

Ein Schema zur Auswahl geeigneter Evaluationsmethoden für die Evaluation von Information Retrieval Systemen mit Visualisierungskomponente

Sonja Hierl

Hochschule für Technik und Wirtschaft, Chur
CH-7000, Chur, Schweiz
sonja.hierl@fh-htwchur.ch

Abstract

Folgende Ausarbeitung begegnet dem Missstand, dass trotz der sich schnell entwickelnden Angebote von Suchmaschinen mit visueller Ergebnisrepräsentation noch kein Konsens gefunden wurde über eine gemeinsame Basis, auf deren Grundlage nachhaltige Evaluationen von Information Retrieval Systemen mit Visualisierungskomponente durchgeführt werden können. Diese Problematik wird anhand einer State-of-the-Art-Analyse aufgezeigt und es wird ein Lösungsvorschlag erarbeitet und exemplarisch getestet, der einen integrierten Ansatz zur Kombination geeigneter Evaluationsmethoden auf Grundlage eines morphologischen Rahmens empfiehlt.

1 Einführung

Auf dem Suchmaschinenmarkt ist ein Trend hin zu Information Retrieval Systemen (IRS) mit integrierter Visualisierungskomponente (VK) erkennbar.

Zum einen berichten diverse Publikationen, wie beispielsweise [Koshman, 2005; Reiterer, 2005; Zwol und Oostendorp, 2004; Reiterer, 2004; Mann, 2002; Sebrechts et al., 1999; Veeresamy und Belkin, 1996] von Forschungsprojekten zur Entwicklung solcher Applikationen, zum anderen steigern auf dem Markt befindliche Systeme wie Glookster, Kartoo, Webbrain oder Liveplasma ihren Bekanntheitsgrad insbesondere innerhalb von Unternehmen als Desktop- bzw. Enterprise Search Engines.

Ziel derartiger IRS, die in der Regel eine Visualisierung der Ergebnisrepräsentation vornehmen, ist die Optimierung der Effektivität während des Suchprozesses aus Nutzersicht. Dies geschieht beispielsweise durch die Repräsentation von Relationen zwischen selektierten Ergebnisobjekten oder der Darstellung von Themenclustern, wodurch eine Aussage über die inhaltlichen Zusammenhänge der Treffer ermöglicht wird.

[Vaughan, 2004] stellt fest, dass die Entwicklung valider Evaluationstechniken derzeit nicht Schritt halten kann mit der rapiden Geschwindigkeit der Neuentwicklungen von Suchmaschinen mit visueller Ausgabe.

Zwar gibt es diverse Studien, in denen die im Rahmen von Forschungsprojekten entwickelten IRS mit VK abschliessend einer Qualitätsprüfung unterzogen wird. Diesen Studien ist jedoch gemein, dass sie mit den angesetzten Evaluationen unterschiedliche Zielsetzungen verfolgen und ihnen weiterhin kein einheitliches und systematisch basiertes Evaluationskonzept oder Untersuchungsdesign zugrunde liegt, anhand dessen kontrol-

lierte Tests durchgeführt werden können, die anschliessend einen Vergleich der Ergebnisse ermöglichen.

Wie [Arnold, 2004] feststellt, liegen derzeit darüber hinaus nur sehr wenige breit abgestützte Untersuchungen vor, die die Wirksamkeit der Visualisierungen von IRS mit VK evaluieren und mit den erzielbaren Ergebnissen konventioneller Suchmaschinen und Information Retrieval Systemen mit Textausgabe vergleichen.

Durch das Fehlen einer gemeinsamen Grundlage und nachhaltiger Untersuchungen können nach [Cugini et al., 2000] folglich kaum repräsentative Schlüsse und empirisch abgestützte Aussagen zur generellen Wirksamkeit von Visualisierungen in der Ergebnisrepräsentation von IRS getroffen werden:

„One of the lessons of our experience is that no matter how much intuitive appeal a given interface might have, without some systematic testing, its real value remains unknown. Especially in the field of visualization, it is all too common for technical wizardry to be unaccompanied by any gain in efficiency.”

2 State-of-the-Art der Evaluation von IRS mit integrierter Visualisierungskomponente

[Chen und Yu, 2000] führten erstmals eine umfassende Studie durch, die eine Meta-Analyse aktueller, empirischer Evaluationen von visuellen Informationssystemen zum Ziel hatte und identifizierten dabei u. a. die bereits oben skizzierten Problemstellungen:

„Empirical studies on information visualization are still very diverse and it is difficult to apply meta-analysis methods (...). A larger homogenous sample of studies would be needed to expect conclusive results”

Die Autoren kommen zum Schluss, dass die Schaffung einer einheitlichen Grundlage für die Evaluation von IRS mit VK in Form eines Referenzframeworks dringend erforderlich ist:

“This is the first attempt in raising the awareness that it is crucial to conduct empirical studies concerning information visualization systematically within a comparable reference framework”

[Plaisant, 2004] untersucht vier Jahre später auf dieser Metastudie aufbauend die gängige Evaluationspraxis für IRS mit Visualisierungskomponente und kommt zum Schluss, dass sich auch dann noch kein einheitliches Evaluationsdesign durchgesetzt hat.

Diese Feststellung wiederholen [Shneiderman und Plaisant, 2006] im Mai 2006. Die Autoren fordern einen Trend weg von Labor-geprägten Usabilitystudien hin zu ethnographischen Studien in der üblichen Arbeitsumge-

bung der Probanden, bei denen Unterbrechungen, Arbeitsplatz, Hilfeleistungen und der soziale Austausch wie gewohnt vorliegen. Weiterhin betonen sie die Relevanz der Durchführung von Studien, bei denen Probanden reale Aufgaben ihrer täglichen Arbeit durchführen und nicht vorgegebene Testaufgaben.

Auch in dieser Publikation erfolgt die Aufforderung zur Schaffung eines Evaluationsframeworks, anhand dessen IRS mit Visualisierungskomponenten untersucht und die Ergebnisse verglichen werden können.

Den Ansatz von Langzeitstudien sehen die Autoren hierbei als mögliche Grundlage für abgesicherte und generalisierbare Evaluationsergebnisse:

„Observations of dozens of users over months and years would do much to improve the reliability, validity, and generalizability of the results. If dozens of software engineers using a source code static analysis visualization tools were found to eagerly adopt this new tool and increased their usage over several months this would be compelling evidence“.

3 Interdependenzen zwischen Visualisierungs- und Retrievalkomponenten

Die Auswertung diverser Evaluationsstudien, u. a. von [Vaughan, 2004; Greymy *et al.*, 2006; Beg, 2005; Buckley und Voorhees, 2000; Jensen *et al.*, 2005; Kwan und Venkatsubramanian, 2006; Hawking *et al.*, 2001; Mann, 2002; Reiterer *et al.*, 2005; Zwol und Oostendorp, 2004] ergibt, dass zur Evaluation von IRS mit VK im wesentlichen Methoden aus den Bereichen der Retrievalperformanzmessung und der Gebrauchstauglichkeitsmessung angesetzt werden. So können neben den von [Plaisant, 2004] aufgeführten häufig vertretenen Ansätzen

- „Kontrollierter Experimente zum Vergleich von Designelementen
- Usabilityevaluation einer Anwendung
- Kontrollierter Experimente zum Vergleich zweier oder mehrerer Anwendungen, sowie
- Case Studies in realistischen Szenarien.“

vor allem Ansätze identifiziert werden, in denen anhand klassischer Retrievaleffektivitätsmaße wie Recall und Precision die Retrievaleffektivität von IRS mit VK gemessen wird.

In den angesetzten Evaluationen wird jedoch meistens nicht auf eine bewusste Kombination geeigneter Methoden geachtet, vielmehr werden angewendete Methoden voneinander losgelöst betrachtet.

Es wird folglich nicht berücksichtigt, dass zwischen Visualisierung, Usabilitymessung und Retrievaleffektivitätsmessung durchaus Interdependenzen bestehen, auf die es zu achten gilt.

So haben aus Sicht der Autorin beispielsweise Interaktionen von Nutzer auf der visuellen Oberfläche einen Einfluss auf die Retrievalfunktionalität eines Systems, was wiederum Auswirkungen auf die Usability des IRS hat.

Weiterhin lassen sich in Abhängigkeit von Retrievalfunktionalität und Gebrauchstauglichkeit Ansprüche an die Gestaltung der Nutzeroberfläche identifizieren. Durch die Kombination von Retrieval- und Visualisierungskomponente ergeben sich folglich Interdependenzen, die sich auf die Qualität des Systems auswirken, wie in Abbildung 1 dargestellt.

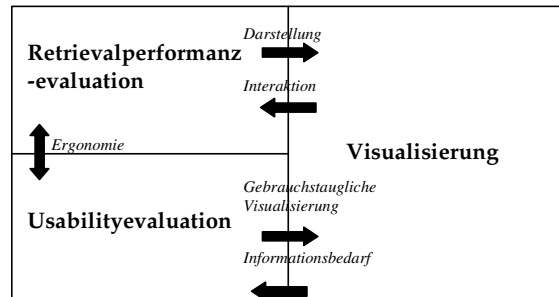


Abbildung 1: Interdependenzen Retrievalperformanz-Visualisierung-Usabilityevaluation

Die gegenseitigen Einflüsse der drei Aspekte lassen sich wie folgt erläutern.

3.1 Interdependenzen zwischen Retrievaleffektivitätsevaluation und Visualisierung

Die visuelle Darstellung von Retrievalergebnissen ermöglicht in der Regel eine Interaktion zwischen Nutzer und IRS. Durch Interaktions- und Verzerrungstechniken kann der Anwender des Systems die Darstellung der Treffer verändern, eine Auswahl oder Einschränkung vornehmen und somit nicht nur die Trefferdarstellung, sondern auch die durchgeführte Suchanfrage verändern [Keim, 2004].

Die Visualisierung hat folglich unmittelbaren Einfluss auf die im System implementierten Retrievalalgorithmen: In Abhängigkeit der Interaktionsmöglichkeiten des Nutzers, muss das IRS die entsprechenden Veränderungen der Suchanfrage und der Ergebnismenge verarbeiten um anschließend die erneut ermittelten und verfeinerten Ergebnisse visuell zu präsentieren.

Im Gegenzug hat auch das Retrieval auf die Visualisierung einen unmittelbaren Einfluss, da beispielsweise Datamingalgorithmen zu Ergebnissen führen, bei denen Cluster, Hierarchien, Beziehungen etc. vorhanden sind, die es in geeigneter Form zu Visualisieren gilt.

Die Retrievalperformanzevaluation kann somit als interdependent zur implementierten Visualisierung gewertet werden und umgekehrt.

3.2 Interdependenzen zwischen Visualisierung und Usabilityevaluation

Eine der zentralen Aufgaben einer Visualisierung zur Retrievalergebnisrepräsentation ist mitunter die Hilfestellung der Definition des Informationsbedürfnisses des Nutzers. Wie die viel zitierten Arbeiten [Belkins *et al.*, 1982] belegen, besteht bei der Formulierung von Suchanfragen durch Anwender eines IRS das Problem des Anomalous State of Knowledge: Der Nutzer kennt lediglich sein subjektiv empfundenes Informationsbedürfnis, oder Teile davon, nicht jedoch den tatsächlich vorhandenen objektiven Informationsbedarf.

Durch diese fehlende Kenntnis des Informationsbedarfs ergibt sich ein Defizit in der Suchanfrageformulierung, die unter Umständen zu irrelevanten Treffermengen führt.

Ziel eines IRS mit Visualisierungskomponente sollte es sein, diesem Defizit zu begegnen und dem Nutzer durch die Darstellung beispielsweise semantisch belegter Relationen von Treffern oder geclusterten Ergebnissen eine Hilfestellung zu bieten in der Erkenntnis, welche Informa-

tion seinen Bedarf hinsichtlich der zu erfüllenden Aufgabe er tatsächlich decken könnte.

Der Erfolg einer Visualisierung, diesen Mehrwert für den Nutzer zu bieten, ist in einer Evaluation des IRS zu bewerten. Mithilfe Methoden der Usabilityevaluation sollte also nicht nur grundsätzlich festgestellt werden, ob ein IRS mit VK für den Nutzer allgemein gebrauchstauglich gestaltet ist, sondern auch ob die gewählten Visualisierungen dem Zweck der Definition des Informationsbedürfnisses dienlich sind.

Im Gegenzug hat auch das Ziel der Erreichung einer höchstmöglichen Gebrauchstauglichkeit Einfluss auf die VK, da die gewählten Visualisierungen nicht nur geeignet sein sollten im Hinblick auf die zugrunde liegende Datenbasis und die zu visualisierenden Elemente. Vielmehr sollten Visualisierungskomponenten auch den Anforderungen entsprechen, gebrauchstauglich zu sein in Bezug auf ihre Intuitivität und Verständlichkeit im eingesetzten Kontext (also beispielsweise die Verwendung intuitiv interpretierbarer Metaphern als Symbole bei einer Relationenvisualisierung).

3.3 Interdependenzen zwischen Usabilityevaluation und Retrievaleffektivitätsevaluation

Hinsichtlich der Retrievalperformanz- und Usabilityevaluation sind ebenfalls Interdependenzen zu beachten.

Einerseits kann festgestellt werden, dass viele Instrumente, die zur Messung der Gebrauchstauglichkeit verwendet werden, bei entsprechendem Einsatz auch wertvolle Ergebnisse hinsichtlich der Retrievalperformanzevaluation erbringen können.

Z. B. kann das automatisierte Mitloggen von Interaktionen eines Probanden, das für die Aufmerksamkeitsanalyse in der Usabilitymessung eingesetzt wird, ebenfalls für die Datengewinnung genutzt werden, auf deren Basis sich Recall- oder Precisionwerte errechnen lassen.

Durch eine breite methodische Abstützung kann somit ein Phänomen anhand unterschiedlicher Vorgehensweisen gemessen und die Ergebnisse besser interpretiert werden.

Bei der Planung und Durchführung einer umfassenden Evaluation sollte folglich auf eine gewissenhafte Zusammenstellung der eingesetzten Instrumente geachtet werden.

Zum anderen kann festgestellt werden, dass Anforderungen an ein System bzw. dessen Oberfläche, die sich aufgrund von Aspekten der Gebrauchstauglichkeit ergeben, durchaus einen Einfluss haben können auf die Gestaltung der Retrievalperformanz und umgekehrt.

Soll z. B. bei einem IRS die für den Nutzer erforderliche Art des Informationsbedürfnisses berücksichtigt werden, um einen darauf optimierten Retrievalalgorithmus anzuwenden, hat dies zur Folge, dass das System aus Nutzersicht komplexer und dadurch schwerer zu verwenden wird. Während für versierte Anwender die Auswahl von Vorteil ist, könnte sie sich für ungeübte Nutzer aus Sicht der Gebrauchstauglichkeit überflüssig oder gar nachteilig gestalten. Es müssen folglich Wege gefunden werden, diese aus Retrievalperformanzsicht sinnvolle Ergänzung auf ergonomische Weise in die Systemoberfläche zu integrieren um Einbußen in der Gebrauchstauglichkeit zu vermeiden.

Allgemein müssen Aspekte der Retrievaleffektivität, wie beispielsweise Metaphern, die für die Relationenvisualisierung eingesetzt werden, in der Usabilityevaluation berücksichtigt werden und deren Eignung für die

Durchführung der Recherche überprüft werden.

Die wechselseitigen Interdependenzen zwischen Usabilityevaluation und Retrievaleffektivitätsmessung dürfen folglich nicht unberücksichtigt bleiben und es muss ein gezielter Einsatz von Methoden vorgenommen werden, der einen ganzheitlichen und integrierten Ansatz verfolgt.

3.4 Auswirkungen auf das Evaluationsdesign

Zusammenfassend stellt sich bei einer Evaluation eines IRS mit VK nicht nur die Frage, ob eine Visualisierungsart für den Einsatz in einem bestimmten IRS geeignet ist, sondern auch, ob sie den implementierten Retrievalalgorithmen entspricht, diese in geeigneter Form abbilden kann und weiterhin ob sie gebrauchstauglich ist im Sinne der Unterstützung des Nutzers bei der Identifikation und Deckung seines Informationsbedarfs.

Diese Aspekte müssen bei einer Evaluation ebenso berücksichtigt werden, wie die klassischen Fragestellungen der Retrievalperformanz oder der Eignung eingesetzter Metaphern für die Ergebnisrepräsentation, was einen breit angelegten Ansatz zur Folge hat.

Die Herausforderung besteht somit darin, anhand eines integrierten Designs ein Evaluationsframework zu entwickeln, das nicht nur Methoden kombiniert, sondern auch den oben genannten neu auftretenden Fragestellungen begegnet.

Weiterhin sollte eine Grundlage geschaffen werden, auf der ein Vergleich durchgeführter Evaluationen möglich ist um Synergieeffekte zu erzielen und Unterschiede sowie Ähnlichkeiten in den Resultaten identifizieren, analysieren und daraus generalisierbare Schlüsse ziehen zu können.

4 Morphologischer Kasten für die Auswahl geeigneter Methoden zur Evaluation von IRS mit Visualisierungskomponente

4.1 Spannungsfeld bei der Kombination geeigneter Methoden

Bei der Kombination von Methoden zur Evaluation von IRS mit VK befinden sich die Instrumente im Spannungsfeld zwischen Laborstudien einerseits, die durch eine künstlich geschaffene Umgebung sehr präzise aber unter Umständen auch leicht verfälschte Ergebnisse erzielen lassen und Feldstudien andererseits [Plaisant, 2004]. Bei letzteren werden Probanden für eine Evaluation möglichst authentische und reale Arbeitsbedingungen geboten, wobei die Störfaktoren nicht behoben werden und Ursachen für die Artung der Ergebnisse aus diesem Grund nicht immer identifizierbar sind.

Als weitere Dimension sind die Objektivität der Ergebnisse einerseits und ein hoher Grad an Nutzerbeteiligung andererseits zu berücksichtigen, wie aus dem Portfolio in Abbildung 2 ersichtlich wird.

Die Zusammenstellung eines integrierten Methoden-Mix für die Evaluation eines IRS mit VK sollte dieses Spannungsfeld berücksichtigen und das Portfolio zur Begegnung oben identifizierter Herausforderungen stets umfassend abdecken.

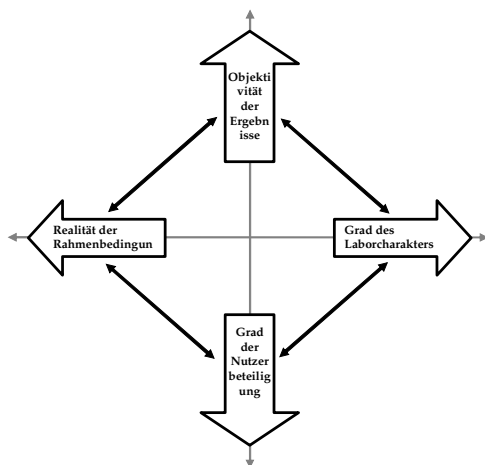


Abbildung 2: Spannungsfeld der Evaluation von IRS mit Visualisierungskomponenten

Auf Grundlage der untersuchten Studien zur Evaluation von IRS mit VK wurden rund 40 Methoden in verschiedenen Variationen aus den Bereichen der Retrievaleffektivitäts- und der Gebrauchstauglichkeitsmessung zusammengetragen und charakterisiert sowie hinsichtlich des oben erläuterten Spannungsfelds klassifiziert. Eine Zusammenstellung der Methoden findet sich unter folgender URL:

<http://www.informationswissenschaft.ch/index.php?id=299>

4.2 Morphologischer Kasten für die Klassifikation von Evaluationsmethoden

Für die Klassifikation von geeigneten Methoden und Instrumenten, die den Zielsetzungen einer Evaluation entsprechen, wurde in Anlehnung an [Mussnug und Stowasser, 2004] ein morphologischer Kasten entwickelt, anhand dessen sich unterschiedliche Ansätze zwischen Usability- und Retrievaleffektivitätsevaluation einordnen lassen und gleichzeitig dem oben aufgezeigten Spannungsfeld Rechnung trägt (Vergleiche Tabelle 1).

| DIMENSION | SKALA bzw. Ausprägung | | | |
|---|--------------------------------------|---------------------|------------------------|--------|
| Zielsetzung | | | | |
| Zielsetzung und Art der Evaluation | Vergleichend | Funktionalität | Leistung (Algorithmen) | Design |
| Rahmenbedingungen | | | | |
| Kosten | Gering | Mittel | Hoch | |
| Trainingsaufwand | Gering | Mittel | Hoch | |
| Menge erforderlicher Probanden | Keine | Wenige | Viele | |
| Erforderliche Apparaturen bzw. Software | Auflistung (m) = Muss, (k) = Kann | | | |
| Durchführung | | | | |
| Zeitlicher Aufwand | Gering | Mittel | Hoch | |
| Untersuchungsort | Feld | Kontrolliertes Feld | Labor | |
| Auswertungsdimensionen | | | | |

| | | | | |
|-------------------|------------|---------------|-------------|---------|
| Ergebnisdimension | Qualitativ | | Quantitativ | |
| Bezugsdimension | Subjektiv | | Objektiv | |
| Untersuchungsart | Analytisch | Experimentell | Beobachtend | Fragend |

Tabelle 1: Morphologischer Kasten für die Klassifikation von Evaluationsmethoden für IRS mit Visualisierungskomponente

Auf Grundlage dieses morphologischen Kastens wurde eine Morphologie erstellt mit den identifizierten Methoden und Evaluationsinstrumenten, die ebenfalls unter folgender URL aufgerufen werden kann: <http://www.informationswissenschaft.ch/index.php?id=299>

4.3 Ergebnisse einer exemplarischen vergleichenden Evaluation

Auf den bisherigen Erkenntnissen aufbauend wurde im Rahmen einer kleinen Beispielevaluation, die keinen Anspruch erhebt, repräsentativ zu sein, der vorgeschlagene Ansatz in der Empirie exemplarisch überprüft.

Hierzu wurde eine vergleichende Evaluation der Suchmaschinen Yahoo (<http://search.yahoo.com>) mit konventioneller Listenausgabe und Grotker (www.grotker.com) mit einer visuellen Ergebnisrepräsentation mit fünf Probanden vorgenommen, bei der unter anderem folgende Methoden zum Tragen kamen:

- Kontrolliertes Experiment (Usability-Test) mit vorgegebenen Aufgabenstellungen
- „Lautes Denken“
- Screen-Capturing mit Auswertung
- Fragebogen
- Tagebuchstudie
- Retrievaleffektivitäts-Evaluation zur Erhebung u. a. folgender Masse: Relative Recall@n, Precision@n, Jewel Measure, First Retrieved Document Rank nach [Kwan und Venkatsubramanian, 2006]

Bei der Auswertung der Evaluation kann festgestellt werden, dass sich die in einem Fragebogen von Probanden geäußerten Bewertungen nicht immer decken mit den anderweitig erhobenen Kennzahlen.

Beispielsweise gaben alle Probanden an, sie hätten das Gefühl, die Anzahl der erforderlichen Interaktionen seien bei Grotker höher gewesen, als bei Yahoo. Die Auswertung des Screen-Capturing ergibt jedoch, dass während des Usability-Tests zur Erfüllung der Aufgaben im Schnitt rund 20% mehr Interaktionen auf der Oberfläche von Yahoo vorgenommen wurden, als auf der visuellen Oberfläche von Grotker.

Ähnliche scheinbare Widersprüche ergaben sich hinsichtlich der Qualität der erzielten Treffer, die mit Methoden der Retrievaleffektivitätsmessung und der Auswertung des Screen-Capturing erhoben und gleichzeitig durch Befragung der Probanden eingeschätzt wurde.

In diesen Beispielen kann anhand der breiten methodischen Abstützung festgestellt werden, dass sich durch das IRS mit VK zwar bessere Ergebnisse erzielen liessen, sich diese Feststellung jedoch nicht in der Einschätzung aus Probandensicht widerspiegelte.

Die exemplarische Evaluation macht somit deutlich, dass bei gezieltem Einsatz und der integrierten Kombination geeigneter Methodenansätze durchaus Synergieeffekte

fekte zum Tragen kommen. Weiterhin wird daraus ersichtlich, dass sich bereits mit wenigen Mitteln eine breit abgestützte Evaluation durchführen lässt, die durch den Einbezug qualitativer und quantitativer Masse eine gute Ausgangsbasis bieten für die Interpretation von Ergebnissen sowie die Identifikation von Auslösern scheinbarer Differenzen im Erhebungsmaterial.

Die Schwächen einiger Evaluationsmethoden werden somit durch die Stärken anderer Methoden ausgeglichen und die erhobenen Ergebnisse weisen insgesamt eine höhere Qualität auf und lassen sich besser auswerten und interpretieren.

5 Ausblick

Die Auswahl und Kombination geeigneter Methoden für eine nachhaltige und umfassende Evaluation von IRS mit Visualisierungskomponente bedeutet nach wie vor eine grosse Herausforderung.

Anhand einer Morphologie lässt sich ein integrierter Ansatz verfolgen, der eine möglichst breite Abstützung der Ergebnisse hinsichtlich aller relevanten Aspekte eines IRS mit visueller Ausgabe gewährleistet. Künftig gilt es auf Grundlage des Evaluationsframeworks eine konkrete Evaluationsumgebung für die Durchführung von Evaluationen von IRS mit VK zu gestalten, die als Ausgangsbasis für den evaluationenübergreifenden Vergleich dient. Auf diesen Ergebnissen beruhend lassen sich langfristig repräsentative und allgemeingültige Aussagen zur Eignung von Visualisierungen im Information Retrieval ableiten.

Das vorgelegte Framework gilt es hierzu in weiteren Schritten anhand der Ergebnisse empirischer Erprobung sukzessive zu verfeinern und zu optimieren. Vorliegender Vorschlag ist somit als erster Schritt in Richtung einer einheitlichen Evaluationsgrundlage zu verstehen, die gemäss der Aussage diverser Autoren langfristig gesehen unabdingbar ist.

Referenzen

- [Arnold, 2004] C. Arnold. Visualisierung im Information Retrieval. Magisterarbeit in der Philosophischen Fakultät IV (Informationswissenschaft) der Universität Regensburg: Regensburg, 2004.
- [Bar-Ilan *et al.*, 2004] J. Bar-Ilan, M. Levene, M. Mat-Hassan. Dynamics of search engine rankings - a case study. In *Proceedings of the 3rd international workshop on web dynamics*, New York, 2004.
- [Beg, 2005] M. M. S. Beg. A subjective measure of web search quality. In *Information Sciences* Volume 169, Issues 3-4, 1 February 2005, S. 365-381.
- [Belkin *et al.*, 1982] N. Belkin, R. Oddy, H. Brooks. ASK for information retrieval: Part I. background and theory. In *Journal of Documentation*, 38(2) Seiten 61-71, 1982.
- [Buckley und Voorhees, 2000] C. Buckley, E. M. Voorhees. Evaluating Evaluation Measure Stability. In *SIGIR 2000*. Belkin, N. J., Ingwersen, P., und Leong, M.-K. (eds.); ACM, S. 33-40, Athen, 2000.
- [Chen und Yu, 2000] C. Chen, Y. Yu. Empirical studies of information visualization: a metaanalysis. In *International Journal of Human-Computer Studies*, 53 (2000) 5, Seiten 851-866, 2000.
- [Cugini, 2000] J. Cugini. Presenting Search Results: Design, Visualization and Evaluation. In: *Workshop: Information Doors - Where Information Search and Hypertext Link*, San Antonio
- [Gremy *et al.*, 1999] F. Gremy, J. M. Fessler, M. Bonnin. Information systems evaluation and subjectivity. In *International Journal of Medical Informatics* Volume 56, Issues 1-3, December 1999, S. 13-23.
- [Grokker] Grokker: Suchmaschine Grokker, URL: <http://www.grokker.com>, Stand: 25.06.2006.
- [Hawking *et al.*, 2001] D. Hawking, N. Craswell, P. Bailey *et al.* Measuring search engine quality. In *Information Retrieval*, 4 Nr.1, S. 33-59, 2001.
- [Jensen *et al.*, 2005] E. C. Jensen, S. M. Beitzel, O. Frieder, O. *et al.* A framework for determining necessary query set sizes to evaluate web search effectiveness. In *Special interest tracks and posters of the 14th international conference on World Wide Web*, May 10-14, 2005, Chiba, Japan.
- [Kartoo] Kartoo: Suchmaschine Kartoo, URL: <http://www.kartoo.com>, Stand: 25.06.2006.
- [Keim, 2002] D. A. Keim. Information Visualization and Visual Data Mining. In *IEEE Transactions on visualization and computer graphics*, Vol. 7, No. 1, January-March 2002
- [Koshman, 2005] Koshman. Testing user interaction with a prototype visualization-based information retrieval system. In *Journal of the American Society for Information Science and Technology*, 56(8) 2005, Seiten 824-833, 2005.
- [Kwan und Venkatsubramanian, 2006] S. K. Kwan, S. Venkatsubramanian. An Economic Model for Comparing Search Services. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*. S. 107-116, 2006.
- [Liveplasma] Liveplasma: Suchmaschine Liveplasma URL: <http://www.liveplasma.com/>, Stand: 25.06.2006.
- [Mann, 2002] T. M. Mann. Visualization of Search Results from the World Wide Web, Dissertation, Universität Konstanz, 2002.
- [Mussgnug und Stowasser, 2004] J. Mussgnug, S. Stowasser. Ein Schema zur Auswahl geeigneter Usability-Methoden - Dargestellt am Beispiel der Blickbewegungsanalyse. In *Proceedings of the 2nd annual GC-UPA Track Paderborn*, September 2004, Paderborn.

- [Plaisant, 2004] C. Plaisant. The Challenge of Information Visualization Evaluation. In *Proceedings of Conference on Advanced Visual Interfaces AVI'04*.
- [Reiterer *et al.*, 2005] H. Reiterer, G. Tullius, T. M. Mann. INSYDER: a content-based visual-information-seeking system for the Web. In *International Journal on Digital Libraries*, Volume 5, Issue 1, Mar 2005, Seiten 25 - 41.
- [Reiterer, 2004] H. Reiterer. Visuelle Recherchesysteme zur Unterstützung der Wissensverarbeitung. In Hammwöhner, R.; Rittberger, M.; Semar, W. (Hrg.): *Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen*. Konstanz, 2004. Seiten 1–21.
- [Shneiderman und Plaisant, 2006] B. Shneiderman, C. Plaisant. Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies. In *Proceedings of the BELIV'06 Workshop*, Venice Seiten 61–77, 2006.
- [Sebrechts *et al.*, 1999] M. Sebrechts, J. Vasilakis, M. Miller, et al. (1999): Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, S. 3-10, 1999.
- [Vaughan, 2004] Vaughan, L. (2004): New measurements for search engine evaluation proposed and tested. In *Information Processing and management* 40 (2004) S. 677-691, 2004.
- [Veerasingam und Belkin, 1996] A. Veerasingam, N. J. Belkin. Evaluation of a tool for visualization of information retrieval results. In *Proceedings of the 19th Annual international ACM SIGIR Conference on Research and Development in information Retrieval* (Zurich, Switzerland, August 18 - 22, 1996). SIGIR '96. ACM Press, New York, NY, Seiten 85-92, 1996.
- [Webbrain] Webbrain: Suchmaschine Webbrain, URL: <http://www.webbrain.com>, Stand: 25.06.2006.
- [Zwol und Oostendorp, 2004] R. Van Zwol, H. Van Oostendorp. Google's "I'm feeling lucky", Truly a Gamble?. In *Zhou, X. et al. (Hrg.) (2004): Web Information Systems - WISE 2004, Proceedings of the 5th International Conference on Web Information Systems Engineering*. Brisbane, Australia, Seiten 378-390, 2004.