

Nutzerzentrierte Entwicklung einer Suche für den Life-Science-Bereich

Sonja Hamann, Thomas Weinhold, Bernard Bekavac, Daniel Streiff

Schweizerisches Institut für Informationswissenschaft (SII), HTW Chur

Zusammenfassung

Suchmaschinen wie Google oder Bing setzen heutzutage den Standard bzgl. der Bedienung von Recherchewerkzeugen im Internet. Eine Herausforderung bei der Entwicklung von Suchinstrumenten für spezifische Domänen respektive für Fachdatenbanken ist darin zu sehen, dass die einfache und intuitive Bedienbarkeit von Websuchmaschinen für gewöhnlich nicht unmittelbar auf professionelle Recherchewerkzeuge übertragen werden kann. Dies begründet sich durch den Umstand, dass Anwender bei professionellen Recherchen oftmals deutlich komplexere Informationsbedürfnisse aufweisen und zudem in der Regel auch die Vollständigkeit des Ergebnisraumes von größerer Bedeutung ist. Folgender Beitrag illustriert die iterative, nutzerzentrierte Entwicklung eines komplexen, modular aufgebauten Suchsystems zur professionellen Recherche in Businessdaten aus dem Life-Science-Bereich. Im Vordergrund stehen dabei die Ergebnisse aus mehreren nutzerbasierten Evaluationen, auf deren Basis letztendlich die getroffenen Designentscheidungen für die entwickelte Suchkomponente begründet und erläutert werden.

1 Ausgangslage und Motivation

Analog zu der Entwicklung neuer Technologien haben sich in den letzten Jahren die Anforderungen von Informationssuchenden stark verändert. Anwender erwarten einfach bedienbare Recherchewerkzeuge, die in der Lage sind, ihre spezifischen Informationsbedürfnisse adäquat zu befriedigen. Dementsprechend haben sich sowohl ein einfacher Suchschlitz, wie man ihn bspw. von Google kennt, als auch leicht bedienbare Filtermechanismen wie z.B. eine „Faceted Search“ (Tunkelang 2009; Hearst 2008) am Markt durchgesetzt. Zudem hat der vermehrte Einsatz von dynamischen Komponenten (z.B. mittels AJAX) dazu geführt, dass sich Möglichkeiten der direkten Manipulation (Hutchins et al. 1985; Shneiderman 1982) mittlerweile auch im Web als wichtiges Qualitätsmerkmal etabliert haben (Hearst 2009). Trotz dieser Fortschritte sind die Entwicklungen im Bereich von Suchmaschinen jedoch noch lange nicht abgeschlossen, wie auch ein Blick auf die einschlägige Forschung verdeutlicht: einen Einblick in den äusserst umfangreichen State-of-the-Art von Suchmaschinen findet

man u.a. bei Hearst (2009), Spree et al. (2011) und Weinhold et al.(2009). Dort angeführte Konzepte sind jedoch nicht immer angemessen und unmittelbar auf domänenspezifische IR-Systeme übertragbar, da z.B. bei professionellen Recherchen der Vollständigkeit des Ergebnisraums sowie der Möglichkeit, komplexe Suchanfragen zu formulieren, deutlich mehr Bedeutung zukommt. Gerade bei letzterem haben Anwender große Probleme (Sutcliffe et al. 2000; English et al. 2002), nicht zuletzt da etablierte UI-Elemente wie ein einfacher Suchschlitz ihnen hierbei keinerlei Unterstützung bieten. In Bezug auf die Vollständigkeit des Ergebnisraums liegt die Herausforderung darin, Synonyme oder artverwandte Begriffe (z.B. „cancer“ und „oncology“) bei der Suche sinngemäß zu berücksichtigen.

Gegenstand des Projektes SERENA – SEmantic RETrieval in busiNess contActs¹ war es daher, ein zugleich leicht bedienbares wie auch effizientes Werkzeug zur professionellen Recherche in Businessdaten aus dem Life-Science-Bereich zu entwickeln, welches die oben spezifizierten Problemstellungen aufgreift. Hierbei stand ein nutzerzentrierter, iterativer Entwicklungsprozess im Vordergrund, bei dem erarbeitete Konzepte jeweils mit Endanwendern evaluiert und entsprechend weiter optimiert wurden.

2 Anforderungsdefinition

Ausgangsbasis des Projektes waren die beiden Datenbanken Biotechgate (BTG)² und PartneringOne (P1)³ sowie die Suchtechnologie der Firma Infocodex⁴. Während BTG Informationen zu knapp 24'000 Unternehmen aus dem Life-Science-Bereich sowie rund 25'000 Medtech- und Biotechprodukten bereit stellt, handelt es sich bei P1 um eine Partnering-Software, welche auf Life-Science-Konferenzen eingesetzt wird, um es Konferenzbesuchern zu ermöglichen, schnell und einfach Meetings mit anderen Teilnehmern zu vereinbaren („Business-Speed-Dating“).

Im Rahmen von SERENA sollten die Recherchemöglichkeiten von BTG und P1 mittels einer semantischen Suche erweitert werden. Hierfür wurden zu Projektbeginn im Rahmen einer umfassenden Anforderungsanalyse die vorhandenen Produkte auf Schwachstellen untersucht, die zu Grunde liegenden Datenbanken und entsprechende Logfiles analysiert und zusätzlich Interviews mit den Projektpartnern und Endanwendern geführt. Als Hauptproblem hat sich dabei der Umstand herauskristallisiert, dass sich Nutzer bislang bereits im Vorfeld einer Recherche entscheiden mussten, ob sie nach einem Unternehmen, einem Produkt oder einer Technologie suchen wollen. Eine gezielte Suche nach einem Unternehmen mit spezifischen Produkten war jedoch nicht möglich, da die Suchmasken eine solche Kombination

¹ Gefördert von der Kommission für Technologie und Innovation unter der KTI-Nr.10767.1 PFES-ES; Laufzeit 01/2010 bis 02/2012

² Venture Valuation AG, <http://www.biotechgate.com>, 07.03.2012

³ EBD Group, <http://www.partneringone.com>, 07.03.2012

⁴ <http://www.infocodex.com>, 07.03.2012

unterschiedlicher Attribute nicht zuließen. Auch die vorhandenen Filtermechanismen wurden aufgrund der fehlenden Option zur Mehrfachauswahl teilweise als unzureichend deklariert. Ein weiteres Problem war der Umstand, dass Begriffe wie „neoplasm“, „cancer“ und „oncology“ in den Unternehmensprofilen nahezu synonym verwendet werden, was im Rahmen der Suche jedoch bislang nicht berücksichtigt wurde. Hinzu kam, dass die komplexen Formulare zum Filtern der Treffer einen Großteil der Benutzeroberfläche benötigten, weshalb einerseits in ihnen gescrollt werden musste und andererseits die Trefferliste auf eine zweite Seite ausgelagert werden musste. Eine direkte Manipulation der Suchanfrage bzw. der verwendeten Filter bei der Sichtung der Treffer war demzufolge nicht möglich. Als weitere Schwachstelle wurden bei beiden Produkten fehlende Funktionalitäten zum Vergleich von Unternehmen identifiziert.

Mit Hinblick auf die Optimierung der Suchfunktionen beider Produkte wurden dementsprechend folgende Anforderungen definiert:

- Implementierung einer zentralen Suche über das gesamte Firmenprofil (inkl. Produkten und Technologien)
- Automatische Berücksichtigung artverwandter Suchbegriffe (unscharfe Suche)
- Implementierung eines Filterkonzeptes, welches eine Kombination an Einschränkungsmöglichkeiten unterstützt (Mehrfachauswahl, Kombination von unterschiedlichen Filterkategorien)
- Entwicklung einer Funktionalität zum Vergleich von ähnlichen Firmen
- Visualisierung des Domänenwissens zur Unterstützung bei der Suchanfragenformulierung

Diese Ziele und Interessen der Anwender wurden anschließend mit Hilfe von Personas (Cooper 1999; Goodwin 2009) und Szenarien (Carroll 1994; Goodwin 2009) festgehalten. Diese wurden im gesamten Entwicklungsprozess immer wieder zur Überprüfung der angestrebten Ziele sowie zur Qualitätssicherung herangezogen.

3 Konzeption

Da die zu entwickelnden Komponenten sowohl in BTG als auch in P1 genutzt werden können sollten, war eine modulare Konzeption erforderlich, um sicherzustellen, dass einzelne Funktionalitäten je nach Bedarf in beide Produkte angemessen integriert und ggf. modifiziert werden können. Es galt hierbei, trotz der Unabhängigkeit der Bausteine eine unmittelbare Interaktion zwischen den jeweiligen Komponenten zu ermöglichen, welche nachfolgend genauer beschrieben werden.

3.1 Suche und Trefferliste

Analog dem Standard aktueller Suchmaschinen wurde bei SERENA kein komplexes Suchformular implementiert, sondern lediglich ein einfacher Suchschlitz, der auf die semantische Suche von Infocodex zugreift. Die Basis der Software von Infocodex stellt eine linguistische Datenbank mit einem sprachübergreifenden Wortschatz von mehr als 3 Mio. Begriffen dar, wodurch die Verarbeitung von Synonymen oder zusammengesetzten Begriffen („Named Entity Recognition“) erlaubt wird. Mit Hinblick auf die Anpassung bzw. Optimierung dieser linguistischen Datenbank an die Spezifika der Life-Science-Branche, wurden einerseits relevante Terminologien, wie z.B. der ICD10-Code⁵, in diese integriert und andererseits wurde eine Ontologie erstellt, welche die in diesem Kontext relevanten Begriffe und deren Zusammenhänge aus Business-Sicht widerspiegelt.

Im Rahmen einer Suche wird dabei nun jeweils das gesamte Unternehmensprofil inklusive der jeweiligen Produkte und Technologien berücksichtigt. Daraus resultiert der Umstand, dass die Trefferliste dementsprechend auch diese unterschiedlichen Perspektiven abbilden können muss, weshalb sie grundlegend umgestaltet wurde. Da die grundlegende Seitenstruktur von BTG und P1 beibehalten werden sollten, fiel der Entscheid auf eine reiterbasierte Darstellung der Trefferlisten (vgl. Abbildung 3), welche einen unmittelbaren Wechsel zwischen den jeweiligen Treffern erlaubt.

3.2 Filterkonzept

Das bei SERENA verwendete Filterkonzept besteht aus drei Filterboxen, deren inhaltliche Ausrichtung jeweils über ein Drop-Down-Menü verändert werden kann. Die Default-Einstellungen der einzelnen Filterboxen wurden dabei auf Basis mehrerer Interviews mit Anwendern erhoben. Die Anzahl von drei Filterboxen wurde deshalb gewählt, weil eine Analyse der Logfiles von BTG gezeigt hat, dass Anwender bei ihren Recherchen kaum mit mehr Einschränkungen arbeiten. Gleichzeitig kann die Treffermenge mit Hilfe von drei Filtermöglichkeiten in der Regel auf eine gut überschaubare Anzahl von Treffern reduziert werden, was letztendlich auch das positive Feedback der Testpersonen im Rahmen der durchgeführten Evaluationen bestätigt hat.

Die Auswahloptionen in den Filterboxen selbst sind in Form von Checkboxes umgesetzt. Bei den einzelnen Filteroptionen wird dabei jeweils die zugehörige Treffermenge mitangezeigt, wie man es z.B. bei einer „Faceted Search“ gewohnt ist (vgl. Abbildungen 1, 2 und 3). Innerhalb einer einzelnen Filterbox sind die einzelnen Filteroptionen mit dem Booleschen Operator OR verknüpft, um die von Anwendern geforderte Mehrfachauswahl unterstützen zu können. Die Filterboxen untereinander sind hingegen mit AND verbunden.

Bereits die ersten Mockups haben ein analytisches Potenzial der Filterkombinationen aufgezeigt, weshalb im Folgenden eine lineare Funktionsweise der Filteroptionen ausgetestet wur-

⁵ <http://www.who.int/classifications/icd/en/>, 07.03.2012

de: Einstellungen im ersten Filter beeinflussen somit die Auswahlmöglichkeiten und Mengenangaben im zweiten und dritten Filter, wohingegen Einstellungen im zweiten Filter nur den dritten Filter, nicht aber den vorangehenden beeinflussen. In Abhängigkeit der Reihenfolge kann somit beispielsweise ermittelt werden, welcher Anteil aller Pharmaunternehmen sich in der Schweiz befindet oder auch wie viele der in der Schweiz befindlichen Firmen speziell der Pharmabranche zugeordnet sind. Das Filtersystem kann also bei Bedarf für einfache statistische Auswertungen unabhängig der Trefferliste genutzt werden.



Abbildung 1: Horizontale Anordnung der Filter – Betonung der Reihenfolge durch das Einbinden von Pfeilen

Trotz Versuchen mit unterschiedlichen Darstellungsformen hat sich diese lineare Funktionsweise der Filter als nur schwer vermittelbar erwiesen. So wurde in ersten Versuchen mit einer horizontalen (gemäß der üblichen Leserichtung) Anordnung der Filterboxen experimentiert (vgl. Abbildung 1). Anwender konnten die Reihenfolge in diesem Fall sowohl über die pfeilförmigen Buttons an den Rändern der Boxen als auch per Drag&Drop verändern. Anhand der Pfeile im Hintergrund sollte der lineare Ablauf verdeutlicht werden. Wie eine Zwischenevaluation mit Endanwendern zeigte, konnten bei dieser Version des Prototyp zwar alle Testpersonen die Filterfunktionen problemlos bedienen, aber nur zwei Probanden erkannten den oben beschriebenen, analytischen Wert der Filterfunktionen, welche durch die Interpretation der Mengenangaben und das Vertauschen der Reihenfolge entstehen würde. Deshalb wurde im Folgenden mit verschiedenen Visualisierungen zur Illustration der linearen Funktionsweise der Filter experimentiert, die im Rahmen mehrerer informeller Nutzer-tests überprüft wurden.

Letztendlich zeigte erst die vertikale Anordnung der Filterboxen und die Einbettung zusätzlicher visueller Stützen eine Verbesserung der Verständlichkeit dieser Lösung (vgl. Abbildung 2). Zusätzlich wurde das Umstellen der Filterreihenfolge durch eine Auswahl in den Drop-Down-Feldern ermöglicht. Dadurch fiel es Anwendern anschließend in den beiden durchgeführten Usability-Tests⁶ leichter, die Reihenfolge der Filterboxen zielorientiert anzupassen. Auch die angebotenen statistischen Informationen konnten nun von den Probanden mehrheitlich korrekt interpretiert werden. Allerdings hat sich im Rahmen nachfolgender Testaufgaben

⁶ Neben einer Vielzahl informeller Zwischenevaluationen wurden zwei ausführliche Usability-Tests durchgeführt: Der erste Test wurde im Rahmen einer Bachelorarbeit im Usability-Labor der HTW Chur mit acht Probanden durchgeführt, wohingegen der zweite Test mit neun Endanwendern im Rahmen einer Biotech-Konferenz stattfand. Die evaluierten Versionen der Prototypen unterschieden sich nur geringfügig hinsichtlich der visuellen Gestaltung, die grundlegende Anordnung und Funktionsweise war jedoch identisch.

gezeigt, dass das lineare Konzept trotz aller Optimierungsversuche nicht vollständig verstanden wurde.

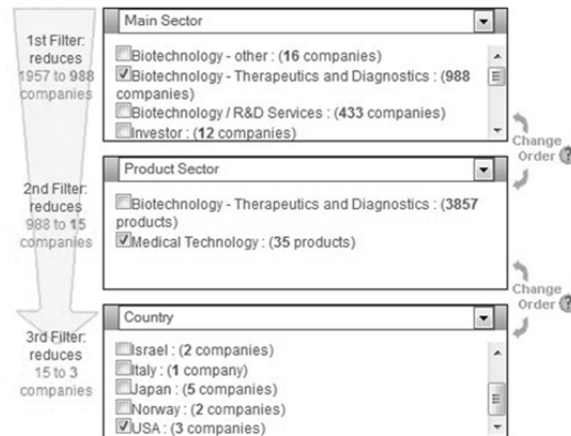


Abbildung 2: Vertikale Anordnung der Filterboxen inklusive visueller Stützen zur Betonung der Reihenfolge

Interpretiert man die Aussagen einzelner Testpersonen, so können hierfür zwei Ursachen angeführt werden: Einerseits war die direkte Manipulation der Filter bzw. der Trefferliste mittels AJAX, welche einen zusätzlichen Button zur Aktivierung des Filters unnötig macht, anfangs etwas ungewöhnlich. Andererseits waren die Testpersonen aus anderen Anwendungen mit keinem vergleichbaren Filterkonzept vertraut, weshalb die implementierte Funktionsweise nicht konform mit ihren Erwartungen war.

In Konsequenz dieser Evaluationsergebnisse wurden die Linearität des Filterkonzepts respektive die Vorgabe der Reihenfolge wieder verworfen. Wählt ein Anwender nun eine oder mehrere Checkboxes innerhalb eines Filters aus, so werden einfach die anderen beiden Filterboxen entsprechend angepasst, indem nicht zutreffende Optionen ausgeblendet und die Mengenangaben der restlichen Optionen angepasst werden (vgl. Abbildung 3). Somit kann der Anwender die Reihenfolge, in der er die Filter nutzen möchte, frei bestimmen. Der Vorteil dieser Lösung ist, dass eine leere Ergebnismenge verhindert wird. Gleichzeitig wird bereits frühzeitig deutlich, welche weiteren Einschränkungen sinnvoll wären. Da nicht nur der Filterbereich selbst, sondern auch die einzelnen Filterboxen bei Bedarf zugeklappt werden können, um mehr Platz für die Trefferliste zu schaffen, wechselt der farbliche Hintergrund der Kopfzeile einer Filterbox bei der Aktivierung eines Filters und gleichzeitig wird auch die entsprechend oben eingebundene Checkbox selektiert. Auf diesem Weg wird dem Anwender verdeutlicht, welche Filterboxen aktiv sind und zugleich eine Option zur Verfügung gestellt, um alle gesetzten Filter in einer Box abzuwählen.

oncology search

Visualization

Filter > reset all filter settings

☒ Company - Main Sector

- ☒ Biotechnology - Therapeutics : (8 companies)
- ☒ Biotechnology / R&D Services : (8 companies)

☐ Product - Main Therapeutic Area

☒ Company - Country

- ☐ Australia : (27 companies)
- ☒ Austria : (16 companies)
- ☐ Belgium : (10 companies)
- ☐ Brazil : (2 companies)
- ☐ Canada : (91 companies)

Companies (16) Products (30) Technologies (6)

Results: 1-16 of 16 1 1 1 1 1 1 1 1 1 1 1 1 1 1

1

"Trimed Biotech GmbH" compare to similar!

Main Sector: "Biotechnology - Therapeutics"

Abbildung 3: Implementierung des Filterkonzeptes ohne vorgegebene Reihenfolge

3.3 Vergleichsfunktion

Im Rahmen von SERENA wurde basierend auf der Ähnlichkeitssuche von Infocodex eine Vergleichsfunktion implementiert, die Anwender im Kontext von Recherchen beim Auffinden potentiell interessanter Unternehmen zusätzliche Unterstützung bieten soll („compare to similar“). Für den Aufruf der zugehörigen Vergleichstabelle (vgl. Abbildung 4) steht bei jedem Objekt der Trefferliste ein entsprechender Button zur Verfügung, wobei diese in Form eines modalen Dialoges im Vordergrund geöffnet wird. Die einzelnen, in der Liste enthaltenen Unternehmen sind hierbei spaltenweise abgetragen, wobei die Firma, welche als Ausgangspunkt der ähnlichkeitsbasierten Suche diene, farblich hinterlegt ist und stets eingeblendet bleibt. Die Eigenschaften der verschiedenen Unternehmen werden zeilenweise abgetragen, wobei es sich um jene Kriterien handelt, die in Interviews und Anwenderumfragen als relevant für einen Vergleich eingestuft worden sind. In entsprechenden Befragungen und Anwendertests hat sich von Beginn an gezeigt, dass diese Funktion leicht und intuitiv bedienbar ist, weshalb diesbezüglich im weiteren Projektverlauf nur kleinere Anpassungen notwendig waren.

Ausgehend von einer Studie von Turpin & Hersh (2001), welche u.a. die Fähigkeiten von Anwendern bei der Selektion relevanter Treffer aufzeigt, wurden unterschiedliche Mechanismen zur Manipulation der Tabelle implementiert. Es steht den Anwendern somit frei,

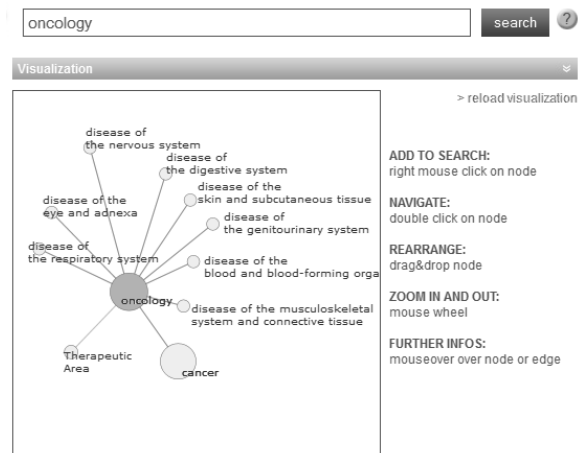


Abbildung 5: Visualisierung der Ontologie zur Suchanfragen-Unterstützung

Der Anwender kann anschließend die Ontologie explorativ erkunden, in dem er durch den Graphen navigieren und die jeweiligen Bezeichnungen der Knoten der Suchanfrage per rechten Mausklick hinzufügen kann. Zusätzliche Informationen zu allen Knoten und Kanten können in Form eines Tooltips abgerufen werden.

4 Fazit & Ausblick

Im Rahmen des Projektes SERENA ist ein modular aufgebauter, funktionsfähiger Prototyp entstanden, dessen Bausteine im nächsten Schritt an unterschiedlichen Stellen in die jeweiligen Produkte der Projektpartner integriert werden. Die modulare Bauweise war während des Projekts als zusätzliche Herausforderung zu betrachten, da üblicherweise ein einzelnes Zielprodukt definiert wird, auf das sich Konzeption, Design und auch die entsprechenden Evaluationen ausrichten. Im Rahmen von SERENA galt es somit, sowohl die einzelnen Bausteine alleinstehend als auch deren Zusammenspiel ausgiebig zu analysieren. Hierbei haben sich insbesondere die Untersuchungen des Filterkonzeptes als sehr interessant erwiesen, da die Grenzen der Bedienbarkeit und Erlernbarkeit zwischen komplexen und einfachen Filtermethoden deutlich herausgearbeitet werden konnten, was in praxisorientierten Projekten nur selten ausführlich getestet werden kann.

Danksagung

Wir danken Lukas Toggenburger, Martin Studer sowie unseren Projektpartnern Venture Valuation, Infocodex AG und EBD Group.

Literaturverzeichnis

Carroll, J.M. (1994). Making use: a design representation. *Communications of the ACM*, 37(12), 28-35.

- Cooper, A. (1999). *The inmates are running the asylum : why high-tech products drive us crazy and how restore the sanity*. Indianapolis: Sams.
- English, J., Hearst, M., Sinha, R., Swearingen, K. & Yee, P. (2002). *Flexible Search and Navigation using Faceted Metadata*. Unveröffentlichtes Manuskript, online verfügbar unter: <http://flamenco.berkeley.edu/papers/flamenco02.pdf>, (08.03.2012).
- Fruchterman, T. & Reingold, E. (1991). Graph drawing by force-directed placement. *Software- Practice & Experience*, 21, 11 (November 1991), 1129-1164.
- Goodwin, K. (2009). *Designing for the Digital Age: How to Create Human-Centered Products and Services*. Indianapolis: Wiley.
- Hearst, M. (2009). *Search User Interfaces*, Cambridge University Press, September, 2009.
- Hearst, M. (2008). UIs for faceted navigation: Recent advances and remaining open problems. *Proceedings of the Second Workshop on Human-Computer Interaction and Information Retrieval – (HCIR 2008)*, Redmond, S. 13-17.
- Hutchins, E., Hollan, J. & Norman, D. (1985). Direct Manipulation Interfaces. *Human-Computer Interaction*, 1(4), 311-338.
- Shneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. *Behavior and Information Technology*, 1(3), 237-256.
- Spree, U., Feißt N., Lühr A., Peisztal, B., Schroeder N., Wollschläger, P. (2011). Semantic Search - State-of-the-Art-Überblick zu semantischen Suchlösungen im WWW. In: Dirk Lewandowski (Hrsg.). *Handbuch Internet-Suchmaschinen 2– Neue Entwicklungen der Web-Suche*. Heidelberg: Akademische Verlagsgesellschaft, S. 77-109
- Sutcliffe, A. G., Ennis, M. & Watkinson, S. J. (2000). Empirical studies of end-user information searching. *Journal of the American Society for Information Science and Technology*, 51(13), 1211-1231.
- Tunkelang, D. (2009). *Faceted Search (Synthesis Lectures on Information Concepts, Retrieval, and Services)*. San Rafael: Morgan and Claypool Publishers.
- Turpin, A.H. & Hersh, W. (2001). Why batch and user evaluations do not give the same results. *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR'01*, 225-231.
- Weinhold, T., Bekavac, B., Hierl, S., Öttl, S. und Herget, J. (2009). Visualisierungen bei Internetsuchdiensten. In: Dirk Lewandowski (Hrsg.). *Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis*. Heidelberg: Akademische Verlagsgesellschaft, S. 249-284.

Kontaktinformationen

Schweizerisches Institut für Informationswissenschaft (SII),
Hochschule für Technik und Wirtschaft (HTW) Chur
Ringstrasse/Pulvermühlestrasse 57
CH-7004 Chur