Editorial

Andreas Henrich · Norbert Fuhr

Online publiziert: 30. September 2011
© Springer-Verlag 2011

Durch die Erfahrung mit einfachen und populären Anfragen erwarten viele Benutzer von Web-Suchmaschinen, dass IR-Systeme für Anfragen mit wenigen Suchbegriffen bereits nach dem ersten Schritt die gewünschten Ergebnisse liefern können. Tatsächlich sind Web-Suchmaschinen für diese Art von Informationsproblemen optimiert. Bei der Analyse von Suchmaschinenlogs bestätigen sich aber auch in diesem Kontext die Ergebnisse früherer empirischer Studien aus anderen Zusammenhängen, wonach Nutzer oft iterativ vorgehen und unterschiedliche Suchstrategien anwenden. Umso mehr werden interaktive Suchstrategien notwendig, wenn die Informationsbedürfnisse komplexer werden. Das vorliegende Themenheft widmet sich daher Modellen, Konzepten und Systemen, bei denen die Interaktion zwischen Nutzer und System bei der Bearbeitung von Informationsbedürfnissen im Vordergrund steht.

Schon der Titel dieses Themenheftes „Interaktives IR und Information Seeking“ macht dabei die verschiedenen Blickwinkel auf das Themenfeld deutlich. Während der Begriff „Interaktives Information Retrieval“ (IIR) der Informatik zugerechnet werden kann und wahlweise einen speziellen Aspekt des Information Retrieval oder Interaktiver Systeme bezeichnet, entstammt der Begriff „Information Seeking“ den Informationswissenschaften. Er bezeichnet den Prozess bzw. die Aktivität der Wissensbeschaffung und ist im Kon-

text von Informationsbedürfnissen und Informationsverhalten zu sehen.¹

Beide Wurzeln reichen dabei weit zurück. Interaktives Information Retrieval (IIR) gibt es als Anwendung schon seit den 1960er Jahren. In der Forschung hat dieses Thema aber lange Zeit kaum eine Rolle gespielt: Im Bereich der systemorientierten Ansätze gab es zwar Forschung an Relevance-Feedback-Methoden – diese konnten sich aber in der Praxis nicht durchsetzen. Ähnliches gilt für die zahlreichen Vorschläge für neue grafische Benutzerschnittstellen. Bei den benutzerorientierten Ansätzen gab es empirische Studien und darauf basierende kognitive Modelle, die aber ebenfalls ohne Auswirkung auf die Entwicklung von praktischen Systemen blieben. Erst seit einigen Jahren gibt es verstärkte Forschungsanstrengungen im Bereich IIR, zum einen im Zusammenhang mit kontextabhängigem Retrieval (wo die beiden Tagungsreihen IiX² und CONTEXT³ entstanden sind), zum anderen bei der Entwicklung und Evaluierung von Benutzerschnittstellen (insbesondere im Zusammenhang mit Evaluierungsinitiativen wie z.B. den „interactive tracks“ bei TREC⁴ und INEX⁵).

In diesem Heft sind nun fünf Arbeiten zum Schwerpunktthema versammelt. Die Beiträge behandeln aktuelle Forschungsfragen und geben so einen breiten Überblick, der – so unsere Hoffnung – zum Überdenken des eigenen Informationsverhaltens und der Nutzung von Suchmaschinen sowie zu einer neuen Sicht auf die uns im privaten und berufli-

A. Henrich (✉)
Fakultät für Wirtschaftsinformatik und Angewandte Informatik,
Uni Bamberg, Bamberg, Deutschland
e-mail: andreas.henrich@uni-bamberg.de

N. Fuhr
Abteilung Informatik und Kognitionswissenschaften, Uni
Duisburg-Essen, Duisburg, Deutschland
e-mail: norbert.fuhr@uni-due.de

¹ Siehe hierzu z. B. auch: Case, Donald O (2007) Looking for information: a survey of research on information seeking, needs, and behavior. Elsevier/Academic Press, Amsterdam.

² <http://www.iix2010.org/>.

³ <http://context-11.teco.edu/>.

⁴ <http://trec.nist.gov/>.

⁵ <http://www.is.informatik.uni-duisburg.de/projects/inex/>.

chen Alltag begegnenden Such- und Informationsbedürfnisse anregt.

Daniel Backhausen, Claus-Peter Klas und Matthias Hemmje geben in ihrem Beitrag *Adaptive Benutzerunterstützung in Interaktiven Informationssuchprozessen* zunächst eine umfangreiche Einführung in das Thema und einen breiten Überblick über die grundlegenden Arbeiten zum Information Seeking. Ein besonderer Schwerpunkt liegt dabei auf den verschiedenen Phasen des Prozesses der Informationssuche und entsprechenden kognitiven Modellen. Die Autoren gehen darauf aufbauend auf die Bedeutung des Kontextes des Informationssuchenden ein und analysieren die verschiedenen Kontextfaktoren sowie deren Erfassung und Nutzung in Systemen zur interaktiven Informationssuche. Die Berücksichtigung des Kontextes erschließt aus der Sicht des Information Retrieval wichtige zusätzliche Hintergründe zum Informationsbedürfnis des Nutzers und erlaubt so eine präzisere Bearbeitung von Anfragen. Die Autoren geben in ihrem Beitrag Hinweise zu den Problemen und Chancen, die sich hierdurch ergeben.

Carola Carstens und Dorothea Mildner berichten im Aufsatz *Query Reformulation Behavior in an Interactive Query Expansion Environment* von einer Nutzerstudie, in der untersucht wurde, wie die Gestaltung der Benutzeroberfläche die Art der gestellten Anfragen in einer interaktiven Suche beeinflusst. Hierzu wurde einer Vergleichsgruppe bei der Nutzung der FIS Bildung Literaturdatenbank eine erweiterte Schnittstelle zur Verfügung gestellt, die Vorschläge zur Modifikation der ursprünglichen Anfrage im Sinne einer Anfrageerweiterung (verwandter Term wird ODER-verknüpft) eines Tausches eines bisherigen Anfragebegriffs gegen einen verwandten Begriff sowie einer Anfrageeinschränkung (verwandter Term wird UND-verknüpft) anbietet. Die präsentierten Ergebnisse zeigen, dass die Unterstützung des Nutzers zu einer stärkeren und auch differenzierteren Nutzung der Möglichkeiten der Anfragesprache führt. Die Arbeit steht daher in diesem Heft als Beispiel für eine Evaluierung interaktiver IR-Systeme und als Beleg, dass eine zielgerichtete Unterstützung des Nutzers Veränderungen/Verbesserungen im Umgang mit den Möglichkeiten des Systems bewirken kann.

Sascha Kriewel beschreibt in seinem Artikel *Interaktives Retrieval und situationsabhängige Vorschläge*, wie Benutzer bei der Informationsrecherche durch das System unterstützt werden können. Ausgangslage ist dabei, dass zwar heute praktisch jedermann Zugang zu IIR-Systemen hat, aber nur wenige über eine solide Recherchekompetenz verfügen, was dann zu mageren Suchergebnissen führt. Das Ziel des von Herrn Kriewel entwickelten Systems ist nun, den Benutzern diese Kompetenz während der Recherche zu vermitteln, indem das System Vorschläge macht, wie die laufende Suche fortgesetzt werden kann. Die experimentellen Ergebnisse zeigen, dass derart unterstützte Benutzer erfolgreicher suchen, und zudem ist ein gewisser Lerneffekt

erkennbar. Dieser Ansatz könnte somit ein Weg sein, um die Informationskompetenz der Benutzer zu steigern.

David Zellhöfer und Ingo Schmitt stellen in ihrer Arbeit *Interaktives Information-Retrieval auf der Basis der Polyrepräsentation* ein Quantenlogik-basiertes Modell für IIR vor, wobei insbesondere das Konzept der Polyrepräsentation unterstützt werden soll: Dieses auf Peter Ingwersen zurückgehende Prinzip besagt, dass man die Erfolgchancen bei der Suche verbessern kann, indem das System die Dokumente auf verschiedene Art und Weise repräsentiert (indem etwa in einem Buchkatalog neben den bibliografischen Angaben auch eine Inhaltsbeschreibung, ein Miniaturbild des Buchdeckels sowie Benutzerrezensionen abgelegt sind). Zellhöfer und Schmitt präsentieren nun ein theoretisches Modell, das sowohl das eigentliche Information Retrieval als auch die Interaktion mit dem Benutzer beschreibt. Relevance-Feedback wird ebenfalls unterstützt, wobei das Modell auch die Verschiebung des Informationsbedürfnisses während der Suche erlaubt. Die Eignung des Modells wird durch einige Experimente aus dem Bereich des Multimedia Retrieval demonstriert.

Gerhard Heyer, Daniel Keim, Sven Teresniak, und Daniela Oelke betrachten in Ihrem Aufsatz *Interaktive explorative Suche in großen Dokumentenbeständen* Verfahren zur Extraktion und zur Visualisierung „auffälliger“ Terme in Textkollektionen, um dadurch den Benutzer bei der Exploration der Datenbasis zu unterstützen. Hierzu definieren sie ein Maß für die Volatilität eines Terms, und können so automatisch Begriffe identifizieren, deren Verwendung zeitabhängig schwankt, insbesondere aufgrund singulärer Ereignisse. Dadurch werden die für eine Kollektion kennzeichnenden Terme extrahiert, und der Benutzer bekommt schnell einen Überblick über den Inhalt des Dokumentenbestandes. Indem die Zeitabhängigkeit der Terme grafisch dargestellt wird, kann der Benutzer zudem Korrelationen zwischen Termen erkennen – die Interpretation dieser Beobachtung muss der Benutzer aber selbst leisten.

Neben diesen Schwerpunktbeiträgen finden Sie unter der Rubrik *Kurz erklärt* einen Beitrag von M. Thiele zu *Echtzeit-Data-Warehouse-Systemen* und einen Beitrag *Der Lehrstuhl „Datenbanken und Informationssysteme“ an der Universität Stuttgart stellt sich vor* unter der Rubrik *Datenbankgruppen vorgestellt*. Die Rubrik *Dissertationen* enthält in diesem Heft vier Kurzfassungen von Dissertationen. Weiterhin erscheinen in der Rubrik *Community* ein Aufruf zur Teilnahme an der Fachkollegienwahl der DFG sowie aktuelle Nachrichten aus der jüngsten Zeit.

Künftige Schwerpunktthemen

1 Data Streams and Event Processing

The processing of continuous data streams has become an important paradigm of modern data processing and management, covering many applications and domains such as

monitoring and controlling networks and complex production systems as well complex event processing in medicine, finance, or compliance.

A special issue of Datenbank-Spektrum wants to publish research contributions providing an overview over ongoing work in the data/event management area. The goal of the special issue is to attract both academic and industrial contributions to foster the exchange of ideas and a discussion about the state of the art and future directions.

Submissions covering topics from the following non-exclusive list are encouraged:

- Data stream management
- (Complex) event processing
- Case Studies and real-life usage
- Foundations, like semantics, maintenance and life cycle, optimization
- Statistical and probabilistic approaches
- Quality of service of continuous data management
- Stream mining
- Provenance in data streams and event processing
- Specific systems for event and stream processing, like „Conventional“ DBMS, main memory databases, sensor networks, cloud computing or hardware acceleration (GPU, FPGA, ...)
- Standardization efforts

Guest editors:

Peter Fischer, ETH Zürich,
peter.fischer@inf.ethz.ch

Daniela Nicklas, Uni Oldenburg,
dnicklas@acm.org

Bernhard Seeger, Uni Marburg,
seeger@Mathematik.Uni-Marburg.de

2 Information Extraction

Database technology as tool to structure, retrieve, and analyze information has proven to be of great value. However, today's information repositories continuously grow and diversify and range from rather unstructured document collections on the Web, over partially structured information in business warehouses, to structured, yet huge experimental data collections, e.g., in the life sciences. Extracting information from such heterogeneous sources and making it available for analysis is a hard problem needing complex algorithms and mining technologies. This special issue aims to provide an overview over the spectrum of techniques addressing different types of information sources.

Submissions covering topics from the following non-exclusive list are encouraged:

- Web information extraction
- Data mining and visualization
- Information extraction from data streams
- Business intelligence and information mining

Expected size of the paper: 8–10 pages

Notice of intent for a contribution: January 5th, 2012

Guest editor:

Wolf-Tilo Balke, TU Braunschweig,
balke@ifis.cs.tu-bs.de

Deadline for submissions: March 1st, 2012

3 Scientific Data Management

The past decade has witnessed a dramatic increase in scientific data being generated in the physical, earth, and life sciences. This development is primarily a result of major advancements in sensor technology, surveying techniques, computer-based simulations, and instrumentation of experiments. In a special issue of the “Datenbank-Spektrum”, we want to publish original work on different aspects related to the management and analysis of scientific data. The objective of this special issue is to exchange ideas between academia and industry and to discuss recent developments, challenges, and future directions in scientific data management.

Topics of interest include (but are not limited to)

- Modeling and representation of data, metadata, ontologies, and processes for scientific application domains
- Integration and exchange of scientific data
- Design, implementation, and optimization of scientific workflows
- Architectures and components for scientific computing and eScience, including Web portals, repositories, and digital libraries
- Annotation and provenance of scientific data
- Mining and analysis of large-scale scientific datasets
- Case studies and applications related to scientific data management in all domains, with a particular focus on biology, physics, chemistry, medicine, and geography

Paper format: 8–10 pages, double column

Notice of intent for a contribution: March 15th, 2012

Guest editors:

Michael Gertz, Heidelberg University,
gertz@informatik.uni-heidelberg.de

Wolfgang Müller, HITS gGmbH,
wolfgang.mueller@h-its.org

Deadline for submissions: June 1st, 2012