**APIs Group at L3S Research Center – Your Need is What Counts**

**Rethinking Information Services**

Databases and information systems are crucial for a large variety of today's networked applications. Together with Web technology they form the major tools to manage the information flood on the Internet and provide pervasive information access anytime anywhere. The emerging techniques of Web services and the Semantic Web already address some of the challenges by trying to standardize semantic descriptions, addressing security issues and fostering interoperability and seamless service integration. However, these problems cannot be solved by information technology alone. Modeling semantics for intended goals relies on human cognition and cultural notions that will need strong interdisciplinary research. With the arrival of the APIs (Advanced Personalization for Information Systems) research group, the L3S research center now has a strong focus on these problems. L3S new Associate Research Director Dr. Wolf-Tilo Balke has a strong background in personalized information systems. After several years' work at the University of California at Berkeley, he has formed the APIs group at L3S based on funding from the German Research Foundation (DFG) within the renowned Emmy Noether excellence program.

A basic belief of the APIS group is that networked services should be accessible from a variety of client devices and offer up-to-date information. However, given the increasing variety of services, individual users will hardly be able to choose adequate services to satisfy their personal needs unless effective personalization techniques can be provided. This is because in interacting with services users are rather more interested in accomplishing some high level goals, than in the technical details of each specific service. Generally speaking, a user's notion of adequate services will strongly depend on explicitly stated preferences or semantic descriptions of the service's desired capabilities. However, also implicating knowledge like a user's history of interaction, the current intention, a user's situation, a user's personal background and general assumptions about the respective domain have to be applied. Anticipated service profiles or usage patterns allow for user-centered service behavior offering support at different stages of the interaction process. Recent research on the nature of commonly used preferences and their integration into information systems has gained broad attention. In today's systems, however, personalization capabilities are mostly restricted to simple keyword matching, subscribing to different modules or customizing features like the layout or presentation of information. The benefits of personalized web-based information services are pervasive information access with services for information filtering, collaboration and information sharing. Together with innovative wireless technologies and powerful client devices these services can help to pave the way towards unrestricted mobility. The APIS project is committed to bring together researchers from different disciplines to improve the personalization capabilities in today's information services. For more information contact Dr. Wolf-Tilo Balke (balke@l3s.de) or visit the project homepage www.l3s.de/~apis

**KnowledgeWeb hosted at L3S Research Center**

The Semantic Web is the main topic of the European Network of Excellence KnowledgeWeb, focusing on strengthening the European industry in this highly important area of current computer technology. Out of the 15 European research institutions involved the L3S Research Center is among the five main partners in KnowledgeWeb. L3S is for instance responsible for the co-ordination of the education area activities. Education is deemed the most important success factor for European industry. Therefore the project emphasizes on spreading knowledge on Semantic Web technologies to current employees in industry as well as to future employees, making the L3S an important factor for information and communication technologies in the region.

Established in 2001 as a national and international center of excellence, the L3S Research Center today brings together research groups from several universities from four German states. We have more than 70 women and men working for the L3S, half of them are Post-docs or PhD students, and a rising number of international researchers among them.

Of our four main research areas eLearning, Semantic Web and Digital Libraries, Industrial Informatics and Mobile / Distributed Computing and Networks we focus in this edition of L3S @work on “Semantic Web and Digital Libraries”.

What are some of the key-projects of L3S in the field of “Semantic Web” and “Digital Libraries”?

There is still a long way to go before the technology of a Semantic Web can transform from a scientific adventure to commercialized technology. The main goal of the NoD Knowledge Web is the support of the current transition process of ontology technology from the world of Academia to the world of Industry. One of the main objectives of Knowledge Web is to establish a Virtual Institute for Semantic Web Education (VISWEB).

The objective of REVERSE is to establish Europe as a leader in reasoning languages for the Web which is an emerging technology that does not exist today. REVERSE will establish itself as the world leading virtual research centre on reasoning languages and methods for the Web. REVERSE will ensure that this novel technology is fully exploited to the advantage of the European industry.

The Network of Excellence PROLEARN (Professional Learning) focuses on two key issues for future eLearning scenarios and contents, namely future technology enhanced professional learning resources and the use of these learning resources for professional training in SME’s and larger companies.

The L3S Research Center with a strong scientific track record in technology enhanced learning and Semantic Web technology also strengthens its research in the field of digital libraries. With ELEENET the L3S will validate its research in a specific field of the eLearning value chain in the general context of digital libraries and improving existing bibliographic technologies.

**General Assembly of the International Network of Excellence**

**What is the L3S Research Center?**

The L3S is a global hub for theoretical and applied research on information technology and knowledge technologies. L3S projects include research, consulting, and technology transfer. These activities as well as the increasing number of network partners from business and industry make the L3S an important factor for information and communication technologies in the region.

On January 24 and 25th 2005, L3S hosted the general assembly of KnowledgeWeb, bringing together over 60 renowned researchers from all over Europe. The major objectives of this meeting were to strengthen the cooperation among the different European research work packages. For this purpose, the agenda focused on “Semantic Web Services”1. Semantic Web Services are one of the major topics in Semantic Web research. They provide machine-readable description of the Web Service, so that an automatic combination and interaction among different Web Services becomes possible.

During the General Assembly results of the first year’s work of the project were presented. Also, it was discussed concretely how the results of the different work packages can be integrated with regard to this single highly important topic.

The workshop was a great success and yet another important step towards realizing the next generation of the worldwide web.
Semantic Grid and Grid Technologies have attracted a lot of attention over the past few years. What are the main ideas behind these areas?

I guess you could say that both the Semantic Web and the Grid are about sharing, and making the most of, resources. They are both fundamentally about building infrastructure for interoperability. But they vary on what they mean by a resource, how they go about sharing, and the importance they give to different parts of the problem, and their target applications. And they are pretty separate communities too.

The Semantic Web starts from the point of view of the Web, unsurprisingly. The primary resource is a web page, and the idea is to build a semantic “web” of metadata that describes and links the contents of web pages. This metadata is used to fuel automatic processing of a machine understandable web; for example to improve the effectiveness of search engines or as a platform for agents. Ontologies play an important role to capture the shared vocabulary of a community, and to replicate the knowledge that a human uses if we are to automate the understanding of a web resource. So the emphasis is on data integration, knowledge representation and the languages needed for metadata and ontologies, reasoning about knowledge and the agents/Al stuff. It very technology-push too. Examples include gathering together web-based information about a film star, or providing the knowledge for booking a trip.

The Grid, on the other hand, is more application pull, and about distributed computing and programmatic integration. The idea is to build middleware for seamless, flexible and coordinated on-demand resource sharing, where resources are applications, computers, sensors, instruments, databases, and so on. The Grid should allow users to log in and access remote resources across organisational boundaries to form a temporary Virtual Organisation to solve a problem. The user shouldn’t need to know the nuts and bolts of where the resources are located, and the service providers don’t know in advance who might use their services. The resource configurations are transient, dynamic and volatile as a consortium of services participating in a complex analysis may be switched in and out. The configurations may be long-lived, for example a protein folding simulation could take weeks.

The underlying principles of the Grid are virtualisation and dynamic provisioning. For example, if we need to call on many computers to pool together to supply enough compute power to run a simulation, those computers provision on-demand resource sharing, where resources are applications, computers, sensors, instruments, databases, and so on. The Grid should allow users to log in and access remote resources across organisational boundaries to form a temporary Virtual Organisation to solve a problem. The user shouldn’t need to know the nuts and bolts of where the resources are located, and the service providers don’t know in advance who might use their services. The resource configurations are transient, dynamic and volatile as a consortium of services participating in a complex analysis may be switched in and out. The configurations may be long-lived, for example a protein folding simulation could take weeks.

The security policies of the different providers and ensure we are compliant. We need to discover different data sets and map between them using a controlled vocabulary; and record exactly what we did and why – a provenance trail – so we can add context to the results and make them interpretable and repeatable. To run the simulation we need to use a broker to find and reserve compute resources to run it and data resources to store the results, and we might need to look at some history of past simulations to see if a parameter needs adjusting if the simulation gets bogged down. We might need to find and substitute resources if they go down. This quick sketch has a semantic metadata that is to do with the domain and the application itself, and metadata that controls the way the middleware operates.

Is there anything that Grid can offer the Semantic Web community?

Yes. Semantic Grid is a bridge between both communities, and is not just one-way traffic! Grid developments are very pragmatic and concerned with scale, robustness, reliability and software engineering. To be honest, the knowledge management, A.I. and agent communities have up to know been more interested in “proof of concept” and elegant theory. Grid approaches have great potential to be incorporated into distributed knowledge management. For example, secure file transfer, transparent replication management, and service lifetime management.

To what extent is the Semantic Grid relevant for researchers, companies or end-users?

It has a semantic discovery using an RDFS on “in silico” experiments in bioinformatics. We are working on a whole range of Grid projects with Semantics as part of them. We lead the myGrid project (http://www.mygrid.org, uk), which has build middleware for data-oriented “in silico” experiments in bioinformatics. It has a semantic discovery using an RDFS ontology of over 1000 remote biology services over a registry, semantic support for finding and constructing workflows, provenance logging of the experiments using RDF, and mining those logs. We also partners on the Geodesia project (http://www.geodesia.org) that build an ad- visory system for aeronautical engineers using Grid-enabled Matlab optimisation simulations. We have just started an EU project, OntoGrid (http://www.ontogrid.net), which is investigating the architecture needed for a Semantic Grid; and at the other end of the spectrum, the UNIGrids project (http://www.unigrids.org) is using an ontology to mediate between two different middleware resource brokers (Globus and Unicom). Plus we have fundamental projects working on using reasoning and scheduling together and the “plumbing” end, and ontology development tools at the “application” end. So we work across the board.

What are the major challenges for bringing Semantic Web technologies into the Grid?

Both are immature. Many of the benefits of Semantic Grid will happen when reuse occurs over time and across large collaborations, and this will happen in the medium to long term. Meanwhile the Grid middleware is immature and unstable, and we need proof that the Semantic Web can scale, and deliver the quality of service and performance that Grid demands. Perhaps the biggest challenge is getting the Semantic Web and Grid communities to talk to one another!
Digital Libraries

“Digital libraries” means the merging of the classical world of libraries, with their huge amount of content and very little technology involved that has hardly changed in the last one hundred years, with the new world of new media and the Semantic Web.

A digital library is, like a traditional library, a collection of books, pictures, animations, audio- and video-recordings and reference materials. Unlike a traditional library, however, the collection of a digital library is digital, and is usually served over the World Wide Web. It is the logical extension and augmentation of physical libraries in the electronic information society. But a digital library is not only the pure collection of information objects. It is furthermore:

• the collection of services
• that support users in dealing with information objects
• and the organization and presentation of those objects
• available directly or indirectly
• via electronic/digital means.

Digital library technologies provide the opportunity for traditional libraries to open their scope far beyond their present services. More content can be offered by including external archives, contents of media like multimedia objects or primary data, and different forms of publications ranging from technical reports to complete online courses.

Furthermore digital libraries enable the creation of pro-active networks for scientific communities and their results, far beyond classical publications. These infrastructures support the knowledge transfer from science to industry creating the background for new innovative services in a connected eScience context. The promise of digital information organization implies the possibility of disseminating materials and information far beyond what has ever been imagined. Digital libraries constitute a relatively young scientific field, whose life spans roughly the last ten years, it represents the merging point of a large number of disciplines and fields, i.e. data management, information retrieval, library science, document management, information systems, the web, image processing, artificial intelligence, human-computer interaction and others. The research field digital libraries provides a huge toolbox for modern Semantic Web technologies and has established itself as a very fast growing discipline.

The German National Library of Science and Technology (TIB)

The L3S Research Center as one of Germany’s foremost research institutes in the field of learning technologies is proud to present the German National Library of Science and Technology (TIB) as an associate partner for research and development in the field of digital libraries. The TIB is the German National Library for all areas of engineering and its basic sciences, in particular architecture, chemistry, information technology, mathematics and physics.

As part of the forward-looking expansion of its digital library, on the one hand the TIB is constantly increasing its collection of electronic publications and on the other the TIB is developing new services to fulfill the requirements of their customers in industry, research and education.

In detail the TIB is offering and developing services respectively:

• TIBORDER: an integrated search and order system – vandalism has been started
• Subject gateway for technology, physics and wood technology: an integrated information infrastructure where content and service can

Further information:

http://www.l3s.de/eleonet
Contact: irina.sens@tib.uni-hannover.de
Jan Braun, E-mail: braun@tib.de

References:

The ELEONET project aims at extending the scope of the DOI to European eLearning environment. The DOI (Digital Object Identifier) is the new identification standard for any Intellectual Property entity launched by the International DOI Foundation.

A DOI can be resolved in every web browser worldwide using the Handle system from the Cooperation for National Research Initiatives (CNRI). The Handle system is a free Java-based comprehensive system for identifying, managing, and resolving persistent identifiers, known as “handles,” for digital objects and other resources on the Internet.

The ELEONET project aims at implementing a specific DOI application profile for Learning Objects (LO), using existing schemas implemented at European and international level. The main service that is provided is the registration of DOIs to learning objects and is based on the existing technologies in order to achieve interoperability with the international e-learning community.

LOs rights holders will be allowed to assign DOIs to their content, and to register metadata onto any content, according to a schema that will be interoperable with the existing metadata standards. As an effect of such registration, educational communities and any citizen for his proper learning purpose will be facilitated in the search and access of European learning objects. Specific search engines will be implemented to enhance this functionality of the system, with and without Internet interface.

Within the eLearning value chain, the ELEONET will thus offer value both to LOs producers and to educational communities, facilitating the interoperability of IT systems of any kind of organisation involved in the process. A special attention is paid to intermediaries between LOs rights holders and educational communities, such as DRM providers and clearing houses, since intermediaries are the most interested organisations in the standardisation of the identification systems.

The international partners for ELEONET are:

• Italian Publishers Association (AIE),
• Nielsen BookData (UK), Indire (Italy)
• EDITRIM (Spain) from the publishing community
• L3S and Cineca (Italy) as research agencies

The project is founded by the European community until the end of 2006.

Scientific Primary Data

In its 2004 report „Data and information“, the International Council for Science (ICSU) strongly recommended a new strategic framework for scientific data and information.

In principle, scientists are prepared to provide data, but for the time being it is unusual to appreciate the necessary extra work for processing, context documentation and quality assurance. The classical mode of distributing scientific results is their publication in professional journals. These articles in journals are recorded in the “citation index”. The index is provided to access all electronic documents as well as print documents.

Registration agency for scientific primary data

ABT in all, the TIB currently handles about 2,000 orders a day. TIB comprises approximately 7 million volumes of books, microforms and CD-ROMs, as well as about 18,000 subscriptions to general periodicals and specialist journals. The stock includes conference reports, research reports, patents in science, norms and standards, and dissertations. Furthermore the TIB holds 8 millions patents (most of them on CD-ROM). The L3S and the TIB have cooperated in the context of the project “Publication and Citation of Scientific Primary Data” funded by the German research foundation (DFG). For this project the German TIB has become the first registration agency worldwide for scientific primary data. This project is a classical example for the usage of modern digital library techniques to achieve yet unreached goals of the classical library community. Furthermore the L3S is currently developing a new cataloguing scheme for the online catalogue of the TIB.

ELEONET

In 2007, the TIB now is the world’s first registration agency for scientific primary data. By this, scientific primary data is now able to cite the data in his work using its DOI, re-ferring to the uniqueness and own identity of the original data.

If a scientist reads a publication where the registered data is used, he might be interested in analysing the data under different aspects. After gaining permission to do so by the research institution maintaining the data, he can cite the data in his own publications using its DOI, re-ferring to the uniqueness and own identity of the original data.

If furthermore a scientist is interested in certain data, he can use the online library catalogue of the TIB to search for scientific primary data. In cooperation with several major German research institutes for earth science, the TIB will provide the world’s first registration agency for primary data in the field of earth sciences.

We have registered 100,000 datasets so far (March 2005). We expect an amount of approximately 1,500,000 datasets to be registered by TIB until the end of 2007. The registration of primary data will be widened to other science fields in 2006.

Publication and Citation of Scientific Primary Data
iSearch – Satisfying Your Information Need

iSearch. Do you? Are you always satisfied with your search results? With iSearch, L3S Research Center comes to seek for improvements for nowadays Internet Search services to better satisfy everyday user information needs. We focus on three specific target areas: The World Wide Web, Peer2Peer Networks, and more recently, Personal Computers.

iSearch on the Web. Our main goal on the Web is to propose new solutions for personalizing search. We usually achieve this by means of an improved ranking algorithm built on top of a regular search solution. The improvements are either straightforward, or realized through various innovations, which make use of the large scale semantic annotations currently existing on the Web, of the social network information collected for the targeted application, etc. Final output results can then be further enhanced with algorithms for finding related pages, which we are also investigating.

iSearch in a Peer2Peer Environment. We cover most of the research areas related to Peer2Peer networks. The classical search was illustrated in the Edutella project, where we were the first ones to bring together the areas of Semantic Web and Peer2Peer, and is now continued within iSearch with our Top-K Retrieval research. We are also interested in introducing personalization into the field, both for the search itself and for other applications, such as trust and reputation algorithms. Finally, we deal with continuous querying environments, also known as publish/subscribe systems.

iSearch on the PC Desktop. We are currently opening up towards the very promising area of Personal Computers. The on-going research is about introducing context-based semantics into the desktop search algorithms, but we also investigate other subjects, such as semantic social networks built on top of the PC desktop.

You will hear from us when it comes to Information Search innovations!

References:
http://www.l3s.de/isearch

Contact:
Paul-Alexandru Chritá
E-mail: chritá@l3s.de

Grid Computing – bringing contradictions together

EScience needs tremendous computing capacity for applications such as high-energy and astro physics or molecular biology. The demand for growing peak performance and widely distributed resources coupled with the availability of network speeds of multiple gigabit per second between research facilities leads to the creation of computing Grids. These Grids are basically computing clusters that may consist of supercomputers or simply hundreds of common PCs. The aim is to let researchers use the Grid transparently, means that without having to know where their data have to be transferred to or in which system they have to log into to get their data processed.

This paradigm of Grid Computing leads to completely new challenges in network, system and data security, because resources have to be tightly coupled across institutional and network boundaries. Systems have to be accessible from outside the institutional network, foreign users must be authenticated and authorised to use resources without being known beforehand and large amounts of data have to be transferred between institutions with strong protection against unauthorised access or alteration.

So how do we achieve these goals? A key role plays the AAAI, the Authentication, Authorisation and Accounting Infrastructure which has to be implemented across all involved institutions. The first “A” makes sure that the identity of the Grid user can be verified. Grids achieve this through the use of digital certificates, which are issued and signed by certification authorities. The resource the user wants to employ for his data processing cryptographically verifies the origin of these certificates. Now the identity of the user is proven, but what resources is he allowed to use? This decision reaches the second “A” in our AAAI, the authorisation part. It knows which computing, storage and possibly network equipment the researcher may use. The third “A”, the accounting part keeps track of the amount of resources the user has consumed, possibly limits the usage or even charges him through a billing component.

While AAAI deals with the legitimate usage of resources, what if bad guys want to abuse the Grid for their needs? This has to be prevented through additional measures which include well known actions that should be taken in every networked system, such as using strong passwords, applying security patches immediately and limiting access through firewalls. The former two measures work almost unchanged within Grids, the latter requires vast modifications because of the increased communication and performance demands of Grid applications.

Researchers at L3S and the Regional Computing Center for Lower Saxony (RRZK) together develop new strategies towards secure Grid environments. The projects include novel extensions of existing approaches to AAAI as well as performance and flexibility improvements of firewalls.

Reference:
http://www.rezg.uni-hannover.de/grid.html

Contact:
Christian Grimm, Stefan Piger
(grimm, piger)@rvs.uni-hannover.de