

ADVANCES ON
INFORMATION
PROCESSING AND
MANAGEMENT

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
2nd 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	50
Manolis Chalaris, Anastasios Tsolakidis and Ioannis Chalaris	
A Web Portal Model for NGOs' Knowledge Management	54
Zuhal Tanrikulu	
The Digital Archives System and Application Optimized for the Tradition Knowledge Archives	58
Jeon Hong. Chan, In Deok. Hwang, Jae Hak. Park, Hyeok. Sim, U won. Gwon and Soon Cheol. Park	
A Semi-automatic Emerging Technology Trend Classifier Using SCOPUS and PATSTAT	62
Seonho Kim, Woondong Yeo, Byong-Youl Coh, Waqas Rasheed, Jaewoo Kang	
Presenting a Framework for Knowledge Management within a Web Enabled Living Lab	66
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	71
Dr. Damianos Sakas	
New Political Communication Practices: No Budget Events Management. The New Challenge	73
Evangelia N. Markaki, Damianios P. Sakas and Theodore Chadjipantelis	
Free Software – Open Source Software. A Powerful Tool for Developing Creativity in the Hands of the Student	78
Nasiopoulos K. Dimitrios, Damianos P. Sakas, Konstantinos Masselos	
Open Source Web Applications. How it Spread Through the Internet and their Contribution to Education.	82
Nasiopoulos K. Dimitrios, Damianos P. Sakas, Konstantinos Masselos	
Culture in Modern Times in the Frame of Luhmann's System Theory	85
Anastasia J. Chournazidis	
Managing Scientific Journals: A Cultural Viewpoint	87
Marina C. Terzi, Damianos P. Sakas, and Ioannis Seimenis	
A Conceptual Framework for Analyzing Knowledge-based Entrepreneurship	92
Nikos S. Kanellos	

SESSION ON INFORMATION HISTORY: PERSPECTIVES, METHODS AND CURRENT TOPICS	96
Prof. Laszlo Karvalics	
Emerging Research Fields in Information History	98
Laszlo Z. Karvalics	
Information Management through Elementary Data Clusters: New Observations on Pridianum-Type Roman Statistical Documents	102
Gergő Gellérfi	
Information and Secrecy on the Silk Road. Methods of Encryption of Legal Documents in Inner Asia (3th-4th century)	106
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	110
Márton Gergő Vér	
Early Warning Systems and the Hospitallers in the Eastern Mediterranean	114
Zsolt Hunyadi	
Information Management as Establishment Dutch Navigational Knowledge on Japan, 1608-1641	118
Gabor Szommer	
Files Everywhere - Register and Training of Men for Military and Civil Purpose in Prussia in the early 18th century	123
Marton Holczer	
SYMPOSIUM ON INTEGRATED INFORMATION: THEORY, POLICIES, TOOLS	126
Prof. Georgios Giannakopoulos	
Approaching Information as an Integrated Field: Educating Information Professionals	128
Georgios Giannakopoulos, Daphne Kyriaki Manesi and Stryidon Zervos	
Special Libraries as Knowledge Management Centers	132
Eva Semertzaki	
Digital Libraries' Developers and their Suitability: A Case Study	136
Maria Monopoli	

A Preliminary Study for the Creation of a Greek Citation index in the Humanities and the Social Sciences (GCI – H&SS)	140
Daphne Kyriaki-Manessi and Evi Sachini	
Archiving as an Information Science. Evidence from a Survey Carried out on a Sample of Greek Students	144
Georgios Giannakopoulos and Ioannis Koumantakis	
Transition Process of E-records Management and Archiving System in Universities: Ankara University	147
S. Özlem Bayram and Fahrettin Ozdemirci	
Government Information: Access and Greece's Efforts for Access	150
Aikaterini Yiannoukakou	
School Archives and their Potentials in Teaching: Aspects of Greek Reality	156
Sonia Geladaki and Panagiota Papadimitriou	
Research on School Libraries in Greece and Suggestions on its Further Development	160
Georgios D. Bikos	
Building Digital Collections for Archeological Sites: Metadata Requirements and CIDOC CRM Extension	164
Georgios S. Gkrous and Mara Nikolaidou	
Museological Claims to Autonomous Knowledge: Rethinking the Conceptual Mode of Display and its Claims to Knowledge	169
Assimina Kaniari and Georgios Giannakopoulos	
Use of Library Loan Records for Book Recommendation	172
Keita Tsuji, Erika Kuroo, Sho Sato, Ui Ikeuchi, Atsushi Ikeuchi, Fuyuki Yoshikane and Hiroshi Itsumura	
Developing a National Database on Librarianship and Information Science. The Case of E-VIVA, the Hellenic Fulltext Database	176
Filippos Ch. Tsimpoglou, Vasiliki V. Koukounidou and Eleni K. Sakka	
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	181
Ms. Mitra Zarei and Ms. Maliheh Farrokhnia	
Attitudes of University Librarians and Information Scientists towards the Draft Code of	185

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	188
Dr. Alexandros Koulouris	
Geographical Collections in Greek Academic Libraries: Current Situation and Perspectives	189
Ifigenia Vardakosta and Sarantos Kapidakis	
Information Seeking Behavior: Factors that Affect the Behavior of Greek Astronomers	194
Hara Brindesi and Sarantos Kapidakis	
Aggregating Metadata for Europeana: The Greek Paradigm	198
Alexandros Koulouris, Vangelis Banos and Emmanouel Garoufallou	
Integrating a Repository with Research Output and Publications: The Case of the National Technical University of Athens	202
Dionysis Kokkinos	
Implementation of Workflows as Finite State Machines in a National Doctoral Dissertations Archive	205
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	209
Michail Agathos and Sarantos Kapidakis	
Integration of Metadata in BWMETA-2.0.0 Format	213
Katarzyna Zamlynska, Jakub Jurkiewicz and Lukasz Bolikowski	
SESSION ON EVIDENCE-BASED INFORMATION IN CLINICAL PRACTICE	216
Dr. Evangelia Lappa	
Applicability of Data Mining Algorithms on Clinical Datasets	218
Wilfred, Bonney	
Changing Roles of Health Librarians with Open Access Repositories	221
Christine Urquhar and Assimina Vlachaki	
From Medical Records to Health Knowledge Management Systems: The Coding to Health Sector	225
Evangelia C. Lappa and Georgios A. Giannakopoulos	

The Survey of Skill, Attitude and Use of Computer and Internet among Faculty Members	229
Hasan Siamian, Azita Bala Ghafari, Kobra Aligolbandi, Mohammad Vahedi and Gholam Ali Golafshani Jooybari	
Trends in Scholarly Communication among Biomedical Scientists in Greece	232
Assimina Vlachaki and Christine Urquhart	
SESSION ON ELECTRONIC PUBLISHING: A DEVELOPING LANDSCAPE	236
Dr. Dimitris Kouis	
E-Journal and Open Access Journal Publishing in the Humanities: Preliminary Results from a Survey among Byzantine Studies Scholars	238
Victoria Tsoukala and Evi Sachini	
Preliminary Results on a Printed VS Electronic Text Books Assessment Through Questionnaire	242
Dimitrios A. Kouis and Kanella Pouli	
An Interpretation of Aristotelian Logic According to George Boole	246
Markos N. Dendrinis	
SESSION ON INFORMATION CONTENT PRESERVATION AS OUTCOME OF CONSERVATION OF CULTURAL HERITAGE: ETHICS, METHODOLOGY AND TOOLS	251
Prof. George Panagiaris and Dr. Spiros Zervos	
Intrinsic Data Obfuscation as the Result of Book and Paper Conservation Interventions	254
Spiros Zervos, Alexandros Koulouris and Georgios Giannakopoulos	
Mass Deacidification: Preserving More than Written Information	258
Michael Ramin, Evelyn Eisenhauer and Markus Reist	
Information Literacy of Library Users: A Case Study of Mazandaran Public Library Users, Iran	260
Hussein Mahdizadeh and Hasan Siamian	
The Narratives of Paper in The Archives of the New Independent Greek State (Mid 19th c.)	264
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	268
Nikolaos Maniatis and Georgios Panagiaris	

Digital images: A valuable scholar's tool or misleading material?	272
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	277
Mahsoomeh Latifi, Fatemeh Zandianand and Hasan Siamian	
Investigation of the Degradation Mechanisms of Organic Materials: From Accelerated Ageing to Chemometric Studies	280
Ekaterini Malea, Effie Papageorgiou and Georgios Panagiaris	
SESSION ON DIVERGENCE AND CONVERGENCE: INFORMATION WORK IN DIGITAL CULTURAL MEMORY INSTITUTIONS	285
Dr. Susan Myburgh	
Extending Convergence and Divergence in Cultural Memory Institutions: The Old Slave Lodge in the New South Africa	287
Archie L Dick	
The Transfer of Knowledge from Large Organizations to Small: Experiences from a Research Project on Digitization in Wales	289
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	293
Enrico Francese	
Varialog : How to Locate Words in a French Renaissance Virtual Library	297
Marie-Hélène Lay	
The Urge to Merge: A Theoretical Approach	301
Susan Myburgh	
SYMPOSIUM ON ADVANCES INFORMATION FOR STRATEGIC MANAGEMENT	304
Professor Nikolaos Konstantopoulos	
Empowerment in the Tax Office of Greece	306
Antonios E. Giokas and Nikolaos P. Antonakas	
Building Absorptive Capacity Through Internal Corporate Venturing	310
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)	314
Catherina G. Siampou, Eleni G. Fassou and Athanassios P. Panagiotopoulos	
Corruption in Tax Administration: The Entrepreneurs View Point	318
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.	322
Anthi Z. Vaxevanou, Nikolaos Konstantopoulos, Damianos P. Sakas	
Contemporary Forms of Ordering Between the Supply Department and Ship Chandler Companies in the Shipping Industry	325
Anthi Z. Vaxevanou, Nikolaos Konstantopoulos, Damianos P. Sakas	
Strategies Implemented and Sources Used for the Acquisition of Information on Foreign Markets	329
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	333
Nikolaos Konstantopoulos and Yiannis Triantafyllopoulos	
Personnel's Absorptive Capacity as a Guiding Concept for Effective Performance in Informative Technology	337
Nikolaos Konstantopoulos and Yiannis Triantafyllopoulos	
SESSION ON CONTEMPORARY ISSUES IN MANAGEMENT: ORGANISATIONAL BEHAVIOUR, INFORMATION TECHNOLOG, EDUCATION & HOSPITAL LEADERSHIP	341
Dr. Panagiotis Trivellas	
Investigating the Importance of Sustainable Development for Hotel SMES	343
Panagiotis Reklitis and Anestis Fotiadis	
Strategic Alignment of ERP, CRM and E-business: A Value Creation	347
Catherine C. Marinagi and Christos K. Akrivos	
The Impact of Occupational Stress on Performance in Health Care	351
Panagiotis Trivellas Panagiotis Reklitis and Charalambos Platis	

The Impact of Emotional Intelligence on Job Outcomes and Turnover Intention in Health Care	356
Panagiotis Trivellas Vassilis Gerogiannis and Sofia Svarna	
SYMPOSIUM ON BUSINESS MANAGEMENT AND COMMUNICATION STRATEGIES SUPPORTING DECISION MAKING PROCESS IN TOURISM SECTOR	360
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	362
Christos K. Akrivos and Panagiotis Reklitis	
Tourist Destination Marketing and Management Using Advanced ICTS Technologies	365
Anastasia Argyropoulou, Panagiota Dionyssopoulou, Georgios Miaoulis	
G.N.T.O. (Greek National Tourism Organization) Communication Strategy in Advertising Campaigns 1991-2006	370
George Stafylakis and Panagiota Dionyssopoulou	
GENERAL PAPERS	375
The role of Environmental Education within the Framework of the Environmental Policy of a Regional Municipality	376
Vassiliki Delitheou and Dimitra Thanasia	
Issues of Social Cohesion: A case study from the Greek Urban Scenery	380
Evgenia Tousi	
Merging Activity and Employee Performance: The Greek Banking System	384
Panagiotis Liargovas and Spyridon Repousis	
Sustainable Development and Corporate Social Responsibility in Higher Education: Some Evidence from Greece	387
Anastasios Sepetis and Fotios Rizos	
Exploring the Effects of Organizational Culture on Collaborative vs. Competitive Knowledge Sharing Behaviors	395
Hanan Abdulla Mohammed Al Mehairi and Norhayati Zakaria	

Preface: Proceedings of the International Conference on Integrated Information (IC-ININFO 2011)

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

Library Science, Archives Science, Museum and Gallery Studies, Information Science, Documentation, Digital Libraries, Electronic Archives, Information Management, Records / Document Management, Knowledge Management, Data Management, Copyright, Electronic Publications, Cultural Heritage Management, Conservation Management, Management of Nonprofit Organizations, History of Information, History of Collections, Health Information

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
Emmanuel Garoufallou, Technological Educational Institute of Thessaloniki
Filippos Tsimpoglou, 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. Dendrinis, Technological Institute of Athens
Milena Dobрева, 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].

Session on Evidence-Based Information in Clinical Practice

Organizer: Dr Evangelia Lappa, *evlappa(at)med.uoa.gr*

Medical Library

General Hospital Attikis K.A.T., 14564 Kiffisia, Athens-Greece.

Description: Getting research into the practice is not a straightforward exercise and as usual with research, a number of questions are raised: where to find the Evidence in MEDLINE, Cochrane Databases, InfoPoems, Diseases Topics, Database Search ? How do you get evidence into practice? The recent policy statement of the MLA takes the position that scientific evidence is the basis for improving the quality of Information Sciences now and in the future. Developments such as total quality of Health Management and continuous quality improvement reinforce the centrality of research and its relationship to efficient and effective Information Practice. Evidence Based Practice (EBP) encourages health professionals to seek out, appraise, and apply the best evidence from the Medical Literature to improve the quality of clinical judgments. The Health Information Scientist interact effectively seeking the best information to assist in decision making. Is it time to change how seek best evidence to solve patient problems?



DR. EVANGELIA LAPPA

Dr Evangelia Lappa is the Director of the Medical Library/KAT General Hospital Attikis (Athens) and member of the contract academic staff at the Technological Educational Institute (TEI) of Athens since 2000. She holds a PhD from the Faculty of Medicine, University of Athens and a M.Sc. Econ in Health Information Science from Aberystwyth University of Wales/UK.

She has studied Library Science at the Technological Educational Institute of Athens and at Ionian University (Corfu, Greece). In the academic year 2009-2010, she has taught courses in Health Management Science to the Postgraduate program at the Uninettuno University (Rome).

As Health Information Scientist, her research interests are focusing on the areas of Medical Information Management, and Evidence-Based Librarianship in clinical practice for physicians, researchers, health care workers and consumers.

She is council member of EAHIL and of Medical Foundation Center/ Documentation Center Continuing Medical Education of the Medical Society of Athens.

Applicability of Data Mining Algorithms on Clinical Datasets

Wilfred, Bonney[†]

[†] School of Business and Technology, Capella University, Minneapolis, Minnesota, USA

Abstract: *The essential need of database management systems to improve the quality of healthcare delivery makes the use of data mining techniques a phenomenon that cannot be ignored. Today, many healthcare providers are in the business of capturing and storing patients' personalized health information such as demographics, family history, allergies, medications, and diagnosis. This information is generally collected not only to make the healthcare practitioner well-informed about the health status of patients but also to improve the efficiency of care delivery and reduce waiting times. This paper aims to discover the applicability of data mining algorithms on clinical datasets. An experimental study was conducted to compare the performance of four different learning algorithms across four clinical datasets using 10 fold cross-validations.*

Keywords: *Data mining, Database management, Clinical datasets, Algorithm.s*

I. INTRODUCTION

The increasing use of hospital information systems and the emerging health information standards have provided a rich platform for the collection and storage of health information. This explosive accumulation of data in hospitals is proving great challenge to healthcare providers and researchers wanting to analyze the data manually (Tan, Yu, Heng, and Lee, 2003). There is, therefore, an increasing demand to apply machine learning algorithms on clinical datasets to help solve diagnostic and prognostic problems (Andreeva, Dimitrova, and Radeva, 2004).

Clinical databases serve as the foundation for an integrated information infrastructure on which data mining algorithms could be applied (Hogl, Mueller, Stoyan, and Stuelingher, 2001; Lush, Henry, Foote, and Jones, 1997). Data mining techniques have the potential to discover hidden knowledge in clinical databases (Han, Kamber, and Pei, 2011). The discovered knowledge could be stored in a knowledge base and used in computational inferences not only to support medical diagnosis but also improve the quality of decision making, and increase the working efficiency of medical practitioners (Abe et al., 2007; Tan et al., 2003). The main objective of this paper is to reveal the applicability of data mining algorithms on clinical datasets.

II. RELATED STUDIES

Many researchers have attempted to study how data mining algorithms could be applied effectively to clinical datasets. For most of these researchers, the focus is on developing unique algorithms to support

clinical datasets. For example, Abe et al. (2007) introduced an integrated data mining approach that classifies the clinical data into multiple categories to discover relationships using health levels. The health levels were used as significant clinical parameters to express the health status of patients. Abe et al. (2007) went further and proposed a framework called Cyber Integrated Medical Infrastructure (CIMI), which is made up of integrated management of clinical data on computer networks. The CIMI framework supports the integration and storage of various types of clinical data in a database prior to analysis by machine learning techniques (Abe et al., 2007). They recommended the need to classify clinical datasets not according to statistical or associated patterns but rather according to their influence on health levels. Abe et al. (2007) concluded that the integration of multiple results from partial data mining will enable us to discover complex relationships between the clinical data and health levels.

In a related study, Tan et al. (2003, p. 129) proposed the use of a two-phase hybrid evolutionary classification technique to “extract classification rules that can be used in clinical practice for better understanding and prevention of unwanted medical events”. Tan et al. (2003) were not convinced that the current existing data mining algorithms are sufficient to reveal hidden knowledge in clinical datasets. The evolutionary classifier (EvoC) algorithm, introduced by Tan et al. (2003), was validated with the hepatitis and breast cancer datasets obtained from the UCI machine-learning repository, and the simulation results showed that the algorithm produces comprehensible rules and good classification accuracy for the medical datasets.

Last but not the least, Andreeva et al. (2004) investigated different learning models and provided a practical guideline on how to select the most suited algorithm for a specific clinical dataset. Andreeva et al. (2004, p. 152) formulated that “for most medical applications the logical rules are not precise but vague and the uncertainty is present both in premise and in the decision”. Hence, a good methodology for medical datasets is the rule-based representation from decision-tree method, which is easily understood by the general public (Andreeva et al., 2004). Andreeva et al. (2004) concluded that the integration of fuzzy set and data mining methods gives a much better and more exact representation of relationship between symptoms and diagnosis.

The above three studies suggest the possibility of applying data mining algorithms on clinical datasets. Although selecting appropriate data mining algorithm for a given dataset is not trivial, selective application of data mining algorithms on selective clinical datasets could provide guidelines to researchers in determining

the best algorithm to use in discovering hidden knowledge from clinical dataset.

III. EXPERIMENTAL STUDY

In this experiment, the WEKA Experimenter (Scuse and Reutemann, 2007; Witten and Frank, 2005) was used to compare the performance of four different learning algorithms across four clinical datasets using 10 fold cross-validation. The Weka Experimenter module enables the researcher to create, modify, and run several schemes against a series of datasets and then analyse the results of the experiment to determine if one of the schemes is statistically better than the other schemes (Scuse and Reutemann, 2007). The four datasets used in the experiment are the Breast Cancer, Heart Disease, Hepatitis and Pima Diabetes datasets. The datasets were obtained from the UCI Machine-Learning repository (Frank and Asuncion, 2010).

A. Data Preprocessing

Data preprocessing is important because quality decisions must be based on quality data (Han et al., 2011). In order to obtain target datasets for the experiment, data preprocessing techniques were applied to two of the datasets by (a) removing instances with missing attribute values and (b) removing attributes with higher missing values. Nine instances were removed from the breast cancer dataset, reducing it from 286 to 277 instances. In the case of the hepatitis dataset, two attributes (ALK PHOSPHATE and PROTIME) with highest missing values were removed from the dataset. Removing instances from the hepatitis dataset was not practical as it has only 155 instances. There were no missing attribute values in the other two datasets.

B. Applied Algorithms

Four data mining algorithms were applied on the target datasets. The essence for the use of the four algorithms could be summarized as follows:

1. BayesNet: BayesNet is a Bayesian Network algorithm that uses various search algorithms and quality measures. It is a probabilistic graphical model that represents a set of variables and their probabilistic independencies (Bouckaert, 2007; Witten and Frank, 2005).
2. LogitBoost: This algorithm class is used for performing additive logistic regression. It performs classification using a regression scheme as the base learner, and can handle multi-class problems (Witten and Frank, 2005).
3. J48: The J48 is a decision tree learning algorithm, which is an extension of the C4.5 algorithm proposed by Ross Quinlan (1993). The algorithm generates a classification-decision tree for the given dataset by recursive partitioning of data. The decision is grown using depth-first strategy. The J48 algorithm “considers all the possible tests that can split the dataset and selects a test that gives the best information gain” (Ali, 2002, p. 1503).

4. ZeroR: This is a rule-based classifier for building and using a Zero-R classifier. It predicts the mean (for a numeric class) or the mode (for a nominal class) (Witten and Frank, 2005)

V. FINDINGS AND ANALYSIS

The experiment consisted of 10 runs, for 4 algorithms, for 4 datasets, for a total of 1600 result lines. The test is performed on the Percent correct field with BayesNet as the baseline algorithm at the significance level of 0.05. The output from the WEKA Experimenter (shown in Table 1) depicts the percentage correct for each of the four learning algorithms in each dataset row. The annotation “v” or “*” indicates that a specific result is statistically better (v) or worse (*) than the baseline algorithm (Scuse and Reutemann, 2007).

	BayesNet	LogitBoost	J48	ZeroR
breast	77.10	71.71	74.87	70.75
heart	82.56	82.22	78.15	55.56*
hepatitis	83.87	81.18	79.23	79.38
diabetes	75.25	74.54	74.49	65.11*
	(v / /*)	(0/4/0)	(0/4/0)	(0/2/2)

Table 1. Performance of algorithms.

Table 1 indicates that there is no statistical difference between the results for BayesNet, LogitBoost and J48 on the datasets. Both LogitBoost and J48 were four times (0/4/0) equivalent to the BayesNet on the datasets. However, there exists a statistically significant difference between BayesNet and ZeroR. ZeroR (0/2/2) was twice the same as BayesNet and twice worse than BayesNet on the datasets. A graph of the mean absolute error is shown in Figure 1. The graph in Figure 1 indicates that the BayesNet algorithm performed well on the datasets with the lowest values of mean absolute error. On the other hand, the mean absolute error values for the ZeroR algorithm are very high, indicating that it is not the best algorithm to be used on the datasets.

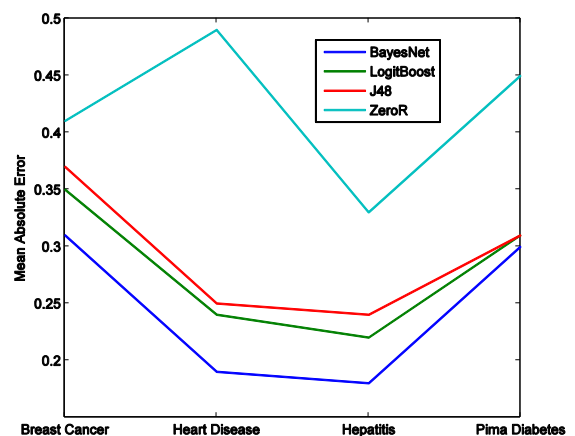


Figure 1. Graph of Mean Absolute Error.

The outcome of this experiment substantiates the findings by Andreava et al (2004). This is evident by the fact that the rule-based algorithm (i.e. ZeroR), used in

this experiment, did not perform well on the clinical datasets. Andreeva et al. (2004, p. 152) concluded in their experiment that “induced decision trees are useful for the analysis of the importance of clinical parameters and their combinations for the prediction of the diagnosis”. This conclusion is also manifested in the experiment by the fact that the decision tree algorithm (i.e. J48) performed well on all the four clinical datasets.

VI. CONCLUSIONS

This paper has demonstrated the feasibility of applying data mining algorithms on clinical datasets. Even though the outcome of this analysis is limited to the quality and quantity of the datasets used, it is evident that proper use of data mining algorithms will help uncover hidden knowledge in clinical datasets. The discovered knowledge from clinical datasets has enormous potential to support evidence-based practice and improve the quality of healthcare delivery.

REFERENCES

- [Abe, A. Hagita, N. Furutani, M. Furutani, Y. and Matsuoka, R. \(2007\). Possibility of integrated data mining of clinical data. *Data Science Journal*. 6\(1\), pp. 104-115.](#)
- Ali, S. (2002). Support vector machine and neural network with others for two dimensional classification problem: An empirical study. *Proceedings of the Sixth International Conference on Knowledge-Based Intelligent Information & Engineering Systems*. Crema, Italy. pp. 1502-1506.
- Andreeva, P. Dimitrova, M. and Radeva, P. (2004). Data mining learning models and algorithms: Application in medical domains. In R. Romansky (Ed.) *Proceedings of 18th Conference on Systems for Automation of Engineering and Research (SAER)*. Varna, Bulgaria. 24-26 Sept 2004. pp. 148-152.
- [Bouckaert, R.R. \(2007\). *Bayesian Network Classifiers in Weka for Version 3-5-6*. \[Online\] Available at: <http://docs.huihoo.com/weka/BayesianNetClassifier-s-3.5.6.pdf> \[Accessed 30 June 2011\]](#)
- [Frank, A. and Asuncion, A. \(2010\). *UCI Machine Learning Repository*. \[Online\] Available at: <http://archive.ics.uci.edu/ml> \[Accessed 30 June 2011\].](#)
- [Han, J. Kamber, M. and Pei, J. \(2011\). *Data Mining: Concepts and Techniques*. 3rd ed. San Francisco: Morgan Kaufmann.](#)
- Hogl, O. Mueller, M. Stoyan, H. and Stuelinger, W. (2001). On supporting medical quality with intelligent data mining. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. January 3-6, 2001, Computer Society Press.
- Lush, M.T. Henry, S.B. Foote, K. and Jones, D.L. (1997). Developing a generic health status for use in a computer-based outcomes infrastructure. *In Proceedings of the 6th International Conference on Nursing Informatics*. Stockholm, Sweden, Sept 26-Oct 1, 1997. pp. 1-6.
- [Quinlan, J.R. \(1993\). *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann.](#)
- [Scuse, D. and Reutemann, P. \(2007\). *WEKA Experimenter Tutorial for Version 3-5-6* \[Online\] Available at: <http://surfnet.dl.sourceforge.net/project/weka/documentation/3.5.x/ExperimenterTutorial-3.5.6.pdf> \[Accessed 30 June 2011\]](#)
- [Tan, K.C. Yu, Q. Heng, C.M. and Lee, T.H. \(2003\). Evolutionary computing for knowledge discovery in medical diagnosis. *Artificial Intelligence in Medicine*. 27\(2\), pp. 129-154.](#)
- [Witten, I.H. and Frank, E. \(2005\). *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed. San Francisco: Morgan Kaufmann.](#)