Transforming PCs into an Intelligent Interlinked Workspace

Transforming PCs into an Intelligent Interlinked Workspace

In daily office work, a plethora of documents and electronic media such as presentations, photos, etc. on your desktop. Does this sound like additional work? It is not.

The goal of the PIM project NEPOMUK is to create a Semantic Desktop, where information items are organized and retrieved according to their relatedness with each other. NEPOMUK is a European project in the area of Semantic Web Technologies, which started in January 2006 and not only involves researchers from the L3S Research Center, but also key international research and industrial players with complementing competences such as DFKI, IBM, DERI, SAP and others.

For realizing the Semantic Desktop, a semantic layer is created on top of the resources (emails, presentations, photos, etc.) on your desktop. Does this sound like additional work? It is not.

NEPOMUK provides users with a powerful machinery to extract this information, to recognize when the same information is referenced in two places and to organize and manage this additional information for search and browsing.

Intelligent search support is a special focus of the L3S team in the NEPOMUK project, building on all aspects of knowledge, information and learning. L3S projects include research, consulting, and technology transfer, where we work together with business and academic project partners not only from the region, but also from all over the world. This diversity is also reflected by a corresponding acceptance rate of 15% or less, from more than one of the most selective conferences, with about 1000 attendees held in September 2007 at L3S, Hannover.

I will also chair, together with Yoelle Maarek, the 2008 IEEE International Conference on Wide Web Conference, taking place 2009 in Madrid. This is the main conference for all Web related research, with about 1000 attendees from academia and industry each year. It is also one of the most selective conferences, with an acceptance rate of 15% or less, from more than 1000 submissions.

About the L3S Research Center
L3S is a global hub for basic and applied research on all aspects of knowledge, information and learning. L3S projects include research, consulting, and technology transfer, where we work together with business and academic project partners not only from the region, but also from all over the world. This diversity is also reflected by a corresponding acceptance rate of 15% or less, from more than one of the most selective conferences, with about 1000 attendees held in September 2007 at L3S, Hannover.

The PHAROS project at L3S is carried out in cooperation with the PHAROS project, which started in January 2007. In PHAROS, we are responsible for devising complex new algorithms for Social Search, i.e. how to improve search results for audio-visual media by taking user interactions and user-provided content into account (commonly known as Web 2.0). Using tagging data to better search songs on Last.FM or robust collaborative filtering algorithms are two examples for new algorithms.

On the other hand, integrating these algorithms into state-of-the-art search engines such as the ones from FAST, or into large online communities managed by France Telecom, are important priorities in PHAROS as well. Another example is providing advanced search capabilities for Digital Libraries in LINSearch, where we work in close cooperation with the German National Library of Science and Technology, TIB, in Hannover, and FITZ Technik in Frankfurt. This L3SPillwork Issue, focusing on Search and Digital Libraries, gives a short glimpse into these two and other related projects at L3S.

Can you tell me about some important events L3S is involved in?

Next July L3S will organize the 5th Adaptive Hypermedia and Web Applications conference, which is the main conference on personalizing hypermedia and web based systems. In September, the 2008 IEEE International Conference on Semantic Web conference, taking place 2009 in Madrid. This is the main conference for all Web related research, with about 1000 attendees from academia and industry each year. It is also one of the most selective conferences, with an acceptance rate of 15% or less, from more than 1000 submissions.

The PHAROS project at L3S is carried out in cooperation with L3S projects include research, consulting, and technology transfer, where we work together with business and academic project partners not only from the region, but also from all over the world. This diversity is also reflected by a corresponding acceptance rate of 15% or less, from more than one of the most selective conferences, with about 1000 attendees held in September 2007 at L3S, Hannover.

Contact: Ling Chen, Ph.D., lachen@L3S.de
Rebecca Pau, rebecca@L3S.de

http://www.pharos-audiovisual-search.eu
Contact: Raluca Paiu, paiu@L3S.de

http://www.pharos-audiovisual-search.eu
Contact: Dr. Claudia Neidert, cneidert@L3S.de

The goal of the PIM project NEPOMUK is to create a Semantic Desktop, where information items are organized and retrieved according to their relatedness with each other. NEPOMUK is a European project in the area of Semantic Web Technologies, which started in January 2006 and not only involves researchers from the L3S Research Center, but also key international research and industrial players with complementing competences such as DFKI, IBM, DERI, SAP and others.

For realizing the Semantic Desktop, a semantic layer is created on top of the resources (emails, presentations, photos, etc.) on your desktop. Does this sound like additional work? It is not.

NEPOMUK provides users with a powerful machinery to extract this information, to recognize when the same information is referenced in two places and to organize and manage this additional information for search and browsing.

Intelligent search support is a special focus of the L3S team in the NEPOMUK project, building on all aspects of knowledge, information and learning. L3S projects include research, consulting, and technology transfer, where we work together with business and academic project partners not only from the region, but also from all over the world. This diversity is also reflected by a corresponding acceptance rate of 15% or less, from more than one of the most selective conferences, with about 1000 attendees held in September 2007 at L3S, Hannover.

The PHAROS project at L3S is carried out in cooperation with the PHAROS project, which started in January 2007. In PHAROS, we are responsible for devising complex new algorithms for Social Search, i.e. how to improve search results for audio-visual media by taking user interactions and user-provided content into account (commonly known as Web 2.0). Using tagging data to better search songs on Last.FM or robust collaborative filtering algorithms are two examples for new algorithms.

On the other hand, integrating these algorithms into state-of-the-art search engines such as the ones from FAST, or into large online communities managed by France Telecom, are important priorities in PHAROS as well. Another example is providing advanced search capabilities for Digital Libraries in LINSearch, where we work in close cooperation with the German National Library of Science and Technology, TIB, in Hannover, and FITZ Technik in Frankfurt. This L3SPillwork Issue, focusing on Search and Digital Libraries, gives a short glimpse into these two and other related projects at L3S.

Can you tell me about some important events L3S is involved in?

Next July L3S will organize the 5th Adaptive Hypermedia and Web Applications conference, which is the main conference on personalizing hypermedia and web based systems. In September, the 2008 IEEE International Conference on Semantic Web conference, taking place 2009 in Madrid. This is the main conference for all Web related research, with about 1000 attendees from academia and industry each year. It is also one of the most selective conferences, with an acceptance rate of 15% or less, from more than 1000 submissions.
**The Future of Search**

Prof. Ricardo Baeza-Yates is Vice President of Yahoo! Research for Europe and Latin America, leading the labs at Barcelona, Spain and Santiago, Chile. Until 2005 he was the director of the Center for Web Research at the Department of Computer Science of the Engineering School of the University of Chile; and ICREA Professor at the Dept. of Technology of Univ. Pompeu Fabra in Barcelona, Spain. He maintains ties with both mentioned universities as a part-time professor. His research interests includes algorithms and data structures, information retrieval, web mining, text and multimedia databases, software and database visualization, and user interfaces.

Yahoo! was founded 1994 as a directory of websites. How would you describe the evolution of Yahoo! from 1994 to 2007? Yahoo! has evolved from a directory to “The Place to Be” in the Web, ranging from a plethora of services for different needs (search, finance, entertainment, etc.) to a communication space (m ail, chat, etc.) and a collaborative environment (Yahoo! Answers, Flickr, Del.icio.us, etc.).

Today, search engines play an important role for every user of the World Wide Web. What do users expect from “their” search engine?

Users expect relevant answers. However, the main problem is that queries range from generic to specific, and from informational to navigational or transactional. Hence, in the short term, search engines will have to predict the intent of the search and get the right answer based on contextual features. For example if the query is navigational the search engine could redirect the user to the most likely answer without showing results.

Web spam (text, links or clicks) constitute a steadily increasing problem. Can companies and end users expect practical solutions on this issue?

Yes, users can expect better anti-spam solution, especially as the technology for collaborative filtering improves each day. At the same time, more usage data allows spam cases to surface, and spot them as data outliers, more easily.

The idea of Web 2.0 is changing passive web consumers to active contributors. How will this influence search engines?

Web 2.0 will influence search engines in many ways. Three examples: (i) we already have a Web 2.0 search tool powered by humans: Yahoo! Answers; (ii) using image tags from Flickr, we can improve image search itself and using known images to find similar images; and (iii) the increasing use of digital cameras will allow GIS systems to tag images based on their location and simple image processing, enlarging the power of image search. Put together examples (i) and (ii), and you can improve generic search.

The web no longer consists of billions of web pages, but includes an increasing amount of semi-structured and structured information, in web catalogues, online-databases and digital libraries. What strategies does Yahoo have regarding the integration of these data into its search engine?

There are several strategies to use structured data. They range from ad-hoc agreements, where the structured data is provided to us by third parties, to information extraction based on templates that can be learned from a few tagged examples. However this is not enough if we cannot recognize the intention of the query and match the associated structure. There are already examples of this such as Yahoo! shortcuts or, how Google displays a structured flight menu if you query for two airport connectors.

One main social concern today is privacy. What is your view on this issue?

The main value in the Web is trust, so we are very careful with data privacy, because trust in the Web can be lost very easily. First, we should qualify what we know from a “user”. If the person is registered, what we know, is what that person told us during the registration process. However, that can be incomplete or (partially) false. Moreover, most people do not register and then we only know an IP (usually dynamic, sometimes a proxy) and sometimes a cookie (could be shared, temporal, etc.). Hence, in most cases the concept of a “user” is very fuzzy.

Then, as user data follows a power law, the only way to help a significant fraction of people is by aggregating data of users doing similar tasks. This means, for example, that we are personalizing search at the task level, not at the user level, providing another level of privacy protection. On the other hand, it is true that the identity of a person can be revealed by only using queries: such was the case last year with the AOL query log. That is why Yahoo! has taken the strongest guidelines of the three major search engines, anonymizing and keeping logs only for 13 months.

Can you mention one of your recent results regarding Web mining?

Sure. My pet research is how to capture the implicit knowledge behind what people do in the Web. That is, capture the “wisdom of the crowds” and use that knowledge to improve usability and user satisfaction. In the last ACM KDD we showed that we can extract similar or related concepts by just using queries and clicks. We are now building the implicit folksonomy behind queries, which can be a valuable resource for many applications, starting with search. One open problem is how to validate our results, as the semantic relations that we can capture are potentially much larger than any similar resource, see ODP or WordNet.

Can you give us a vision for the future? How will web search look in 20 years?

That is a tough question as 20 years is too much time in Internet. So let’s imagine search in 10 years. Today, and in the future, people want to get tasks done; so in the future I would like to see a higher level search engine that uses all the knowledge stored explicitly and implicitly in the Web and can answer complete tasks. For example, using the experience of other people available in blogs, forums, travel sites, Yahoo! Answers, etc. to plan your holidays based on, say, your preferences, economic restrictions and available time.

Prof. Baeza-Yates, thank you very much for this interview.

**VIKEF: Information and Knowledge**

In March 2007 the L3S team successfully concluded the VIKEF project, one of the first integrat- ed Projects in the area of semantic tech- nologies funded by the European Commission. The Integrated Projects were initiated by the European Commission for realizing large-scale application-oriented research endeavors with a large impact for the research community, for related IT industries and for empowering the European information society.

The goal of the VIKEF project, which started in April 2004, was to bridge the gap between the (partly) implicit knowledge conveyed in scientific and business content resources and the explicit representation of knowledge required for semantic level multimedia extraction processing and ontology learning for product catalogues, which exploits and combines content objects, existing structure in catalogue, and results from linguistic information extraction in a synergistic way.

• ontology discovery in support of ontology construction, for supporting re-use of ontologies in ontology construction while re- using the modeling idea of the ontology engineer;

• conceptualization of knowledge bases as a powerful approach for systematically supporting knowledge (in the form of RDF statements) that has been created in different set- tings or with a different background (opinion, state of knowledge);

• approaches to systematic entity manage- ment, which has led to the successful acquisi- tion of a follow up project in this area con- tributing to realizing the “Web of Entities” (OKKAM, see http://www.okkam.org);

• the concept of semantic inference, which feeds extracted, integrated and enriched semi- nematic information back into content objects in a task specific way — easing the interpreta- tion and digestion of such content objects (e.g. scientific publications);

VIKEF components were integrated into an cus- tomizable annotation pipeline for supporting the complete VIKEF knowledge supply chain in a flexible way, covering the way from raw con- tent over extracted information to semantically enriched content objects (annotated content). This is a core functionality for the creation of semantic information and a prerequisite for its systematic exploitation. Intelligently browsing, which has been implemented as a VIKEF showcase, is a good example for showing the potential of the VIKEF results. Intelligent browsing provides semantic navigation and combines VIK Ef components for text and image information extraction, on-
Knowledge Transfer through International Cooperation

The L3S Research Center, in cooperation with the RTS/Institute for Systems Engineering, has started an international cooperation with the Universidad Tecnológica Nacional in Buenos Aires, Argentina. Together, they are working to build distributed laboratories that operate remotely among several campuses of the Argentinean University.

In October 2007, a team of researchers from the L3S Research Center visited the Universidad Tecnológica Nacional (UTN) in Buenos Aires, Argentina under the direction of Prof. Bernardo Wagner. The purpose of this trip was to carry out two training workshops. The first workshop dealt with the topic „Remote Labs“ and took place at the computer center of the UTN. A lecture on „Making Laboratory Experiments through the Internet“ was given by Prof. Wagner, two live demos were executed in which two laboratories at L3S were operated remotely from the conference room in Argentina. In the afternoon, a short introduction of the topic „Building a Remote Lab“ was given to provide the guidelines on the practical part of the workshop. The attendees were given the task of programming a simple application to operate a mobile mini-remote lab, operated by the team at Buenos Aires. The second workshop „Using the Open Source Robotic Application Construction Kit (RACK)“ was performed at the Secretaria de Ciencia y Tecnica (Science and Technology Secretariats) of the academic staff of the university's different branch campuses. During this meeting, the real framework for robotic applications, developed at L3S and RTS of the Leibniz University of Hannover was introduced.

In addition, a presentation was given by Prof. Wagner on “Service Robots in Real Applications”. One lecturer provided a full description of the RACK architecture. Following this, a mobile robot demonstration was given. The demonstration allowed participants to become familiar with technical terms, acquire knowledge about the robot functionality and to understand the importance of using real time systems in robotics.

The last part of the first day was dedicated to hands-on interaction; allowing participants to learn configuration details and receive feedback to their questions. Finally, on the last day of the event, an assignment was given which involved compiling and executing programs written by the participants. This allowed them to use some functions of the system and to guide a mobile robot at their own. At the end, the robot was given to the branch campus of Bahia Blanca for subsequent joint work with L3S.

Along with the two events, fruitful discussions on distance education and international cooperation to promote research were held. These sessions covered technical issues, availability of resources by Latin American countries, quality of education and special language barriers. Since the official language of Argentina is Spanish, English alone is not sufficient in all cases. The language issues were also noticed by the visiting team during their stay in Argentina. Two of them gave lectures in English but sometimes, translation into Spanish was necessary, especially when the attendees posed questions. To overcome the language differences, translations were made by a participating Colombian researcher, and high motivation for further increased cooperation between UTN, L3S and Leibniz University of Hannover.

The two activities were carried out in the context of a project proposal submitted to the German Service for Academic Interchange (DAAD). The main objective of this proposal is to develop and evaluate a remote laboratory based on L3S technology. The project will be distributed among the several campuses of the Universidad Tecnológica Nacional across Argentina.

The results of the stay were very positive. Nevertheless, there is still more to do...

FacetedDBLP – Advanced Visualisation for Digital Libraries

FacetedDBLP provides a faceted search interface for the publication meta-data available on "DBLP", the Digital Bibliography & Library Project (http://dblp.uni-trier.de).

Users of Digital Libraries can have a broad spectrum of objectives when they access a document collection. For example, some users search for a specific document, others search for relevant documents about a specific topic, and still others look for new documents for a specific topic. Furthermore, some users are also interested in summarizing the results, and not necessarily in single documents. This is typically the case when users want to learn about related topics, people devoted to work on a topic, or relevant publishing venues for a particular topic.

In principal, Digital Libraries can provide all such information; however, in reality, current user interface technology makes it difficult to actually satisfy the information needs of all these different users. The faceted search paradigm offers a well-founded compromise to support the diverse information needs of most of these users using a single user interface.

The L3S FacetedDBLP system provides a faceted search interface for the publication meta-data available on "DBLP". It is one of the largest collections of bibliographic metadata related to computer science publications. These facets can be used in three different ways:

1. Faceted browsing: Users can browse through the facets to discover new information,
2. Faceted filtering: Users can filter the results through the facets,
3. Faceted ranking: Users can rank the results through the facets.

Users can use any combination of these facets to filter and rank the search results.

The FacetedDBLP system can be used to issue new queries. For example, labels can be used to query all publications from a specific author or all publications published at a specific publishing venue. The main objective of this project is to develop and evaluate a remote laboratory based on L3S technology. The project will be distributed among the several campuses of the Universidad Tecnológica Nacional across Argentina.

The results of the stay were very positive. Nevertheless, there is still more to do...

The fifth edition of the successful series of Conferences on Adaptive Hypermedia and Adaptive Web-Based Systems will be organized by the L3S Research Center in the city of Hannover, Germany, from July 28 through August 1, 2008.

The Adaptive Hypermedia conferences are the major forums for the scientific exchange and presentation of research on adaptive hypermedia and adaptive web-based systems. The scope of AH2008 includes all forms of Web and Hypermedia personalization from personalized search and browsing to recommender systems to personalized E-Learning.

The AH2008 conference is a unique event, representing the cooperation of the Adaptive Hypermedia (AH) and User Modeling (UM) communities, which will combine the alternating bi-annual AH and UM conferences into one annual event. The conferences typically attract 200-250 participants and are rather competitive, with a selection rate of less than 20% for both full and short papers. Submissions are invited in the fields of Adaptive Hypermedia systems and techniques, Web personalization, adaptive information filtering and retrieval, intelligent tutoring systems, semantic Web, applications of adaptive hypermedia and empirical studies of AH and Web systems.

Contact: Prof. Dr. Bernardo Wagner, Wagner@L3S.de
http://www.rts.uni-hannover.de
Contact: Dr. Jörg Diederich, diederich@L3S.de
www.ah2008.org
Contact: Dr. Eelco Herder, herder@L3S.de
http://www.L3S.de/growbag
Contact: Prof. Dr. Bernardo Wagner, wagner@L3S.de
http://www.rts.uni-hannover.de
Contact: Dr. Jörg Diederich, diederich@L3S.de
http://dblp.L3S.de
BRICKS – Efficient access to European digital cultural assets

In spite of technologies such as digital libraries and web search engines, the retrieval of relevant information is still challenging for end users. The main reasons are that information technologies lack awareness about information sources which are locked behind heterogeneous interfaces and buried in the deep web. The aim of the BRICKS project is the development of a self-organizing digital library infrastructure which allows integrated search among all connected and hidden resources.

The availability of digital libraries does not automatically mean that the retrieval of relevant knowledge is easier than in traditional libraries. One problem users are still often confronted with is finding relevant sources information, like sources for works from the French Renaissance or books from the Middle Ages. The usage of Web search engines can be a starting point. However, a detailed search within the content of the digital library is often not possible, as the catalogues are normally hidden from the search engines. The situation is exacerbated if the content of commercial providers should be taken into account, since their content is only accessible after registration. After the identification of interesting sources, users must query each of them independently. This process is very time consuming and burdensome to the users. Additionally, users have to become familiar with different interfaces in order to gain access to the information. An integrated search among known as well as unknown content providers, would make the retrieval of relevant information much easier for them. To address this issue the BRICKS project (Building Resources for Integrat-ed Cultural Knowledge Services - IST 507457) was conceived.

The results of the project produced during the period of January 2004 through June 2007 include a modular digital library framework that uses self-organization and peer-to-peer technologies to achieve a robust decentralized infrastructure which exposes all functionalities as web services. The framework provides a huge set of services to easily build a digital library which automatically integrates itself into the BRICKS network. The standard services can be complemented with specialized functionalities, such as feature extraction for multimedia objects. For accessing the BRICKS network, generic interfaces as well as user domain specific interfaces can be used. With BRICKS, users are able to conduct integrated queries of diverse knowledge sources such as reconstructing the history of historic documents. “What we provide is a distributed, extremely reliable and inexpensive library network which guarantees access to all digital libraries, museums and comparable institutions as soon as they are online,” explains the BRICKS technical director Dr. Thomas Risse. BRICKS shields the user from the heterogeneous nature of the information sources, thus allowing content to be handled easier. The BRICKS network avoids having any central services for the infrastructure management, instead, a peer-to-peer based approach is used. Such an approach provides highly reliable digital library infrastructures. To make the technology attractive for small institutions, it was required that all costs be kept low, allowing BRICKS to be a financially feasible option for a broad range of potential adopters. Due to the self-organizing infrastructure, the need to have a costly central maintenance organization. The project results are published as open source. The BRICKS project for the preservation and expansion of European cultural heritage is an acclaimed winner among projects in the European 6th Framework Programme. BRICKS is rated by the EU as one of the most important and most innovative projects in the cultural sector.

LinSearch: Indexing and natural-language search for technical and scientific documents

Knowledge workers with a high demand on Information are using commercial information providers to easily access high quality publications. Classifications are an important method to narrow and structure the information space and index terms are describing the content in more detail. The aim of LinSearch is the development, evaluation and usage of an integrated automated system for indexing and searching technical and scientific documents for FIZ Technik and TIB Hannover.

Knowledge workers with demands for high quality information cannot completely rely on the Internet. The reasons are not only the time consuming process to retrieve relevant information, but also lower quality results or unavailability of certain publications. Commercial information providers like FIZ Technik and public libraries such as TIB Hannover are a natural and important information source for knowledge workers and knowledge intensive enterprises such as Daimler-Chrysler.

The requirements of professional users are demanding, as they request high quality and up-to-date information, quickly. Before users are able to pose their queries, it is necessary to prepare the content by indexing and classifying it. Due to the quality requirements for commercial information providers such as FIZ Technik and TIB, indexing and classification in these libraries have been done manually. In a manual classification process, an expert reads the document and assigns the appropriate classifications and index terms. Due to the increasing amount of publications, such a manual process becomes inefficient and ultimately has an impact on the overall amount of documents available for the community.

The aim of the LinSearch project is the development, evaluation and usage of an integrated system for providing automated indexing, classification and search support for technical and scientific documents within FIZ Technik and TIB Hannover libraries. This system should support the natural language of the users as well as other languages, especially German and English. The semi-automatic indexing process will support the human indexers in their job by proposing appropriate index terms, which can subsequently be edited manually. With the help of self-learning algorithms, the indexing algorithm will learn from these feedback, automatically improving its suggestions over time. The user will be assisted in the query formulation by proposing related terms. The LinSearch system will also support the user in query formulation, proposing related terms and classifications. After the query execution, the result set not only contains the description of the documents, but also related index terms and classifications, which can be used to further refine the query. The LinSearch retrieval engine combines both linguistic methods, thesaurus and machine learning algorithms to improve its query results.