CONFERENCE ORGANIZERS INSTITUTES

The International Conference on Integrated Information is supported by the following Institutes:

Emerald Group Publishing Limited
Technological educational Institute of Athens, Greece
University of Peloponnese, Greece
National And Kapodistrian University of Athens, Greece
Mednet Hellas, The Greek Medical Network
2\textsuperscript{nd} AMICUS Workshop

To learn more about I-DAS, including the Book Series, please visit the webpage
http://www.i-das.org/
INTEGRATED INFORMATION

International Conference on Integrated Information

Kos, Greece September, 29 – October, 3 2011

EDITORS

Georgios A. Giannakopoulos
Technological Educational Institute of Athens, Greece

Damianos P. Sakas
University of Peloponnese, Greece

All papers have been peer-reviewed

Piraeus, Greece, 2011
Editors

Georgios A. Giannakopoulos
Technological Educational Institute of Athens
Faculty of Management and Economics
Department of Library Science and Information Systems
Address: Aghiou Spyridonos Street, 12210, Egaleo
E-mail: gian@teiath.gr

Damianos P. Sakas
University of Peloponnese
Faculty of Science and Technology
Department of Computer Science and Technology
Address: End of Karaiskaki St., 22100, Tripolis, Greece
E-mail: D.Sakas@uop.gr

The copyrights will be owned by the authors under the Creative Commons Attribution-Non Commercial license (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted use, distribution, and reproduction in any non-commercial medium, provided the original work is properly cited.

ISSN:

Printed in the Greece, EU
CONTENTS

PREFACE: Proceedings of the International Conference on Integrated Information (IC-INFO 2011) 1
Georgios A. Giannakopoulos, Damianos P. Sakas

Conference Details 3

Keynote Speaker 5

SYMPOSIUM ON INFORMATION AND KNOWLEDGE MANAGEMENT 6
Prof. Christos Skourlas

Towards the Preservation and Availability of Historical Books and Manuscripts: A Case Study 8
Eleni Galiotou

An Extensive Experimental Study on the Cluster-based Reference set Reduction for Speeding-up the k-nn Classifier 12
Stefanos Ougiaroglou, Georgios Evangelidis and Dimitris A. Dervos

Exploiting the Search Culture Modulated by the Documentation Retrieval Applications 16
Nikitas N. Karanikolas and Christos Skourlas

Information and Knowledge Organization: The Case of the TEI of Athens 22
Anastasios Tsolakidis, Manolis Chalaris and Ioannis Chalaris

Providing Access to Students with Disabilities and Learning Difficulties in Higher Education through a Secure Wireless framework 26
Catherine Marinagi and Christos Skourlas

Improving Query Efficiency in High Dimensional Point Indexes 30
Evangelos Outsios and Georgios Evangelidis

Text Segmentation Using Named Entity Recognition and co-Reference Resolution in Greek Texts 34
Pavlina Fragkou

KINISIS, a Graphical XQuery Language 42
Euclid Keramopoulos, Achilleas Pliakas, Konstantinos Tsekos and Ignatios Deligiannis

Dimensionality Curse, Concentration Phenomenon and the KDB-tree 46
Nikolaos Kouiroukidis and Georgios Evangelidi
Applying Balanced Scorecard Strategic Management in Higher Education
Manolis Chalaris, Anastasios Tsolakidis and Ioannis Chalaris

A Web Portal Model for NGOs’ Knowledge Management
Zuhal Tanrikulu

The Digital Archives System and Application Optimized for the Tradition Knowledge Archives

A Semi-automatic Emerging Technology Trend Classifier Using SCOPUS and PATSTAT
Seonho Kim, Woondong Yeo, Byong-Youl Coh, Waqas Rasheed, Jaewoo Kang

Presenting a Framework for Knowledge Management within a Web Enabled Living Lab
Lizette de Jager and Albertus AK Buitendag and Potjie (JS) van der Walt

4TH SYMPOSIUM ON BUSINESS AND MANAGEMENT AND DYNAMIC SIMULATION MODELS SUPPORTING MANAGEMENT STRATEGIES
Dr. Damianos Sakas

Evangelia N. Markaki, Damianios P. Sakas and Theodore Chadipantelis

Free Software – Open Source Software. A Powerful Tool for Developing Creativity in the Hands of the Student
Nasiopoulos K. Dimitrios, Damianos P. Sakas, Konstantinos Masselos

Open Source Web Applications. How it Spread Through the Internet and their Contribution to Education.
Nasiopoulos K. Dimitrios, Damianos P. Sakas, Konstantinos Masselos

Culture in Modern Times in the Frame of Luhmann’s System Theory
Anastasia J. Chournazidis

Managing Scientific Journals: A Cultural Viewpoint
Marina C. Terzi, Damianos P. Sakas, and Ioannis Seimenis

A Conceptual Framework for Analyzing Knowledge-based Entrepreneurship
Nikos S. Kanellos
SESSION ON INFORMATION HISTORY: PERSPECTIVES, METHODS AND CURRENT TOPICS

Prof. Laszlo Karvalics

Emerging Research Fields in Information History
Laszlo Z. Karvalics

Information Management through Elementary Data Clusters: New Observations on Pridianum-Type Roman Statistical Documents
Gergő Gellérfi

Information and Secrecy on the Silk Road. Methods of Encryption of Legal Documents in Inner Asia (3th-4th century)
Szabolcs Felföldi

The Role of Information and Disinformation in the Establishment of the Mongolian Empire: A Re-examination of the 13th century Mongolian History from the Viewpoint of Information History
Márton Gergő Vér

Early Warning Systems and the Hospitallers in the Eastern Mediterranean
Zsolt Hunyadi

Information Management as Establishment Dutch Navigational Knowledge on Japan, 1608-1641
Gabor Szommer

Files Everywhere - Register and Training of Men for Military and Civil Purpose in Prussia in the early 18th century
Marton Holczer

SYMPOSIUM ON INTEGRATED INFORMATION: THEORY, POLICIES, TOOLS

Prof. Georgios Giannakopoulos

Approaching Information as an Integrated Field: Educating Information Professionals
Georgios Giannakopoulos, Daphne Kyriaki Manesi and Sryridon Zervos

Special Libraries as Knowledge Management Centers
Eva Semertzaki

Digital Libraries’ Developers and their Suitability: A Case Study
Maria Monopoli
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Preliminary Study for the Creation of a Greek Citation index in the Humanities and the Social Sciences (GCI – H&amp;SS)</td>
<td>140</td>
</tr>
<tr>
<td>Daphne Kyriaki-Manessi and Evi Sachini</td>
<td></td>
</tr>
<tr>
<td>Archiving as an Information Science. Evidence from a Survey Carried out on a Sample of Greek Students</td>
<td>144</td>
</tr>
<tr>
<td>Georgios Giannakopoulos and Ioannis Koumantakis</td>
<td></td>
</tr>
<tr>
<td>Transition Process of E-records Management and Archiving System in Universities: Ankara University</td>
<td>147</td>
</tr>
<tr>
<td>S. Özlem Bayram and Fahrettin Ozdemirci</td>
<td></td>
</tr>
<tr>
<td>Government Information: Access and Greece’s Efforts for Access</td>
<td>150</td>
</tr>
<tr>
<td>Aikaterini Yiannoukakou</td>
<td></td>
</tr>
<tr>
<td>School Archives and their Potentials in Teaching: Aspects of Greek Reality</td>
<td>156</td>
</tr>
<tr>
<td>Sonia Geladaki and Panagiota Papadimitriou</td>
<td></td>
</tr>
<tr>
<td>Research on School Libraries in Greece and Suggestions on its Further Development</td>
<td>160</td>
</tr>
<tr>
<td>Georgios D. Bikos</td>
<td></td>
</tr>
<tr>
<td>Building Digital Collections for Archeological Sites: Metadata Requirements and CIDOC CRM Extension</td>
<td>164</td>
</tr>
<tr>
<td>Georgios S. Gkrous and Mara Nikolaidou</td>
<td></td>
</tr>
<tr>
<td>Museological Claims to Autonomous Knowledge: Rethinking the Conceptual Mode of Display and its Claims to Knowledge</td>
<td>169</td>
</tr>
<tr>
<td>Assimina Kaniari and Georgios Giannakopoulos</td>
<td></td>
</tr>
<tr>
<td>Use of Library Loan Records for Book Recommendation</td>
<td>172</td>
</tr>
<tr>
<td>Keita Tsuji, Erika Kuroo, Sho Sato, Ui Ikeuchi, Atsushi Ikeuchi, Fuyuki Yoshikane and Hiroshi Itsumura</td>
<td></td>
</tr>
<tr>
<td>Developing a National Database on Librarianship and Information Science. The Case of E-VIVA, the Hellenic Fulltext Database</td>
<td>176</td>
</tr>
<tr>
<td>Filippos Ch. Tsimpoglou, Vasiliki V. Koukoundou and Eleni K. Sakka</td>
<td></td>
</tr>
<tr>
<td>Integrated Access to Cultural Heritage Information Pieces in Iran Astan-Quds Razavi’s Organization of Libraries, Museums and Documents Center: A Theory of Unionization Disparate Information Assets over Imam Reza’s Zarih</td>
<td>181</td>
</tr>
<tr>
<td>Ms. Mitra Zarei and Ms. Maliheh Farrokhnia</td>
<td></td>
</tr>
<tr>
<td>Attitudes of University Librarians and Information Scientists towards the Draft Code of</td>
<td>185</td>
</tr>
</tbody>
</table>
Library Ethics to Present a Model for Final Library Ethical Codes
Mahsoomeh Latifi, Fatemeh Zandian and Hasan Siamian

SESSION ON OPEN ACCESS REPOSITORIES: SELF-ARCHIVING, METADATA, CONTENT POLICIES, USAGE
Dr. Alexandros Koulouris

Geographical Collections in Greek Academic Libraries: Current Situation and Perspectives
Ifigenia Vardakosta and Sarantos Kapidakis

Information Seeking Behavior: Factors that Affect the Behavior of Greek Astronomers
Hara Brindesi and Sarantos Kapidakis

Aggregating Metadata for Europeana: The Greek Paradigm
Alexandros Koulouris, Vangelis Banos and Emmanouel Garoufallou

Integrating a Repository with Research Output and Publications: The Case of the National Technical University of Athens
Dionysis Kokkinos

Implementation of Workflows as Finite State Machines in a National Doctoral Dissertations Archive
Nikos Houssos, Dimitris Zavaliadis, Kostas Stamatis and Panagiotis Stathopoulos

Practices of “Local” Repositories of Legally Protected Immovable Monuments. A Global Scheme for ‘Designation – Significance’ Information
Michail Agathos and Sarantos Kapidakis

Integration of Metadata in BWMETA-2.0.0 Format
Katarzyna Zamlynska, Jakub Jurkiewicz and Lukasz Bolikowski

SESSION ON EVIDENCE-BASED INFORMATION IN CLINICAL PRACTICE
Dr. Evangelia Lappa

Applicability of Data Mining Algorithms on Clinical Datasets
Wilfred, Bonney

Changing Roles of Health Librarians with Open Access Repositories
Christine Urquhar and Assimina Vlachaki

From Medical Records to Health Knowledge Management Systems: The Coding to Health Sector
Evangelia C. Lappa and Georgios A. Giannakopoulos
The Survey of Skill, Attitude and Use of Computer and Internet among Faculty Members  
Hasan Siamian, Azita Bala Ghafari, Kobra Aligolbandi, Mohammad Vahedi and Gholam Ali Golafshani Jooybari

Trends in Scholarly Communication among Biomedical Scientists in Greece  
Assimina Vlachaki and Christine Urquhart

SESSION ON ELECTRONIC PUBLISHING: A DEVELOPING LANDSCAPE  
Dr. Dimitris Kouis

E-Journal and Open Access Journal Publishing in the Humanities: Preliminary Results from a Survey among Byzantine Studies Scholars  
Victoria Tsoukala and Evi Sachini

Preliminary Results on a Printed VS Electronic Text Books Assessment Through Questionnaire  
Dimitrios A. Kouis and Kanella Pouli

An Interpretation of Aristotelian Logic According to George Boole  
Markos N. Dendrinos

SESSION ON INFORMATION CONTENT PRESERVATION AS OUTCOME OF CONSERVATION OF CULTURAL HERITAGE: ETHICS, METHODOLOGY AND TOOLS  
Prof. George Panagiaris and Dr. Spiros Zervos

Intrinsic Data Obfuscation as the Result of Book and Paper Conservation Interventions  
Spiros Zervos, Alexandros Koulouris and Georgios Giannakopoulos

Mass Deacidification: Preserving More than Written Information  
Michael Ramin, Evelyn Eisenhauer and Markus Reist

Information Literacy of Library Users: A Case Study of Mazandaran Public Library Users, Iran  
Hussein Mahdizadeh and Hasan Siamian

The Narratives of Paper in The Archives of the New Independent Greek State (Mid 19th c.)  
Ourania Kanakari and Maria Giannikou

From Macro to Micro and from Micro to Nano: The Evolution of the Information Content Preservation of Biological Wet Specimen Collections  
Nikolaos Maniatis and Georgios Panagiaris
Digital images: A valuable scholar's tool or misleading material?  
Patricia Engel  

Attitudes of University Librarians and Information Scientists Towards the Draft Code of Library Ethics to Present a Model for Final Library Ethical Codes  
Mahsoomeh Latifi, Fatemeh Zandianand and Hasan Siamian  

Ekaterini Malea, Effie Papageorgiou and Georgios Panagiaris  

SESSION ON DIVERGENCE AND CONVERGENCE: INFORMATION WORK IN DIGITAL CULTURAL MEMORY INSTITUTIONS  
Dr. Susan Myburgh  

Extending Convergence and Divergence in Cultural Memory Institutions: The Old Slave Lodge in the New South Africa  
Archie L Dick  

The Transfer of Knowledge from Large Organizations to Small: Experiences from a Research Project on Digitization in Wales  
Clare Wood-Fisher, Richard Gough, Sarah Higgins, Menna Morgan, Amy Staniforth and Lucy Tedd  

The Usage of Reference Management Software (Rms) in an Academic Environment: A Survey at Tallinn University  
Enrico Francese  

Varialog: How to Locate Words in a French Renaissance Virtual Library  
Marie-Hélène Lay  

The Urges to Merge: A Theoretical Approach  
Susan Myburgh  

SYMPOSIUM ON ADVANCES INFORMATION FOR STRATEGIC MANAGEMENT  
Professor Nikolaos Konstantopoulos  

Empowerment in the Tax Office of Greece  
Antonios E. Giokas and Nikolaos P. Antonakas  

Building Absorptive Capacity Through Internal Corporate Venturing  
Ioannis M. Sotiriou and Alexandros I. Alexandrakis
The Monitoring Information System (M.I.S.) - An information and Management System for Projects Co-financed Under the National Strategic Reference Framework (NSRF) and the Community support framework (CSF)

Catherina G. Siampou, Eleni G. Fassou and Athanassios P. Panagiotopoulos

Corruption in Tax Administration: The Entrepreneurs View Point

Nikolaos P. Antonakas, Antonios E. Giokas and Nikolaos Konstantopoulos

Conflicts between the IT Manager and the Software House after the Strategic Choice of Outsourcing of the Information Processes in Maritime Companies.

Anthi Z. Vaxevanou, Nikolaos Konstantopoulos, Damianos P. Sakas

Contemporary Forms of Ordering Between the Supply Department and Ship Chandler Companies in the Shipping Industry

Anthi Z. Vaxevanou, Nikolaos Konstantopoulos, Damianos P. Sakas

Strategies Implemented and Sources Used for the Acquisition of Information on Foreign Markets

Myropi Garri, Nikolaos Konstantopoulos and Michail G. Bekiaris

The Effect of High Performance Working Systems on Informative Technology in Enterprises after Organisation Changes such as Mergers & Acquisitions

Nikolaos Konstantopoulos and Yiannis Triantafyllopoulos

Personnel’s Absorptive Capacity as a Guiding Concept for Effective Performance in Informative Technology

Nikolaos Konstantopoulos and Yiannis Triantafyllopoulos

SESSION ON CONTEMPORARY ISSUES IN MANAGEMENT: ORGANISATIONAL BEHAVIOUR, INFORMATION TECHNOLOG, EDUCATION & HOSPITAL LEADERSHIP

Dr. Panagiotis Trivellas

Investigating the Importance of Sustainable Development for Hotel SMES

Panagiotis Reklitis and Anestis Fotiadis

Strategic Alignment of ERP, CRM and E-business: A Value Creation

Catherine C. Marinagi and Christos K. Akrivos

The Impact of Occupational Stress on Performance in Health Care

Panagiotis Trivellas, Panagiotis Reklitis and Charalambos Platis
The Impact of Emotional Intelligence on Job Outcomes and Turnover Intention in Health Care
   Panagiotis Trivellas Vassilis Gerogiannis and Sofia Svarna

SYMPOSIUM ON BUSINESS MANAGEMENT AND COMMUNICATION STRATEGIES SUPPORTING DECISION MAKING PROCESS IN TOURISM SECTOR
   Dr. Panagiota Dionysopoulou

The Human Factor as a Mediator to the Total Quality in the Tourism Companies. The impact of Employees’ Motivation to Quality Improvements
   Christos K. Akrivos and Panagiotis Reklitis

Tourist Destination Marketing and Management Using Advanced ICTS Technologies
   Anastasia Argyropoulou, Panagiota Dionysopoulou, Georgios Miaoulis

   George Stafylakis and Panagiota Dionysopoulou

GENERAL PAPERS

The role of Environmental Education within the Framework of the Environmental Policy of a Regional Municipality
   Vassiliki Delitheou and Dimitra Thanasia

Issues of Social Cohesion: A case study from the Greek Urban Scenery
   Evgenia Tousi

Merging Activity and Employee Performance: The Greek Banking System
   Panagiotis Liargovas and Spyridon Repousis

Sustainable Development and Corporate Social Responsibility in Higher Education: Some Evidence from Greece
   Anastasios Sepetis and Fotios Rizos

Exploring the Effects of Organizational Culture on Collaborative vs. Competitive Knowledge Sharing Behaviors
   Hanan Abdulla Mohammed Al Mehairi and Norhayati Zakaria

GEORGIOS A. GIANNAKOPOULOS

Department of Library Science and Information Systems, Technological Educational Institute of Athens, Aghiou Spyridonos, Egaleo, 12210, Greece

DAMIANOS P. SAKAS

Department of Computer and Technology Science, University of Peloponnese, Praxitelous 89-91, Piraeus, 18532, Greece

Aims and Scope of the Conference

The International Conference on Integrated Information 2011 took place in Kos Island, Greece, between September, 29 and October, 3, 2011. IC-ININFO is an international interdisciplinary conference covering research and development in the field of information management and integration.

The conference aims at creating a forum for further discussion for an Integrated Information Field incorporating a series of issues and/or related organizations that manage information in their everyday operations. Therefore, the call for papers is addressed to scholars and/or professionals of the fields of Library and Archives Science (including digital libraries and electronic archives), Museum and Gallery Studies, Information Science, Documentation, Information Management, Records Management, Knowledge Management, Data management and Copyright experts the latter with an emphasis on Electronic Publications. Furthermore, papers focusing on issues of Cultural Heritage Management and Conservation Management are also be welcomed along with papers regarding the Management of Nonprofit Organizations such as libraries, archives and museums.

One of the primary objectives of the IC-ININFO will be the investigation of information-based managerial change in organizations. Driven by the fast-paced advances in the Information field, this change is characterized in terms of its impact on organizations that manage information in their everyday operations.

Grouping emerging technologies in the Information field together in a close examination of practices, problems and trends, IC-ININFO and its emphases on integration and management will present the state of the art in the field. Addressed jointly to the academic and practitioner, it will provide a forum for a number of perspectives based on either theoretical analyses or empirical case studies that will foster dialogue and exchange of ideas.

Topics of general Interest


Symposia

The Conference offered a number of sessions under its patronage, providing a concise overview of the most current issues and hands-on experience in information-related fields.

- Symposium on Integrated information: Theory, Policies, Tools
- 4th Symposium on Business and Management and Dynamic Simulation Models supporting management strategies
• Session on Open Access Repositories: Self-archiving, Metadata, Content policies, Usage
• Session on Evidence-Based Information in Clinical Practice
• Session on Business Management and Communication Strategies supporting Decision Making Process in Tourism Sector
• Session on Electronic Publishing: A Developing Landscape
• Session on Information and Knowledge Management
• Session on Information Content Preservation as Outcome of Conservation of Cultural Heritage: Ethics, Methodology and Tools
• Session on Advances Information for Strategic Management
• Session on Information History: Perspectives, Methods and Current Topics
• Session on Divergence and Convergence: Information Work in Digital Cultural Memory Institutions
• Session on Contemporary issues in Management: Organisational Behaviour, Information Technology, Education & Hospital leadership.

The wide range of aspects that the sessions covered, highlighted future trends in the Information Science.

**Paper Peer Review**

More than 300 papers had been submitted for consideration in IC-ININFO 2011. From them, 91 were selected for presentation, after peer review in a double blind review process. The accepted papers were presented at IC-ININFO 2011.

**Thanks**

We would like to thank all members that participated in any way in the IC-ININFO 2011 Conference and especially:

• The famous publishing house Emerald for its communication sponsorship.
• The co-organizing Universities and Institutes for their support and development of a high-quality Conference scientific level and profile.
• The members of the Scientific Committee that honored the Conference with their presence and provided a significant contribution to the review of papers as well as for their indications for the improvement of the Conference.
• All members of the Organizing Committee for their help, support and spirit participation before, during and after the Conference.
• The Session Organizers for their willing to organize sessions of high importance and for their editorial work, contributing in the development of valued services to the Conference.
• PhDc Marina Terzi for her excellent editorial work, contributing in the production of the Conference proceedings.
CONFERENCE DETAILS

Chairs

Georgios A. Giannakopoulos, Technological Educational Institute of Athens, Greece
Damianos P. Sakas, University of Peloponnese, Greece

Co-Chairs

Daphne Kyriaki – Manesi, Technological Educational Institute of Athens, Greece
Dimitrios Vlachos, University of Peloponnese, Greece

Scientific Committee

Amanda Spink, Queensland University of Technology
Andreas Bagias, European Court
Andreas Rauber, Vienna University of Technology
Astrid van Wesenbeeck, SPARC Europe
Christine Urquhart, Aberystwyth University
Christos Schizas, University of Cyprus
Christos Skourlas, Technological Educational Institute of Athens
Claire Farago, University of Colorado at Boulder
Claus-Peter Klas, FernUniversität in Hagen
Costas Vassilakis, University of Peloponnese,
Dimitris Dervos, Technological Educational Institute of Thessaloniki
Eelco Ferwerda, OAPEN
Elena Garcia Barriocanal, University of Alcalá
Emmanouel Garoufallou, Technological Educational Institute of Thessaloniki
Filippos Tsimoglou, University of Cyprus
Fillia Makedon, University of Texas at Arlington
George Korres, University of Newcastle
Georgios Evangelidis, University of Macedonia
Georgios Panagiaris, Technological Educational Institute of Athens
Johan Oomen, Netherlands Institute for Sound and Vision
José Aldana, University of Malaga
Konstantinos Masselos, University of Peloponnese
Luciana Duranti, The University of British Columbia
Markos N. Dendrinos, Technological Institute of Athens
Milena Dobreva, University of Strathclyde
Prodromos Tsiavos, London School of Economics and Political Science
Sándor Darányi, University of Borås
Sarantos Kapidakis, Ionian University
Sirje Virkus, Tallinn University
Spiros Zervos, Technological Educational Institute of Athens
Susan Myburgh, University of South Australia
Theodoros Pitsios, University of Athens, Faculty of Medicine

Organizing Committee

Alexandros Koulouris (Chair), Technological Educational Institute of Athens
Christos Christopoulos, SCEV Scientific Events Ltd
Marina Terzi, University of the Aegean, Greece
Evangelia Markaki, Aristotle University of Thessaloniki
Assimina Kaniari, Athens School of Fine Arts
Evangelia Lappa, General Hospital Attikis K.A.T.
Dimitris Kouis, Greek Ministry of Education, Lifelong Learning and Religious Affairs
Dionysis Kokkinos, National Technical University of Athens
KEYNOTE SPEAKER

Professor Amanda Spink

Professor Amanda Spink has published over 340 scholarly journal articles, refereed conference papers and book chapters, and 6 books. Many of her journal articles are published in the Journal of the American Society for Information Science and Technology, Information Processing and Management, and the Journal of Documentation. She is Editor of the Emerald journal Aslib Proceedings. Amanda’s research has been published at many conferences including ASIST, IEEE ITCC, CAIS, Internet Computing, ACM SIGIR, and ISIC Conferences. Her recent books include Information Behavior: An Evolutionary Instinct and Web Search: Multidisciplinary Perspectives, both published by Springer. Amanda’s research focuses on theoretical and empirical studies of information behavior, including the evolutionary and developmental foundations. The National Science Foundation, the American Library Association, Andrew R. Mellon Foundation, Amazon.com, Vivisimo. Com, Infospace.com, NEC, IBM, Excite.com, AlltheWeb.com, AltaVista.com, FAST, and Lockheed Martin have sponsored her research. In 2008 Professor Spink had the second highest H-index citation score in her field from 1998 to 2008 [Norris, M. (2008)]. Ranking Fellow Scholars and their H-Index: Preliminary Survey Results. Loughborough University, Dept of Information Science Report].
Implementation of Workflows as Finite State Machines in a National Doctoral Dissertations Archive

Nikos Houssos, Dimitris Zavaliadis, Kostas Stamatis and Panagiotis Stathopoulos

1 National Documentation Centre / National Hellenic Research Foundation, 11635, Athens, Greece.
nhoussos (at) ekt.gr, dimzava (at) ekt.gr, kstamatis (at) ekt.gr, pstath (at) ekt.gr

Abstract: Workflows for submission and processing of digital material is one of the important aspects of open access repositories and the major concerns in their implementation. The subject of the present paper is the modelling of processing workflows in the Hellenic National Archive of Doctoral Dissertations as Finite State Machines and their implementation through an original software library that enables client applications to create and utilise FSMs.

Keywords: Repositories, Workflows, Finite state machines, Electronic theses and dissertations, Object-oriented design.

I. INTRODUCTION

In this article we present the implementation of workflows for cataloguing and processing doctoral theses in the Greek National Archive of Doctoral Dissertations (HEDI) based on an original software library for Finite State Machines that has been developed by the Hellenic National Documentation Centre (EKT).

EKT has been granted by law 1566/1985 the responsibility of developing and maintaining the Greek National Archive of doctoral theses. The archive contains the doctoral dissertations produced in Higher Education Institutions as well as a number of PhD theses awarded to Greek scholars by universities outside Greece (USA, UK, Canada, and Germany, among others), in total about 24,500 theses as of February 2010, 2,75M pages of digitised and born-digital dissertations, with 1200 -1400 new dissertations arriving every year.

HEDI is supported by IT systems since 1986 when EKT developed the bibliographic database “National Archive of PhD Theses” employing for cataloguing the home-grown library automation software, ABEKT. Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε. Initially, EKT has been collecting from individual universities and cataloguing theses solely in print form. The database was made available to the public via the mainframe host computer ‘Hermes’ for more than a decade from 1986 until 1999. Thereafter, a new version of ABEKT has been used, including support for metadata standards like UNIMARC, UNIMARC Authorities and ISO 2709 and the Z39.50 protocol for search and retrieval of bibliographic records. This system, later integrated into the ARGO digital library portal (argo.ekt.gr) (Sfakakis, 2003) that is still in operation, provides free web-based access to metadata as well as advanced services to librarians, through a library catalogue-like user interface. Meanwhile, EKT proceeded with executing a major digitisation project for the majority of the dissertations in the – until then – print-only archive, which enabled open access to theses full text for Web users – realised through a specialised presentation application (Loverdos, 2001). Furthermore, in later years universities have been submitting theses to the archive, in both print and electronic form. As a result, today more than 75% of the theses in HEDI are available online in full text.

In 2009, a decision was made to re-build and modernise the information infrastructure supporting HEDI – a project that was completed in 2010. The following main choices were made:

- Create an e-theses repository on the DSpace platform, mainly targeting end-users.
- Create a separate IT application for the administration and management of the EKT internal workflows that are necessary for the processing of the material submitted to HEDI. The output of these workflows for every thesis is a quality-controlled metadata record and a searchable full-text file (or set of files). An important aspect to consider is that these workflows should be able to handle dissertations in both print and electronic form – note that the print archive is still maintained although a thesis is normally submitted in both print and electronic form. This application supports processing workflows for theses, acting as a service consumer for the workflow engine presented in this article.
- Maintain and constantly update the ARGO version of HEDI as a bibliographic system of choice for librarians offering to the latter additional important services like record export and transformation among various formats (e.g., MARC21 to UNIMARC). Ceasing to provide the HEDI database through this portal was considered to be a non sound option, given the popularity of ARGO among the Greek library community as well as the support of UNIMARC and related services.
- Select and reuse open source main software components, from the operating system to the repository and the middleware/database layer, while exploiting EKT’s virtual infrastructure (Stathopoulos, 2009) in order to provide HEDI services with higher availability, scalability and total cost of ownership. The same software...
components and technology have been previously used by EKT while developing institutional and subject repositories.

II. PROCESSING WORKFLOW FOR THESES
The internal workflow followed for the processing of the incoming dissertations material is depicted, at a high level of abstraction, in Figure 2 and can be described as follows:

1. The first step is the recording of the incoming material upon its receipt by EKT. This material includes the full-text dissertation document in print and electronic form (only print is currently mandatory); sometimes supplementary files (e.g., data sets) are provided. Furthermore, a declaration sheet by the doctoral degree holder about copyright issues is normally received – authors are being given the option to either make their full-text thesis publicly available immediately or after an embargo period (at most three years), as is common in other e-thesis systems. This sheet contains also basic descriptive metadata, comprising also author-defined keywords and abstract. Recording of physical (e.g., thesis hard copies) and digital object (files) is performed, as well as creating a very short metadata record including fields like author, degree awarding institution, date of degree award, date of submission to HEDI, public availability/copyright status as specified by the author, etc.

2. The second step in the workflow is the cataloguing of the thesis, to produce a detailed metadata record in the UNIMARC format; this is performed by qualified EKT/NHRF library personnel, starting either from scratch or from an ingested institutional repository record through a user-friendly interface. Furthermore, author-defined metadata is also taken into account, when available, and could be utilised after quality control. It is during the cataloguing phase that a unique identifier (number) is assigned to this particular thesis; this number is permanent and is used to identify all material related to the thesis within the archive. The assignment of the unique identifier is also a strict pre-requisite for starting the processing of the corresponding digital material (see step 3). After cataloguing, the metadata record (without the corresponding digital files – e.g. thesis full text) is published online in the repository and the bibliographic database.

3. The third phase is the processing of the material for the generation of a high quality, fully searchable digital file containing the dissertation full-text, as well as image files for individual pages to allow online page-by-page reading. This phase typically includes one of the following alternatives:
   - Digitisation, if the thesis document is available only in print form, which includes Optical Character Recognition (OCR) processing for making the resulting PDF file full-text searchable.
   - Production, based on the initially submitted digital file(s), of a thesis full-text file (in PDF format) ready for publication and subsequent preservation. This could entail, among others, transformation to PDF from other formats like .doc and .odp and checks on whether the dissertation text is fully searchable (otherwise, OCR is applied).
4. The last step in the workflow concerns the submission and storage of the digital material in a repository and the corresponding detailed record at the bibliographic database. Subsequent updates in the metadata and/or digital files are performed through the theses administration application and are propagated to both the repository and the bibliographic database. Future preservation actions are enacted on the digital files in the repository.

Figure 2 depicts the processing workflow modelled as a Finite State Machine (FSM) - a simplified version of the workflow is depicted for economy of presentation.

III. A SOFTWARE ENGINE FOR FINITE STATE MACHINES

To address the implementation of the aforementioned workflow in the system supporting HEDI, we have developed a software library, call FSM engine, for executing workflows modelled as Finite State Machines.

In particular, the FSM Engine is a Java API that allows definition and execution of workflows represented as state machines. It helps in cases where the behaviour of an object needs to be changed at runtime depending on its state, eliminating the need for extensive use of if/else and switch statements which make code unreadable and difficult to maintain. The FSM Engine is inspired by the State Design Pattern (Gamma, 1995) and has been developed by EKT.

The FSM Engine provides a set of reusable Java classes and interfaces that apply the abstract concepts of finite state machines at the source code level. By extending and implementing those classes and interfaces, a client application can systematically define how a system reacts to certain events avoiding at the same time cluttered case statements. This is achieved by specifying a set of valid states for an object along with possible transitions between states which are triggered by events. In addition, guard conditions can determine whether a transition should be executed or not and a set of possible actions can be associated with a given transition in order to be fired as a side effect of executing the transition. Branching is also supported, which enables multiple discrete transitions for a given event and at a given state – the evaluation of a boolean expression determines the specific transition to be followed in such a case.

The FSM engine domain model is depicted in Figure 3. The main classes comprising the engine are the following:

**StateContext**
Maintains a reference to the current state. It should be implemented by one or more client classes that can have different internal states and whose behavior changes depending on those states.

**State**
Represents the different states of the state machine. Each possible state of a class implementing the StateContext interface should correspond to a concrete implementation of this interface.

**StateName**
A Java enum acting as a bridge between StateContext and States.

**Event**
Represents an external or internal event such as a button click or a date expiration that can change the State of an object, thus leading to a state Transition in the system. Typically, there is only one Transition corresponding to a given Event but it is also possible to accommodate more than one Transitions for an Event via Branches and GuardConditions.

**Transition**
Defines the transition from one state to another as a response of the state machine to an occurrence of a specific Event. A Transition can be guarded by a GuardCondition which determines whether the execution of the Transition should proceed or not. In addition, a Transition can be associated with one or more Actions to be performed once the Transition has been executed.

**GuardCondition**
A GuardCondition is actually a Boolean expression that affects the behavior of the state machine by enabling Transitions only when it evaluates to true and disabling them when it evaluates to false. A GuardCondition can be associated with many Transitions but each Transition can have only one GuardCondition.

**Action**
Defines an activity that is to be performed when executing a certain Transition. An Action can be associated with more than one Transitions and a Transition can have more than one Actions. Also note that an Action can well trigger an Event having as a result another Transition.
**Branch**
Acts as a container of Transitions. It is used in cases where there is no one-to-one mapping between an Event and a Transition but instead there are more than one possible Transition for a given Event. In principle, the Transitions contained in a Branch are mutually exclusive, meaning that only one Transition will be executed each time and this will be determined once the GuardConditions of all Transitions have been evaluated.

**FiniteStateMachine**
An abstraction encapsulating the internal details of the state machine, presenting a single interface to the outside world. Implementation classes should be used as the main point where the lifecycle management of a StateContext takes place, whenever a triggering Event occurs.

In practice, to use the engine, the implementor of a workflow has to create concrete classes implementing interfaces like State, StateContext, Event, Transition, Action or the corresponding default classes offered by the library.

An important feature of the engine is that a specific workflow based on the implemented classes is specified outside the code in an XML configuration file – in particular an application context file of the Spring framework. The Spring dependency injection mechanisms are employed for an instantiation of a workflow/FSM engine. Therefore, modifications in a workflow (additions of states, transitions, events, actions, etc.) can be injected into the system without modifying source code – just by editing the configuration file.

**III. SUMMARY AND FUTURE WORK**
The main subject of the present article is the modelling of processing workflows in the Hellenic National Archive of Doctoral Dissertations as Finite State Machines and their implementation through an original software library that enables client applications to create and utilise FSMs. The adopted modelling and development approach proved to be appropriate for this use case.

Plans for further work regarding HEDI include the support of online submission workflows by authorised parties outside EKT (e.g. graduating doctoral students, university personnel) and automated ingest of records from external systems. These services require, among others, more sophisticated workflows for quality control both at the metadata and digital file level and further automation of digital file checking and processing.

Regarding the FSM engine, future plans include publishing and maintaining it as an autonomous open source project under a license that facilitates re-use in third-party applications. In terms of features, a priority is to make the engine even less intrusive for client applications, probably through mechanisms like Java annotations, aspect-oriented programming, and potentially metaprogramming in cases of integration with other than Java languages running on the Java Virtual Machine such as Groovy. Furthermore, support of advanced workflow features that go beyond standard FSM functionality will be investigated, such as clustering and orthogonality as supported by schemes like statecharts (Harel, 1987).

**REFERENCES**