

Zum hohen Stellenwert einer Usability-Untersuchung in der Evaluierung einer interaktiven Informationsvisualisierung.

Markus Rester, Margit Pohl, Slaven Banovic

Technische Universität Wien, Institut für Gestaltungs- und Wirkungsforschung

Klaus Hinum, Silvia Miksch

Technische Universität Wien, Institut für Softwaretechnik und interaktive Systeme

Susanne Ohmann, Christian Popow

Medizinische Universität Wien / AKH Wien, Universitätsklinik für Neuropsychiatrie des Kindes- und Jugendalters

Zusammenfassung

Im Rahmen einer Therapie von Mädchen mit Anorexia nervosa (Magersucht) im AKH Wien werden sehr viele hochstrukturierte und zeitabhängige Daten erhoben. Gravi++, eine interaktive Informationsvisualisierung (InfoVis), wurde entwickelt, welche die PsychotherapeutInnen bei der Auswertung dieser Daten unterstützen soll.

Ein Abschätzen der Usability von Info-Vis Tools ist deshalb besonders schwierig, weil sowohl die Angemessenheit der Visualisierungsmethode für die gegebenen Daten als auch die leichte Benutzbarkeit das Verhalten der AnwenderInnen beeinflussen können. Diese Einflussfaktoren sollten analytisch getrennt werden.

Eine ausführliche Untersuchung der Usability ist erforderlich. Diese soll möglichst früh im Entwicklungsprozess beginnen und diesen begleiten. Sie soll einen eigenständigen und wichtigen Teil der Evaluierung der Visualisierungsmethode darstellen und als solche klar von anderen Teilen einer umfassenden Studie getrennt sein.

Dieser Beitrag beschreibt die Usability-Untersuchung. Verwendete Methoden waren Usability Inspektion, Heuristische Evaluierung und Fokusgruppen. Der große Umfang der erhobenen Daten machte eine systematische Auswertung notwendig. Hierzu verwendete Vorgangsweisen und dabei aufgetretene Schwierigkeiten werden im folgenden Artikel vorgestellt.

1 Einleitung

Die Evaluierung von InfoVis ist ein viel diskutiertes Thema, da klare Richtlinien sowohl für die Gestaltung als auch für die Bewertung von InfoVis fehlen, oft widersprüchlich sind oder aufgrund von besonderer NutzerInnen- und Aufgaben-Zentriertheit nicht anwendbar sind. Es gibt aber verschiedene Ebenen, auf denen angesetzt werden kann und auch muss: kontrollierte Experimente zu alternativen Elementen im Design von InfoVis, Langzeitstudien mit den wirklichen BenutzerInnen eines Systems, Usability-Untersuchungen und Vergleichsstudien von unterschiedlichen Informationsvisualisierungssystemen (Plaisant, 2004).

Gerade bei interaktiver InfoVis steigt die Bedeutung von Usability, weil ein besonders starker Zusammenhang von Usability und Utility/Usefulness auf der Hand zu liegen scheint (vgl. Grinstein et al. 2003). Aussagen über Nützlichkeit und Brauchbarkeit von InfoVis in der Anwendung können also nur mit einer expliziten Bewertung der BenutzerInnenfreundlichkeit gemacht werden. Eine an sich innovative und viel versprechende InfoVis-Methode könnte ansonsten zu Unrecht falsch eingeschätzt werden.

Methoden zur Überprüfung der Usability sollten möglichst früh in einem iterativen Design- und Entwicklungsprozess eingesetzt werden um die Qualität zu steigern. Die Erkenntnisse aus solchen Untersuchungen stellen gleichsam eine notwendige Sichtweise für eine umfassende Evaluierung einer InfoVis dar. Aber auch bei einer nachträglichen Vergleichsstudie verschiedener Systeme (z.B. Saraiya et al., 2005) müssen Fragen zur Bedienbarkeit mitbedacht werden. Eine Trennung der beiden Ebenen und eine explizite Untersuchung dieser mit dafür jeweils brauchbaren Methoden scheint ein gangbarer Weg zu sein (vgl. Tory & Möller, 2004).

Gravi++ (Hinum et al. 2005) ist eine Informationsvisualisierung zur Unterstützung von ÄrztInnen bei der Auswertung von Daten, welche im Rahmen einer Therapie von anorektischen Mädchen am AKH Wien erhoben werden.

Dies war notwendig, weil statistische Methoden in diesem Zusammenhang aufgrund der geringen Stichprobengröße problematisch sind, und beim Fehlen von Hypothesen nicht zielführend sind, da sie die Hypothesenbildung nicht adäquat unterstützen. Es handelt sich um eine interaktive Methode, die auf einer Feder-basierten Positionierung von Icons beruht. Die Studie zur Usability war Teil eines partizipativen Softwareentwicklungsprozesses.

2 Methoden

Um der Wichtigkeit einer ausführlichen Bewertung der Usability gerecht zu werden, kamen drei unterschiedliche Methoden zur Anwendung: Usability Inspektion, Heuristische Evaluierung und Fokusgruppen. Diese Auswahl ermöglicht unterschiedliche Sichtweisen auf die Fragestellung und die Ergebnisse ergänzen sich.

2.1 Usability Inspektion

Eine sehr informelle Inspektion durch nur eine Person half vor der weit aufwendigeren Heuristischen Evaluierung dabei, die offensichtlichsten Probleme zu erkennen, so dass diese im Vorfeld noch behoben werden konnten. Damit wurde sichergestellt, dass sich die späteren EvaluatorInnen nicht mit unzähligen leicht zu findenden Kleinigkeiten beschäftigen müssen, sondern sich auf das Aufspüren komplexerer Probleme konzentrieren können.

Nielsens Heuristik (Nielsen, 1994) und Tognazzinis Leitsätze (Tognazzini, 1992) bildeten die Grundlage für diese Überprüfung. Aber auch Guidelines zur Gestaltung von Multiple Views (Wang Baldonado, 2000) wurden herangezogen. Zum Beispiel stellt Konsistenz ein anerkanntes Prinzip im User Interface Design dar. Aber im Zusammenhang mit Informationsvisualisierung, und dort insb. bei Verwendung von Multiple Views, erwächst diesem Leitmotiv zusätzliche Bedeutung (vgl. Wang Baldonado, 2000).

So wurden 31 Probleme dokumentiert, die für eine Korrektur noch vor der Heuristischen Evaluierung in Frage kamen. Eines war z.B. die Verwendung von inkonsistenten Bezeichnungen (27 Fälle), die alle korrigiert wurden, weil sie leicht zu finden sind und die späteren EvaluatorInnen, welche keine Domain-ExpertInnen sein würden, nur unnötig verwirrt hätten.

2.2 Heuristische Evaluierung

27 StudentInnen untersuchten Gravi++ mit der weit verbreiteten Methode der Heuristischen Evaluierung (Nielsen, 1994). Wenngleich die gängige Meinung vorherrscht, dass 3-5 ExpertInnen als EvaluatorInnen genügen, um die meisten Probleme zu entdecken (vgl. Holzinger, 2005), entschieden wir für eine wesentlich größere Anzahl. Zum einen können die Versuchspersonen nur als Semi-ExpertInnen in Usability beschrieben werden, obwohl viele bereits Lehrveranstaltungen in den Bereichen User Interface Design oder Usability Engineering besucht hatten. Zum anderen bestand großes Interesse daran, nicht nur die meisten, sondern möglichst viele Probleme zu entdecken.

Diese große Zahl an EvaluatorInnen macht Methoden, die auf einem face-to-face Setting beruhen, unmöglich. Auch die Menge an erhobenem Material erfordert eine systematische Auswertung. Zu diesem Zweck wurde ein Report-System entwickelt (Perl, MySQL), welches folgende Eigenschaften anbietet: Screenshot-Upload, Angabe der verletzten Regel(n) der Heuristik und Problembeschreibung. Für die Auswertung der dokumentierten Probleme steht ein administrativer Zugang zu den Reports zur Verfügung (s. Abb. 1). Die Reports mit den beschriebenen Problemen wurden auf mehreren Ebenen klassifiziert:

1. Eine dreistufige Lokalisierung der Reports beschreibt den Screenshot und das dazugehörige Hauptproblem. So kann damit abgebildet werden, ob es sich z.B. um Personen-Icons auf der Arbeitsfläche oder in der Übersichtsvisualisierung handelt. Auch kann damit bei zwei sehr ähnlichen Screenshots festgehalten werden, ob sich evtl. um sehr unterschiedliche Interaktionsmuster handelt (z.B. das Hinzufügen/Entfernen von Personen mittels Drag&Drop, Kontext-Menüs oder Hauptmenü).
2. Eine eindeutige Identifikation von dokumentierten Problemen stellte die wichtigste Information dar. Einem Report können beliebig viele identifizierte Fehler zugeordnet werden, da häufig mehrere Fehler in einer Problembeschreibung erwähnt wurden.
3. Eine Differenz der Anzahl der Problembeschreibungen mit jeweils angegebener verletzter Regel in einem Report und den tatsächlich darin dokumentierten Fehlern ist möglich.
4. Eine Checkbox dient zum Markieren besonders interessanter und/oder wichtiger Reports.

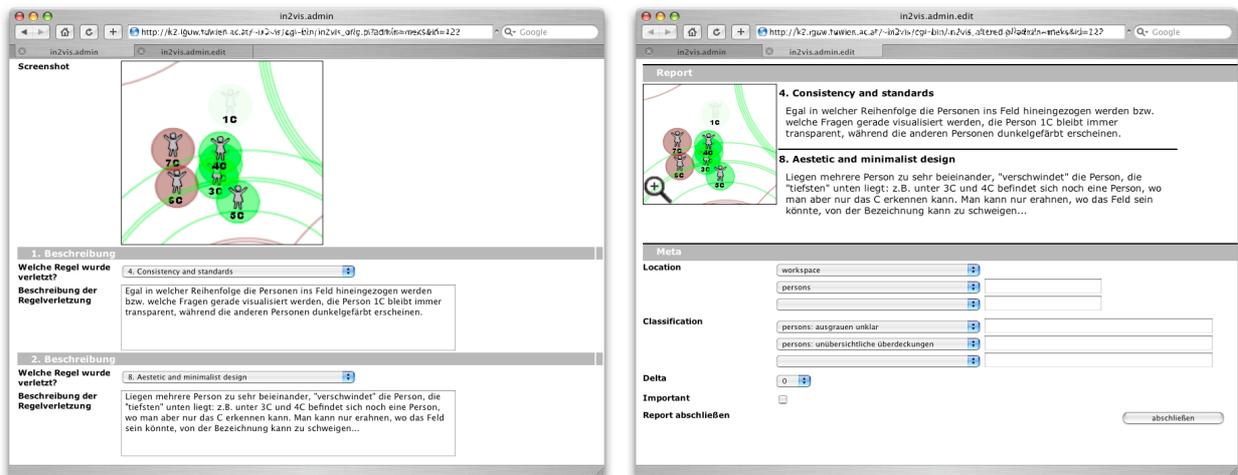


Abbildung 1: Typischer Report mit Screenshot, zwei Problembeschreibungen und Angabe der jeweils verletzten Regeln (links) und administrativer Zugang zu den Reports zum Zwecke der Klassifikation (rechts).

Eine effiziente Klassifizierung wird durch „intelligentes“ Befüllen der jeweiligen Drop-Down-Leisten jeweils in Abhängigkeit bereits ausgewählter Einträge und der Auto-Completion von Formularefeldern ausreichend unterstützt und erleichtert.

2.3 Fokusgruppen

Fokusgruppen können als strukturierte Gruppendiskussionen und -interviews beschrieben werden (s. z.B. Kuniavsky, 2003). Eine Stärke dieser Methode ist die Möglichkeit, dass durch den Diskussions- und Interaktionsprozess in der Gruppe interessante Bereiche angesprochen werden können, die keine der TeilnehmerInnen vorher – und natürlich auch nicht in einem Einzelinterview – antizipieren hätte können. Der Gruppenprozess stellt gleichermaßen eine Gefahr dar, da es z.B. durch eine dominante TeilnehmerIn zu einer starken Beeinflussung der Gruppenmeinung kommen kann. Eine ModeratorIn leitet die Diskussion anhand eines Leitfadens zu vorher festgelegten Themen.

3 Ergebnisse

27 Versuchspersonen dokumentierten in 447 Reports 513 vermeintliche Regelverstöße. Die Interpretation und Klassifikation der Problembeschreibungen ergibt 221 unterschiedliche Probleme, die in Summe 576 mal dokumentiert wurden. Ein hoher Anteil an gefundenen Bugs ergibt sich aus dem frühen Stadium des Entwicklungsprozesses, in dem die Evaluierung angesetzt war. In den Fokusgruppen gab es auf die Frage nach dem größten Usability-Problem 46 Wortmeldungen zu 27 unterschiedlichen Problemen. Detaillierte Ergebnisse zu den Verteilungen der in den Reports angegebenen verletzten Regeln und den Wortmeldungen zu den größten Usability-Problemen sind anderenorts beschrieben (s. Rester et al. 2005).

4 Diskussion

Die dreistufige Lokalisierung ist am ehesten – im Sinne eines Hilfsmittels – für die eindeutige Identifizierung von dokumentierten Problemen sinnvoll, weil mit ihrer Unterstützung sehr ähnliche Beschreibungen als dennoch unterschiedliche Interaktionsmuster erkannt werden konnten (z.B. war das Hinzufügen von zu visualisierenden Daten mittels Drag & Drop, Kontextmenü und Hauptmenü möglich). Aus den daraus folgenden Verteilungen ist jedoch kaum etwas Brauchbares für die Verbesserung der Software abzuleiten.

Die Kennzeichnung ausgewählter interessanter bzw. wichtiger Reports erwies sich als sehr praktisch, da hiermit auf exemplarischer Ebene eine Liste vorrangiger notwendiger Verbesserungen zusammengestellt werden kann. Die

Auswahl eines der aussagekräftigsten Reports zu einem bestimmten Problem enthält in vielen Fällen implizit oder explizit eine Anleitung oder einen Vorschlag für eine Korrektur.

Das Erheben der Differenz von Anzahl der eindeutigen Fehler und Anzahl der Problembeschreibungen musste ziemlich schnell eingestellt werden, da es in sehr vielen Fällen zu Doppelnennung von gleichzeitig verletzten Regeln ohne weitere Problembeschreibung kam. Gleichzeitig war die Anzahl der identifizierbaren dokumentierten Fehler teilweise viel größer als die der Beschreibungen. Die Versuchspersonen nutzten das Report-Systems zu unterschiedlich, als dass eine Auswertung dieser Größe interessante Ergebnisse hätte liefern können.

Nielsens Heuristik stellt ein brauchbares „Transportmittel“ sowohl für die Einschulung der Versuchspersonen im Bereich Usability, als auch zur Durchführung der Evaluierung dar. Sie kann als grobes Raster dienen, um zu verhindern, dass bestimmte Bereiche keine Berücksichtigung durch die Versuchspersonen finden. Die einzelnen Regeln sind jedoch sowohl in ihrer Breite als auch in ihrer Detailliertheit bzw. Abstraktheit einigermaßen unterschiedlich, sodass eine Auswertung der Verteilung wenig bis keinen Sinn macht.

Die Notwendigkeit mehr als einer Fokusgruppe durchzuführen ist durch unsere Ergebnisse erneut bestätigt worden. Die Wortmeldungen zum größten Usability-Problem in den beiden Gruppen waren teilweise disjunkt. Das kann auf einen möglicherweise vorhandenen Gruppendruck oder aber einfach auf die Unterschiedlichkeit der teilnehmenden Personen zurück zu führen sein.

Die Verwendung unterschiedlicher Methoden, deren Perspektiven sich ergänzen und damit helfen, ein besseres Gesamtverständnis zu erlangen, erwies sich als äußerst vorteilhaft. Gerade für die Gewichtung der Probleme – ein wesentlicher Faktor für konkrete Verbesserungsschritte – sind Fokusgruppen notwendig, denn die Ergebnisse der Heuristischen Evaluierung können diesbezüglich in die Irre führen.

Das Projekt „Interactive Information Visualization: Exploring and Supporting Human Reasoning Processes“ wird finanziert vom Wiener Wissenschafts-, Forschungs- und Technologiefonds [Projektnummer WWTF CI038].

5 Literaturverzeichnis

- Grinstein, G., Kobsa, A., Plaisant, C., Stasko, J. T. (2003). Which comes first, usability or utility? In *Proc. IEEE Visualization 2003*. IEEE, 605–606.
- Hinum, K., Miksch, S., Aigner, W., Ohmann S., Popow, C., Pohl, M., Rester, M. (2005). Gravi++: Interactive Information Visualization to Explore Highly Structured Temporal Data. In: *Journal of Universal Computer Science (J.UCS), Special Issue on Visual Data Mining*, Vol. 11(11), 1792–1805.
- Holzinger, A. (2005). Usability Engineering Methods for Software Developers. In *Communications of the ACM*, 48(1), 71–74.
- Kuniavsky, M. (2003). *Observing the User Experience. A Practitioner's Guide to User Research*. San Francisco, San Diego, New York: Morgan Kaufmann Publishers.
- Nielsen, J.(Hrsg.), Mack, R.L.(Hrsg.) (1994). *Usability Inspection Methods*. New York: John Wiley & Sons.
- Plaisant, C. (2004). The Challenge of Information Visualization Evaluation. In *Proc. AVI'04*, ACM Press, 109–116.
- Rester, M., Pohl, M., Hinum, K., Miksch, S., Ohmann, S., Popow, C., Banovic, S. (2005). Assessing the usability of an interactive information visualization method as the first step of a sustainable evaluation. In *Empowering Software Quality: How can Usability Engineering reach these goals?*, Vol. 198 of books@ocg.at, Wien: OCG, 31–44.
- Saraiya, P., North, C., Duca, K. (2005). An insight-based methodology for evaluating bioinformatics visualizations. In *IEEE Transactions on Visualization and Computer Graphics*, 11(4), 443–456.
- Tognazzini, B. (1992). *Tog on Interface*. Reading, MA: Addison-Wesley Publishing, 301–315.
- Tory, M., Möller, T. (2004). Human Factors in Visualization Research. In *IEEE Transactions on Visualization and Computer Graphics*, 10 (1), 72–84.
- Wang Baldonado, M. Q., Woodruff, A., Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. In *Proc. AVI'00*, ACM Press, 110–119.