S. Boll, S. Maaß & R. Malaka (Hrsg.): Workshopband Mensch & Computer 2013
München: Oldenbourg Verlag, 2013, S. 277–281

Towards Metric-based Usability Evaluation of Online Web Interfaces

Maximilian Speicher[1,2], Andreas Both[2], Martin Gaedke[1]

Chemnitz University of Technology[1]
R&D, Unister GmbH[2]

Abstract

Current approaches to web interface evaluation are tedious or do not provide sufficient information. Thus, we propose a new metric-based method building on interaction data and usability models. This would enable internet companies to evaluate interfaces at faster iteration cycles but poses new requirements to usability instruments. As a first step, we present *INUIT*—an instrument aiming at this specific purpose. A confirmatory factor analysis showed that INUIT can reasonably well describe real-world perceptions of usability while being compatible with the desired metric-based approach.

1 Motivation

In e-commerce, customer loyalty and satisfaction highly depend on web interface usability (Sauro 2010). Thus, it is crucial to perform usability evaluations for product optimization. Common approaches are inspection methods (Nielsen 1994), controlled user studies (Jeffries et al. 1991) or A/B testing. From own experience we know that the first two options are *effective* but mostly cumbersome and only performed *before* a new website or major redesign is launched. The latter option is more *efficient* and can be used with *online* web interfaces but lacks insights into users' actual behavior (Nielsen 2005). Conversions such as the number of registrations, on which A/B tests are based, can even be contradictory to usability (Nielsen 2005).

We propose a new methodological approach to usability evaluation that is both efficient and *effective by inferring a quantitative measure from user interactions.* Given interaction data and explicit usability judgments, it is possible to provide a model predicting such a measure from interactions with a web interface. Contrary to user studies, explicit usability judgments are only required during an initial step in which we collect training data for learning the model. Further evaluations can be based on real users' interactions that are fed into this model. Developers can deploy variations of an online web interface during A/B tests and are provided with usability values that are derived from automatically tracked user interactions. These values can be directly compared to identify the potentially better interface, without tedious inspections of actual interaction data (Fig. 1).

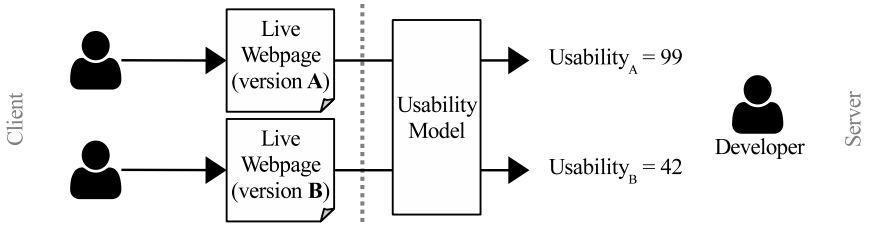

Figure 1: Concept of metric-based usability evaluation with sample values.

Existing solutions (Atterer et al. 2006; *m-pathy*[1]) leverage interaction data but do not infer usability values from these. Rather, interactions are visualized and have to be inspected by experts/developers. If metrics are provided, they are rather oriented towards conversion optimization. As another example, De Vasconcelos & Baldochi Jr. (2012) compare interactions against pre-defined optimal patterns in a remote user study setting.

The design of an adequate quantitative usability measure is a major challenge that is addressed in the remainder of this paper. The development of a complete system (Fig. 1) is our current work-in-progress.

2 Inuit: The Interface Usability Instrument

Providing a measure of usability requires an adequate instrument describing this latent variable. *Yet, it is not possible to simply take an existing one since these were not developed w.r.t. the specific purpose described above.* Particularly, our approach poses the following specific requirements:

(R1) The number of items is kept to a minimum so that users asked to fill out a questionnaire are not deterred and we can obtain a large amount of high-quality training data.

(R2) The items have the right level of abstraction so that they can be meaningfully correlated with user interactions. For example, "user confusion" can be mapped to interactions such as unstructured cursor movements.

(R3) The items can be applied to a webpage as a stand-alone entity since we want to enable correlations with client-side interactions, which are difficult to track and put into context across multiple webpages.

Existing instruments do not meet all of these specific requirements. Green & Pearson (2006) describe an instrument to measure the usability of e-commerce websites. Yet, several of their items do not have the right level of abstraction (R2) or cannot be applied to individual webpages (R3). Questionnaires such as *AttrakDiff*[2] or *UEQ*[3] show similar problems. They are rather oriented towards user experience evaluation of complete interactive products (e.g., online shops), which is considerably different from usability evaluation of a stand-alone web

1 http://www.m-pathy.com/
2 http://www.attrakdiff.de/
3 http://www.ueq-online.org/

interface (R3). For example, it is not possible to map items like *product novelty* (UEQ) onto concrete interactions (R2).

In the following we present *INUIT*—the new INterface Usability InstrumenT that aims at meeting R1–R3 above.

Item Design

We have determined the items of INUIT in a two-step process. First, we have reviewed popular usability guidelines (i.e., heuristics and checklists) with over 250 rules for usability. After eliminating all rules not consistent with R1–R3, we extracted a set of underlying factors, i.e., we investigated which of the remaining rules were different expressions of a common underlying principle. From these, we have derived a "structure" of usability (Fig. 2) based on ISO 9241-11 (1998).

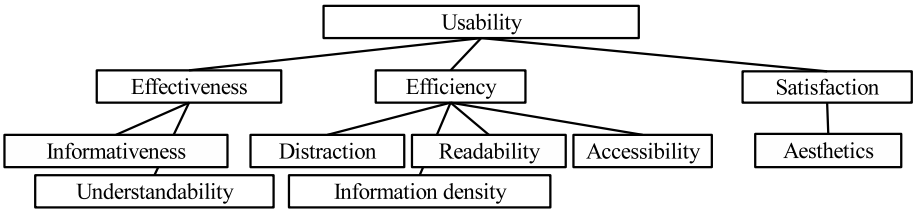

Figure 2: Structure of usability derived from guideline reviews.

Second, we conducted interviews with nine frontend/usability experts working in e-commerce (avg. age 30.44, σ=2.96). We asked them to name driving factors of usability while showing corresponding examples on the web. Subsequently, we presented them with pen and paper showing Fig. 2 and asked them to modify it so that it reflected their perception of usability with R1–R3 in mind. All of the above factors were mentioned by the experts but 38 similar statements like "a not clearly structured layout leads to user confusion" remained that did not fit into the existing set. Thus, it was necessary to add *user confusion* as a subconcept of *efficiency*. After also removing *satisfaction*, which does not comply with R2, we formulated corresponding questions to form INUIT:

Usability factor	**Question**
Informativeness	Did you find the content you were looking for?
Understandability	Could you easily understand the provided content?
Confusion	Were you confused using the webpage?
Distraction	Were you distracted by elements of the webpage?
Readability	Did typography and layout add to readability?
Information density	Was there too much information presented on too little space?
Accessibility	Was your desired content easily accessible?

Table 1: The Interface Usability Instrument.

Evaluation

To evaluate INUIT, we conducted a Confirmatory Factor Analysis (CFA; Arbuckle 2011) with a model in which all of the items directly load on the latent variable *usability*. Data for eval-

uation was obtained in a study with 81 non-unique participants (66 male, avg. age 28.43, σ=2.37) who interacted with one of four news articles and rated its usability using a questionnaire based on the above items. It was possible to take part a maximum of four times with different articles.

Results suggest that our model is a reasonably good fit to the data set, with χ^2=15.817 (df=12, p=0.2), a comparative fit index of 0.971 and a root mean square error of approximation (Arbuckle 2011) of 0.063.

3 Demo

A demo is available at "http://vsr.informatik.tu-chemnitz.de/demo/inuit". We provide the study set-up for inspection (i.e. the specifically prepared news articles including the questionnaire), the complete guideline review results and all necessary resources for reproducing the CFA with IBM SPSS Amos 20.

4 Conclusions

This paper focused on the methodological challenge of providing a new metric-based approach to usability evaluation. As a first step, we presented INUIT—an instrument aiming at this specific purpose. Results of a user study and CFA suggest that INUIT reasonably well describes real-world perceptions of usability. We are aware of the fact that usability is a difficult-to-grasp concept that cannot be forced into yes/no questions in its entirety. Thus, we plan to repeat the user study and CFA with a more complex scenario, i.e., a larger amount of users and Likert scale–based ratings of the usability items. Still, INUIT is an important step into the direction of metric-based usability evaluation, which enables internet companies to optimize their web interfaces at faster iteration cycles compared to inspections and user studies. However, due to the subjective nature of usability, the proposed approach cannot be a complete solution. It should be combined with established methods during interface development and if usability values are not significantly different between evaluated interfaces.

Acknowledgments

This work has been supported by the ESF and the Free State of Saxony.

Gefördert aus Mitteln der Europäischen Union

Europa fördert Sachsen. ESF Europäischer Sozialfonds

References

Arbuckle, J. L. (2011). *IBM SPSS Amos 20 User's Guide.* Armonk: IBM Corporation.

Atterer, R., Wnuk, M. & Schmidt, A. (2006). Knowing the User's Every Move – User Activity Tracking for Website Usability Evaluation and Implicit Interaction. In *Proc. WWW.*

De Vasconcelos, L. G. & Baldochi Jr., L. A. (2012). Towards an automatic evaluation of Web applications. In *Proc. SAC*.

Green, D. & Pearson, J. M. (2006). Development of a Website Usability Instrument based on ISO 9241-11. *JCIS, 47*(1).

ISO 9241-11 (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*. ISO, Geneva.

Jeffries, R., Miller, J., Wharton, C. & Uyeda, K. (1991). User Interface Evaluation in the Real World: A Comparison of Four Techniques. In *Proc. CHI*.

Nielsen, J. (1994). Usability Inspection Methods. In *CHI '94: Conference Companion on Human Factors in Computing Systems*.

Nielsen, J. (2005). Putting A/B Testing in Its Place. http://www.nngroup.com/articles/putting-ab-testing-in-its-place/ (2013/03/14).

Sauro, J. (2010). Does Better Usability Increase Customer Loyalty? http://www.measuringusability.com/usability-loyalty.php (2013/06/05).

Contact Information

Maximilian Speicher

TU Chemnitz, Professur VSR
Straße der Nationen 62
D-09111 Chemnitz

E-Mail: maximilian.speicher@unister.de