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 (ICINFO 2011)
Georgios A. Giannakopoulos, Damianos P. Sakas

<table>
<thead>
<tr>
<th>Conference Details</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keynote Speaker</td>
<td>5</td>
</tr>
</tbody>
</table>

### SYMPOSIUM ON INFORMATION AND KNOWLEDGE MANAGEMENT
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 Kouirioukidis 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 Chadjipantelis

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. Koukounidou 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>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Library Ethics to Present a Model for Final Library Ethical Codes</td>
<td>188</td>
</tr>
<tr>
<td>Mahsoomeh Latifi, Fatemeh Zandian and Hasan Siamian</td>
<td></td>
</tr>
<tr>
<td>SESSION ON OPEN ACCESS REPOSITORIES: SELF-ARCHIVING, METADATA,</td>
<td>189</td>
</tr>
<tr>
<td>CONTENT POLICIES, USAGE</td>
<td></td>
</tr>
<tr>
<td>Dr. Alexandros Koulouris</td>
<td></td>
</tr>
<tr>
<td>Geographical Collections in Greek Academic Libraries: Current Situation and Perspectives</td>
<td>194</td>
</tr>
<tr>
<td>Ifigenia Vardakosta and Sarantos Kapidakis</td>
<td></td>
</tr>
<tr>
<td>Information Seeking Behavior: Factors that Affect the Behavior of Greek Astronomers</td>
<td>198</td>
</tr>
<tr>
<td>Hara Brindesi and Sarantos Kapidakis</td>
<td></td>
</tr>
<tr>
<td>Aggregating Metadata for Europeana: The Greek Paradigm</td>
<td>202</td>
</tr>
<tr>
<td>Alexandros Koulouris, Vangelis Banos and Emmanouel Garoufallou</td>
<td></td>
</tr>
<tr>
<td>Integrating a Repository with Research Output and Publications: The Case of the National Technical University of Athens</td>
<td>205</td>
</tr>
<tr>
<td>Dionysis Kokkinos</td>
<td></td>
</tr>
<tr>
<td>Implementation of Workflows as Finite State Machines in a National Doctoral Dissertations Archive</td>
<td>209</td>
</tr>
<tr>
<td>Nikos Houssos, Dimitris Zavaliadis, Kostas Stamatis and Panagiotis Stathopoulos</td>
<td></td>
</tr>
<tr>
<td>Practices of “Local” Repositories of Legally Protected Immoveable Monuments. A Global Scheme for ‘Designation – Significance’ Information</td>
<td>213</td>
</tr>
<tr>
<td>Michail Agathos and Sarantos Kapidakis</td>
<td></td>
</tr>
<tr>
<td>Integration of Metadata in BWMETA-2.0.0 Format</td>
<td></td>
</tr>
<tr>
<td>Katarzyna Zamlynska, Jakub Jurkiewicz and Lukasz Bolikowski</td>
<td></td>
</tr>
<tr>
<td>SESSION ON EVIDENCE-BASED INFORMATION IN CLINICAL PRACTICE</td>
<td>216</td>
</tr>
<tr>
<td>Dr. Evangelia Lappa</td>
<td></td>
</tr>
<tr>
<td>Applicability of Data Mining Algorithms on Clinical Datasets</td>
<td>218</td>
</tr>
<tr>
<td>Wilfred, Bonney</td>
<td></td>
</tr>
<tr>
<td>Changing Roles of Health Librarians with Open Access Repositories</td>
<td>221</td>
</tr>
<tr>
<td>Christine Urquhar and Assimina Vlachaki</td>
<td></td>
</tr>
<tr>
<td>From Medical Records to Health Knowledge Management Systems: The Coding to Health Sector</td>
<td>225</td>
</tr>
<tr>
<td>Evangelia C. Lappa and Georgios A. Giannakopoulos</td>
<td></td>
</tr>
</tbody>
</table>
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 Urge 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 TECHNOLOGY, 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 Dionyssopoulou, Georgios Miaoulis

George Stafylakis and Panagiota Dionyssopoulou

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, Ekle, 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 Skoutras, 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 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. Dendrinos, Technological Institute of Athens
Milena Dobreva, University of Strathclyde
Prodromos Tsiavos, London School of Economics and Political Science
Sándor Darányi, University of Borás
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].
A Semi-Automatic Emerging Technology Trend Classifier Using SCOPUS and PATSTAT

Seonho Kim, Woondong Yeo, Byong-Youl Coh, Waqas Rasheed†, Jaewoo Kang†

Technology Information Analysis Lab. Department of Information Analysis, Korea Institute of Science and Technology Information, 66, Hoeugi-ro, Dongdaemun-gu, Seoul 130-741, Korea
{haebang, wdyeo, cohby}@kisti.re.kr
† Data Mining and Information System Lab. College of Information and Communication, Korea University, 5-ga, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea
{waqas, kang}@at.korea.ac.kr

Abstract: Identifying Emerging Technology Trends is crucial for decision makers of nations and organizations in order to use limited resources, such as time, money, etc., efficiently. Many researchers have proposed emerging trend detection systems based on a popularity analysis of the document, but this still needs to be improved.

In this paper, an emerging trend detection classifier is proposed which uses both academic and industrial data, SCOPUS [1] and PATSTAT [2]. Unlike most previous research, our emerging technology trend classifier utilizes supervised, semi-automatic, machine learning techniques to improve the precision of the results. In addition, the citation information from among the SCOPUS data is analyzed to identify the early signals of emerging technology trends.

Keywords: SCOPUS, PATSTAT, Emerging trend detection, Machine learning, Artificial neural network

I. INTRODUCTION

Research on emerging trend detection has been conducted by many researchers and organizations, because identifying emerging trends and predicting the near future can improve efficient resource distribution and effective policy establishment. Traditional research on detecting emerging trends is based on massive data analysis, literature review, and brainstorming of intellectuals [3]. However, as the amount of information to analyze increases and computing technology develops, recent trends of this research are focused on automating the processing by utilizing text mining and data analysis techniques [4, 5]. In this study, we propose a semi-automatic machine learning technology for recognizing emerging technology trends from academic and industrial text data.

II. METHODS

A. Previous Research

Traditional research on emerging trend detection in text data is conducted by two major methods, fully automatic detection and semi-automatic detection. The fully automatic methods, such as TimeMines [6], VUDM [7], and on-line event detection by Allan [8], have employed unsupervised learning methods, in which learning is achieved in a fully automatic manner by detecting topics describing technologies to analyze their trends. These approaches tend to show good performance in recall, which represents the portion of emerging trends found over the total number of emerging trends in the system, but only medium performance in precision, which expresses the rate of correctly determined emerging trends over the total number of trends found. The reason for the good recall of the method is that it uses bottom-up methods to test every possible combination of noun phrases as topics describing emerging trends. After finding topics, the trends of each topic are analyzed, and emerging trends are selected from among the topics based on certain criteria. However, the results of this approach can be noisy for some decision makers because, in most cases, the identification of several correct targets is more meaningful than identifying as many correct targets as possible without missing.

For this reason, in order to increase precision rather than the recall, semi-automatic methods which use supervised learning are attracting attention. These approaches, such as CIMEL [9], PatentMiner [10], and HDDI [11], use experts to provide guidance to the machine along with the training data. For example, the training data of the machine learning is tagged by the experts to identify whether or not the data is about an emerging trend, or the experts interact with the software during learning. Therefore, the machine can learn more correct patterns of emerging trends and classify new trends more precisely.

B. Semi-automatic Supervised Machine Learning

Our proposed method is a semi-automatic system in which experts provide guidance to an artificial neural network machine learning system. The guidance from the experts in our system is the classification tags attached to the training data, so that the machine is directed as to whether or not the information currently being learned is about emerging technology. In a machine learning system, selecting the correct features for learning is most critical. HDDI [11] used the frequencies of concepts within a fixed length of periods and the count of concepts in semantically defined regions, which can be obtained automatically. Oh [12] also provided the results of a comparison experiment for testing the strength of impact of the most famous features of trends generally used for emerging trend
detection, such as change, persistency, stability, and volume. According to Oh's result, the strongest impact feature of a trend is expressed by the change, stability, volume, and persistency in descending order. However, because every emerging trend detection systems have different purposes, available source data, and training methods, the features to be learned need to be selected and tested themselves. Our system uses both academic data, SCOPUS [1], and industrial data, PATSTAT [2]. Thus, in addition to the usual term-frequency based trend changes in academic data, the latency between the weak academic signal and the industrial legal activity (patent application) could be trained. Also, our system uses a citation half-life feature, which is described in the next section, to train the aging factor of technology. We assume that the age of the emerging technology is relatively young and young technology tends to have a later citation half-life than mature technology.

This study consists of three stages, 1) emerging trend detection, 2) training data generation, and 3) learning and experimentation. Figure 1 shows the brief structure of the proposed system.

![Figure 1. Emerging technology trend classifier](image)

At stage 1, a set of emerging technology trends are identified from SCOPUS. The process contains clustering, filtering, and naming. The field experts intervene during the process. After that, in the second stage, the training data is generated by extracting learning features from SCOPUS and PATSTAT. Experts also join in this process to add their guidance to the learning data.

C. Emerging Technology Trends

Detecting emerging trend from SCOPUS is conducted to generate tagged feature vectors of training data for neural network learning. Only 1% of the highly cited papers are used. The papers are clustered using the scientometrics techniques, which use the citation information, classification code, core paper identification, and shared term frequency. As a result of clustering, 512 clusters are identified. Then, about 50 field experts are selected for the naming of the clusters and voting on the prospects of the technologies as expressed by their names. Clusters that do not have any consistent topic, based on the knowledge of the field experts, are filtered out, and final 87 clusters are selected as emerging technology trends. From this result, 60 technologies are selected for positive training data. In addition, 26 matured, non-emerging technologies, such as compact discs, fountain pens, optical character recognition, and Bluetooth technology, are prepared as negative training data.

D. Feature Extraction and Tagging

An artificial neural network is trained with positive training data, which consist of features from emerging technology trends, and negative training data, which consist of features from matured, non-emerging, technology trends.

The training features are extracted from SCOPUS and PATSTAT data from 1996 to 2010. Two types of feature, linguistic and statistical, are used. For the linguistic features, the trend "change" for 2 different periods from SCOPUS and PATSTAT are used. For statistical features, the citation "half-life" in SCOPUS is used. The change value represents the slope of the change between the start and end point of the trend curve. A positive change value means that the trend is emerging, while a negative change value means that the trend is submerging [12].

\[ \text{diff}(f) = LR_f(T_{end}) - LR_f(T_{start}) \]  
\[ \text{Change}(f) = \frac{\text{diff}(f)}{\sqrt{\text{diff}(f)^2 + (T_{end} - T_{start})^2}} \]

The change feature is obtained by calculating the linear regression, which is represented by equations (1) and (2). In these equations, \( f \) is a trend function expressed by an independent value, time, and a dependent value, the populations in SCOPUS and PATSTAT. \( LR_f \) is the linear regression function about \( f \). \( T_{end} \) expresses the population at the end point of the period of interest, and \( T_{start} \) is the population at the start point of the period. For the machine learning part of this study, four change features, 1) the change in the first 11 years in SCOPUS, 2) the change in the last 4 years in SCOPUS, 3) the change in the first 11 years in PATSTAT, and 4) the change in the last 4 years in PATSTAT, are extracted from one technology trend curve.
The reason for dividing the period into two sections, 11 years and 4 years, is based on our basic interpretation about the data. Figure 2 and figure 3 show some examples of trend curves of emerging technologies in SCOPUS and PATSTAT, respectively, and both charts show that the trend curves changed noticeably in the last 4–5 years.

According to the figure 2, the population of emerging technology trends had increased steadily in the first 11 years and its changes escalated during the last 4 years in SCOPUS. In figure 3 the speed of change dropped rapidly during the last 4 years in PATSTAT.

The other type of feature selected for this study is the citation half-life, which is the number of years that have passed after half of the total number of citations concerning a particular technology. The citation half-life represents the center of the total group of citations. Figure 4 illustrates the concept of citation half-lifetime.

The number of citation on "nano drug delivery materials" increased steadily until its peak in 2008, and then it decreased rapidly until the present year, 2010. The total number of citations on the technology during the observation period in SCOPUS is 24,225 and the half of the total number of citations, 12,112.5, is met around 2005. Therefore, the citation half-life of the "nano drug delivery material" technology is 4.93 years.

Every technology has a distinct trend in citation. Figure 5 shows the citation trends of some example emerging technologies. The trend curves tend to bias to the left or right based on the age of the technology. Thus, we assumed that an emerging technology is a relatively new technology, and the citation half-life value represents the age of technology, and that learning this value may help detect emerging technologies.

E. Learning

The goal of machine learning in this study is to educate the neural network with the features of emerging technologies and non-emerging technologies, and evaluate the performance of the trained neural network as an emerging technology trend classifier.

During learning, a total 86 feature vectors, 60 from positive data and 26 from negative data, and tagged as to whether or not it is emerging technology, are entered into the neural network for training. In the evaluation session, the precision of the classification performance is measured. In order to evaluate the precision of the learned neural network, 64 new feature vectors, which have not been used for training, are entered. The evaluation data is not tagged but are already classified by field experts and compared with the decisions made by the neural network.

F. Evaluation and Results

The final status of a neural network, the learned status, is different in every trial, even if it is trained with identical training data, because the neural network assigns random initial weights to the links among the neurons in the hidden layers. Therefore, we repeated the training and evaluations 10 thousand times, with the same evaluation data, and determined the average precision and a standard deviation.

The average precision of classification of the neural
network is 69.92 and the standard deviation is 21.93. Regarding precision only, this result is better than the normal precision rate, which is about 45, for normal unsupervised learning for emerging trend technology.

The precision of learning can be improved by iterative tuning of the training data by filtering out noisy data and adding correct data, and the deviation can be decreased by enlarging the size of training data. However, the tuning process is not included in this study.

III. CONCLUSIONS AND FUTURE WORK

An emerging technology trend identification system is proposed, which is based on semi-automatic supervised machine learning technology, and its performance is evaluated. Our method learns from both academic data, SCOPUS, and industrial data, PATSTAT. The change in trend and citation half-life is selected as the learning feature. By using a supervised learning method, our system places the focus on enhancing the precision of classification rather than recall. This system can be applied for fast classification of emerging technology, such as a real time online technology evaluator or competitive intelligence system.

For future work, we will tune the training data to make the learning process more efficient. Also, a larger volume of training data and evaluation data may improve the precision of the classifier. Furthermore, we plan to apply the standard test set for emerging trend detection, such as the Topic Detection and Tracking (TDT) corpora [13], to make our system applicable as a general purpose topic detection system. Finally, we will test for various trend features, such as changes and statistics about search queries.

ACKNOWLEDGEMENTS

Thanks go to Korea Ministry of Education, Science And Technology for support of grants K-11-L05-C03-S02 for this research.

REFERENCES


Swan, R. and D. Jensen. Timelines: Constructing timelines with statistical models of word usage, in the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000. Boston, MA, USA: Association for Computing Machinery.

http://www.cs.cmu.edu/~dunja/KDDpapers/Swan_TM.pdf


http://dimacs.rutgers.edu/~billp/pubs/ETDMетодологияs.pdf


http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.9475


http://ir.kaist.ac.kr/anthology/2009.03-%EC%98%A4%ED%9D%8C%EC%84%A0.pdf