

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

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].

Exploiting the Search Culture Modulated by the Documentation Retrieval Applications

Nikitas N. Karanikolas[†] and Christos Skourlas[†]

[†] *Technological Educational Institute of Athens. Department of Informatics, Aigaleo 12210, Athens, Greece.*
{nnk, cskourlas} (at)teiath.gr

Abstract: *Users Interaction with various well-established tools of documentation retrieval and the influence from each tool to the other ones have contributed in the creation of some search culture which is shared between the design assumptions (existing behind) of search interfaces and the user skills. The purpose of this work is to exploit the hidden design assumptions of the user interfaces in order to define search scenarios that combine concrete features for specific retrieval. One such interesting application case is the definition of scenarios for finding the citations that a researcher has got, excluding self citations*

Keywords: *Abstracting and indexing; Citation; Documentation retrieval; Search engines; Search culture.*

I. INTRODUCTION

It is generally acceptable that query languages should be simple and easy to learn for the end-user, and they should provide facilities for the experienced user / information scientist (Gebhardt and Stellmacher 1978). Search behaviour is an area that is often oversimplified, and it is important to highlight different kinds of search activity (Ruthven, 2008). In recent years, the focus has been on search on the Web. Search culture took hold in the late 1990s with Yahoo and Google and a key development occurred with the movement of services onto the Web (Brown and Dumouchel, 2007).

An effective query language is simple enough for novices/end-users but also offers possibilities of complex query languages, to cover specific user needs. An alternative solution is to use a combination of simple form-based searches of a tool, and also templates of ready to use sophisticated and precise queries for specific applications.

Previous studies (Karanikolas, 2011) have focused on the modulated search culture shared between the various applications of documentation retrieval, their interfaces and their users. In this paper, we summarize a portion of web and GUI based search culture to exploit it for evaluating the research work of researchers, on the base of citation counts (NUS Libraries, 2011). In order to achieve our goal, we will concentrate to two concrete documentation retrieval systems/sources (Google Scholar and Scopus) and to a number of customized query templates. It is worth mentioning that there are also other suggestions for evaluating academic

reputation of authors (more refined than the simple citation count) (Hou, Li and Niu, 2011; Burns, 2011).

The motivation for using concrete custom query templates, instead of the standard research evaluation methods provided by the specific applications and tools of documentation retrieval systems/sources, is to improve the unsatisfactory results returned by the standard methods. These custom query templates are based on both the complete bibliographical records and also to the full text of articles. Various methods for the evaluation of research activities, provided by documentation retrieval systems/sources, usually underestimate the individual academic's actual impact (underestimate the actual citations count) (Harzing, 2008). There are a lot of reasons for such an underestimation (Chapman, 1989).

The paper is organized as follows: In section 2 we discuss the design assumptions and the related search culture we came across while in section 3 we describe and discuss some experiments. Finally, in section 4 we draw conclusions and point to future work.

II. SEARCH CULTURE AND METHODS

From the diversity of documentation retrieval applications' interfaces and their behavior, previous studies (Karanikolas, 2011) have concluded a set of (hidden) assumptions, which constitute a search culture modulated through time. The set of assumptions includes:

Possibility for selection of search fields (from the pool of the available ones) by the user for the construction of simple (basic) criteria. An alternative possibility is the apposition of a fixed set of search fields, that can be used into the formation of a query (if the user enters a required value) or remain inactive.

Possibility for selection of Boolean operators by the user, for building composite queries from simple criteria. An alternative possibility is the suppression of Boolean operators, with the hidden assumption that logical conjunction (AND) is always used.

Possibility of using (the implied) left to right nesting of simple criteria. There is not any alternative solution based on form-type interfaces. Only the search command languages permit the definition of nested criteria, by using parentheses.

Semantics of equality is dependent from context. The meaning of equality can vary from "exact match" to "contains" and this is implicitly determined by the used search field in the criterion. Command language based interfaces do not (usually) provide handlers (expressions) for altering the context dependent semantics of equations. Form based interfaces can

provide handlers for defining (explicitly by user) the semantics of equality. The user defined semantics of equality can be global or specifically defined for any simple criterion.

Possibility to use meta-fields in order to search simultaneously, in a set of fields, for the value of a given criterion. Alternatively, for the same purpose, the system can automatically perform query transcription for creating expanded queries that search the specific (search) value in a set of fields.

Whenever the data collection contains full-text documents, the search mechanism can provide some kind of word normalization (stemming, lemmatization, etc). Documentation retrieval applications' interfaces can provide handlers for activation of this feature by the user, or ignore it.

Use of Relevance Ranking. It is mainly applied to free text collections. The documents are ranked (using a similarity function) according to their similarity to the given question and are presented (usually) in descending order of similarity.

III. EXPERIMENTS & PRELIMINARY DISCUSSION

An interesting and useful exploitation of the above mentioned search culture can be applied for evaluating the research work of academicians and researchers. Some interesting types of queries that can be used for evaluating research work are the following:

1. Citations of an author,
2. Citations of an author, without self-citations,
3. Citations of an author, without self-citations from the author or specific collaborator,
4. Citations that have received two (or more) researchers in their common publications,
5. Citations that have received two (or more) researchers in their common publications, without self-citations,
6. Citations of a specific article title, without self-citations from the first author,
7. Citations of a specific article title, without self-citations from any of the contributing authors.

In the following we give templates and examples of queries that can be applied to the selected systems/sources (Google Scholar and Scopus). Query types 1 and 6 are not examined. Using the term Scopus we assume that "Scopus Advanced search" is used and whenever we mention Google Scholar we actually use "Google Scholar search box" (the simple – default – search of Google Scholar). In the following, numbers at end of place holder names (as is number 1 in the following "author name 1" and number 2 in the following "author name 2") are used for permitting the user to provide different values for the same entity (e.g. name variations of the same researcher).

Query type 2 – Google Scholar – template:

```
"author name 1" -author:"author name 2"
```

This template searches for the phrase "author name 1" inside any field in the bibliographic record and in the full text of articles. It excludes from the results all the articles that contain the phrase "author name 2" in the authors' field of the bibliographic record. In Google Scholar this is the only way for finding citations of an author, without self-citations, because the alternative interface (Advanced Scholar search) provides only the possibility for defining that the phrase exist either "anywhere in the article" or "in the title of the article".

Query type 2 – Google Scholar – examples:

```
"n. karanikolas" -author:karanikolas
```

```
"n.n. karanikolas" -author:karanikolas
```

```
"nikitas karanikolas" -author:karanikolas
```

Since the author names are not always written in the same way, the query type 2 should be executed in Google Scholar with all variations used for filling the template's place holder "author name 1". However, using only the authors surname in the template's "author name 1" will retrieve too many irrelevant articles. On the contrary, the value used for filling the place holder "author name 2" (the value that accompanies the exclusion operator "-author:") should be as simple as possible (only the surname of author). These suggestions have been used in the previous examples.

Query type 2 – Scopus – template:

```
( REF("author name 1")  
  AND NOT  
  AUTHOR-NAME("author name 2")  
)
```

Query type 2 – Scopus – example:

```
( REF(Karanikolas)  
  AND NOT  
  AUTHOR-NAME(Karanikolas)  
)
```

On the contrary to Google Scholar, queries of type 2 can be executed in Scopus by providing only the author's surname as the required value for the REF field. This is a consequence of the structured nature of Scopus and, more precisely, it occurs because the provided surname should be contained in a specific structured field (REF). However, it is a matter of experimentation whether the query will work properly by providing only the author's surname for the REF field or something more precise (e.g. surname with initials) will be needed.

Query type 3 – Google Scholar – template:

```
"author name 1" -author:"author name 2"  
-author:"collaborator name"
```

Query type 3 – Google Scholar – example:

```
"n. karanikolas" -author:karanikolas  
-author:skourlas
```

Query type 3 – Scopus – template:

```
( REF("author name 1")
  AND NOT
  ( AUTHOR-NAME("author name 2")
    OR
    AUTHOR-NAME("collaborator name")
  )
)
```

Query type 3 – Scopus – example:

```
( REF(Karanikolas)
  AND NOT
  ( AUTHOR-NAME(Karanikolas)
    OR
    AUTHOR-NAME(Skourlas)
  )
)
```

Unfortunately, query type 4 is not supported by Google Scholar, since Google Scholar does not provide fields for structured queries.

Query type 4 – Scopus – template:

```
( REF("author name")
  AND
  REF("other author name")
)
```

Query type 4 – Scopus – example:

```
( REF(Karanikolas)
  AND
  REF(Skourlas)
)
```

Query type 5 – Google Scholar – template:

```
"author name 1" +"other author name 1"
-author:"author name 2"
-author:"other author name 2"
```

Query type 5 – Google Scholar – examples:

```
"n. karanikolas" +"c. skourlas"
-author:karanikolas -author:skourlas

karanikolas +skourlas
-author:karanikolas -author:skourlas
```

The previous two examples have the only difference that the first one uses surnames with initials in the “author name 1” and in the “other author name 1”, while the second one uses only surnames. During the time of writing the present article, these queries returned 3 and 13 results, respectively. Consequently, we should be very careful when using name variations. Using only surnames in place of the “author name 1” and the “other author name 1” will result too many irrelevant articles. But using only one variation of surname with initials in “author name 1” and in the “other author name 1” will drive in loss of relevant articles (loss of citations). The best result will be achieved by repeating the query with different variations of surnames with initials. It is almost the same technique used in the query type 2 examples in Google Scholar. This applies also for query type 3.

Query type 5 – Scopus – template:

```
( ( REF("author name 1")
  AND
  REF("other author name 1")
)
  AND NOT
  ( AUTHOR-NAME("author name 2")
    OR
    AUTHOR-NAME("other author name 2")
  )
)
```

Query type 5 – Scopus – example:

```
( ( REF(Karanikolas)
  AND
  REF(Skourlas)
)
  AND NOT
  ( AUTHOR-NAME(Karanikolas)
    OR
    AUTHOR-NAME(Skourlas)
  )
)
```

Query type 7 – Google Scholar – template:

```
"article title" -author:"author1 name"
-author: "author2 name"
-author: "author3 name" ...
```

Query type 7 – Google Scholar – examples:

```
"Computer Assisted Information Resources
Navigation" -author:karanikolas
-author:skourlas
```

```
"CUDL language semantics: updating FDB data"
-author:karanikolas -author:skourlas
-author:nitsiou -author:yannakoudakis
```

Query type 7 – Scopus – template:

```
( REF("article title")
  AND NOT
  ( AUTHOR-NAME("author1 name")
    OR
    AUTHOR-NAME("author2 name")
    OR
    AUTHOR-NAME("author3 name")
  )
  ...
)
```

Query type 7 – Scopus – example:

```
( REF("Computer Assisted Information
Resources Navigation")
  AND NOT
  ( AUTHOR-NAME(Skourlas)
    OR
    AUTHOR-NAME(Karanikolas)
  )
)
```

In both cases of using these systems/sources, Google Scholar or Scopus, the last type (7) of queries decreases (and maybe nullifies), in comparison to the queries of type 5, the percentage of irrelevant returned articles. For example, submitting the following query (of type 5) to

Scopus has the consequence of returning two articles that do not contain reference to any common (co-authored) article of the mentioned authors (Karanikolas and Vassilakopoulos):

```
( ( REF(Vassilakopoulos) AND
  REF(Karanikolas)
  )
  AND NOT
  ( AUTHOR-NAME(Vassilakopoulos) OR
    AUTHOR-NAME(Karanikolas)
  )
)
```

Obviously, the same problem (of finding false citations) can also happen for queries of type 7. However, in our opinion it extremely rarely occurs in this case. The reference list of each returned paper should be examined in order to verify that the returned document contains valid citations for the provided authors (in case of queries of type 5) or for the provided article title (in case of queries of type 7). The superiority of type 7 queries, that decrease (or even nullify) the false citations, coexists with the possibility of reduced recall of citations, in some cases. The main reason for the reduced recall of citations is usually caused by slight modifications of articles' titles (e.g. substitutions of punctuation marks). Such modifications often occur when authors write their article's reference list. To overcome this drawback of type 7 queries, we should try, instead of the whole title of the article, subsets of consecutive words of the title. For example, the next two (type 7) queries can be used as alternative ways for finding citations for the same publication, through Scopus:

```
( REF("Computer Assisted Assessment (CAA) of
  Free-Text: Literature Review and the
  specification of an alternative CAA system")
  AND NOT
  AUTHOR-NAME(Karanikolas)
)
```

```
( REF("Literature Review and the specification
  of an alternative CAA system")
  AND NOT
  AUTHOR-NAME(Karanikolas)
)
```

One considerable conjecture that should be investigated is the following:

The citations that are recognized automatically by the Scopus system/source (the citations included in the index card of a researcher) are always less than the citations that can be found by using either queries of type 5 (for each group of researchers that participates the researcher of interest) or queries of type 7 (for each one of the titles of author's articles). A similar conjecture is also investigated in the case of Google Scholar.

In order to evaluate our conjecture, we suggest a methodology based on type 5 or type 7 queries

(templates) for finding citations (excluding self-citations) of an author/researcher. Two alternative versions/variations of this methodology (one for each type of used queries) are synoptically presented in the following.

Methodology variation for type 5 queries:

- Create query of type 5 for finding citations (excluding self citations) that has received one group of (two or more) co-researchers (co-authors) in their common publications and in which participates the researcher under investigation,
- Repeat such kind of queries for anyone of the groups in which the researcher under investigation participates,
- Unify the results of the previous queries, for finding the citation count for the researcher under investigation.

Methodology variation for type 7 queries:

- Create query of type 7 for finding citations (excluding self citations) that has received one concrete publication (on the basis of the publication title),
- Repeat such kind of queries for every article co-authored by the researcher under investigation,
- Unify the results of the previous queries, for finding the citation count for the researcher under investigation.

In the following, we focus on the methodology variation with queries of type 7 and evaluate it for finding citations of an author. We prefer methodology variation for type 7 queries, because it permits us a simpler comparison of systems/sources. It also easily combines results retrieved by equivalent queries submitted in two or more systems/sources. For our experimentation, we can choose a small set of publications (journal articles and conference papers) and create queries of type 7 for each publication and for each system/source: Google Scholar and Scopus. The chosen set of publications contains eight (8) journal articles and ten (10) conference papers. Before providing measurements of efficiency for each system/source, we will concentrate in few of the thirty-six (36, 18 for Google Scholar and 18 for Scopus) queries. These queries have a behavior that should be focused. The following four (4) queries submitted to Google Scholar return only irrelevant results:

```
"Automatic Diagnosis Classification of
patient discharge letters"
-author:karanikolas -author:skourlas
0/3 (relevant retrieved / retrieved, precision)
```

```
"Conceptual Universal Database Language: Moving
Up the Database Design Levels"
-author:karanikolas -author:vassilakopoulos
0/1
```

"Bootstrapping the Albanian Information Retrieval" -author:karanikolas
0/1

"Computer Assisted Assessment (CAA) of Free-Text: Literature Review and the specification of an alternative CAA system" -author:karanikolas
0/1

The following two (2) queries submitted to Google Scholar return only relevant results, while the next adjunct four (4) queries submitted to Scopus (2 queries equivalent with the following 2 queries submitted to Google Scholar and two variations) return nothing:

"Interconnection of Laboratory Information System and Hospital Information System. The case of ARETEION University Hospital" -author:karanikolas -author:skourlas
1/1

"Strengthening the Security of E-Banking Transactions. The case of NBG" -author:karanikolas -author:marinakis
2/2

```
( REF("Interconnection of Laboratory Information System and Hospital Information System. The case of ARETEION University Hospital") AND NOT ( AUTHOR-NAME (Skourlas) OR AUTHOR-NAME (Karanikolas) ) )
```

0/0

Variation with part of the title

```
( REF("Interconnection of Laboratory Information System and Hospital Information System") AND NOT ( AUTHOR-NAME (Skourlas) OR AUTHOR-NAME (Karanikolas) ) )
```

0/0

```
( REF("Strengthening the Security of E-Banking Transactions. The case of NBG") AND NOT ( AUTHOR-NAME (Marinakis) OR AUTHOR-NAME (Karanikolas) ) )
```

0/0

Variation with part of the title

```
( REF("Strengthening the Security of E-Banking Transactions") AND NOT
```

```
( AUTHOR-NAME (Marinakis) OR AUTHOR-NAME (Karanikolas) ) )
```

0/0

From the above groups of queries (with 4, 2 and 4 queries), we have an indication of the increased recall of Google Scholar versus Scopus, and the decreased precision of Google Scholar versus Scopus.

In the following table, we summarize the efficiency of three approaches (the methodology variation with type 7 queries applied to Scopus, using the first mentioned set of 18 queries; the same methodology variation applied to Google Scholar, using the second mentioned set of 18 queries; the standard method automatically provided by Scopus for an author's citation count). The well known measures of precision and recall are used for estimating the performance of the proposed search methodology. Precision is defined as the fraction which is equal to the number of the relevant documents retrieved divided by the number of all the retrieved documents. Recall is defined as the fraction which is equal to the relevant documents retrieved divided by the number of all the documents that are relevant. Next table depicts the results of the proposed methodology and shows the calculation of precision and recall for 18 publications of a specific researcher (Karanikolas) without taking into account the self-citation.

Approach	precision	recall
Automatically measured by Scopus	5/12	5/18
Using queries of type 7 in Scopus	8/8	8/18
Using queries of type 7 in Google Scholar	10/16	10/18

It is obvious from the previous table that 8 relevant documents are retrieved by queries of type 7 in Scopus and 10 relevant documents are retrieved by queries of type 7 in Google Scholar. Taking into account that 4 of the relevant retrieved documents are common, we can calculate the performance of methodology for the combination of results: precision=14/20, recall=14/18.

It is worth of mentioning that when we use queries of type 5 in Scopus for the same researcher (Karanikolas) and for each of the groups in which he contributes, the following values for precision are calculated: 6/6, 1/1, 0/2 και 0/0. Therefore, the total performance in terms of precision is equal to 7/9 and in terms of recall is equal to 7/18. Comparing the results of using queries of type 5 with the results of using queries of type 7, both in Scopus, we conclude that the calculated precision and recall is better in the later variation of the proposed methodology (the variation with queries of type 7). This is another reason (except the already mentioned simplification of overall combination of results by two or more systems/sources) for suggesting the variation with queries of type 7.

IV. CONCLUSIONS

Our experimentation gave us evidence that our conjecture that “the proposed methodology (either variation with type 5 queries either variation with type 7 queries) improves the estimation of a researcher’s citation count versus the automatically measured by Scopus” is valid. It also seems that there is a slight advantage of using templates of queries of type 7 instead of using queries of type 5. Apart from this, the methodology of using queries of type 5 applies fewer and simpler queries and it is therefore an effective and fast solution for estimating the citation count. The combination of results from the use of the methodology variation with type 7 queries in both Scopus and Google Scholar seems to approach very close to the actual number of a researcher’s citations. Future work involves performing further experimentation using a corpus of publications organized per research groups, to uncover research policies. Another research direction is oriented towards the examination of other systems/sources and methods.

REFERENCES

- [David Brown and Bernard Dumouchel, “Understanding user behaviour and its metrics,” *Information Services & Use*, **27**, 3–34 \(2007\), IOS Press.](#)
- [C. Sean Burns, “Collecting bibliographic references: A bibliometric analysis of CiteULike's collection as grounds for in-depth interviews,” *Research round table at the Canadian Association for Information Science*, Fredericton, N.B. Canada, June 2011.](#)
- [Chapman, A.J., “Assessing research: citation-count shortcomings,” *The Psychologist: Bulletin of the British Psychological Society*, **8**, 336-344 \(1989\).](#)
- [Friedrich Gebhardt, Imant Stellmacher, “Opinion paper: Design criteria for documentation retrieval languages,” *Journal of the American Society for Information Science*, **29**, Issue 4, 191–199 \(1978\).](#)
- [Anne-Wil Harzing, “Google Scholar - a new data source for citation analysis,” 2008.](#)
- [Wen-Ru Hou, Ming Li, Deng-Ke Niu, “Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution,” *BioEssays*, **33**, 724–727 \(2011\).](#)
- [Nikitas N. Karanikolas, “Search Culture,” *Proc. 15th Panhellenic Conference on Informatics*, Kastoria, Greece, 229-234 \(2011\).](#)
- [NUS Libraries, “Citation Count Workshop for NUS Staff,” August 2011.](#)
- [Ian Ruthven, *Interactive information retrieval, Annual Review of Information Science and Technology*, Volume 42, Issue 1, pages 43–91, 2008.](#)