

09101 Abstracts Collection
Interactive Information Retrieval
— Dagstuhl Seminar —

Nicholas J. Belkin¹, Norbert Fuhr², Joemon Jose³ and C.J. van Rijsbergen⁴

¹ Rutgers Univ. - New Brunswick, USA

nick@belkin.rutgers.edu

² Universität Duisburg-Essen, D

norbert.fuhr@uni-due.de

³ University of Glasgow, GB

jj@dcs.gla.ac.uk

⁴ University of Glasgow, GB

keith@dcs.gla.ac.uk

Abstract. From 01.03. to 06.03.2009, the Dagstuhl Seminar 09101 “Interactive Information Retrieval” was held in Schloss Dagstuhl – Leibniz Center for Informatics. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Interactive Information Retrieval, Evaluation, Context, Modeling

09101 Workshop Report – Interactive Information Retrieval

There is need for more foundational research in the development of interactive information retrieval systems. The results of a week long discussion by a group of multidisciplinary researchers have reported here. A brief description of main activities and major recommendations of the workshop are reported here.

Keywords: Interactive Information Retrieval, Evaluation, Context, Modeling

Joint work of: Fuhr, Norbert; Belkin, Nicholas; Jose, Joemon M; van Rijsbergen, Cornelis J.

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2009/2152>

Evaluation of Interactive Retrieval

Maristella Agosti (Università di Padova, IT)

Log is a concept commonly used in computer science; in fact, log data are collected by an operating system to make a permanent record of events during the usage of the operating system itself. This is done to better support its operations, and in particular its recovery procedures. Due to the experience gained in the management of operating systems and many application systems that manage permanent data, log procedures are commonly put in place also to collect and store data on the usage of an application system by its users. Initially, these data were mainly used to manage recovery procedures of the application system, but over time it became apparent that they could also be used to study the usage of the application by its users, and to better adapt the system to the objectives the users were expecting to reach. In the context of the Web, the storage and the analysis of Web log files are mainly used to gain knowledge on the users and improve the services offered by a Web portal, without the need to bother the users with explicit collection of information on use of the portal.

In general, users studies and logs are used in a separate way, since they are adopted with different aims in mind. Ingwersen and Jarvelin reported that it seems more scientifically informative to combine logs together with observation in naturalistic settings. Pharo and Jarvelin suggested a systematic use of triangulation of different data collection techniques as a general approach in order to get better knowledge of the Web information search process. Taking inspiration from this general approach, we have conceived a method of combining implicit and explicit user interaction data to gain information to be used for personalization purposes.

So, data log analysis can be combined with the results of data derived from user studies to evaluate information access services.

Keywords: Evaluation of interactive retrieval, log analysis, Cranfield approach, test collection

A REVOLUTION IN BROWSING

Marcia J. Bates (Univ. of California - Los Angeles, US)

Browsing capabilities in interactive information systems have traditionally consisted simply of the capability to scan down a list, moving a cursor to display more information to scan. Indeed, many standard definitions of browsing state or imply that browsing consists, in its essence, of scanning. It is argued here that this conception of browsing grossly mis-states the nature of the process. Our misunderstanding of the nature of browsing has led to the provision of the wrong sort of browsing capabilities for the information searcher in online systems. Providing a scanning capability is fine, but it does not support browsing, which is a more complex and interesting behavior.

Drawing on psychological and anthropological literature, the author has recently made the argument that:

Browsing is seen to consist of a series of four steps, iterated indefinitely until the end of a browsing episode: 1) glimpsing a field of vision, 2) selecting or sampling a physical or informational object within the field of vision, 3) examining the object, 4) acquiring the object (conceptually and/or physically) or abandoning it. Not all of these elements need be present in every browsing episode, though multiple glimpses are seen to be the minimum to constitute the act.

In other words, browsing consists of glimpses, not scanning. In a glimpse, the browser sees a wide area of the environment all at once, catches sight of something of interest within that area, then focuses in on the object of interest, examines it more closely (visually or haptically), then engages in a new glimpse. Based on psychological research on visual search, this way of browsing is actually a surprisingly efficient way of searching an unknown environment. The detailed research is described that has been done by psychologists on visual search. They have demonstrated that visual search is actually a two-stage process, which this common human behavioral pattern for browsing perfectly captures, and thus optimally utilizes.

Implications are drawn for interactive information system design.

Keywords: Browsing; Scanning, Interactive browsing capabilities

Full Paper:

<http://InformationR.net/ir/12-4/paper330.html>

See also: Bates, Marcia J. (2007). "What is browsing-really? A model drawing from behavioural science research" *Information Research*, 12(4) paper 330.

Cognitive and Context Modeling for Interactive Information Retrieval

Nicholas J. Belkin (Rutgers Univ. - New Brunswick, US)

I give an historical overview of how cognitive (and other) models of users of interactive IR (IIR) systems have been elicited, constructed and used. Cognitive models in the domain of interactive information retrieval (IIR) are understood as models that a system (or a person) constructs of a(nother) person's "information need"; these are called "user models". Context models are models that a system (or a person) constructs of the conditions that led a(nother) person to engage in information seeking behavior, various characteristics of that person, and various aspects of that person's environment, broadly construed. Both types of models are used in personalization of IIR.

The review begins with discussion of Robert Taylor's 1968 article in which he proposed four levels of "information need" or "query", and five "filters" according to which the librarian and the information seeker identify and clarify (i.e. model)

various aspects of the user, the user's goals, the topic of interest and so on. I then review various approaches to understanding of why people engage in information seeking behavior, and of automatically constructing cognitive and contextual models.

Significant change points include the "cognitive turn" in the early to mid 1970s, modeling the human intermediary in the 1980s, and the "interactive turn" in the 1990s. The review concludes with a brief description of, and comments on the current state of cognitive and contextual modeling in IIR.

Keywords: Cognitive models, Context models, Interactive information retrieval

Getting Personal: Personalization of Support for Interaction with Information

Nicholas J. Belkin (Rutgers Univ. - New Brunswick, US)

An overview of personalization for interactive information retrieval, including a classification of facets of personalization. This was presented at the First Workshop on Adaptive Information Retrieval, Glasgow, 2006.

Keywords: Personalization, Contextualization, Interactive Information Retrieval

Summary of TREC Interactive Track Data Collection Protocol

Nicholas J. Belkin (Rutgers Univ. - New Brunswick, US)

This is an abbreviated, from memory, description of the general conduct of the TREC model for IIR experiments, with special reference to data types collected.

Interacting with information

Ann Blandford (University College London, GB)

We live in an information age", but information is only useful when it is interpreted by people in the context of their goals and activities. The volume of information to which people have access is growing rapidly, vastly outstripping people's ability to find, assimilate and manage it. In order to design technologies that better support information work, it is necessary to better understand the details of that work. In this talk I will discuss how people interact with information in terms of an "information journey", in which people iteratively: recognise a need for information; find information; interpret and evaluate that information in the context of their goals; and use the information to support

their broader activities. People's information needs may be explicit and clearly articulated, but may be exploratory and evolving. Widely used tools supporting information access support clearly defined information requirements well, but provide limited support for other information needs. Most other stages of the information journey are poorly supported at present. Novel design solutions are unlikely to be purely digital, but to exploit the rich variety of information resources, digital, physical and social, that are available. Theories of information interaction and sensemaking can highlight new design possibilities that augment human capabilities. In this talk, I will discuss relevant theories and findings for understanding information behaviours, to both assess existing tools and identify requirements for the future.

A Probability Ranking Principle for Interactive IR

Norbert Fuhr (Universität Duisburg-Essen, DE)

The classical Probability Ranking Principle (PRP) forms the theoretical basis for probabilistic Information Retrieval (IR) models, which are dominating IR theory since about 20 years.

However, the assumptions underlying the PRP often do not hold, and its view is too narrow for interactive information retrieval (IIR).

In this talk, a new theoretical framework for interactive retrieval is proposed: The basic idea is that during IIR, a user moves between situations. In each situation, the system presents to the user a list of choices, about which s/he has to decide, and the first positive decision moves the user to a new situation. Each choice is associated with a number of cost and probability parameters. Based on these parameters, an optimum ordering of the choices can be derived - the PRP for IIR. The relationship of this rule to the classical PRP is described, and issues of further research are pointed out.

Full Paper:

<http://dx.doi.org/10.1007/s10791-008-9045-0>

See also: N. Fuhr: A Probability Ranking Principle for Interactive Information Retrieval. Information Retrieval 11(3), 2008.

Interactive Information Access in Linked Annotated Media

Lynda Hardman (CWI - Amsterdam, NL)

The Web of data enables fragments of information to be identified, described and connected together in a rich information environment. Users requiring information are faced with the problem of finding out what information is available, and obtaining sufficient fragments to successfully carry out their task. Systems

supporting these tasks can use the fragments, descriptions of them and relationships among them, to improve query formulation, selection and presentation of information. Questions to be answered are: which information needs can be better supported, and how can the Web of data help.

We have constructed prototype information systems to investigate both the desired functionality from, and user interfaces to, these repositories. Our target user group has been expert cultural heritage users.

Our experience is that grouping and ordering are useful basic concepts, where: "ranking" is an often used means of ordering; grouping provides a means of indicating the range of concepts present in the result set; and different groups may require different orderings. Mapping thesauri concepts to more general concepts, such as person, location and date/time, has also been effective in increasing interface accessibility.

Keywords: Semantic web, information exploration, information gathering, user interaction

Understanding cultural heritage experts' information seeking needs

Lynda Hardman (CWI - Amsterdam, NL)

This may serve as inspiration for identifying user tasks that can usefully be supported.

We report on our user study on the information seeking behavior of cultural heritage experts and the sources they use to carry out search tasks. Seventeen experts from nine cultural heritage institutes in the Netherlands were interviewed and asked to answer questionnaires about their daily search activities. The interviews helped us to better understand their search motivations, types, sources and tools. A key finding of our study is that the majority of search tasks involve relatively complex information gathering. This is in contrast to the relatively simple fact-finding oriented support provided by current tools. We describe a number of strategies that experts have developed to overcome the inadequacies of their tools. Finally, based on the analysis, we derive general trends of cultural heritage experts' information seeking needs and discuss our preliminary experiences with potential solutions.

Keywords: Cultural heritage experts, information seeking

Joint work of: Amin, Alia K.; van Ossenbruggen, Jacco; Hardman, Lynda; van Nispen, Annelies

Full Paper:

<http://doi.acm.org/10.1145/1378889.1378897>

See also: Alia K. Amin, Jacco van Ossenbruggen, Lynda Hardman, Annelies van Nispen: Understanding cultural heritage experts' information seeking needs. JCDL 2008: 39-47

Understanding Interactive IR

Kalervo Jarvelin (University of Tampere, FI)

Abstract. The motivation for this talk is the view that the ultimate goal of information retrieval (IR) is to support humans to better access information in order to better carry out their tasks. Therefore IR research should provide methods and techniques to improve the retrieval/access process. I have three messages. First I will argue that IR research is not focusing its efforts properly to serve the goal above. The Cranfield paradigm of evaluation tends to lose its power as soon as one looks at human performance instead of system performance. Secondly, I will argue that searchers using IR systems make use of rather unorthodox queries (from the Cranfield point-of-view) and sessions. Their search strategies have not been sufficiently described and cannot therefore be properly understood, supported or evaluated. Thirdly, I will argue that searchers engage in an information seeking process which they have found effective enough based on their searching experiences. They try not to optimize the search result alone but the entire process (and effort) and its expected contribution to their primary task. This can be better understood in terms of the management science theory "incrementalism" than in terms of rationalism.

Keywords: Interactive information retrieval, theory, evaluation

User Interfaces for Interactive Information Retrieval

Hans-Christian Jetter (Universität Konstanz, DE)

Users of digital libraries are nowadays confronted with information spaces that are rapidly growing in quantity, heterogeneity, relationality and dimensionality. Therefore, more effective tools are required to facilitate the exploration and search in these information spaces. MedioVis is a flexible application for the visual exploration of such data that is especially designed for users without prior professional experience in search, retrieval or visualization. To give them a satisfying search experience, different views on the data space are provided. MedioVis supports analytical and browsing oriented exploration strategies through the usage of multiple coordinated visualizations and a consistent and supportive interaction design. Further, we intentionally decided to use and combine visualizations that base on well-known and straightforward concepts (e.g. tables, browser or scatter plots). These visualizations provide visual filtering mechanisms (e.g. table filters or zooming into a region of a scatter plot) to support a natural way of query formulation and refinement. We called our table visualization HyperGrid, because it combines the functionality of a table, a browser and avoids information overload, by accessing detail information through a continuous semantic zoom into a table cell. Figure 1 shows the two coordinated views of MedioVis 1.0 called HyperGrid and the HyperScatter.

MedioVis 1.0 is funded science 2005 by the German Research Foundation (DFG) in the domain Scientific Library Services and Information Systems (LIS). The first system of MedioVis 1.0 was launched four years ago and still undergoes iterative development and evaluation cycles. To gain continuous end-user feedback and insights in real interaction behavior, we are running MedioVis 1.0 for over three years in the media library of the University of Konstanz.

The latest version called MedioVis 2.0 relies primarily on the paradigm of zoomable user interfaces and object-orientation. In consequence, no windows, icons, menus, files or dialogs are used. For the wide varied activities of knowledge workers, we further integrated different techniques to search and explore the information space via different visualizations. To keep and manage information objects, MedioVis 2.0 offers personalization functionalities.

MedioVis 2.0 is based on a new user interface paradigm called Zoomable Object Oriented Information Landscape (ZOIL). Thereby, an information landscape of infinite size serves as basic visualization and starting point. MedioVis 2.0 arranges each media object, corresponding to their primary genre on the landscape. Users are able to navigate in this landscape by zooming and panning. This navigation technique takes advantage of the human abilities of visual-spatial orientation and remembering visual "landmarks". By means of this concept, users are able to utilize zooming and panning operations as search strategy in the media collection. Analytical search methods are supported by MedioVis 2.0 as another way to formulate information needs. Users are able to enter text queries into a search field on the upper right of the screen. With each key press the visual representation of matching objects expands. We use the concept of Dynamic Queries and Sensitivity that declare a direct highlighting of objects, which still match the current query instead of removing all non-matching objects. Portals provide a supplementary way of exploration. By selecting an arbitrary region of the information landscape via a bounding box, the user creates a portal, providing a special view on the underlying media objects. Through this concept, MedioVis 2.0 offers multiple visualization techniques - for understanding, filtering and querying - ranging from a Rapid Serial Visual Presentation over the HyperScatter to the HyperGrid.

MedioVis 2.0 provides a unified user experience, not only on a desktop PC, but also on different devices such as shown for multitouch tabletops. To check the capability of the concept, we will transfer MedioVis 2.0 to complementary devices like large high-resolution displays (e.g. public walls, digital "advertising pillars") or mobile gadgets (e.g. smart phones, netbooks).

Joint work of: Reiterer, Harald; Jetter, Hans-Christian

Measuring Result Utility through Interactive Experiments

Thorsten Joachims (Cornell University, US)

The goal of a retrieval system is to provide the users with results of maximum utility. But how can utility be measured?

The conventional approach is to equate utility with some measure (e.g. Avg. Precision, NDCG) derived from expert relevance judgments. Since this has several well-know problems (e.g. divergence between user and expert judgments, ignorance of user context, cost, availability), it raises the question of whether utility can be elicited directly from the user?

While users provide observable feedback in terms of, for example, clicks and query reformulations, it is unclear how these relate to utility (e.g. do more clicks indicate higher utility). The relationship between clicks and utility to the user becomes clearer, however, when moving from a passive observation model to an interactive experiment setting. Interactive retrieval systems can conduct such interactive experiments, making implicit feedback data interpretable.

In this talk, I will discuss interactive experiment designs under which clicks accurately reveal ordinal statements about the relative utility of two retrieval functions. These experiment designs are evaluated in a controlled study on the E-Print ArXiv. It is shown that such interactive experiments can be used to directly compare different retrieval functions based on the user's clicking behavior. I also discuss how similar experiment designs can be used to adapt system behavior during or search session, and for long-term learning of improved retrieval functions.

Further details:

F. Radlinski, M. Kurup, T. Joachims, How Does Clickthrough Data Reflect Retrieval Quality?, Proceedings of the ACM Conference on Information and Knowledge Management (CIKM), 2008.

Keywords: Evaluation, Implicit Feedback, Learning

Layout, Content and Structure in Interactive IR

Matthias Jordan (Universität Duisburg-Essen, DE)

We present a new generic framework for designing user interfaces for interactive IR. Based on the distinction between the logical, layout and content view on documents, we regard IR as a combination of selection, projection, organization and visualization on these aspects. These transformations should be configured according to the information seeking strategy currently chosen by the user.

Keywords: Polyrepresentation, interaction, retrieval, Hypergrid

Joint work of: Fuhr, Norbert; Jordan, Matthias; Frommholz, Ingo

Affective Search

Joemon Jose (University of Glasgow, GB)

By analyzing explicit & implicit feedback information retrieval systems can determine topical relevance and tailor the search criteria to the user needs. In this talk, I will explore whether it is possible to infer what is relevant for a user in an information retrieval task, through the observation of their affective behavior.

The sensory data employed in our study vary between facial expressions and other peripheral physiological signals, which are all regarded as indicative of the user's current affective state. The modeling goal is to predict with reasonable accuracy topical relevance.

We extract a set of features from the signals and explore the data using classification methods, such as SVM and KNN.

The results of our initial evaluation indicate that prediction of relevance is possible, to a certain extent, and models can benefit from taking into account user affective behavior.

Keywords: Emotion, information seeking, relevance

Full Paper:

<http://www.dcs.gla.ac.uk/~jj/publications/>

Multi-faceted Exploratory Search and Users' Cognitive Process

Noriko Kando (National Institute of Informatics - Tokyo, JP)

This talk outlines MEW (Multi-faceted Exploratory search system for the Web) and the lessons learned from a series of cognitive studies of the users' exploratory information seeking on the Web. MEW provides functions of "Search", "Navigation" along a multi-faceted directory, and "View" in which the search results were classified into a multifaceted scheme and displayed. A user can start a session either of search, navigate or view and seamlessly continue the flow of exploratory using these three modes. And according to the search results, it suggest a few "Viewpoint Frames" which are sub-directories relevant to the search results in order to assist the user to navigate on a specific sub-directory relevant to the context, to suggest the user to directions to investigate more or to provide the user a kind of meta-knowledge to suggest what types information shall be searched to accomplish the task or purpose of the search, etc. The fundamental system design was derived from a series of users studies investigating the users' cognitive process during the exploratory searches on the Web with varies types of users with different types of users information seeking tasks like "Report Writing", or "Trip Planning". The talk concludes some thoughts on future direction of the evaluation of such types of interactive information retrieval systems.

Keywords: Multi-faceted search, exploratory search, context aware, viewpoint, cognitive process, users task, serendipity

Use cases as a bridge between benchmarking and validation

Jussi Karlgren (SICS - Kista, SE)

Benchmarking or system evaluation has a strong tradition as a research method for information retrieval.

Details of aggregation of performance over data sets and test buckets and of quantification of results with respect to the target notion of relevance are discussed regularly, and sometimes parametrized to cover several usage scenarios.

Success of future information systems is only in part due to system qualities of the type typically tested in benchmarking exercises. Collection coverage, interface design, uptime, branding and marketing concerns all have as strong or stronger effect on system take-up and eventual success than the qualities of the retrieval algorithm. The carefully designed evaluation schemes cannot predict usefulness of the enveloping system: they are only intended to test some central components.

Moving from topic- and task-oriented retrieval to other usages such as diversion and entertainment and from text to other media it is even clearer that the bare retrieval component of the system is only a small factor in predicting usefulness of a system. Aggregated quantified measures of whatever type do not model the various usage scenarios that a system may be intended to address.

Validating the various tests in terms of predicted usage is the crucial further step for establishing system effectiveness and usefulness. This requires system developers to engage in user studies, preferably situated in realistic or near-realistic usage contexts, but including partners competent in the craft of designing and performing user studies is not usually a part of multimedia system development efforts.

Bridging the two activities of parametrized system benchmarking and realistic or near-realistic usage evaluation is the formulation of use cases. Use cases could be formulated to lock down parameters for informed system evaluation and to establish appropriate contextual constraints for validating usage scenarios for the informed user studies. This would enable information flow between system developers and studies of usage and would allow systems to be compared even across various system architectures.

The various commission-funded projects on multimedia information access have been asked by CHORUS, a commission-funded coordination action, to describe the use cases they intend to address. The reports from the participating projects has been composed into a use case typology, intended to capture the most significant dimensions of variation with respect to system evaluation. This work is still under progress, but will be presented at Dagstuhl.

An Examination of Different Delivery Modes for Interactive IR Studies

Diane Kelly (University of North Carolina- Chapel Hill, US)

While the laboratory was once the primary location where researchers studied IIR, researchers are increasingly using different delivery modes to gather data from human subjects. Researchers have used mechanical Turk to gather relevance assessments (<https://www.mturk.com/mturk/welcome>), online games such as ESP to generate training data for machine classifiers (see <http://www.gwap.com/gwap/>

for more examples) and remote usability testing to interact with geographically dispersed clients. One of the primary benefits of conducting a study in a laboratory is that the researcher has increased control over the study situation. Many sources of variance can be identified and controlled and alternative explanations can be ruled-out. Furthermore, the researcher can ensure that a standard protocol is followed and that subjects are not engaged in extracurricular activities. When studies are conducted online the researcher is no longer able to fully monitor the subject and cannot even be certain that there is a real human participant. However, an online delivery mode allows researchers to study a wider range of more diverse users and researchers do not always have to be present when subjects are participating (in other words, you can collect data while you eat, sleep, run and holiday!). It could further be argued that studies delivered via the Internet are particularly well-suited to IIR since a large amount of information seeking and retrieval takes places in online environments.

In this talk, I will review different study delivery modes that have been used in IIR, as well as in other related research areas. I will present preliminary results of an IIR experiment that was completed by two cohorts of subjects: the first cohort completed the study online and the second completed the study in the laboratory. Finally, I will discuss the benefits and limitations of these various delivery modes with respect to different types of IIR studies (e.g., experiment, evaluation, usability test). It is hoped that this talk will generate further consideration and discussion of which techniques are appropriate for IIR research.

Context Aware Interactive Information Retrieval

Claus-Peter Klas (FernUniversität in Hagen, DE)

To effectively support the user during the interactive IR process we need to capture the actual context, task or cognitive space of the user.

Based on the logging of the information dialog of the user during the information search process we propose to use the implicit and explicit feedback give to either provide better query terms or to re-rank the result list accordingly.

The context model about the user consists of several actions like entered query terms, viewed and stored documents or associated journals and conferences. All information over time lead to a rich context representation which can be used to solve the actual task.

Supporting Interactive Search with Adaptive Suggestions

Sascha Kriewel (Universität Duisburg-Essen, DE)

The problem of supporting end users in finding and deriving good strategies for their information search is well known. During an interactive search session users have to make many decisions regarding the best way to proceed from their current situation. This is a challenge for many inexperienced searchers.

Using the successful searches of previous users, an adaptive support system can suggest appropriate next steps at any point during an interactive search. Case-base reasoning techniques have been used to implement such a support system, which uses the situations of previous users as cases and suggests search actions and stratagems as solutions (adapted to the situation of the current users). The system can learn from user participation.

Versions of the suggester have been integrated into a digital library search system (Daffodil) as well as the open source browser Firefox, where it is used to support Google searches with interactive suggestions.

Contextual Information Retrieval Using Vector Spaces

Massimo Melucci (Università di Padova, IT)

The talk will introduce a geometric model and its investigation for contextual information retrieval. The geometric model leverages recent advances of vector space-based information retrieval. Some issues of the implementation of the model will also be discussed in the context of implicit feedback tailored for each task and personalized for each user. Besides the theoretical and methodological issues, the results of some experiments using rich interaction logs will also be described to show both the potential benefits of the proposed model and the value of incorporating multiple sources of interaction evidence when developing implicit feedback algorithms.

Keywords: Contextual information retrieval

Interactivity, Learning, Trust, and Revisitation

Natasa Milic-Frayling (Microsoft Research UK - Cambridge, GB)

Interactive IR amplifies the issues related to learning from interaction with the data, trust, and reconciliation of multiple perspectives on relevance issues. Recently we conducted experiments that involve relevance assessment of large documents (books) through extensive interaction with the content, search engine, and multiple human assessors. This exercise made us revisit the notion of relevance, trust, and authoritativeness in relevance judgements. I suggest that we discuss this theme as it touches on the very foundations of interactivity: evolving view of the topic scope and relevance assessment.

I would also like us to give some attention to discussing interactive retrieval as part of the content management workflows. For example, lots of information access is aimed at revisiting the content that has been used in the recent or more distant past (desktop data, mail, online blogs and community data, web, etc.). Some content revisitation is better supported by browsing while other requires search. In order to support efficient access to content we need to look at the practices around interaction with the content during complex tasks that

users perform (not restricted to search and browsing). We may be able to capture additional context by leveraging user's need to manage their work efficiently and thus provide more metadata about their activities. We can then use the data for better data revisitation. I can share results of our recent work on introducing tagging into the desktop environment and showing how new ways of browsing and search are enabled.

Proactive Context Aware Retrieval with Mobile Devices

Stefano Mizzaro (Università di Udine, IT)

I will present the Context Aware Browser, a novel paradigm for context-aware access to Web contents with mobile devices. The idea is to allow automatic download of Web pages, and even automatic execution of Web applications, on user's own mobile device.

The Web resources are not simply pushed on the mobile device; rather, they are selected on the basis of the context the user is in: context data (mainly location, but not only) are used to build a query sent to an external search engine, that selects the most relevant Web content.

I will describe the idea, provide some examples, show a video of a recently built prototype, present implementation issues, discuss our specific evaluation methodology and the results, and sketch future work and problems. This is an ongoing project, started about five years ago; it is joint work with the Context Aware and Mobile Systems laboratory (smdc.uniud.it) and the MoBe spinoff (www.mobe.it) at Udine University.

Keywords: Context-aware, information retrieval, mobile devices

Full Paper: <http://drops.dagstuhl.de/opus/volltexte/2009/2151>

See also: Paolo Coppola, Vincenzo Della Mea, Luca Di Gaspero, Danny Mischis, Stefano Mizzaro, Elena Nazzi, Ivan Scagnetto, Luca Vassena, Context-Aware Browser, Technical Report UDMI/9/2008, University of Udine, Mathematics and Computer Science Dept., 2008

Demo: CREDO, Credino, SmartCREDO, CloudCREDO

Stefano Mizzaro (Università di Udine, IT)

Demo of four clustering search engines. CREDO clusters the retrieved documents and presents, besides the usual ranked list, a set of hierarchically organized clusters; it is adequate for a desktop (large screen) computer. Credino and SmartCREDO are CREDO versions for PDAs (smaller screens) and cellphones (even smaller), respectively. CloudCREDO presents the hierarchy of clusters as a tag cloud and is designed for small screen devices.

Context matters(?) - user behaviour and element retrieval

Ragnar Nordlie (Oslo University College, NO)

Over a 5-year period, the INEX (INitiative for the Evaluation of XML retrieval) interactive track has gathered data (logs, questionnaires, observation notes) from more than 400 search sessions where users interact with systems which present search results on different levels of granularity.

This presentation will summarize and discuss findings on the use and value of document context when users attempt to decide the relevance of retrieved material. In particular, this will be discussed in the context of search effort, and the value and ability of log data to determine search effort will be considered.

Multi-modal and Multi-functional Aspects of Information

Andreas Rauber (TU Wien, AT)

IR systems in many cases implement an ad-hoc combination of modalities, features, and interpretation of object semantics with interpretations of user needs and queries. In this talk I want to raise questions concerning a systematic approach as to

- where the information resides in a digital object and its environment, i.e. which aspects/dimensions/modalities of information here are in any specific setting,
- why it exists and is being searched for, i.e. for which purpose it was created by whom, which context it is living in, and what the intention of a searcher might be, and
- ultimately, what information actually is, which incarnations of representations it exists in, how it can be indexed and described in various modalities even if the task at hand may seem inherently single-modal

I will use scenarios and show examples from three different domains to illustrate the above questions, namely

- **Music Retrieval, with a focus on the multi-modal aspects of music information, including textual, image and video dimensions, and potential combinations of those to capture the complexities of seemingly mono-modal (acoustic) information in real-world settings**
- **Web Archiving, with a focus on ethical aspects of Web Archive access and consequences of/for search functionalities and machine learning algorithms analyzing Web Archive content**
- **Digital Preservation, focussing on what actually is the information represented by and in an object that is to be preserved (and thus the target of retrieval) in terms of significant properties, context of information and authenticity, going beyond the simple mono/multi-modal search of primary content**

These examples demonstrate numerous issues calling for a consolidated view of information and the information retrieval process. They call for an integrated model for matching information representation, structuring and interpretation across a range of orthogonal dimensions, with use cases detailing user needs and intentions. Only by providing such an integrated view and structured matching we may be able to achieve a consolidated means of benchmarking and evaluating interactive information retrieval for the complex tasks encountered in real-world settings.

Keywords: Music Retrieval, Web Archive Mining, Genre Analysis, Multi-modal Retrieval

User Interfaces for Interactive Information Retrieval

Harald Reiterer (Universität Konstanz, DE)

Users of digital libraries are nowadays confronted with information spaces that are rapidly growing in quantity, heterogeneity, relationality and dimensionality. Therefore, more effective tools are required to facilitate the exploration and search in these information spaces. MedioVis is a flexible application for the visual exploration of such data that is especially designed for users without prior professional experience in search, retrieval or visualization. To give them a satisfying search experience, different views on the data space are provided. MedioVis supports analytical and browsing oriented exploration strategies through the usage of multiple coordinated visualizations and a consistent and supportive interaction design. Further, we intentionally decided to use and combine visualizations that base on well-known and straightforward concepts (e.g. tables, browser or scatter plots). These visualizations provide visual filtering mechanisms (e.g. table filters or zooming into a region of a scatter plot) to support a natural way of query formulation and refinement. We called our table visualization HyperGrid, because it combines the functionality of a table, a browser and avoids information overload, by accessing detail information through a continuous semantic zoom into a table cell. Figure 1 shows the two coordinated views of MedioVis 1.0 called HyperGrid and the HyperScatter.

MedioVis 1.0 is funded since 2005 by the German Research Foundation (DFG) in the domain Scientific Library Services and Information Systems (LIS). The first system of MedioVis 1.0 was launched four years ago and still undergoes iterative development and evaluation cycles. To gain continuous end-user feedback and insights in real interaction behavior, we are running MedioVis 1.0 for over three years in the media library of the University of Konstanz.

The latest version called MedioVis 2.0 relies primarily on the paradigm of zoomable user interfaces and object-orientation. In consequence, no windows, icons, menus, files or dialogs are used. For the wide varied activities of knowledge workers, we further integrated different techniques to search and explore the information space via different visualizations. To keep and manage information objects, MedioVis 2.0 offers personalization functionalities.

MedioVis 2.0 is based on a new user interface paradigm called Zoomable Object Oriented Information Landscape (ZOIL). Thereby, an information landscape of infinite size serves as basic visualization and starting point. MedioVis 2.0 arranges each media object, corresponding to their primary genre on the landscape. Users are able to navigate in this landscape by zooming and panning. This navigation technique takes advantage of the human abilities of visual-spatial orientation and remembering visual "landmarks". By means of this concept, users are able to utilize zooming and panning operations as search strategy in the media collection. Analytical search methods are supported by MedioVis 2.0 as another way to formulate information needs. Users are able to enter text queries into a search field on the upper right of the screen. With each key press the visual representation of matching objects expands. We use the concept of Dynamic Queries and Sensitivity that declare a direct highlighting of objects, which still match the current query instead of removing all non-matching objects. Portals provide a supplementary way of exploration. By selecting an arbitrary region of the information landscape via a bounding box, the user creates a portal, providing a special view on the underlying media objects. Through this concept, MedioVis 2.0 offers multiple visualization techniques - for understanding, filtering and querying - ranging from a Rapid Serial Visual Presentation over the HyperScatter to the HyperGrid.

MedioVis 2.0 provides a unified user experience, not only on a desktop PC, but also on different devices such as shown for multitouch tabletops. To check the capability of the concept, we will transfer MedioVis 2.0 to complementary devices like large high-resolution displays (e.g. public walls, digital "advertising pillars") or mobile gadgets (e.g. smart phones, netbooks).

Keywords: Human Computer Interaction, Usability, Interaction Design, Visualization

Joint work of: Reiterer, Harald; Jetter, Hans-Christian

Visual resource discovery: Can user interaction and automated processing replace expensive meta-data?

Stefan Rueger (The Open University - Milton Keynes, GB)

This provocative position talk looks at the task of visual resource discovery, viz image and video search engines, and asks the provocative question whether we need expensive meta-data at all. Instead it is proposed to use automated visual annotation to produce unreliable but still useful keywords and at the same time put the user in the information retrieval loop, so their feedback narrows down the wanted resources with a combination of searching and browsing.

Keywords: Metadata, visual resource discovery, user interaction

Managing Social Wisdom: Efficient and Effective Search in Social Tagging Networks

Ralf Schenkel (Universität des Saarlandes, DE)

Online communities have become popular for publishing and searching content, as well as for finding and connecting to other users. User-generated content includes, for example, personal blogs, bookmarks, and digital photos. These items can be annotated and rated by different users, and these social tags and derived user-specific scores can be leveraged for searching relevant content and discovering subjectively interesting items. Moreover, the relationships among users can also be taken into consideration for ranking search results, the intuition being that you trust the recommendations of your close friends more than those of your casual acquaintances.

Queries for tag or keyword combinations that compute and rank the top-k results thus face a large variety of options that complicate the query processing and pose efficiency challenges. We present a scoring framework and a corresponding incremental top-k algorithm that addresses these issues with two-dimensional aggregations: social expansion considers the strength of relations among users, and semantic expansion considers the relatedness of different tags. Based on principles of threshold algorithms, it folds friends and related tags into the search space in an incremental on-demand manner. We additionally discuss ideas for meaningful evaluation in such dynamic environments.

Keywords: Social networks, tagging, efficiency

Human-centered information access

Nicu Sebe (University of Amsterdam, NL)

Human Centered Computing (HCC) is an emerging field that aims at bridging the existing gaps between the various disciplines involved with the design and implementation of computing systems that support people's activities. HCC aims at tightly integrating human sciences (e.g. social and cognitive) and computer science (e.g. human-computer interaction (HCI), signal processing, machine learning, and computer vision) for the design of computing systems with a human focus from beginning to end.

This focus should consider the personal, social, and cultural contexts in which such systems are deployed.

In this presentation, I discuss the existing challenges in HCC and describe what I consider to be the three main areas of interest: media production, analysis (especially retrieval issues), and interaction. In addition, I will identify the core characteristics of HCC, describe example applications, and propose a research agenda for HCC.

Evaluating information retrieval in complex tasks

Pertti Vakkari (University of Tampere, FI)

Evaluation of interactive information retrieval (IIR) systems has typically focused on outputs, neither on outcomes nor on search process. Outputs are products delivered by systems, outcomes are the benefits systems produce for the users. In evaluations outputs have been the number of relevant items retrieved. The aim of this presentation is to sketch an evaluation framework for IIR, which takes into account search process, and outcomes also to some extent.

The goal of information searching is to support actor's task performance. Information searching consists of search goals and actions for accomplishing these goals. We suppose that the goals of IR systems are to support actor's to proceed in the task performance for finding useful information for the task. These consist of the following sub-goals: support in structuring the search topic, supporting search process, and supporting evaluation of the search results.

Evaluation of IIR should focus on assessing to which extent systems support searchers in attaining these three goals, not only on the output of the systems. In the evaluation model search process is divided into five stages, structuring information needs, query formulation, querying, evaluation of results, and the use of information in results. For the first four stages exemplary search tools are introduced, and evaluation measures for these stages are proposed.

It is proposed that the evaluation concerns both to which extent the system supports attaining search goals and attaining goals of the task. In order to be able to relate search process with its output and outcome, it is necessary to analyze how process variables influence output and outcome variables. It is discussed also to which extent it is possible to evaluate the outcomes of the system.

Keywords: Interactive information retrieval, evaluation, output, outcome, complex tasks, evaluation criteria

Tagging and Information retrieval - Looking at the Users' Role

Christian Wolff (Universität Regensburg, DE)

Especially in the area of web-based and multimedia information, tagging platforms have become increasingly popular in recent years. Depending on system as well as media type, user contributed tags can play various roles in the information interaction and retrieval process, being used for organising one's own information items as well as for information sharing purposes. In the talk, a model of information behavior in tagging systems is given along with some data on typical tag usage and the relationship between tags and document content. In a broader perspective, I want to discuss some assumptions on future user needs in the context of "digital totality": Given that all media are digitally available, how can the average user contribute towards interactive information retrieval?

Literature

- Heckner, M., Mühlbacher, S., & Wolff, C. (2008). Tagging Tagging. Analysing user keywords in scientific bibliography management systems. *Journal of Digital Information (JoDI)*, 9(27); <http://journals.tdl.org/jodi/article/view/246/208>
- Heckner, M., Neubauer, T., & Wolff, C. (2008). Tree, funny, to_read, google:
- What are Tags Supposed to Achieve? A Comparative Analysis of User Keywords for Different Digital Resource Types Proc. ACM Workshop Search in Social Media, SSM '08, Held in conjunction with ACM 17th Conference on Information and Knowledge Management (CIKM), Napa Valley/CA. New York: ACM, <http://doi.acm.org/10.1145/1458583.1458589>
- Heckner, M., & Wolff, C. (2009). Towards Social Information Seeking and Interaction on the Web. In G. Hahn & R. Kuhlen (Eds.), *Information: Droge, Ware oder Commons? Wertschöpfungs- und Transformationsprozesse auf den Informationsmärkten*. Proc. 11. Internationales Symposium für Informationswissenschaft, Konstanz, April 2009. Boizenburg: vwh Verlag Werner Hülsbusch [im Druck].

Keywords: Tagging, information behavior, multimedia

Interactive Patent Retrieval

Christa Womser-Hacker (Universität Hildesheim, DE)

Patent retrieval is an important domain with very specific requirements. On the one hand, there are very experienced professionals at the overlap of information retrieval and their own discipline. They have a high expertise concerning terminology, database coverage, database content etc. Because of this knowledge, they do not trust in fully automatical systems and demand for transparency and controllability at any point of the information process.

This fact has led to a strong preference for Boolean retrieval systems for a long period. On the other hand, there are more and more other issues submitted to patent databases (e.g. competitive patent search, patentability search, patent watch) where users do not have this elaborated knowledge. At the moment, we find first steps to integrate partial match system into the intellectual property search.

I will report on a project where a patent retrieval system based on partial match models was developed and evaluated on the basis of real users. This system allows patent retrieval to benefit from the advantages of ranking technology and relevance feedback. The user interface suggests expansion terms but the user controls the modification of the query. For ranking systems, it is inherently hard to visualize how they reach their results. Our system calculates several rankings in order to compare them transparently. It displays the relation between the rankings and allows the user to explore the position of documents in all lists.

Keywords: Patent retrieval, ranking, relevance feedback, user-oriented search

Context-Adaptive Information Interfaces Based on Spreading Activation in Ontologies

Jürgen Ziegler (Universität Duisburg-Essen, DE)

Adaptive user interfaces offer the potential for making information access in Websites and other information rich applications more efficient. To achieve successful adaptations, a range of context factors such as the user's interest profile, task, location and others have to be taken into account in an integrated fashion. We present an approach that represents domain entities and external context factors in an integrated ontological model. A spreading activation technique is used to determine the context-dependent relevance of information items and to enable system learning. The results of the spreading activation-based reasoning process can be used to perform various adaptations such as content or service selection as well as changes in the navigation structure or presentation. A prototype system called SPREADR is presented that implements this approach in an adaptive Website.

Furthermore, we will discuss first steps towards a generalized framework for context in interactive and collaborative applications.

Keywords: Adaptive User Interfaces, Context, Ontology, Spreading Activation