: Klassifikation von asynchronen Remote-Usability-Tools nach Albert 2010 (eigene Darstellung: Meier, 2012, S. 28).

2.2 Crowdsourcing als Rekrutierungsstrategie

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. 37). Der von Jeff Howe geprägte Begriff 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. Anders als bei zum Crowdsourcing verwandten Konzepten wie dem der *Human Computation* oder dem *Social Computing*, bei denen stärker spiel- und motivationsorientierte Anreize vorliegen, erfolgt die Motivation beim Crowdsourcing durch einen unmittelbaren monetären Anreiz (die über die Crowdsourcing-Plattform rekrutierten Teilnehmer erhalten für die Teilnahme am Test eine kleine finanzielle Entlohnung).

Bei der Nutzung von Crowdsourcing als Rekrutierungsstrategie lassen sich prinzipiell drei Varianten unterscheiden, welche aus der Vorgehensweise der Usability-Test-Applikationen ableitbar sind (Meier, 2011, S. 37):

1. Erstellung eines eigenen Panels mit Hilfe von Crowdsourcing
2. Direkte Nutzung von Crowdsourcing-Plattformen durch API-Anbindung
3. Eigene Distribution der Nutzertests auf Crowdsourcing-Plattformen

Um die Eignung von Crowdsourcing-Plattformen als Rekrutierungsinstrument zu überprüfen wurde eine eigene Studie durchgeführt, bei der die dritte Variante genutzt wurde. Hierbei wurde der Link zum Nutzertest über *Crowdflower* bzw. *Amazon Mechanical Turk* (MTurk) an die Testteilnehmer distribuiert. Es existieren bereits Studien die Crowdsourcing im Kontext von Usability-Untersuchungen verwenden (z. B. Tidball & Stappers 2011), jedoch nicht zur unmittelbaren Bewertung von Systemparametern, wie dies im Falle der vorliegenden Arbeit geschieht (Kittur, Chi & Suh 2008).

3 Informationsarchitektur und Navigations-Stress-Test

Untersuchungsgegenstand der Vergleichsstudie ist die Informationsarchitektur einer Hochschulwebsite. Die Disziplin der Informationsarchitektur beschäftigt sich mit der Organisation und Strukturierung von Inhalten, sowie dem Zugang zu diesen Inhalten über die angebotenen Navigationsmöglichkeiten (Morville & Rosenfeld 2006, S. 4). Für die gezielte Bewertung der Informationsarchitektur existieren nur wenige Methoden. Ein bekanntes Verfahren stellt der von Keith Instone entwickelte Navigations-Stress-Test dar (Instone 2000), bei dem der Nutzer wesentliche Frage zu Struktur und den Navigationsmöglichkeiten einer Website auf einem Papierausdruck kennzeichnen muss. Da der Navigations-Stress-Test ursprünglich papiergebunden konzipiert ist, wurde für den Nutzertest im Crowdsourcing-Szenario eine auf *HTML 5* und *JavaScript* basierende Onlinevariante des Navigations-Stress-Tests entwickelt. Hierbei konnten die Testteilnehmer die Maus zur Annotation nutzen und somit die Markierungen auf einem Screenshot der Website durchführen. Tabelle 1 zeigt die Fragen sowie die Annotationsanweisungen für den Navigations-Stress-Test, wie sie sowohl im Labor- als auch im Crowdsourcing-Szenario verwendet wurden.

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

4 Vergleichsstudie: Crowdsourcing-Labor

Um die Eignung von Crowdsourcing-Plattformen in Kombination mit asynchronen Remote-Usability-Werkzeugen zu überprüfen wurde der Navigations-Stress-Test sowohl im klassischen Laborszenario als auch im Crowdsourcing-Szenario durchgeführt. Dabei wurden neben demographischen Daten auch Ressourcenaspekte (zeitlicher und finanzieller Aufwand) sowie Effektivitätsaspekte (Aufgabenerfolg und Bearbeitungsdauer der Probanden) erhoben. Zusätzlich erhielten die Probanden einen Nachbefragungsbogen mit Bewertungsfragen zur Informationsarchitektur und Fragen zu Aspekten der Aufgabe im Allgemeinen. Der Labortest erfolgte wie für den Navigations-Stress-Test üblich papiergebunden und wurde von einem Testleiter begleitet und per Video aufgezeichnet. Der digitale Navigations-Stress-Test wurde in Form eines MTurk HITs (*Human Intelligence Task*) auf der Plattform angeboten. Die Crowdsourcing-Probanden wurden dabei strukturiert durch den Test geführt. So wurden sie mit Hilfe eines Startscreens in die Funktionsweise des Tests eingeführt, über den Zweck aufgeklärt und mit einem finalen Screen verabschiedet.

4.1 Auswertung

Bei den Testszenarien wurden Aufgabenerfolg und Bearbeitungsdauer festgehalten. Die Auswertung des Navigations-Stress-Tests erfolgt bei beiden Untersuchungsszenarien intellektuell, indem die Annotationen der Laborprobanden sowie der Crowdsourcing-Probanden mit einer Musterlösung verglichen und entsprechend festgehalten wurden. Die Datenauswertung der Versuche sowie der Vor- und Nachbefragung (demographische Angaben, IT-Kompetenz, Nachbewertung der Verfahren) erfolgte in *SPSS* und *R*. Abbildung 3 zeigt einen Ausschnitt der Musterlösung, ergo Bereiche des Untersuchungsgegenstandes mit korrekter Annotation. Abbildung 4 zeigt die aggregierte Mausinterkation (Heatmap) der Crowdsourcing-Teilnehmer.

4.2 Ergebnisse

Hinsichtlich der Stichproben (Crowdsourcing $n = 28$, Labor $n = 10$) konnte in Bezug auf die demographischen Aspekte eine gute Vergleichbarkeit erzielt werden. Die Bearbeitungsdauern sowie der Aufgabenerfolg ist bei beiden

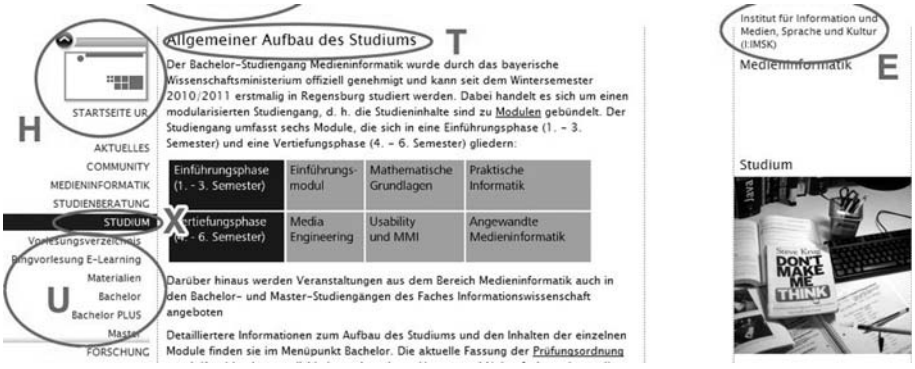

Abb. 2: Musterlösung des Navigations-Stress-Tests (Ausschnitt).

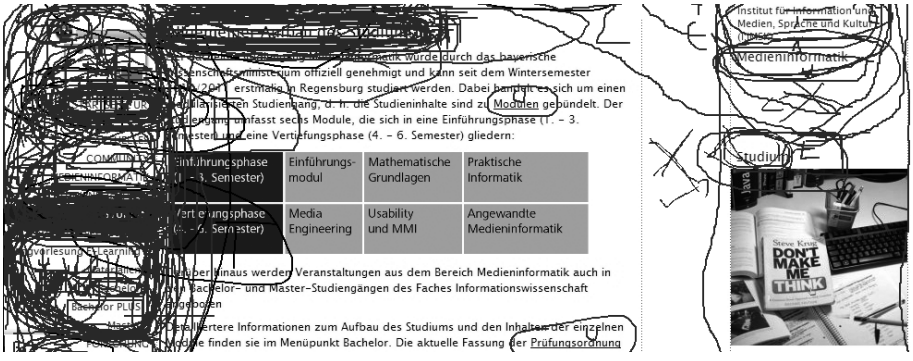

Abb. 3: Heatmap der Mausinteraktion aller Crowdsourcing-Teilnehmer (Ausschnitt).

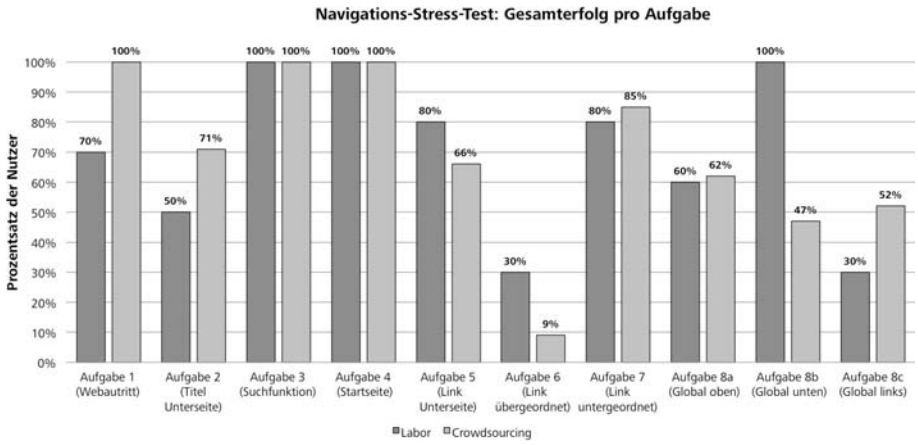

Abb. 4: Gesamterfolg pro Aufgabe für die Erhebungsszenarien Labor und Crowdsourcing im Vergleich.

Verfahren ebenfalls vergleichbar (siehe Tabelle 2 und Abbildung 4).

Mit Hilfe des Exakten Tests nach Fisher respektive dem Qui-Quadrat-Test wurden die Daten der Erhebungsszenarien in Bezug auf die jeweilige abhängige Variable (Aufgabenerfolg) auf Unabhängigkeit überprüft, welche mit einem p-Wert von 0,953 bestätigt werden konnte. Bei den Bewertungsfragen zur Informationsarchitektur fällt auf, dass die Crowdsourcing-Teilnehmer zur Informationsarchitektur bessere Noten vergeben als die Laborprobanden. Eine Tatsache, die sich mit Erfahrungen aus asynchronen Remote-Tests deckt (Sarodnick & Brau, 2011, S. 158). Freitextkommentare, welche ebenfalls im Nachbefragungsbogen erfasst wurden, lieferten sowohl Erkenntnisse über Schwächen in der Informationsarchitektur sowie Hinweise zur Verbesserung des Online-Navigations-Stress-Test (z. B. Drag-and-Drop-Funktion der verwendeten Buchstaben). Es lässt sich feststellen, dass mit Hilfe des Crowdsourcing-Verfahrens innerhalb kürzerer Zeit mehr Probanden zu einem geringeren finanziellen Aufwand rekrutiert werden konnten, als dies für den Laborversuch der Fall war. Darüber hinaus zeigt sich, dass die Rate an unverwertbaren Daten (aufgrund von Spam oder Absprüngen) niedriger lag (etwa 10 bis 25 %), als dies bei vergleichbaren asynchronen Remote-Studien der Fall ist (Meier, 2012, S. 186).

	Bearbeitungsdauer Mittelwert	**SD**	**T-Test**
Navigations-Stress-Test **Crowdsourcing**	4,2 min	2,0 min	$p = 0{,}296$
Navigations-Stress-Test **Labor**	4,8 min	0,8 min	

Tabelle 2: Bearbeitungsdauern, deren Standardabweichung sowie das Ergebnis des T-Tests für den Navigations-Stress-Test.

5 Fazit

Der Beitrag zeigt, dass die Wahl von Crowdsourcing als Rekrutierungsstrategie neue Perspektiven für die Evaluation der Informationsarchitekturen im Speziellen und der Durchführung von Usability-Studien im Allgemeinen bietet, da sie unabhängig von der kritischen Ressource Laborinfrastruktur (hoher zeitlicher und finanzieller Aufwand) durchgeführt werden können und vergleichbare Testergebnisse liefern.

Literatur

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-unmoderatedusability-testing-tools/ [19.7.2012].

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.

Hartson, H. R. & Castillo, J. C. & Kelso, J. & Neale, W. C. (1996). Remote evaluation: the network as an extension of the usability laboratory. In: Proceedings of the SIGCHI conference on Human factors in computing systems: common ground (CHI '96). New York, NY, USA: ACM, S. 228–235.

Instone, K. (2000). Navigations-Stress-Test. http://instone.org/navstress [19.7.2012].

Kittur, A.; Chi, E. H. & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing sys-

tems (CHI '08). New York, NY, USA: ACM, S. 453–456. http://doi.acm.org/10.1145/1357054.1357127 [19.7.2012].
Meier, F. (2012). Crowdsourced Information Architecture. Evaluation der Informationsarchitektur von Websites mit Hilfe von Crowdsourcing und asynchronen Remote-Usability-Tools. Masterarbeit, Universität Regensburg.
Morville, P. & Rosenfeld, L. (2006). Information Architecture for the World Wide Web. Designing Large-Scale Web Sites. 2. Auflage. Beijing, Cambridge, Köln [u. a.]: O'Reilly.
Sarodnick, F. & Brau, H. (2011). Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung. 2. Auflage. Bern: Verlag Hans Huber.
Sauro, J. (2011). The Methods UX Professionals Use. http://www.measuringusability.com/blog/uxmethods.php [19.7.2012].
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. http://crowdresearch.org/chi2011-workshop/papers/tidball.pdf [19.7.2012].

Florian Meier, M. A.
Lehrstuhl für Informationswissenschaft
Universität Regensburg
Universitätsstraße 31
93053 Regensburg
Telefon 0941 943–3464
florian.meier@ur.de

Florian Meier studierte Medienwissenschaft (Bachelor) und Informationswissenschaft (Master) an der Universität Regensburg. Seit April 2012 ist er wissenschaftlicher Mitarbeiter am Lehrstuhl für Informationswissenschaft und plant eine Dissertation zum Themengebiet Human Computation mit Fokus *gameswith a purpose*.

2. DGI-Praxistage 2012

Faktor „i" - Strategischer Umgang mit Informationen in Unternehmen

8. bis 9. November 2012, Frankfurt am Main

Donnerstag, 8. November 2012

18:30 Podiumsdiskussion
Vom wertvollsten Gut zum nervtötenden Zeitfresser ...Viel Information | Wenig Kompetenz?

19:30 Get-together

Freitag, 9. November 2012

09:15 Verschmelzung von Zeit, Raum und Information. Trends und Herausforderungen der virtuellen Arbeitswelt

11:00 Enterprise 2.0: Beispielhafter Einsatz neuer Verfahren bei der Informations- und Wissensvermittlung

13:30 „Job Title Generator" oder Job-Maschine: Schaffen neue Medien neue Berufsbilder?

Weitere Informationen zu den 2. DGI-Praxistagen finden Sie unter: http://www.dgi-info.de/

Einladung zur außerordentlichen Mitgliederversammlung

8.11.2012, 14:30 Uhr, Frankfurt am Main

Die Veranstaltung steht allen Mitgliedern der DGI offen. Die Einladung, Tagesordnung und Anlagen zur Tagesordnung finden Sie im Mitgliederbereich der DGI-Website.

Einladung zum Workshop „Die DGI der Zukunft - Entwicklung unserer gemeinnützigen Fachgesellschaft"

8.11.2012, 10:30 Uhr, Frankfurt am Main

Gestalten Sie die Kernthemen und Angebote Ihrer Fachgesellschaft mit! Welche Wünsche haben Sie für „Ihre DGI", welche Angebote sind interessant, wo können Sie oder Ihr Unternehmen sich einbringen?

Nachweis für Beitragsermäßigung

Um auch 2013 in den Genuss der ermäßigten Beiträge für die Mitgliedschaft in der DGI zu kommen, müssen alle, die diese beanspruchen wollen, bis zum 31.12.2012 per Post oder elektronisch ihre entsprechende Bescheinigunge bei der DGI-Geschäftsstelle einreichen. Das gilt vor allem für studentische Mitglieder. Vielen Dank!