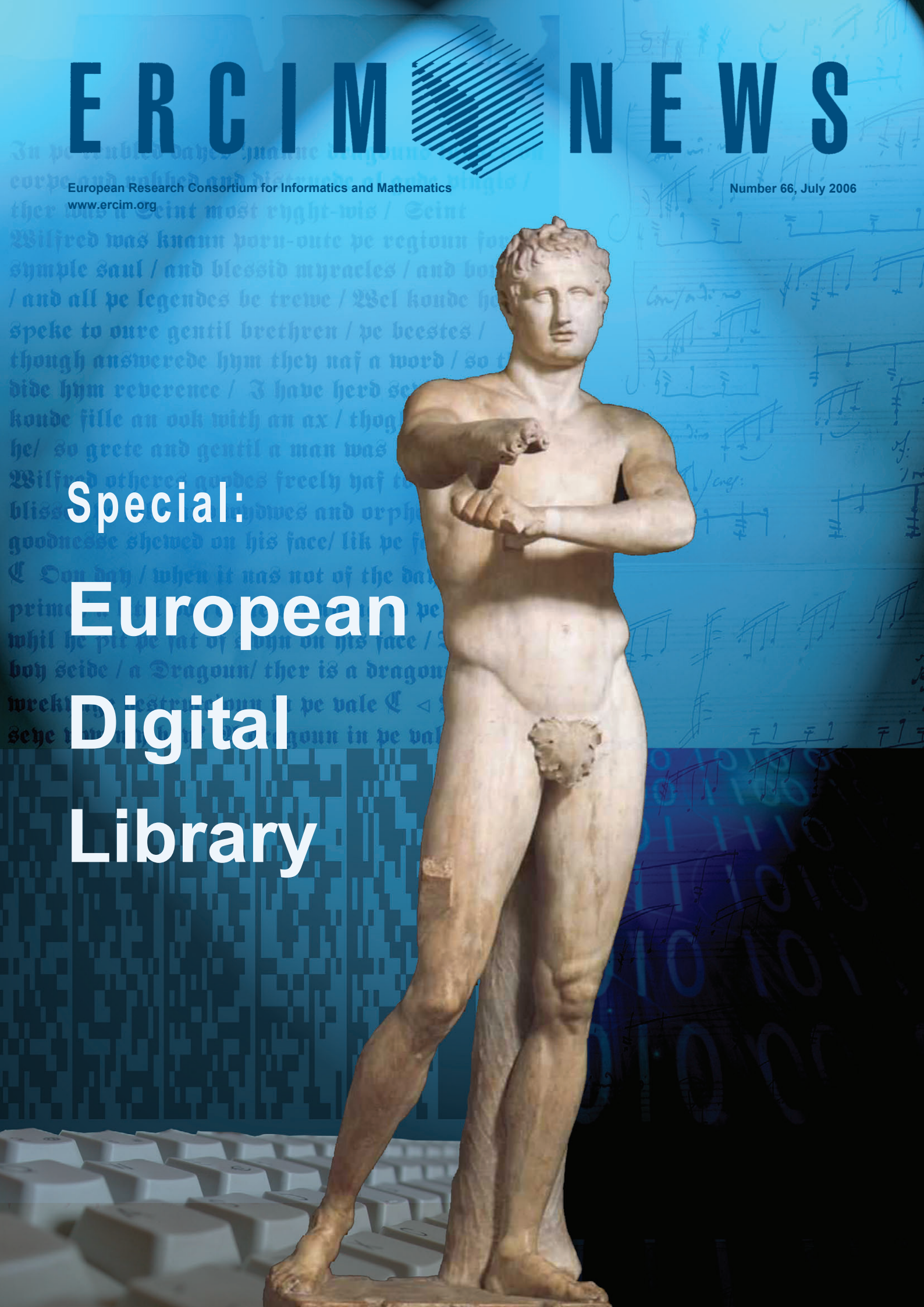
ERCIM NEWS

European Research Consortium for Informatics and Mathematics
www.ercim.org

Number 66, July 2006



Special:

European Digital Library

CONTENTS

Next issue: October 2006 — Special theme: Embedded Intelligence

Building Europe's Digital Library

Horst Forster, Director Content, Directorate General Information Society and Media, European Commission

Earlier this year, the Commission unveiled a roadmap that will see to the realization by 2010 of a distributed European library presence (also involving archives and museums). This roadmap sets clear targets: full cooperation between Europe's national libraries by the end of this year, digitized objects increasing from 2 million in 2008 to 6 million in 2010, and support for multilingual access. It builds on The European Library, a partnership between national libraries that provides an organizational nucleus for further developments.

Large-scale digitization requires that Member States and institutions ramp up their efforts, as well as making improvements in the efficiency and sophistication of processing and indexing methods for digitized data.

Most current digitization initiatives concentrate on material in the public domain. This creates the risk of fostering a 20th century black hole in our organised knowledge: in a world where increasingly the Web is the sole source of information, we would banish 20th century works to digital oblivion. For the audiovisual world, which was essentially born in the 20th century, this is particularly critical. Accepted ways must therefore be found that will ensure digitized copyrighted and orphan materials are accessible on the Internet. The High-Level Group on Digital Libraries set up by Commissioner Reding started its work earlier this year by looking into this subject. There is also substantial research work to be carried out.

What are the research issues to be addressed in this area, and how can the research community contribute? Over the past four years, the European Commission has already invested more than 100 M€ in funding for digital libraries research. This needs to be stepped up in the 7th Framework Programme. A solid body of existing research is creating the building blocks for a European digital library. Topics of immediate concern revolve around more sophisticated treatment of digitized materials; automated indexing of text, sound and images; improved multilingual search engines; services supporting annotations and collaborative work. The work and the results of the research community need however to be anchored in the needs of users, whether these be the owners and creators of content or the end user.

The first universal library in Alexandria burnt down 2000 years ago. Our digital libraries may face an equivalent loss if the question of how to preserve digital content is not adequately addressed. A number of ongoing Commission-funded projects are working on the integration of digital preservation tools into workflows and the preservation of different digital formats, including high-volume scientific data and multimedia music. This creates a baseline for future research.

I have no doubts that we shall reach our targets by 2010. The future will tell whether the European Digital Library will be more durable than the Library of Alexandria.

Horst Forster

ERCIM Beyond-the-Horizon Action Coordinates European ICT Research for the Future

by Peter Kunz

The results of theBeyond-the-Horizon action, which identified futuer ICT research challenges, was highlighted at a meeting with members of the European Parliament on 10 May in Brussels.

The results of the Beyond-the-Horizon action, which identified future ICT research challenges, were highlighted at a meeting with members of the European Parliament on 10 May in Brussels.

Bits, atoms and genes define the scene of future European research into Information and Communication Technology (ICT). Computers, the physical world and living organisms will increasingly merge, leading to entirely new methods of computing and communication. Large-scale interdisciplinary research efforts in this direction will be crucial for Europe's competitiveness in the long term.

This picture is the driving force behind the Beyond-the-Horizon action. Coordinated jointly by ERCIM and ICS-FORTH, it has identified six key research areas for developing the ICT of tomorrow's world. The action is funded by IST-FET, the Future and Emerging Technologies activity of the EU Information Society Technologies programme. Several workshops held across Europe during 2005 were followed by extensive consultation using the Internet, thus involving all relevant European research communities in the action.

Beyond-the-Horizon was presented at a meeting with members of the European Parliament on 10 May in Brussels, where researchers and officials from ERCIM and the EU further elucidated the action.

ICT has always profited from cross-fertilization with other scientific disciplines, including mathematics, biology, materials science and psychology. This is reflected in the wide range of problems and challenges to be addressed in the identified research areas. For example:

- the miniaturization of components on a chip requires new materials and new designs, as well as a search for alternative computing methods, eg quantum computing
- since future ICT systems need greater 'intelligence' in order to function properly, a promising way to achieve this is to study how living organisms – from a single cell to animal colonies and the human brain – process information
- the rapidly increasing volume and complexity of data and networks, in which humans interact with many small, embedded, mobile devices, requires penetrating studies of complex systems (Nature may teach us here too)
- mechanisms should be devised to ensure security for, and trust in the use of future technologies, which offer dazzling possibilities but also serious threats.

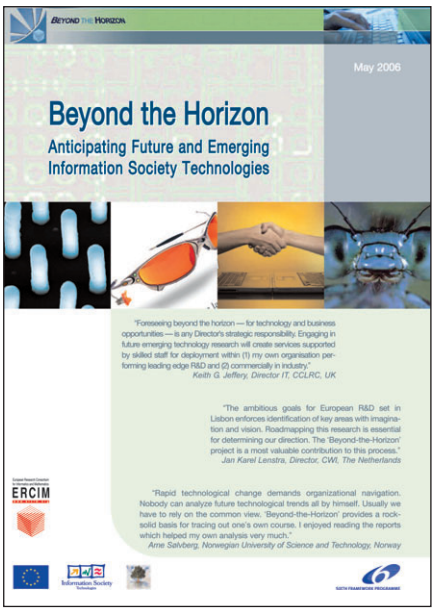
ERCIM has edited a booklet summarizing the results of the six thematic research areas identified by the Beyond-The-Horizon action. These are:

- Pervasive Computing and Communications
- Nano-Electronics and Nanotechnologies
- Security, Dependability and Trust
- Bio-ICT Synergies
- Intelligent and Cognitive Systems
- Software-Intensive Systems.

Three additional research areas were brought in by IST-FET and are included in the report:

- Quantum Information Processing and Communication
- Complex Systems
- Tera-Device Computing.

'IST Results', an online news service provided by the European Commission, has recently published a feature article on the Beyond-the-Horizon action citing Dimitris Plexousakis, scientific coordinator of the action from ICS-FORTH: "ICTs provide the glue that binds together multiple themes in European research. The time to address this multiplicity of themes and their inter-relationships is now.



ERCIM has published a booklet summarising the results of the Beyond-The-Horizon action - a European coordination action to identify ICT-related research trends and strategic areas that require support. The booklet is available for download from the ERCIM web site.

Links:

Beyond-The-Horizon home page:
http://www.beyond-the-horizon.net

Booklet presenting the results of the Beyond-The-Horizon action:
http://www.ercim.org/publication/policy/BTH-booklet-MAY2006.pdf

Feature article about Beyond-The-Horizon "ICTs - the glue that binds future research" published by "IST Results"
http://istresults.cordis.lu/index.cfm/section/news/tpl/article/ID/82433

Please contact:
Dimitris Plexousakis, ICS- FORTH
Tel: +30 2810 391637
E-mail: dp@ics.forth.gr

Stelios Orphanoudakis ERCIM Memorial Seminar

by Erzsébet Csuhaj-Varjú

In honour of Stelios Orphanoudakis, the fourth president of ERCIM, a memorial seminar was held on 30 May 2006 in Budapest during the ERCIM meetings at SZTAKI.

The premature passing of Stelios Orphanoudakis on 18 March 2005 was a great loss for the ERCIM community. A commemorative event was organized by Constantine Stephanidis, director of FORTH-ICS, and was held in Budapest during the ERCIM meetings.

The memorial seminar was opened by Keith Jeffery, president of ERCIM, whose speech highlighted the accomplishments of Stelios Orphanoudakis and the outstanding role he played in ERCIM. On behalf of the ERCIM community, Keith Jeffery thanked the organizers of the meeting for their efforts, and Jose Koster (CWI) from the Human Resources Managers Task Group of ERCIM, for commissioning a sculpture for permanent exhibition at FORTH-ICS, as a symbol of ERCIM community's high esteem for Stelios Orphanoudakis. He also thanked Cor Baayen, Dennis Tzichritzis and Gerard van Oortmerssen, former presidents of ERCIM, Alain Bensoussan, one of the 'founding fathers' of ERCIM, and the Orphanoudakis family members for their participation in the event.

The director of ICS-FORTH, Constantine Stephanidis, then gave an overview of the oeuvre of Stelios Orphanoudakis, briefly outlining his distinguished academic career and outstanding scientific achievements, his fundamental role in the life of FORTH and in the past and the future of ICT in Greece, and his influential pioneering activities in international scientific cooperation. Stelios Orphanoudakis deeply believed in ERCIM's promotion of European scientific research, and was actively committed to the achievement of this objective. His absence is a great loss for FORTH, for Greece, and for the international scientific community. As a former colleague, Constantine Stephanidis also warmly recalled their collaboration and friendship.

Technical presentations by distinguished speakers and former collaborators of Stelios Orphanoudakis followed, reflecting his main fields of research. Stelios Orphanoudakis had dedicated many years of teaching and research to the fields of computational vision and robotics, intelligent image management and retrieval by content, medical informatics, and medical imaging. The first talk was entitled 'Stelios Orphanoudakis, The European and The Greek', and was given by Niels Rossing (Danish Centre for Health Telematics, Denmark). The talk addressed past and future developments in e-Health, with reference to the pioneering ideas of Stelios Orphanoudakis in this area. Special mention was given to HYGEIAnet, the Integrated Regional Health Information Network of Crete, which is one of Orphanoudakis' most significant achievements.

The second talk, 'Computer Vision and Intelligent Systems', was presented by Jan-Olof Eklundh (KTH, Sweden), who summarized the trends in computational vision and robotics. Kostas Daniilidis (University of Pennsylvania, USA) gave the third presentation, '3D Visualization, 3D Navigation, 3D Content Creation'. The pioneering work of Stelios Orphanoudakis in all these areas was emphasized.

In addition to his research activities, Orphanoudakis had served on various committees and Working Groups of the European Commission, and was active in numerous European R&D programs.

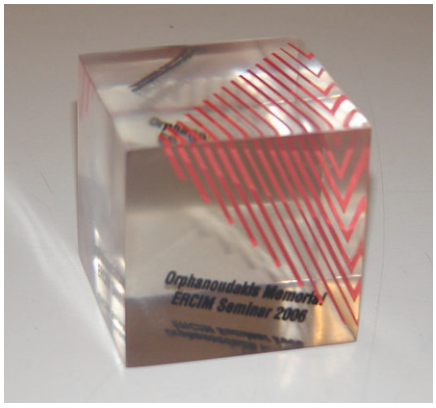
After the break, Ilias Iakovidis (Deputy Head of Unit-ICT for Health, European Commission) gave a talk entitled

From left: Keith Jeffery, ERCIM president, Constantine Stephanidis director of ICS-FORTH, Ava and Eleni Orphanoudakis, wife and daughter of Stelios Orphanoudakis.

From left: The former ERCIM presidents Gerard van Oortmerssen, Dennis Tsichritzis and Cor Baayen.

'e-Health: Achievements and Future Plans of the European Union'. Iakovidis also highlighted the important role played by Stelios Orphanoudakis in this area and in the European cooperation. All the presentations offered a unique blend of science and personal reflections.

In the second part of the meeting, Cor Baayen, Dennis Tscihritzis, Gerard van Oortmerssen and Alain Bensoussan discussed the extraordinary merits of Stelios Orphanoudakis as a scientist and as an individual, expressing their high estimation of his role in promoting ERCIM. They described Stelios with



A replica of the memorial artwork for permanent exhibition at ICS-FORTH as a sign of the high esteem in which the ERCIM community holds its former president Stelios Orphanoudakis.

great warmth, giving tribute to his charismatic personality, and his vision as a scientist who realized his ideas and plans with vigour. He had been a strong and pragmatic advocate of cooperation in Europe, and a great supporter of ERCIM since its inception.

At the end of the meeting, the commemorative sculpture was presented to Constantine Stephanidis for permanent exhibition at FORTH-ICS, and small replicas were presented to Ava Orphanoudakis and the invited speakers. The seminar concluded with a heartfelt speech by Ava Orphanoudakis, which was greatly appreciated by the audience.

The ERCIM "Alain Bensoussan" Fellowship Programme

ERCIM offers fellowships in leading European information technology research centres. Fellowships are available for PhD-holders from all over the world.

The Fellowship Programme has been established as one of the premier activities of ERCIM. Since its inception in 1991, over 180 fellows have passed through the programme. ERCIM has now named the programme to honour Alain Bensoussan, one of ERCIM's 'founding fathers. As presidient of INRIA, Alain Bensoussan initiatiated the creation of ERCIM in 1989 together with Cor Baayen from CWI and Gerhard Seegmueller from GMD (now part of Fraunhofer Institute).

Conditions

Applicants must:

- have obtained a PhD degree during the last 4 years prior to the application deadline or be in the last year of the thesis work with an outstanding academic record
- be fluent in English
- be discharged or get deferment from military service
- start the grant before October 2007 (for the September 2006 application deadline)
- have completed their PhD before starting the grant.

Fellowships are of 18 month duration, spent in two of the ERCIM institutes. In particular cases a fellowship might be of 12 month duration spent in one insitute.

The fellow will receive a competitive monthly allowance which varies depending on the country. In order to encourage mobility a member institute will not be eligible to host a candidate of the same nationality. Further, a candidate cannot be hosted by a member institute, if he or she has already worked in this institute for a total of 6 months or more, during the last 3 years.

Topics

The programme focuses on topics defined by the ERCIM working groups and projects administrated by ERCIM. Topics include: Applications of Numerical Mathematics in Science, BioMedical Informatics, Constraints, Control and System Theory, E-Learning, Dependable Software-Intensive Embedded Systems, Digital Libraries, Environmental Modelling, Formal Methods for Industrial Critical Systems, Grids, Image and Video Understanding, IT and Mathematics applied to Interventional Medicine, Matrix Computations and Statistics, Rapid Integration of Software Engineering Techniques, Security and Trust Management, Semantic Web, Soft Computing, Software Evolution, User Interfaces for All.

In addition, applications are also welcome for other areas in which ERCIM institutes are active. Detailed description of the topics is available on the ERCIM web site.

ERCIM does not only encourage researchers from academic institutions to apply, but also scientists working in industry.

Deadlines for Application

Deadlines for applications are 30 April and 30 September every year.

More Information

Detailed information, conditions and online application form is available at:
http://www.ercim.org/fellowship/

ERCIM Workshop on Software Evolution

by Tom Mens, Maja D'Hondt and Laurence Duchien

The ERCIM Working Group on Software Evolution organised its annual two-day workshop 6-7 April 2006 at Université des Sciences et Technologies de Lille (USTL) in France. The workshop reported on the theoretical, practical and empirical research on software evolution carried out by the working group members, and discussd new opportunities for collaboration.

The workshop gathered 40 researchers from ten European countries. In total, 25 position papers were submitted to the workshop, all of which were peer-reviewed by an international programme committee consisting of 17 well-known researchers. The best submissions were selected for inclusion in a special issue of Elsevier's Electronical Notes in Theoretical Computer Science. Eleven submissions were invited for a long presentation and six for short presentation. These workshop presentations covered a wide variety of research topics. Among others, the following topics were addressed, with the aim to provide either better formal support or better tool support: model-driven software evolution, aspect-oriented software evolution, component-based software evolution, architectural evolution, runtime software evolution, empirical analysis, software restructuring, and software quality measurement.

Arie Van Deursen, Delft University of Technology in the Netherlands, gave an invited talk on 'The Software Evolution Paradox: An Aspect Mining Perspective'. During this talk, he explored the relation between software evolution and the exciting research domain of aspect-oriented software development.

In addition to the scientific purpose, the workshop also hosted the annual steering committee meeting of the ERCIM Working Group on Software Evolution. The committee discussed the current status of the network, which includes over 35 members from research institutes all over Europe, 17 of which belong to ten different ERCIM partner institutes and plans about future activities such as opportunities for proposing new initiatives within the IST domain of the EU 7th Framework Programme, in particular within the strategic objective "Adaptive Software Intensive Systems". It appears that the need for supporting software adaptation and software evolution is becoming increasingly important within this strategic objective.

The workshop was co-organised by Tom Mens (WG chair), Laurence Duchien and Maja D'Hondt (ERCIM postdoctoral fellow) who offered to host the workshop at the Laboratoire d'Informatique Fondamentale de Lille (LIFL) and INRIA Futurs in Lille, France.

Links:
Working Group website:
http://w3.umh.ac.be/evol/
Workshop website:
http://w3.umh.ac.be/evol/meetings/evol2006.html

Please contact:
Tom Mens, Institut d'Informatique,
Université de Mons-Hainaut, Belgium
Tel: +32 65 37 3453
E-mail: tom.mens@umh.ac.be

Workshop participants.

ERCIM-Sponsored Events

ERCIM sponsors up to ten conferences, workshops and summer schools per year. The funding for all types of events is 2000 Euro.

Conferences

ERCIM invites sponsorship proposals from established conferences with an international reputation, where substantive overlap can be shown between the conference topic and ERCIM areas of activity. Typical cases would include annual conferences in computer science with international programme committees, substantial international participation, and proceedings published with an established international science publisher.

Workshops and Summer Schools

ERCIM sponsors workshops or summer schools (co-) organised by an ERCIM institute. The additional funding provided by ERCIM should be used to enhance the workshop by, for example, increasing the number of external speakers supported.

Application Deadlines

- *Conferences:*
  15 July 2006 for conferences later than 15 May 2007
  15 October 2006 for conferences later than 15 August 2007
- *Workshops and summer schools:*
  15 July 2006 for workshops and schools later than 15 October 2006
  15 October 2006 for workshops and schools later than 15 December 2006

Events sponsored in 2006

- ***World Wide Web Conference 2006***, Edinburgh, UK, 22-26 May 2006
- ***CAiSE 2006*** - 18th Confernce on Advanced Information Systems Engineering, Luxembourg, 5-9 June 2006
- **COOP 2006** - European Conference on Object-Oriented Programming, 20th edition, Nantes, France, 3-7 July 2006
- ***CONCUR 2006*** - 17th International Conference on Concurrency Theory, Bonn, Germany, 27-30 August 2006
- ***DISC 2006*** - International Symposium on Distributed Computing, Stockholm Sweden, 19-21 September 2006

More information:

http://www.ercim.org/activity/sponsored.html

It is with great sorrow that CNR announces the death of Franco Denoth, Director of the Institute for Informatics and Telematics of CNR, Pisa, and former member of the ERCIM Board of Directors.

Professor Denoth had a long and distinguished scientific career at CNR, spreading over almost fifty years, and beginning with his participation in the design of the first Italian computer to be built for scientific computing activities. He was director of three Institutes in Pisa: the Istituto di Elaborazione dell'Informazione (1979-1994), the Istituto per le Applicazioni Telematiche (1999-2002), and the Istituto di Informatica e Telematica (IIT) (from 2002) and was President of the CNR National Committee for Information Sciences and Technologies. He was also Italian representative at the European Commission for the IST programme under FP5. From 1999 he was responsible for the Italian Registry of Internet Domain names and, since 2002, CNR delegate on the board of EURid, the European Registry of Internet Domain names. From 1991 - 2001, he was the CNR representative on the ERCIM Board of Directors.

Denoth was a firm believer in multidisciplinary research and his main scientific interests were in the application of electronics and information technologies to medicine and biology. He was awarded three prestigious prizes for his research in these areas. Author of many scientific articles, he was member of the editorial board of several international journals, and editor of a number of books.

Franco Denoth will be deeply missed by all his colleagues and friends, in Italy and abroad, not only for his scientific merits but also for the warmth of his personality, his willingness to share a joke and his readiness to offer advice or assistance when necessary.

EC Expert Group on Next Generation GRIDs

by Keith Jeffery

The third Next Generation Grid expert group (NGG3) convened by the European Commission has completed its work and reported. The report (and much relevant documentation on GRIDs) is available at http://www.cordis.lu/ist/grids/

In the past few years, a group of high-level experts, named the Next Generation Grid (NGG) expert group, has developed a vision that has emerged as the European vision for Grid research. Driven by the need and opportunity of bringing Grid capabilities to business and citizens, the NGG vision underpins the evolution of Grid from a tool to solve compute- and data-intensive problems towards a general-purpose infrastructure enabling complex business processes and workflows across virtual organisations (VOs) spanning multiple administrative domains.

The NGG vision, articulated by NGG1 (2003) consists of three complementary dimensions: the end-user perspective where the simplicity of access to and use of Grid technologies is exemplified; the architectural perspective where the Grid is seen as a large evolutionary system made of billions of interconnected nodes of any type; and the software perspective of a fully programmable and customisable Grid. In order to realise the Next Generation Grid vision, numerous research priorities were identified in terms of properties, facilities, models, tools, etc. which have inspired national and international research programmes for Grid research. Almost half of the NGG1 experts were from ERCIM member organisations. The EC FP6 Call2 in the area of GRIDs resulted in projects aligned with the NGG1 vision, including the Network of Excellence managed by ERCIM: 'CoreGRID'.

NGG2 (2004) went further and elaborated the middleware required for GRIDs and considered the requirements of operating systems to support a GRIDs environment. Particular attention was paid to the need for self-* (self-managing, self-organising, self-healing, self-tuning etc) systems. There was initial consideration of the need for semantic description of service components to permit the construction of such systems. Almost 20% of the NGG2 experts were from ERCIM member organisations: this reflects the broadening acceptance of GRIDs in other organisations represented by the majority of experts. Projects resulting from EC FP6 Call 5 in the area of GRIDs are currently under negotiation.

NGG3 (2005 reporting January 2006) built upon these foundations and concentrated on a service-oriented architecture where the services have strong semantic descriptions allowing self-choreography (composition with flexibility and dynamism). Again approximately 20% of the experts were drawn from ERCIM member organisations and this team had a much stronger participation from industry, indicating the take-up of commercial interest in GRIDs. The SOKU (Service-Oriented Knowledge Utility) vision identifies a flexible, powerful and cost-efficient way of building, operating and evolving IT intensive solutions for use by businesses, science and society. It builds on existing industry practices, trends and emerging technologies and gives the rules and methods for combining them into an ecosystem that promotes collaboration and self-organisation. The benefits are increased agility, lower overhead costs and broader availability of useful services for everybody, shifting the balance of power from traditional ICT (Information and Communication Technology) players towards intermediaries and end-consumers of ICT. It is fortunate that SOKU may also be read as Self-Organising Knowledge Utility and Semantic Oriented Knowledge Utility.

The need for developing the SOKU vision stems from the necessity of effectively bringing knowledge and processing capabilities to everybody, thus underpinning the emergence of a competitive knowledge-based economy. The

SOKU vision builds on and extends the Next Generation Grids vision. It captures three key notions:

- *Service Oriented* - the architecture comprises services which may be instantiated and assembled dynamically, hence the structure, behaviour and location of software is changing at run-time;
- *Knowledge* - SOKU services are knowledge-assisted ('semantic') to facilitate automation and advanced functionality, the knowledge aspect reinforced by the emphasis on delivering high level services to the user;
- *Utility* - A utility is a directly and immediately useable service with established functionality, performance and dependability, illustrating the emphasis on user needs and issues such as trust.

The primary difference between the SOKU vision and earlier approaches is a switch from a prescribed layered view to a multi-dimensional mesh of concepts, applying the same mechanisms along each dimension across the traditional layers.

Thanks to the substantial investments and the numerous initiatives launched at the Member States and European levels, Europe has succeeded in establishing a leading worldwide position in Grids. The consistent portfolio of Sixth Framework Programme (FP6) Grid research projects will further contribute to the realisation of the NGG vision, thus boosting European competitiveness in Grid technologies and applications. It is no accident that ERCIM experts have been involved heavily in this strategic work. Three ERCIM personnel have been involved in each of the 3 expert groups: Thierry Priol (INRIA), Domenico Laforenza (CNR) and the author. We have worked closely and productively with our EC colleagues, particularly Franco Accordino, Max Lemke and Wolfgang Boch.

Link:
http://www.cordis.lu/ist/grids/

Please contact:
Keith G Jeffery
Director, IT CCLR and ERCIM president
E-mail: K.G.Jeffery@rl.ac.uk

EURO-LEGAL

News about legal information relating to Information Technology from European directives, and pan-European legal requirements and regulations.

European Commission Consulting on Copyright Levy

Under the Copyright Directive, EU member states were given a choice: either allow private copying and give 'fair compensation' to rights holders or ban private copying. Most European countries (except 5 member states) allowed copying of music for private use. These countries add a levy to the cost of items which are likely to be used to make private copies.

The Commission is now consulting with industry so that it can change the laws around this 'copyright levy' to suit the world of digital copying. An initiative on copyright levies is in the commission work program 2006. The additional follow-up consultation focuses on a series of salient points and will run from 6 June to 14 July 2006.

The commission sought to address the issue by posing various questions from several aspects, which also indicted the possible changes in the new law:

- the efficiency of applying the levy to equipment or media that consumers use, rather than the party that carries out and controls the private copying
- the necessity and the way of improving the accountability of collecting societies with respect to the application, collection and distribution of copyright
- the distribution of levies among right owners
- the efficiency of current copyright levy system with regard to the development of digital sales in Europe
- copyright levies and the notion of harm based on private copying
- the criteria for establishing whether a levy is imposed on particular equipment or media
- the phenomenon of convergence and copy right levies. It pointed out that levies that were applied to photocopies or cassette decks are being increasingly deployed on digital equipment, multifunction devices such as personal computers, hard disks and even printers
- the internal market and differences in copyright levy systems
- opinions from several sectors that are affected by copyright levies such as rights holders, collecting societies, the record and film industries, the ICT industry, consumers of digital equipment and /or blank media.

In the consultation document, the commission stated that in the digital media world "it would no longer be possible to hold only liable the manufacturers or importers of equipment and media. The logic of levies would also have to be applied to broadband and infrastructure service providers including telecommunications providers that carry content." This statement may indicate the possibility to impose copyright levy on ISPs in the future. However the commission also recognized that clarifying the complex situation is not an easy task, and it may cause "a serious risk of a backlash against the rights holder community and consumer welfare".

Link:
The consultation paper is available at http://www.ec.europa.eu/internal_market/copyright/docs/levy_reform/stakeholder_consultation_en.pdf

By Yue Liu, NRCCL, Oslo, Norway

W3C to Participate in Advisory Board of Internet Governance Forum

United Nations Secretary-General Kofi Annan established an Advisory Group to assist him in convening the Internet Governance Forum (IGF), a new forum for a multi-stakeholder dialogue on Internet governance. Daniel Dardailler, W3C's Associate Chair for Europe, will represent W3C on the new Advisory Board. W3C looks forward to sharing its experience in distributed consensus-building within this new international environment for standardization.

Link:
http://www.un.org/News/Press/docs//2006/sga1006.doc.htm

W3C Workshop on a Device Description Repository

Madrid, Spain, 12-13 July 2006

With ever-increasing diversity of Web-enabled devices, it is expected that content adaptation will play a significant role in the delivery of content. The successful adaptation of content to the capabilities of a device depends on reliable knowledge about the target device. For example, the selection of columns of a table may depend on the physical width of the screen. The goal of this Device Description Repository workshop is to discuss the design, the implementation and use of a repository of device information (DDR) for content and service providers, as part of W3C's Mobile Web Initiative.

Link:
http://www.w3.org/2005/MWI/DDWG/workshop2006/

Second Incubator Group to Explore Semantic Web for Multimedia Content

As part of W3C's Incubator Activity, W3C is pleased to announce the creation of the Multimedia Semantics Incubator Group, chartered to show how metadata interoperability can be achieved by using the Semantic Web technologies to integrate existing multimedia metadata standards. This new Incubator group (XG) is sponsored by W3C Members IVML-NTUA, CWI, University of Aberdeen, University of Maryland and DFKI.

Links:
Multimedia Semantics XG: http://www.w3.org/2005/Incubator/mmsem/
Incubator Activity: http://www.w3.org/2005/Incubator/

In Memoriam: Alan Kotok

Alan Kotok, W3C Associate Chair, MIT site manager and head of the W3C Systems Team, passed away mid-May in Cambridge, Massachusetts, USA. He was 64. Tim Berners-Lee, W3C Director, and Steve Bratt, W3C CEO, expressed their deep sorrow: "Our great friend, colleague, and mentor Alan Kotok has passed away. Alan's W3C involvement goes back before its formal inception, when he was still employed at Digital Equipment Corporation. His early ideas shaped W3C, and helped lead it to what it is today.

Long before Alan came to W3C, his experience established him as one of the early wise men of computer science.

One of Alan's undergraduate creations was the first video game, Spacewar, which he and several classmates created for the PDP-1 in 1962. Alan was also part of the team which invented the joystick, an icon of many young computer gamers' experiences.

Alan spent 34 years with Digital Equipment Corp. in numerous leadership roles. He served as Technical Director for product strategy and development groups in Telecommunications, Storage, and Internet. Alan provided thought leadership as a member of the Corporate Strategy Group which advocated early adoption and integration of Internet and Web-based technologies.

Alan held a wide range of roles at W3C. He carried the title of Associate Chairman, but he also served as the MIT site manager, managed the Systems Team, and worked closely with the Advisory Board. His contributions to membership and financial issues were highly valued. Alan shone as a problem solver, especially in important and complex areas: patent policy development, Patent Advisory Groups, whatever processes, policies and procedures were needed to improve the W3C as a standards body.

We have opened a publicly archived mailing list, public-memoria@w3.org, http://lists.w3.org/Archives/Public/public-memoria/ to which remembrances and photographs are welcome to be sent.

The W3C Team and our organization was immeasurably better for his presence. We will all miss him for who he was, and all that he achieved."

Link: http://en.wikipedia.org/wiki/Alan_Kotok

W3C Launches WebCGM Working Group

Computer Graphics Metafile, or CGM, is an ISO standard for tree-structured, binary graphics format that has been adopted especially by the technical industries (defense, aviation, transportation, etc.) for technical illustration in electronic documents. W3C started a Working Group which is chartered to develop a W3C Recommendation for WebCGM 2.0, starting with WebCGM 2.0 Submission.

Link:
WebCGM WG: http://www.w3.org/Graphics/WebCGM/WG/

Call for Implementations of Mobile Web Best Practices 1.0

W3C reached an important milestone toward its mission of making it as easy to use the Web on a mobile device as on a desktop computer. Written for designers of Web sites and content management systems, the 'Mobile Web Best Practices 1.0' guidelines describe how to author Web content that works well on mobile devices. W3C invites the designers of Web sites and content management systems to read the guidelines, make implementations, and test their results with the alpha version of a guidelines checker (http://www.w3.org/2006/05/mwbp-check/).

Also, in order to build a strong community of mobile Web developers, W3C has launched a wiki (http://www.w3.org/2005/MWI/BPWG/techs/TechniquesIntro) to collect observations and suggestions on techniques and implementation experience of Mobile Web Best Practices 1.0. Thirty organizations participating in the Mobile Web Initiative achieved consensus and encourage adoption and implementation of these guidelines to improve user experience and to achieve the goal of 'One Web.'

Links:
Mobile Web Best Practices 1.0 in Candidate Recommendation:
http://www.w3.org/TR/2006/CR-mobile-bp-20060627/
W3C Mobile Web Initiative: http://www.w3.org/Mobile/

Latest W3C Recommendations

- Web Services Addressing 1.0 - Core
  9 May 2006, Marc Hadley, Tony Rogers, Martin Gudgin
- Web Services Addressing 1.0 - SOAP Binding
  9 May 2006, Tony Rogers, Martin Gudgin, Marc Hadley, Tony Rogers, Marc Hadley, Martin Gudgin

A complete list of all W3C Technical Reports:
http://www.w3.org/TR/

W3C Web Security Workshop Report

W3C held a workshop on 'Transparency and Usability of Web Authentication' 15-16 March 2006, in order to identify steps W3C can take to improve Web Security from the user-facing end of the spectrum. Most workshop participants came from the security and browser vendor community, such as Google, HP, IBM, KDE, Microsoft, Mozilla, Nokia, Opera, VeriSign, Yahoo!, etc., as well as leaders of the online finance actors.

The workshop program was structured into seven sessions and an open discussion of next steps. Participants considered shortcomings in the usability of current browser-based authentication technologies. Requirements for and limitations of possible

Workshop participants.

improvements were also presented by a number of speakers. Approaches for concrete improvements included leveraging (secure) metadata; a number of proposals for changes to browser user interfaces and behaviors; protocol changes; and new approaches to identity online.

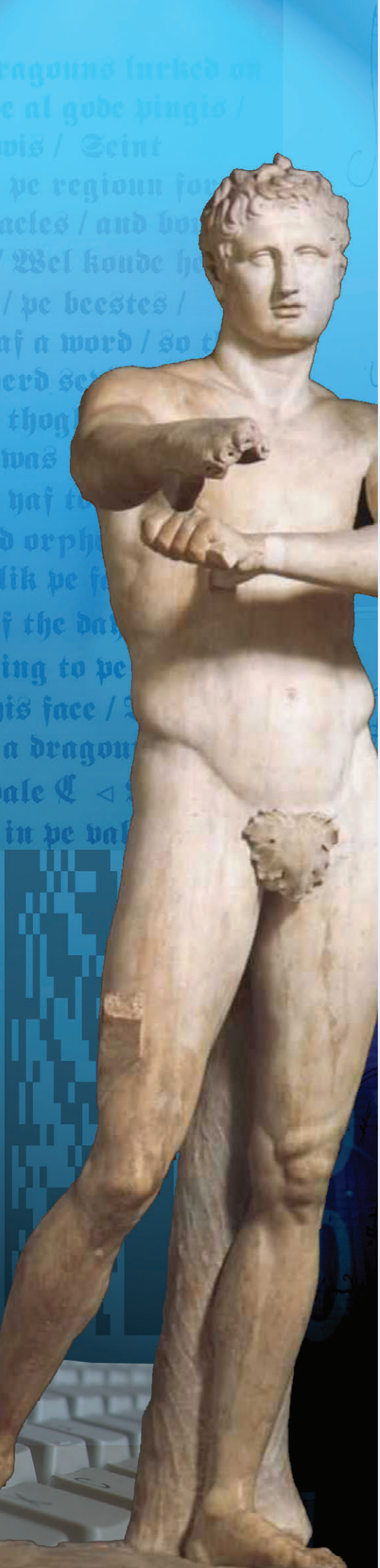
Based on the discussions, W3C staff is currently engaging those present at the workshop and other W3C Members in discussions that may lead to Working Group charters in three areas: Form-filler support; Secure Chrome; and Secure Metadata.

'Form-filler support' would enable browsers to reliably recognize log-in forms. This ability would allow browser-side credential management that is more reliable and usable than current heuristics-based form filling mechanisms. Browsers could also use this capability to launch advanced and security-focused user interfaces for credential entry.

Work on secure chrome and secure metadata would identify a baseline set of security context information that should be presented to the user, and best practices for the display of this information to the user. Work in this area may also cover restrictions on scripting capabilities that are known to make faking security indicators particularly easy.

The workshop, hosted by Citigroup, was chaired by Dan Schutzer (FSTC) and Thomas Roessler (W3C).

Link:
Workshop report: http://www.w3.org/2005/Security/usability-ws/report



Towards the European Digital Library - Introduction to the Special Theme

by Ingeborg Torvik Sølvberg and Costantino Thanos

The recent events in the European scene concerning the digital library field - the initiatives announced by Google and Yahoo aiming at making accessible vast online libraries of books, and the European Commission's plan for a European Digital Library recently unveiled - have put 'digital libraries' in the center of the debate and interest of the European research community.

That all citizens, anywhere, anytime, should have access to Internet-connected digital devices to search all of human knowledge, regardless of barriers of time, place, culture or language has also been the vision of DELOS, the European Network of Excellence on Digital Libraries, since its inception. DELOS believes that, in the near future, networked virtual libraries will enable anyone from their home, school or office to access the knowledge contained in the digital collections created by traditional libraries, museums, archives, universities, governmental agencies, specialized organizations, and individuals around the world. These new libraries will offer digital versions of traditional library, museum and archive holdings including text, documents, video, sound and images. But they will also provide powerful new technological capabilities that enable users to refine their requests, analyze the results, access collections in other languages, share resources, and work collaboratively. No matter where the digital information resides physically, sophisticated search software can find it and present it to the user on demand.

Having said this, we are not talking about a 'googlization' of digital libraries. Digital libraries should be much more than search engine portals. They should extend traditional libraries dramatically. They should provide services, including search, that facilitate the use of their resources by their target community.

The European Digital Library will be a major step towards making this vision a reality. The European Digital Library should support the interoperability of the different eContent holders – where by interoperability we intend the ability to store and retrieve information across collections in diverse media and languages, administered independently. Clearly, techniques for querying across languages that also take cultural differences into account must be available and the results of cross-language searches must be presented in a form that is easily comprehensible to the user.

The European Digital Library must also support the storage and preservation of its digital collections. Long-term storage technologies and efficient procedures for migration of contents and processes within a digital library to new environments should be developed so that they remain available to the user. The European Digital Library must also be able to manage complex intellectual property rights which will involve both legal and cost issues.

We are convinced that the time for the European Digital Library has come. The European digital library community has carried out a relevant amount of work during the last years. DELOS, in particular, has engaged the major European teams and expertise to help this become a reality and is ready to work with the eContent holders to make this vision a reality.

This issue stands in witness of the considerable amount of research activities carried out by the European digital library community and is organized in six sections. The first contains five invited articles. It begins with a descrip-

tion by Yannis Ioannidis of the long term vision for digital libraries that has been developed by DELOS during the last five years. The next article is by Elisabeth Niggemann who presents the views of CENL (Conference of European National Libraries) on the European Digital Library. Edward Fox then makes some considerations about the future European Digital Library on the basis of a theory-based approach to the field of digital libraries (the 5S model). The fourth article is co-authored by John Lervik and Svein Arne Brygfjeld and focuses on search engine technology applied in digital libraries. The final contribution in this section is by Jane Hunter and describes the shifting landscape of digital library R&D in Australia.

The rest of this section contains a selection of submitted articles on a variety of research topics within the Digital Library domain. We have grouped these articles under the following headings: digital library architectures and related concepts; ontology and metadata issues; information access and multimedia; repositories and preservation. The final sub-section presents three new projects addressing different aspects of the digital library paradigm. Overall we feel that these articles give a good picture of current trends in digital library R&D not only in ERCIM institutes but in the European research community at large.

Links:

http://www.delos.info/

http://europa.eu.int/information_society/activities/digital_libraries/

Please contact:

Ingeborg Torvik Sølvberg, NTNU, Norway

E-mail: ingeborg.solvberg@idi.ntnu.no

Costantino Thanos, ISTI-CNR, Italy

Tel.: +39 050 3152910

E-mail: costantino.thanos@isti.cnr.it

ARTICLES IN THIS SECTION

From Digital Libraries to Knowledge Commons

by Yannis Ioannidis

Digital Libraries began as systems whose goal was to simulate the operation of traditional libraries for books and other text documents in digital form. Significant developments have been made since then, and Digital Libraries are now on their way to becoming 'Knowledge Commons'. These are pervasive systems at the centre of intellectual activity, facilitating communication and collaboration among scientists or the general public and synthesizing distributed multimedia documents, sensor data, and other information.

Digital Libraries represent the confluence of a variety of technical areas both within the field of informatics (eg data management and information retrieval), and outside it (eg library sciences and sociology). Early Digital Library efforts mostly focused on bridging some of the gaps between the constituent fields, defining `digital library functionality', and integrating solutions from each field into systems that support such functionality. These have resulted in several successful systems: researchers, educators, students and members of other communities now continuously search Digital Libraries for information as part of their daily routines, decision-making processes, or entertainment.

Most current Digital Library systems share certain characteristics. They are content-centric, motivated by the need to organize and provide access to data and information. They concentrate on storage-centric functionality, mainly offering static storage and retrieval of information. They are specialized systems, built from scratch and tailored to the particular needs and characteristics of the data and users of their target environment, with little provision for generalization. They tend to operate in isolation, limiting the opportunities for large-scale analysis and global-scale information availability. Finally, they concentrate on material that is traditionally found in libraries, mostly related to cultural heritage. Hence, despite the undisputed advantages that current Digital Library systems offer compared to the pre-1990s era, the above restrictions limit the role that Digital Libraries can play in Knowledge Societies, which will serve as important educational nuclei in the future.

Together with the general community, the DELOS Network of Excellence on Digital Libraries has initiated a long journey from current Digital Libraries towards the vision of 'Knowledge Commons'. These will be environments that will impose no conceptual, logical, physical, temporal or personal borders or barriers on content. They will be the universal knowledge repositories and communication conduits of the future, common vehicles by which everyone will access, analyse, evaluate, enhance and exchange all forms of information. They will be indispensable tools in the daily personal and professional lives of people, allowing everyone to advance their knowledge, professions and roles in society. They will be accessible at any time and from anywhere, and will offer a user-friendly, multi-modal, efficient and effective interaction and exploration environment.

Achieving this requires significant changes to be made to past development strategies, which shaped the functionality, operational environment and other aspects of Digital Libraries. Knowledge Commons will have different characteristics. They will be person-centric, motivated by needs to provide novel, sophisticated, and personalized experiences to users. They will concentrate on communication and collaboration functionality, facilitating intellectual interactions on themes that are pertinent to their contents, with storage and retrieval being only a small part of such functionality. They will remain specialized systems that will nevertheless be built on top of widely-available, industrial-strength, generic management systems, offering all typically required functionality. In general, they will be managed by globally distributed systems, through which infor-

Related Reports

"Recommendations and Observations for a European Digital Library (EDL)": Brainstorming Report, Juan-Les-Pain, France, December 2005
http://www.delos.info/index.php?option=com_content&task=view&id=344&Itemid=125

"A Future Vision for Digital Libraries": DELOS Brainstorming Report, Corvara, Italy, July 2004
http://www.delos.info/files/pdf/events/2004_Jul_8_10/D8.pdf

"Digital Libraries at a Crossroads", DELOS FP5 Final Report, July 2003
In International Journal on Digital libraries, Volume 5, Number 4, August 2005
ISSN: 1432-5012 (Paper) 1432-1300 (Online) DOI: 10.1007/s00799-004-0098-4,
Pages: 255 - 265, http://www.informatik.uni-trier.de/~ley/db/journals/jodl/jodl5.html

"Digital Libraries: Future Directions for a European Research Program", DELOS Brainstorming Report, San Cassiano, Italy, June 2001:
http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf

mation sources across the world will exchange and integrate their contents. Finally, they will be characterized by universality of information and application, serving all applications and comprehensively managing all forms of content.

There are several key milestones to be achieved on the way towards Knowledge Commons. In particular, a Reference Model for Digital Libraries/Knowledge Commons must be obtained, that is, a framework with a set of interrelated concepts that will collectively capture the essence of the field and help everyone understand its basic elements. An appropriate system architecture (eg Grid or Peer-to-Peer) must also be identified. Other critical steps include devising sophisticated similarity search techniques, handling complex audio-visual content, personalizing user experiences, facilitating semantic interoperation among systems and modelling curation and preservation of content. Research within and outside DELOS is advancing steadily towards these so that a first version may become reality by the end of the decade.

It is serendipitous that, as part of its 'i2010 – a European Information Society for growth and jobs' initiative, the European Commission has recently announced its plans to foster the development of European Digital Libraries, so that Europe's written and audiovisual heritage becomes widely available. This represents a significant step towards Knowledge Commons. Extensive digitization of materials and the formation of many individual Digital Libraries should be followed by incorporation of the latter into a single, universal system that will provide unified access to all content across Europe. This will be the realization of the short-term vision of `The European Digital Library'. If appropriately advanced and openly expandable technology is used, it can also serve as the ideal springboard for realizing the longer-term vision of Knowledge Commons.

Link:
DELOS: http://www.delos.info

Please contact:
Yannis Ioannidis, University of Athens, Greece
E-mail: yannis@di.uoa.gr

The European Digital Library – A Project of the Conference of European National Librarians

by Elisabeth Niggemann

The Conference of European National Librarians (CENL) shares the vision of a European Digital Library and has been working towards this goal by creating The European Library service. The Communication published in September 2005 by the European Commissioner for Information Society and Media, Viviane Reding, on her plans for European digital libraries provided the impulse to think in a much broader way of a more comprehensive European Digital Library.

A true European Digital Library should serve all types of user needs: present and future, up-to-date or historical information, science and humanities, education, research and everyday, 'normal' information needs. It must comprise all types of media from the full gamut of Europe's cultural heritage institutions: libraries, museums and archives.

This is a highly diverse universe of knowledge, information and creativity: it exists as print, sound or image; in traditional analogue form or increasingly as 'born digital' or digitized form; and held by institutions with different professional backgrounds and traditions and in all the member states with their different institutional structures, responsibilities and financing. It would therefore be extremely difficult to organize a central, comprehensive super-structure. Instead, a networked structure is required, allowing faster and slower partners to proceed at their own speeds, while all benefit from the process.

The 'hub' of the network should be a central entry point to all the participating gateways. This network of networks must be scalable, and will rely heavily on common rules, standards and procedures. On the other hand, it must be built with diversity and heterogeneity in mind. From the users' point of view, cross-searching of all data and cross-services offered by all or most of the network members is of great importance. From the participating institutions' point of view it is the technical interoperability for limitless data exchange and high-performance access that are critical.

Partner networks need not be homogenized, but can continue to be organized in many ways. They will create among themselves a network of subnetworks, with nodes and substructures, that reflect the diverse needs of the different user communities, media types, institution types, and eventually also reflect legal frameworks.

Since the Europe of the future will be larger than it is now, it is important that all European countries are taken into consideration from the very beginning, not only those that are today's EU Member States. Building a scalable system means not only technical scala-

bility but also functional scalability: this includes all the European languages with their different character sets.

The European Digital Library should also – from the very beginning – try to build bridges to those global or regional networks outside Europe that provide additional resources for Europe's citizens and researchers.

One European gateway built on these principles is already in existence: TEL, The European Library. TEL is an ambitious and pioneering collaboration between European national libraries, supported by the EU and created under the auspices of CENL. It offers a professionally designed and maintained single access point to their holdings, spanning a range of collections in all the partner national libraries. It already allows access to more than one million digital items, as well as millions of catalogue records. Of the 45 CENL member libraries, about thirty will be full partners at the end of 2007, including all EU member states. All 45 CENL members were involved as partners from the start.

CENL believes that TEL is a model platform and model organizational network for building the European Digital Library. As a group, the members of CENL own Europe's cultural published heritage – many of them according to legal deposit, many for the whole period of time of their nation's history. In those cases where sections of their nation's memory are not part of their holdings, they are usually part of a national network of libraries that, as a group, cooperatively own the complete national corpus.

To take up the challenge of creating a European Digital Library, CENL adopted the Resolution on Digitization of European Cultural Heritage at its last annual meeting in Luxembourg in September 2005. CENL aims at examining how best to use the organizational model of The European Library to develop and coordinate, as a common effort, the existing strategies for digitization and digital libraries, including technical, content selection and funding issues. As it is particularly important to focus on content selection, CENL established the Content Working Group which will work on how content for mass digitization in Europe can be selected and created for the European Digital Library.

The next step towards a European Digital Library will be the start of a new EU-funded project where CENL is one of the coordinators. The project will deal with the enlargement of The European Library service by new partners, with multilingual issues and content development. Negotiations with the European Comission are underway and the project should commence in summer 2006: the European Digital Library (EDL).

Links:

http://www.cenl.org

http:// www.theeuropeanlibrary.org

Please contact:

Elisabeth Niggemann, Chair of CENL

Die Deutsche Bibliothek, Frankfurt, Germany

E-mail: niggemann@dbf.ddb.de

A Forward-Looking European Digital Library? Hence 5S?

by Edward A. Fox

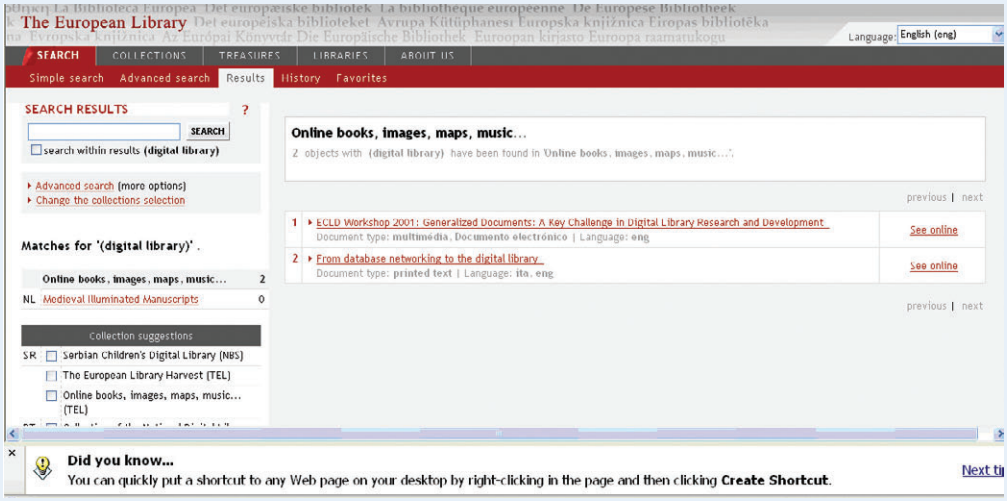
Will The European Digital Library (TEDL), as it expands beyond current efforts with The European Library, be forward-looking? Will it consider the 5S checklist: Societies, Scenarios, Spaces, Structures, and Streams? Will it draw upon the expertise that DELOS helps to make visible, and that is apparent in conferences like the European Conference on Digital Libraries?

Europe has a wonderful opportunity to become a world leader in developing a comprehensive next-generation digital library that will serve the entire European Community. Doing so will require courageous leadership, careful investment, hard work and a willingness to share for the common good.

The EU has much to draw upon as it works toward TEDL. One key resource is DELOS, the Network of Excellence on Digital Libraries, partially funded by the European Commission's Information Society Technologies Programme (IST). Another resource is the annual conference on digital libraries in Europe; eg, ECDL 2006, the 10th European Conference on Research and Advanced Technology for Digital Libraries. (Though current holdings of The European Library are limited, ECDL does come up as one of the two things found when searching for 'digital library', as can be seen in the figure.) Even beyond the borders of Europe, there exist professional groups like the IEEE-CS Technical Committee on Digital Libraries (TCDL), and conferences like JCDL and ICADL. The field of Digital Libraries, which began around 1991, has led to many projects, services, systems, theoretical results and a great deal of advanced technology. I sincerely hope that the EU will ensure that this work and all its related successes will lead to a forward-looking TEDL.

One way to focus this discussion is to use the 5S framework, a theory-based approach to the field of digital libraries. First, we consider the Society perspective. What social (and cultural, economic, legal, linguistic, national, to

The European Library - http://www.theeuropeanlibrary.org - the portal for searching the content of European national libraries.



political, etc.) concerns apply? What is the target audience? Who are the users? How can their collaboration, membership in groups, and myriad other social relationships be made use of, for example through collaborative filtering? How can we move beyond TEL's current focus on librarians (as is clear from the interface, query language, use of federated search technology and provision mostly of catalogue records)? How can the citizens, students, teachers, scholars, researchers, businesses and other institutions in the EU be properly supported?

This leads us to think about Scenarios. For what purposes, goals, and objectives can TEDL be used? For what tasks and activities? Through what classes and types of services? How can such a system respect privacy, while at the same time remembering a user for longer than the brief time it takes to handle a single WWW transaction: that is, into sessions, investigations and lifelong learning activities? Will Open Access be fully supported, to all available content?

Considering Spaces, we must consider all the aspects of context, including the location of the user, the effects of time and space, the presentation of results using information visualization, integration with geographic information systems, and the simplification of searching that comes from reducing the dimensionality of a vector space (with features that range from terms, phrases and categories to concepts and color histograms)? Can TEDL work in 'Semantic Space' as we move toward Web 2.0 and the Semantic Web? Will search results be grouped into useful clusters based on similarity? Will results be presented, if desired, on a map (eg, of Europe), so that locations are understood, such as the site of an author's institution?

In the case of Structures, will TEDL make use of organization in its many forms? Of ontologies and thesauri? Of appropriate category systems? Of presentation using trees or graphs? Of fields and markup structure? Of database schema and related records? Of facts, snippets (ie structures atop content streams) and extracted information?

Considering Streams, will TEDL go beyond streams or sequences of characters and consider large books? Audio? Video? Animation? Sensor streams from surveillance or satellites? Will streams be managed in concert with structures in order to find suitable video frames or book sections? Will spatial considerations help in stream selection and subsetting (eg finding a scene in a news video covering a particular heritage site or important event or personage)? Will helpful scenarios allow users to work with all media types and mixtures, providing all the support afforded for textual content, and beyond?

Clearly, TEDL could serve millions, with tens of millions of resources, including theses, reports, papers, newspapers, images from museums and courseware. Will institutional repositories at all colleges, universities, centres, and even agencies and businesses support Open Access? Will the holdings of the regional, federal and national libraries of Europe catalogue both information and full-text (and full-content), and be opened as well? Will publishers cooperate to vastly increase their market and visibility?

This is a vision to challenge the entire R&D community worldwide. I hope the EU will work in this direction and call for broad involvement and assistance in a forward-looking digital library – of and for Europe (and beyond).

Links:

DELOS: http://www.delos. info

5S framework:
http://www.dlib.vt.edu/projects/5S-Model

ECDL 2006: http://www.ecdl2006. org

Please contact:
Edward A. Fox
Virginia Tech, USA
Tel: +1 540 231 5113
E-mail: fox@vt.edu

Search Engine Technology Applied in Digital Libraries

by John M Lervik and Svein Arne Brygfjeld

Digital libraries are experiencing rapid growth with respect to both the amount and richness of available digital content. This is the result of a range of large-scale digitization projects on books and periodicals that are occurring locally, nationally and internationally. Furthermore, a number of libraries are digitizing other information carriers such as still images, audio and video. Much of the information in libraries has good metadata; to supplement this, OCR technology is used to extract text-based information from textual information within digitized images.

As a consequence of the huge amounts of digital content becoming available, modern search engines are now being introduced in digital libraries. Online users are accustomed to Internet search engines: they expect simplicity, speed and cross-collection searching. Given the complexity and amount of digital content, third-generation search technology is required to provide highly relevant search capabilities. This search technology also scales in a cost–efficient manner and provides proven methodologies for the operation and maintenance of these systems. Hence, as digital libraries grow in content volume and diversity, this third-generation search technology provides a convenient platform to give users what they expect with respect to relevance, functionality, and speed.

The exponential growth in the volume and diversity of digital library content represents both challenges and opportunities. One obvious challenge is to keep the query response time low. While conventional search technology gives acceptable query response times when searching is limited to metadata only, it cannot handle the addition of large amounts of OCR text. With a third-generation platform, searching can be performed across multiple systems and on all the available content.

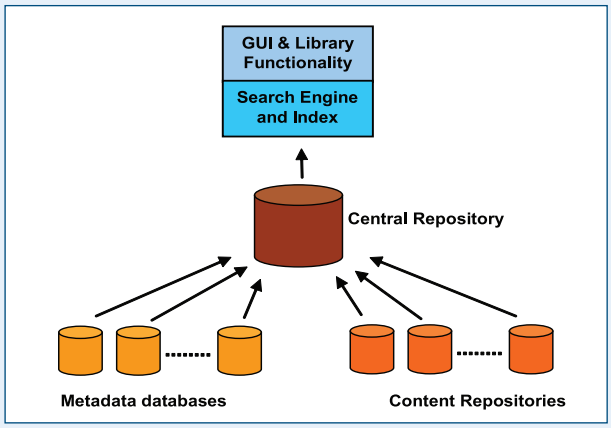


Figure 1: Functional model at the National Library of Norway

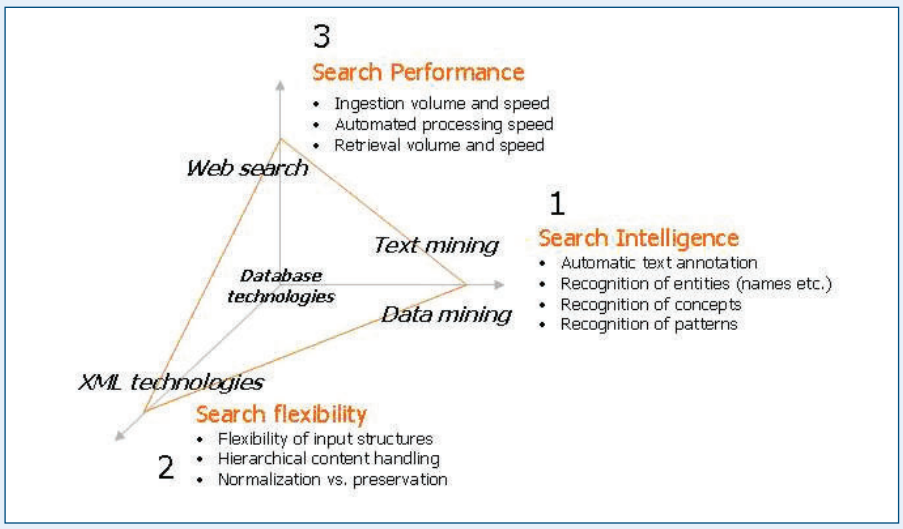


Figure 2: Contextual searching.

Another significant challenge is to provide the most relevant information to the user. This involves finding and ranking hits from large amounts of information, as well as retrieving relevant information that the user does not know exists or that she/he may not even know how to search for. The most relevant information may exist in another database and be described in ways that do not show up in traditional searches. Cross-database searching is another area where traditional systems are insufficient. Some systems provide distributed searching through the use of Z39.50, OpenURL and other protocols. However, they are very slow due to protocols and long response times in source systems, and do not provide one consistent relevancy function across all of the content sources.

A third challenge is searching a rich set of information types, such as the content of still images or audio. With such a diversity of information, there is a clear need for a more dynamic inclusion of new search methods on various information types.

Third-generation search engine technology is designed to combine the scalability of Internet search engines with new and improved relevance models.

This will include contextual relevance, allowing searches to be performed across any type of content and any type and number of sources. Libraries already have a wealth of experience in harvesting information to build centralized repositories of both metadata and content. Such environments are well suited for exploiting the capabilities of search engine technology and thus meeting the challenges above. The National Library of Norway has recently implemented infrastructure based on these principles, with the FAST search platform as a core component. This is shown in Figure 1.

This model makes it possible to perform very fast and relevant searches of large amounts of information residing in disparate databases. In this case, more than thirty metadata sources are included in addition to content from digitized newspapers, books and journals. The search service provides access to the complete palette of information at the library, including books, periodicals, still images, audio and video. It also serves as a base for statistical purposes, as well as management information.

A major advance in third-generation search technology is the introduction of contextual searching. The current Web search approach provides links to documents based on a hard-wired ranking method, such as the number of inbound links or scientific citations. This approach has some fundamental limitations for use by digital libraries. First, researchers cannot find what they don't know exists, so all queries must be specified by the user and there are no tools to crate data-driven content analysis. Second, researchers will only get access to the 'most popular' or 'newest' articles as defined by a black-box relevance function. It is possible that neither of these forms of ranking will illustrate the results or trends the researcher is interested in. Third, the user interaction does not provide any learning for the researcher. The researcher must open and read the full articles in order to assess them, and this time investment is usually significant. Hence, the overall approach offered by most library services does not make the best use of the offered content.

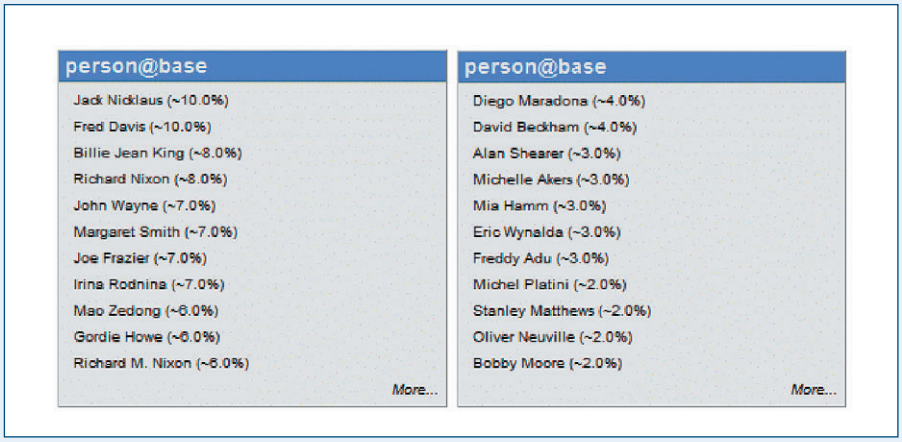


Figure 3:
Comparison of results from a standard Web search (left) and a contextual search (right).

Contextual searching has been designed to address these challenges. Figure 2 illustrates some of the key components in the foundation of a contextual search.

Contextual searches introduce new metaphors for user interaction: each document is further decomposed into semantic components that can be retrieved and analysed across billions of articles. Even ambiguous, open-ended queries can therefore be answered with highly accurate 'table-of-contents'-type results relating to factual information contained inside the relevant documents. Hence, in addition to getting access to highly relevant search results, researchers can:

- discover concepts/facts they did not know existed
- understand trends and get access to the 'long tail' of information which is invisible in traditional searches
- become more efficient and understand factual patterns across different documents.

As an example of how contextual searching improves search precision, we have run two test queries against the online encyclopedia Wikipedia (see Figure 3). Results on the left are from the query "persons that appear in the same document as the word 'soccer'"; those on the right are from the query "persons that appear in the same context (ie paragraph) as the word 'soccer'". The improvement in result quality is striking, yet the first list of results corresponds to a standard Web search result which until recently was considered state-of-the-art!

Contextual searching allows the information provider to preserve all the original information and spend less time annotating and classifying the content. To end users, it gives ease-of-use through better content and features, and creative freedom to ask questions the providers may not have planned for. In essence, the value of information should not be judged by the ability to store it, but by the ability to use it.

In summary, contextual searching enables deep semantic analysis and refinement across structured, unstructured and rich media, and dynamic interpretation of contextual meaning of the content. The overall results of the contextual combination are vastly improved discovery, schema exploration and disambiguation capabilities. Contextual search is all about turning information buried inside digital libraries into value for researchers.

Please contact:
John M Lervik, CEO
Fast Search & Transfer ASA (FAST), Norway
E-mail: John.Lervik@fast.no

Svein Arne Brygfjeld, Digital Libraries,
The National Library of Norway
E-mail: sveinb@oline-atm.nbr.no

The Shifting Landscape of Digital Libraries Research and Development in Australia

by Jane Hunter

Over the past ten years, Australia has seen steady growth in research and development in Digital Libraries (DLs). The number, size, sophistication and recognition of related projects and initiatives has increased, leading to better funding from government and private organizations. Specific strategies have been put in place to improve Australian skills, expertise and technologies in this field. The aim is to ensure that digital content being developed in the Australian creative, cultural, educational, scientific and academic sectors is accessible in the long term. This requires that it be stored and maintained in digital libraries with robust management and preservation processes.

The first phase of DL research began in the mid- to late 1990s with Australian involvement in metadata standards designed to facilitate the discovery of digital resources, ie Dublin Core. The key Australian participants in this work included the National Library (through its Digital Services Project), the National Archives and the Resource Discovery Unit at DSTC in the University of Queensland. The main focus of this work was on metadata input tools and Dublin Core-based search engines. During the late 1990s, this research activity expanded to include metadata standards for multimedia content: images, video, audio and composite multimedia objects. However, it was still primarily focused on support for managing digital collections within libraries, archives and cultural institutions. Funding was ad hoc and fragmented, and a coordinated national effort did not exist.

In the past five years, the Australian government has attempted to develop a more structured approach to the funding of DL activities through the Systemic Infrastructure Initiative (SII) and the Australian Research Information Infrastructure Committee (ARIIC). The main focus of ARIIC is to improve the access of Australian researchers to relevant information, thereby aiding their research and making their results widely available and easily accessible. In 2003, ARIIC funded four FRODO projects (Federated Repositories of Online Digital Objects) to the tune of $AUD22 million. These are:

- Australian Partnership for Sustainable Repositories (APSR)
- Meta Access Management System Project (MAMS)
- Australian Research Repositories Online to the World (ARROW)
- Australian Digital Theses Program Expansion (ADT).

These projects focused primarily on traditional scholarly publications and on evaluating existing DL technologies or middleware developed overseas that could be promoted and deployed within Australia. These included DSpace, Fedora and Shibboleth. Although some local refinements and extensions were made, the FRODO projects have not delivered any original DL research.

In 2005, ARIIC provided funding to $AUD nine MERRI projects (Managed Environment for Research Repository Infrastructure). By this stage, the global trend was towards the adoption of e-Research infrastructure; hence, rather than dealing with traditional scholarly publications, the MERRI projects are focusing on long-term access to raw research data (including scientific, medical, financial and sensor data) and analytical services by collaborating teams of scientists. The nine projects are:

- BlueNet: the Australian Marine Science Data Network
- Molecular Medicine Informatics Model (MMIM): a multi-institutional, multi-disciplinary research and training platform for clinical research
- Time Sync: mapping the global financial system
- Australian Service for Knowledge of Open Source Software (ASK-OSS)
- Middleware Action Plan and Strategy (MAPS)

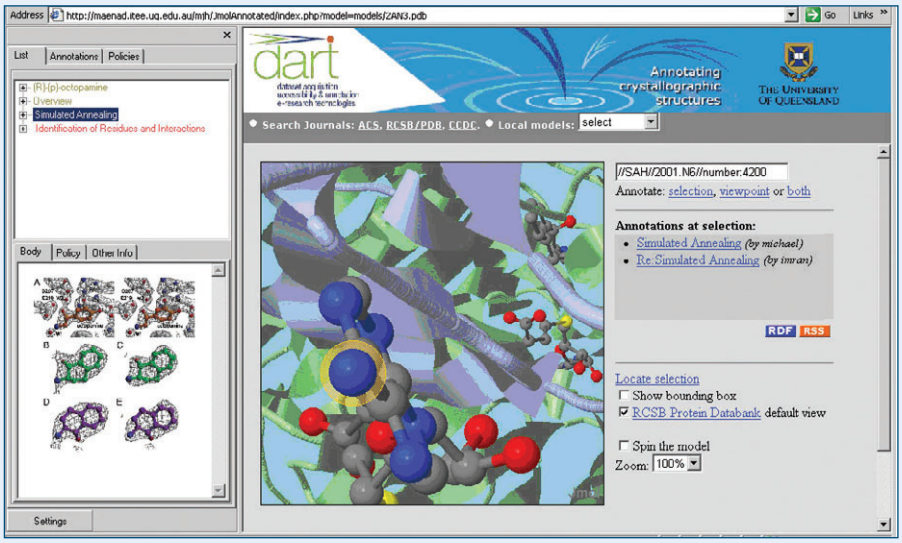


Figure 1: DART's secure annotation system for the collaborative annotation of 3D crystallographic structures.

- Legal Protocols for Copyright Management: facilitating open access to research at the national and international levels (OAKLAW)
- Dataset Acquisition, Accessibility and Annotation e-Research Technology (DART)
- E-Security Framework for Research
- Regional Universities Building Research Infrastructure Collaboratively (RUBRIC).

Most of these projects are in an early stage of development, and have not yet produced any substantial results. The DART project has, however, developed and demonstrated sophisticated annotation services for communities including protein crystallographers, climate modellers and social scientists. Figure 1 illustrates a screen shot from the secure annotation system developed by the University of Queensland to enable collaborative annotation of 3D crystallographic structures.

Apart from the ARIIC-funded projects, there are a number of additional DL research projects being undertaken within Australia that are specifically focusing on preservation. The National Library's PANDORA project is investigating digital preservation technology. In addition, the PANIC project at the University of Queensland is investigating automatic obsolescence detection and notification services, and the application of Semantic Web/GRID services to the discovery of optimum preservation services.

Please contact:
Jane Hunter
University of Queensland, Australia
E-mail: jane@dstc.edu.au

A Reference Architecture for Digital Library Systems

by Leonardo Candela, Donatella Castelli and Pasquale Pagano

The building of a European Digital Library requires a cooperative and distributed development model that, as far as possible, promotes the sharing and reuse of current Digital Library products. In order to support this model, an abstract solution to the problem of implementing a Digital Library - in other words, a reference architecture – is fundamental.

After about fifteen years of Digital Library (DL) research, development and deployment the research community is now well versed in the successes and weaknesses of the field. Until now a pragmatic approach has been adopted in developing systems via specialized methodologies, usually by adapting techniques borrowed from other disciplines. This approach has produced a plethora of heterogeneous entities and systems – commonly classified as 'digital library systems' – and has resulted in a lack of agreement on what should constitute the fundamental aspects of DL technology. This makes the interoperability, reuse, sharing, and cooperative development of DLs extremely difficult. Moreover, the role played by DLs undergoes continuous evolution, making current systems inadequate for future applications. Modern DLs are conceived as systems to support the whole process of dealing with human knowledge production, maintenance and communication.

A Digital Library Reference Model

To overcome these limitations and lay the foundations upon which to build future DL systems, the DELOS Network of Excellence on Digital Libraries is promoting an activity, lead by ISTI-CNR, aimed at producing a Digital Library Reference Model.

A reference model is an abstract framework for describing and understanding the significant relationships between entities in an environment. It consists of a minimal set of unifying concepts and relationships within a problem domain and is usually independent of specific standards, technology, implementations or other concrete details. Its goal is to enable the development and integration of systems by using consistent standards or specifications supporting that environment.

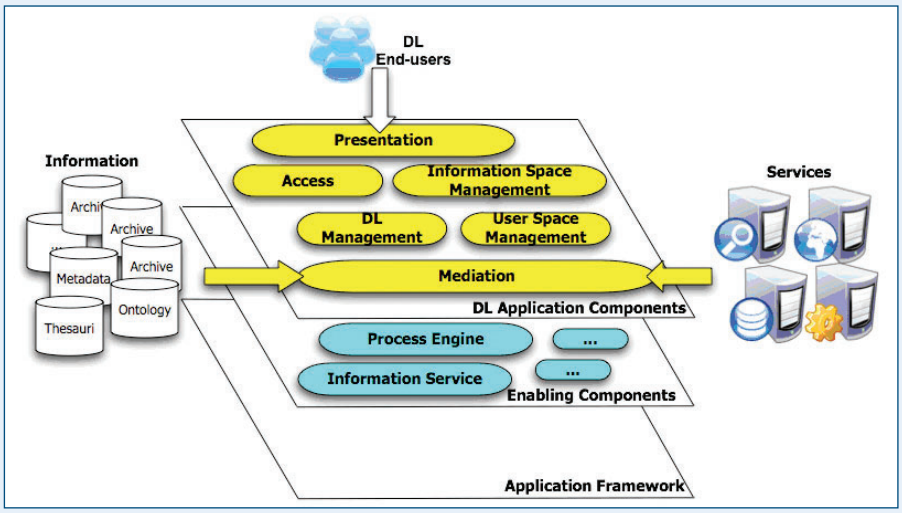
The core of this model is being developed by a consortium that includes, in addition to ISTI-CNR, four other partners: the University of Athens (GR), the University of Glasgow (UK), the University for Health Informatics and Technology (AU), and the University of Basel (CH).

The DELOS Digital Library System Reference Architecture

One of the main outcomes of this work is the Digital Library System Reference Architecture depicted in Figure 1, which provides an abstract solution to the problem of organizing and implementing DL systems. This reference architecture is based on a loosely coupled component-oriented approach. Such an approach is fundamental for the purposes of the reference architecture, since it allows for: (i) easy tailoring of the DL through component selection and replacement; (ii) reuse of the components in different contexts; (iii) distributed installation (since each component can be independently implemented); and (iv) easy support for heterogeneity issues by using or providing an appropriate component dealing with the particular issue.

Together with the component-oriented approach we also adopt a layered approach, and organize the constituent components into three tiers: (i) the Application Framework, ie the set of libraries and subsystems supporting the operation of the other DL system components; (ii) the Enabling Components, which provide the functionality required to support cooperation among the components implementing the DL application; and (iii) the DL Application Components, which provide the DL functionality specific to end users.

The components that constitute the DL Application Components tier are further organized into functional areas. Mediation components deal with and provide access to third-party information and services that vary in structure, format, media and physical representation. Information Space Management components implement and manage the DL information objects by providing storage and preservation facilities. Access components support the discovery of DL information objects via search and browse facilities. User Space Management components provide for registration and activities concerning administration of the users. DL Management components aid the administration of the DL in terms of the librarian activity, eg review processes, policy management and preservation activities. Finally, Presentation components provide users with simple access to the DL information and services, namely the graphical user interface.



The Digital Library System Reference Architecture.

Exploitation and Future Work

In order to test this model and prove its effectiveness, a DELOS DL system prototype is being implemented. The goal is to combine existing software and technologies produced by members of DELOS, and use the guidelines provided by the Reference Architecture to build a prototype. One of the first outcomes of this activity is identifying the most appropriate current technologies: the proposed component-oriented approach dovetails with the service-oriented approach and promotes the usage of P2P and Grid technologies.

These activities, ie the Digital Library Reference Model and the DL system Reference Architecture with its preliminary implementation, are the topics of a DELOS Workshop. The workshop took place in June 2006 in Rome, and involved leading DL researchers. The objective of the workshop was to discuss the current model and to plan future activities and collaborations to build on this foundational work.

Link:
DELOS website: http://www.delos.info

Please contact:
Donatella Castelli, ISTI-CNR, Pisa, Italy
E-mail: donatella.castelli@isti.cnr.it

DelosDLMS - Infrastructure for the Next Generation of Digital Library Management Systems

by Hans-J. Schek and Heiko Schuldt

The overall goal of the DelosDLMS is the implementation of a prototype of a next-generation digital library management system. This system combines text and audio-visual searching, offers personalized browsing using new information visualization and relevance feedback tools, allows retrieved information to be annotated and processed, integrates and processes sensor data streams, and finally, from a systems engineering point of view, is easily configured and adapted while being reliable and scalable. The prototype will be built by integrating digital library functionality provided by the DELOS partners into the OSIRIS/ISIS platform, a middleware environment developed by ETH Zürich and now being extended at the University of Basel.

In the first two years of DELOS - the EC-funded Network of Excellence on Digital Libraries, work has mainly focused on improving digital libraries (DLs) by developing independent, powerful and highly sophisticated prototype systems. Current integration activities began in early 2006 and are coordinated by the architecture work package of

DELOS. These prototype systems are being integrated as building blocks into OSIRIS/ISIS, an existing middleware environment that was developed at ETH Zurich. The result of the integration – that is, the middleware infrastructure together with all the advanced DL functionality – will constitute the DelosDLMS.

ISIS and OSIRIS

A central task in the second phase of DELOS is the development of a global prototype. The objective is to build a joint prototype for the future Digital Library Management System that makes available results from many groups in DELOS. This will be based on the OSIRIS/ISIS middleware, the development of which began at ETH Zürich for ETHWorld, the virtual campus of ETH. It was further developed for data streams and for medical objects at UMIT, and is currently being extended at the University of Basel. The OSIRIS middleware (Open Service Infrastructure for Reliable and Integrated process Support) supports programming-in-the-large; ie, the combination of arbitrary application services into so-called processes. This is realized by a set of generic (application-independent) services that include the registration of services and processes, interfaces for application development, an engine for decentralized execution of processes, and services for load balancing. In addition, it features reliable execution by applying advanced database concepts – essentially for failure handling and concurrency control – at the level of processes. ISIS (Interactive SImilarity Search). is a set of DL services that have been developed on the basis of the OSIRIS middleware. ISIS includes a sophisticated index structure for similarity searching, which is particularly well suited for high-dimensional vector spaces. Furthermore, in terms of Digital Library functions, ISIS features rudimentary support for textual and content-based audiovisual searching. It also provides basic support for relevance feedback and visualization.

With the DelosDLMS, existing ISIS services will be significantly enriched by other specialized DL services that have been developed within the DELOS network. This will be achieved by integrating these services into the OSIRIS infrastructure, thereby combining them with other ISIS and non-ISIS services into advanced, process-based DL applications.

The DelosDLMS

The plan for the DelosDLMS includes the upgrading of existing ISIS components and services and the integration of new functionalities. The final product will support multi-object multi-feature queries over collections of different media types. Personalized browsing and information access, relevance feedback and object annotation will also be considered. Since information (for instance in e-Science Digital Library applications) increasingly originates from software or hardware sensors, sensor datastream processing will also be integrated in the DelosDLMS. Essentially, all this DL functionality will be made available by means of services. The challenge of DelosDLMS is therefore to provide a scalable and reliable infrastructure where these services can be plugged in and used as building blocks.

Two alternatives exist for integrating services with OSIRIS. First, there are tightly coupled services, which are tightly integrated into the OSIRIS runtime infrastructure. Advanced failure handling and load balancing are among the main advantages of this arrangement. In terms of failure handling, compensating services can be registered which are automatically invoked in case of failures. In terms of load balancing, ORISIS can automatically choose the node carrying the lightest load to invoke a service that is deployed several times. This is particularly important for computationally expensive services like feature extraction.

Second, services can be loosely coupled with OSIRIS, meaning that services are described and invoked by standard Web service interfaces (SOAP and WSDL). This reduces the effort needed for integration but does not provide the benefits of tight coupling.

Recently, a 'call for services' has been issued to both members and non-members of DELOS. The goal is to identify services which are best suited for integration into the DelosDLMS. For the first version of the integrated DelosDLMS prototype, services will in most cases be loosely coupled. The final version will then support a higher degree of reliability by tightly coupling as many services as possible. Services will be integrated from the following areas:

- sophisticated term extraction from text, text indexing and collection management
- annotation services
- reliable sensor data management
- multimedia indexing
- automatic search process generation and personalization services
- image feature extraction
- 3D shape recognition
- special indexing techniques for video retrieval
- audio feature extraction and audio retrieval
- advanced visualization services and visual relevance feedback
- Self-organizing maps visualization
- active paper (linking digital information and paper)
- services for transformations between standards
- ontology services and natural language access
- preservation services
- services for multi-lingual access.

This list will be extended and revised after the evaluation of the call for services and during the actual integration work. Nonetheless, it highlights examples of building blocks that will be considered for DelosDLMS.

Links:

The DELOS Project: http://www.delos.info

The DELOS Architecture Work Package: http://dbis.cs.unibas.ch/delos_website/

The ISIS/OSIRIS Homepagel: http://dbis.cs.unibas.ch/research/isis_osiris

Please contact:

Hans-Jörg Schek, Department of Computer and Information Science, University of Konstanz, Germany

Heiko Schuldt,

Database and Information Systems Group, University of Basel, Switzerland

Tel: +41 61 267 0558

E-mail: heiko.schuldt@unibas.ch

A Powerful and Scalable Digital Library Information Service

by Henri Avancini, Leonardo Candela, Andrea Manzi and Manuele Simi

The implementation of a Digital Library capable of putting Europe's memory on the Web demands a service-oriented, federated and distributed approach. Supporting such an approach requires the introduction of a new type of enabling service, usually called an Information Service, which can collect and disseminate information on the resources that constitute the federation. In large and distributed Digital Libraries, the key features of this service are scalability and availability.

The DLib group of the Networked Multimedia Information System Laboratory at ISTI-CNR has extensive experience in building digital libraries (DLs). This experience arises from the participation with scientific leadership in a series of EU IST projects such as SCHOLNET. It also stems from the development of OpenDLib, a highly flexible Digital Library Service System that has been shown to be suitable for building and operating a range of digital libraries. One of these is the DELOS DL, which manages the documentation of the DELOS Network of Excellence. Another is the BELIEF DL, which serves the eInfrastructure community by collecting and providing focussed views over multimedia documents, as well as presenting the latest details relating to projects, initiatives and events.

Our experience leads us to believe that the service-oriented approach with loosely coupled services is the most appropriate architectural approach for building highly distributed systems. This approach relies on independent services that provide the expected functionality by cooperating with other services of the federation. In order to produce distributed DL infrastructures of a high quality, we find that an effective discovery phase of the constituent components and careful monitoring of the infrastructure are mandatory. Supporting these features means relying on two kinds of information about services: (i) static information, which includes data that remains fixed during the service lifetime, eg location, usage policies and configuration parameters; and (ii) dynamic information, which contains data on the operational state of the service, eg the set of properties that keeps track of events during a given sequence of interactions with a user, a device or another service.

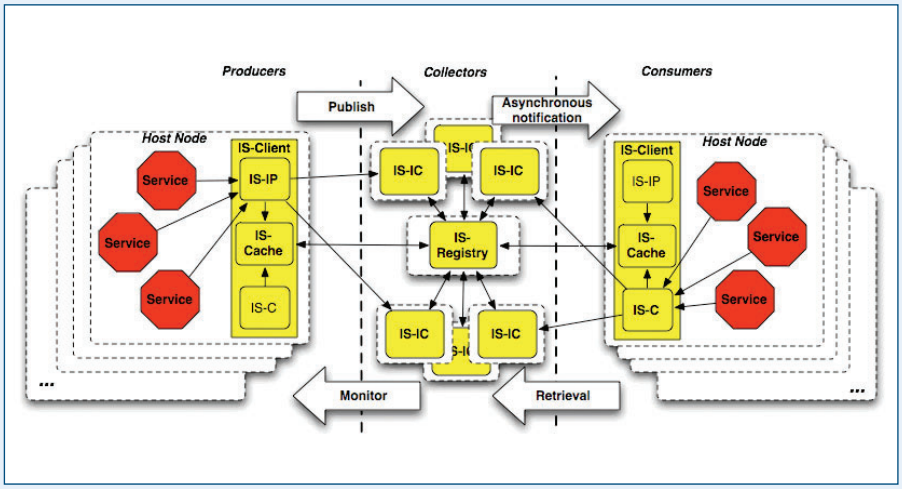
The DILIGENT Information Service

These features were recently included in the DILIGENT infrastructure, developed as part of the project of the same name. DILIGENT is a testbed infrastructure that will allow members of dynamic virtual eScience communities to create on-demand transient DLs based on shared computing, storage, applications, and multimedia and multi-type content. It is designed as a service-oriented architecture over Grid technology and relies on WS-* family standards, namely the WSRF framework and the WS-Addressing, WS-Security and WS-Notification specifications.

In this infrastructure, discovery and monitoring occur through a specific service, called an Information Service (IS), depicted in Figure 1. This service is organized in three logical parts, each serving the needs of a class of actors: information producers, collectors, and consumers.

Producers and consumers are supported in interacting with the IS via a lightweight component that is distributed on each hosting node of the infrastructure. This component is called an IS-Client, and supports three main features: (i) publication of the information (IS-IP library); (ii) access to information and discovery via querying and subscription/notification mechanisms (IS-C); and (iii) the local storage and maintenance of useful and constantly updated information (IS-Cache). The IS-Client allows information in the distributed infrastructure to be efficiently accessed and published, while hiding any detail of the routing process that could identify the collectors involved.

The collectors aggregate the producers' information. This part is composed of two components, the IS-Registry and the IS-IC. The former acts as a classical registry and maintains the list of available



Information Service Logical Architecture.

services and their static information. The latter maintains the dynamic information and is based on a highly distributed architecture.

From an operational point of view, it is important to note that each time one of the federation's services is deployed, it is first registered on the IS-Registry, and then starts producing its dynamic information via the local IS-IP. In parallel, the IS-Cache takes care of maintaining the set of minimal information needed by locally hosted services for both publishing and querying. The IS-Registry continuously monitors the service instances, thereby maintaining an overall 'picture' of the infrastructure in line with the actual status.

As well as designing this logical organization, we are currently evaluating and comparing various caching strategies and the distribution and selection algorithms for the IS-ICs. For instance, we are investigating the use of distributed information retrieval techniques like CORI.

Next Steps

The viability of the proposed approach will be further tested in the context of the forthcoming IST project: 'Digital Repository Infrastructure Vision for European Research – DRIVER'. The objective of DRIVER is to build a testbed for a future knowledge infrastructure of the European Research Area. Existing digital repositories spread over the Net will be federated, and a set of cross-repository services will be set up to provide seamless access to the DL content, regardless of which repository owns the content. Concretely, the project will start by federating 51 institutional repositories from The Netherlands, United Kingdom, Germany, France and Belgium. Each of these repositories will be considered as an element of the component-oriented infrastructure. Other components will provide digital library functionality, eg search and browse, personalized information access through recommendations, and virtual collections. In this context the Information Service will play a key role, since it will allow the other services to become aware of each other and to dynamically discover new repositories and services as they join the infrastructure.

This work would not have been possible without the help of colleagues at the NMIS Laboratory. Special thanks go to Davide Bernardini and Pasquale Pagano for their invaluable support in designing and developing this distributed and scalable Information Service.

Links:

Networked Multimedia Information System Laboratory website:
http://www.isti.cnr.it/ResearchUnits/Labs/nmis-lab/

OpenDLib website:
http://www.opendlib.com/

DILIGENT project website:
http://www.diligentproject.org/

BELIEF project website:
http://www.beliefproject.org/

Please contact:
Leonardo Candela – ISTI-CNR, Pisa, Italy
E-mail: leonardo.candela@isti.cnr.it

Semantic Search in Peer-to-Peer-Based Digital Libraries

by Hao Ding and Ingeborg Torvik Sølvberg

Advances in peer-to-peer overlay networks and Semantic Web technology will have a substantial influence on the design and implementation of future digital libraries. However, it remains unclear how best to combine their advantages in digital library construction. Research in the IF group at the Norwegian University of Science and Technology (NTNU) is evaluating possible solutions to advance developments in this field.

One of the most important features of the digital library of the future will be that it is accessible from anywhere, by anyone and at any time. Achieving this goal requires that the digital library be investigated as an integrated whole rather than as the sum of its individual parts. The approaches used in peer-to-peer overlay networks and Semantic Web technology show promises for aspects of communication infrastructure and semantic processing respectively. However, little work has been done to determine how best to combine these two technologies to form a total solution for digital library construction. NTNU researchers, under the framework of the IKT/WEB-TEK project sponsored by the Research Council of Norway, have developed a semantic search framework for peer-to-peer based digital libraries.

Our work, as illustrated in Figure 1, has involved comparing and identifying the strengths and weaknesses of both peer-to-peer and Semantic Web technology. Based on our analysis, we concluded that these two fields are complementary, and that there are great advantages to be gained by combining them in conducting semantic searches in a large-scale distributed environment. One major weakness in the current peer-to-peer systems is their limited search capabilities, which is due to their lack of power in responding to queries. The Semantic Web and ontologies as a semantic tool provide a basis for a shared understanding across a group of individuals,

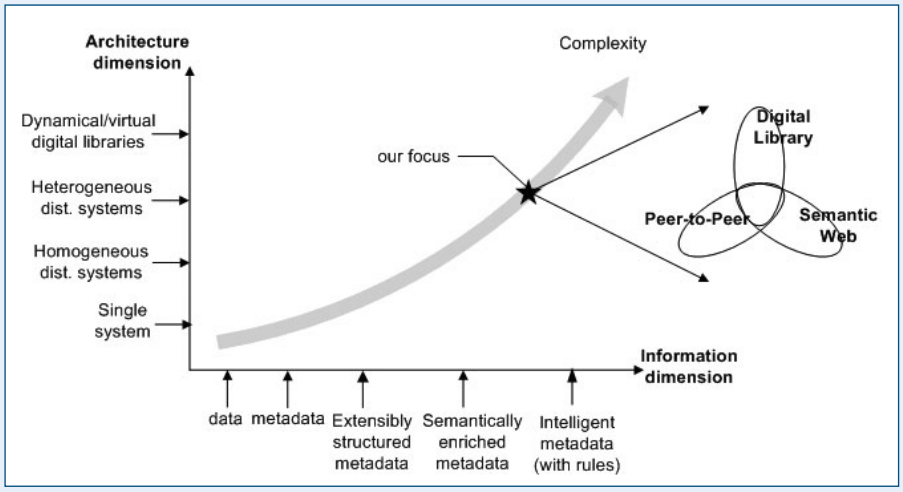


Figure 1. Combining P2P and Semantic Web for Constructing Digital Libraries.

such as in detecting similar concepts among ontologies and integrating multiple ontologies at no cost to the end users. By applying ontologies, the search capability in peer-to-peer networks can be strengthened via semantic information processing. The inference engine can also be specifically adapted to achieve more reliable results by deducing predefined rules.

However, while the Semantic Web and ontologies provide us with a mechanism for facilitating semantic information management and processing, they focus more on local and static situations, rather than a distributed and dynamic environment. Because they are innately decentralized, peer-to-peer systems can help exploit the full potential of the Semantic Web's capabilities. In other words, peer-to-peer systems can act as a fundamental platform for the searching and sharing of distributed information by using the Semantic Web technology.

In our survey of existing peer-to-peer systems, our project has concentrated mainly on scalability and autonomy. From a technical perspective, digital libraries need a common infrastructure that is highly scalable, customizable and adaptable. To this end, peer-to-peer systems have been suggested as one method for facilitating cooperation among digital libraries and for improving the accessibility of library services. Another critical goal of digital libraries is the sharing of resources with a wider audience. However, many inconsistencies exist across platforms, applications and capabilities. This means that library systems must often sacrifice autonomy to reach agreement with each other, so as to enable better searching and sharing of information. In comparison with client/server architecture, peer-to-peer systems provide a more open architecture by decentralizing the control from servers, allowing nodes (eg digital libraries) to be loosely coupled. As a consequence, system scalability and robustness can be improved with a small overhead in running specific communication protocols on these nodes.

As an intermediate goal, a tentative benchmark has been proposed for selecting an appropriate peer-to-peer networks for information searching in various digital library applications. In particular, our project has extended classic super-peer-based networks with load-balancing and self-organizing functionalities, thereby catering for dynamic situations that characterize digital libraries, such as continuous departures of peers, overload caused by the joining of peers, or even a system catastrophe. Evaluation results are illustrated in Figure 2.

In studying the use of Semantic Web technology to enhance search performance in digital libraries, this project investigates not just ontology-enriched metadata searching, but also the use of rules to express more complicated relations that exist among metadata records. We have compared the performance of a super-peer-based digital library by applying searches based on global schemas and ontology mapping. Currently, we are evaluating the potential introduction of rule-based reasoning in all applications.

Link:
http://www.idi.ntnu.no/grupper/if/

Please contact:
Hao Ding, NTNU, Norway
Tel: +47 73594168
E-mail: hao.ding@idi.ntnu.no

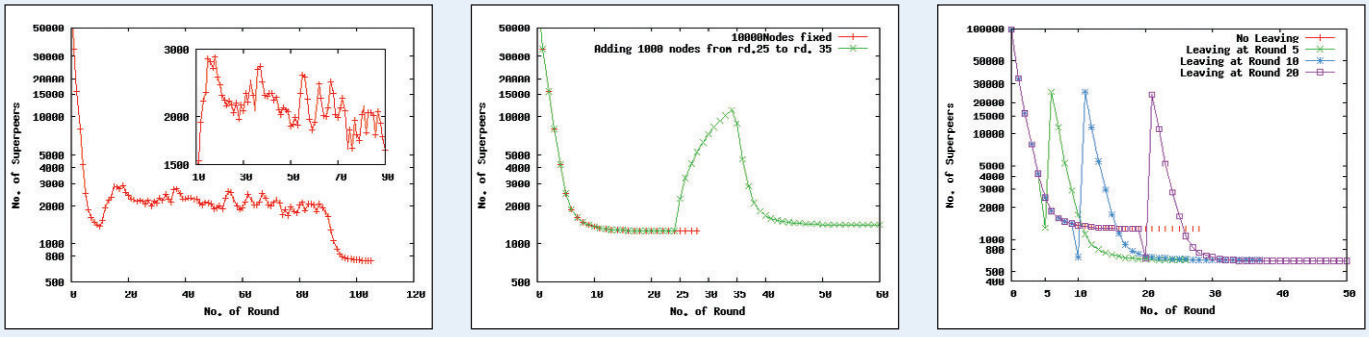


Figure 2. Evaluation Results: from left: (a) Self-organizing under a scenario of continuous leaving of peers; (b) Load-balancing under a scenario of continuous joining of peers; (c) Catastrophe Recovery.

XPeer: A Digital Library for the European Higher Education Area

by Mark Roantree and Zohra Bellahsène

The Bologna Process was initiated in 1999 by 29 Education Ministers of the EU, with the aim of providing a single European Higher Education Area (EHEA). Its goals were to promote mobility across EU states by standardizing the education programmes in all states, and by 2010, to provide a system of constructing degree programmes across multiple states. In effect, this means that the EHEA System becomes a vast digital library with the appropriate data-management functionality. This presents a major challenge to educational institutes as they prepare to modify their systems for incorporation into the EHEA Digital Library.

There are five broad goals to the Bologna Agreement: a system of easily readable and comparable degrees; a system with two study terms; the establishment of a system of credits; the promotion of mobility and European cooperation in quality assurance; and the promotion of the necessary European dimensions in higher education. To achieve this, the issues that must be overcome are integrating the very large number of databases that contain education programmes, dealing with the heterogeneity of the systems involved, and designing a single query interface to the EHEA Digital Library.

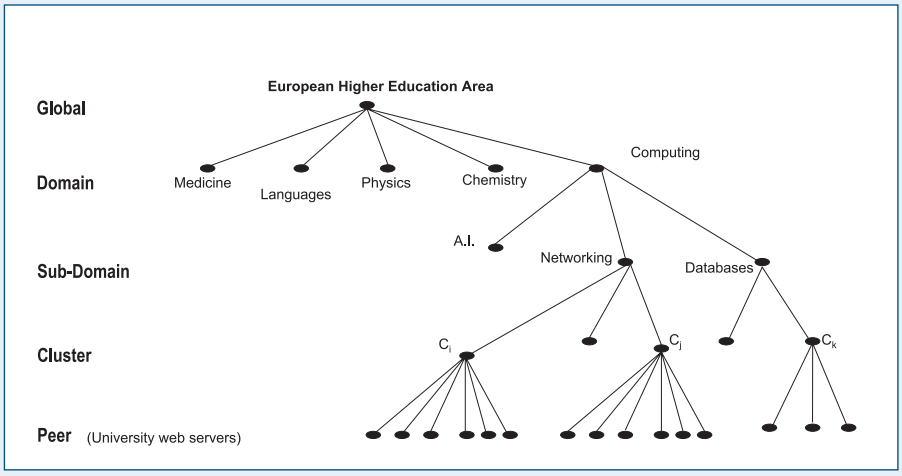
The XPeer Architecture

The XPeer Framework emerged from a collaborative effort between Dublin City University and the University of Montpellier to design a large-scale database architecture together with Query and Metadata Services. This research examined scenarios in which it was necessary to write complex queries for large numbers of heterogeneous pre-existing databases and information systems. This project faced the same issues as Bologna institutes: management of an extremely large system of databases where old information disappears and new information is made available on a regular basis. This is essentially what takes place when old courses or subjects are decommissioned and new ones emerge.

Data integration is a significant challenge: relevant data objects are split across multiple information sources and are often owned by different organizations. The sources represent, maintain and export the information using a variety of formats, interfaces and semantics. XPeer is a peer-to-peer (P2P) architecture with each peer representing a single database through an XML interface. In the Bologna context, a peer may contain a single course module or a group of modules within the educational institute and it is assumed that each institute provides any number of information peers. The XPeer Architecture is novel in that it adopts the 'super-peer' concept, which was first employed in networking terms to denote a peer of greater importance. In XPeer, we use this concept to integrate common peers, and thus define groups such as 'All modules in Europe for Final Year Computer Graphics students', 'Database courses in Europe delivered in English', 'Networking Courses in French Universities', or '1st Year Java Programming in Ireland, Greece and Germany'. In this way, we have introduced a two-tiered system with databases (peers) on one level and super-peers (integrated systems) on the second level. In Figure 1, the clusters C_i, C_j and C_k are formed using course module databases from various institutes.

Managing the Large-Scale Digital Library

The super-peer is the access point for users in the XPeer System, and for other super-peers that wish to communicate within the system (to resolve queries internally). It is the underlying P2P framework that supports the large-scale element to the system, as there is no central control point and therefore no bottleneck or limit to the number of participating systems. The super-peer concept permits the creation of 'clusters' of interesting systems for end-users. However, as the system grows, it is necessary to introduce a form of classification to assist with query optimization. As peers



XPeer framework for the EHEA Digital Library.

join the system, they are classified within the the existing set of domains. When XPeer is used to model the Bologna Process, these domains become the major disciplines such as medicine, computing, history and engineering. Thus, while the P2P approach allows for an infinitely large digital library, the classification process ensures that the library is ordered, and the integration process (super-peer) provides a global interface to collections of related databases. Thus, peers are accessed through super-peers; super-peers belong to a specified domain; and finally, a global peer (replicated to avoid performance issues) is used to manage the set of domains.

Current and Future Work

Current research efforts are focused on the provision of query and metadata services for the EHEA Digital Library. The Query Service uses the XPath Query Language, boosted by a fast indexing system and result-merging process to facilitate distributed querying on a large scale. The Metadata Service also uses the XPath language, extended to allow network creation, the addition and removal of peers, the promotion of common peers to the role of super-peer, the addition and removal of domains and so on. The project is expected to conclude at the end of 2008 with a set of fully specified services to support the large EHEA digital library.

Links:
XPeer publications:
http://www.computing.dcu.ie/~isg
Bologna Agreement:
http://www.euractiv.com/en/education/bologna-process/article-117448

Please contact:
Mark Roantree
Dublin City University, Ireland
E-mail: Mark.Roantree@Computing.DCU.ie

Zohra Bellahsène
University of Montpellier II, France
E-mail: bella@lirmm.fr

Increasing the Power of Semantic Interoperability for the European Library

by Martin Doerr

With the support of the DELOS Network of Excellence, IFLA and ICOM are merging their core ontologies. This is an important step towards semantic interoperability of metadata schemata across all archives, libraries and museums, and opens new prospects for advanced information integration services in the European Digital Library. The first draft of the combined model will be published in June 2006.

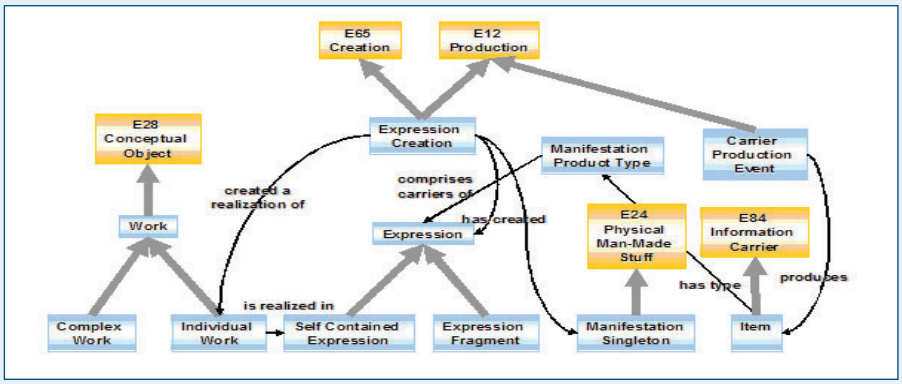
Semantic interoperability of Digital Libraries (DLs) requires compatibility of both the employed Knowledge Organization Systems (KOS; eg classification systems, terminologies and authority files) and of the employed metadata schemata. Currently, the notion and scope of DLs covers not only traditional publications, but also scientific and cultural heritage data. The grand vision is to see all these data integrated so that users are effectively supported in searching for and analyzing data across all domains. Even though the Dublin Core Metadata Element Set is well accepted as a general solution, it fails to describe more complex information assets. These include multimedia and learning objects, and data from characteristic domains such as archaeological finds or observational data from geosciences.

Core ontologies describing the semantics of metadata schemata are the most effective tool to drive global schema and information integration, and provide a more robust, scalable solution than tailored 'cross-walks' between individual schemata. Information and queries are mapped to and from the core ontology, which serves as a virtual global schema and has the capability to integrate complementary information from more restricted schemata. Many scientists question the feasibility of such a global ontology across domains. On the other side, schemata like Dublin Core reveal the existence of overarching concepts. Ideally, the European Digital Library would be based on one sufficiently expressive core ontology, not by selection, but by harmonization and integration of the relevant alternatives. The challenge is to explore practically the limits of harmonizing conceptualizations from relevant domains.

The CIDOC Conceptual Reference Model (CRM) has been developed since 1996 under the auspices of the International Committee on Documentation (CIDOC) of the International Council for Museums (ICOM) Documentation Standards Working Group. This is occurring with the initiative and support of ICS-FORTH, Heraklion, and the CRM is about to be accepted as ISO standard (currently ISO/DIS 21127) in 2006. It is a core ontology aiming to integrate cultural heritage information. It already generalizes over most data structures used by highly diverse museum disciplines, archives, and site and monument records. Even the common library format MARC ('MAchine Readable Cataloguing') can be adequately mapped to it. Its innovation is to centre descriptions not around the things, but around the events that connect people, material and immaterial things in space-time. Further, it explicitly describes the discourse on relations between identifiers and the identified, a powerful feature for the integration of information assets.

Quite independently, the FRBR model ('Functional Requirements for

Bibliographic Records') was designed as an entity-relationship model by a study group appointed by the International Federation of Library Associations and Institutions (IFLA) during the period 1991-1997. It was published in 1998. Its innovation is to cluster publications and other items around the notion of a common conceptual origin – the 'Work' in order to support information retrieval. Its focus is domain-independent and can be regarded as the most advanced formulation of library conceptualization.



Partial model of the intellectual creation process.

Initial contacts in 2000 between the two communities eventually led to the formation in 2003 of the International Working Group on FRBR/CIDOC CRM Harmonisation. It is headed by Martin Doerr from ICS-FORTH and Patrick LeBoeuf from BNF Paris, and brings together representatives from both communities. The common goals are to express the IFLA FRBR model with the concepts, ontological methodology and notation conventions provided by the CIDOC CRM, and to merge the two object-oriented models thus obtained. This Working Group is now being supported by the DELOS NoE, and in June 2006 will publish the first complete draft of $FRBR_{OO}$, ie the object-oriented version of FRBR, harmonized with CIDOC CRM. This formal ontology is intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation and interchange of bibliographic and museum information. Its major innovation is a realistic, explicit model of the intellectual creation process (see Figure). Work will continue with modelling information about authority records and performing arts.

The potential impact can be high. The domains explicitly covered by the combined models are already immense. Further, they seem to be applicable to the experimental and observational scientific record for e-science applications. From a methodological perspective, the endeavour experimentally proves the feasibility of finding viable common conceptual grounds even if the initial conceptualizations seem incompatible. Even though this process is intellectually demanding and time-consuming, we hope the tremendous benefits of nearly global models will encourage more integration work on the core-ontology level. A recent practical application of these models is the derivation of the CRM Core Metadata schema, which is compatible and similar in coverage and complexity to Dublin Core, but much more powerful. It allows for a minimal description of complex processes, scientific and archaeological data, and is widely extensible in a consistent way by the CRM-FRBR concepts. CRM Core can be easily used by Digital Libraries.

Links:
IFLA: http://www.ifla.org
ICOM: http://icom.museum
Definition of the CIDOC CRM:
http://cidoc.ics.forth.gr.
Definition of CRM Core:
http://cidoc.ics.forth.gr/working_editions_cidoc.html
Definition of FRBR:
http://www.ifla.org/VII/s13/frbr/frbr.htm
DELOS NoE deliverable 5.3.1:
http://delos-wp5.ukoln.ac.uk/project-outcomes/SI-in-DLs

Please contact:
Martin Doerr, ICS-FORTH, Greece
Tel: +30 2810 391625
E-mail: martin@ics.forth.gr

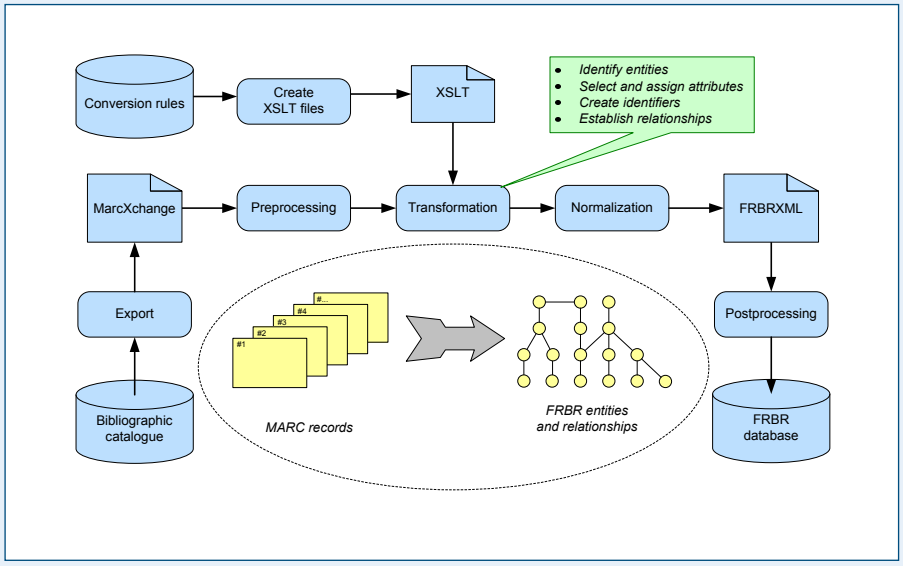
A Tool for Converting Bibliographic Records

by Trond Aalberg

The FRBR model for bibliographic information enables libraries to accommodate a broad range of user needs, and is considered to be an important step towards the next generation of library information systems. To support the application of FRBR in current library catalogues, solutions are needed to the problem of interpreting or converting MARC-based information. At the Norwegian University of Science and Technology, we have developed a conversion tool for this purpose.

The Functional Requirement for Bibliographic Records (FRBR) was published by the International Foundation for Library Associations and Institutions in 1998 and is widely acknowledged within the library community as an important contribution to the modernizing of library cataloguing and information systems. The ER-model proposed by the FRBR Working Group is a formal conceptualization of the entities, attributes and relationships of concern in bibliographic information. For the end user, the model promises to support a broad range of expectations and needs.

The heart of the FRBR model is a set of entities that represent the key objects of interest to users of bibliographic information. The products of intellectual or artistic endeavour that are named or described in bibliographic records are represented by the entities work, expression, manifestation and item. The entities person and corporate body represent those responsible for the content, production, dissemination or custodianship of the product entities. An additional set includes entities that serve as the subjects of works. For each entity a set of attributes is defined and the model includes an extensive set of possible relationships that may exist between the entities.



The process of transforming MARC records into a representation based on the FRBR model.

Although many projects have explored the use of FRBR in different contexts and some tools exist, there is little support for the systematic processing of all information in all MARC records into a proper representation that directly reflects the entities, attributes and relationships of the FRBR model. Due to a paucity of reusable solutions, researchers beginning work in this area typically need to reinvent the conversion process and write their own interpretation system. The transformation from MARC to FRBR is a complex task that in many ways is different from a simple transformation such as the conversion from MARC to Dublin Core. Entities such as work or person may have duplicate descriptions in numerous records, and to be able to create a consistent set of entities with a proper set of relationships, the process needs to be based on an extensive set of rules and conditions. The final output of the process should be a normalized set of unique entities with a proper set of attributes and relationships. Additionally, the conversion process needs to support solutions to numerous problems and exceptions. These may be caused by inconsistencies and errors resulting from erroneous registrations, data imported from low-quality sources or changes made to the catalogue.

This issue – that of transforming MARC records into a representation that directly reflects the FRBR model – has been investigated by the Norwegian University of Science and Technology in a joint project together with the Norwegian library service center BIBSYS and the National Library of Norway. The purpose of this project has been to support and explore the application of the FRBR model on existing MARC-based library catalogues. The project has two major goals: the identification and modelling of the various tasks required in a conversion process, and the development of a conversion tool that is based on the use of rules and conditions to define the transformation from MARC to FRBR. The tool is based on the use of XML, includes the automatic generation of the XSL transformation files used in the conversion, and the solution is independent of any specific MARC format and cataloguing rules. Because of this, the tool is reusable across catalogues and MARC formats. It uses records in the MarcXchange format as input and produces output in a format that is based on MarcXchange, but has specific attributes for describing the types defined in the FRBR model and elements for describing the relationships between entities. The conversion tool has successfully been used to transform the 4 million records of the BIBSYS database into an FRBR-ized prototype that is available on the Web. This prototype database is primarily intended to demonstrate the results of the transformation and can be used to search and navigate the BIBSYS database in the shape of FRBR entities and relationships. The actual conversion performed on this particular catalogue is still far from perfect, and the set of conditions and rules must be extended to support the exceptions and errors in the initial data. However, the tool enables librarians to easily specify and test various rules and conditions, and the overall result can be evaluated by inspecting or querying the resulting database.

The application of the FRBR model as a common ground for interchange and integration between libraries fits well with the current focus on cross-domain semantic interoperability in digital libraries. NTNU is participating in the DELOS NoE activity on development of the $FRBR_{OO}$ ontology, and future activities include adapting the conversion system to produce bibliographic information encoded as RDF using the $FRBR_{OO}$ ontology, for the purpose of cross-domain integration and interoperability using semantic Web technology.

Link:
BIBSYS FRBRized prototype:
http://november.idi.ntnu.no/frbrized

Please contact:
Trond Aalberg, Norwegian University of Science and Technology, Norway
Tel: +47 7359 7952
E-mail: Trond.Aalberg@idi.ntnu.no

Information Patterns for Digital Cultural Repositories

by Chryssoula Bekiari, Panos Constantopoulos and Martin Doerr

Digital cultural repositories emerge from the ever-increasing digitization of documents and images, digital photography, analogue-to-digital conversion of audio or video recordings, and digital transcription of object information recorded in various ways. Adding original digital cultural products and digital recordings of cultural information to these digital surrogates leads to an impressive collection of digital cultural material.

As digital collections are created independently by autonomous organizations, the emergence of a unified digital space is neither automatic nor easy. Aside from the legal and organizational issues, certain conditions are required for repositories to be interoperable. These are primarily concerned with documentation data and processes. Cultural documentation comprises a wide spectrum of information on the objects themselves, physical or informational, as well as related processes ranging from data acquisition to various scientific studies, conservation, exhibition design and publication. These processes may be separately documented and multiple relevant data sets may exist. If all this information is to be truly useful, we must ensure the ability exists to easily access and analyse information from disparate sources.

Interoperability has a syntactic and a semantic aspect. Syntactic interoperability is achieved by conforming to standards for information encoding and exchange. Semantic interoperability is the ability of different information systems to provide information consistent with the intended meaning. In practice, semantic interoperability aims to associate knowledge dispersed in various carriers and forms, thereby allowing related concepts to be automatically identified. To do this, standards for representing objects, functions and content must be adhered to, during both documentation and 'productive' uses of digital information.In order to build a framework for developing interoperable cultural digital repositories we follow a dual strategy. First, we draw on standard (meta)data structures recommended by established national and international bodies concerning archaeological, ethnological, museological, archival, geographical, terminological and digital preservation data for specific application areas. Second, all structures in the framework are related to a common ontology (namely CIDOC CRM of the International Council of Museums (also ISO Draft Standard 21127)) for long-term semantic interoperability, and are written in XML for syntactic interoperability. The CIDOC CRM is an ontology describing the concepts and relations involved in cultural documentation. It provides a common base for the interpretation of various forms of documentation, but does not dictate the documentation elements. Thus it plays a pivotal role in building interoperable digital cultural repositories.

Instruction in ancient Athens, red-figured attic vase, 5th c. BC, Berlin Archaeological Museum.

On this basis we define a set of information patterns, ie fundamental types of information unit such as time, place, object composition, event etc. This reduces the problem of designing cultural object records to one of designing a set of information patterns and a general, flexible record structure. As there are fewer information patterns than record fields, the design and the conformance with relevant standards are much more closely controlled. Rather than independent records, we thus obtain a family of records, conformant at the information pattern level. This allows different needs to be addressed and systematic data-entry procedures to be adopted, and ensures interoperability. Examples of information patterns are given in Table 1. Furthermore, for the description of museum objects and site monuments we have produced a comprehensive, CIDOC CRM-compliant, common

Date from until	**Object composition** number of parts part name kind code or cardinal number
Chronology within throughout cultural period social time justification	**Dating** chronology time measurement value method laboratory
Place name code cadastral number kind geopolitical hierarchy address coordinates values reference point precision of measurement geodesic coordinate system link to design	**Event** name kind chronology place description persons involved organizations involved objects involved comprises events
Person name biographical data communication data role/capacity/social group	**Organization** Title legal address communication data department role/capacity/social group

Table 1: Examples of information patterns.

XML DTD, resulting in a new, dynamic cultural object record of unprecedented genericity. Supporting a pragmatic object documentation workflow model, this object record can accommodate everything from minimal to highly detailed object information in structured and unstructured forms, thus gracefully adapting to the needs of various working contexts.

The XML DTD is available as a ready-to-use CIDOC CRM-compliant solution, along with higher-level guidelines and the CIDOC CRM ontology itself. Further work includes a number of application-specific extensions and, most importantly, domain extensions arising from current advances in harmonizing the CIDOC CRM with the FRBR (the bibliographic record model endorsed by IFLA).

This work, carried out at the Centre for Cultural Informatics, FORTH-ICS, was part of the compilation of guidelines for cultural digitization projects under the Greek Information Society Programme. It draws heavily on previous and concurrent experiences from several cultural information modelling and integration projects. It also benefits from the development of the CIDOC CRM ontology and the activities within the DELOS Network of Excellence on Digital Libraries, which are devoted to the harmonization of the CIDOC CRM and ABC (DELOS 1) and the CIDOC CRM and FRBR (DELOS 2) models.

Please contact:
Chryssoula Bekiari
ICS-FORTH, Greece
Tel: +30 2810 391631
E-mail: bekiari@ics.forth.gr

Towards a Semantic Information Platform for Subsea Petroleum Processes

by Jon Atle Gulla

The Integrated Information Platform project is defining an international semantic standard – or ontology – that will assist oil companies in making decisions and organizing collaborations. This article describes how the standard is developed and how it is used in intelligent information retrieval and reasoning.

The subsea petroleum industry is a technically challenging business with complex projects and operational structures. The projects are expensive, and they often include several large companies and span disciplines like drilling, reservoir, production, operations and maintenance. The European petroleum business is now facing a number of challenges that threaten its profitability. The costs of large older fields increase as they enter the decline phase, and new fields tend to be smaller and less scalable. We also produce more oil and gas than is added through exploration and improved oil recovery. For the Norwegian Continental Shelf for example, the addition and production numbers are at about 100 million Sm3 and 250 million Sm3 respectively. Finally, there has been an increase in the number of small and highly specialized service companies that need to collaborate closely both with each other and with the traditional bigger companies. All of this suggests that future petroleum projects need to be more cost-effective, make better use of small fields, and take advantage of the distributed and specialized skills of new service companies.

In 2004 the Norwegian Oil Industry Association launched an Integrated Operations program that proposed the use of new information and communication technology to integrate processes onshore and offshore. OLF's own estimates indicated that the implementation of this program on the Norwegian Continental Shelf would increase oil recovery by 3-4%, accelerate production by 5-10%, and lower operational costs by 20-30%. Central to this program was the semantic and uniform treatment of heterogeneous data, which originate from various disciplines and companies, at various locations, and with various degrees of precision and formality.

The Integrated Information Platform (IIP) project was initiated in June 2004 and is a collaborative project involving academic institutions and companies active on the Norwegian Continental Shelf. Led by Det Norske Veritas, the project has a budget of about 22 million NOK (about 2.8 million Euro), and is closely coordinated with various standardization efforts in the petroleum industry. The overall objective is to use semantic technology to improve decision-making processes and reduce risks and costs in petroleum projects. In particular, the project will result in an open platform that supports semantic interoperability and intelligent information management for subsea production systems.

IIP is now completing one of the largest industrial ontologies for the terminology used in the petroleum business. Support from central industrial partners on the Norwegian Continental Shelf has been secured, and the project intends to propose this ontology as part of a new ISO standard. As such, it will also be available to companies and institutions that are not currently part of the project. Parts of the new ontology have been converted from ISO 15926 Integration of life-cycle data for oil and gas production facilities, but after looking into the terminologies used in selected petroleum projects, we also included concepts from other ISO standards. More than 40 000 concepts have now been defined and modelled in

hierarchical conceptual structures. Figure 1 shows some of the concepts that must be defined just for the representation of wellheads. A Christmas tree is the set of valves, spools and fittings connected to the top of the well and used to control it's the fluid flow. The final ontology, which will be available in 2007 in OWL (Web Ontology Languages) with all the properties and rules incorporated, will be used to integrate petroleum applications, interpret real-time data from subsea installations, and access the information needed in decision processes.

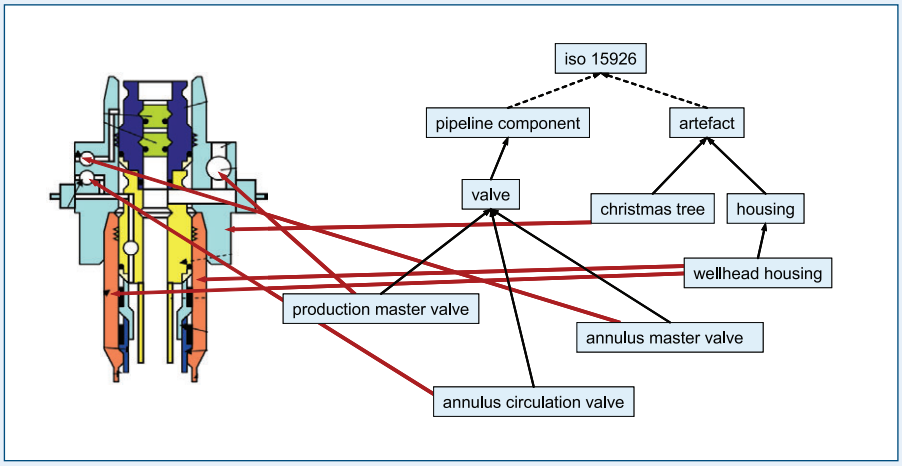


Figure 1: Wellhead with Christmas tree and associated ontology classes.

Finding information quickly is important in petroleum processes. With the vast number of sensors and amount of communication equipment added to new installations, the problem is more to do with relevance than lack of information. There is an overwhelming amount of information available, from project documentation to streams on real-time data from subsea installations, and it may all be relevant to decisions that need to be made quickly and accurately. The project is implementing an ontology-driven approach to searching that interprets the user's query in terms of ontological concepts and associates these concepts with weighted terms used in documents and data records. The idea is to let these weighted terms define the concepts with respect to both the information available and the preferences of the user. They may be constructed on the basis of a training set provided by the user and/or based on her past behaviour. As the user is interacting with the system, her behaviour is observed and the system refines its understanding of the user's perception of the concepts. Rather than using the ontological hierarchies to expand the search and simply increase the amount of retrieved data, the system builds personalized descriptions of ontological concepts that help us better rank the documents with respect to users' interests and needs.

Future work includes a rule-based notification component that will be used to analyse anomalies in real-time data coming from sensors measuring the production of oil and gas. Rules in OWL specify properties of the equipment used and what actions should be taken if a constraint has been violated. Since the ontology is specified in OWL Full however, it is not clear how the reasoning capabilities can be added. OWL Full does not lend itself to standard Description Logic reasoning. We are therefore looking into alternative representations of these parts or other ways of adding reasoning capabilities to our semantic information management framework.

Links:

NTNU: http://www.ntnu.no

Norwegian Oil Industry Association : http://www.olf.no

Please contact:

Jon Atle Gulla, NTNU, Norway

Tel: +47 73591847

E-mail: jon.atle.gulla@idi.ntnu.no

Towards The European Metadata Registry

by László Kovács, András Micsik and Jill Cousins

UKOLN and SZTAKI are partners with The European Library in a new research collaboration that aims to create a European Metadata Registry.

The existing European Library service is an ambitious and pioneering collaboration between European national libraries. It is supported by the EU and was created under the auspices of the Conference of European National Librarians (CENL). It offers a single access point to their holdings and spans a range of collections in all the partner National Libraries. This means that researchers or informed citizens in any country can use – in a single search session – not only the resources of their own national library but alsothose of any other partner national libraries. The European Library adds value to content by offering indexing services through the individual national libraries. It creates a pan-European platform and is a strategic initiative in European content enrichment.

To extend this work, the European Library has commenced a new collaboration with UKOLN (United Kingdom) and SZTAKI, which aims to create a European Metadata Registry (EMR).

This resulted in some software development and a prototype of a running registry service, which is capable of registering application profiles and their semantic connections. The reuse of profiles, terms, elements and encoding schemas is effectively supported by implementing an open platform for a collaboration of users from the CORES registry. A registry is single place to register new metadata terms, inspect terms already in use by different partners, and to propose new terms for properties that require them.

The need for this type of joint European registry is obvious: National Libraries in Europe apply different legacy metadata schemas, including Unimarc and Marc21. The European Library uses the TEL Application Profile, which is Dublin Core with extensions, in order to ensure interoperability when performing a search across libraries and collections.

The European Metadata Registry would provide a set of services:

- It would describe different metadata schemas and/or application profiles, as well as the aims, target audiences, application circumstances and scope of the schemas.
- It would represent internal semantic structure, the hidden model of schemas. Model descriptions aim both to understand and document the terms hierarchy. Because the metadata schemas of partners are based on different background semantical models, model mappings are non-trivial and require further scientific investigations.
- It would register tools and/or on-line services available for mappings, inferencing, translations, versioning and access.
- Finally, it would register semantic connections and relations between different schemas, thereby fostering the reuse of profiles, terms, elements and encoding schemas.

With these services the EMR will support the production of relevant crosswalks between legacy metadata schemas and/or The European Library schema. In addition, it will provide collaborative services for the development and improvement of new metadata schemas. Previous metadata registry approaches have enforced strict model-based methods in order to reach precise one-to-one schema mappings. This new approach will investigate a more relaxed mapping method, as well as a new collaboration technique - a Schemapedia. This is in the style of a Wikipedia, and is for metadata professionals who either need to develop schemas for their projects or to cross metadata access tools/services. The registry can grow organically on the basis of functional granularity and bilateral mappings. The scope and scalability of the registry are also under study, since the number of registered schemas and mappings cannot be predicted.

We intend the EMR to be a standardized, yet flexible and user-friendly tool. It will administer European metadata from different European cultural heritage communities including libraries, museums and archives, and its aim is to ensure transparency, access and interoperability.

Registry collaboration started with an evaluation of the available registry from The European Library. Some issues were uncovered and suggestions for its improvement have been collected. A functional implementation plan for this improvement is under development, with a likely implementation date of September 2006. At the same time, initial attempts to define the 'whys' and 'whats' of a European Metadata Registry have commenced, and are likely to be funded under the European Digital Library project in eContentPlus.

Link:
http://www.TheEuropeanLibrary.org

Please contact:
Jill Cousins, The European Library Office,
The Koninklijke Bibliotheek, National Library of the Netherlands
Tel: +31 70 314 0952
E-mail: Jill.Cousins@kb.nl

László Kovacs, SZTAKI, Hungary
Tel: +36 1 279 6212
E-mail: laszlo.kovacs@sztaki.hu

Personalizing Digital Library Access with Preference-Based Queries

by Periklis Georgiadis, Nicolas Spyratos, Vassilis Christophides and Carlo Meghini

Searching a Digital Library (DL) using traditional database (DB) or knowledge base (KB) query-answering techniques is constrained by the precision and completeness of answers: precision may lead to an empty answer, while completeness may result to a huge answer. Yet reformulating the query-filtering conditions to avoid one handicap may lead to the other: that is, relaxing the filtering conditions may lead to a huge answer, while strengthening the conditions may lead to an empty answer. Using preference-based queries to personalize user access to a Digital Library allows result sets to be tuned between these two extremes.

In general, a DL can be seen as a collection of documents residing in various information sources (on the Web). Each document can be naturally represented by a single row in the DL catalogue. A document's row contains its ID and description. The description involves a number of multi-valued attributes, called in the sequel DL columns. The values for

each column (referred to as terms of that column), come from a specified sort associated with the column, and may be organized in a hierarchy (eg taxonomy of subject headings). A traditional DB or KB query over the DL catalogue filters its rows according to the Boolean conditions on columns as defined in the query, but offers limited support for ordering the documents in the query answer.

In order to personalize DL access we advocate an enriched form of queries, called preference-based queries. A preference over a column C of the DL catalogue is any reflexive and transitive binary relation ? (preorder) over the terms of C. In other words, each pair t → t' of the relation denotes that t is preferred to t'. The presence of both t → t' and t' → t makes the terms t and t' equivalent, meaning that if a document described by t is not in the DL then one described by t' is an acceptable alternative (and vice versa).

A preference-based query comprises three parts:
- the term-filtering part *q*, which is Boolean conditions of terms (with an option of transitive closure due to the underlying term hierarchies)
- the preference part P which consists of preferences over the columns (ie preferences on the data level) as well as priorities over the columns (ie preferences on the schema level)
- optionally, the top-*k* part, ie a positive integer *k* denoting the maximum desired number of returned documents.

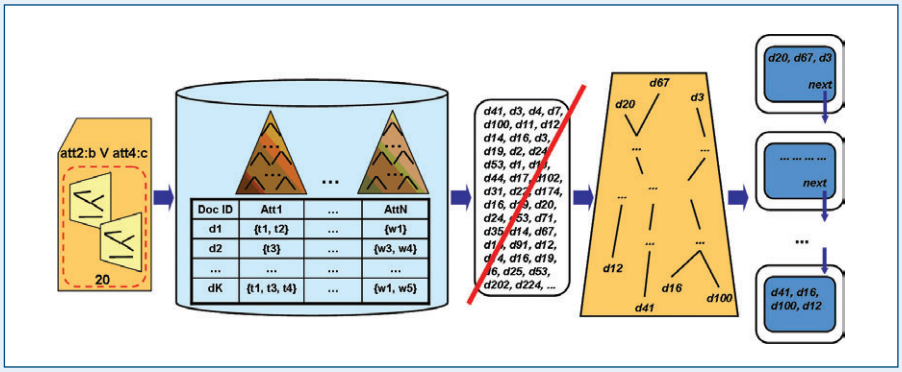
Of these three parts only the term-filtering part q is always submitted online. The remaining two parts can either be submitted online (together with *q*), or taken from a stored user profile and appended to *q* automatically. In either case such an access to a DL is a personalized access.

As DL columns are in general multi-valued, and if we consider different power domain orders, a partial order preference relation over the terms of one column may define a partial preorder over the documents in many ways. The choice is application-dependent, but in general the Hoare and Smyth relations are good candidates, considering that they preserve the initial order properties.

The evaluation of a preference-based query begins by computing the set of documents *ans(q)* which satisfies the query part q. Then, for each preference relation over a DL column a partial order over the set *ans(q)* is induced. As a partially ordered set of documents is not convenient to return as a response to a user query, we can employ variants of topological sorting to induce an ordered partition of the set *ans(q)*. This is a collection of mutually disjoint blocks of documents in a linear order that respects the partial order - and thus the initial user preferences. In this linear order each block would correspond to a screen of relevant documents that is shown to the user, with the most preferred documents appearing first. The parameter *k* is used as a stop condition to end the output process (when *k* documents have been shown to the user).

A user may express preferences over different DL columns and each preference incurs a different ordered partition. Combining preference relations over different columns and taking into account user prioritization over the columns boils down to defining a partial order over the Cartesian product of n partially ordered sets. This is a well-known problem for which various solutions exist, such as the lexicographic ordering or the Pareto orderings. In our case, we compute the product of all partitions and we use one of the known orderings on the topological distances of documents to generate the required final ordered partition.

The problem of preference-based queries for digital libraries is studied jointly by the Institute of Computer Science, FORTH, Greece; the Laboratory of Research in Informatics, LRI, France; and the Institute of Science and Technology in Informatics, CNR, Italy, within the framework of the DELOS European Network of Excellence in Digital Libraries. With respect to similar approaches, the main contribution of our work lies in the combination of ordered partitions coming from preferences expressed over multi-valued attributes, rather than over functional attributes describing traditional database tuples or objects. Our framework is expressive enough to produce sequences of documents from descriptions expressed in diverse data models (eg XML, RDF/S) with respect to a variety of user preferences, while also including priorities over the preferences.



A preference-based query over a Digital Library.

Link:
http://www.ics.forth.gr/isl

Please contact:
Nicolas Spyratos, LRI, University of Paris South, France
Tel: +33 1 69156629
E-mail: spyratos@lri.fr

Vassilis Christophides, ICS-FORTH, Greece
Tel: +30 281 0391628
E-mail: christop@ics.forth.gr

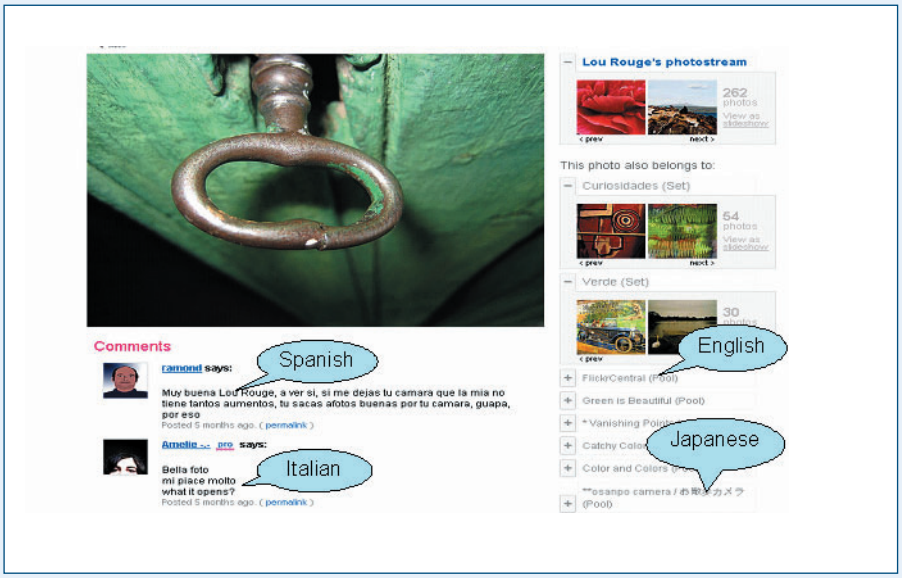
Carlo Meghini, ISTI-CNR, Italy
Tel: +39 050 3152893
E-mail: meghini@isti.cnr.it

Multilingual Interactive Experiments with Flickr

by Jussi Karlgren, Paul Clough and Julio Gonzalo

The Cross-Lingual Evaluation Forum (CLEF) in 2006 will feature a track on interactive image retrieval from dynamic target data taken from the popular Flickr photo-sharing service. In the past, interactive tracks at CLEF have addressed applications such as information retrieval and question answering. This year however, the focus has turned to text-based image retrieval from Flickr.

Information retrieval systems, especially text retrieval systems, have in the last few decades benefited greatly from a fairly strict and straight-laced evaluation scheme, which enables system designers to run tests on versions of their system using a test collection of pre-assessed data. These tests, based on the target notion of topical relevance, with system-oriented evaluation of performance, have served the text retrieval field well. However, system evaluation only addresses some of the bottlenecks in building a successful system.



An example photo from Flickr with multilingual annotations.

As a complement, experiments such as iCLEF – the interactive track at CLEF –aim to investigate real-life cross-language searching problems in a realistic scenario, and to give indications of how best to aid users in solving them. This crucially involves developing new evaluation methodologies and new target notions: relevance does not cover all the aspects that make an interactive session successful.

Over the past five years, the CLEF interactive track has studied various cross-language search tasks, including retrieval of documents, answers and annotated images. All tasks involve the user interacting with information systems in a language different from that of the document collection, and have been evaluated using conventional evaluation methodologies. This involves a fairly elaborate experimental setup.

This year we introduced some major changes. We want to find a collection where the cross-language search necessity arises more naturally for average users. We have chosen Flickr, a large-scale, Web-based image database serving a large social network of WWW users. It has the potential to offer both challenging and realistic multilingual search tasks for interactive experiments.

We want to use the iCLEF track to explore alternative evaluation methodologies for interactive information access. For this reason, we have decided to fix the search tasks, but to keep the evaluation methodology open. This allows each participant to contribute with their own ideas about how to study interactive issues in cross-lingual information access.

Additionally, we will lower the threshold for entry to attract more participants.

This year, the tasks given to participants are:

- *Topical ad-hoc retrieval over many languages:* find pictures of as many different European parliaments as possible.
- *Creative open-ended retrieval:* illustrate a short text on a given topic with five pictures (the text is provided separately to the experiment subjects).
- *Example-based retrieval:* determine the name of the place shown in a given photo.

The majority of Web image searching is text-based, and the success of such an approach often depends on reliably identifying relevant text associated with a particular image. Flickr is an online tool for managing and sharing personal photographs and currently contains over five million freely accessible images. These are available via the Web, and are updated daily by a large number of users. The photos are annotated by authors with freely chosen keywords in a naturally multilingual manner. Most authors use keywords in their native language; some combine keywords in more than one language. This sort of emerging, unsupervised and distributed semantic structure is known as a folksonomy and provides a modelling challenge for traditional knowledge-based retrieval approaches.

Participants will access images and metadata in Flickr through the open API

provided by Flickr, and are encouraged to log as many details as possible about every search session. A skeleton questionnaire will be provided to collect some of the evaluation metrics, and we will aim to probe notions related to user satisfaction and confidence:

- Satisfaction (all tasks): are you satisfied with how you performed the task?
- Completion (creative task, ad-hoc task): did you find sufficient results, or would you have continued if there had not been a time limit? How long would you have continued?
- Score (example-based task, ad-hoc task): one point for each relevant image found.

This year's workshop, held in Alicante in September 2006, will discuss the efficiency of search strategies, the usefulness of tested methods, and the utility of the projected evaluation methodologies.

Links:
CLEF: http://www.clef-campaign.org
Track home page: http://nlp.uned.es/iCLEF/
Data source: http://www.flickr.com

Please contact:
Jussi Karlgren, SICS, Sweden
Tel: +46 8 633 1500
E-mail: jussi@sics.se

Multimedia Ontologies for Video Digital Libraries

by Alberto Del Bimbo, Marco Bertini and CarloTorniai

A research activity is under way at the Department of Informatics, Florence University, which aims at creating a framework for the automatic annotation of soccer videos and the semantic retrieval of videos based on highlights and other high-level concepts.

Effective usage of multimedia digital libraries has to deal with the problem of building efficient content annotation and retrieval tools. Video digital libraries require annotation at the level of pattern specification in order to retrieve multimedia content according to specific user preferences and high level semantic content description. We have implemented multimedia ontologies, which include both visual and linguistic concepts, showing how they can be used for video annotation and retrieval and for the creation of user interfaces that accept complex queries, such as the visual prototypes of actions, their temporal evolution and relations.

Broadcasters need tools to annotate their video asset archives in order to exploit them to produce better TV programmes, and to lower the costs of indexing and search. Usually the video annotation process is carried out manually, using predefined vocabularies and taxonomies defined by the TV archivists.

The basic idea behind multimedia ontologies is that the concepts and categories defined in a traditional ontology are not rich enough to fully describe the plethora of visual events that can occur in a video. In fact, although linguistic terms are appropriate to distinguish event and object categories, they are inadequate to describe specific patterns of events or video entities. We are investigating the representation of events that share the same patterns by visual concepts, instead of linguistic concepts, in order to capture the essence of the event visual development. In this case, high level concepts, expressed through linguistic terms, and pattern specifications, represented through visual concepts, can be both organized into new extended ontologies that couple linguistic terms with visual information.

Using visual prototypes it is possible to group different video clips according to their visual features and at the same time classify them according to the linguistic high-level semantic concepts related to the visual prototype.

We have implemented a multimedia ontology for the soccer domain. A simplified schema is shown in Figure 1. Visual concepts for the different subclasses of 'Shot on Goal' are shown. The ontology is expressed using the Web Ontology Language OWL so that it can be shared and used in a search engine to perform content-based retrieval from video databases or to provide video summaries.

The creation process of the multimedia ontology is performed by selecting a representative set of sequences containing highlights described in the linguistic ontology, extracting the visual features and performing an unsupervised clustering. The clustering process, based on visual features, generates clusters of sequences representing specific patterns of the same highlight, which are regarded as specialization of the highlight. Visual concepts for each highlight specialization are automatically obtained as the centres of these clusters. Reasoning on the ontology is used in order to refine the annotation according to temporal and semantic relations between events.

The Multimedia Ontologies Annotator is the framework that allows users to import basic ontology schemas, generate the multimedia ontology, annotate video clips according to the ontology, and perform complex queries in order to retrieve videos containing specific visual concepts and other high-level linguistic concepts.

Figure 2 shows the interface of the Multimedia Ontologies Annotator.

It should be noted that users are able not only to browse, with a single interface, soccer and other video footage, but can also easily access the visual specifications of the linguistic concepts.

When users wish to see the different visual specifications of the linguistic

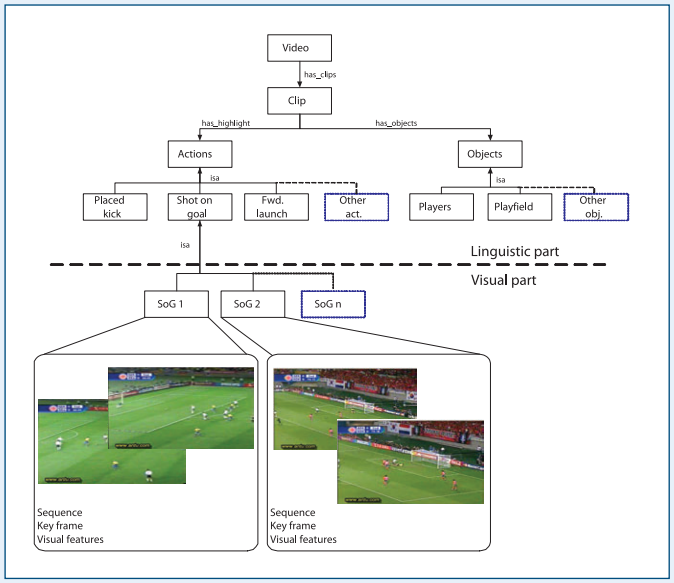


Figure 1: Multimedia Ontology (partial view).

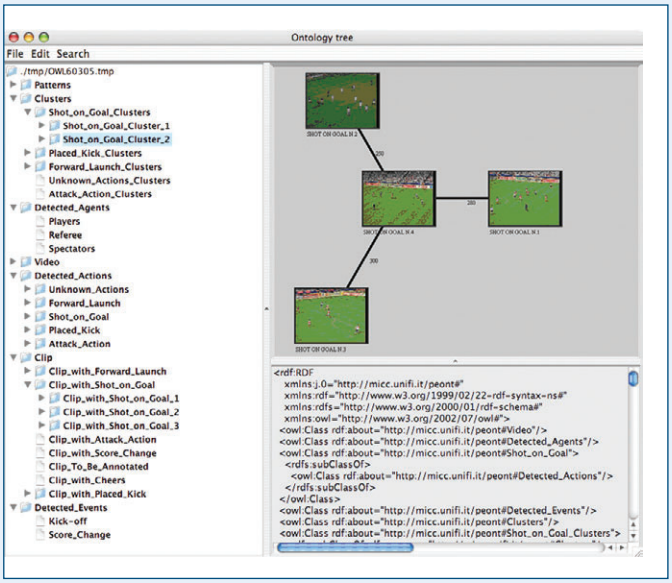


Figure 2: Multimedia Ontologies Annotator Interface.

concept 'Shot on Goal', they simply select the concept and the interface provides the clips that represent that concept. Moreover, a cluster view of similar visual concepts related to the linguistic concept is provided.

The queries that can be performed by the system involve both visual and high-level concepts. For instance a user can query for a sequence that starts with a forward launch, finishes with a shot on goal and contains a placed kick. He can also require that all actions are visually similar to a certain video clip or that all actions took place in a given location. The video produced should contain any type of attack action or placed kick visually similar to the selected models that occurred in soccer games played in the specified location.

Our future work will deal with the automatic generation of textual and vocal descriptions for video content based on visual features and temporal and semantic relations between concepts.

This work is partially supported by the Information Society Technologies (IST) Programme of the European Commission as part of the DELOS Network of Excellence on Digital Libraries.

Please contact:
Marco Bertini, University of Florence, Italy
Tel: +39 055 4796540
E-mail: bertini@dsi.unifi.it

Alberto Del Bimbo, University of Florence,
Tel: +39 055 4796540
E-mail: delbimbo@unifi.it

Carlo Torniai, University of Florence
Tel: +39 055 4237408
E-mail: torniai@micc.unifi.it

Structured Multimedia Description for Simplified Interaction and Enhanced Retrieval

by Stephane Marchand-Maillet, Eric Bruno and Nicolas Moënne-Loccoz

Multimedia management is a challenge common to several user groups, from individual users to corporate groups. It is therefore of high importance to define a platform that will resolve the contradictory issues of simplified description tasks and enhanced querying capabilities. We show that, based on technology common to Digital Libraries and the Semantic Web, we can propose such a framework.

The description of visual documents is a fundamental aspect of an efficient multimedia information management system. This is supported by the fact that a significant part of the information contained in the document can only be captured via the explicit description of a human operator. However, creating such a description is known to be both expensive and incomplete by nature. The importance of context may lead to ambiguous or even contradictory content descriptions. It is therefore critical that visual content description is done in the most favourable environment.

In the context of multimedia retrieval, MPEG-7 seems to be a good candidate as a description schema, since it caters for manual and automated description processes. However, MPEG-7 description tools are still largely not operational. Further, although based on XML, MPEG-7 hardly unifies with other schemes proposed along the Semantic Web route. Rather than pursuing the MPEG-7 direction therefore, we have

constructed a generic ontology-based annotation framework, in close relationship with the emerging W3C standards RDF (Resource Description Framework) and OWL Web Ontology Language. These developments are therefore aligned with the Semantic Web initiative, which shows that rich semantic annotations are needed for automating truly useful processes, including multilingual support.

Structured Description Model

The base of our framework is DEVA, a description scheme that has many favourable properties. DEVA acknowledges the power and utility of both the Dublin Core (DC), largely used in the Digital Libraries community, and the combination of RDF and OWL as description framework and knowledge management framework respectively, both arising in the frame of the Semantic Web.

The latest version of the Dublin Core vocabulary is composed of fifteen properties, among which we can find dc:title, dc:creator, dc:format or dc:subject. DEVA uses and extends these properties within the deva:Document wrapper class, and preserves their semantics. Being a Dublin Core extension allows the DEVA model to take advantage of a recognized standard, making it compatible with most of the software tools already available.

Wrapping DC elements has also been done in W3C's RDFPic tool using the PhotoRDF, where the subject property is constrained by a content schema. By contrast, in the DEVA model the subject property is associated with the Subject class. This class allows the semantic content of a visual document to be described. The specification of the content description is an OWL ontology designed according to the four expressive levels seen earlier:

- the Document class and its Dublin Core extended properties
- the Element class, representing the most significant elements of the scene
- the Property class, representing the properties attached to an element
- the Relationship class, representing the relationships between elements.

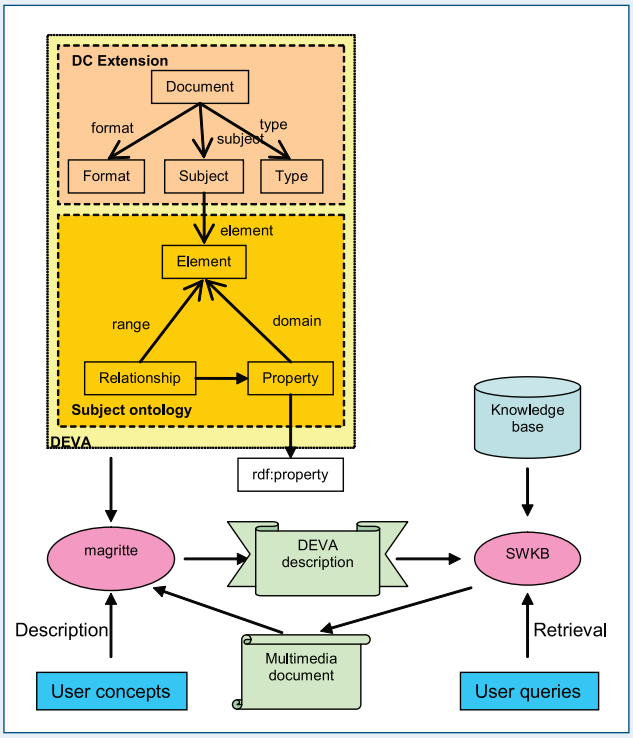


Figure 1: Architecture of the proposed structured document management system.

Interaction Principle

An annotation tool prototype, called magritte, has been implemented to evaluate our DEVA model and to validate its relevance in the frame of image annotation. The prototype is written in Java and makes use of the open-source Jena Semantic Web Framework developed by HP Laboratories. The interaction is simplified in many ways. The action of describing a multimedia document is almost reduced to the baseline classical keywording operation. However, the tool fully exploits the context in which keyword input takes place. Hence, every action is aligned against the knowledge base. Also, the definition of properties related to enumerable content is controlled by restricted choices. In this way, the document description is incrementally constructed by narrowing the scope of possible entries.

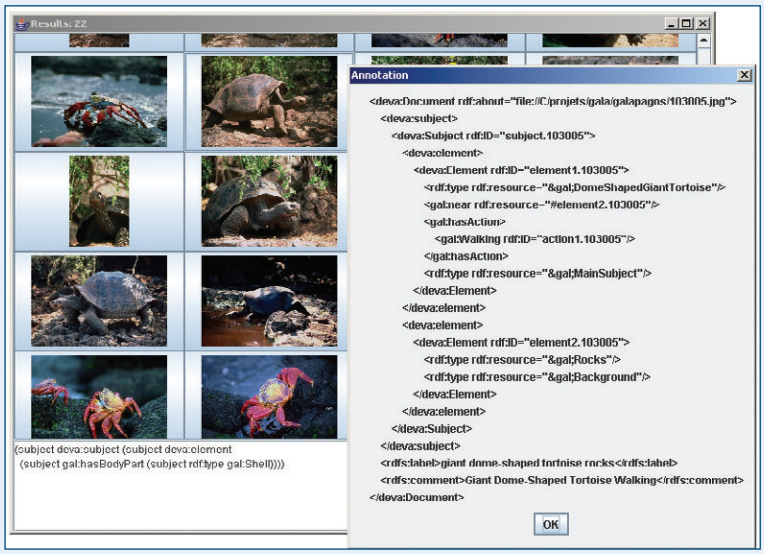


Figure 2: The SWKB demo tool illustrating the query "something with a shell".

Interaction Principle

A structured focused description is useful only when related to corresponding queries. However, in the context of an economical description, queries should, in principle, target known properties. For example, a picture whose subject is said to be a 'bird' would never respond to a query related to 'animal' in a direct search system. In fact, a 'bird' would be as close as a 'car' from such a query. The solution to this is clearly to make use of an external knowledge base to extend the 'bird' concept to that of an "animal".

This is the aim of the Semantic Web Knowledge Base (SWKB) that we have created to complement our description framework with a reasoning engine capable of processing high-level queries.

SWKB is an abstract framework embedding a reasoning engine to process DEVA (RDF) data against a classical OWL-based knowledge base using RDF/S semantics. By default, SWKB embeds the Jess (Java Expert System Shell) as a reasoning engine and may well be extended to other types of reasoning.

Extension

Thus, the above framework makes it possible to create multimedia descriptions aligned with typical assets found in digital libraries, while opening up the possibility of extended queries. We are currently extending it in directions related to content-based analysis of multimedia documents with auto-annotation as a bootstrap procedure, and the use of retrieval to organize the description at the collection level. This extension will partly take place in the domain of the Cultural Heritage asset management via the MultiMATCH project (http://www.multimatch.eu).

The support of the Swiss National Science Foundation is gratefully acknowledged.

Link:
Viper Group on Multimedia Information Management: http://viper.unige.ch

Please contact:
Stéphane Marchand-Maillet
University of Geneva, Switzerland
Tel: +4122 379 7631 / +41 22 379 7660
E-mail: marchand@cui.unige.ch

Taking a New Look at News

by Arne Jacobs and Nektarios Moumoutzis

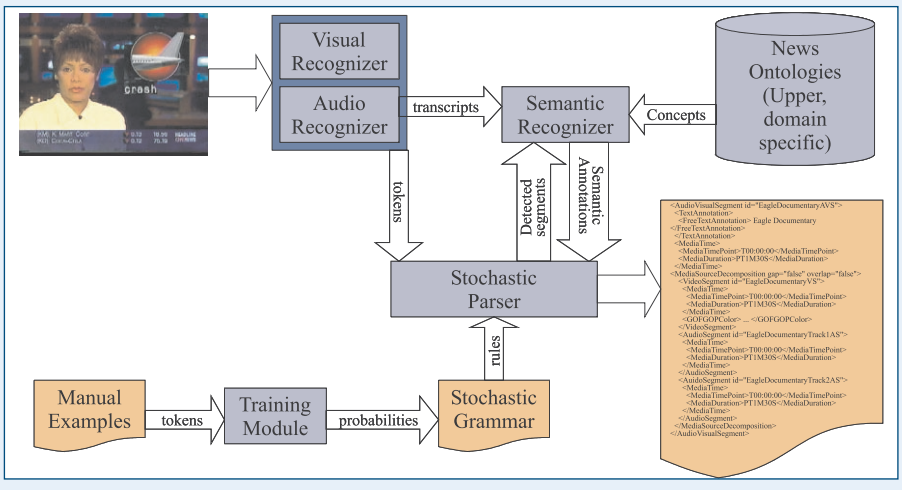
Although video search technology is making rapid strides forward, video search engines continue to be challenged by the semantic gap. This is the difficult problem of relating low-level features to the higher-level meaning that corresponds to the human-like understanding of video content, and a solution to it is necessary for effective retrieval performance. In the Delos Network of Excellence, specifically in task 3.9, "Automatic, context-of-capture based Categorization, Structure Detection, and Segmentation of News Telecasts", our approach to bridging this semantic gap is twofold. First, we restrict the application domain to news videos, and second, we exploit the combination of multimodal analysis with semantic analysis based on ontologies.

A key observation in bridging the semantic gap in the news video domain is that semantic concepts in news videos are conventionalized in many ways. This fact can be exploited. For example, segments in news telecasts do not appear in arbitrary order, but rather follow a relatively strict scheme that determines the order of segments. Different news stories are also usually separated by anchor shots containing the presentation of the story that follows. This telecast structure allows the viewer to easily recognize different segments. Each news format has its own structural model.

We assume that these news format models can be described with context-free grammars. Such a grammar can aid the segmentation process by removing ambiguities in classification, or by associating certain audiovisual cues with segment classes (eg news story, presentation). It models all interesting constraints in visual, audio and textual features, that are partly due to the way news programs are produced and partly to the habits and preferences of journalists and other agents related to the news production.

For parsing of a news video following a corresponding model, we propose a system consisting of two main types of interoperating units: the recognizer unit consisting of several modules, and a parser unit. The recognizer modules analyse the telecast and each one identifies hypothesized instances of 'events' in the audiovisual input. Such events can be higher-level concepts like a specific person appearing in a shot (eg the anchor), the appearance of a certain series of frames (eg the introduction sequence, with which many news broadcasts commence), or low-level concepts, eg the similarity of two frames.

The system contains three distinct recognizer modules: the audio recognizer,



The system architecture shows the interoperation between regognizers, grammar, and parser.

the visual recognizer and the semantic recognizer. The visual recognizer identifies video events in the news stream, such as a face appearing at an expected position in the video or the presence of a familiar frame according to the expected structure of the broadcast. The audio recognizer identifies audio events such as the presence of speech or music, the detection of predetermined keywords, and clustering of speakers. Finally, the semantic recognizer identifies the semantics involved in the telecast. This includes topic detection, high-level event detection, discourse cues and possible story segmentation points. The figure shows a sketch of the system architecture.

The recognizers normally only communicate with the parser in a one-way communication, providing a sequence of predetermined event 'tokens'. However, in the case of the semantic recognizer there may be an exception, since that module requires a transcript of the telecast in order to perform its analysis. In the case where the transcript is not provided through the input (eg in the form of closed captions), the audio recognizer provides this information.

A stochastic parser using a probabilistic grammar analyses the identifications provided by the recognizers. In essence, the recognizers provide the parser with actual lexical tokens just as a lexical analyser would provide to a programming language parser. The grammar represents the possible structures of the news telecast, so the parser can identify the exact structure of this telecast. When the parsing is complete and all the structural elements of the input have been analysed, the semantic recognizer uses that information to identify story topics and events, and to assign all required semantics to the structure tree.

The grammar for each broadcast station, even for different news programs of the same station, is distinct. This is because the grammar captures the directing elements of the broadcast, and no two programs have exactly the same directional structure. Therefore, a grammar must be produced manually for each program examined.

To determine the probability values of the rules in the grammar, it is necessary to (currently manually) complete a training process, which uses a set of correctly labelled news recordings in the form of a sequence of tokens.

Finally, the semantic recognizer has access to an upper ontology, covering all necessary aspects required for multimedia content description, as well as domain-specific ontologies created for news. The concepts acquired from these ontologies will define those detectable semantics that can be identified in the telecast.

We are currently investigating methods for automatically determining the audio-visual cues that characterize a given news format, by analysing a set of example recordings of that format. Based on the experience that these audiovisual cues do not change frequently, we expect videos from the same format to have many nearly identical video and audio sequences at similar time-points. We are trying to exploit this by using an inter-video, intra-format similarity analysis. In the future, this will overcome the limitation of manual creation of format models.

Please contact:
Arne Jacobs,Universität Bremen, Germany
E-mail: jarne@tzi.de

George Ioannidis
Universität Bremen, Germany
E-mail: george.ioannidis@tzi.de

Radio Relief: Radio Archives Departments Benefit from Digital Audio Processing

by Martha Larson, Thomas Beckers and Volker Schlögell

The archives departments of radio broadcasters are currently facing face two significant challenges, namely, how to store rapidly increasing amounts of radio content, and how to satisfy the rising demand for easy retrieval of audio clips that can be recycled into new programs. A pilot project demonstrates that digital audio processing techniques have the potential to provide much-needed support.

Radio broadcasters rely on highly specialized staff to archive broadcast content and respond to requests from journalists and editors for audio content on certain topics. As radio expands rapidly into the digital world, the amount of radio content produced and the demand for a convenient way to access this content for recycling has been growing at a rate that threatens to overwhelm archives departments. The pilot project Audiomining is being undertaken in Germany by Westdeutscher Rundfunk (WDR) and Deutsche Welle (DW) in cooperation with the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS). It has developed an audio archive prototype that demonstrates that automatic audio processing methods have a clear and concrete potential to provide critical support for archivists, journalists and editors in the face of these challenges.

Currently, many radio broadcasters maintain extensive databases containing annotations of analogue radio recordings, painstakingly compiled by the archive staff. When the archives department receives a request from a journalist or editor, the metadata in these databases is searched and the corresponding analogue recording can be located in storage. Information concerning the recorded content that is not noted in the annotations is effectively 'lost' in the archive, since it

cannot be retrieved. As radio broadcasters move towards completely digital workflows, it becomes possible to use automatically generated metadata to supplement the human-produced annotations.

The Audiomining prototype system provides both an indexing interface – which allows archivists to load new radio content into the system for processing – and a search interface. The search interface allows archivists not only to retrieve programs from the archive using titles and production dates, but also to type in keywords, which are then searched for in speech recognition transcripts. This option means that the content of radio broadcasts is directly searchable. The search interface returns a hit list, and individual hits can be opened with a simple click in the graphic audio browser. The audio browser displays a radio program as a series of cuts corresponding to segments of the program containing music or speech. Those containing speech are further divided into segments spoken by the individual speakers, who are assigned speaker index numbers. The graphic audio browser displays keywords that have been found in the radio program at their relative positions, and it is possible to click on keywords and jump into the audio at the exact point when the keyword is spoken.

The interfaces of the Audiomining system were developed in close cooperation with archivists from WDR and DW. The project blended tried-and-true techniques used by the archives departments with new digital audio technology in order to created a concept for a new integrated workflow, which would provide comfortable and intuitive support for archivists for both annotation and retrieval of radio content. Archivists feel that the structured browsing offered by the graphic audio interface will allow them to listen to radio programs in a targeted way, using their annotation time to concentrate on adding high-level semantic labels to targeted radio segments. The keyword search also has clear potential to help archivists locate sections of radio broadcasts, in particular interviews that are relevant to user requests.

The indexing module of the Audiomining stand-alone prototype produces metadata in MPEG7 format. First, it uses audio segmentation, based on the well-known Bayesian Information Criterion, to determine boundaries at which the quality of the audio changes (for example at a speaker turn). It then applies a classifier that separates speech from non-speech, which is generally music. In the next step, it groups all the speech segments into classes that are acoustically similar. These classes correspond to speakers and are assigned a speaker index. Finally, the speech segments are sent to the speech recognizer for the generation of speech recognition transcripts on which the keyword search is carried out.

The Audiomining project is in its final evaluation stage and has accomplished its goal of demonstrating that digital audio processing technology can be smoothly incorporated into archivists' workflows. Automatic systems will not replace human archivists in the foreseeable future. However, the potential inherent in automatic structuring and audio keyword search demonstrates promise to provide significant relief for radio broadcasters, who are inundated with audio content and sorely in need of techniques that make spoken audio as easily accessible as text.

Please contact:
Martha Larson, Fraunhofer
Institute for Intelligent Analysis and
Information Systems (IAIS), Germany
Tel: +49 2241 14 1980
E-mail: martha.larson@iais.fraunhofer.de

Thomas Beckers
WDR Dokumentation & Archive, Germany
Tel: +49 221 220 4799
E-mail: thomas.beckers@wdr.de

Volker Schlögell, Deutsche Welle Archive-
Bibliothek-Dokumentation, Germany
Tel: +49 228 429 4368
E-mail: volker.schloegell@dw-world.de

Self-Organizing Distributed Digital Library Supporting Audio-Video

by László Kovács, András Micsik, Martin Schmidt and Markus Seidl

The StreamOnTheFly network combines peer-to-peer networking and open-archive principles for community radio channels and TV stations in Europe. StreamOnTheFly demonstrates new methods of archive management and personalization technologies for both audio and video. It also provides a collaboration platform for community purposes that suits the flexible activity patterns of these kinds of broadcaster communities.

Community broadcasters are non-profit, are open to the general public and have a local or regional scope. They provide access to radio and TV production facilities for organizations, groups and individuals aiming to make their own programs or shows. Community radio stations and TV channels would like to be able to archive, exchange and stream content over the Internet, but appropriate tools are not yet available at an affordable price. They have therefore expressed a desire to set up a robust distributed infrastructure with a technological solution that is flexible, effective and low-cost.

In 2002, the StreamOnTheFly project with the participation of Public Voice Lab (Austria), SZTAKI and Team Teichenberg (Austria) obtained grants

from the European Commission's IST Programme to build a middleware application for radio with various front ends. StreamOnTheFly was focused on next-generation audio content management and broadcasting offering a customizable community radio program. It is now able to demonstrate new methods of management and personalization technologies for both audio and video. It also provides a collaboration platform that suits the flexible activity patterns of local broadcaster communities.

In a time when access to broadband Internet and the demand for online video material are rapidly increasing, video compatibility has become an important criterion for online media archives. With this in mind, the development of the video extension was initiated in January 2006 by a group of developers at the University of Applied Sciences St. Pölten.

In the case of video, an emphasis was placed on compatibility with a wide variety of user devices, such as mobile phones, iPods, PDAs and Sony's Playstation Portable. The realization of this goal was achieved with the help of ffmpeg, the leading open-source video transcoding tool. Another enhancement of the archive was the adaptation of StreamOnTheFly's RSS (Really Simple Syndication) functionality to make video podcasting an integrated part of the archive.

StreamOnTheFly Architecture

The original StreamOnTheFly architecture consisted of three main components:

- station control: a Web application to manage radio broadcasting, programme scheduling, and the archiving and publishing of selected programmes to node servers
- node server: a distributed network of servers containing the archived programmes with their metadata, access statistics and usage history; users may browse for interesting materials, and compile personalized radio streams for listening
- portal server: personal or community selections of content can be presented in a customized format; the archived content is revitalized through various subjective filters.

The core of the StreamOnTheFly network combines peer-to-peer networking and open-archive principles, while other services are realized as separate network components communicating through open interfaces (APIs). There is no central server in the network, and metadata is exchanged in a peer-to-peer manner. The content is stored on the node of the publisher while all content metadata is available on all node servers. This enables fast searching and browsing with reasonable storage requirements.

A simple exchange format called XBMF (Exchange Broadcast Binary and Metadata Format) is a core element of the network. XBMF enforces the coupling of metadata with content, and also creates the possibility of providing content in different audio formats, or attaching different media (eg images, text) to the audio content.

During the evolution of the StreamOnTheFly network the node server component (archive) remained in focus, while the other two components (station control, portal) were made replaceable using standard communications. A plug-in for free, professional, radio station management software is under development: this will support the archiving of programmes on node servers. RSS 2.0 or OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) are used to promote archived content to portal engines (eg Typo3, Manila, Plone) and search engines. Via podcasting, selected content can be easily transferred to PDAs, iPods and other handheld devices.

StreamOnTheFly in Operation

The StreamOnTheFly network is operational since October 2003. The core of the network (five network nodes are up currently) is accessible in four languages (English, German, French and Hungarian), and contains more than 1800 hours of audio content stored in 1700 programmes (http://radio.sztaki.hu). Other applications of the software include an exhibition, school and fair radios, e-learning and linguistic pro-

Figure 1: StreamOnTheFly Node Server Interface.

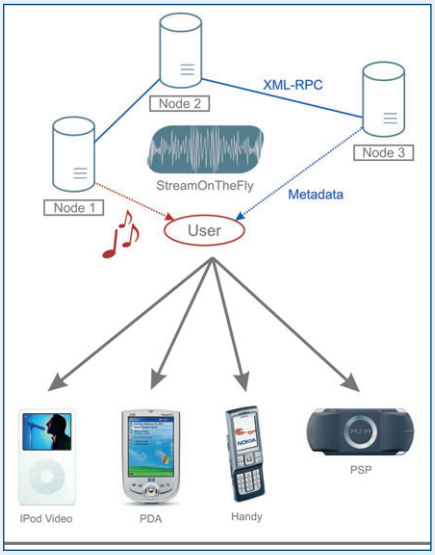


Figure 2: StreamOnTheFly Network Protocols and Access Methods.

jects. The StreamOnTheFly software suite is open-source and licensed under the GNU General Public License.

Although the StreamOnTheFly project officially finished in June 2004, voluntary work is continuing to implement various extensions and use cases. Since that time StreamOnTheFly has proved its flexibility and extensibility in several cases. The architectural decisions made in StreamOnTheFly justified themselves as an open and scaleable base for further developments.

Links:
Project home page:
http://www.streamonthefly.org
The Hungarian node: http://radio.sztaki.hu

Please contact:
László Kovács, SZTAKI, Hungary
Tel: +36 1 279 6212
E-mail: laszlo.kovacs@sztaki.hu

Markus Seidl
University of Applied Sciences
St. Pölten, Austria
Tel: +43 2742 313 228 - 245
E-mail: markus.seidl@fh-stpoelten.ac.at

Repositories and Preservation in the UK

by Neil Jacobs

In recent years, universities and colleges have begun to think more carefully about the management of their intellectual output. Influenced by enlightened self-interest and prompted by the encouragement of funding and regulatory bodies, these higher education institutions are making greater efforts to showcase their research. One of the potential tools in their armoury is the repository, which allows digital objects to be managed locally and accessed globally. In the UK, the Joint Information Systems Committee (JISC) is investing some £13.8m over three years to make this a reality.

JISC's vision is to establish a network of digital resources and services to improve content use and curation. Through a significant investment programme, JISC will build on work on the 'Information Environment' undertaken by UKOLN. This has resulted in a national architecture that provides easier discovery of, and access to, digital content. However, significant development is still required in a number of areas, eg in preserving digital content.

A new JISC development programme will fund initiatives to develop the Information Environment and support digital repositories and preservation, including cross-searching facilities across repositories. It will also fund institutions to develop a critical mass of content, and will provide preservation solutions and advice for the development of repositories. The programme builds on existing JISC work, in particular the 'Digital Repositories' programme and the 'Supporting Digital Preservation and Asset Management in Institutions' programme.

The new programme comprises the following areas:

- Digital Repositories: projects will develop repositories for universities and colleges. This includes a major project that will aid higher education institutions in establishing and developing repositories.
- Digital Preservation: the programme will develop a distributed environment for digital preservation, in which services, roles and responsibilities are defined.
- Discovery to Delivery: this includes a searching service across UK repositories, and development projects to achieve agreement on standards for searching and semantic interoperability.
- Tools and Innovation: the programme will develop new software and tools which will lead to innovative approaches to repository use and digital preservation.
- Shared Infrastructure: in support of both national and international developments, the programme will develop shared infrastructure services such as user profiling services, digital rights management, registries, identifier services, terminology and preservation services.

If these are the instruments, then what will be the outcomes? Firstly, institutional repository services will be created, enhanced and (perhaps most importantly) populated. It will be possible to search in increasingly sophisticated ways, based on a range of effective and practical interoperability standards. Such standards will also underpin preservation services, and universities and national bodies will share the responsibility for preservation. We will have a much clearer idea of how repositories should be used to support education and research. Pilots and demonstrators will illustrate the potential in this programme, and software and tools will make it practical.

It is anticipated that the programme will result in a range of benefits for universities and colleges, including an increased capability to manage intellectual property for education and research, and an infrastructure that will support the sector into the future.

Link:
http://www.jisc.ac.uk/

Please contact:
Neil Jacobs, Joint Information Systems Committee, UK
Tel: +44 117 33 10772
E-mail: n.jacobs@jisc.ac.uk

Digital Library of Historical Newspapers

by Martin Doerr, Georgios Markakis, Maria Theodoridou

A management system for historical newspapers that supports both digital library functionality and archival management of original newspaper articles is being developed for the needs of the Vikelea Municipal Library of Heraklion. It includes OCR-based page analysis and article clipping, article-level metadata generation, semantic indexing and multifaceted classification of articles using a built-in thesaurus. We aim to improve the classification, completeness and precision of retrieved information - supporting both metadata and full-text searching - and to provide user-friendly Web access.

An important part of the study of historical newspapers consists of classifying the material and annotating it such that its future retrieval is made easier. The system has a variety of goals, including supporting the preservation, documentation and study of historical newspapers. It also aims to protect people from exposure to potential health hazards and to assist in the production and dissemination of electronic versions of publications, thereby promoting cultural education.

The structural particularities of digitized newspaper documents pose a significant challenge in creating an efficient digital library system interface. A newspaper page consists of articles (text blocks), pictures and advertisements that refer to a variety of real-world events, activities, actors and/or objects. Consequently, the page itself is not the basic conceptual unit of information and is therefore not suitable for a thorough metadata-based description of the material. Instead we focused on the notion of the segment as a basic conceptual unit. A segment may consist of one or more parts of the newspaper document that are conceptually relevant (ie an article, a group of articles or advertisements etc).

The historical newspaper management system implements a 'hybrid' form of classification and searching based on the following elements:

- user-generated metadata for each annotated segment of the original newspaper based on the CIDOC Conceptual Reference Model ISO/DIS 21127
- full text of the annotated segment of the newspaper produced by an OCR (optical character recognition) session.

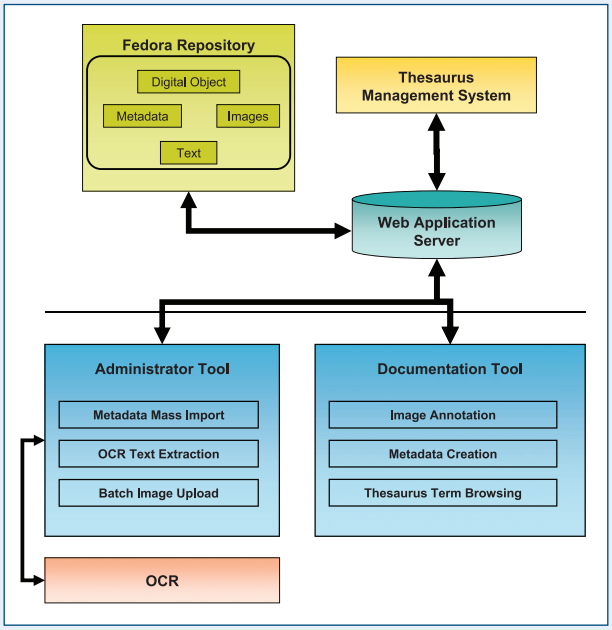
The large volume of the material that needs to be digitized and classified poses another important challenge. The system will be used to digitize approximately 100.000 pages. Given the fact that each page generally contains between five and twenty articles, we need to create an efficient and flexible interface as well as a mass import/OCR mechanism in order to reduce the time and cost of the digitization process.

The historical newspaper management system consists of the following subsystems:

The Digital Library deals with the management of the archival catalogue and information on the contents of the newspapers. It therefore supports thematic indexing and classification based on concepts retrieved from appropriate thesauri.

At the core of the Historical Newspaper Digital Library is the Fedora open-source digital repository system, which is a flexible content repository system that provides organizations with flexible tools for managing and delivering their digital content. Fedora is jointly developed by Cornell University and the University of Virginia Library.

The functionality of the digital repository is enhanced by the use of SIS Thesaurus Management System, which is a semantic network used to store, develop and access multiple thesauri and their interrelations under one database schema. The semantic interoperability of the digital repository with the thesaurus management system aids users in classifying and retrieving newspaper articles.



Historical newspaper management system architecture.

The Documentation Tool provides an efficient Web-based user interface for the insertion, filing, documentation and classification of material, and follows international standards for information modelling and interoperability.

We have created a flexible, easily deployable and user-friendly Web interface for this system to enable the researcher to isolate a specific conceptual entity within the document and perform an on-the-fly creation, description and storage of the produced metadata.

In addition to the creation of the segment, the system performs an

extraction of the text included in the annotated segment of the document and stores it for full-text search purposes.

Graphical terminology visualization techniques enable the user to annotate the document according to appropriately developed thesauri. The combination of thesauri visual graphs and auto-complete algorithms significantly reduces the time needed for the creation of metadata and supports the efficient sharing of knowledge among the members of a community of annotators.

The Administrator Tool allows the mass storage of digitized material (JPEG images) into the digital repository, and the transformation of this material into a format that can be annotated and indexed by the experts via the documentation tool.

The historical newspaper management system is currently being used in the Vikelea Municipal Library of Heraklion to upload a significant part of the historical archive of newspapers and magazines regarding the history of Crete.

Link:
http://www.ics.forth.gr/isl/cci.html

Please contact:
Maria Theodoridou, ICS-FORTH, Greece
Tel: +30 2810 391 731
E-mail: maria@ics.forth.gr

DML-CZ: Czech Digital Mathematics Library

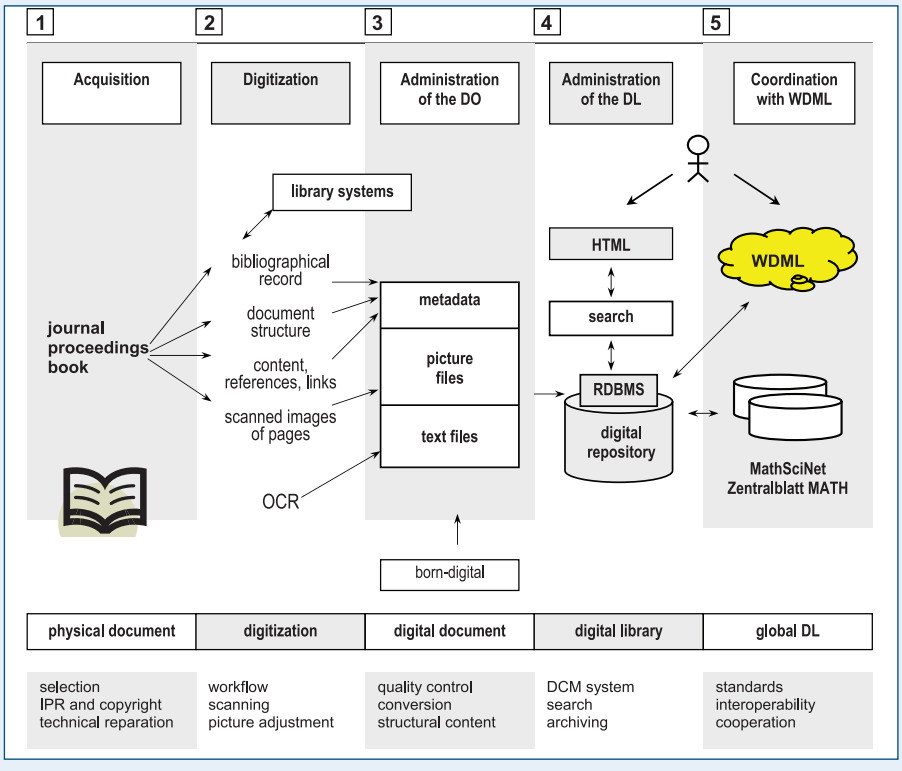
by Jirí Rákosník

Mathematics, much more than any other area of science, depends on access to literature that may be tens of or even hundred years old. The rapidly increasing extent of this kind of information makes efficient searching and navigation difficult, especially if the majority of works remain accessible only in paper form. Contemporary scholarly literature is commonly available in electronic form online, which enables information to be stored, organized, searched and accessed in a digital environment. It would be highly advantageous if this were also possible for the older body of literature.

A number of recent projects worldwide – JSTOR and NUMDAM, for example – were set up with the aim of digitizing historical mathematical literature. Having different initiatives working on the same problem might result in many different formats and interfaces. To avoid such mess, discussions started in order to define common standards and best practices. In addition, conditions were set for interlinking the individual projects in an ambitious system called the World Digital Mathematical Library (WDML). The entire mathematical literature is estimated to consist of approximately 50 million pages.

Encouraged by these activities, the Czech Mathematical Society initiated a national digitization project called DML-CZ: Czech Digital Mathematics Library (see http://dml.cuni.cz for more details about the project and other digitization initiatives). Proposed for the period 2005–2009, it is supported by the Academy of Sciences of the Czech Republic within the framework of the national research programme Information Society.

The aim of the project is to investigate, develop and apply techniques, methods and tools that would allow the creation of a suitable infrastructure and conditions for establishing what will become the DML-CZ. It will consist of the historical mathematical literature published in the Czech lands, and upon completion it will be incorporated into the WDML. The project will involve launching the digitization process and providing end users with access to the digitized material. It will also involve research into advanced technologies for searching



The proposed scheme of the DML-CZ.

mathematical documents, and for including both existing and future 'born-digital' materials. Presumably, in view of the common history and lingual similarity, suitable Slovak mathematical literature will also be included.

Creating an adequate digital library is a complex task and requires numerous problems to be solved. These include the following areas, which will be tackled within the project:

- Acquisition: technical preparation of materials to be digitized; intellectual property and copyright issues.
- Digitization: setting technical parameters compatible with the WDML Best Practice Statements; setting the digitization workflow; selection and adaptation of software supporting the digitization process; OCR processing and post-processing; provision of metadata.
- Digital documents: Digital Objects structure specification; defining standards for descriptive, structural and administrative metadata; global persistent identification; archiving and presentation formats; conversions between formats and generation of digital derivatives; inclusion of born-digital materials; automatic conversions of visually marked OCR data into logically structured documents.
- Digital library: implementation of the Content Management System; providing access to the digitized material; interlinking the content with the reference databases ZMATH and MathSciNet; research and implementation of advanced search techniques; the DML-CZ administration including long-term preservation of the digital content.
- Integration of the DML-CZ in the WDML.

The testbed for the DML-CZ is being built upon digitized documents from the Czechoslovak Mathematical Journal. The electronic material created within the DIEPER project (Mathematica Bohemica and Commentationes Mathematicae Universitatis Carolinae) offers another possibility.

The complexity of this task requires the expertise of specialists in distinct fields. The team therefore consists of five groups from different institutions:

- Mathematical Institute AS CR, Prague (project co-ordination, selection and preparation of materials for digitization, IPR and copyright issues)
- Institute of Computer Science, Masaryk University, Brno (technical integration, development of the digital library for the DML-CZ, metadata provision coordination, incorporation of the DML-CZ into the WDML)
- Faculty of Computer Science, Masaryk University, Brno (OCR post-processing, techniques for searching and presenting digital documents)
- Faculty of Mathematics and Physics, Charles University, Prague (user requirements, metadata specifications, links to ZMATH and MathSciNet)
- Library of the Academy of Sciences, Prague (digitization, OCR, storage and presentation of digitized content within the Kramerius digital library system).

Links:

http://dml.cuni.cz
http://www.jstor.org
http://www.numdam.org
http://gdz.sub.uni-goettingen.de/dieper

Please contact:
Jirí Rákosník, Mathematical Institute AS CR, Prague, Czech Republic
Tel: +420 221403446
E-mail: rakosnik@kav.cas.cz

CASPAR and a European Infrastructure for Digital Preservation

by David Giaretta

The preservation of digitally encoded information is a difficult task, requiring long-term commitment and collaboration. CASPAR, a new EU FP6 Integrated Project, addresses this problem. Together with other major European initiatives, it will form the basis of a continent-wide preservation infrastructure, and will benefit both current and future users. CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) is an EU Integrated Project, which began in April 2006 with a budget of around 16 MEuro (8.8 MEuro from the EU).

One of the challenges currently facing modern society is the vast amount of intrinsically fragile digital information upon which it is increasingly becoming dependent. CASPAR intends to address this problem by building a pioneering framework – based on existing and emerging standards – to support the end-to-end preservation 'lifecycle' for scientific, artistic and cultural information.

The ambitious goal is to build up a common preservation framework for heterogeneous data, along with a variety of innovative applications. This will be achieved through the following high-level objectives:

- establishing a foundation methodology applicable to an extensive range of preservation issues
- researching, developing and integrating advanced components to be used in preservation activities. These components will be the building blocks of the CASPAR Framework
- creating the CASPAR framework: the software platform that enables the-building of services and applications that can be adapted to multiple areas.

The CASPAR consortium will demonstrate the validity of the CASPAR framework through heterogeneous testbeds.

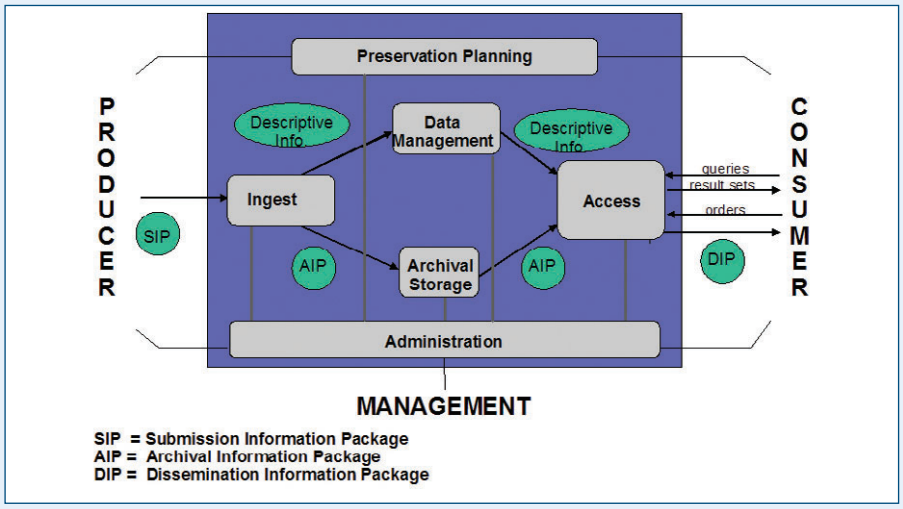


Figure 1: OAIS functional model.

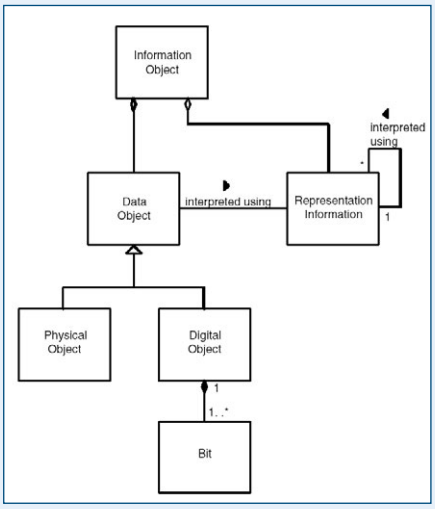


Figure 2: OAIS information model.

These will cover a wide range of disciplines from science to culture, contemporary arts and multi-media, and will provide a reliable common infrastructure that can be used or replicated in other areas.

CASPAR proposes a set of tough metrics by which it, and any other project which claims to be doing something useful for digital preservation, may be judged.

The CASPAR consortium will also seek to guarantee the future evolution of CASPAR in the following ways:

- the CASPAR preservation user community will be built to create consensus around the initiative and gather a critical mass of potential users
- the CASPAR framework and components will be embedded within key memory organizations, both national and international.

To achieve this, CASPAR brings together a consortium covering important digital holdings, with the appropriate extensive scientific (CCLRC – the lead partner and ESA), cultural (UNESCO) and creative expertise (INA, CNRS, University of Leeds, IRCAM and CIANT). This is combined with commercial partners (ACS, ASemantics, MetaWare, Engineering, and IBM/Haifa), experts in knowledge engineering (CNR and FORTH) and other leaders in the field of information preservation (University of Glasgow and University of Urbino).

Models

The Reference Model for an Open Archival Information System (OAIS, ISO 14721) which forms the basis of CASPAR contains a number of models, including a functional model (Figure 1) and an information model (Figure 2).

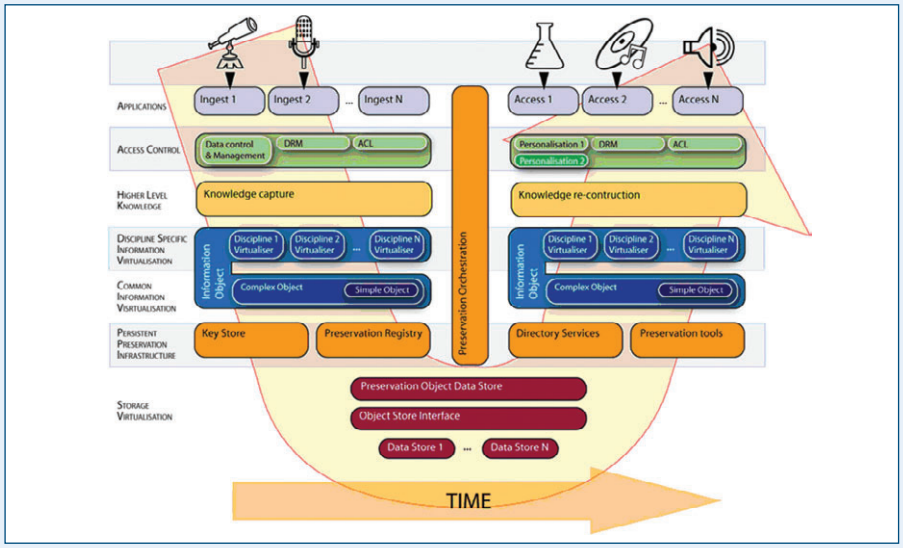


Figure 3: CASPAR virtualization model.

CASPAR adds to these a high-level model of virtualization and a number of high-level components.

The components of infrastructure that CASPAR will produce must themselves be preservable. To this end the project will put 'knowledge' at the heart of preservation. By this we mean that besides simple data semantics, CASPAR will also capture higher-level semantics. Furthermore, we will use Semantic Web techniques to enable the infrastructure components to survive changes over time.

Regardless of how successful CASPAR is as a project, it nevertheless has a limited life. In order to provide long-term support we aim to embed CASPAR results into the production processes of long-lived organizations such as CCLRC, ESA, UNESCO and INA, as well as many related archives.

In addition, the Task Force on Permanent Access to the Records of Science has produced a research programme and strategic plan, the former being consistent with that of CASPAR. Part of this strategic plan is to create an 'alliance' consisting initially of major data holders across Europe. Members of the alliance can, among other things, seek to align their individual infrastructures to form the basis of a Europe-wide preservation infrastructure. It is also hoped that a European Digital Information Infrastructure for Preservation and Access (EDIIPA) will be added to the ESFRI Roadmap, to further embed these activities.

Immediate Benefits from Digital Preservation

While many reasons exist for preserving digitally encoded information, a large proportion – such as legal requirements – are transitory. Longer-term reasons tend to be very worthy (eg for the good of future generations) but do not fare well in competition with other activities that seek support from cash-limited funders. In addition, benefits are hard to quantify.

Yet an immediate benefit can be identified, as long as the preservation is successful. The OAIS view is that the test of preservation is that digitally encoded information should remain comprehensible and useful for future users to whom that data is unfamiliar. However, potential users exist right now to whom the data is unfamiliar. Pulling current data from the Internet (eg for use in a GRID application) has many analogies with retrieving archived data, and indeed it may be hard to distinguish between the two. While it is true that for current data it may be possible to communicate with the data producer, it would be much more convenient not to rely on that but to have automated processes that can use the data correctly.

The virtualization techniques needed for preservation can in many cases provide exactly that capability. They also offer the opportunity to support generic applications that can deal with data from any source, by using the appropriate virtualization information.

We believe CASPAR to be the first project with the aim of producing broadly applicable components and a framework for digital preservation. The confluence of events at a European level offers the opportunity for CASPAR to make more than an ephemeral contribution, even when measured on the long timescales of data relevance. Furthermore, the techniques being adopted offer immediate benefits to current users.

Links:

CASPAR: http://www.casparpreserves.eu
DCC Development: http://dev.dcc.ac.uk
Task Force on Permanent Access: http://tfpa.kb.nl
OAIS Reference Model: http://public.ccsds.org/publications/archive/650x0b1.pdf
Digital Curation Centre (DCC): http://www.dcc.ac.uk
ESFRI: http://cordis.europa.eu/esfri/

Please contact:
David Giaretta, Associate Director (Development) Digital Curation Centre and co-ordinator of CASPAR project
Tel: +44 1235 446235
E-mail: d.l.giaretta@rl.ac.uk

PROBADO – Non-Textual Digital Libraries put into Practice

by Thorsten Steenweg and Ulrike Steffens

In the PROBADO project, librarians and computer scientists are collaborating to produce workflows, systems and tools that enable libraries to professionally handle non-textual documents alongside their traditional textual documents.

According to a study at the University of Berkeley, the information produced globally in 2002 amounts to 5 billion terabytes. Information on paper only represents 0.001% of new information recorded in all media and is very often simultaneously stored in digital format. Meanwhile, the relevance of non-textual digital documents is increasing. This is obvious in our private lives, where digital cameras and music downloads are becoming ubiquitous. However it is also true for professionals, such as architects who produce and combine digital 2- or 3D graphical models of monuments, musicians and composers who create and reuse digital audio recordings, or teachers who work with e-learning material. This last example also illustrates another development: digital documents are becoming more and more complex, ie they may consist of a variety of partial documents, possibly based on different types of media.

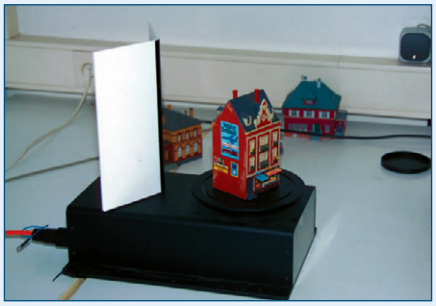


Figure 1:
Acquisition of 3D content by scanning.

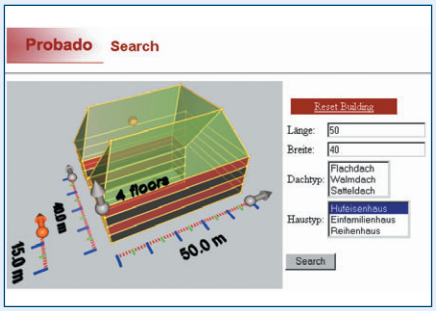


Figure 2: Interactive 3D search interface.

Digital libraries bring together proven expertise in typical library workflows and technical know-how on large, distributed information systems. Hence, they are also candidates for the management of complex, non-textual documents. However, today's libraries are mainly associated with the provision of literature and texts. Furthermore, they usually offer no way of taking into account a document's internal structure. For instance, it is possible to retrieve whole books, but there is no means of accessing single chapters or illustrations.

The PROBADO project started in February 2006 and is being conducted by the University of Bonn, the Technical University of Graz and the OFFIS Research Institute, as well as by the German National Library of Science and Technology in Hannover and the Bavarian State Library in Munich. It aims to support libraries in profession-

ally handling non-textual, complex documents alongside their traditional text documents. The resulting information system will be the basis for a sustainable operational library service, which provides access to non-textual documents for scientists and professionals. Initially, PROBADO will provide services for music, 3D graphics and e-learning content. The underlying digital library system is, however, highly generic. Mechanisms to extend the PROBADO services to different media types will be devised in future project activities.

The challenges to be met by PROBADO can be best explained along the workflow typically implemented by a scientific library:

- In the acquisition phase, documents in various formats are collected and brought into the library.
- In the indexing phase, the documents are prepared for usage by the library patrons. Information on content, form and bibliographic data is deduced and included into the library's catalogues.
- Indexed documents can be retrieved from the library by search or browse facilities.
- A retrieved document can be accessed in some way by the library patron.
- Finally, archiving activities ensure the long-term availability of the library's documents.

Although this workflow is well understood for text documents, it raises new requirements if non-textual documents are also to be managed. PROBADO users will for example expect to be able to search for 3D models of buildings with Gothic windows, for pieces of music containing a certain musical theme or melody, or for e-learning material for students in the first year. To support them PROBADO has to offer enhanced content-based indexing and retrieval methods as well as advanced, flexible user interfaces.

In the area of music, score images are analysed by Optical Music Recognition algorithms and are later synchronized with the respective audio recordings. Among other things, the user interface enables the user to type in note representations or whistle or hum a theme into a microphone. The music index is used to retrieve pieces of music matching the user's request. It can highlight the requested part within the score and synchronously play its audio interpretation.

In the area of 3D graphics, a catalogue is developed of basic architectural shapes, which are then used to index architectural 3D models. Users can then search the model database by giving a textual description like 'buildings with Doric columns', by choosing a basic shape from the catalogue and interactively parameterizing it, or by sketching the architectural shape they are interested in. Search results can be adequately rendered in a 3D browser.

In contrast, e-learning content cannot be restricted to certain media types, and semantically combines different media in ever-changing formats. Hence, PROBADO is developing extensible indexing and retrieval algorithms. These allow existing content-based retrieval methods for different media types to be integrated and enriched, enabling searching by didactic aspects.

The PROBADO project is funded by the German Research Foundation and also collaborates with the DELOS Network of Excellence, ensuring European dissemination. The project has a tentative duration of five years.

Link:
PROBADO home page:
http://www.probado.de

Please contact:
Ulrike Steffens, OFFIS, Germany
Tel: +49 441 9722 176
E-mail: ulrike.steffens@offis.de

Jochen Meyer, OFFIS, Germany
Tel: +49 441 9722 185
E-mail: jochen.meyer@offis.de

WIKINGER - Semantically Enhanced Knowledge Repositories for Scientific Communities

by Lars Broecker

While many scientific communities use the Internet for the exchange of scientific knowledge, it is only rarely used for the collaborative creation of it. The WIKINGER project is working on semantically enhanced knowledge repositories that support the collaborative generation of knowledge by providing a semi-automatically generated semantic net of the topics contained. A Web application is being developed using front-end building on Wiki technology.

As the success of the Wikipedia project shows, collaborative knowledge creation on the Internet is possible, even viable. This is an interesting result, taking into account that the user base generally acts anonymously and is spread all over the world. However, disadvantages exist to the Wiki approach to knowledge creation, especially for scientific communities. First, there is the problem of attaining a critical mass. The domains are often highly specialized, leading to only small numbers of people interested in (or even qualified for) participation, which makes attaining a critical mass of information very difficult. Second, there is a problem in the way HTML handles the linking of pages. Hyperlinks are one-way only and do not carry any semantics besides 'go there from here'. This becomes a problem as soon as the need

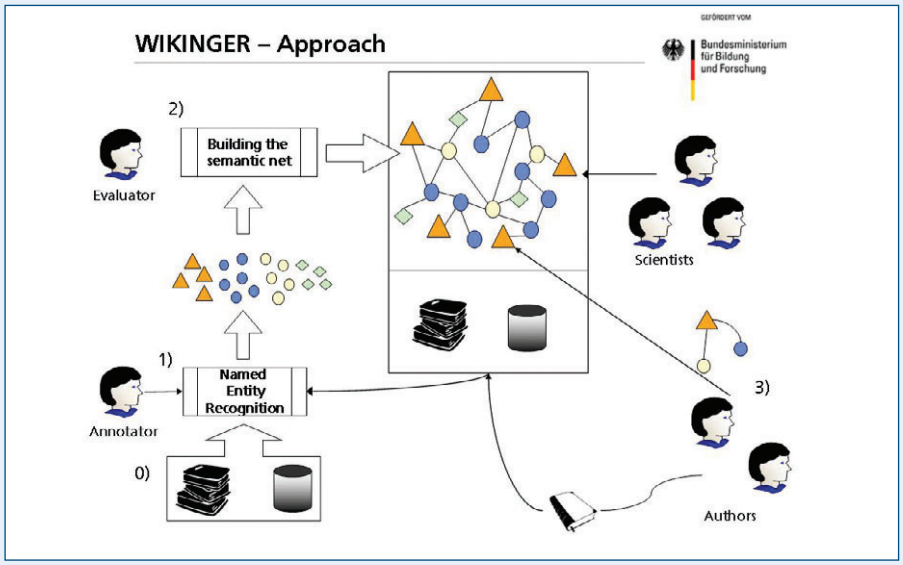
to assign semantic labels to associations arises, eg in order to enable more sophisticated search tools than full-text retrieval.

The scenario illustrated above is typical for academic communities, especially in the humanities. In general, a variety of publications exist that deal with special facets of the discipline. Each of these contains a multitude of pieces of information on people, institutions, places and events, as well as the associations between them. Unfortunately, organized knowledge repositories available in digital format are rare. This problem is recognized in the community, as the efforts necessary to find these pieces of information among the publications grow. Such information would be well suited for publication in a Wiki system, provided that a) the process of identifying articles and their relationships can be automated to a high degree, and b) that the problem of missing semantics in hyperlinks can be solved, since there are many different types of relationships that need to be expressed in hyperlinks.

The goal of the WIKINGER (Wiki Next-Generation Enhanced Repositories) project is the creation of a semantic Wiki containing both the entities relevant to the domain and the qualified associations connecting them. The main difference to other projects dealing with semantic Wikis is the level of automation. The project is working towards the semi-automatic creation of a base for the semantic Wiki from the digital repository, thus reducing the amount of work necessary to attain critical mass.

The process used by the project is shown in the Figure. The initial phase (labelled 0) can be seen as a bootstrapping phase for the system. An initial collection of digitally available data sources including publications, articles or databases is assembled and converted to a format suitable for further processing. The WIKINGER system stores both the original data as well as the derived format in a document repository. The data is then processed by a module doing Named Entity Recognition (NER, labelled 1), which gathers entities according to entity classes. A human annotator provides the module with examples for the designated classes, thus aiding the system in learning those classes. The advantage of this approach is the flexibility to include new classes: given specific examples, the system can learn to recognize them.

The output is a collection of recognized entities which serves as the input for stage 2. Stage 2 tries to identify the associations between the different entities. The result of this stage is a semantic net forming a hypothesis of the knowledge contained in the data sources. This hypothesis is evaluated by human experts, and this evaluation is used as the input for another iteration of the net-building process. This in turn is evaluated, and so the process continues. When the experts are satisfied with the results, the semantic net is deployed for use in the WIKINGER-repository.



Creation of a semantically enhanced knowledge repository.

This repository combines the functionality of a Wiki system with the expressiveness of associations found in languages of the Semantic Web. Nodes in the Semantic Net translate to articles in the Wiki; the different types of association between them form the hyperlinks connecting the articles. Since the Wiki is simply a user interface to the Semantic Net, the semantics behind the hyperlinks are retained and can be used for intelligent software assistants. The Net is kept in sync with the Wiki through use of a feedback loop that subjects all changes in the articles to the same process as the original data. This allows the identification of new topics or associations that come up in daily work with the Wiki.

The project is conducted by the University of Duisburg-Essen and the Fraunhofer Institute for Media Communication in cooperation with the Commission for Contemporary History (KFZG) in Bonn. The pilot project focuses on the domain of Contemporary History, in particular on the social and political history of German Catholicism. While still in its early stages, the first results from the project are very encouraging. At the moment we are working on a prototype offering basic functionality, which will enable users to test-drive the system early in the development cycle.

The project is funded by the German Federal Ministry of Research and Education in the program 'eScience'. Work on the project commenced in October 2005, and will be completed in September 2008.

Link:
WIKINGER project:
http://www.wikinger-escience.de (as of yet German only)

Please contact:
Lars Broecker, Fraunhofer-Institute for Media Communication, Germany
Tel: +49 2241 14 1993
E-mail: Lars.Broecker@imk.fraunhofer.de

Visualization

Computer measures Coral Structures

by Chris Kruszynski and Annette Kik

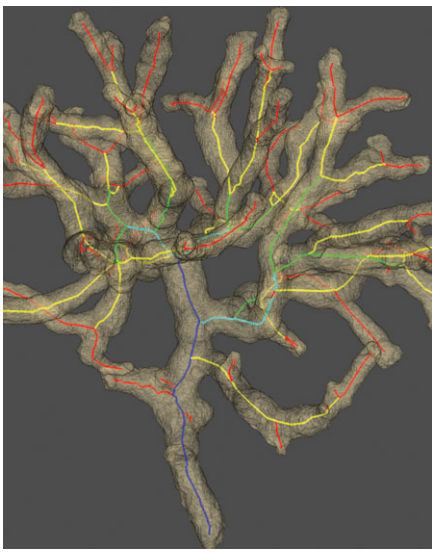
To conserve the Earth's coral reefs, biologists need to study them. In order to make better coral measurements possible, CWI and the Universiteit van Amsterdam developed sophisticated visualization methods to detect thickness, angles, lengths, spacing and branch ordering.

Coral reefs are important for the oceans' biodiversity and the growing nature-based tourism industry sector. They show a variety of forms depending on the environment, such as light, water flow and nutrients. Corals can, for instance, be ball-shaped or formed like a tree. The same species of coral can look completely different depending on environmental factors, which influence the growth pattern. To compare and classify specimens, it is important to make very precise measurements of coral thickness and branch distances. Biologists used to do this by hand, which takes a lot of time, and can cause unwanted errors. Due to the large number of branches, and complexity of corals, such manual measurements are generally only performed for a single metric – for example, branch length – or only for a small number of branches of a specimen.

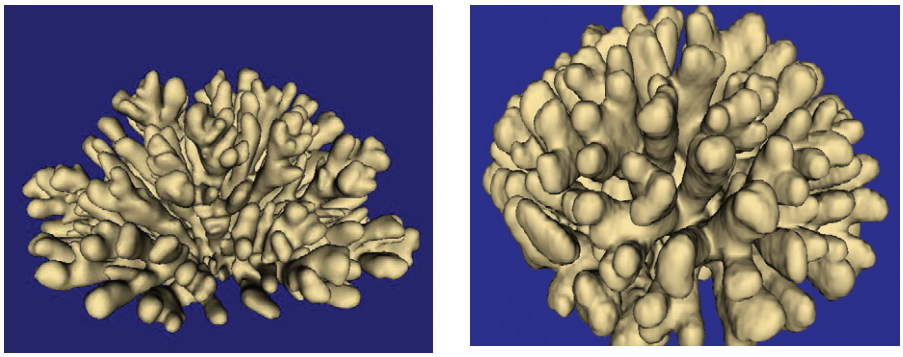
To make these coral measurements easier, quicker, more accurate and robust, and more comprehensive, Krzysztof Kruszynski (CWI) and Jaap Kaandorp (Universiteit van Amsterdam) developed a method for the quantification of branching coral shape, for which coral specimens are scanned in a CT scanner. The scan data are filtered, segmented and transformed to a centerline skeleton. This method simplifies detection of features like branches, branching locations, and endpoints. The skeleton is then measured by the computer, and the results are subjected to statistical analysis; the measurements include thickness, angles, lengths, and spacing of branches.

However, several problems can occur. Noise comes from CT scanner ray scattering, decayed parts of the coral and creatures growing on it. Branches broken during transport are reattached with glue, influencing the measurements. It is difficult to fill holes - made by worms or human drilling –making it hard to distinguish the inside of the coral from the background. 'Skeleton loops' – due to low scanner resolution or branches growing back together - make branch ordering, and thus measurement, impossible. Noise filtering might affect the shape, but the noise itself influences the measurements, and must thus be reduced.

Despite these problems, an interactive visualization system has been created and is being used. It is easier and quicker than older systems. The system has been created using existing open source software, such as the NLM Insight Segmentation and Registration Toolkit (ITK) for image processing, and the Visualization Toolkit (VTK) for visualization and interaction. New software was written for performing the actual measurements, using the VTK framework. The application area – coral biology – is novel for these techniques, some of which are quite advanced. For



Volume data from a CT scanner with a computed skeleton inside. Colours indicate branche numbers, counted from the outside. Picture CWI.



Surface of a real, scanned coral and a simulated specimen. The right one is the real coral. Pictures CWI.

example, the Curvature Flow filter from ITK is an advanced image filtering technique which reduces the amount of noise with minimal impact on the shape of the coral.

Some of the tasks in the new system are performed manually. Segmentation – dividing the 3D image into objects of interest – is mostly automatic, but it produces hollow coral with holes in it. The inside is filled, and the holes are patched, but the results of this automatic process must be checked for correctness. The skeleton can be extracted without user intervention, but any loops in the skeleton must be disconnected by hand; there is no known technique to reliably determine where a part of the skeleton should be removed to disconnect the loop, or how much of the skeleton should actually be removed. The computer assists the user, and the number of loops is typically very small, making this an easy and quick task.

In the future, the researchers want to create more advanced result visualizations; these are currently shown as a large collection of graphs and numbers. The accuracy of the computer measurements also still needs to be quantified. Clustering algorithms could detect similarities between species. Biologists can examine similarities and differences in shape, and study the correlation between shape and environment. This might help biologists with the important conservation challenges of coral reefs.

This work was partially carried out in the context of the Virtual Laboratory for e-Science project (http://www.vl-e.nl). This project is supported by a Bsik grant from the Dutch Ministry of Education, Culture and Science (OC&W) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ).

Links:
http://www.cwi.nl/ins3
http://www.science.uva.nl/research/scs/GF2004/
http://homepages.cwi.nl/~kruszyns/

Please contact:
Krzysztof Kruszynski, CWI
Tel: +31 20 592 4325
E-mail: K.J.Kruszynski@cwi.nl

Pattern Recognition

Fast Synthesis of Dynamic Colour Textures

by Jiøí Filip, Michal Haindl and Dmitry Chetverikov

The textural appearance of many real-world materials is not static, but changes over time. If such change is spatially and temporally homogeneous, these materials can be represented by means of dynamic textures (DT). DT modelling is a challenging problem that can improve the quality of computer graphics applications. As part of the MUSCLE Network of Excellence, collaboration between three ERCIM members – CRCIM-UTIA, SZTAKI and CWI – has developed a novel hybrid method for colour DT modelling.

Dynamic or temporal textures (DT) can be defined as spatially repetitive motion patterns exhibiting stationary temporal properties and having indeterminable spatial and temporal contents. The surface of water, fire and straw in the wind are typical DT examples. As a basic representation of DTs, a video sequence has finite duration. This limits the use of DTs in virtual reality systems of any kind, making temporally unconstrained modelling of DT a challenging problem for research areas such as computer vision, pattern recognition and computer graphics. Previous DT modelling approaches were based either on video editing techniques or time-consuming mathematical models, which were generally restricted to greyscale DT modelling.

The proposed method shows good performance for most of the tested DTs: this depends mainly on the properties of the original sequence. Moreover, this method significantly compresses the original data and enables high-speed synthesis of unlimited artificial sequences, which is easily performed by means of contemporary graphics hardware.

The method, illustrated in Figure 1, is based on a combination of input data dimensionality reduction using the eigen-analysis, and the subsequent modelling of resulting temporal coefficients by means of a causal simultaneous autoregressive random field model (CAR). The model is learned from real measured DTs (typically 250-frame video sequences). Measured data often show

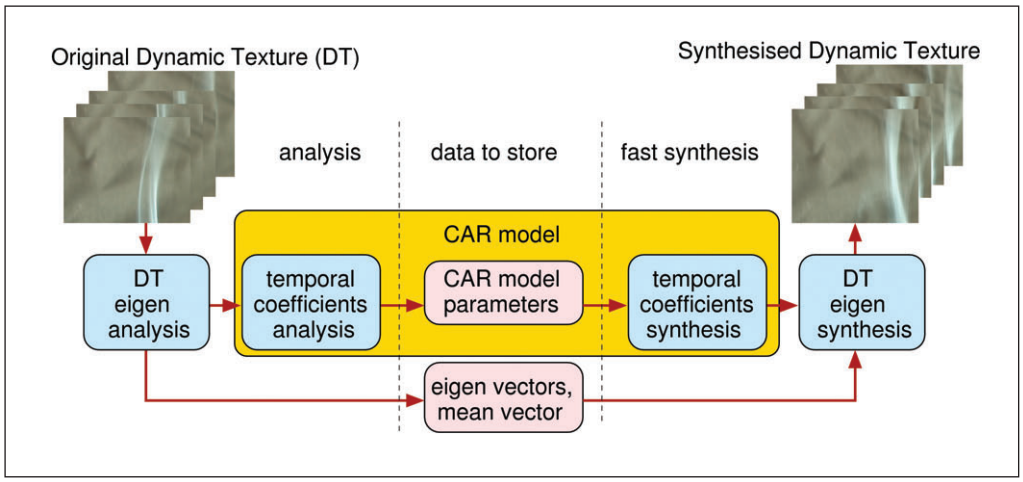


Figure 1: Scheme of the proposed dynamic texture hybrid model.

spatial discontinuity between successive images in DT sequences of very fast processes. Furthermore, available sequences are usually too short for robust statistical estimation of model parameters. We therefore performed the interpolation of individual temporal coefficients by means of cubic splines. This pre-processing step generates additional frames between each pair of original frames and improves the learning quality of the underlying random field model. The major advantage of the CAR model is that it can be solved analytically under several additional and acceptable assumptions.

The CAR model synthesis is very simple. New temporal mixing coefficients of individual eigen-images can be directly generated from the model equation using the estimated model parametric matrix and a multivariate Gaussian generator with estimated noise variance. Both the synthesis of new temporal coefficients and the following interpolation of eigen-images can be performed at even faster rates using contemporary graphics hardware programming. Moreover, this technique enables significant compression of the original DT data, typically at a ratio of between 1:5 and 1:10 depending on the length and the character of the DT sequence. The method was verified visually and by using two proposed statistical similarity measures on dynamic texture data sets. These include fire, boiling water, moving escalators, smoke, straw etc, and are taken from the DynTex texture database maintained by our partners at CWI. The comparison of original and synthesized DT frames of natural textures is shown in Figure 2. The corresponding results for the man-made textures are illustrated in Figure 3. The analysis time of the original DT was about three minutes. The synthesis of a new DT sequence is very fast (about 60 frames/s using non-optimized CPU software implementation on a PC with an Athlon 2GHz processor), and the generation time can be further improved using the programmable processing unit of a contemporary graphics card.

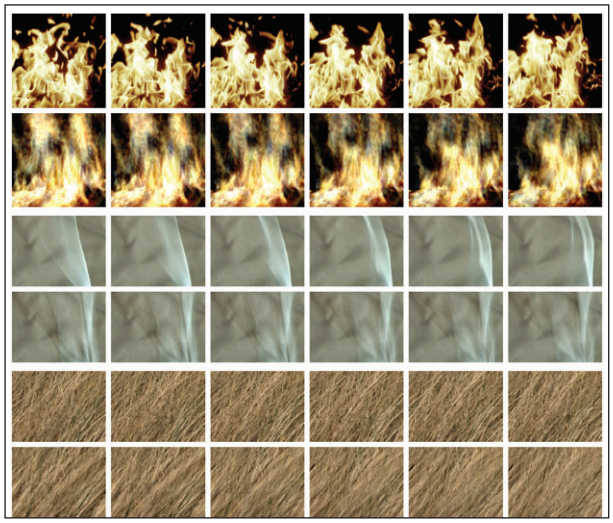


Figure 2: Examples of frames from original DT (odd rows) and the corresponding synthesised frames using the proposed model (even rows) for three natural DTs.

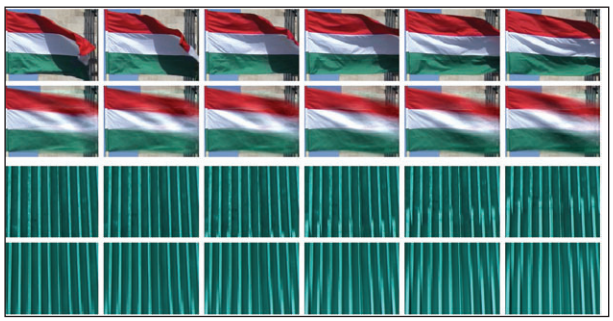


Figure 3: Examples of frames from original DT (odd rows) and the corresponding synthesised frames using the proposed model (even rows) for two man made DTs.

Links:
DTdemos:
http://ro.utia.cz/demos/DTsynth.html
DynTex database:
http://www.cwi.nl/projects/dyntex/
http://ieeexplore.ieee.org/Xplore/guesthome.jsp
(see IEEE Digital Library for the article).

Please contact:
Jiøí Filip, Institute of Information Theory and Automation, Academy of Sciences / CRCIM, Czech Republic
Dmitry Chetverikov, SZTAKI, Hungary
Tel: +420 266052365
E-mail: filipj@utia.cas.cz

Multilingual/Multimodal Information Retrieval

MultiMATCH - Multilingual/Multimedia Access to Cultural Heritage

by Carol Peters

MultiMATCH, a 30 month specific targeted research project under the Sixth Framework Programme, plans to develop a multilingual search engine for the access, organisation and personalised presentation of cultural heritage information.

Cultural heritage content is everywhere on the web, in traditional environments such as libraries, museums, galleries and audiovisual archives, but also in popular magazines and newspapers, in multiple languages and multiple media. The aim of the MultiMATCH project is to enable users to explore and interact with online accessible cultural heritage content, across media types and language boundaries.

The MultiMATCH search engine will be able to:

- identify relevant material via an in-depth crawling of selected cultural heritage institutions, accepting and processing any semantic web encoding of the information retrieved;
- crawl the Internet to identify websites with cultural heritage information, locating relevant texts, images and videos, regardless of the source and target languages used to write the query and/or describe the results;
- automatically classify the results in a semantic-web compliant fashion, based on a document's content, on its metadata, on its context, and on the occurrence of relevant cultural heritage concepts in the document;
- automatically extract relevant information which will then be used to create cross-links between related material, such as the biography of an artist, exhibitions of his/her work, critical analyses, etc.;
- organise and further analyse the material crawled to serve focused queries generated from information needs formulated by the user;
- interact with the user to obtain a more specific definition of initial information requirements;
- the search results will be organised in an integrated, user-friendly manner, allowing users to access and exploit the information retrieved regardless of language barriers.

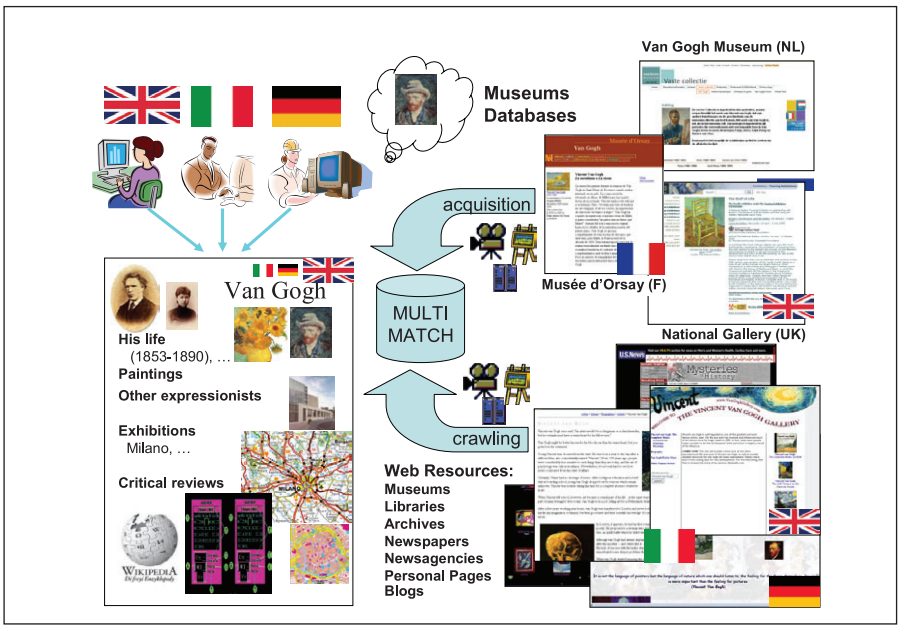


Figure 1: The MultiMATCH idea.

The concepts underlying the system are depicted in Figure 1. On the left-hand side of the figure, we show users querying the system in different languages for a range of information on the Dutch artist Vincent van Gogh, including critical analyses, biographies, details of exhibitions. The system displays the retrieved information in an integrated fashion, and in a format determined by the particular user profile. On the right-hand side, we show possible sources of this information and the ways in which it can be acquired.

The project aims at developing a system prototype that can be demonstrated for at least four languages: Dutch, English, Italian and Spanish, and extendible to others. Figure 2 gives an idea of the workflow for the system development.

The R&D work is organised around three activities:

- User-oriented research activities will primarily investigate the user requirements and consequent definition of the required functionality of the system, content selection and preparation, studies on the ontologies adopted by cultural heritage institutions and the semantic encoding to be adopted by the system.
- System-oriented research activities include the study and development of software components for the acquisition, indexing, classification, retrieval and presentation of multilingual cultural heritage information in diverse and mixed media and their integration in the system prototypes.
- Validation activities will include evaluation of the system and its components. User groups of cultural heritage institutions and cultural heritage consumers will be formed to test the system and provide feedback.

The consortium comprises eleven partners, representing the relevant research, industrial and application communities. Each member will play a significant role in the design and development of the system, providing a part of the necessary know-how; the blend of competences will be a key factor for the success of the project. The six academic research partners (ISTI-CNR, U.Amsterdam, UNED-Madrid, U.Geneva, U.Sheffield, Dublin City U.) have already worked closely together collaborating in coordination of the Cross-Language Evaluation Forum (CLEF). CLEF focuses on stimulating advances in research in multilingual/multimedia information retrieval and on information extraction and user/system interaction in the cross-language context. The industrial partners, OLCC PICA, UK, and WIND, Italy will play a major role in the design of the system architecture and the integration of the various components, also with a view to the future industrialisation and commercial exploitation of the system. The cultural institutions, Casa de América, Spain, Alinari, Italy, and Sound and Vision, the Netherlands, each represent a different type of cultural institution with content in diverse media and languages but all three groups have in common the desire to improve and extend their information dissemination capabilities, and to work towards the development of standards for interoperability and metadata in the cultural heritage domain. The intention of these institutions is to be able to exploit the results of the project in their future information dissemination activities.

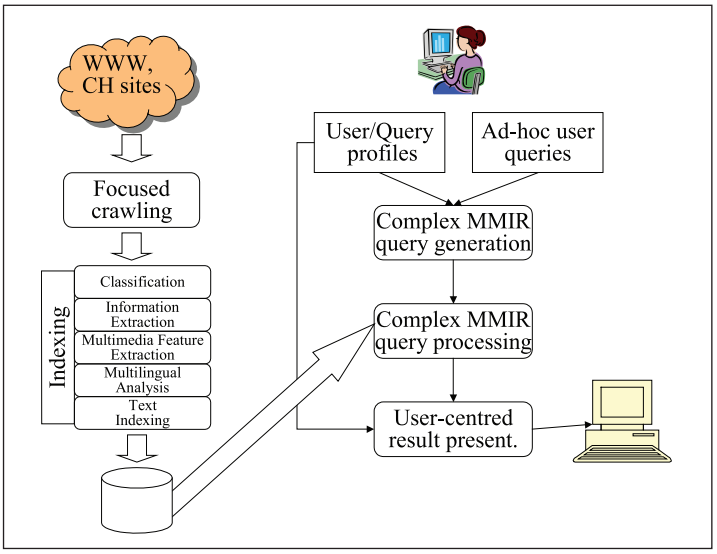


Figure 2: Workflow for development of MultiMATCH search engine.

MultiMATCH is supported by the unit for Content, Learning and Cultural Heritage (Digicult) of the Information Society DG and is coordinated by ISTI-CNR, Pisa, Italy.

The project kick-off meeting was held in Pisa, 10-12 May 2006. The meeting was mainly dedicated to a detailed planning of the activities for the first year. The first system prototype is scheduled for release in November 2007. MultiMATCH will issue a quarterly newsletter providing information on the project activities, events and results. For further information, please see the project website.

Links:
http://www.multimatch.inf
http://www.clef-campaign.org

Please contact:
Carol Peters, ISTI-CNR, Italy
MultiMATCH Project Coordinator
Tel: +39 050 3152897
E-mail: carol.peters@isti.cnr.it

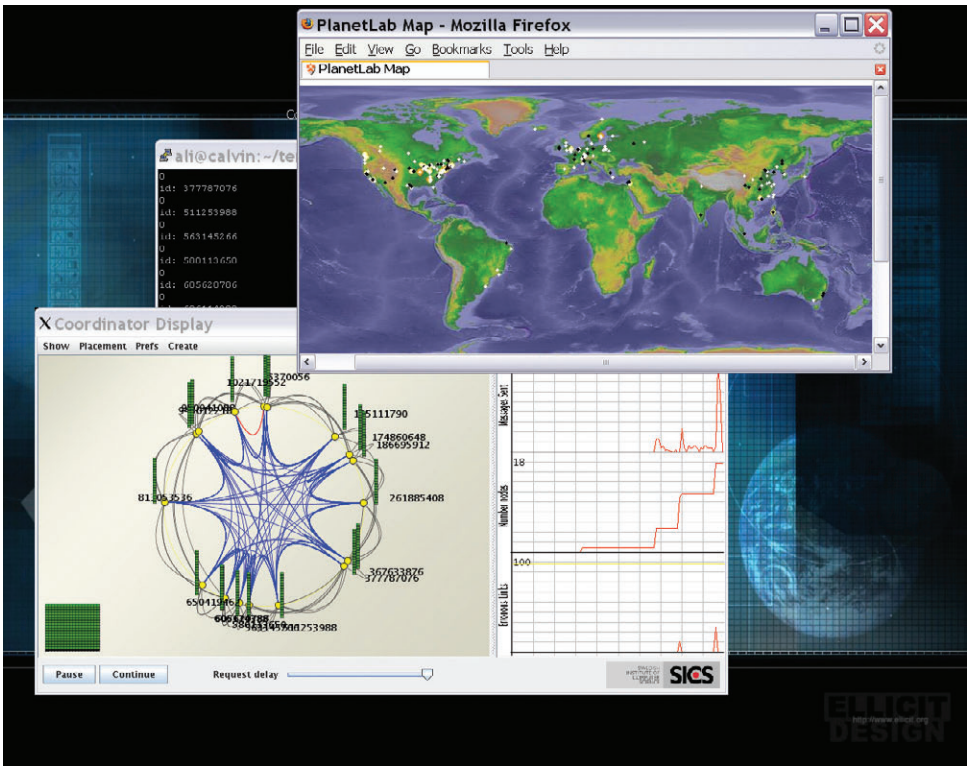
Security

Access Control and Data Distribution Solutions for the Swedish Network Based Defence

by Frej Drejhammar, Ali Ghodsi, Erik Klintskog, Erik Rissanen and Babak Sadighi

Network Based Defence (NBD) is a national Swedish military funded project with the goal of developing the next generation command and control system. The main focus of the project is to develop a system that is so scalable, flexible, robust, decentralized and interoperable that it can handle the needs of tomorrow's battlefield. SICS has in close cooperation with FMV (Swedish Defence Materiel Administration) and Saab Systems developed a role based access control system for NBD.

Access control is about deciding who gets access to what resources in the system. NBD is a plan for a military information system which can be highly flexible, resilient and provide information superiority. The system is built as a highly decentralised and dynamic system of systems. In this environment with high demands on mobility and



DKS provides fault tolerance and enables parts of the system to continue to function autonomously in case of loss of network communications.

autonomy, traditional centralised solutions for access control can no longer be applied. SICS has previously used the Delegent authorisation server, which is based on research done at SICS, in proof of concept demonstrators for NBD. During the year Delegent has been redesigned to be based on XACML (eXtensible Access Control Markup Language), a standard for access control policies. To support the NBD requirement we have extended XACML functions for delegated decentralised administration of policies. Decentralised administration provides more resilience to failures and faster reaction times when adapting to new situations.

Also, in order to further adapt Delegent to the NBD requirements, we have coupled Delegent to a structured peer-to-peer-system (P2P) called DKS, which provides decentralised storage of the access control policies. The DKS system, implemented by SICS and KTH, provides a decentralised data management system, with additional support for a Publish and Subscribe service. The DKS enhances Delegent to no longer rely on a centralised policy repository, as it distributes policies via the DKS storage system and the Publish and Subscribe service can work out which information is needed where in the network. This provides fault tolerance and enables parts of the system to continue to function autonomously in case of loss of network communications.

The DKS system is designed to connect a large number of machines with dynamic behaviour in an overlay network. Dynamic behaviour includes machines joining and leaving the overlay, as well as machines failing and connections to machines failing. With the minimal requirement of point to point connectivity, aggregated functionality such as reliable data storage, name-based communication and multicast are provided.

The updated version of Delegent and DKS have been installed at the Swedish Defence Material Administration proof of concept facility and successfully used in experiments during the autumn of 2005.

The DKS enhanced Delegent system is a potential core component of NBD. Access control functions are no longer located on one single machine, but distributed to the edge of the network with the additional benefit of reduced bandwidth consumption and removal of a single-point-of-failure.

The successful coupling of Delegent and DKS is just one example of where structured P2P-systems can be applied within the NBD project. We foresee a multitude of other applications that could benefit from the usage of a structured P2P-system, such as service repositories, user databases and flat name space resolution services, some of which will be explored in the future. We will also continue the research on access control solutions for dynamic systems with more research on administration and revocation models and how to best present information to users.

Link:
http://www.sics.se/spot/

Please contact:
Erik Rissanen and Frej Drejhammar, SICS
Tel: +46 633 1500
E-mail: mirty@sics.se and frej@sics.se

Ambient Intelligence

Bringing Ambient Computing out of the Labs - INRIA's Agreement with JCDecaux

by Michel Banâtre

INRIA and the JCDecaux group, a worldwide leader in street furniture, recently signed a technology transfer agreement. This may come as a surprise, since this group's business areas make it an unusual partner for INRIA. Nevertheless, the agreement fits well with the research on ambient computing being carried out by INRIA's ACES (ambient computing and embedded systems) research team.

ACES became involved in this exciting area in 1998, working on Spontaneous Information Systems (SIS). Our SIS research involved dynamic information systems shared by proximal mobile devices that communicate through short-range radio transmission. This led to our groundbreaking work on ambient computing, which in turn led to the development and study of a variety of novel concepts, including 'spatial computing'.

In this kind of architecture, physical objects are data symbols and physical space is the basis for addressing. In other words, such an architecture supports implicit computation using the flow of data from the physical motion of the associated objects. When we proposed such concepts, there were already 'popular' solutions in the 'Ubicomp' community. Essentially they were based on 'logically centralized' approaches, built around information systems independent of the physical environment. Such concepts have only very recently emerged as relevant focus areas, thanks to the growing interest in sensor-related themes (electronic labels, sensor networks, smart dust etc).

It is important to note that despite the wealth of new ideas generated since 2000, there have been no major innovations resulting in core applications or widespread use. In other words, ambient computing hasn't left the lab. Overcoming this is a real challenge, and one that the ACES project team wanted to tackle. Our approach, which has already been used successfully several times in the past, is to "go the distance", as we did recently with Texas Instruments. This is even more critical in the domain of ambient computing, which in essence is based on information technology that is tightly coupled with the real world. If we ignore this aspect of ambient computing, we will overlook the real challenges, the very ones we must address to ensure the emergence and application of our ideas as researchers. That's why we've had numerous discussions, productive to varying degrees, with a broad range of enterprises, including equipment makers, mobile telephony operators and end users. Although these discussions did not lead to actual collaboration, they allowed us to identify and understand what was preventing ambient computing from taking hold in these enterprises. As is often the case with innovation, it isn't so much the technical difficulties that block progress, but rather the challenge of identifying the 'missing piece'.

The bottom line is that integrating 'context awareness' into mobile terminals is still very difficult. Even though it seems simple from a technical perspective, there are numerous unavoidable obstacles, not least of which is negotiating agreements between the various partners on wireless standards (Bluetooth, RFID, IR etc) and software. The situation is complicated by the current absence of an application that will bring the players together and motivate them to overcome the obstacles they face.

There exists an alternative however, which is the natural outgrowth of our proposed spatial machine and involves integrating context awareness into the environment. This approach has been tested since 2001 at INRIA Rennes with the 'WebWalker' application, which among other things allows users to move physically through the Web. With this technology, the challenge of producing an effect is linked to the quality and quantity of the sites encountered.

Integrating 'context awareness' into the environment has been tested since 2001 at INRIA Rennes with the 'WebWalker' application, which among other things allows context-driven navigation as the user moves.

Within this context, street furniture represents a very attractive vector. The question of computing aside, street furniture is already at the centre of an environment-based information system, tied to transport, tourism, events and maps/directions, not to mention advertising. Such an information system, built around physical objects, is particularly well suited to the application of our solutions, which are also based on spatial distribution and management of information.

Currently three large global groups – ClearChannel, Viacom and JCDecaux – are the main players tapping into this application domain. It was during 'Les Transports au XXIième Siècle' – an event organized by the French Senate in April 2004 around the theme of transport in the 21st century – that JCDecaux learnt of our research, from a presentation about our 'Ubibus' system.

From the underlying principles alone, and in light of existing technology, JCDecaux immediately saw how they could benefit from our ambient computing solutions. However, convincing the group to formally adopt these solutions required significant effort on our part, not only from a technical perspective (creation and demonstration of pilots for real-world situations, evaluation of performance and development of extensibility) but also in terms of financial criteria (costs of deployment and exploitation, long-term survivability). One of JCDecaux's concerns is the 'intrusiveness' of the implicit way in which ambient computing systems function. This could meet resistance from the public, and that would represent a real problem for a group whose revenues come mainly from advertising. Our expertise and broad perspective on these problems as well as the relevance of our solutions have all been critical factors in JCDecaux's decision to work with us. We have also taken the important step of obtaining patents to protect certain core aspects of our solutions.

Link:
http://www.inria.fr/recherche/equipes/aces.en.html

Please contact:
Michel Banâtre
INRIA Rennes – Irisa
Tel : +33 2 99 84 72 85
E-mail: Michel.Banatre@inria.fr

Text Research

New Text - New Conversations in the Media Landscape

by Jussi Karlgren

New text - that is, new forms of textual communication - such as blogs, instant messages, and Wikis contrast with traditional textual genres in some respects and remain true to them in others. This calls for new research methodologies and provides new challenges for text research.

Recent advances in publication and dissemination systems have given rise to new types of text - dynamic, reactive, multi-lingual, with numerous cooperating or even adversarial authors and little or no editorial control. Many of these new types of text remain true to established existing textual genres. Others break new ground, moving towards new emergent textual genres made possible by the dramatically lowered publication threshold and faster distribution mechanisms.

These new forms of text, with a considerable amount of attention from traditional media, most notably include blogs - texts written as a timely running commentary of public or private matters. Another well-established and remarkable new genre is the Wikipedia - an encyclopaedia built through the cooperative efforts of its readers. New forms of communication such as these raise questions for researchers in a variety of fields, and this past spring has seen no less than two international workshops held on the analysis of new texts - bringing together several topically similar research projects around Europe.

One of the first questions in this research field is how new text is different. How new is 'new'? Have we never had new text before? What, in fact, is the difference between 'new' and 'old'? It is quite clear that authors of both newand traditional texts are aware of linguistic styles of various sorts and use them in ways they deem appropriate. When new genres emerge, such as blogs or Wikipedias, they may pattern them-

A newspaper reader - how new is 'new'?

selves on existing ones, such as diaries or encyclopaedias, thereby drawing on the prestige and position of those existing genres. Alternatively, they may cast around for forms suitable for their intended impact and stature. How to achieve this form where none exists is a matter yet to be resolved!

New texts are more than simply revamped traditional texts however: they have features that traditional texts lack. They are interconnected by a network created by authors and readers in a complex interplay of explicit textual references; they also position themselves much more explicitly within a context of other texts than has previously been the case. Studying this fabric of textuality is just the first step in this area of research.

In view of the less formalized publication process, the credibility of new texts can be called into question. When traditional texts are published in paper form, a number of steps - variable from one mode of publication to another - involve satisfying editors or publishers of the veracity, relevance, quality and impact of a text. (Whether this is a good or a bad thing is a different discussion entirely!) New texts lack this guarantee of having passed many pairs of eyes en route from author to reader. There is no simple measure of the impact, the variable perceived intellectual status and quality of new texts. Understanding credibility, authority and other facets of quality are central to any attempt at analysis of the impact of new texts.

Underlying the issue of credibility and authority is the question of who the author is and why. What makes a blogger blog? Why do people devote time and energy to editing Wikipedia pages? Understanding the motivations and intentions of authors is not incidental to the task of understanding the texts. Integral to the blog is who and why; integral to the Wiki is purpose; and no-one can pretend that the texts are analysable in isolation. While texts remain texts, even with new syntactic patterns and new lexical items, their contextuality is so great as to dominate many other content features. And this, in fact, is truly new!

What services can be expected to emerge from the analysis of new text? Several information access services already use Wikipedias to extract facts and relationships for better understanding of other texts. The analysis of public opinion on issues, or of consumer attitudes towards products and services on the market, has found a rich vein of data in blogs. To do this with any level of reliability however, our processing tools, tuned to newsprint and other well-edited texts, need to address the challenges of variable or multi-lingual texts, containing register swings and formality melanges - not shoddy, but New!

We are currently in a transition phase, which is exceedingly interesting both philologically and industrially. Similar phases have been seen before, for example with the introduction of inexpensive printing processes, publishers put out compilations of private correspondences as one form of written communication assumed to be of public interest. The only certainty we have today is that in the future, people will find creative ways of using the technology we are introducing today ? again, not unpredictable, but New!

These new movements will be discussed in coming research events. Those interested are welcome to join the discussion at newtext@sics.se!

Links:

New Text Workshop:
http://www.sics.se/jussi/newtext

AAAI Blog Symposium:
http://www.umbriacom.com/aaai2006_weblog_symposium/

Int. Conference on Weblogs and Social Media, March 26-28, 2007, Boulder, Colorado, USA: http://www.icwsm.org

Please contact:
Jussi Karlgren, SICS, Sweden
Tel: +468 633 15 00
E-mail: jussi@sics.se

Logistics

LOG4SMEs: Improving the Logistics Performance of SMEs in the Automotive Sector

by Imre Czinege, Elisabeth Ilie-Zudor and András Pfeiffer

The LOG4SMEs project aims at improving the logistics performance of small and medium-sized enterprises (SMEs) in the automotive sector.

Small and medium-sized automotive businesses are threatened by a number of pressing issues, including saturation of the market, fierce competition and the reduction of entry-barriers. Among these problems are the identification of weaknesses in logistics and production processes and the finding of appropriate action lines or IT tools to overcome them. These issues do not depend exclusively on companies themselves, but are also heavily affected by the economic, logistical and social character of the regions in which the SMEs are located.

In January 2006, partners from three regions of the European Union launched the project Log4SMEs, a Regins project (see http://www.regins.org). The University of Bergamo from Lombardy Region (Italy) leads the consortium. Széchenyi István University represents West Pannonia (Hungary), while the Fraunhofer Institute for Manufacturing Engineering and Automation (IPA) represents the region Baden Württemberg (Germany).

One of the project's goals is to enable SMEs to compare their individual current logistics performances with industry and regional averages as well as with the best performer. A second goal is to enable each company to identify its current performance gaps and to determine whether, disregarding the specific company's actions, there are regional factors that affect its logistical performance. Special emphasis is put on identification

strategies that will allow the three regions to develop the location factors for their local automotive industries.

LOG4SMEs will provide companies with the ability to acknowledge the best practices in their region/industry and will encourage the exchange of good logistics practices among companies. From an extensive survey throughout the three regions, a Web service addressed to all registered companies will provide a unified database of logistical performance indicators and practices in the automotive sector. The project will directly involve SMEs operating in the automotive sector by the provision of a survey, as well as phone and direct interviews. The design of the survey is based on the internationally acknowledged standard SCOR-model and SCOR–indicators. The results will also be distributed to local industrial associations or industry clusters.

Through the project web site, automotive companies will be able to compare their logistics and production performance with other SME- and industry-specific indicators as well as to derive their strengths and weaknesses. Developing a catalogue of logistics practices for each identified cluster of companies and describing the main regional location factors that foster or inhibit their logistical strategies is also in the scope of Log4SMEs.

Please contact:
Imre Czinege, Széchenyi István University, Győr, Hungary
Tel: +36 96 613673
E-mail: czinege@sze.hu

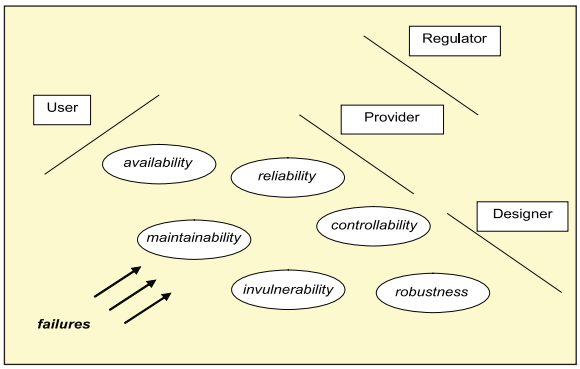
Networks

VTT Develops Dependability Evaluation Methods for IP Networks

by Ilkka Norros

The Finnish research project 'Dependability evaluation methods for IP networks' - IPLU aims to create a conceptual framework and methods for assessing the complex problem "Can one rely on IP technology?" The research is done by VTT and funded by several organisations, including the Ministry of Traffic and Communications, the National Emergency Supply Agency and four telecom operators.

Since the project aims at a comprehensive view of the topic, it has a multidisciplinary character combining VTT's expertise in telecommunications technology, teletraffic and network modelling, and reliability analysis. In the international workshop 'Dependability of all-IP networks', organised by the project 18-19 May 2006 (http://iplu.vtt.fi/workshop-06.html), the multidisciplinary nature was further enriched by contributions from human activity research and theoretical computer science.



A preliminary conceptual framework for network dependability, discussed in the baseline paper of the IPLU project.

This approach has proven fruitful. The established methods of safety and reliability assessment used in contexts like nuclear power are not mechanically transferable to the highly dynamic world of IP networking, and, on the other hand, researchers with telecom background are mostly unaware of the experience accumulated in the reliability research traditions. The IPLU project aims at inventing new methods, but not at inventing the wheel anew from scratch.

As one of its first tasks, the research team produced a baseline paper that sets the scene for structured discussion of the dependability problems in IP networking. The internet is recognized as a new medium the character of which is more generic than traditional electronic communication media like telephone and television. The baseline paper proposes a preliminary conceptualization of dependability, where a traditional set of dependability attributes is augmented by aspects that reflect the self-regulation features of the internet architecture. The paper is available from the project's webpage.

The main aim for the autumn of 2006 is to propose an initial set of criteria, indicators, procedures and recommendations for consideration by network operators and other actors on the telecommunication scene. VTT is prepared to continue IPLU's work in future projects.

Link:
http://iplu.vtt.fi.

Please contact:
Ilkka Norros, VTT-Technical Research Centre of Finland
Tel: +358 20722 5627
E-mail: ilkka.norros@vtt.fi

Software Engineering

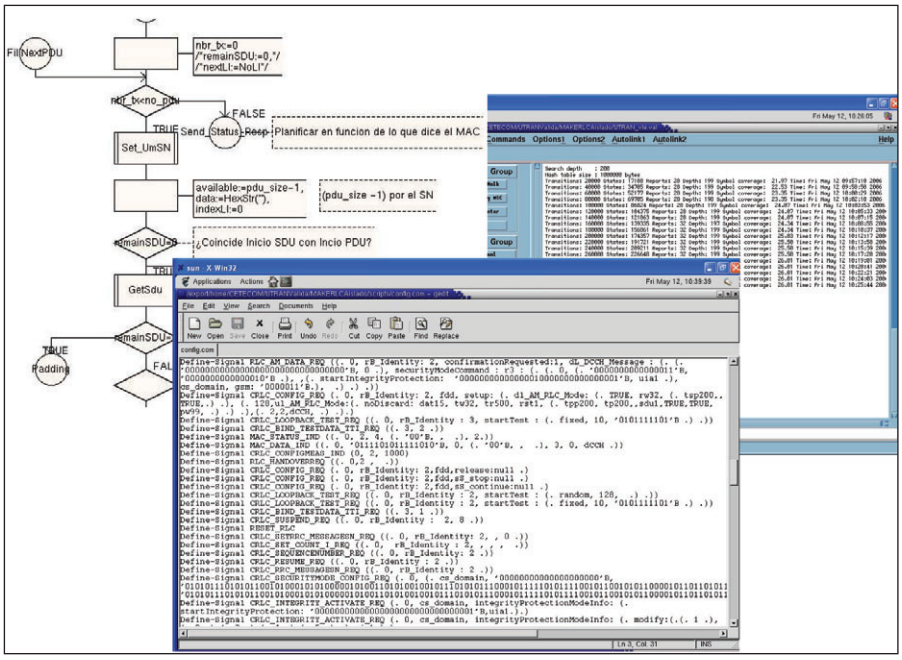
Validating Complex Telecommunication Software

by Sergio Contreras, María del Mar Gallardo, Pedro Merino, David Sanán, Javier Rivas and Joaquín Torrecilla

In a collaboration with the telecommunication company Centro de Tecnología de las Comunicaciones, S.A. (CETECOM), a research group at University of Malaga is validating communication software written in languages such as SDL, ASN.1 and C.

Automatic validation of many critical systems has usually been done with model-oriented techniques, like model checking. This approach requires the previous construction of a specific model (an abstraction) of the problem, using very specific languages oriented to academic tools like SPIN. In general terms, however, this abstraction is not well suited to automatically obtain an implementation. This is why this model-based methodology has only limited success within industry and is not generally employed, in particular, by telecommunication companies.

Many companies employ classic C and C++ languages for their critical communication software, and they have traditionally replaced validation by testing and/or debugging. Other companies use development languages that preserve validation facilities, as long as they are also powerful enough to automatically obtain the final software. In particular, they use standard description languages like the notations SDL (Specification and Description Language) and ASN.1 (Abstract Syntax Notation One), which are promoted by ITU-T (Telecommunication Standardization Sector of the International Telecommunication Union). These languages offer an acceptable formal basis and they are linked to other standard notations, like TTCN (Tree and Tabular Combined Notation) and UML. In theory, using ITU-T languages makes it possible to perform validation at the same time that the software is developed; note that the executable code is obtained automatically by translating SDL to C/C++. However, these validation facilities seem to be underexploited, and more efforts should be devoted to obtaining a methodology for effective validation of telecommunication software. This is one of the objectives of a joint project of the company



Validating with Tau: SDL diagram (left), validation script (center) and validation results (right).

Centro de Tecnología de las Comunicaciones, S.A. (CETECOM), the research institute Centro Andaluz de Innovación y Tecnología de la Información y las Comunicaciones (CITIC) and the Software Engineering Group of the University of Malaga (GISUM).

In particular, in the context of the project, CETECOM and GISUM are working on a methodology to validate existing complex software, using UMTS signalling as a case study. This software is mainly based on SDL to implement the protocol state machines, ASN.1 to describe the message types (as defined by standardization committees) and C to implement critical parts (like encoding/decoding algorithms). The whole software was produced with the tool Telelogic Tau. It contains more than 100,000 SDL symbols and it produces an implementation with more than 800,000 lines of C code. It is worth noting that such software has been previously tested with traditional methods and that it is complex enough to have Tau validation tool running for hours.

As expected, the first problem in performing validation is the size of the system. It produces millions of global states each having more than 20 Kbytes, considering only the SDL structures. However, there is a second more interesting problem. As current implementation is only a part of the whole signalling system, we need to complete the SDL description with information from the environment. The Tau validator can automatically produce many messages simulating the environment; realistic messages, however, have a great number of parameters that cannot be efficiently produced by the tools.

We have designed a methodology to deal with both problems at the same time. The approach is based on the partitioning method proposed in the literature, but we have adapted it to existing heterogeneous software (SDL, ASN.1 and C). We isolate processes and blocks and construct a realistic environment for each partition. The environment is constructed as a validation script that limits the set of messages and the range of values in the messages. Then, the validator uses its internal mechanism to generate messages automatically.

Using validation scripts, we can perform validation of separate parts; however, we still need more optimization methods to deal with complexity. One of them is bit-state respresentation. Another one is variable hiding, which is applied to big C structures.

The project has been successful in terms of quality of the validation results (including the confirmation of the robustness of the code) and also in terms of the methodology generated. The work has also been useful to propose some extensions in the validation tool. We have shown that validation is a valuable task for existing telecommunication code, even when validation was not a primary aim during development. This is made possible by the use of standard languages like SDL with tool support.

Links:
CETECOM: http://www.cetecom.es
CITIC: http://www.citic.es
GISUM: http://www.lcc.uma.es/~gisum/

Please contact:
Maria del Mar Gallardo
University of Málaga/SpaRCIM, Spain
E-mail: gallardo@lcc.uma.es

Pedro Merino
University of Málaga/SpaRCIM, Spain
E-mail: pedro@lcc.uma.es

Joaquín Torrecilla, CETECOM, Spain,
E-mail: jtorrecilla@cetecom.es

Insight into EU R&D Achievements

More than 7 billion Euro has been invested by the European Commission in the Information Society Technologies (IST) research priority. Now, thanks to a dedicated online news service – IST Results – technology users and research teams can read about results and innovations emerging from this considerable R&D activity which have potential for further development or exploitation.

Launched in 2003 by DG Information Society and Media, this free service reports on the latest achievements from IST projects via in-depth feature articles and news-in-brief stories. There is also a calendar of events dedicated to IST project events.

Through its editorial approach, IST Results is establishing an international profile in the mainstream press with articles syndicated to leading publications such as the Financial Times, New Scientist, Wired and ZDNet as well as to specialist online information portals.

Feature articles produced by IST Results are also regularly circulated through press wire services – an article describing results from a robotics project attracted over 93,000 hits from media professionals worldwide, demonstrating the power of this approach for communicating R&D more widely than traditional Commission dissemination tools. IST Results' own website attracts around 250,000 visits a month from technology users, researchers, media and investors in more than 50 different countries.

IST Results covers nearly thirty technology and market application areas so ERCIM members are certain to find articles relevant to their own interests. Browsing through recent news reveals a wide range of achievements in just one week's reporting, such as:

- A range of GRID services which allow the exchange of data and job information between different Grid systems and give a single client access to different Grid infrastructures
http://istresults.cordis.europa.eu/index.cfm/section/news/tpl/article/ID/82315/BrowsingType/Features
- Open source software for museums, archives and libraries, allowing them to utilise their cultural content and resources in novel ways while drastically reducing the costs of deploying digital library services
http://istresults.cordis.europa.eu/index.cfm/section/news/tpl/article/ID/82247/BrowsingType/Features
- A sports broadcasting platform linking the existing media channels, internet, TV and phone to offer a wide variety of services for journalists, VIPs, broadcasters, advertisers and, of course, the fans
http://istresults.cordis.europa.eu/index.cfm/section/news/tpl/article/ID/82199/BrowsingType/Features
- Professional wiki-based collaboration platform that has enabled the creation of a world-leading scientific network in archaeology.
http://istresults.cordis.europa.eu/index.cfm/section/news/tpl/article/ID/82290/BrowsingType/Features

IST Results follows a monthly Editorial Calendar and for July 2006 the theme is Cultural heritage, including digital libraries.

To find out more, browse the website at http://istresults.europa.eu/ and subscribe to the free e-alerts, RSS feed or weekly e-bulletin. If you are interested in republishing IST Results articles on your own website or would like more information, contact helpdesk@istresults.info.

Links:
IST Results home page:
http://istresults.europa.eu/

Editorial Calendar:
http://istresults.cordis.europa.eu/index.cfm?section=press&tpl=editorial_themes

CALL FOR PARTICIPATION

2nd International Workshop on Automated Specification and Verification of Web Systems

Cyprus, 15-16 November 2006,

The increased complexity of Web sites and the explosive growth of Web-based applications has turned their design and construction into a challenging problem. Nowadays, many companies have diverted their Web sites into interactive, completely-automated, Web-based applications (such as Amazon, on-line banking, or travel agencies) with a high complexity that requires appropriate specification and verification techniques and tools. Systematic, formal approaches to the analysis and verification can address the problems of this particular domain with automated and reliable tools that also incorporate semantic aspects.

The WWV series provides a forum for researchers from the communities of Rule-based programming, Automated Software Engineering, and Web-oriented research to facilitate the cross-fertilization and the advancement of hybrid methods that combine the three areas.

Topics

We solicit paper on formal methods and techniques applied to Web sites, Web services or Web-based applications, such as:

- rule-based approaches to Web site analysis, certification, specification, verification, and optimization
- formal models for describing and reasoning about Web sites
- model-checking, synthesis and debugging of Web sites
- abstract interpretation and program transformation applied to the semantic Web
- intelligent tutoring and advisory systems for Web specifications authoring.

WWV'06 will be held as a Special Track of the 2006 International Symposium on Leveraging Applications of Formal Methods, Verification, and Validation (ISoLA 2006)

More information:
http://www.dsic.upv.es/workshops/wwv06

CALL FOR PARTICIPATION

Loco Mummy Contest 2006

Develop a new Interface Program and win a laptop computer!

The goal of the contest is to find one or more clever and creative ways to use standard hardware to design a software application with a natural, intuitive way of interaction between the machine and its user.

The application software should be a new or updated free of rights multimodal User Interface software, allowing an interaction between one or several user and a computer software of any kind through ordinary devices such as keyboard, mouse, screen, microphone, loudspeakers and webcams. Complex and uncommon interfaces are explicitely banned.

The software should present a clear advantage to the user compared with standard mouse and keyboard applications. If the progam is an update to an existing software, copyright issues concerning the base software must be confirmed and clearly specified by the candidate before making the registration.

Important Dates

- Entry Registration Deadline: 15 September 2006
- Software Submission Start: 15 September 15th, 2006
- Software Submission Deadline: 30 October 30th, 2006
- Award Ceremony: 14 December 2006

Prizes

- Best PC User Interface Software Award: Laptop computer
- Best PDA User Interface Software Award: Pocket PC (PDA)
- Innovation award: USB Scanner

More information:
http://www.locomummy.net/

CALL FOR PARTICIPATION

TEL-CoPs'06: First Workshop on Building Technology-Enhanced Learning Solutions for Communities of Practice

Crete, Greece, 2 October 2006

The ITEL-CoPs'06 workshop, held in conjunction with the First European Conference on Technology Enhanced Learning, focuses on current research trends in technology enhanced learning solutions that adress the multiplicity and complexity of needs of 'Communities of Practice' (CoP). It advocates for approaches that build on the synergy of concepts such as multimedia information authoring and reuse, knowledge management, argumentation and negotiation. It will bring together scientists and engineers who work on designing and/or developing the abovementioned solutions, as well as practitioners who evaluate them in diverse real environments. Particular interest will be given to approaches built according to well-established pedagogical principles.

Topics of interest include:

- software engineering issues in tools supporting CoPs
- multimedia authoring and reuse in CoPs
- knowledge management services for CoPs
- mediation services for CoPs
- computer-supported collaborative argumentation
- learning issues and CoPs
- evaluation issues and case studies
- user profiling and awareness issues in tools supporting CoPs
- adaptability issues in tools supporting CoPs
- visualization issues in tools supporting CoPs
- Web-based interactive applications

The workshop is organised in the frame of the EC funded 'Palette' project.

More information:
Workshop web page:
http://palette.cti.gr/workshops/telcops06.htm

Palette project:
http://palette.ercim.org/

CALL FOR PARTICIPATION

The 9th ERCIM Workshop "User Interfaces for All"

Königswinter (Bonn), Germany, 27-28 September 2006

In the tradition of its predecessors, this workshop aims to consolidate recent work, and to stimulate further discussion on the state of the art in the field of User Interfaces for All, and its increasing range of applications in the emerging Information Society. The emphasis of this year's event is on "Universal Access in Ambient Intelligence Environments"

The workshop will therefore focus on the new HCI challenges that Ambient Intelligence brings about in a Universal Access perspective, with the aim to envisage new scenarios of use of Ambient Intelligence technologies by users with diverse needs and requirements, and to identify some of the critical issues that will have to be addressed throughout all phases and aspects of the development life-cycle of interactive applications and services.

Keynote speakers:

- Norbert Streitz, FhG-IPSI, Germany
- Alois Ferscha, Institut für Pervasive Computing, Johannes Kepler Universität Linz, Austria

More information:
http://ui4all.ics.forth.gr/workshop2006/

CALL FOR PARTICIPATION

TED 06: Towards e-Democracy: Partticipation, Deliberation, Communities

Mantova, Italy, 24-26 October 2006

For the past four years, the European Science Foundation programme Towards Electronic Democracy (TED) has focused on the development of methods to address societal issues via the Web and favour e-participation using the methodologies of modern decision analysis and support to involve citizens and stakeholders in the actual process of decision making: a true step towards e-democracy rather than the e-administration techniques that, by and large, have been emphasised by e-government initiatives. At TED's heart is a vision to develop methodologies which enable multiple decision analyses to be communicated, explored and, indeed, built over the WWW, thus providing the mechanism by which stakeholders may be drawn more closely into the decision making process.

This conference occurs at the end of TED's funding cycle and aims both to reflect on progress over the project and to set future research agendas.

More information:
http://www.mi.imati.cnr.it/conferences/ted06.html

CALL FOR PARTICIPATION

FMCO 2006 - 5th International Symposium on Formal Methods for Components and Objects

Amsterdam, 7-10 November 2006

Large and complex software systems provide the necessary infrastucture in all industries today. In order to construct such large systems in a systematic manner, the focus in the development methodologies has switched in the last two decades from functional issues to structural issues: both data and functions are encapsulated into software units which are integrated into large systems by means of various techniques supporting reusability and modifiability. This encapsulation principle is essential to both the object-oriented and the more recent component-based sofware engineering paradigms.

The symposium, hosted by CWI, is a four days event organized to provide an atmosphere that fosters collaborative work, discussions and interaction. The program consists of keynote and tutorial presentations.

Keynote speakers include:

- Gul Agha (The University of Illinois at Urbana-Champaign, USA)
- Sophia Drossopoulou (Imperial College, UK)
- Radu Iosif (Verimag, France)
- Thierry Jeron (INRIA Rennes, France)
- Erik Meijer, (Microsoft research, USA)
- Jayadev Misra (University of Texas at Austin, USA)
- Vijay A. Saraswat (Penn State University, USA)
- Vladimiro Sassone (University of Sussex, UK)
- Jan Tretmans (Radboud University Nijmegen, The Netherlands)
- Moshe Vardi (Rice University, USA)
- Philip Wadler (University of Edinburgh, UK)

More information:
http://www.mi.imati.cnr.it/conferences/ted06.html

ERCIM NEWS

ERCIM News is the magazine of ERCIM. Published quarterly, the newsletter reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of 10,500 copies. The printed version of ERCIM News has a production cost of 8 Euro per copy. It is available free of charge for certain groups.

Advertising
For current advertising rates and conditions, see http://www.ercim.org/publication/ERCIM_News/ or contact office@ercim.org

ERCIM News online edition
http://www.ercim.org/publication/ERCIM_News/

ERCIM News is published by ERCIM EEIG, BP 93, F-06902 Sophia-Antipolis Cedex
Tel: +33 4 9238 5010, E-mail: office@ercim.org
ISSN 0926-4981

Director: Jérôme Chailloux, ERCIM Manager

Central Editor:
Peter Kunz, ERCIM office
peter.kunz@ercim.org

Local Editors:
AARIT: n.a.
CCLRC: Martin Prime
M.J.Prime@rl.ac.uk
CRCIM: Michal Haindl
haindl@utia.cas.cz
CWI: Annette Kik
Annette.Kik@cwi.nl
CNR: Carol Peters
carol.peters@isti.cnr.it
FORTH: Eleni Orphanoudakis
eleni@ics.forth.gr
Fraunhofer ICT Group:
Michael Krapp
michael.krapp@scai.fraunhofer.de
FNR: Patrik Hitzelberger
hitzelbe@lippmann.lu
FWO/FNRS: Benoît Michel
michel@tele.ucl.ac.be
INRIA: Bernard Hidoine
bernard.hidoine@inria.fr
Irish Universities Association:
Ray Walshe
ray@computing.dcu.ie
NTNU: Truls Gjestland
truls.gjestland@ime.ntnu.no
SARIT: Harry Rudin
hrudin@smile.ch
SICS: Kersti Hedman
kersti@sics.se
SpaRCIM: Salvador Lucas
slucas@dsic.upv.es
SZTAKI: Erzsébet Csuhaj-Varjú
csuhaj@sztaki.hu
VTT: Pia-Maria Linden-Linna
pia-maria.linden-linna@vtt.fi
W3C: Marie-Claire Forgue
mcf@w3.org

Subscription
Subscribe to ERCIM News by:
- sending e-mail to your local editor
- contacting the ERCIM office (see address above)
- filling out the form at the ERCIM website at http://www.ercim.org/

INRIA - **Michel Cosnard appointed as INRIA's new chairman**. By an order of the French president signed on 2 May 2006, Michel Cosnard has been named as Chairman of INRIA. In accordance with INRIA's internal regulations, the chairman will also assume the responsibilities of the Institute's CEO. He succeeds Gilles Kahn who passed away 9 February 2006. Michel Cosnard represents INRIA on ERCIM's board of Directors.


© INRIA / C. Lebedinsky

Michel Cosnard.

Michel Cosnard is a Professor at the Polytechnic School of the University of Nice-Sophia Antipolis, Director of the INRIA Sophia Antipolis research unit and member of the board of the 'Communicating Secured Solutions' cluster in the Provence Alpes Côte d'Azur region. He already served as INRIA's CEO from December 2003 to May 2004, but he asked to step down from this position for personal reasons.

Michel Cosnard is a worldwide specialist in algorithm, especially in the design and analysis of parallel algorithms and Grid computing. He has also worked on the automaton and neural network complexity. He is the author of over 100 published works in the most prestigious journals of the field. He is Editor-in-Chief of Parallel Processing Letters, Member of the Editorial Board of Journal of Parallel Computing, and he has served as editor of IEEE Transactions of Parallel and Distributed Systems. He has written two books and supervised 27 theses. He has been awarded the following prizes: the Academy of Science Alfred Verdaguer Award (1994), the IFIP Silver Core Award (1995), the Charles Babbage Award from the Institute of Electrical and Electronics Engineers Computer Society (2003).

SpaRCIM - **The 2006 Spanish National Awards in Informatics** were announced in May 2006. The José García Santesmases Award to the most outstanding professional career was given ex-aequo to Prof. Isidro Ramos Salavert from the Technical University of Valencia (UPV), Prof. Fernando Sáez Vacas from the Technical University of Madrid (UPM) and Prof. Martí Vergés Trías from the Technical University of Catalonia (UPC). The Aritmel Award for the researcher developing the most significant scientific contributions to Informatics Engineering was given to Emilio López-Zapata from the University of Granada.

From left: Alberto Prieto Espinosa, Isidro Ramos, Emilio López-Zapata, and Fernando Sáez Vacas.

Also, two national awards recognized the activity of private and public institutions in the area. The Mare Nostrum Award was given to Grupo Telefónica (http://www.telefonica.es/acercadetelefonica/eng/index.shtml). The Ramón Llull Award for the institutional activity in Informatics Engineering was given ex-aequo to Prof. Alberto Prieto Espinosa, from the University of Granada and Juan José Moreno-Navarro, from the Technical University of Madrid (UPM). Juan José Moreno-Navarro is also the Director of SpaRCIM.

SICS - **Center for Networked Systems established** - a new joint industry-academia research center for the reliable Internet. SICS is one of the winners when Swedish government agencies invest 33 million Euro in eight centers for research important for the country's future competitive strength. Each of these Institute Excellence Centre will receive funding over a six-year-period, matched by corresponding funding from the business community, to build up an internationally competitive environment for research, development and innovation. The vision of SICS Center for Networked Systems is the Reliable Internet, a secure and reliable infrastructure which provides predictable service, enables robust applications on heterogeneous networks, is secure and at the same time easier to manage. SICS Center for Networked Systems is led by Bengt Ahlgren.

Photo by Fredrik Olsson.

Bengt Ahlgren, leader of SICS Center for Networked Systems.

CWI - **Peter Boncz wins ICTRegie Award 2006**. Analyzing complex databases in record time: For his searching techniques Peter Boncz from CWI won the Dutch ICTRegie Award 2006 on May 16. Boncz developed the fast MonetDB database system. It has applications in CRM, digital forensics, science databases and ambient intelligence. With this technology, CWI could launch a successful spin-off company: Data Distilleries ('95), now taken over by SPSS. Martin Rem, chair of the jury at the Nationale ICT Awards 2006 event said: "The challenges for Peter Boncz were not only scientific. The real art was coupling research results to a convincing business model - and he succeeded." Peter Boncz works at CWI in the MultimediaN Bsik program. See: http://monetdb.cwi.nl.

Photo by Gerrit Sarne.

Peter Boncz with the ICTRegie Award.

Francesco Beltrame.

CNR - **Francesco Beltrame** has been nominated Director of the Department for Information and Communications Technologies of the Italian National Research Council, one of the eleven macro research areas resulting from the recent restructuring of CNR. The Department is responsible for the coordination and evaluation of the scientific and technical activities of the seven CNR Institutes working in the ICT sector. Professor Beltrame holds the Chair of BioEngineering at the University of Genoa and he is President of the Scientific and Technical Committee for Industrial Research of the Italian Ministry of Education, University and Research. He is also Italian representative at the European Commission for the IST programme under FP6.

CWI - **CWI was rated 'excellent'** by an international evaluation committee in March 2006. "The combination of mathematics and computer science and fundamental and applied research gives the institute a strong and unique position in the international research landscape," the Netherlands Organisation for Scientific Research NWO said in its press release. NWO subjected six Dutch institutes to an external evaluation. The evaluation committee comprised Frank den Hollander (Technische Universiteit Eindhoven), Christopher Baker (University of Manchester), Susan Graham (University of California, Berkeley), Wendy Hall (University of Southampton) and Kurt Mehlhorn (Max Planck Institute for Computer Science, Saarbrücken).

CWI - **Krzysztof Apt** elected as Member of the Academia Europaea. Krzysztof Apt (CWI and Universiteit van Amsterdam) has been elected as Member of the Academia Europaea in the Informatics Section on 26 April. This section has 66 members of whom nine scientists come from ERCIM member institutes. The Academia Europaea is a European, non-governmental association acting as an Academy. Its members are scientists and scholars who collectively aim to promote learning, education and research. See: http://www.acadeuro.org/

ERCIM – The European Research Consortium for Informatics and Mathematics is an organisation dedicated to the advancement of European research and development, in information technology and applied mathematics. Its national member institutions aim to foster collaborative work within the European research community and to increase co-operation with European industry.

ERCIM is the European Host of the World Wide Web Consortium.

Austrian Association for Research in IT
c/o Österreichische Computer Gesellschaft
Wollzeile 1-3, A-1010 Wien, Austria
Tel: +43 1 512 02 35 0, Fax: +43 1 512 02 35 9
http://www.aarit.at/

Council for the Central Laboratory of the Research Councils, Rutherford Appleton Laboratory
Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom
Tel: +44 1235 82 1900, Fax: +44 1235 44 5385
http://www.cclrc.ac.uk/

Consiglio Nazionale delle Ricerche, ISTI-CNR
Area della Ricerca CNR di Pisa,
Via G. Moruzzi 1, 56124 Pisa, Italy
Tel: +39 050 315 2878, Fax: +39 050 315 2810
http://www.isti.cnr.it/

Czech Research Consortium
for Informatics and Mathematics
FI MU, Botanicka 68a, CZ-602 00 Brno, Czech Republic
Tel: +420 2 688 4669, Fax: +420 2 688 4903
http://www.utia.cas.cz/CRCIM/home.html

Centrum voor Wiskunde en Informatica
Kruislaan 413, NL-1098 SJ Amsterdam,
The Netherlands
Tel: +31 20 592 9333, Fax: +31 20 592 4199
http://www.cwi.nl/

Fonds National de la Recherche
6, rue Antoine de Saint-Exupéry, B.P. 1777
L-1017 Luxembourg-Kirchberg
Tel. +352 26 19 25-1, Fax +352 26 1925 35
http:// www.fnr.lu

FWO
Egmontstraat 5
B-1000 Brussels, Belgium
Tel: +32 2 512.9110
http://www.fwo.be/

FNRS
rue d'Egmont 5
B-1000 Brussels, Belgium
Tel: +32 2 504 92 11
http://www.fnrs.be/

Foundation for Research and Technology – Hellas
Institute of Computer Science
P.O. Box 1385, GR-71110 Heraklion, Crete, Greece
Tel: +30 2810 39 16 00, Fax: +30 2810 39 16 01
http://www.ics.forth.gr/

Fraunhofer ICT Group
Friedrichstr. 60
10117 Berlin, Germany
Tel: +49 30 726 15 66 0, Fax: ++49 30 726 15 66 19
http://www.iuk.fraunhofer.de

Institut National de Recherche en Informatique
et en Automatique
B.P. 105, F-78153 Le Chesnay, France
Tel: +33 1 3963 5511, Fax: +33 1 3963 5330
http://www.inria.fr/

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and
Electrical Engineering, N 7491 Trondheim, Norway
Tel: +47 73 59 80 35, Fax: +47 73 59 36 28
http://www.ntnu.no/

Spanish Research Consortium for Informatics
and Mathematics c/o Esperanza Marcos, Rey Juan Carlos
University, C/ Tulipan s/n, 28933-Móstoles, Madrid, Spain,
Tel: +34 91 664 74 91, Fax: 34 91 664 74 90
http://www.sparcim.org

Swedish Institute of Computer Science
Box 1263,
SE-164 29 Kista, Sweden
Tel: +46 8 633 1500, Fax: +46 8 751 72 30
http://www.sics.se/

Swiss Association for Research in Information Technology
c/o Prof. Dr Alfred Strohmeier, EPFL-IC-LGL,
CH-1015 Lausanne, Switzerland
Tel +41 21 693 4231, Fax +41 21 693 5079
http://www.sarit.ch/

Magyar Tudományos Akadémia
Számítástechnikai és Automatizálási Kutató Intézet
P.O. Box 63, H-1518 Budapest, Hungary
Tel: +36 1 279 6000, Fax: + 36 1 466 7503
http://www.sztaki.hu/

Irish Universities Association
c/o School of Computing, Dublin City University
Glasnevin, Dublin 9, Ireland
Tel: +3531 7005636, Fax: +3531 7005442
http://ercim.computing.dcu.ie/

Technical Research Centre of Finland
P.O. Box 1200
FIN-02044 VTT, Finland
Tel:+358 9 456 6041, Fax :+358 9 456 6027
http://www.vtt.fi/tte

Order Form

If you wish to subscribe to ERCIM News ***free of charge*** *or if you know of a colleague who would like to receive regular copies of ERCIM News, please fill in this form and we will add you/them to the mailing list.*

Send, fax or email this form to:

ERCIM NEWS
2004 route des Lucioles
BP 93
F-06902 Sophia Antipolis Cedex
Fax: +33 4 9238 5011
E-mail: office@ercim.org

Data from this form will be held on a computer database. By giving your email address, you allow ERCIM to send you email

I wish to subscribe to the

☐ ***printed edtion*** ☐ ***online edition (email required)***

Name:

Organisation/Company:

Address:

Postal Code:

City:

Country

E-mail:

You can also subscribe to ERCIM News and order back copies by filling out the form at the ERCIM website at