Newsletter
Issue 1 : April 2004

Contents of Issue 1

[PDF version of this issue]

Feature article

Cluster Reports

All issues of the Newsletter will carry a report from each of the seven clusters operating within the DELOS Network of Excellence. These reports will seek to keep you informed of the developments being made by groups within the cluster and keep you up to date with current interests and the direction in which research and implementation work is proceeding.

The seven different clusters reporting are as follows:

Update

The latest news from round DELOS

Each issue of the Newsletter will carry the most recent news items from the DELOS website. The full listing will grow over time.

Return to top

Publication Date: April 2004
File last modified: Wednesday, 27-Oct-2004 18:33:18 CEST

The Delos Newsletter is published by the DELOS Network of Excellence
and is edited by Richard Waller of UKOLN, University of Bath, UK.

Newsletter
Issue 1 : April 2004

DELOS: A Network of Excellence on Digital Libraries

Costantino Thanos introduces the new DELOS Network of Excellence on Digital Libraries and describes it in terms of its objectives, activities and expected results.

Introduction

Digital libraries represent a new infrastructure and environment that have been brought about by a number of factors, principally the integration and use of a number of Information and Communication technologies, the availability of digital content on a global scale and a strong demand from users who are now online. There are destined to become an essential part of the information infrastructure of the 21st Century [1].

In Europe, partly stimulated by US activity, the digital libraries field started to emerge as a distinct area of research in the middle of the 1990s with the funding of some important national initiatives, (for example the eLib Programme in the UK and the Medoc Project in Germany). In addition, the role played by the 5th Framework Programme of the European Commission in the emergence of digital libraries as a research discipline has been particularly important as it has funded a large number of European digital library projects. In particular, the EC recognized the need to stimulate the creation of an integrated European digital library research community and, for this reason, from 1997 on, supported first a working group and then a fully fledged 5FP Network of Excellence on Digital Libraries: DELOS. DELOS has had considerable success in stimulating European research activity and promoting the building up of expertise in DL-related fields in order to maintain European R&D at a globally competitive level in this important area.

Objectives

The DELOS 10-year grand vision is that "Digital Libraries should enable any citizen to access all human knowledge any time and anywhere, in a user-friendly, multi-modal, efficient and effective way, by overcoming barriers of distance, language, and culture and by using multiple Internet-connected devices" [2]. However, the new generation of digital libraries should not just be regarded as mere repositories of static information. Rather they should be regarded as the initial nuclei of what, at a future stage, will constitute a substantial part of human knowledge, which will depend as much on communication as on information, and which will grow in an integrated fashion and be used both interactively and collaboratively.

In recent years, a large number of digital library systems have been developed. Typically, however, each system is built from scratch, develops its own techniques, focuses on a specific type of information or services and addresses the needs of a specific application domain. In the light of these initial experiences, it has become clear that the future of digital libraries goes beyond what these initial efforts have individually achieved [3].

DELOS aims to develop generic digital library technology to be incorporated into industrial-strength Digital Library Management Systems (DLMSs), offering advanced functionality through reliable and extensible services.

The main objectives of 6FP DELOS are, thus, to:

- define unifying and comprehensive theories and frameworks over the life cycle of digital library information,

and
- build interoperable multimodal/multilingual services and integrated content management ranging from the personal to the global for the specialist and the general population.

6FP DELOS will build on top of the achievements of the 5FP DELOS NoE and will pursue the objective of structuring the European research space in the field of digital libraries by defining and conducting a joint programme of activities which will integrate the current research activities of a large number of active European research teams. We expect that the proposed programme of activities will have a strong and durable structural impact upon the European digital library research community and contribute to the growth of knowledge in the digital library domain.

Specific technical objectives of the DELOS NoE are to:

1. develop a common architectural digital library infrastructure that can be customized to meet the requirements of different sectors and applications;
2. provide a common foundation for several forms of information seeking, searching and querying in digital libraries, so that it becomes possible for all resources to be provided by digital library systems in a cohesive way;
3. establish a sound framework for defining and managing unconventional information handling processes that are critical to many applications;
4. provide both a theoretical framework and practical technologies for supporting comprehensively personalized, profile-based access to diverse forms of information in digital libraries;
5. establish and demonstrate integrated methodologies from different scientific fields for capturing the meaning of audio-visual data in different domains of knowledge and contexts, building on the frameworks of established or emerging multimedia standards;
6. manage multimedia data integration and highly effective retrieval, taking into account domain knowledge and context, as well as community-related information and interactions;
7. establish a theoretically motivated and empirically supported frame of reference for designers and researchers in the field of user interfaces and visualization techniques;
8. develop mechanisms for the preservation of complex and dynamic objects;
9. provide a theoretical and practical framework for the evaluation of digital libraries and their components.

Other important objectives of DELOS are to:

- contribute towards improving the effectiveness of European research in the digital library field;
- provide a forum where researchers, practitioners, and representatives of interested applications and industries can exchange ideas and experiences;
- contribute towards improving international cooperation in digital library research areas.

Expected Results

The major milestones for the DELOS NoE are:

The DELOS portal

This will maintain and make accessible the collection of all the results and reports made available by the Network. It will also provide information on the latest research results in the field of digital libraries as well as the latest information about international projects, initiatives, conferences, etc. in the digital library domain. In addition, the website will provide tools to facilitate the exchange of information and the discussion of research topics among the Network participants.

A cross-referenced survey of all DL-related fields

An extensive survey of the technologies and the state of the art in all the DL-related fields will be carried out and made available to the research community. The different enabling technologies will be examined and cross-referenced, in order to provide indications for the main components of a Digital Library Management System.

A reference architecture of a Digital Library Management System (DLMS)

A reference architecture for a generic DLMS will jointly be defined by the Network, based on the concepts and indications provided by the survey. The main components of the DLMS will be identified and defined, having

determined the most appropriate technologies in the different DL-related fields.

A DLMS prototype and joint evaluation

An implementation of the new architecture will be carried out jointly by some of the Network participants. A limited-functionality DLMS prototype will be tested and evaluated in some specific application areas in order to identify the strengths and weaknesses of the reference architecture. The results and the refinements of the demonstration system will be made available to the research community.

Activities

The Joint Programme of Activities is organized into seven research clusters (each one is a Work Package) and is composed of three types of activities: integration, research and dissemination activities. The Work Packages and the scientific institutions coordinating their activities are as follows:

- WP1: Digital Library Architecture, coordinated by the Swiss Federal Institute of Technology and the University for Health Informatics and Technology, Tyrol;
- WP2: Information Access and Personalization, coordinated by the University of Athens;
- WP3: Audio/Visual and Non-traditional Objects, coordinated by the University of Florence and the Technical University of Crete;
- WP4: User Interfaces and Visualization, coordinated by the University of Rome 1;
- WP5: Knowledge Extraction and Semantic Interoperability, coordinated by UKOLN, University of Bath;
- WP6: Preservation, coordinated by the University of Glasgow;
- WP7: Evaluation, coordinated by the University of Duisburg.

Digital Library Architecture (DLA)

A core requirement for digital libraries is a common infrastructure. From a technical viewpoint, this infrastructure has to support state-of-the-art and promising innovative models and techniques, and has to be highly customizable, configurable and adaptive. To this end, various activities and developments have to be seamlessly integrated into a coherent whole to develop such a generic and modular digital library infrastructure. Thus, within the DLA Work Package the following activities will be carried out:

- development of surveys that collect the most significant contributions in DL architectures,
- development of prototype software modules and components and
- testing the solutions developed on an ongoing prototype application.

Information Access and Personalization (IAP)

Information stored in digital libraries needs to be accessed, integrated and individualized for any user anytime and anywhere. Consequently, within the IAP Work Package, information access in digital libraries will be studied from three different aspects:

- access of information stored in an individual source,
- integrated access of information distributed in different sources and
- access of information for different users.

Audio/Visual and Non-traditional Objects (AV)

Digital libraries will organize, store and manage large amounts of human knowledge in different application domains, for a variety of uses and communities. They will provide access to their content in different contexts and from a range of delivery channels. Consequently, the work of the AV Work Package will focus on metadata capture for audio-visual content, universal access and interaction with audio-visual libraries together with the management of audio-visual content in digital libraries.

User Interfaces and Visualization (UIV)

The notion of a "Digital Library" is currently associated with technological and scientific efforts to build, maintain, and use large collections of electronic documents. However this particular perspective brings with it a variety of problems, which will have to be solved in order to ensure the usability and accessibility of this environment to

different users with varying needs and capabilities -and both for professional and recreational purposes. The ultimate goal of the UIV Work Package is to develop methodologies, techniques and tools to enable future DL designers and developers to meet not only the technological, but also the user-oriented requirements in a balanced way.

Knowledge Extraction and Semantic Interoperability (KESI)

The KESI Work Package has two strategic goals:

- to co-ordinate a programme of activity which brings together research excellence from a range of inter-related knowledge engineering and information management areas, and which facilitates the sharing of experience and expertise amongst practitioners from both DL and Grid computing science backgrounds, and
- to explore the potential of new models, algorithms, methodologies and processes in a variety of technical applications, institutional frameworks and cross-sectorial environments, which will lead to the creation of guidelines and recommendations of best practice for dissemination to the widest possible community of interest.

Preservation

Integrated research in the Preservation Work Package will provide the methodological framework and theory for ensuring that digital library research properly addresses preservation issues and that digital libraries incorporate preservation processes in their design as a matter of course.

The research agenda in digital preservation is very broad and the objective will be to tackle a small number of aspects:

- the establishment of file format registries and the relationship between the typology of file formats and preservation methods,
- the identification of preservation functions which need to be incorporated into the digital library,
- the creation of testbeds and validation metrics to test preservation strategies,
- the enhancement of methods for evaluating material for inclusion in a digital library and for processing such material, and
- the definition of guidelines for establishing, auditing, and certifying the digital repository functions of a digital library.

Evaluation

Digital libraries need to be evaluated as systems and as services to determine how useful, usable, and economical they are, i.e. whether they achieve reasonable cost-benefit ratios. The findings of evaluation studies are expected:

- to provide strategic guidance for the design and deployment of future systems,
- to assist in determining whether digital libraries address the appropriate social, cultural, and economic problems that arise, and
- to determine whether they are as maintainable as possible.

Consistent evaluation methods will also permit comparison between systems and services. The evaluation Work Package will work both on evaluation methodologies in general as well as on providing the infrastructure for specific evaluations.

The Virtual D-Lib Competence Center

In order to coordinate the dissemination effort of the wide range of activities carried out by the Network, a Virtual D-Lib Competence Center (VDLCC) has been established. In addition to providing support for the dissemination of the Network activities , the Center can provide education, training and technology transfer to research, memory institutions and industrial organizations in the field of digital libraries. The Virtual D-Lib Center is implemented by the coordinated efforts of three institutions participating in the Network, strategically located in Europe: ISTI-CNR in Italy, UKOLN (University of Bath) in the UK, and Netlab (University of Lund) in Sweden.

The Network is managed administratively and financially by the European Research Consortium for Informatics and Mathematics (ERCIM) and scientifically by the Institute for Information Science and Technologies of the Italian

National Research Council (I STI-CNR).

Conclusions

So, by way of conclusion, what can be said to sum up this new FP6 undertaking? Firstly we would recognise that the DELOS Network of Excellence is, by any measurement within this field, a big and indeed ambitious project with a wide range of deliverables to realize. Moreover while the project brings together a large number of experienced partners, those of us who have worked in distributed communities before would readily accept that such distributed working inevitably has implications for our ability not only to communicate effectively but also to achieve the results that we have in our collective sights.

These considerations however also serve to underline the importance that we attach to the work of the VDLCC. The Center is essential to the support for the dissemination work of the DELOS Project for two different but complementary reasons. Firstly it is central to the process by which reciprocal knowledge will be made available to all members of the project right across the Network. Secondly, and equally important, it will assume a central role in raising the profile of DELOS activity and its findings throughout the rest of the digital library community around the world.

Indeed in the latter regard the VDLCC will lead the way in developing outreach activity, which will seek to involve an increasing number of representative user communities (e.g. libraries, archives, museums) as well as industrial partners in the discussions, investigations and developments that DELOS has already begun.

References

1. Excerpts from the "Technical Annex to the DELOS contract", December 2003 http://www.delos.info/workprogramme.html
2. Digital Libraries: Future Directions for a European Research Programme" DELOS Brainstorming Report, San Cassiano, Italy, June 2001 http://delos-noe .iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf
3. "Digital Libraries Information Infrastructures" DELOS Joint Working Group Report, January 2003 http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/DLinfoinfras/Infrastructures.pdf

Author Details

Costantino Thanos
Research Director at the Institute of Information Science and Technologies of the Italian National Research Council (ISTI-CNR)
Head of the "Networked Multimedia Information Systems" Lab of ISTI-CNR
Email: Costantino.Thanos@isti.cnr.it
Tel: +39 050 315 2910
Fax: +39 050 315 3464

Return to top

DELOS Home | DELOS Newsletter Front Page | Delos Newsletter Contents

Publication Date: April 2004
Publication: Delos Newsletter Issue 1
Originating URL: /newsletter/issue1/feature1/intro.html
File last modified: Wednesday, 27-Oct-2004 18:27:34 CEST

The Delos Newsletter is published by the DELOS Network of Excellence
and is edited by Richard Waller of UKOLN, University of Bath, UK.

Newsletter
Issue 1 : April 2004

Reports from the DELOS Clusters

Each issue of the DELOS Newsletter will carry a report from each of the clusters working within the DELOS Network of Excellence. In this issue clusters are reporting inital objectives and decisions made about their work.

Digital Library Architecture

Can Türker and Hans-Jörg Schek provide us with an overview of the aims and objectives of their cluster activity and explain the outcomes of their kick-off meeting in terms of their workpackage's first steps.

Introduction

The DELOS Project commenced January 2004 and so at this early stage in proceedings we intend to make use of our first Newsletter report to highlight briefly the main objectives of the workpackage and to report on the activities fixed in the kick-off meeting of this workpackage.

Objectives of our Workpackage

Citizens of the future should be able, through the medium of better designed digital libraries to gain access to a myriad of forms of knowledge from anywhere and at any time and in an efficient and user-friendly fashion. But for this to happen those digital libraries will need to arrive at a common infrastructure which is highly scalable, customizable and adaptive. Ideally, such an infrastructure would combine concepts and techniques from peer-to-peer data management, grid computing middleware, and service-oriented architectures.

Peer-to-peer architectures allow for loosely coupled integration of information services and sharing of information such as recommendations and annotations. Different aspects of peer-to-peer systems (e.g. indexes, and P2P application platforms) will need to be combined. Grid computing middleware is needed because certain services within digital libraries are complex and computationally intensive, (for example the extraction of features in multimedia documents to support content-based similarity search or for information mining in bio-medical data). The service-oriented architecture provides mechanisms to describe the semantics and usage of information services. Moreover, it supports mechanisms to combine services into workflow processes for sophisticated search and maintenance of dependencies.

The main objective of this workpackage therefore is the conceptual and experimental evaluation of the impact of these three main directions on a digital library architecture. A thorough evaluation of existing approaches will reveal the advantages and disadvantages of either approach. Moreover in order to be able to quantify that

thorough evaluation satisfactorily, we will have to develop a robust set of benchmarks.

Digital Library Architecture Cluster Activity

A joint kick-off meeting was held in Zurich over 5-6 February 2004 with colleagues working in Information Access and Personalization which meant some 24 staff in all. We felt that it was very important to take the opportunity to bring together members of both workpackages and discuss the first steps of the various tasks. After general guidelines on procedures and an overview of the project as a whole and our workpackages in particular, each colleague spoke on the work of their group and how they intended to contribute to the work of their cluster This was the principal aim of the kick-off meeting and we felt we had largely achieved it.

We then proceeded to put some flesh on the bones of a number of activities that had been outlined in the project proposal. The three main task areas within the DLA Cluster emerged refined and scheduled as follows:

- Architectures for digital libraries - UMIT
- Standards and protocols - UKOLN
- Mobile information and information dynamics - ETHZ

Outlined briefly below, the following activities are planned for the first 18 months of the workpackage:

- Organisation of workshops on digital library architectures
- Development of surveys that collect most significant contributions of peer-to-peer data management, grid computing, and service orientation for digital library architectures
- Execution of a comparison and feasibility study on the adoption of a set of common standards and protocols
- Evaluation of approaches to connection management and information synchronisation
- Development of a benchmark for the evaluation of digital library architectures
- Execution of experiments to find out the strengths and weaknesses of different architectures

With respect to the first item, it was agreed to hold the first workshop on digital library architectures over 23-24 June 2004. This workshop will take place at Cagliari (Italy) in conjunction with the Italian symposium on advanced database systems. Maristella Agosti has taken on the general and local organization of the workshop. Hans-Jürg Schek and Can Türker are co-chairing the programme committee. Meanwhile, a call for papers has been distributed through the various email channels.

This first workshop will be devoted to the architectural infrastructure of future digital libraries. The primary objectives of the workshop are to bring together European researchers interested in the architecture and related basic services that make it possible to build and operate digital libraries, and secondly, to identify the directions in which further research should go. This workshop will provide a forum for discussing the development and integration of building blocks and services for digital library infrastructures particularly in the following areas:

- Peer-to-Peer Data Management
- Grid Computing Middleware
- Service-oriented Architecture

More details about this workshop are available at http://www.dbs.ethz.ch/delos which will also hold the presentations of this workshop on its completion.

A second workshop, again a joint one between DLA and IAP, has already been fixed for 29 March - 1 April 2005 in Dagstuhl, Germany. This workshop will be organized by Gerhard Weikum, MPI Saarbrücken.

Author Details

Can Türker
Member
ETH Zurich,
Institute of Information Systems,
CH-8092 Zurich
E-mail: tuerker@inf.ethz.ch
Telephone: +41 1 6327248

Fax: +41 1 6321172
URL: http://www.dbs.ethz.ch

Hans-Jörg Schek
Leader of WP1
ETH Zurich,
Institute of Information Systems,
CH-8092 Zurich
E-mail: schek@inf.ethz.ch
Telephone: +41 1 6327241
Fax: +41 1 6321172 URL: http://www.dbs.ethz.ch

DELOS	Network of Excellence	email: schek@inf.ethz.ch
Name of Cluster:	Digital Library Architecture	phone: +30 210 727-5224
		fax: +30 210 727-5214
Name of Leader:	Hans-Jörg Schek	url: http://www.dbs.ethz.ch/

Return to top

Information Access and Personalization

Georgia Koutrika provides a general description of the Information Access and Personalization cluster, its objectives and the current activities towards the completion of these objectives.

Introduction

Information stored in digital libraries needs to be accessed, integrated and individualized for any user anytime and anywhere in possibly multiple comprehensive and efficient ways. Within DELOS, Information Access in Digital Libraries is studied from three different aspects:

Information Access: interaction with a single information provider

Information stored in a source comes in different types and formats, each one with its own characteristics and peculiarities. Organization of data within an individual source and efficient and effective search performance are the key issues and are actually very closely related to each other. Different approaches exist but there is a general trend towards richer representations and languages both at the structural and at the semantic level.

Information Integration: interaction with multiple information providers

Integrated access of different sources presents particular problems due to information heterogeneity, redundancy etc. Issues such as source selection and results fusion must be considered in a range of different settings. Data provenance is often crucial to the trust that is placed in data; hence it has to be managed on a very sound basis.

Personalization: customization of information and interaction to the user

Different users have different characteristics and preferences concerning the information in which they are interested when they access a digital library. Even users with a common information need can expect to receive different results, different functionality or a different interface. Moreover, the relevant contents and interface of a digital library may be dependent on other factors as well, for example there may be device- or network-specific issues to consider.

Cluster Objectives

The cluster's objectives with respect to the aforementioned aspects are the following:

1. Promotion of knowledge about available practices in the fields of information access and personalization in digital libraries is the first goal being pursued. This will lead to a uniform understanding of problems among

researchers.
2. Construction of a common, comprehensive framework for information access and personalization approaches is essential. This framework is intended to serve as a reference point for the DL area and to stimulate research.
3. Promotion of research on new information access and personalization models and methodologies.

Current Activities

The cluster has begun work towards the establishment of a common foundation for European researchers in all the aforementioned areas. Current activities are organized into the following tasks:

Task 1: Creation of Common Foundation for Information Access. Leader: CNR

This task seeks to create a common conceptual and infrastructural foundation in respect of Information Access, and is subdivided into three sub-tasks.

1. T1.1: Information and Interaction Models and Processing Schemes (CNR/ UPSXI)
2. T1.2: Metadata Models (IICM)
3. T1.3: Accessing time/locality information (CWI)

Task 2: Creation of Common Foundation for Information Integration. Leader: FORTH-ISL

Here the aim is to achieve a common conceptual and infrastructural foundation in respect of Information Integration which deals with multiple, heterogeneous DLs that need to be treated in a cohesive fashion. It is subdivided into three sub-tasks.

1. T2.1: Integrated Interaction Management Schemes (FORTH-ISL)
2. T2.2: Integrated Metadata (FORTH-ISL)
3. T2.3: Data Provenance (UEDIN)

Task 3: Creation of Common Foundation for Personalization. Leader: UOA

This task aims to create a common conceptual and infrastructural foundation with regard to Personalisation and Customisation of the behaviour of a DL system. It is subdivided into two sub-tasks.

1. T3.1: User Modelling for Personalization (TUC)
2. T3.2: Content, Service, and Interaction Personalization (UOA)

Important activities within the above tasks include the organisation of two thematic workshops. The first one will be on Personalization, (dates and location to be determined), while the second one will be on Information Access and Integration and will be organised in cooperation with the Digital Library Architecture Cluster in Dagstuhl over 29 March to 1 April 2005.

Author Details

Georgia Koutrika
University of Athens
Email: koutrika@di.uoa.gr
Telephone: +30 210 727 5242
Fax: +30 210 727 5214

DELOS	Network of Excellence	email: yannis@di.uoa.gr
Name of Cluster:	Information Access and Personalization	phone: +30 210 727-5224 fax: +30 210 727-5214
Name of Leader:	Yannis Ioannidis	url: http://www.di.uoa.gr/~yannis

Return to top

Audio/Visual and Non-traditional Objects

Stavros Christodoulakis describes the principal aims and activities for this cluster.

Introduction

Digital libraries will capture, organize, store and manage the access to large amounts of digital information regarding human knowledge, culture, and history in various, possibly interconnected, presentation forms like video, audio, images, etc. These types of non-traditional content will often be highly structured into segments and/or semantic units (objects) which will be indexed and interconnected with other objects in a variety of ways allowing flexible access, transcoding, browsing, semantic integration, presentation, and personalization according to the application functionality, the domain of knowledge described, the presentation device and the user's preferences and goals. The WP3 cluster entitled "Audio-visual and non-traditional objects" will focus on metadata capture for audio-visual content, universal access and interaction with audio-visual libraries, together with the management of audio-visual content in digital libraries.

Aims and Objectives

The overall objectives of this cluster are to establish a common ground of knowledge for European researchers about the state of the art, to identify the direction that research will take and realise important new applications for digital libraries with audio-visual and non-traditional objects. In particular, the cluster aims at:

- Establishing common functionalities and advancing the state of the art in the area of metadata capture from audio-visual content, including the investigation of issues related to multimodal information extraction, and the use of domain-specific, context-specific and historical information in the extraction process.
- Establishing common foundations and advancing the state of the art in the area of information access and interactions with audio-visual digital libraries exploring multimedia content standards, domain- and context-specific knowledge as well as investigating advanced interactions and interfaces to multimedia content.
- Establishing common foundations and improving performance in the area of management of audio-visual content, including new database models and data structures for storage, retrieval, and dissemination of multimedia data in emerging architectures and applications.

Cluster Activities

For the first 18 months period of the project, the foreseen activities are organized in 6 distinct tasks:

- **Cluster Management** including the setting up and maintenance of cluster website and the organization of cluster workshops for the dissemination of cluster results and the discussion and development on new proposals, solutions and achievements for audio-visual digital libraries. *(Task leaders: Technical University of Crete, Università di Firenze)*
- **Establishment of a Forum on Audio-Visual Digital Libraries** to enable international networking, communication between the cluster participants, bilateral collaboration and prototype evaluation as well as the organization of overviews and surveys of developments and activities in the area of audio-visual content. *(Task leader: University of Amsterdam)*
- **Assembly of Demonstrators and Testbeds** to demonstrate innovative solutions and applications for audio-visual digital libraries, to create domain- and context-specific testbeds and to demonstrate multi-channel delivery scenarios for audio-visual content. *(Task leaders: Technical University of Crete, Università di Firenze)*
- **Common Foundation and Joint Research on Metadata Extraction** to investigate the state of the art in the automatic extraction of audio-visual metadata and produce new models and experiments with prototype systems for automatic and semi-automatic extraction models in domain-specific and context-specific audio-visual application environments. *(Task leader: Università di Firenze)*
- **Common Foundation and Joint Research on Information Access and Interaction** to investigate the state of the art in audio-visual information access and interaction. In particular, it will explore models based on the integration of existing frameworks of multimedia content (e.g. MPEG-7, TV_Anytime) and domain-specific extensions of those models to enhance the effectiveness of retrieval and provide value-added services. It will also explore models of user interaction based on the concept of context and context-based retrieval in audio-visual digital library applications. Moreover, this task will study models and interfaces for advanced applications of audio-visual digital libraries, develop new solutions and make experiments with prototype systems for processing, semantic transcoding, interaction and presentation of audio-visual content for different delivery media and interaction devices so as to adapt the users' preferences about content to the specific characteristics of the delivery media and the users' terminals. *(Task leader: Technical University of Crete)*
- **Common Foundation and Joint Research on Management of Audio-visual Content** to study the

management aspects of audio-visual data. In particular this task will investigate the state-of-the-art database models and data structures for the storage, retrieval and dissemination of audio-visual metadata and will conduct experiments with prototype systems. It will also investigate the state-of-the-art models for profiling and stereotyping the users of the most important multimedia applications and current dissemination strategies. *(Task leader: Consiglio Nazionale delle Ricerche)*

Work in the cluster was launched by the kick-off meeting held in Paris at INRIA on 16 January 2004. During this meeting the work scheduled was specified in detail and the cluster steering committee was nominated. The cluster also nominated colleagues to be responsible for liaison with the other DELOS II clusters in order to facilitate integration within the network. We also decided that the first cluster workshop would be held in Chania, Crete during September 2004.

The A/V-NTO Cluster website http://delos.music.tuc.gr/

and the Cluster mailing list delos-wp3@ced.tuc.gr

have been established to facilitate communication between the partners and the dissemination of cluster results.

The complete list of A/V-NTO partners is also available.

The cluster is currently working on the preparation of three state of the art reports on metadata extraction, information access and interaction as well as on management of audio-visual content.

Author Details

Stavros Christodoulakis
Cluster Leader A/V-NTO
Director Lab. of Distributed Multimedia Information Systems & Applications (TUC/MUSIC)
Dept. of Electronic & Computer Engineering
Technical University of Crete
University Campus, Kounoupidiana
73100, Chania, Greece
Phone: +30 2821 037390, 37396
Fax: +30 2821 037567
Website: http://www.music.tuc.gr

DELOS	Network of Excellence	
Name of Cluster:	Audio/Visual and Non-traditional Objects	
Name of Leaders:	Stavros Christodoulakis	email: stavros@ced.tuc.gr phone: +30-2821-037399 fax: +30-2821-037399 url: http://www.music.tuc.gr
	Alberto del Bimbo	email: delbimbo@dsi.unifi.it phone: +39-055-479 6262 fax: +39-055-479 6363 url: http://viplab.dsi.unifi.it/

Return to top

User Interfaces and Visualization

Tiziana Catarci describes the aims and objectives of this cluster and puts us in the picture about work already underway in the area of user requirements relating to different groups of stakeholders.

Introduction

In current discussions on the notion of what is a digital library, it is generally associated with the technological drive to build, use, and maintain large collections of electronic documents. However, it could also be considered as

a linchpin in the construction of an information-enriched environment. It should nonetheless be recognized that this higher-level vision of a digital library, however perceptive, nonetheless brings with it a range of problems which will require solutions if we are to guarantee the usability and accessibility of this environment for the different classes of users with varying needs and capabilities. The ultimate goal of the User-Interface and Visualization cluster is to develop methodologies, techniques and tools to enable future digital library (DL) designers and developers to meet not only the technological, but also the user-oriented requirements in a balanced way. The UIV cluster will address this goal by pursuing several activities:

User Requirement-related Activities

User requirements will be studied systematically. The different perspectives on a digital library will be analyzed to relate them to the requirements and technical implementation options that emerge from the ongoing development projects being undertaken by the partners of the Network of Excellence (NoE).

It is our intention to analyze methodically all the aspects and phases pertaining to the development and the usage of a DL system. The analysis will focus not only on the DL end-user, for it will also need to take into account other DL stakeholders such as librarians, content providers and maintainers. The DL life cycle will be related to both functional and non-functional requirements.

It is also part of our plan to develop a set of character profiles of DL users and we need to bear in mind that the user interface should afford accessibility to all categories of users, including those with special needs. In addition, this cluster will also explore how users can make use of a multi-modal DL-user interface which will meet their particular needs.

For further details see the update page
UIV Ongoing Activities: Collection of User Requirements - Questionnaire Formulation.

User Interface and Visualization Design Activities

We will explore the impact and corresponding DL opportunities with respect to contextual information. The cluster will carry out studies towards developing a taxonomy of relevant context models. A language specification will then be proposed which will encompass the pertinent characteristics and requirements of context models identified during the development of the taxonomy. The cluster will then model context-dependent DL aspects.

As a consequence of taking the usage situation/context into account in our work, we find ourselves rethinking the basic assumptions underlying most of the current approaches to information filtering and retrieval. This should lead to more realistic definitions of 'relevance'. It will then be possible to move to the development of a comprehensive model for relevance criteria.

The UIV cluster will investigate the exploitation of existing visualizations as well as consider novel visualizations and how they present DL results/views; it also expects to examine certain aspects of the DL life cycle. Furthermore it will also give thought towards the possibility of extending current visualizations through the suitable application of both text and multimedia data. We will therefore seek to gauge the efficacy, clarity and degree of interactivity of visualization in the digital library context. The cluster will build a theoretical framework from which user interface designers and developers can design interfaces for DL-users. With the framework in place, it should be possible for developers to gather together the various resources provided by the theoretical framework (for example tools and methodologies) and design a DL-user interface specific to a particular application domain.

Furthermore we need to bear in mind that digital library solutions of the future will have to offer components that are both integrated and customizable and which are able to provide the necessary functionality required by the environment they serve. Ultimately our goal is to begin by developing a generic user interface which will allow us to produce a design methodology, and associated guidelines, which will help us to define appropriate technical solutions. Not only should it be possible to implement the latter in a given scenario but also to ensure that the needs of users have been met as a first priority. Therefore we aim to create an integrated DL architecture which combines both user- and application-oriented functions, for example query and navigation features, and which will adapt to the needs of its users.

Therefore, for the UIV cluster, the chief outcomes, i.e. the final integrated results of this work should be the

theoretical framework for user interface developers together with the design methodology and supporting documentation which will allow designers to create specific technical solutions. However additionally we are looking to provide profiles of the different classes of users; together with analyses of the DL life cycle, of functional and non-functional requirements and of the efficacy of visualization; as well as ensuring the development of both a relevance and a context model.

A temporary website for this cluster is available at: http://www.dis.uniroma1.it/~delos/

Author Details

Tiziana Catarci
Cluster Leader
User Interfaces and Visualization Cluster (UIV)
Università degli Studi di Roma "La Spienza"
E-mail: catarci@dis.uniroma1.it
url: http://www.dis.uniroma1.it/~catarci/
Telephone: +39-06-4991 8331
Fax: +39-06-4991 8331

DELOS	Network of Excellence	email: catarci@dis.uniroma1.it
Name of Cluster:	User Interfaces and Visualization	phone: +39-06-4991 8331
Name of Leader:	Tiziana Catarci	fax: +39-06-4991 8331 url: http://www.dis.uniroma1.it/~catarci/

Return to top

Knowledge Extraction and Semantic Interoperability

Liz Lyon describes the range of research activities being planned by the partners working in this challenging area.

Introduction

The thematic area of Semantic Interoperability is growing in importance in digital library (DL) research (taking the interpretation of "digital library" at its broadest). It applies to the application of different vocabularies and terminology used in descriptions of digital objects for both learning and research, collections of those objects, collections of datasets and resources used in the wider cultural heritage sector and in e-research. Indeed, cross-sectoral and cross-domain shared understanding of semantic descriptions is one of the goals of the Semantic Web as envisaged by Tim Berners-Lee in his "roadmap" published in 1998, (for further details, see http://www.w3.org/DesignIssues/Semantic.html). This vision has more recently (2001) been applied to "Grid computing" and e-science / e-research initiatives in the Semantic Grid approach, (see http://www.semanticgrid.org/).

In addition, the application of algorithms for the mining and analysis of digital resources (text, data, complex objects), offers exciting opportunities for the extraction of new knowledge and the re-use of data and information in new ways.

Today, we are beginning to address some of the issues and challenges in this complex area and the Delos Network of Excellence has the opportunity to carry out some important research to move the Semantic Web/Grid vision forwards towards implementation.

Aims and objectives of the cluster

The Knowledge Extraction & Semantic Interoperability research cluster has two key strategic goals:

- To co-ordinate a programme of activities which brings together research excellence from a range of inter-related knowledge engineering and information management areas, and which facilitates the sharing of experience and expertise amongst practitioners from both DL and Grid/computing science backgrounds.

- To explore the potential of new models, algorithms, methodologies and processes in a variety of technical applications, institutional frameworks and cross-sectoral environments, which will lead to the creation of guidelines and recommendations of best practice for dissemination to the widest possible community of interest.

We can examine some of the themes underlying this area in more detail.

Open Access to Digital Repositories of Data and Information

The development of digital repositories for the support of research and learning is at a critical stage. There has been a concerted effort to promote open access to the research literature with the success of the Open Archives Initiative, the development of the ePrints software from the University of Southampton, UK, the establishment of the European-focused Open Archives Forum and national initiatives such as DARE (Digital Academic Repositories). There has also been a drive to promote institutional repositories as the location for e-print deposit e.g. the DSpace project at the University of Cambridge, UK. These developments have all been made possible through the implementation of the OAI-Protocol for Metadata Harvesting within the information architectures. Digital resources published in this way may also include primary research data, experimental data generated by Grid-enabled applications, gene and protein structure data, statistical data, satellite data, census data and environmental modelling data. The current increase in Grid -enabled applications is resulting in large volumes of data being collected in data libraries and this trend is likely to continue in the future. These large datasets need to be managed, curated and made accessible to the research community.

In parallel to the development of repositories of research data and derived information, many institutions are creating learning objects for manipulation and inclusion in learning programmes and curriculum-based activities. Learning Management Systems are being deployed as vehicles for the development and distribution of online courses as part of e-learning initiatives. Repositories of learning objects are being developed, both at national and institutional level, to enable the access to and deposit of discrete learning objects for wider use by the community.

Provenance

The integrity, authenticity and value of the mass of information and knowledge derived from original data are actually dependent on a number of critical factors. For example in science, the provenance or origin of a particular set of data is essential to determining the likely accuracy, currency and validity of derived information and any assumptions, hypotheses or further work based on that information. Significant research has been carried out on describing the provenance of scientific data in molecular genetics databases and the topic has been explored in the Global Grid Forum (GGF6) in relation to Grid data. The Open Archives Initiative has carried out work to describe the provenance of harvested metadata records and the concept is included as an element in the administrative metadata which is part of the METS metadata standard. The critical factors include the definition and acceptance of appropriate frameworks for metadata description, a shared understanding of the concept of provenance, the widespread use of unique identifiers, appropriate linking technology and the application of common ontologies for discrete domains. These concepts are relatively new but have the potential for significant impact on the way in which research and learning is conducted in the future and on the ability to integrate and re-use digital resources in a variety of ways.

Semantic Web, Ontologies and Metadata Schema Registries

In order to achieve semantic interoperability between descriptions of services, collections and items, there needs to be a shared understanding of the meanings of subject terms and descriptors. Frequently, discrete subject domains have their own shared vocabulary, however specific terms may have different meanings within another subject domain. Additionally, one particular domain may have multiple vocabularies which are used by the different communities of interest. The myriad of existing vocabularies both at domain and high level is a major challenge to implementers and users of digital libraries who are trying to locate resources and services.

There is now an increasing number of developments in the broad area of Semantic Web/Grid technologies, ranging from the development of Semantic Web-enabled Web Services to the scoping of terminology servers to provide services to distributed digital libraries. There is also a growing body of work on registries and their use in the publication and validation of metadata schemas.

Knowledge Extraction

Finally, the increasing richness of both data and the descriptive metadata contained in digital libraries offers great potential for the application of a variety of tools to extract additional information to contribute to knowledge. The research community has a growing requirement for data manipulation tools to facilitate spatial change (federation, aggregation, dis-aggregation, replication, manipulation, linking, annotation, editing/versioning, transformation) and for knowledge extraction which can include analysis (textual, musical, statistical, mathematical, visual, chemical, gene), mining (text, data, structures), and modelling (economic, mathematical, biological). Taking an example, text mining techniques have been applied to resources in various domains and in particular to biomedical materials. Similarly, data mining techniques have been applied to domain datasets such as biomedical and physical data and this form of analysis is becoming increasingly important in the understanding of outputs from Grid-enabled projects and associated data repositories.

Together these themes form a rich contextual background to the research programme of this cluster.

Cluster Partners

A number of organisations and institutions are currently involved in this Work Package:

- Department of Electronics & Computer Science, University of Southampton, UK
- ETH, Swiss Federal Institute of Technology, Zurich, Switzerland
- FORTH, Crete, Greece
- Netlab Knowledge Technologies Group, Lund University, Sweden
- School of Informatics, University of Edinburgh, UK
- Technical University of Crete, Greece
- UKOLN, University of Bath, UK
- UNIMI, University of Milan, Italy
- University for Health Informatics & Technologies, Tyrol, Austria

Start-up activities

We have identified a number of activities to initiate a programme of work which is currently being explored in terms of the definitions and scope of the various themes.

A Forum is being created to provide a physical and virtual arena for the exchange of experience and research in all the areas/themes of this cluster. The first meeting of the Forum is planned to coincide with the European Conference on Digital Libraries (ECDL) to be held in September in Bath, UK. It will provide an opportunity to integrate systematically other relevant groups into the cluster and will take the format of a one-day state-of-the-art workshop. This development is being supported by a moderated virtual forum or discussion list for the expansion of discussion on selected topics. It is also intended to maximise opportunities to harmonise with other relevant initiatives such as CIDOC and FRBR. The activity will culminate with an evaluative report and a second Forum workshop to disseminate the findings of the Report.

In the area of Knowledge Extraction, initially a study will be produced to determine the requirements for and usage of extracted knowledge for biblio-metrics, domain analysis, issue tracking and community modelling.

Semantic Interoperability is being addressed initially by scoping the area with the aim of producing a state-of-the-art overview of DL semantic issues including the application of standards, thesauri, ontologies, Knowledge Organisation Systems and the implementation of metadata schema registries.

It is intended that the discussions and various reports produced will inform the future research programme for the cluster.

Further information will shortly be available on the cluster Web site.

Author Details

Dr Liz Lyon
Knowledge Extraction & Semantic Interoperability Cluster leader
Director UKOLN

University of Bath
UK
E-mail: e.lyon@ukoln.ac.uk
Tel: +44 (0) 1225 386580
Fax: +44 (0) 1225 386838

DELOS	Network of Excellence	email: e.lyon@ukoln.ac.uk
Name of Cluster:	Knowledge Extraction and Semantic Interoperability	phone: +44 1225 386580 fax: +44 1225 386838
Name of Leader:	Elizabeth Lyon	url: http://www.ukoln.ac.uk

Return to top

Preservation

Digital Libraries depend on preservation of the digital materials they contain and the ability to build successful digital libraries depends upon methodological and technical solutions. Two years ago, an international workgroup brought together by DELOS and the National Science Foundation defined a research agenda for Digital Preservation and Archiving in broad terms. Beginning 2004 the DELOS Preservation Cluster will thus aim to:

- (inter-)connect people (researchers, stakeholders, suppliers, vendors etc.), organisations and projects that can deliver this research agenda;
- coordinate and promote research and projects;
- enable identification, collection, and dissemination of information, knowledge and expertise;
- provide mechanisms for developing testbeds and metrics for assessing the effectiveness of preservation strategies, and tools for evaluating digital preservation strategies;
- create a coherent platform for proactive cooperation, collaboration, exchange and dissemination of research results and experience in the preservation of digital objects;
- relate the research agenda more directly to the development of exploitable product opportunities and develop links with the industrial sectors;
- eliminate the duplication of effort between the various research activities by creating an integrating framework;
- and ensure that the work of the digital preservation cluster has a direct impact on digital library architecture and development work of other clusters in the DELOS network.

DELOS	Network of Excellence	email: s.ross@hatii.arts.gla.ac.uk
Name of Cluster:	Preservation	phone: +44-141-330-3635 fax: +44-141-330-2793
Name of Leader:	Seamus Ross	url: http://www.hatii.arts.gla.ac.uk

Return to top

Evaluation

Norbert Fuhr gives a survey of current activity in the Evaluation cluster.

Introduction

While supporting ongoing evaluation initiatives like the INEX and CLEF campaigns, the evaluation cluster is also working towards the development of new evaluation models and methods.

CLEF

CLEF (Cross-language Evaluation Forum, http://www.clef-campaign.org/) provides test-beds for the evaluation of cross-language information retrieval. For the 2004 campaign, CLEF will consist of 8 different evaluation tracks: Mono-, Bi- and Multi-lingual Information Retrieval, Mono- and Cross-Language Information Retrieval on Structured Scientific Data, Interactive Cross-Language Information Retrieval, Multiple Language Question Answering, Cross-Language Retrieval in Image Collections, and Cross-Language Spoken Document Retrieval. Supported languages will be Dutch, English, Finnish, French, German, Italian, Portuguese, Spanish, Swedish, Russian, Japanese and

Chinese. At the moment, data and topics have been distributed to the participants, who will submit their runs by 15 May. Results will be presented at the CLEF workshop during ECDL over 16-17 September 2004 in Bath, UK.

INEX

INEX (Initiative for the evaluation of XML Retrieval, http://inex.is.informatik.uni-duisburg.de:2004/) deals with the evaluation of information retrieval methods for XML documents. The major tasks involved have been identified as retrieval for content-only queries and retrieval for queries referring both to content and structure of the target elements. During 2004 INEX will undertake 4 additional tracks dealing with relevance feedback, natural language queries, heterogeneous collections and interactive retrieval. 49 participating groups have registered for this year's INEX campaign and are currently involved in topic creation. The results for 2004 will be presented at a workshop in Schloss Dagstuhl, Germany, over 6-8 December.

Evaluation Models and Methods

This task focuses on the specification of standard evaluation methods for digital libraries, starting with a comparison and evaluation of existing evaluation methodology, and then developing new techniques, methods and measures. The first step will be a workshop on DL evaluation over 4-5 October 2004 in Padova, Italy. The workshop will concentrate upon a survey on the state of the art in digital library evaluation and will identify major issues for further research in the area of DL evaluation.

Conclusion

Most of the current activities in this cluster are targeted towards providing the necessary infrastructure for the INEX and CLEF campaigns. As a general infrastructure for DL evaluation, an evaluation forum is under development which will support communication between DL researchers and evaluation specialists. In the future, research on evaluation models and methods will be enforced in DELOS, along with the development of appropriate evaluation toolkits and test-beds.

Author Details

Norbert Fuhr
Cluster Leader Evaluation
University of Duisburg-Essen
E-mail : fuhr@uni-duisburg.de
Telephone: +492033792524
Fax: +492033792549

DELOS	Network of Excellence	email: fuhr@uni-duisburg.de
Name of Cluster:	Evaluation	phone: +49-203-379-2524
Name of Leader:	Norbert Fuhr	fax: +49-203-379-2549 url: http://is.informatik.uni-duisburg.de

Return to top

DELOS Home | DELOS Newsletter Front Page | Delos Newsletter Contents

Publication Date: April 2004
Publication: Delos Newsletter Issue 1
Originating URL: /newsletter/issue1/cluster-reports/intro.html
File last modified: Wednesday, 27-Oct-2004 18:43:26 CEST

The Delos Newsletter is published by the DELOS Network of Excellence
and is edited by Richard Waller of UKOLN, University of Bath, UK.

DELOS
Network of Excellence on Digital Libraries

Newsletter
Issue 1 : April 2004

UIV Ongoing Activities: Collection of User Requirements - Questionnaire Formulation

Tiziana Catarci provides an update on ongoing activity into the approach adopted by the User Interfaces and Visualization (UIV) Cluster to gather user requirements.

Introduction

While it is understood that user requirements are a central plank of most successful designs, some thought also has to be given to the effective collection of those requirements. As mentioned in the User Interface and Visualization Cluster report, activity on this aspect of planning and design has already begun. This update provides a brief look at the approach adopted. The general aim is to obtain different perspectives on a digital library and thus flesh out the list of options for implementation that will need to be considered.

It is our intention to analyze methodically all the aspects and phases pertaining to the development and the usage of a DL system. The analysis will focus not only on the DL end-user, for it will also need to take into account other DL stakeholders such as librarians, content providers and maintainers. The DL life cycle will be related to both functional and non-functional requirements.

Questionnaire Formulation

As mentioned above, one of the main objectives of this cluster is to collect and analyze user requirements in order to relate them to the different research perspectives and options for technical implementation within a digital library.

In this effort, various colleagues within the cluster were asked to formulate a questionnaire as it related to their task area. In particular, task leaders were assigned responsibility to deliver a questionnaire relevant to their activities, namely:

- Task 1: Provision of an Empirical Basis
- Task 2: Digital Library Life Cycle Support
- Task 3: Characterisation of Digital Library Users
- Task 4: Context Consideration and Exploitation

Based on the four questionnaires, the cluster came up with:

1. a refined and integrated questionnaire for DL end-users
2. a refined and integrated questionnaire for DL stakeholders

1. Integrated questionnaire for digital library end-users:

This questionnaire has been designed for the express purpose of gathering data from digital library end-users. The questionnaire for digital library end-users consists of four parts:

- Part 1 collects data regarding users' background and demographics;
- Part 2 solicits data pertaining to users' current experience with digital library services;

- Part 3 collects data pertaining to the functional requirements of the digital library from the perspective of end-users; and
- Part 4 contains questions for collecting data regarding the non-functional requirements of the digital library, again from the end-users' viewpoint.

2. Integrated questionnaire for digital library stakeholders:

This questionnaire is intended to solicit input from digital library stakeholders. The questionnaire has therefore been designed with content that is relevant to the stakeholders such as librarians, archaeologists, content providers, service providers, software developers, etc.

The questionnaire for the stakeholders is made up of five parts. In particular:

- Part 1 collects data pertaining to the stakeholders' background and demographics;
- Part 2 gathers data regarding the stakeholders' current experience with digital libraries;
- Part 3 collects data pertaining to the functional requirements of the digital library from the stakeholders' point of view;
- Part 4 gathers data regarding the non-functional requirements of the digital library from the stakeholders' perspective; and
- Part 5 collects data pertaining to content life cycle and the user interface of the digital library.

The integrated questionnaires have been set up for online collection and processing of data. The cluster is currently enlisting appropriate digital library end-users and stakeholders. The questionnaires can be found at http://www.dis.uniroma1.it/~delos

Author Details

Tiziana Catarci
Cluster Leader
User Interfaces and Visualization Cluster (UIV)
Università degli Studi di Roma "La Spienza"
E-mail: catarci@dis.uniroma1.it
url: http://www.dis.uniroma1.it/~catarci/
Telephone: +39-06-4991 8331
Fax: +39-06-4991 8331

Return to top

DELOS Home | DELOS Newsletter Front Page | Delos Newsletter Contents

Publication Date: April 2004
Publication: Delos Newsletter Issue 1
Originating URL: /newsletter/issue1/update1/intro.html
File last modified: Wednesday, 27-Oct-2004 18:04:48 CEST

The Delos Newsletter is published by the DELOS Network of Excellence
and is edited by Richard Waller of UKOLN, University of Bath, UK.

Newsletter
Issue 1 : April 2004

Latest News from DELOS

Each issue of the Newsletter will carry the most recent news items from the DELOS website. The full listing will grow over time.

- Digital Library Architectures: Peer-to-Peer, Grid, and Service Orientation
- INEX 2004

Digital Library Architectures: Peer-to-Peer, Grid, and Service Orientation

First International Workshop of the EU Network of Excellence DELOS on Digital Library Architectures
S. Margherita di Pula (Cagliari), Italy, 24-25 June, 2004

Call for papers

DELOS is a new interdisciplinary EU FP6 Network of Excellence with a broad vision: Future digital libraries (DLs) should enable any citizen to access human knowledge any time and anywhere, in a user-friendly, multi-modal, efficient and effective way. The main objective of the DELOS network is thus to define and conduct a joint programme of activities in order to integrate and coordinate the ongoing research activities of the research teams in the field of digital libraries for the purpose of developing the next generation DL technologies. This first workshop is devoted to the architectural infrastructure of future DLs. The objective of the workshop is to bring together researchers interested in the architecture and related basic services that allow us to build and operate DLs and to identify the direction ongoing research will take. Ideally the infrastructure combines concepts and techniques from the following fields:

- Peer-to-Peer Data Management
- Grid Middleware
- Service-oriented Architecture

Peer-to-peer (P2P) architectures allow for loosely coupled integration of information services and sharing of information such as recommendations and annotations. Different aspects of peer-to-peer systems (e.g. indexes, and P2P application platforms) must be combined. Grid computing middleware is needed because certain services within digital libraries are complex and computationally intensive (e.g., extraction of features in multimedia documents to support content-based similarity search or for information mining in bio-medical data). The service-oriented architecture (SoA) provides mechanisms to describe the semantics and usage of information services. Moreover, in a SoA we have mechanisms to combine services into workflow processes for sophisticated search and maintenance of dependencies.

It is obvious that elements of all three directions should be combined in a synthesis for future DL architectures. Therefore, a main goal of this workshop is to provide a forum for discussing the development and integration of building blocks and services for DL infrastructures from these and related areas. In the spirit of a workshop we ask for extended abstracts describing ongoing research and development.

Abstract Submission

Submissions in form of extended abstracts, not exceeding two pages based on Springer's LNCS style, should be sent in PDF to tuerker@inf.ethz.ch before April 24, 2004. Accepted papers will be published in the DELOS workshop proceedings.

Important Dates

Abstract Submission Deadline:	**April 24, 2004**
Notification of Acceptance:	**May 10, 2004**
Camera-Ready Full Version:	**June 14, 2004**

This workshop is co-located with SEBD 2004 - the 12th Italian Symposium on Advanced Database Systems.

Programme Committee

- Maristella Agosti (University of Padua, Italy)
- Elisa Bertino (University of Milano, Italy)
- Donatella Castelli (CNR-ISTI, Italy)
- Stavros Christodoulakis (Technical University of Chania, Greece)
- Wilhelm Hasselbring (OFFIS Oldenburg, Germany)
- Yannis Ioannidis (University of Athens, Greece)
- Martin Kersten (CWI Amsterdam, Netherlands)
- Liz Lyon (UKOLN, University of Bath, United Kingdom)
- Erich Neuhold (FhG Darmstadt, Germany)
- Hans-Jörg Schek (ETH Zürich, Switzerland/UMIT Innsbruck, Austria)
- Heiko Schuldt (UMIT Innsbruck, Austria)
- Can Türker (ETH Zürich, Switzerland)
- Gerhard Weikum (MPI Saarbrücken, Germany)
- Pavel Zezula (Masaryk University Brno, Czech Republic)

Workshop Organization

Maristella Agosti, Chair
University of Padua
Department of Information Engineering
Via Gradenigo 6/a
I-35131 Padova, Italy
Email: maristella.agosti@unipd.it

Hans-Jörg Schek, PC Chair
ETH Zurich and UMIT Innsbruck
Email: hans.joerg.schek@umit.at

Can Türker, Co-Chair
ETH Zurich
Email: tuerker@inf.ethz.ch

Return to top

INEX 2004

INitiative for the Evaluation of XML Retrieval

April 2004 - December 2004

Call for participation

The DELOS Network of Excellence for Digital Libraries invites participation in an evaluation initiative for XML document retrieval.

The widespread use of the extensible Markup Language (XML), especially the increasing use of XML in scientific data repositories, Digital Libraries and on the Web, brought about an explosion in the development of XML tools, including systems to store and access XML content. The aim of such retrieval systems is to exploit the logical structure of documents, which is explicitly represented by the XML markup, and retrieve document components, instead of whole documents, in response to a user query. Implementing this, more focused, retrieval paradigm means that an XML retrieval system needs not only to find relevant information in the XML documents, but also determine the appropriate level of granularity to return to the user. In addition, the relevance of a retrieved component is dependent on meeting both content and structural conditions.

Evaluating the effectiveness of XML retrieval systems, hence, requires a test collection where the relevance assessments are provided according to a relevance criterion, which takes into account the imposed structural aspects. A test collection as such has been built as a result of two rounds of the Initiative for the Evaluation of XML Retrieval (INEX 2002 and INEX 2003). This initiative provides an opportunity for participants to evaluate their XML retrieval methods using uniform scoring procedures and a forum for participating organizations to compare their results. As part of a large-scale effort to improve the efficiency of research in information retrieval and digital libraries, this project initiated an international, coordinated effort to promote evaluation procedures for content-oriented XML retrieval.

In INEX 2004, participating organizations will be able to compare the retrieval effectiveness of their XML document retrieval systems and will contribute to the continuous construction of a large XML test collection. The test collection will also provide participants a means for future comparative and quantitative experiments. Due to copyright issues, only participating organizations will have access to the constructed test collection.

INEX test collection

The test collection consists of a set of XML documents, topics and relevance assessments. The topics and the relevance judgments are obtained through a collaborative effort from the participants. Detailed guidelines on the on-line topic submission, retrieval result submission, relevance assessment task, and evaluation metrics will be provided by INEX.

Documents

The INEX document collection is so far made up of the full texts, marked up in XML, of 12,107 articles of the IEEE Computer Society's publications from 12 magazines and 6 transactions, covering the period 1995-2002, and totalling 494 megabytes in size. The collection has a suitably complex XML structure (192 different content models in DTD) and contains scientific articles of varying length. On average an article contains 1,532 XML nodes, where the average depth of a node is 6.9.

Topics

Each participating group will be asked to create a set of candidate topics, which are representative of the range of real user needs over the XML collection. The queries may be content-only (CO) or content-and-structure (CAS) queries, and broad or narrow topic queries. CO queries are free-text queries, like those used in TREC, for which the retrieval system should retrieve relevant XML elements of varying granularity, while CAS queries contain explicit structural constraints, such as containment conditions. From the pooled set of candidate topics INEX will select a final set of topics to form part of the INEX test collection

Tasks

The general task, to be performed with the data and the final set of topics, will be the ad-hoc retrieval of XML documents. Similarly to information retrieval, we regard ad-hoc retrieval as a simulation of how a library might be used, where a static set of documents is searched using a new set of queries (topics). The main differences are that, in INEX, the library consists of XML documents, the queries may contain both content and structural conditions and, in response to a query, arbitrary XML elements may be retrieved from the library. Participants will be able to submit up to a fixed number of runs, each containing the top 1500 retrieval results for each of the selected topics.

INEX will have four tracks this year:

1. Relevance feedback track
2. Natural language track
3. Heterogenous collection track
4. Interactive track

Relevance assessments

Relevance assessments will be provided by the participating groups using INEX's on-line assessment system. Assessors will judge 1-2 topics, either the topics that they originally created or if these were removed from the final set of topics, then topics that were similar to their original queries. Please note that assessments will take about one person week per topic. Participating groups will gain access to the completed INEX test collection only after they have completed their assessment task.

Evaluation

The evaluation of the retrieval effectiveness of the XML retrieval engines used by the participants will be based on the constructed INEX test collection and uniform scoring techniques, including recall/precision measures, which will take into account the structural nature of XML documents, and overlap of answers.

Participants will be able to present their approaches and final results at the INEX 2004 workshop to be held in December 2004 in Dagstul. All results will be published in the INEX workshop proceedings and on the Web.

Data Handling Agreement

In order to have access to the data designated as the IEEE Computer Society XML Retrieval Research Collection, organizations (which did not sign the agreement in 2003) must first fill in a data release Application Form (to be obtained from the INEX 2004 website).

Schedule

April 2: Deadline for the submission of "Application for Participation".

April 02 - 16: The collection of XML documents will be distributed to all participants on the receipt of their signed data handling agreement. Participants will also be provided with detailed instructions and formatting criteria for candidate topics/queries.

May 03: Submission deadline for candidate topics.

May 24: Distribution of final set of topics/queries to participants along with detailed information on the formatting requirements of the search results.

August 09: Submission deadline of search results.

August 23: Distribution of merged results to participants for relevance assessments.

October 08: Submission deadline for relevance assessments.

November 01: Distribution of XML test collection and evaluation scores to participants.

December 1 (tbc): Submission of papers for the workshop pre-proceedings

December 13-15 (tbc): Workshop in Schloss Dagstuhl (http://www.dagstuhl.de/).

Organisers

Project Leaders

Norbert Fuhr
University of Duisburg-Essen
Email: fuhr@uni-duisburg.de

Mounia Lalmas
Queen Mary University of London
Email: mounia@dcs.qmul.ac.uk

Contact person

Saadia Malik
University of Duisburg-Essen
Email: malik@is.informatik.uni-duisburg.de

Topic format specification

Börkur Sigurbjörnsson
University of Amsterdam
Email: borkur@science.uva.nl

Andrew Trotman
University of Otago
Email: andrew@cs.otago.ac.nz

Online relevance assessment tool

Benjamin Piwowarski
Université Paris 6, France
Email: Benjamin.Piwowarski@lip6.fr

Metrics

Gabriella Kazai
Queen Mary University of London
Email: gabs@dcs.qmul.ac.uk

Arjen P. de Vries
CWI, The Netherlands
Email: Arjen.de.Vries@cwi.nl

Return to top

DELOS Home | DELOS Newsletter Front Page | Delos Newsletter Contents

Publication Date: April 2004
Publication: Delos Newsletter Issue 1
Originating URL: /newsletter/issue1/news/intro.html
File last modified: Wednesday, 27-Oct-2004 21:04:48 CEST

The Delos Newsletter is published by the DELOS Network of Excellence
and is edited by Richard Waller of UKOLN, University of Bath, UK.