Foresight Academy of Technology
fatech.org.uk/press/

International Transactions on Systems Science and Applications
Volume 6 Number 2/3 August 2010 pp. 146-157

iTSSA

An Analysis of Research on Information Reuse and Integration (2003-2008)

Min-Yuh Day [1,2], Chorng-Shyong Ong [1] and Wen-Lian Hsu [2,*]

[1] Department of Information Management , National Taiwan University, Taiwan
[2] Institute of Information Science, Academia Sinica, Taiwan
Email: {myday; hsu}@iis.sinica.edu.tw; ongcs@im.ntu.edu.tw

Abstract: Information Reuse and Integration (IRI) plays a pivotal role in the capture, representation, maintenance, integration, validation, and extrapolation of information. Both information and knowledge are applied to enhance decision-making in various application domains. The objective of this paper is to provide a summary and analysis of research devoted to advancing the field of information reuse and integration. To this end, we identify the most popular research topics, together with the most productive researchers and institutions associated with the majority of research publications of the International Conference on Information Reuse and Integration during the past six years (2003-2008). Based on those publications, we have identified the most popular research topics, as well as the top researchers and institutions in the field of Information Reuse and Integration.

Keywords: content analysis; information reuse and integration; IRI topics; meta analysis.

1. Introduction

Information Reuse and Integration (IRI) plays a pivotal role in the capture, representation, maintenance, integration, validation, and extrapolation of information. Both information and knowledge are applied to enhance decision-making in various application domains. The study of IRI is a relatively new research area; thus, an analysis of the latest literature would be useful to show what researchers are doing and what can be done to improve our approach [8]. This paper analyzes articles published by the IEEE International Conference on Information Reuse and Integration between 2003 and 2008. Specifically, we consider topics, authors, and institutions.

IRI research can be broadly categorized into application development, IRI theory building, and the study of reference disciplines. Since IRI applications are the outcomes of IRI research activities, it is important to periodically survey IRI applications in order to monitor progress in the field as a basis for setting the direction of future research. The objective of this survey is to inform both academicians and practitioners about the areas in which IRI applications have been reported, as well as to provide insight into major historical trends, the implications of this study, and future directions for new theoretical developments [9]. We also provide a comprehensive reference list of literature on IRI.

The International Conference on Information Reuse and Integration (IRI) is currently sponsored by the IEEE Systems, Man, & Cybernetics Society. The first Information Reuse and Integration (IRI) conference was held in 1999. IEEE IRI 2003, 2004, 2005, 2007, 2008 were held in Las Vegas, and IRI 2006 was held in Waikoloa, Hawaii. Since it started in 2003, IEEE-IRI has become a major, highly respected international conference. The proceedings of the IEEE IRI conferences are indexed and included in IEEE Xplore, DBLP, Conference Proceedings Citation Index (CPCI), Scopus, and EI compendex citation databases.

We selected IRI conferences over other sources, such as computer science (CS) journals and other CS conferences for a number of reasons. First, we consider IEEE IRI to be the most prestigious

tai: itssa.2010.08.034 ISSN 1751-1461 (print) ISSN 1751-147X (cd-rom)

international conference in the field of IRI. For example, IRI has consistently received more than 200 submissions from researchers all over the world (IRI 2007), and it has become the primary forum for IRI researchers to interact, communicate, and demonstrate their research. Second, as a conference, IRI has a faster turnaround time than journals, i.e., its review process is much shorter than that of CS journals [28]. As a result, papers published in IRI can report new ideas and thoughts, the latest research directions, and up-to-date research results in a timely manner.

In line with our objective, we provide a summary and analysis of research that has advanced the field of information reuse and integration. We identify the most popular research topics, together with the most productive researchers and institutions associated with the majority of research publications of the International Conference on Information Reuse and Integration during the past six years (2003-2008).

The remainder of this paper is organized as follows. In Section 2, we introduce the background to this research. Section 3 describes the research methods and data we used in this study. In Section 4, we detail the results of our analysis and discuss their implications. Finally, in Section 5, we summarize our conclusions and consider future research avenues.

2. Research Background

We review related works on literature-based discipline assessment [3-4, 7, 22, 24-25, 28-29]. Two major approaches, classification and citation analysis, are used to assess a scientific discipline based on the domain's literature.

The classification approach categorizes the literature along different dimensions, such as topics and research methods [14, 23-25]. For example, Glass et al. [14] presented an analysis of research in the computing disciplines. They divided the computing field into the three most common academic subdivisions: computer science (CS), software engineering (SE), and information systems (IS). Vessey et al. [25] proposed a unified classification system for research in the computing disciplines that also consists of three academic subdivisions, namely CS, SE, and IS. They also introduced a multi-faced system based on five research-focused characteristics: topic, approach, method, unit of analysis, and reference discipline. Similarly, in the software engineering (SE) field, Glass [12] has provided an annual series of assessments of systems and software engineering scholars and institutions since 1994 [12-15]. Claver et al. [8] analyzed the articles published in the MIS Quarterly and Information & Management journals between 1981 and 1997 based on the most frequent topics, research methods, and top authors. Palvia et al. [20] identified the most productive authors and universities associated with the majority of research publications in Information & Management between 1992 and 2005. Arnott and Pervan [2] identified eight key issues for the decision support systems discipline by performing content analysis of 1093 DSS articles published in 14 major journals from 1990 to 2004.

The citation analysis approach is a bibliometric perspective that generates and analyzes networks of related papers (or authors) based on literature citation information [4, 11, 22, 26-27]. This approach, which has been used in several studies to measure the quality of journals and conferences [4, 11], is often used to study the research specialties and the intellectual structure of a wide range of scientific disciplines such as economics [18], computer science [5], and information science [26]. It has also been used to evaluate the productivity and reputation of individual researchers, academic programs and institutions, and journals [16]. For example, Chan et al. [4] employed a citation analysis approach to identify information systems citation patterns based on a three-year (2000-2002) set of International Conference on Information Systems (ICIS) articles. They provided the citation rankings of journals and conferences in the IS research area. In contrast to US-based conferences like ICIS, Whitley and Galliers [27] presented an alternative perspective on citation data from the first 10 years of the European Conference on Information Systems (ECIS). Moreover, Serenko and Cocosila [22] used a scientometric approach to investigate the state and evolution of information systems (IS) research in Canada as reflected in the annual conference of the Administrative Sciences Association of Canada from 1974 to 2007.

In addition to the above approaches, collaboration network analysis provides another important

literature-based perspective that considers co-authorship patterns. Under this approach, which has received more attention in recent years [7, 17, 19, 28], two authors are connected in a collaboration network if they have written one or more papers together. A collaboration network is a partial representation of the social network among authors, because authors who co-write papers are generally acquainted with each other. However, using co-authorships as a convenient substitute for social relations among researchers is not without methodological risk. For example, two researchers may work together on a project, but choose to publish their work separately. In addition, Xu and Chao [28] observed that directors of large laboratories or research centers may be listed as co-authors in papers on multiple projects, but they may not have participated in the projects directly. Their study investigated the social identity of IS by applying social network analysis to collaboration (co-authorship) networks for ICIS conference papers over 26 years (1980-2005) and identified the critical mass of the community, as well as the most productive authors and institutions. Subsequently, Cheong and B. J. Corbitt [7] used a social network analysis model to investigate the co-authorship network of the Pacific Asia Conference on Information Systems (PACIS) from 1993 to 2008.

Literature-based discipline assessment can be conducted by an automatic approach [21] or as a manual process [28]. Ren and Taylor [21] developed an automatic and versatile framework to support the ranking of publications by research institutions and scholars, but it is not suitable for the classification approach, i.e., topic classification. Hence, most literature-based discipline assessments have been conducted via a manual process.

Numerous works have examined the topics addressed in IS research [1, 6, 8, 10, 23] as well as in sub-fields of the IS domain, such as human computer interaction [29] and decision support systems [2, 9]. However, to the best of our knowledge, the topics addressed in IRI research have received little attention. To fill this gap in the literature, we consider the following research questions:

RQ1: What are the most popular research topics in IRI?

RQ2: Who are the most productive researchers in the IRI research community?

RQ3: Which institutions are the most productive in terms of IRI publications?

3. Method

Researchers have used two major approaches to assess the state of the IS field empirically [24, 28]. The first approach is based on classification studies that code categories of interest, typically topics and research methods, either manually or by using keywords [1, 6, 10]. The second approach is based on citation studies, which examine references to cited articles. It has been used almost exclusively to assess the degree of reliance of IS on other disciplinary areas [4]. Because we wished to conduct a comprehensive study that captured the diversity of IRI research, we adopted the classification approach in this study. To answer the above research questions, we analyzed research papers published by IRI conferences over the past six years (2003-2008).

Our analysis of IRI literature was based on the study of the IEEE International Conference on Information Reuse and Integration (IEEE IRI). The main reason for this decision was that IEEE IRI is a well established conference that focuses on research in the IRI area.

To compile an IRI literature list, we used two well-known online data sources: IEEE Xplore and DBLP. The data sources supply the following information about each article: title, authors, title of conference, year of publication, number of pages, keywords and Abstract. IEEE INSPEC provides affiliation information about the first author. We consider that this information is sufficient for our study, at least for the classification of articles belonging to one subject or topic and for identifying the authors.

We developed a simple web focused crawler program to download literature information about all IRI papers published between 2003 and 2008 from IEEE Xplore and DBLP. Note that we only included research papers; keynotes addresses and panel discussions were excluded.

After downloading all the data, we integrated the literature information in a spreadsheet. During data processing, we found some inconsistencies in the data provided by IEEE Xplore and DBLP. First, some authors' names were inconsistent; for example, the name "Stuart Harvey Rubin" (cited 14 times) was sometimes written as "Stuart H. Rubin"(6 times). Second, the first author's affiliations

were inconsistent, for example, “Central Michigan University, Mount Pleasant, USA” was sometimes written as “Central Michigan University, Mt Pleasant, MI”. We therefore checked the data manually and rectified the problem. After preprocessing the data, we obtained a list of 601 papers authored by 1301 authors in 42 countries. Table 1 shows the locations of IRI conferences from 2003 to 2008 and the number of papers published each year.

Table 1. Locations the IRI conferences and the number of papers published each year

Year	Location	Number of Papers
2003	Las Vegas, Nevada, USA	86
2004	Las Vegas, Nevada, USA	104
2005	Las Vegas, Nevada, USA	100
2006	Waikoloa, Hawaii, USA	107
2007	Las Vegas, Nevada, USA	120
2008	Las Vegas, Nevada, USA	84
TOTAL		601

Table 2. The most popular research topics in IRI (2003-2008)

Rank	Topics	IRI 2003	IRI 2004	IRI 2005	IRI 2006	IRI 2007	IRI 2008	Frequency	%
1	Data Mining and Knowledge Discovery	4	19	8	4	12	6	53	8.82%
2	Component-Based Design and Reuse	5	8	4	8	9	3	37	6.16%
3	Reuse in Software Engineering		7	10	4	4	3	28	4.66%
4	Fuzzy Neural Systems and Soft Computing		7	6	4	8		25	4.16%
5	Knowledge Acquisition and Management		8	8	4	4		24	3.99%
6	Agent Based Information Systems	8	3	3		5	3	22	3.66%
7	Information Assurance		6	7	8			21	3.49%
8	Heuristic Optimization and Search				8	8	3	19	3.16%
9	Multimedia Reuse and Integration	1	4	3	4	4	3	19	3.16%
10	AI & Decision Support Systems	4	3	7		4		18	3.00%
11	Biomedical & Healthcare Systems				7	4	6	17	2.83%
12	Large Scale Data Integration		3	4	4		6	17	2.83%
13	Natural Language Understanding		3	4	4	3	3	17	2.83%
14	Software Stability	7		3		6		16	2.66%
15	Manufacturing Systems & Business Process Engineering	4				3	6	13	2.16%
16	Software Development, Frameworks and Tools for Reuse		6	6				12	2.00%
17	Database Systems and Integration	5					6	11	1.83%
18	HW & SW Engineering for Reuse		3	7				10	1.66%
19	Knowledge Management and E-Government		3		4	3		10	1.66%
20	Knowledge Management and Ontological Engineering			3		4	3	10	1.66%

4. Data Analysis and Discussion

The analysis is based on the data derived from the papers published by IRI conferences during the past six years (2003-2008). The purpose of this analysis is to address the research questions mentioned earlier.

It should be noted that the published IRI literature is not representative of IRI applications in

practice, as many IRI applications go unreported. Therefore, we should not interpret the results of this survey as if they reflect real world practice [9].

4.1. Topics

An analysis of the topics most frequently addressed in the 2003-2008 period is presented in Table 2. The table shows the number of articles dedicated to each topic and the percentage of the total number of articles considered. Prior research [20] indicated that the total count of all topic frequencies was higher than the total number of articles for this period, simply because an article often dealt with multiple topics. Although there were 69 distinct topics for the period studied, Table 2 only shows the top 20 topics.

The table also shows that the most popular IRI topics during the 6-year study period were Data Mining and Knowledge Discovery (8.82% of the articles), followed by Component-Based Design and Reuse (6.16%), Reuse in Software Engineering (4.66%), Fuzzy Neural Systems and Soft Computing (4.16%), and Knowledge Acquisition and Management (3.99%). Taken together, these topics represent 27.79% of the articles analyzed.

- **Controlled Index Terms.** A controlled index term is an assignment of subject descriptors derived from a regulated set of words and phrases (a list of authorized descriptors) found in the INSPEC thesaurus. INSPEC, the world's leading bibliographic information service, provides access to the world's scientific and technical literature. The thesaurus is produced by The Institution of Engineering and Technology (IET)(http://www.theiet.org/). In the information retrieval field, a controlled index term is called controlled indexing. Controlled index terms are useful for searching core subject areas and topics with numerous synonyms; while controlled indexing is used for standardized spelling (e.g., analysers – analyzers, analysers), punctuation (e.g., online—on line, on-line, online), and terminology (e.g., Internet – world wide web/www.).

Table 3. The most popular controlled index terms in IRI (2003-2007)

Rank	Controlled Index Term	Frequency	%
1	software reusability	85	3.95%
2	object-oriented programming	69	3.20%
3	Internet	61	2.83%
4	data mining	58	2.69%
5	formal specification	35	1.63%
6	software architecture	34	1.58%
7	knowledge management	28	1.30%
8	learning (artificial intelligence)	28	1.30%
9	XML	25	1.16%
10	information retrieval	24	1.11%
11	knowledge based systems	23	1.07%
12	ontologies (artificial intelligence)	21	0.98%
13	decision making	19	0.88%
14	genetic algorithms	18	0.84%
15	decision support systems	17	0.79%
16	Java	17	0.79%
17	pattern classification	15	0.70%
18	software metrics	15	0.70%
19	software quality	15	0.70%
20	multi-agent systems	14	0.65%

Note: The controlled index terms were obtained from IEEE INSPEC. (The controlled index terms for IRI 2008 are not yet available)

Note that the controlled index terms for IRI 2008 are not yet available from IEEE Xplore. For the period 2003 to 2007, there were 594 distinct controlled index terms in papers published by the IRI conferences. Table 3 shows the most popular controlled index terms for that period. The top 10 controlled index terms were software reusability, object-oriented programming, Internet, data mining, formal specification, software architecture, knowledge management, learning (artificial intelligence), XML, and information retrieval.

- **Non-Controlled Index Terms.** Non-controlled index terms are keywords and phrases selected from document titles, abstracts, and full texts provided by INSPEC. The definition of an uncontrolled index (uncontrolled terms) is "one field of an INSPEC record that includes key words and phrases from the original document selected by indexers to represent the key concepts of the document" (IEEE INSPEC). Non-controlled index terms are useful for describing new concepts, non-scientific/non-technical topics, and terms that are also common words.

There were 3410 distinct non-controlled index terms in IRI publications from 2003 to 2007. The data for IRI 2008 is not yet available from IEEE Xplore. Table 4 shows the most popular non-controlled index terms in IRI (2003-2007). The top 10 non-controlled index terms for that period were data mining, Internet, software development, knowledge representation, software reuse, information reuse, XML, genetic algorithm, information retrieval, and knowledge management. We observe that "Data mining" is the most popular research topic (Rank 1) and also the most popular non-controlled index term (Rank 1).

Table 4. The most popular non-controlled index terms in IRI (2003-2007)

Rank	Non-Controlled Index Term	Frequency	%
1	data mining	22	0.56%
2	Internet	17	0.43%
3	software development	15	0.38%
4	knowledge representation	13	0.33%
5	software reuse	13	0.33%
6	information reuse	11	0.28%
7	XML	11	0.28%
8	genetic algorithm	10	0.25%
9	information retrieval	10	0.25%
10	knowledge management	10	0.25%
11	software architecture	10	0.25%
12	software engineering	10	0.25%
13	machine learning	9	0.23%
14	software stability	9	0.23%
15	information integration	8	0.20%
16	decision making	7	0.18%
17	Java	7	0.18%
18	knowledge acquisition	7	0.18%
19	knowledge reuse	7	0.18%
20	semantic Web	6	0.15%

Note: The non-controlled index terms were obtained from IEEE INSPEC. (The non-controlled index terms for IRI 2008 are not yet available)

- **Title Words.** In total, 1578 distinct words were used in the titles of 601 IRI papers between 2003 and 2008. A word cloud is a quick way to obtain an overall impression of the themes in articles based on the frequency of words in the titles. We used the titles of IRI papers listed by the Wordle word cloud generator (Wordle.net). Figure 1 shows the word cloud analysis of IRI paper titles (2003-2008). The top 10 title words were based, using, knowledge, system, data, model, software, design, web, and reuse.

Figure 1. Word cloud analysis of IRI paper titles (2003-2008).
Note: Word cloud visualization from Wordle (http://wordle.net)

Table 5. Authors with the most articles published in IRI (2003-2008)

Rank	Author Name	Current Affiliation	IRI 2003	IRI 2004	IRI 2005	IRI 2006	IRI 2007	IRI 2008	Frequency	%
1	Stuart Harvey Rubin	SPAWAR Systems Center, USA	4	4	3	4	3	2	20	1.12%
2	Taghi M. Khoshgoftaar	Florida Atlantic University, USA		2	4	5	4	3	18	1.01%
3	Mohamed E. Fayad	San José State University, USA	4	3	4		5		16	0.89%
4	Shu-Ching Chen	Florida International University, USA	4	2	2	3	1	1	13	0.73%
5	Mei-Ling Shyu	University of Miami, USA	3	2	3	1	1	1	11	0.61%
6	Wen-Lian Hsu	Academia Sinica, Taiwan			1	2	2	4	9	0.50%
7	Reda Alhajj	University of Calgary, Canada	1	2	2		1	3	9	0.50%
8	Min-Yuh Day	National Taiwan University , Taiwan			1	2	2	3	8	0.45%
9	Du Zhang	California State University, USA	1	2	1	2	1	1	8	0.45%
10	Narayan C. Debnath	Winona State University, USA		5	3				8	0.45%
11	Jason Van Hulse	First Data Corp., USA			1	2	2	2	7	0.39%
12	Eric Gregoire	Universiti d'Artois, France	1	1	1	1	2	1	7	0.39%
13	Meiliu Lu	California State University, USA		3	1	2		1	7	0.39%
14	Waleed W. Smari	University of Dayton, USA	3	1	1	2			7	0.39%
15	Viktor K. Prasanna	University of Southern California, USA			1	1	1	3	6	0.34%
16	Roumen Kountchev	University of Sophia, Bulgaria	1	1		1	2	1	6	0.34%
17	Chengcui Zhang	University of Alabama at Birmingham, USA	3			1	1	1	6	0.34%
18	Chia-Chu Chiang	University of Arkansas at Little Rock, USA	1	1	1	1	1	1	6	0.34%
19	Seung-yun Kim	Shepherd University, USA	3	1		1		1	6	0.34%
20	Eduardo Santana de Almeida	Federal University of Pernambuco, Brazil		3	2		1		6	0.34%

4.2. Authors

As noted in prior research [28], the most productive authors or leaders in a scientific discipline are

important assets to that discipline. Treating scientific growth as a process of knowledge diffusion, leaders are often the ones who introduce new ideas and thus have a great impact on their respective research communities.

There were 1301 distinct authors in IRI between 2003 and 2008. Table 5 shows the authors who published the most articles during that period. The most productive authors were Stuart Harvey Rubin (20), followed by Taghi M. Khoshgoftaar (18), Mohamed E. Fayad (16), Shu-Ching Chen (13), Mei-Ling Shyu (11), Wen-Lian Hsu (9), Reda Alhajj (9), Min-Yuh Day (8), Du Zhang (8), and Narayan C. Debnath (8). Note that the productivity data reported here is based on IRI records only.

Table 6 and Figure 2 show the distribution of the number of co-authors in IRI (2003-2008).The top 3 categories were 3 co-authors (32.61%), 2 co-authors (29.45%), and 4 co-authors (18.30%). In total, 90.52% of the papers were written by more than one author. Hence, single-author papers accounted for only 9.48% of the total. These results support the commonly held view that co-authorship is the predominant trend in the field of IRI.

Table 6. The number of co-authors in IRI (2003-2008)

Number of co-authors	IRI2003	IRI2004	IRI2005	IRI2006	IRI2007	IRI2008	Frequency	%
1	16	5	13	7	12	4	57	9.48%
2	19	35	29	30	43	21	177	29.45%
3	30	33	33	37	30	33	196	32.61%
4	15	17	12	22	29	15	110	18.30%
5	4	12	5	5	2	6	34	5.66%
6		1	4	4	2	4	15	2.50%
7	1	1	3	1	1	1	8	1.33%
8	1			1	1		3	0.50%
9							0	0.00%
10			1				1	0.17%
TOTAL	86	104	100	107	120	84	601	100.00%

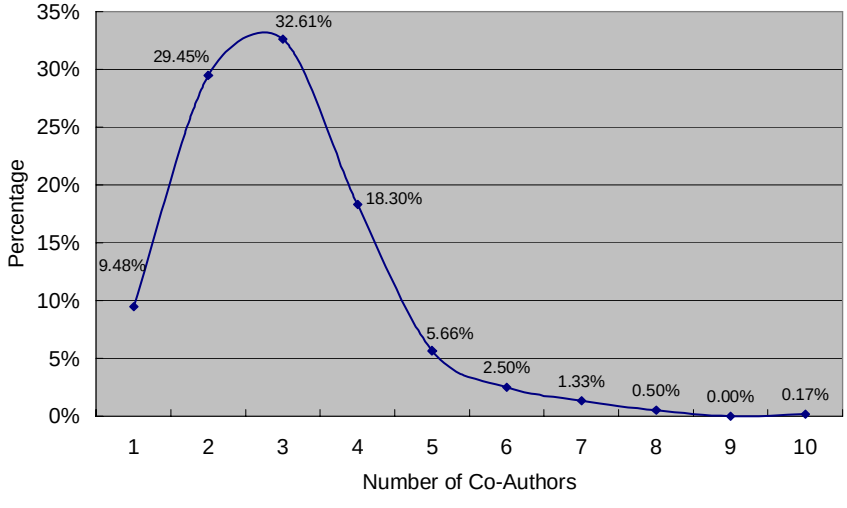

Figure 2. The distribution of the number of co-authors in IRI (2003-2008)

4.3. Institutions

In total, 381 institutions published papers in IRI between 2003 and 2008. Table 7 shows the

institutions that published the most articles during that period. The most productive institutions were California State University, Sacramento, USA (17), Florida Atlantic University, USA (16), San José State University, USA (13), Florida International University, USA (9), Tsinghua University, China (9), University of Arkansas at Little Rock, USA (8), SPAWAR Systems Center, USA (8), Academia

Table 7. Institutions with the most articles published in IRI (2003-2008)

Rank	Institution	IRI 2003	IRI 2004	IRI 2005	IRI 2006	IRI 2007	IRI 2008	Frequency	%
1	California State University, Sacramento, USA	1	4	5	4	1	2	17	4.45%
2	Florida Atlantic University, USA		2	3	4	3	4	16	4.19%
3	San José State University, USA	4	1	3		5		13	3.40%
4	Florida International University, USA	3	1	1	2	1	1	9	2.36%
5	Tsinghua University, China	1	1		5	1	1	9	2.36%
6	University of Arkansas at Little Rock, USA	1	1		2	3	1	8	2.09%
7	SPAWAR Systems Center, USA	2	2	4				8	2.09%
8	Academia Sinica, Taiwan			1	2	1	2	6	1.57%
9	Dayton University, USA	2	1	1	2			6	1.57%
10	Southeast University, China	4	2					6	1.57%
11	University of Southern California, USA		1			1	3	5	1.31%
12	National Chiao Tung University, Taiwan		1	3			1	5	1.31%
13	CRIL, University d'Artois, France		1	1	1	2		5	1.31%
14	Calgary University, Canada		2	2	1			5	1.31%
15	Carleton University, Canada		3	1	1			5	1.31%
16	Nebraska University, USA	2	2		1			5	1.31%
17	Central Michigan University, USA	2		2	1			5	1.31%
18	National Taiwan University, Taiwan			1		1	2	4	1.05%
19	Wuhan University, China		2				2	4	1.05%
20	National Cheng Kung University, Taiwan	2				2		4	1.05%

Table 8. The most frequent presentation of IRI (2003-2008) papers by country (Total of 601 papers)

Rank	Country	IRI2003	IRI2004	IRI2005	IRI2006	IRI2007	IRI2008	Frequency	%
1	USA	32	43	53	51	49	30	258	42.93%
2	China	21	18	4	11	7	11	72	11.98%
3	Taiwan	2	5	8	4	6	7	32	5.32%
4	France	4	6	4	9	5	2	30	4.99%
5	Canada	1	6	8	2	5	4	26	4.33%
6	Japan	3	5	3	5	3	6	25	4.16%
7	Brazil	5	4	3	3	4	2	21	3.49%
8	UK	3	1	4	3	5	2	18	3.00%
9	South Korea	1	1	1	6	8	1	18	3.00%
10	Germany		2	1		6	1	10	1.66%
11	Australia	4	1		2	2	1	10	1.66%
12	India				2	2	3	7	1.16%
13	Iran	1		1		2	1	5	0.83%
14	Czech Republic				3	1	1	5	0.83%
15	Thailand		2	1		1	1	5	0.83%
16	Italy	2		2			1	5	0.83%
17	Spain	1	2	1		1		5	0.83%
18	Israel					2	2	4	0.67%
19	Bulgaria	1	1			1	1	4	0.67%
20	Singapore			1	1	1	1	4	0.67%

Sinica, Taiwan (6), Dayton University, USA (6), Southeast University, China (6). Note that the productivity of the various institutions detailed here is based on the first author's affiliation according to IRI data.

4.4. Country

During the study period IRI published 601 papers from 42 countries. Table 8 shows the most frequent representation of IRI (2003-2008) papers by country. Eleven countries had at least 10 papers published by IRI over the 6 years. As shown in Table 8, the USA was the biggest contributor (42.93%), followed by China (11.98%), Taiwan (5.32%), France (4.99%), and Canada (4.33%). The USA has had a consistently large number of papers published by each IRI conference. The top 10 countries accounted for 84.86% of the total contributions in IRI (2003 to 2008).

5. Conclusion

In this paper, our goal has been to provide a summary of research in order to advance the field of information reuse and integration. To this end, we have identified the most popular research topics, as well as the most productive researchers and organizations associated with the majority of research publications in the International Conference on Information Reuse and Integration (IRI) during the past six years (2003-2008). Based on a detailed analysis of those publications, we have identified the research topics that were investigated most often. Our findings show that the research conducted in the field of IRI covers a wide variety of topics.

The major contribution of this paper is that it is the one of the few empirical studies that focuses on research in the field of information reuse and integration. Specifically, we present the findings of a six-year study of the most popular research topics, as well as the top researchers and institutions in the IRI field. The top 10 topics were Data Mining and Knowledge Discovery, Component-Based Design and Reuse, Reuse in Software Engineering, Fuzzy Neural Systems and Soft Computing, Knowledge Acquisition and Management, Agent Based Information Systems, Information Assurance, Heuristic Optimization and Search, Multimedia Reuse and Integration, AI & Decision Support Systems. The top 10 controlled index terms were software reusability, object-oriented programming, Internet, data mining, formal specification, software architecture, knowledge management, learning (artificial intelligence), XML, and information retrieval. The top 10 non-controlled index terms were data mining, Internet, software development, knowledge representation, software reuse, information reuse, XML, genetic algorithm, information retrieval, and knowledge management. The top 10 title words were based, using, knowledge, system, data, model, software, design, web, and reuse. The top researcher was Stuart Harvey Rubin of the SPAWAR Systems Center, USA, and the top institution was California State University, Sacramento, USA. The top 10 countries were the USA, China, Taiwan, France, Canada, Japan, Brazil, UK, South Korea, and Germany.

In the future, we will extend our study to investigate the IRI scientific network though co-authorship analysis. By so doing, we hope to develop a more comprehensive view of the IRI research community.

Acknowledgements

This research was supported in part by the National Science Council of Taiwan under grants NSC 96-2752-E-001-001-PAE and 95-2416-H-002-047, and the thematic program of Academia Sinica under grant AS95ASIA02.

References

[1] M. Alavi and P. Carlson, A review of MIS research and disciplinary development, Journal of

Management Information Systems, vol. 8, pp. 45-62, 1992.
[2] D. Arnott and G. Pervan, Eight key issues for the decision support systems discipline, Decision Support Systems, vol. 44, pp. 657-672, 2008.
[3] R. D. Banker and R. J. Kauffman, The evolution of research on information systems: A fiftieth-year survey of the literature in Management Science, Management Science, vol. 50, pp. 281-298, Mar 2004.
[4] H. C. Chan, H. W. Kim, and W. C. Tan, Information systems citation patterns from international conference on information systems articles, Journal of the American Society for Information Science and Technology, vol. 57, pp. 1263-1274, Jul 2006.
[5] C. M. Chen and R. J. Paul, Visualizing a knowledge domain's intellectual structure, IEEE Computer, vol. 34, pp. 65-71, Mar 2001.
[6] M. J. Cheon, V. Grover, and R. Sabherwal, The evolution of empirical research in IS: A study in IS maturity, Information & Management, vol. 24, pp. 107-119, 1987.
[7] F. Cheong and B. J. Corbitt, A social network analysis of the co-authorship network of the Pacific Asia Conference on Information Systems from 1993 to 2008, in Proceedings of Pacific Asia Conference on Information Systems (PACIS 2009) Hyderabad, India, 2009.
[8] E. Claver, R. Gonzalez, and J. Llopis, An analysis of research in information systems (1981-1997), Information & Management, vol. 37, pp. 181-195, Jun 2000.
[9] S. Eom and E. Kim, A survey of decision support system applications (1995-2001), Journal of the Operational Research Society, vol. 57, pp. 1264-1278, Nov 2006.
[10] A. F. Farhoomand, Scientific Progress of Management Information Systems, Data Base, vol. 18, pp. 48-56, 1987.
[11] G. A. Forgionne and R. Kohli, A multiple criteria assessment of decision technology system journal quality, Information & Management, vol. 38, pp. 421-435, Aug 2001.
[12] R. L. Glass, An Assessment of Systems and Software Engineering Scholars and Institutions, Journal of Systems and Software, vol. 27, pp. 63-67, Oct 1994.
[13] R. L. Glass, An Assessment of Systems and Software Engineering Scholars and Institutions, 1993 and 1994, Journal of Systems and Software, vol. 31, pp. 3-6, Oct 1995.
[14] R. L. Glass, V. Ramesh, and I. Vessey, An analysis of research in computing disciplines, Communications of the ACM, vol. 47, pp. 89-94, Jun 2004.
[15] R. L. Glass and T. Y. Chen, An assessment of systems and software engineering scholars and institutions (1999-2003), Journal of Systems and Software, vol. 76, pp. 91-97, Apr 2005.
[16] H.-H. Huang and J. S.-C. Hsu, An Evaluation of Publication Productivity in Information Systems: 1999 to 2003, Communications of the AIS, vol. 14, pp. 555-564, 2003.
[17] X. M. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel, Co-authorship networks in the digital library research community, Information Processing & Management, vol. 41, pp. 1462-1480, Dec 2005.
[18] K. W. McCain, Mapping authors in intellectual space: A technical overview, Journal of American Society of Information Science, vol. 41, pp. 433-443, 1990.
[19] M. E. J. Newman, Coauthorship networks and patterns of scientific collaboration, in Proceedings of the National Academy of Sciences. vol. 101, 2004, pp. 5200-5205.
[20] P. Palvia, P. Pinjani, and E. H. Sibley, A profile of information systems research published in Information & Management, Information & Management, vol. 44, pp. 1-11, Jan 2007.
[21] J. Ren and R. N. Taylor, Automatic and versatile publications ranking for research institutions and scholars, Communications of the ACM, vol. 50, pp. 81-85, Jun 2007.
[22] A. Serenko, M. Cocosila, and O. Turel, The State and Evolution of Information Systems Research in Canada: A Scientometric Analysis, Canadian Journal of Administrative Sciences-Revue Canadienne Des Sciences De L Administration, vol. 25, pp. 279-294, Dec 2008.
[23] E. B. Swanson and N. C. Ramiller, Information Systems Research Thematics: Submissions to a New Journal, Information Systems Research, vol. 4, pp. 299-330, 1993.
[24] I. Vessey, V. Ramesh, and R. L. Glass, Research in information systems: An empirical study of diversity in the discipline and its journals, Journal of Management Information Systems, vol. 19, pp. 129-174, Fal 2002.

[25] I. Vessey, V. Ramesh, and R. L. Glass, A unified classification system for research in the computing disciplines, Information and Software Technology, vol. 47, pp. 245-255, Mar 15 2005.
[26] H. D. White and K. W. McCain, Visualizing a discipline: An author co-citation analysis of information science, 1972-1995, Journal of the American Society for Information Science, vol. 49, pp. 327-355, Apr 1998.
[27] E. A. Whitley and R. D. Galliers, An alternative perspective on citation classics: Evidence from the first 10 years of the European Conference on Information Systems, Information & Management, vol. 44, pp. 441-455, Jul 2007.
[28] J. Xu and M. Chau, The social identity of IS: analyzing the collaboration network of the ICIS conferences (1980-2005), in Proceedings of the International Conference on Information Systems, Milwaukee, 2006, pp. 569-589.
[29] P. Zhang and N. Li, The intellectual development of Human-Computer Interaction research: A critical assessment of the MIS literature (1990-2002), Journal of Association for Information Systems, vol. 6, pp. 227-292, 2005.