Scientia Iranica E (2021) 28(3), 1830–1852

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

Evolution of IT, management and industrial engineering research: A topic model approach

M. Rabiei[a], S.-M. Hosseini-Motlagh[a,*], A. Haeri[a], and B. Minaei Bidgoli[b]

a. *School of Industrial Engineering, Iran University of Science and Technology (IUST), University Ave. Narmak, 16846-13114, Tehran, Iran.*
b. *School of Computer Engineering, Iran University of Science and Technology (IUST), University Ave. Narmak, 16846-13114, Tehran, Iran.*

Received 15 May 2019; received in revised form 24 July 2019; accepted 22 February 2020

KEYWORDS
Research evolution;
Topic modeling;
Trend analysis;
Information Technology (IT);
Industrial engineering;
Management.

Abstract. Information Technology (IT), management and industrial engineering are correlated academic disciplines whose publications have risen significantly over the last decades. The aim of this study is to analyze the research evolution, determine the important topics and areas, and depict the trend of interdisciplinary topics in these domains. To accomplish this, text mining techniques are used and a combination of bibliographic analysis and a topic modeling approach are applied to relevant publications in the Web of Science (WoS) repository over the last 20 years. In the topic extraction process, a heuristic function was suggested for key extraction, and some new applicable criteria were defined to compare the topics. Moreover, a novel approach was proposed to determine the high-level category for each topic. The results determined the hot-important topics, and incremented, decremented and fixed topics are identified. Subsequently, a comparison between high-level research areas confirmed strong scientific relationships between them. This study presents a deep knowledge about the internal research evolution of domains and illustrates the effect of topics on each other over the past 20 years. Furthermore, the methodology of this study could be applied to determine interdisciplinary topics and observe the research evolution of other academic domains.

1. Introduction

Management, *Information Technology (IT)* and *Industrial Engineering* are academic research disciplines with a high amount of correlation. In many universities, Industrial Engineering and Management are blended together and consist of both engineering programs (e.g. mathematics, engineering analysis and design, manufacturing processes, and quality assurance and control) and managerial courses (e.g. cost analysis, Human Resources (HR), and business and marketing).

The close link between Industrial engineering and management has led to the creation of business schools and some new academic disciplines, such as MBA degrees, in top U.S. universities and institutes, to produce e managers trained in the new science of business administration [1]. On the other hand, IT and computer science have been regarded as undeniable parts of these disciplines. Porter and Rafols [2] illustrated that science is indeed becoming more interdisciplinary and the connections between distant cognitive areas are modestly increasing.

*. *Corresponding author. Tel.: +98 21 73225000-1; Fax: +98 21 73225098*
E-mail addresses: mo_rabiei@ind.iust.ac.ir (M. Rabiei); motlagh@iust.ac.ir (S.-M. Hosseini-Motlagh); ahaeri@iust.ac.ir (A. Haeri); b_minaei@iust.ac.ir (B. Minaei Bidgoli)

doi: 10.24200/sci.2020.53582.3312

The strength of linkage between these domains and the dramatic growth of their publications over the last decades [3–7] has created a diversity of research into several investigative lines and has led to constructing new paradigms in the research areas of these domains [5,8]. Creation of new paradigms creates difficulty for researchers and policymakers in understanding and monitoring the evolution of these related domains.

The aim of this study is the analysis of the evolution of *Management*, *Industrial Engineering*, and *IT* studies and determination of the important topics and trends of each domain over the last 20 years. Moreover, this research tries to illustrate the growth, decline and change in common research areas in these three interrelated domains in a single window. To do this, a hybrid method and some new criteria are introduced to make the inter-relation and intra-relation of the research domains and topics visible and comparable. The results of the present investigation are applicable for scholars, students, journal editors and scientific policymakers of these three domains, in order to analysis the historical evolution and future development of these disciplines. They could use the results of this study to obtain a deep knowledge about the most important topics and extract the common topics which bind these areas together. The method and process of analyzing the correlated research domain in this study are also applicable to other domains.

The structure of this research is as follows: In the next section, the previous scientometric studies in these domains are reviewed. Then, the methodology of this research is described. After that, a bibliographic analysis of relevant publications is presented. In the "Topic modeling" section, the steps of topic modeling are explained. Then, the results of the previous section are reported and discussed in the "Results and discussion" section. Finally, the main conclusions are summarized and some suggestions for future studies are outlined.

2. Literature review

Wagner et al. [9,10] claim that interdisciplinary research (IDR) derives from a typology presented at the first international conference on IDR and teaching in 1970. They also believe that IDR could be considered as an evolution, rather than a state. Therefore, many of studies have been undertaken to illustrate the evolution of science in the various domains.

More than half century ago, direct citation linkages were used to show the evolutionary pathways within a research [11]. Later, Small [12] clustered the highly co-cited documents to detect the hot topics. Co-citation is not the only clustering approach that has been used to identify emerging topics. Using bibliographic information is another common method that is applied widely to detect the emerging topics in a domain. Refs. [4,13,14] are just some cases of this method. In recent years, some studies applied text mining or a topic modeling approach to detect important topics in the research studies [15–19].

Furrer et al. [20] used the content analysis method to investigate the strategic management research published in 1980–2005, and illustrated the relationships between the subfields of strategic management. They then depicted the evolution of the literature in five periods of time. The bibliographic analysis is applied to the forty year publication of the Computers & Industrial Engineering (CIE) journal for identification of the leading topics, institutions and countries in the industrial engineering domain [21]. In another research [22], the associations between standardization and evolution of the Information and Communications Technology (ICT) industry is examined. To do this, co-occurrence between ICT research areas and ICT standards are analyzed and the relation of ICT openness or concentration with the number of standards in this domain is investigated.

Studies on the evolution of the intellectual structure of management as an academic discipline are lacking in the literature, and the dynamics of the internal evolution of its research topics is also vague for the researchers [6]. In the same way, this gap is apparent for the IT and industrial engineering studies. In addition, the search to find a comprehensive research that compares the topics and trends of these related domains in a single window has had no results. Table 1 classifies some important previous scientormetric studies in these domains.

Regardless of the method or approach for topic detection, which have been mentioned above, previous research analyses in these domains are limited to narrow research areas or are undertaken for a specific location. For example, international business [23], industrial productivity [5], supply chain resilience [24], public private partnership [14], information privacy [16], cloud computing [4], health IT [25], green building based studies [26], and multiple criteria decision making [27] were analyzed. In some cases, analyzing the research status of a domain in a specific location has been considered; environmental science in Iran [28], biomedical engineering in Thailand [18] and management in Latin America and the Caribbean [6] are some samples.

In comparison with previous studies, this research applies a combination of bibliographic analysis and a topic modeling approach to identify hot and important topics and to monitor the evolution of internal research lines in three correlated domains. In addition, because of the close relationship between management, industrial engineering and IT disciplines, the scope of this study is wide and all three domains are considered.

Table 1. Previous scientometric studies in these domains.

Domain	Method	Sources	Duration	Ref.
General science	Citation Analysis	SCI-EXPANDED	1970–1974	[12]
Green building based	Citation & Bibliographic Analysis	Web of Science	2002–2018	[26]
Cloud computing	Bibliographic Analysis	Scopus	2008–2013	[4]
Antarctic	Bibliographic Analysis	SCI-EXPANDED	1900–2012	[13]
Public-private partnership	Bibliographic Analysis	Scopus	1990–2013	[14]
Industrial engineering	Bibliographic Analysis	Journal of CIE	1976–2016	[21]
Medical big data	Bibliographic Analysis	Web of Science	1991–2017	[29]
MCDM	Bibliographic Analysis	Web of Science	1993–2018	[27]
Supply Chain Resilience (SCR)	Bibliographic & Content Analysis	Google Scholar & Scopus	2002–2017	[24]
Strategic management	Content Analysis	Four related Journals	1980–2005	[20]
ICT	Content Analysis	Korean patents & standards	1977–2010	[22]
Management in Latin America	Content Analysis	Web of Science	1988–2013	[6]
International Business (IB)	Content Analysis	Top eight IB journals	1991–2015	[23]
Industrial productivity	Content Analysis & Interviews	Scopus	1970–2013	[5]
Biomedical engineering	Bibliographic Analysis & Topic Modeling	Scopus	1980–2010	[18]
Information privacy	Topic Modeling	Scopus	1972–2015	[16]
Big data in marketing	Topic Modeling	Science Direct	2010–2015	[15]
7 scientific areas	Topic Modeling	Web of Science	Not Given	[19]
Statistics	Topic Modeling	Three related journals	2000–2010	[17]

This study will try to answer these questions: What are the bibliographic characteristics of research studies in these three domains over the past 20 years? What are the important topics in each domain? What happened to important topics of each domain within 20 years? Which high-level research areas are covered alone by these domains or what do they have in common with them?

3. Methodology

To make the corpus of publication for management, industrial engineering, and IT domains, the Web of Science (WoS) repository was used. The WoS is a quality controlled repository which is used as a data source to make the corpus in a wide range of scientometric studies [6,19,26,27,29,30]. At the first step, in order to choose an appropriate dataset for each domain and after consulting with domain experts, some indicator terms are selected for the *Organization* and the *Sub organization* fields of WoS. To do this, "information", "computer", "knowledge", and "software" are used as indicators of organizations' names that have done research in the IT domain. In the same way, "industr*" and "management*" are searched, respectively, as indicators of the industrial engineering and the management domains. All queries are restricted to the *English* language and *Articles* published in the last 20 years (from 1998 to 2017). Moreover, documents that are indexed in the Science Citation Index Expanded (SCI-EXPANDED) or the Social Science Citation Index (SSCI) are considered. The result of this query contains a large number of articles from various subject categories. To make the results more homogenous and avoid the effects of outlier data, the top 20 WoS subject categories of each domain are identified. Every journal and book covered by WoS is assigned to at least one of the subjects listed in this category [31]. Using WoS subject categories in the process of dataset selection is widely used in scientometric analysis studies [14,32–35]. Table 2 reports these subject categories for each domain. In this table, "#" and "%" indicate, respectively, the number and percentage of articles covered by each category. Obviously, some articles are assigned to more than one subject and some subjects are common for both or all three of the domains. Therefore, the sum of the "%" column is more than 100. Consequently, 43

Table 2. Top 20 Web of Science (WoS) subject categories for each domain.

		Mng		Ind		IT	
	WoS categories	#	%	#	%	#	%
1	Engineering, Industrial	13999	4.0	72165	83.7	–	–
2	Operations research management science	26533	7.6	30046	34.8	1063	2.4
3	Management	3347	12.4	7954	9.2	–	–
4	Engineering, Manufacturing	9016	2.6	27961	32.4	–	–
5	Environmental sciences	27434	7.8	–	–	–	–
6	Economics	23024	6.6	2673	3.1	–	–
7	Business	24107	6.9	–	–	–	–
8	Computer Science, Interdisciplinary Applications	9111	2.6	9895	11.5	2441	5.5
9	Computer Science, Information Systems	10601	3.0	–	–	6919	15.5
10	Computer Science, Artificial Intelligence	10432	3.0	–	–	5767	13.0
11	Engineering, Electrical & Electronic	9088	2.6	-	-	6159	13.8
12	Psychology, Applied	7919	2.3	6744	7.8	–	–
13	Industrial Relations & Labor	-	-	13068	15.2	–	–
14	Public, Environmental & Occupational Health	12008	3.4	–	–	923	2.1
15	Ecology	11918	3.4	–	–	–	–
16	Materials Science, Multidisciplinary	–	–	11236	13.0	–	–
17	Water resources	10302	2.9	–	–	–	–
18	Health Care sciences & Services	9119	2.6	–	–	893	2.0
19	Environmental studies	8602	2.5	–	–	–	–
20	Engineering, Environmental	8272	2.4	–	–	–	–
21	Business, Finance	8068	2.3	–	–	–	–
22	Ergonomics	–	–	7729	9.0	–	–
23	Health Policy & Services	7264	2.1	–	–	–	–
24	Computer Science, Software Engineering	–	–	–	–	7019	15.8
25	Computer Science, Theory & Methods	–	–	–	–	6715	15.1
26	Engineering, Multidisciplinary	–	–	3604	4.2	987	2.2
27	Psychology	–	–	3695	4.3	–	–
28	Multidisciplinary sciences	-	-	2199	2.6	1061	2.4
29	Engineering, Civil	–	–	3098	3.6	–	–
30	Telecommunications	–	–	–	–	3075	6.9
31	Business	–	–	2828	3.3	–	–
32	Computer Science, Hardware & Architecture	–	–	–	–	2255	5.1
33	Construction & Building Technology	-	–	2249	2.6	–	–
34	Automation & Control Systems	–	–	1354	1.6	806	1.8
35	Mathematics, Applied	–	–	–	–	2033	4.6
36	Statistics & Probability	–	–	1811	2.1	–	–
37	History	–	–	1348	1.6	-	–
38	Robotics	–	–	1291	1.5	-	–
39	Mathematics	–	–	–	–	846	1.9
40	Mathematics, Interdisciplinary Applications	–	–	–	–	807	1.8
41	Mathematical & Computational Biology	–	–	–	–	697	1.6
42	Medicine, General & Internal	–	–	–	–	690	1.6
43	Information Science & Library Science	–	–	–	–	672	1.5
44	Other categories	142636	40.7	12217	14.1	13868	31.0

unique subject categories are identified and the results of previous queries are restricted to these categories.

The results of the above queries contain articles with a variety of citations. Some have been cited more than 1000 times. In contrast, many articles have not yet received any citation. To balance this heterogeneity and to ensure repeatability of the paper, which is an indicator of the accuracy of its information [36], Total Citation (TC) is applied. The TC is one of the most corresponding indicators in bibliographic analyses [5,12,13,32]. Therefore, this paper has concentrated on the first quarter of the most cited articles. Because of the accumulative feature of TC, the papers which were published in recent years will have a lower portion of the results. To solve this problem, a quarter of the most cited articles published in the last 4 years are extracted separately for each year. For the last year, if the number of papers which have been cited is less than 25% of all papers, only the cited articles are considered. Finally, duplicated records in each domain are removed and the "Mng" dataset (73,840 documents) for the management domain, the "Ind" dataset (25,687 documents) for the industrial engineering domain, and the "IT" dataset (10,246 documents) for the IT domain are constructed. The schematic for data selection is shown in Figure 1.

The combination of bibliographic analysis and topic modeling is applied in this study to extract the topics and monitor the research evolution in these three domains. Topic models are probabilistic models that could observe the terms from a generative probabilistic process and the hidden topics are identified using posterior inference from a textual corpus [37]. Topic discovery, document clustering, information retrieval, and predicting influential research studies are just some examples of topic modeling applications [38]. In this respect, there are various algorithms for topic modeling. The first one was presented as Latent Semantic Analysis (LSA) [39] and, after one year, the Probabilistic Latent Semantic Analysis (PLSA) [40] was proposed. Latent Dirichlet Allocation (LDA) [41] was presented to solve the over fitting problem of PLSA and some derivatives of LDA, such as Hierarchical Latent Dirichlet Allocation (HLDA) [42], Correlated Topic Models (CTM) [43], and Relational Topic Modeling (RTM) [44], have been proposed in recent years to address some constraints in the LDA. In this paper, the LDA algorithm is relied upon because it provides more than a topic explanation for each paper. Additionally, the overlapping between topics that are extracted by LDA helps one to find the topic correlations, which is one of the main goals of this study. All analyses are undertaken using R software, which is one of the most powerful tools for data and text mining, statistical analysis, and Natural Language Processing (NLP). Figure 2 shows the steps of the methodology of this research.

4. Bibliographic analysis

Table 3 shows the top 10 countries based on the affiliation of publications on the selected datasets. In this case, if the authors of the same paper come from different countries, that paper is counted separately for each country. Therefore, the sum of the number of all countries is more than the total number of papers in the corpora.

As is shown in Table 3, USA, China, England, Canada, and Taiwan are the 5 top countries in the research of both management and industrial engineering domains. In the IT domain, China has taken first position, with the USA, Canada, Germany, and France following. An important note is that either America or China is present in about 50% of the highly cited papers and American researchers participate in 45% of the highly cited management papers.

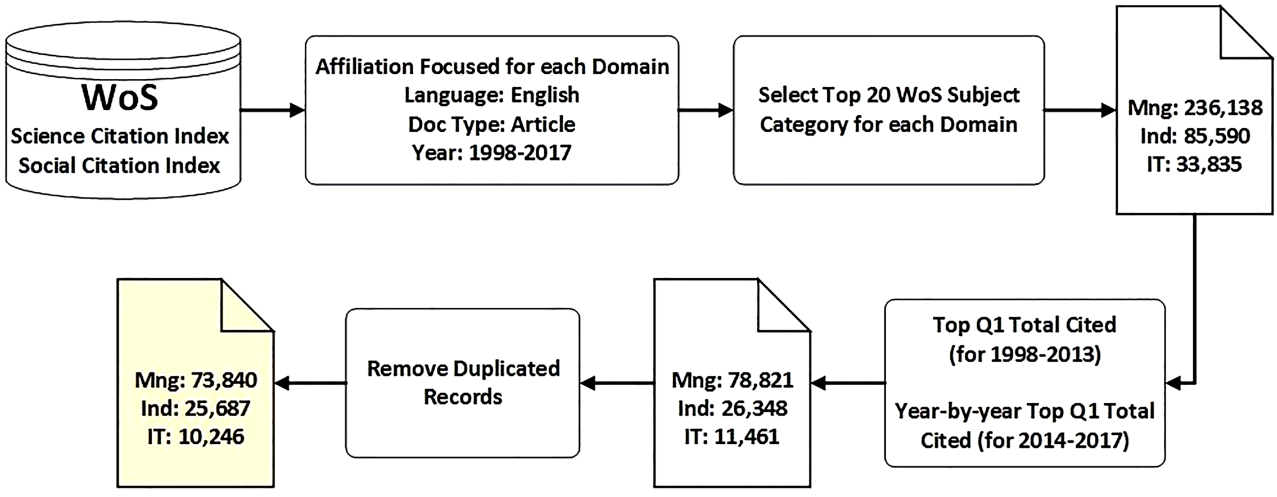

Figure 1. Schematic of dataset selection for each domain.

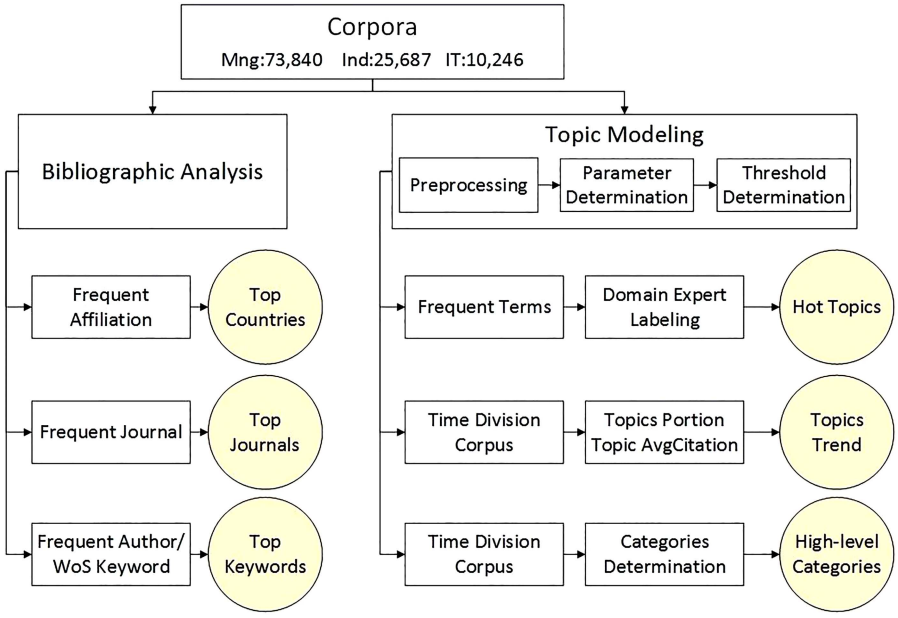

Figure 2. The steps of the research method.

Table 3. Top 10 countries based on the affiliation of publications in each domain.

Mng			Ind			IT		
Country	**#**	**%**	**Country**	**#**	**%**	**Country**	**#**	**%**
USA	33390	45	USA	7600	30	China	3100	30
China	10167	14	China	3173	12	USA	2864	28
England	8390	11	England	2524	10	Canada	1428	14
Canada	5654	8	Canada	1560	6	Germany	1068	10
Taiwan	5449	7	Taiwan	1354	5	France	1034	10
Germany	5109	7	France	1259	5	England	911	9
Italy	4838	7	South Korea	1090	4	Spain	543	5
France	4471	6	Australia	1064	4	Italy	484	5
South Korea	2637	4	Denmark	1014	4	Norway	442	4
Australia	2544	3	Germany	1012	4	Denmark	344	3
Others	42858	58.0	Others	13254	51.5	Others	7001	68.3

Table 4 lists the top 5 countries with the highest growth rate in the number of publications in each domain. This table reports the number of publications in four periods of time. The top countries are identified among those that have at least 1% of publications of a domain and their *Growth Rate* is maximized. The *Growth Rate* is defined as the amount of publication in the last period divided by the first periods' publications.

As seen in Table 4, China and France are identified as countries with the highest growth in all three domains. Spain, Italy, and Iran are also marked as high growth countries in management and industrial engineering science. A fact to consider is that Iran, as a developing country, is identified as having more than 54% and 13% growth in management and industrial engineering domains, respectively.

Analyzing the journal names of the datasets indicates that 3572 journals are present in the selected datasets. In this list, some journals such as *JAMA*, *Science, Nature, Strategic Management Journal*, and *New England Journal of Medicine* have the highest number of articles with more than 500 citations. Table 5 shows the top 10 journals based on the number of articles in each domain.

Moreover, there are 82 journals that contain articles in all three domains. Table 6 presents the top 5 journals with the most articles in all three domains.

WoS uses two fields to cover the document keywords. The first one is *DE* (Descriptor) that contains

Table 4. Top 5 countries with the highest growth rate.

Domain	Country	1998–2002	2003–2007	2008–2012	2013–2017	Growth rate	Percent of total publication
Mng	Iran	10	29	334	548	54.8	1.2
	Italy	115	259	854	1409	12.2	3.5
	Spain	81	151	612	949	11.7	2.4
	China	636	1158	3365	5008	7.8	13.7
	France	164	271	890	1219	7.4	3.4
Ind	Iran	18	69	260	241	13.3	2.2
	China	321	443	1091	1318	4.1	12.3
	Spain	65	166	307	204	3.1	2.8
	Italy	124	200	407	359	2.8	4.2
	France	154	197	375	338	2.1	4.1
IT	China	64	248	1037	1751	27.4	30.2
	Norway	9	28	124	100	11.1	2.5
	Hong Kong	23	37	125	239	10.4	4.1
	France	16	26	117	139	8.7	2.9
	India	16	22	74	135	8.4	2.4

Table 5. Top 10 journals based on the number of articles in each domain.

Mng		Ind		IT	
Journal	**#**	**Journal**	**#**	**Journal**	**#**
European journal of operational research	1434	Journal of materials processing technology	4492	PLOS one	215
Expert systems with applications	972	International journal of production economics	2174	Information Sciences	179
PLOS one	936	International journal of production research	1797	Journal of systems and software	173
International journal of production economics	887	Computers & operations research	1555	IEEE transactions on software engineering	144
Journal of applied psychology	763	Reliability engineering & system safety	1338	Expert systems with applications	141
Journal of cleaner production	735	Computers & industrial engineering	1210	Neurocomputing	141
Environmental science & technology	684	IEEE transactions on industrial informatics	778	Information and software technology	121
Management Science	661	Ergonomics	778	IEEE software	118
Journal of business ethics	624	CIRP annals-manufacturing technology	778	Pattern recognition	112
Tourism management	614	Safety Science	755	Knowledge-based systems	103

Table 6. Top 5 journals based on the most articles in all three domains.

No.	Journal	Mng	Ind	IT	Total
1	Journal of materials processing technology	83	4492	9	4584
2	International journal of production economics	887	2174	8	3069
3	International journal of production research	578	1797	11	2386
4	Computers & operations research	417	1555	18	1972
5	Computers & industrial engineering	385	1210	11	1606

the authors original keywords and the second one is *ID* (Identifier) or *Keyword Plus* that consists of words and phrases harvested from the title of the cited articles [45]. A comparative study of semantic similarities between keywords generated automatically from the database and controlled vocabulary (e.g. thesaurus) illustrates that both usually cover the main concepts in an article [46]. To extract the most frequent keywords, author keyword (DE) and keyword plus (ID) are analyzed separately. In the analyzing process, it is found that some keywords are presented in the dataset in both plural and singular forms (e.g. Algorithm and Algorithms). Therefore, to solve this problem, if the singular form of a keyword exists in the dataset, the plural form has been converted to the singular and its frequency changed to the sum of its plural and singular frequencies. The comparison between the author keywords (DE) and the keyword plus (ID) shows that ID is more generalized. The number of unique DE keywords in all three domains is at least 1.5 times more than the number of unique ID keywords (106,651 versus 70,510 for Mng, 39,904 versus 19,601 for Ind, and 22,850 versus 14,672 for IT). Consequently, the average amount of ID frequency is higher than DE (7.0 versus 2.6 for Mng, 5.7 versus 2.3 for Ind, and 2.9 versus 1.6 for IT). Therefore, the amount of the *Total* column in Table 7 is affected more by ID and the result of top keywords based on *Total* is almost same as the ID. Table 7 presents the most frequent keywords (DE and ID) in each domain.

As Table 7 shows, "*Model*", "*System*", "*Performance*", "*Scheduling*", "*Genetic algorithm*", and "*Management*" are common frequent keywords in all three domains. Moreover, there are some keywords that are frequent in two domains. "*Supply Chain Management" (SCM), "Innovation", "Impact"*, and "*Behavior*" are frequent in the management and industrial engineering domains. Similarly, "*Design*" is common in IT and industrial engineering and "*Information*" is seen in both IT and management. A note to consider is that these common terms may have a different meaning or denote different concepts in different domains. For example, the "*Genetic algorithm*" is widely used as an applicable tool and a solution approach in various areas of production and operation management [47], and it is also used in a wide range of optimization problems in industrial engineering [48]. It is, however, used in IT and computer science as an important concept and fundamental part of artificial intelligence and evolutionary algorithms [49]. In the same way, the term "*Scheduling*" in management and industrial engineering denote the time management in job scheduling or the production line. However, in the IT domain, it is often referred to the CPU or memory allocation as a basic task of the operating system. Another consideration is the inconsistency of the term "*China*" as the name of a country with other keywords in Mng. As mentioned before, "ID" is a descriptor term assigned by the author to the article. Therefore, it is not unpredictable for Chinese authors to use "*China*" as a descriptor in less than 10% of their publications.

Because of the close relationship between the management and industrial engineering disciplines, many keywords are common in these domains. But some terms such as "Cloud computing", "Wireless sensor networks" and "Security" are identified as frequent keywords, especially for the IT domain.

5. Topic modeling

5.1. Preprocessing

In order to cluster the documents into appropriate groups (topics), each document is represented by a list of conjunct terms (bag-of-words). The quality of this Bag-of-Words (BoWs) and the process of selecting the representative terms are very important to the result of clustering [38]. To generate a high quality BoWs, a method called term clumping is used. This method includes three steps: a) *fields selection*, b) *appropriate term extraction*, and c) *terms cleaning* [50].

In the first step of term clumping, the appropriate fields are selected as a source of term extraction. In the scientometric studies, based on the research aims and scopes, sometimes the descriptive fields (Authors' Organization, Department, Journal name, Conference name,...) or content fields (Title, Abstract, Keywords, Results,...) or a combination of them are chosen as the source fields. The second step is the term extraction process. In this phase, the terms or phrases are extracted from the texts. After the first and second steps of term clumping, a large number of terms are extracted. Therefore, in the third step, many of stepwise methods are prescribed for terms cleaning and

Table 7. Top 10 keywords based on the DE and ID in each domain.

Domain	Based on DE frequencies				Based on ID frequencies			
	Keyword	DE	ID	Total	Keyword	DE	ID	Total
Mng	China	860	649	1509	Model	128	7936	8064
	Supply chain management	798	195	993	Performance	468	6200	6668
	Innovation	651	1866	2517	Management	192	5138	5330
	Climate change	589	1	590	System	30	4052	4082
	Genetic algorithm	518	536	1054	Impact	22	3266	3288
	Performance	468	6200	6668	Behavior	44	2961	3005
	Trust	432	728	1160	Perspective	7	2774	2781
	Sustainability	428	309	737	Firm	12	2334	2346
	Scheduling	416	0	416	Information	51	2264	2315
	Uncertainty	368	1012	1380	Organization	78	2170	2248
	Total	5528	11496	17024	Total	1032	39095	40127
Ind	Supply chain management	678	372	1050	Model	43	2735	2778
	Genetic algorithm	509	509	1018	System	14	2216	2230
	Scheduling	494	0	494	Performance	142	1911	2053
	Simulation	377	487	864	Design	90	1478	1568
	Optimization	314	845	1159	Management	39	1390	1429
	Heuristic	311	0	311	Algorithm	36	979	1015
	Microstructure	240	225	465	Optimization	314	845	1159
	Reliability	233	255	488	Impact	12	730	742
	Innovation	231	533	764	Industry	7	597	604
	Inventory	223	153	376	Behavior	6	590	596
	Total	3610	3379	6989	Total	703	13471	14174
IT	Machine learning	126	0	126	System	0	771	771
	Cloud computing	117	0	117	Model	18	663	681
	Wireless sensor net	110	59	169	Algorithm	62	577	639
	Data mining	106	0	106	Design	50	363	413
	Classification	81	299	380	Classification	81	299	380
	Genetic algorithm	81	110	191	Network	15	282	297
	Security	76	74	150	Information	2	263	265
	Scheduling	70	0	70	Management	23	258	281
	Support vector machine	68	70	138	Performance	50	226	276
	Feature selection	65	0	65	Framework	6	214	220
	Total	900	612	1512	Total	307	3916	4223

reducing the huge dataset to a more meaningful and user-friendly dataset. These methods are: applying a thesaurus for removing common terms (stop words, non-alphabetic characters, and meaningless terms), using NLP techniques to combine terms with a similar structure (e.g. stemming terms, eliminating plural forms of the words, and coordinating different spelling), combining (removing the top general terms, replacing similar terms with shorter ones, and term clustering), pruning (eliminating the terms which appear in a single record), screening (e.g. using Term Frequency (TF) as a weighting method for screening the common and unimportant terms), and finally, the clustering step (using dimension-reduction tools such as PCA (Principal Component Analysis) or term relevancy detection models such as Topic Modeling to dramatically reduce the number of terms) [50].

In this study, in the term clumping phase, to

construct an appropriate BoW for each document, the focus was on keywords (DE and ID). The challenge is the managing of missing values. In about 20% of records (24% of Mng, 19% of Ind and 25% of IT) the DE field is null, and in about 15% of records (7% of Mng, 17% of Ind and 20% of IT) the ID has no value. Choosing a combination of ID and DE creates the better condition, but the result is not satisfactory (2% of Mng, 3% of Ind and 7% of IT still are missing). To resolve this problem, a keyword extraction method is required to extract BoW from the title of articles. There are several automated or semi-automated ways proposed for keyword extraction. These ways are complicated and some are based on expert knowledge or require domain specific thesauri or ontologies [51,52], and some need a large corpus of the related text to learn the semantic of terms [53,54]. To avoid these complexities, a simple heuristic key extraction function is developed in this study. Because of the length and nature of the title and keywords, the terms that occur in these fields are more valuable than those seen in the abstract. Therefore, this function constructs a BoW set for each document from its keywords or title. The extracted terms should exist in a predefined dictionary (Dic). This dictionary consists of all the DE and ID which are edited (e.g. eliminating punctuation, converting plural terms to singular and so on) to become more valuable. To extract the keywords from the title, the biggest adjacent sequence of words (n-gram) that existed in the Dic is considered. An n-gram is a contiguous sequence of n items from a text or speech. The idea comes from the fact that the greater the number of components of a phrase (n), the more important it is [55]. In this approach, a term is located in the BoWs of a record if it exists in the ID or the DE keywords. In cases where both ID and DE are empty, the $Term$ is selected as BoW if it is a part of the document's title and exists in the Dic and there is not another term ($Term'$) with the above conditions and the biggest n in its n-gram. The BoWs for record i is defined formally by Eq. (1) as shown in Box I. After using the above key extraction function and making a BoWs for each record, almost every record of all domains are covered by at least one keyword (99.99% of Mng, 99.99% of Ind and 99.94% of IT).

5.2. The input parameters for LDA

The first input object of LDA is a document-term-matrix that contains the TF value of each term for each document that describes the frequency of terms that occur for a document in the corpus. In this matrix, rows correspond to documents (papers) and columns denote terms. To avoid using scarce terms, the terms which have been seen more than 5 times in the BoWs set are selected.

The second variable has to be set in an LDA algorithm, which is the number of topics. There are different solutions to address this problem [37,56]. Choosing a high number of topics leads to covering all the themes and research areas of a domain. On the other hand, selecting a limited number of topics can be more easily understood and interpreted by domain experts [17]. In order to choose this parameter, the perplexity measure has been introduced [57]. The perplexity is a measurement of how well a probability model predicts a sample set. It can be applied to compare probability models and a lower perplexity indicates that the distribution is better at predicting the sample. Perplexity is a measure for the quality evaluation of the model. A lower perplexity score indicates better generalization performance [41] and higher values of perplexity in LDA indicate a higher misrepresentation of the terms of test documents by the trained topics [17]. In this paper, the LDA algorithm is run on the whole collection of papers in all three domains with a variable number of topics, ranging from 2 to 200. This process is highly CPU bounded. Therefore, the parallel solution [58] is applied in R software to use multicore functionality. Subsequently, the perplexity values of each execution are collected. The perplexity plot is presented in Figure 3 for each domain.

Figure 3 shows that the perplexity dramatically

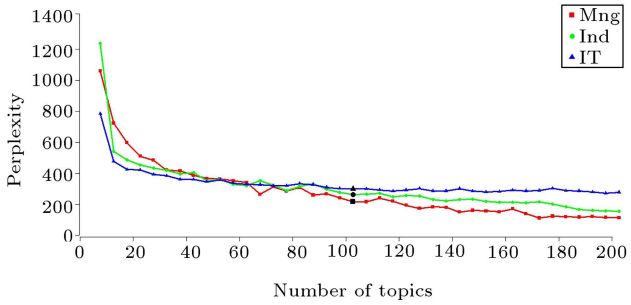

Figure 3. Perplexity plot for various numbers of topics in each domain.

$$Bow_i = \left\{ \begin{array}{l} DE_i \cup ID_i \\ \left\{ \begin{array}{c} Term | Term \in n\_gram_i \wedge Term \in Dic \wedge \\ (\nexists Term' | Term' \in N\_gram_i \wedge Term' \in Dic \wedge N > n) \end{array} \right\} \end{array} \right. \begin{array}{l} if\ DE_i \cup ID_i \neq \emptyset \\ O.W \end{array} \quad (1)$$

Box I

decreases from 5 topics to 20 topics. Then, the decrement continues slowly from 20 to 100 topics and becomes more or less stable for values higher than 100. Since it is preferred to have a low level of perplexity; on the other hand, keeping the number of topics as low as possible, 100 was chosen for the number of topics in all three domains.

5.3. Explanation threshold

Each topic returned by the LDA is associated with each paper with a specific amount of relevancy. In other words, paper p is explained by topic t with the explanation level of $\rho(p,t)$. In this approach, a threshold needs to be set for the explanation level to determine which papers are explained by a specific topic. If this threshold is set at a low amount level, the topics can explain more papers and the coverage of the corpus become acceptable; on the other hand, many papers are explained by a topic with a low level of explanation relevancies. Therefore, the topics and papers will be connected loosely to each other. Applying the explanation threshold, it can be said that p is explained by topic t if $\rho(p,t) \geq th_e$. To determine the appropriate th_e two parameters are addressed. These two parameters are *corpus coverage* and *explanation quality* [17]. The corpus coverage is the fraction of papers in the corpus that is explained by at least one extracted topic. Therefore, for the corpus P, extracted topics set T and explanation threshold th_e, the explained set E contains all explained papers and the relevance set R is defined as all couples of (p_i, t_j) that p_i is explained by t_j. More formally:

$$E = \Big\{ p_i | i \in \{1,2,3,\ldots,|P|\}, \exists t_j, j \in \{1,2,3,\ldots,|T|\} \quad \rho(p_i,t_j) \geq th_e \Big\},$$

$$R = \Big\{ (i,j) | i \in \{1,2,3,\ldots,|P|\}, j \in \{1,2,3,\ldots,|T|\}, \quad \rho(p_i,t_j) \geq th_e \Big\}. \tag{2}$$

And $C_{P,T}^{th_e}$, which is the coverage of corpus P, extracted topics set T and explanation threshold th_e, is defined as:

$$C_{P,T}^{th_e} = \frac{|E|}{|P|}. \tag{3}$$

Consequently, explanation quality $EQ_{P,T}^{th_e}$, which is the average relevance level of papers that are explained by at least one topic, is defined as:

$$EQ_{P,T}^{th_e} = \frac{\sum_{(i,j)\in R} \rho(p_i,t_j)}{|R|}. \tag{4}$$

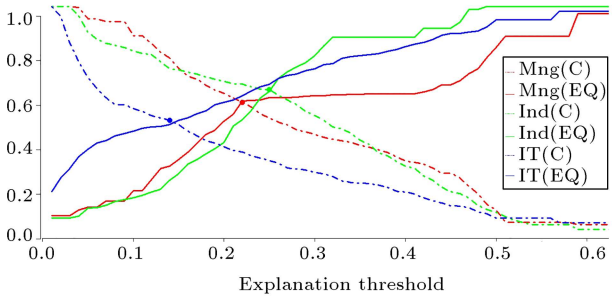

Figure 4. Corpus coverage (C) and Explanation Quality (EQ) of domains for different explanation threshold.

In the above equation, the numerator is the sum of all relevancies between all topics and papers that are higher than or equal to the explanation threshold. And the denominator is the count of these relevancies between papers and topics that pass the explanation threshold. The aim is to find an appropriate explanation threshold that can maximize both the corpus coverage and the explanation quality. To do this, the LDA algorithm was applied with $k = 100$ topics on the document-term-matrix of each domain. To apply LDA to the term-document-matrix, some papers which have no terms in their BoWs that appear more than 5 times in the term dictionary are ignored. Therefore, the total numbers of papers in these matrices are lower than the total number of papers in the corpora. Figure 4 presents the corpus coverage (C) and the Explanation Quality (EQ) for each domain. As illustrated in Figure 4, to determine the appropriate th_e, the points that maximize both the corpus coverage and the explanation quality are selected for each domain. In this way, $th_e = 22$, which covers 57% of the Mng corpus (41,520 papers), is selected for the management domain. In the same way, for industrial engineering, $th_e = 0.25$ covers 63% of the Ind corpus (15,656 papers), and for the IT domain, $th_e = 0.14$ covers 49% of its corpus (4,576 papers).

6. Results and discussion

6.1. Hot-important topics

As reported in the previous section, the LDA algorithm is applied to each domain to extract 100 topics. To distinguish some topics and to explain in more detail, ten hot-important topics were chosen in each domain. Because all corpora in this research contain 100 topics, an important topic is defined as a topic which covers more than 1% of its corpus. It can be assumed that studies receiving more than an average number of citations have high importance and representativeness [5,13]. Klavans and Boyack [59] reported that [12] identified the hot fields based on the number of citations and from 1997–2007, new thresholds were determined to identify the hot, warm and cold topics based on

the number of citations. Therefore, in this study, the hot-important topics are assumed as important topics whose average amount of citations is maximized. Hence, topics will be analyzed at separate periods of time; this assumption also being applicable for new phenomena in these domains. In the other words, $E_{P,t}$ is defined as the set of explained papers of corpus P by topic t, $I_{P,T}$ is the important topics of corpus P, $AvgCite_{P,t}$ is the average citation of papers in topic. Finally, $HI_{P,k}$ contains k number of important topics with the maximum amount of average citation. More formally:

$$E_{P,t}=\{p_i|i\in\{1,2,3,\ldots,|P|\},\rho(p_i,t)\geq th_e\},$$

$$I_{P,T}=\{t_i|i\in\{1,2,3,\ldots,|T|\},|E_{P,t_i}|\geq(|P|/|T|)\},$$

$$AvgCite_{P,t}=\frac{\sum_{p_i\in E_{P,t}}Citation(p_i)}{|E_{P,t}|}. \quad (5)$$

Now, suppose that $f: I\rightarrow I$ is a set function as follows:

$$f(I)=I\backslash\{argmax_{t\in T}AvgCite_{P,t}\},$$

$$f^k(I)=\underbrace{fofo\cdots of}_{k}(I).$$

Therefore:

$$HI_{P,k}=I\backslash f^k(I).$$

Based on the above measurement, ten hot-important topics ($HI_{P,10}$) are reported in Table 8.

In Table 8, topics are labeled by domain experts to make them meaningful and understandable for more analysis. The experts choose the labels by focusing on the related terms of each topic. The most relevant terms of a topic enable the researcher to identify the content of that and manually assign it a label [19]. Nine experts (three in each domain) were chosen from the

Table 8. Ten hot-important topics in each domain.

	Topics	#	Most relevant terms
Mng	T38: Cooperation & partnership	1068	Trust, cooperation, strategic alliance, outsourcing, joint venture
	T26: Technology management	929	Adoption, user acceptance, planned behavior, acceptance
	T11: HR	875	Personality, job performance, individual differences, 5-factor model
	T6: Health management	1151	Intervention, cancer, quality of care, blood pressure, self-management
	T43: Environmental health	1269	Mortality, risk factor, smoking, environmental factor
	T20: Financial management	794	Investment, acquisition, profitability, finance, market value
	T97: Stock market	1360	Stock returns, volatility, liquidity, stock price, fluctuations
	T10: Resource & capacity	1126	Competitive advantage, resource-based view, absorptive capacity
	T99: Occupational health	846	Prevalence, prevention, depression, mental health, telemedicine
	T94: Job performance	1441	Job satisfaction, social support, procedural justice, employee turnover
Ind	T25: R&D	204	R&D, Development projects, development cycle time, Japanese
	T9: Manufacturing	373	Machining, surface integrity, tyool life, chip formation, hard turning
	T94: MCDM	402	Analytic hierarchy process, supplier selection, criteria, TOPSIS
	T17: Information systems	191	Information system, enterprise system, Internet of Things (IOT), Web
	T54: HR (sociology)	290	Gender, earnings, inequality, sex differences, discrimination
	T59: HR (psychology)	211	Job satisfaction, ownership, Korea, performance appraisal
	T83: Safety management	196	Safety climate, safety culture, safety management, accident prevention
	T5 : Innovation	210	Product development, creativity, radical innovation, innovation
	T97: Environmental Mng	241	Green, environmental management, environmental performance
	T18: CRM	189	Behavior, perception, personality, loyalty, planned behavior
IT	T18: Health technology	65	Clinical practice guideline, systematic reviews, health services research
	T31: Machine learning	73	Machine learning, neural network, ensemble learning
	T72: Health care	61	Quality of life, care, public health, evidence-based medicine, doctors
	T26: Data mining	40	Data mining, classification, clustering, boosting, pattern discovery
	T30: Computational biology	47	Discovery, microarray data, human genome, RNA, ant colony
	T8: General medicine	103	Therapy, prevention, glycemic control, diabetes, insulin
	T3: Cloud computing	58	Cloud computing, Web service, Service composition, Virtualization
	T29: Occupational health	113	Risk factors, coronary heart disease, blood pressure, mental health
	T100: Internal medicine	49	Death, Breast cancer, Surgery, tumor, pregnancy
	T61: Health services	52	Care, intervention, clinical trial, engagement, empirical evidence

faculty members of management, industrial engineering and IT departments. The topics of each domain were assigned independently to two related experts. For each topic, a list of the first 20 more related terms was presented to the domain experts. The rank of the terms is important for experts to choose the appropriate label. Therefore, sometimes two topics with similar related terms get different labels. Moreover, the similarity between topics, which is calculated based on the *Cosine measure* [60], informed experts about the relation between the topics and helped them to determine those labels that could explain the topics in the best way. In this process, if both domain experts suggest a similar label for a topic (42 topics for Mng, 38 topics for Ind and 47 topics for IT), that label is chosen for that topic. But in cases where labels are different, the third expert chooses one of the suggested labels. For two topics of Mng, three topics of Ind and five topics of IT, the third expert could not make a certain decision. For these topics, the appropriate label is assigned after the aggregation session that is held by all three domain experts.

An important point in Table 8 is the role of "*Health Science*" in the IT and management domains. There are two important reasons for this strong participation. The first reason is that many studies [61,62] show that the use of Electronic Medical Records (EMRs) by physicians has increased over the past decade and Health Information Technology (HIT) has been widely used in medical centers since 2000. Most published HIT implementation studies report positive effects on the quality, safety, and efficiency of healthcare [25]. For example in the IT domain, keywords such as "Classification", "Feature Selection", "Genetic Algorithm", "Support Vector Machine", etc., which could extract pure machine learning or data mining publications in the older publications of the used dataset, led to the extraction of health related topics after 2008. It means that machine learning or data mining have shifted from a technical concept in computer science to applicable tools and approaches in other domains (especially in the health domain). A noted fact is that health related topics in the IT domain are extracted via computer science based keywords (such as Support Vector Machine (SVM), Clustering, Sensor, Information System ...), but the most relevant terms of these topics in the above table are obtained after the process of topic extraction.

The second reason is that based on Eq. (5), citation is a critical factor to choose the hot important topics of each domain. As known, publications of the health domain often have high citations in comparison to the computer science domain. Therefore, 6 hot important topics of the IT domain in the above table are associated with the health domain.

Table 8 indicates that these three domains have some similar or common topics. For example, *human resources (HR)* is common between industrial engineering and management. Some concepts such as *work improvement, occupational safety*, and *health management* are mixed together, and are mentioned in both domains. Some studies illustrated that there has been a growing trend in occupational health research, risk management and occupational safety over the last 20 years [63,64].

6.2. Comparing topics trend

To compare topic trends, the corpus of each domain is divided into four periods of time. The total corpus contains the papers of the last 20 years (from 1998 to 2017). Therefore, each divided corpus contains five years of publications. Analyzing the research trends based on five year periods is prevalent [65]. After creating the time-based corpora, a document-term-matrix is constructed for each corpus and the topic distributions for papers and the term distributions for topics are obtained for all corpora based on the previous LDA results of total papers in each domain. Then, $Portion_{P,t}$ is defined for each topic in a corpus P as the number of papers explained by t divided by the size of P as follows:

$$Portion_{P,t}=\frac{|E_{P,t}|}{|P|}. \tag{6}$$

The properties of each divided corpus are reported in Table 9. It is notable that the Average citation of a divided corpus in Table 9 refers to the average amount of all citations of that corpus and does not refer to that part of citations of papers cited in that Time period. Because of the increment feature of citation, Table 9 indicates that the average citation for older corpus is higher than the average citations of recent publications. This limitation of citation analysis motivates researchers to work on the concept called the "*Time window*" [66,67] for research evaluation based on citation criteria.

Figures 5–7 show the trends of hot-important topics mentioned in Table 8 separately. In these figures, each circle represents a topic in a certain time period. The size of the circles indicates the portion of the topic (number of papers explained by topics in that period of time divided by all papers of the corpus in that period) that shows the share of a topic from the whole corpus in a specific period of time. The number in each circle is the average citation of papers in the related topic. Based on changes in the size of the circles, each topic could be marked as an increment, fixed or decrement topic. An important note is that the size of each circle only, which shows the portion of papers on that topic in the corpus, could not determine the trend of a topic, and the changes in its average citation is also important.

Table 9. Properties of divided corpus for each domain.

	Divided corpus	Time period	# of papers	Portion	AvgCite
Mng	Mng-Q1	1998 – 2002	10,924	15%	76.5
	Mng-Q2	2003 – 2007	17,855	24%	62.3
	Mng-Q3	2008 – 2012	21,888	30%	44.9
	Mng-Q4	2013–2017	23,176	31%	13.8
	Total	1998 – 2017	73,840	100%	44.0
Ind	Ind-Q1	1998 – 2002	5,781	23%	45.4
	Ind-Q2	2003 – 2007	7,507	29%	41.5
	Ind-Q3	2008 – 2012	6,777	26%	34.0
	Ind-Q4	2013–2017	5,622	22%	11.8
	Total	1998 – 2017	25,687	100%	33.9
IT	IT-Q1	1998 – 2002	1,195	12%	45.3
	IT-Q2	2003 – 2007	2,020	20%	39.5
	IT-Q3	2008 – 2012	3,335	32%	36.7
	IT-Q4	2013–2017	3,696	36%	14.8
	Total	1998 – 2017	10246	100%	30.3

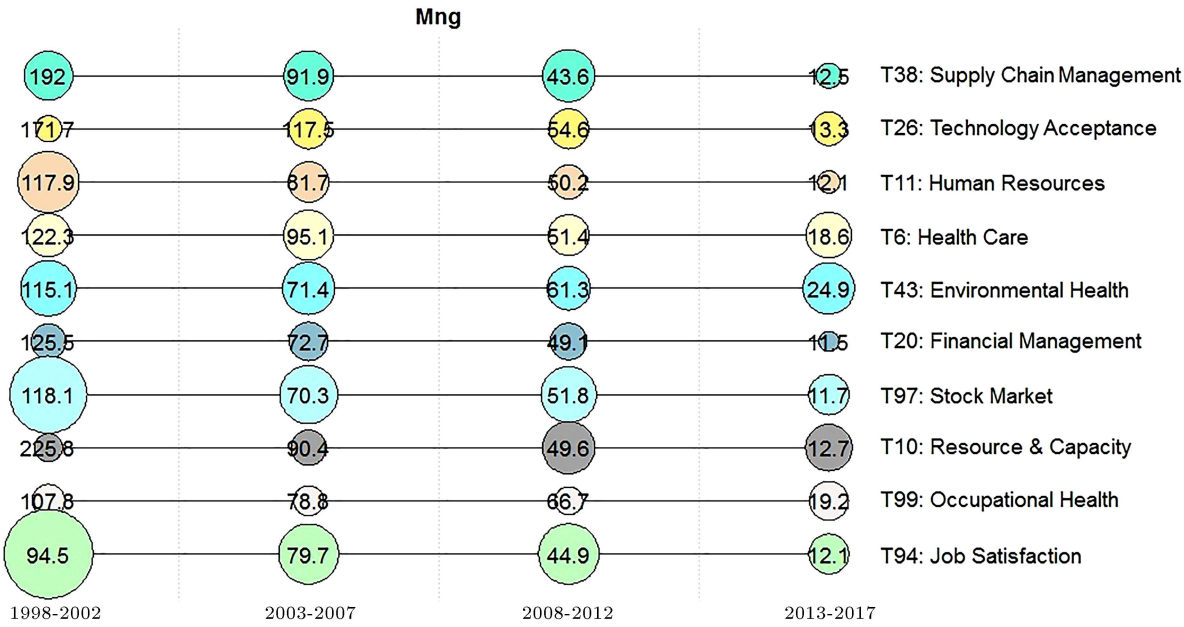

Figure 5. Trends of ten hot-important topics in management.

As Figure 5 shows, in the management domain, some topics such as HR, financial management, *Stock Market* and *Job Performance* (as a branch of HR) are identified as decrement topics. Nevertheless, the high amount of average citation of these topics is reliable evidence that this reduction does not mean a decrease in the importance of these topics. In other words, these topics play a fundamental role in the management domain. In contrast, the health-related topics have a considerable share in management research and *Environmental, Occupational Health*, and *Resource & Capacity* are marked as increment topics in the management domain. Although *Technology Management* has the highest amount of citation, it could be identified as a fixed topic based on its portion.

Cooperation and *Partnership* is a major research interest in the management domain [14], and different areas of partnership, especially between public and

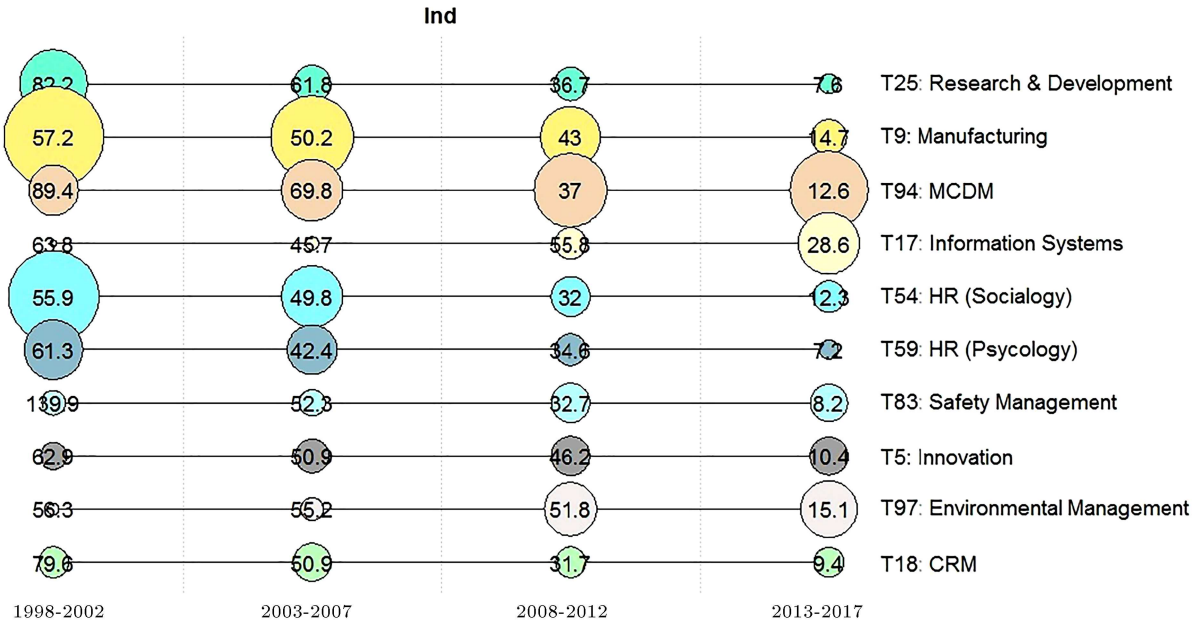

Figure 6. Trends of ten hot-important topics in industrial engineering.

private sectors, have been explored and investigated since the late 1990s. Previous research reported that the total number of papers published from 1990 to 2013 increased; the rapid increase in publication began in 2010 and then steadily progressed in 2013 [14]. Moreover, the partnership is an important research content of the SCM and in the face of the complex relations of cooperation, risk management and establishing joint venture strategies are highlighted in recent studies on management [68]. Figure 5 shows the importance of this topic.

Focusing on the trends of topics in Figure 6 illustrates that industrial engineering studies concentrated on *Manufacturing* and *Human Resources (HR)* topics before 2007. However, this paradigm changed and today these topics are identified as decrement topics, because *Information systems, Multiple Criteria Decision Making (MCDM), Safety Management*, and *Environmental Management* took their place in recent papers and are being tagged as increment topics. Traditional manufacturing studies have turned to new concepts when confronting *IT* and *Environmental Science* [5]. As evident, *predictive manufacturing systems* [69] and *cloud manufacturing* [3,70] that are founded based on *big data analysis*, or *green manufacturing* [8] and *sustainable manufacturing* [71] that are related to *environmental science* are new interdisciplinary subject areas of manufacturing.

Jin et al. [5] in their research about new issues in industrial productivity over the past 44 years, have emphasized that concern about *cleaner production (CP)* and *green technology* has been increased by reinforcement of environment-related regulations, and *Computer Integrated Manufacturing (CIM), intelligent manufacturing systems* and *Flexible Manufacturing Systems (FMS)* are applied widely in this area. They also found that after 2009, industrial engineering has highlighted innovations, such as those involved in the development and use of green IT and environmental protection related studies.

A considerable fact highlighted in Figure 6 is the increasing usage of *Decision Making* approaches in industrial engineering. The MCDM is applied widely in the energy fuels, Operation Research (OR), management and environmental science and its usage in research papers has risen since 2006 [27]. Supplier selection, as an important decision in SCM, is an important usage of MCDM in industrial engineering [72].

The late 1980s included studies about the utilization of computers and IT for automation in manufacturing and operational research. But, the period between 2000 and 2013 was marked by studies related to the connection between manufacturing, decision making and IT in this domain [5]. IT alignment and IT advancement affect supply chain capabilities positively [73]. Therefore IT is used as an enabler, tool, approach, and infrastructure in all elements of the supply chain and sometimes it switches the traditional paradigm of industrial engineering to a new paradigm that is based on information-oriented systems [74].

Finally, *R&D* is identified as a decrement topic but *Innovation* that is an output of *R&D* activities is marked as an increment topic. And *Customer Relationship Management (CRM), since* its growth from 1998 to 2007, could be tagged as a fixed topic.

As Figure 7 presents, a considerable part of hot-important topics in the IT domain is allocated for

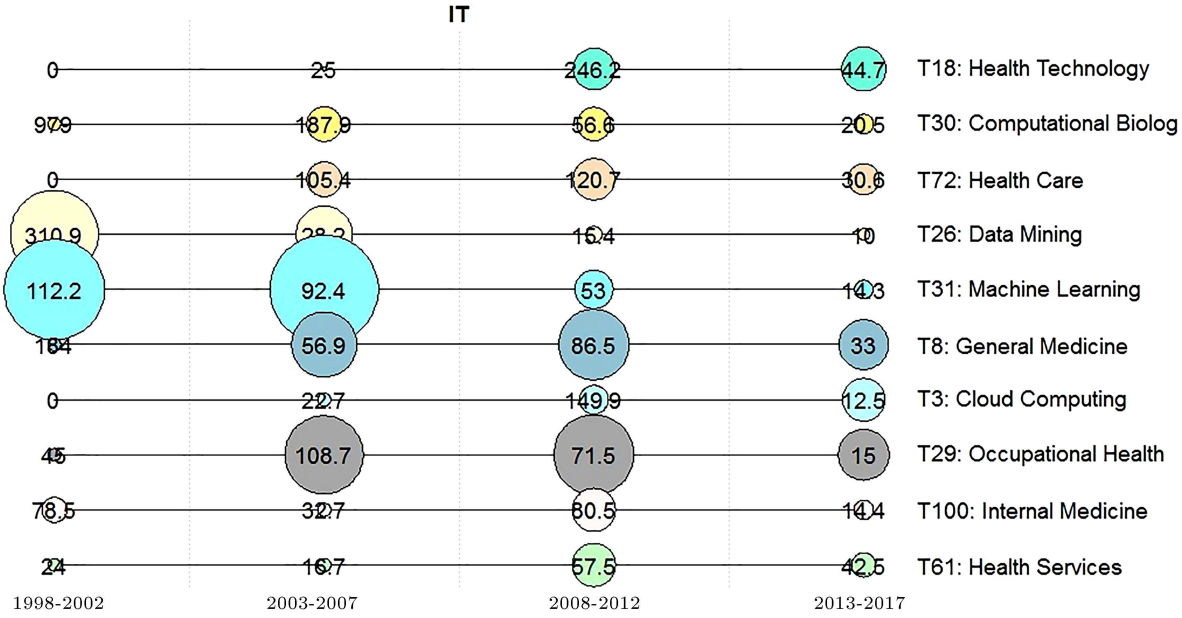

Figure 7. Trends of ten hot-important topics in IT.

Health. This figure indicates that publications of this domain have been shifting from pure computational topics (such as data mining and machine learning) to application of computer science in other domains (especially in the health domain, for example) [75–77]. Moreover, In the last years, individuals working at the intersection of IT and medicine have developed and computer applications to improve health care services have increased [16]. Moreover, a growing body of scientific evidence supports the use of this technology in the clinical decision-making process [25,34,61]. On the other hand, the movement of health care to patient-centered care has empowered patients and changed the role of IT in healthcare [78]. The development of new technology such as Web 2.0 and 3.0 and the rapid growth of social media provide new opportunities for individuals to access and exchange health information [78].

Jones et al. [25] in research about the investigation of HIT studies between 1995 and 2013 found that the number of published health IT evaluation studies is increasing rapidly. This study explored that such areas in HIT increased by approximately 13% per year before 2007, and roughly 25% per year from 2008 to 2012. Approximately, the same result is figured in health-related topics in Figure 7.

Although Figure 7 shows that the portions of *Data Mining* and *Machine Learning* topics become smaller in previous periods, the high amount of average citations of these topics indicates that data mining and machine learning are applied widely as useful approaches or tools in the other topics or domains. For example, life sciences, biomedicine, computational biology, and health care are increasingly turning into a data-intensive science. Therefore, data mining and consequently machine learning should be used to discover knowledge from the huge amount of generated data [79]. Moreover, some new terms have been merged with basic concepts and new topics or concepts have been constructed. For example, *Big Data Mining* is introduced instead of traditional data mining [80].

Cloud Computing as a new phenomenon that has joined with other concepts such as *Big Data*, *Internet of Things (IoT)*, and social media is rising in recent studies and the huge number of publications in this area of research is undertaken by interdisciplinary researchers [4]. Cloud computing is one of the most significant shifts in ICT and service that eliminates the need for expensive hardware and dedicated space [81].

In conclusion, *Health-related* topics and *Cloud Computing* are known as increment topics; conversely, *Machine Learning* and *Data Mining* with considerable amounts of citation are marked as decrement topics and *Computational Biology* could be tagged as a fixed topic.

6.3. Comparing high-level categories

To compare all three domains in a single perspective, it is necessary to assign a high-level category to the topics. To do this, the field "SC" that denotes the research areas of WoS is applied. The high-level categories for each topic are identified by counting the SC content of the documents which are explained by that topic. To make these high-level categories comparable among different domains, some applicable criteria are required. Previous criteria are usually used to describe the trends, importance, density or evolution of topics in a domain. This study endeavours to describe these,

in their correlated domains. Each domain has its specific characteristics. For example, Table 8 indicates that average citation of the first time period for the Management domain is more than 75 instead of about 45 for Industrial Engineering and IT. Therefore, using "Citation" as a common measure for topic evaluation is not appropriate, especially for evaluation of a category among two or three domains. In this approach, to respect the TC of a paper, this attribute is used as a factor of counting, and to normalize the different amounts of TC in each domain and time period, the TC is divided into its related $AvgCite_P$, as mentioned in Table 9. WC_P is defined as all research areas in the corpus P. Based on this approach, $Fr_p\,(cat)$, as the frequency of a research area category for paper p and $Rel_{P,t}\,(cat)$ for topic t in corpus P, are proposed formally as:

$$Fr_p\,(cat) = \begin{cases} \frac{TC}{AvgCite_P} & \text{if cat } \in WC_P \\ 0 & \text{if cat } \notin WC_P \end{cases}$$

$$Re\ l_{p,t}(cat) = \frac{\sum\limits_{pi \in E_{p,t}} Fr_{pi}(cat)}{\sum\limits_{cat \in WC_P} \sum\limits_{pi \in E_{P,t}} Fr_{pi}(cat)} \tag{7}$$

Therefore, $Cat1_{P,t}$ and $Cat2_{P,t}$ as the first and second high-level categories for topic t in corpus P are identified as:

$$Cat1_{P,t} = argmax_{cat \in WC_P} Rel_{P,t}\,(cat)$$

$$Cat2_{P,t} = argmax_{cat \in (WC_P \setminus Cat1(P,t))} Rel_{P,t}\,(cat),$$

After that, for topic t, the relevance of a high-level category $HlRel_{P,t}$ is assumed as its $Rel_{P,t}\,(cat)$ when the cat is marked as $Cat1_{P,t}$ or $Cat2_{P,t}$. Consequently, $HlRel_p\,(cat)$ that is the relevance of a high-level category for corpus P is proposed as:

$$HlRel_{P,t}\,(cat) = \begin{cases} Rel_{P,t}\,(cat) & cat \in \{Cat1_{P,t}, Cat2_{P,t}\} \\ 0 & O.W \end{cases}$$

$$HlRel_P(cat) = \sum_{t \in T} HlRel_{P,t}(cat) \tag{8}$$

Finally, the high-level categories with the maximumamount of relevancy are extracted for each domain in separate periods. The result is depicted in Figure 8. In this figure, to make the high-level categories comparable to three domains and four periods, the square root of $HlRel_P\,(cat)$ is used as the ratio of the circles and the biggest circles are drawn backwards.

In the above approach, to simplify the results, "*Basic Science*" is applied instead of *Physics, Chemistry, Biology*, and *Mathematics*. In some cases, a high-level category of a corpus is only seen in the second high-level category of topics ($Cat2_{P,t}$). For example, *Sociology* and *Automation* are extracted from the second level categories of industrial engineering topics. Extending the relevancy to the third and fourth levels has not changed the results considerably.

As Figure 8 shows, *Telecommunication* is identified exclusively as a high-level category of the IT

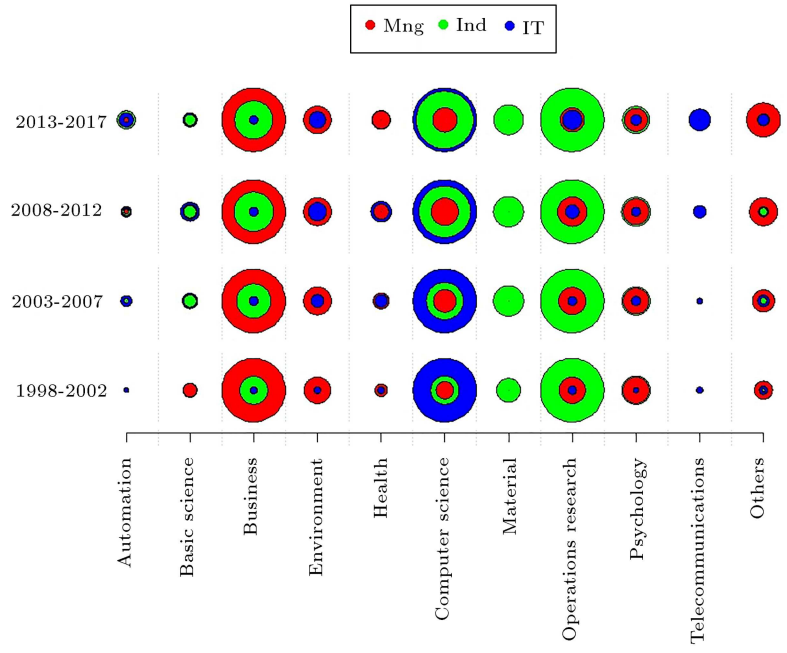

Figure 8. The high-level categories for each domain in the separate periods.

domain and relevant research studies have increased significantly. *Material science* is also explored solely for industrial engineering with no serious change in its importance. In the *Automation* category, IT and industrial engineering have the more important role. The previous research [5] indicates this role as the utilization of computers and IT systems for automation in the manufacturing process. *Basic Science* is covered similarly by all three domains. *Biology* is addressed more by management and IT, whereas *Physics* and *Chemistry* are used widely in industrial engineering studies. Finally, *Mathematics* is applied in the research studies of these domains as an essential tool or theoretical approach.

Although the *Business* category is one of the core studies in management science, and previous research [6] indicates that it has high centrality and medium or high density in studies during the last 25 years for this domain, industrial engineering researchers are trying to undertake further research in this category. In the same way, *Operational Research* is marked as a fundamental research area of industrial engineering. However, management has a considerable share in this category and recently many studies in the IT domain have focused on this subject.

Computer Science as an essential field of IT is applied by management and industrial engineering. As mentioned before, industrial engineering researchers have applied computer science in many aspects of their research. Its usage is undeniable in many studies in the field of industrial engineering that are related to automation, problem-solving, decision making, simulation, evaluation, forecasting, planning, and information management [5].

In the *Environment* and *Health* category, management and IT have a more important role. In both of them, IT usage has a rapid increment and sometimes new paradigms are born in these areas. Research on the important role of social media and knowledge management systems in the enhancement of public knowledge regarding global environmental changes [82,83], and applying new IT-based tools for the prediction and monitoring of environmental research issues [28,82] are some examples of this increment in environmental studies in recent years. In the same way, the use of computer science in the research management of the health domain [18], applying a data mining approach in disease prediction [84–86] and the widespread adoption of healthcare information systems for health management [87] are clear evidence for the IT increment trend in the health field.

Lastly, in *Psychology* science, all three domains exist, but management and industrial engineering have a dominant presence. This category contains studies that are more about the psychological aspects of *human resource management* [88] or concepts about *technology acceptance, customer relationship management (CRM)*, and *Organizational Psychology* [89].

The results depicted in Figure 8 indicate that in the 9 categories, all three domains are present and some areas in a domain are common or very close to other domains. This fact confirms the strong relationships between these domains and verifies the efficacy of topics in these domains on each other.

7. Conclusions

The purpose of the current study was to illustrate the evolution of Information Technology (IT), management and industrial engineering studies as three academic and research domains. To achieve this purpose, the bibliographic data of 20 years (from 1998 to 2017) of publications from the Web of Science (WoS) repository in each domain were extracted. A new approach, which consists of a combination of bibliographic analysis and a topic modeling approach, was applied in this research to identify topic trends, determine the important topics and describe the research evolution in each domain. In the topic modeling process, a novel simple heuristic method was suggested for key extraction for each record. In addition, some new criteria were defined to compare the topics. Proposing the new method to determine the high-level category for each topic is another innovation of this study. The bibliographic analysis indicated that the USA, China, England, Canada and Taiwan have the most publications in these domains. USA researchers participate in 45% of the management studies and USA or China is present in about half of these three domain studies. In addition, China and France are marked as countries with high growth in all three domains and Iran is identified as a country with the maximum growth rate in management and industrial engineering research. Moreover, bibliographic analysis identified the journals with the maximum number of articles in each domain. In this process, 82 journals were discovered to contain articles in all three domains. The keyword analysis was the last step of bibliographic analysis for these corpora. This step reported that "*Model*", "*System*", "*performance*", "*Scheduling*", "*Management*", and "*Genetic algorithm*" are common frequent keywords in all three domains. In addition, "*Supply chain management*", "*Innovation*", "*Impact*", and "*Behavior*" were identified as common frequent keywords in the management and industrial engineering domains. In the same way, "*Design*" was marked as a common keyword in IT and industrial engineering and "*Information*" was recognized in both IT and management. On the other hand, some frequent keywords such as "*Cloud computing*" and "*Wireless sensor networks*" are specific to the IT domain.

To determine the Bag-of-Words (BoW) for each

record, the key extraction method constructed the BoW from the combination of keywords and the biggest n-gram of the title. After that, the Latent Drichlet Allocation (LDA) algorithm is used to extract the topics and the hot-important topics were extracted for each domain. The most obvious finding of this part is the greater importance of the role of "*Health Science*" in IT and management domains over the last 20 years. Another result of extracting hot-important topics was the identification of common or similar topics in these domains. For example, *Human Resources* (HR) and its related topics are common in both management and industrial engineering research. Similarly, *Health* is an incremental research topic in both IT and management domains.

To make the topics comparable, two new criteria (*Portion* and *AvgCite*) were proposed. The corpus of each domain was divided based on four periods of time, and topic trends were described based on these criteria for each topic in a divided corpus. *Cooperation* and *Partnership* was marked as the most hot-important topic of management research. In addition, *Health*-related topics were recognized as incremental topics in recent years for management and IT domains. In the industrial engineering domain, the concentration of researchers has changed from *Manufacturing* and *HR* to *Information Systems, Multiple Criteria Decision Making (MCDM)* and *Environmental Management*. The trends also illustrated that *Computer Science* has been widely applied in industrial engineering studies in the last 10 years. In the IT domain, *Data Mining* and *Machine Learning* are applied widely in other topics or domains and *Cloud Computing*, as a new phenomenon, has changed some traditional research topics.

Finally, to answer the last question of this study, high-level categories were explored. In the 9 categories from the 11 explored categories, all three domains are presented and it confirms that there is a strong relationship between these three domains. The results also indicated that *Telecommunication, Automation, Health* and *Computer Science* are incremental research areas in these domains.

Some topics extracted in these domains are similar to each other or sometimes topics with the same label in two domains have a different meaning or cover different subjects. For example, "*Data Mining*" in the IT domain focuses on pure algorithms and concepts of knowledge extracted from raw data. But this topic in industrial engineering denotes application of these algorithms as tools or approaches in industrial information management. Moreover, some keywords have ambiguity in the meaning. For instance, NLP in the IT domain refers to *Natural Language Processing*, but in industrial engineering, it is the abbreviation of *non-linear planning*. Managing the challenges above and analyzing the correlation between topics and determining the multidisciplinary topics between these domains are suggested for future studies.

Acknowledgement

The authors gratefully acknowledge the Iranian Research Institute for Information Science and Technology (IRANDOC) and especially acknowledge the support of the IRANDOC Text Mining and Machine Learning Laboratory.

References

1. Pfeiffer, A., *Close Link Between Engineering and Business Management*, in The New York Times (2009).
2. Porter, A. and Rafols, I. "Is science becoming more interdisciplinary? Measuring and mapping six research fields over time", *Scientometrics*, **81**, pp. 719–745 (2009).
3. He, W. and Xu, L. "A state-of-the-art survey of cloud manufacturing", *International Journal of Computer Integrated Manufacturing*, **28**, pp. 239–250 (2015).
4. Heilig, L. and Voβ, S. "A scientometric analysis of cloud computing literature", *IEEE Transactions on Cloud Computing*, **2**, pp. 266–278 (2014).
5. Jin, J., Leem, C., and Lee, C. "Research issues and trends in industrial productivity over 44 years", *International Journal of Production Research*, **54**, pp. 1273–1284 (2016).
6. Ronda-Pupo, G. "Knowledge map of Latin American research on management: Trends and future advancement", *Social Science Information*, **55**, pp. 3–27 (2016).
7. Sedighi, M. and Jalalimanesh, A. "Mapping research trends in the field of knowledge management", *Malaysian Journal of Library & Information Science*, **19**, pp. 71–85 (2017).
8. Seth, D., Seth, D., Shrivastava, R., Shrivastava, S., and Shrivastava, S. "An empirical investigation of critical success factors and performance measures for green manufacturing in cement industry", *Journal of Manufacturing Technology Management*, **27**, pp. 1076–1101 (2016).
9. Wagner, C., Roessner, J., Bobb, K., et al. "Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature", *Journal of Informetrics*, **5**, pp. 14–26 (2011).
10. Wagner, A., Roessner, J., Bobb, K., et al. "Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature", *Journal of Informetrics*, **5**, pp. 14–26 (2011).
11. Garfield, E., Sher, I., and Torpie, R. "The use of citation data in writing the history of science," Institute for Scientific Information Inc Philadelphia, PA (1964).

12. Small, H. "A co-citation model of a scientific specialty: A longitudinal study of collagen research", *Social Studies of Science*, **7**, pp. 139–166 (1977).
13. Fu, H. and Ho, Y. "Highly cited Antarctic articles using science citation index expanded: a bibliometric analysis", *Scientometrics*, **109**, pp. 337–357 (2016).
14. Osei-Kyei, R. and Chan, A. "Review of studies on the critical success factors for Public-Private Partnership (PPP) projects from 1990 to 2013", *International Journal of Project Management*, **33**, pp. 1335–1346 (2015).
15. Amado, A., Cortez, P., Rita, P., and Moro, S. "Research trends on big data in marketing: A text mining and topic modeling based literature analysis", *European Research on Management and Business Economics*, **24**, pp. 1–7 (2018).
16. Choi, H.S., Lee, W.S., and Sohn, S. "Analyzing research trends in personal information privacy using topic modeling", *as Computers & Security*, **67**, pp. 244–253 (2017).
17. De Battisti, F., Ferrara, A., and Salini, S. "A decade of research in statistics: a topic model approach", *Scientometrics*, **103**, pp. 413–433 (2015).
18. Gerdsri, N., Kongthon, A., and Puengrusme, S. "Profiling the research landscape in emerging areas using bibliometrics and text mining: A case study of biomedical engineering (BME) in Thailand", *International Journal of Innovation and Technology Management*, **14**, p. 1740011 (2017).
19. Yau, C.-K., Porter, A., Newman, N., and Suominen, A. "Clustering scientific documents with topic modeling", *Scientometrics*, **100**, pp. 767–786 (2014).
20. Furrer, O., Thomas, H., and Goussevskaia, A. "The structure and evolution of the strategic management field: A content analysis of 26 years of strategic management research", *International Journal of Management Reviews*, **10**, pp. 1–23 (2008).
21. Cancino, C., Merigó, J.M, Coronado, F., Dessouky, Y., and Dessouky, M. "Forty years of Computers & Industrial Engineering: A bibliometric analysis", *Computers & Industrial Engineering*, **113**, pp. 614–629 (2017).
22. Lee, W.S. and Sohn, S.Y. "Effects of standardization on the evolution of information and communications technology", *Technological Forecasting and Social Change,* **132**, pp. 308–317 (2018).
23. Gaur, A. and Kumar, M. "A systematic approach to conducting review studies: An assessment of content analysis in 25 years of IB research", *Journal of World Business*, **53**, pp. 280–289 (2018).
24. Hosseini, S., Ivanov, D., and Dolgui, A. "Review of quantitative methods for supply chain resilience analysis", *Transportation Research Part E: Logistics and Transportation Review*, **125**, pp. 285–307 (2019).
25. Jones, S.S., Rudin, R.S., Perry, T., and Shekelle, P.G. "Health information technology: an updated systematic review with a focus on meaningful use", *Annals of Internal Medicine*, **160**, pp. 48–54 (2014).
26. Shi, Y. and Liu, X. "Research on the literature of green building based on the web of science: A scientometric analysis in CiteSpace (2002-2018)", *Sustainability*, **11**, p. 3716 (2019).
27. Morkūnaitė, Ž., Kalibatas, D., and Kalibatienė, D. "A bibliometric data analysis of multi-criteria decision making methods in heritage buildings", *Journal of Civil Engineering and Management*, **25**, pp. 76–99 (2019).
28. Rabiei, M., Hosseini-Motlagh, S.-M., and Haeri, A. "Using text mining techniques for identifying research gaps and priorities: a case study of the environmental science in Iran", *Scientometrics*, **110**, pp. 815–842 (2017).
29. Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., and Zeng, X.J. "A bibliometric analysis and visualization of research", *Sustainability*, **10**, p. 166 (2018).
30. Sedighi, M. and Jalalimanesh, A. "Mapping research trends in the field of knowledge management", *Malaysian Journal of Library & amp; Information Science*, **19**, pp. 71–85, 2017-03-22 (2017).
31. Thomson-Results (2017, 8/12/2017). *Web of ScienceTM Core Collection Help*. Available: http://images.webofknowledge.com/WOKRS524B8/help/WOS/hp_subject_category_terms_tasca.html
32. Elango, B. and Ho, Y.S. "Top-cited articles in the field of tribology : A bibliometric analysis", *Journal of Scientometrics and Information Management*, **12**, pp. 289–307 (2018).
33. Kim, M.C. and Zhu, Y. "Scientometrics of scientometrics: mapping historical footprint and emerging technologies in scientometrics", in *Scientometrics*, Ed: IntechOpen, p. 9 (2018).
34. Klarenbeek, T. and Boshoff, N. "Measuring multidisciplinary health research at South African universities: a comparative analysis based on co-authorships and journal subject categories", *Scientometrics*, **116**, pp. 1461–1485 (2018).
35. Lin, H., Zhu, Y., Ahmad, N., and Han, Q. "A scientometric analysis and visualization of global research on brownfields", *Environmental Science and Pollution Research*, **26**, pp. 17666–17684 (2019).
36. Fu, H-Z., Wang, M-H., and Ho, Y-S. "The most frequently cited adsorption research articles in the Science Citation Index (Expanded)", *Journal of Colloid and Interface Science*, **379**, pp. 148–156 (2012).
37. Blei, D.M. "Probabilistic topic models", *Communications of the ACM*, **55**, pp. 77–84 (2012).
38. Hu, Z., Fang, S., and Liang, T. "Empirical study of constructing a knowledge organization system of patent documents using topic modeling", *Scientometrics*, **100**, pp. 787–799 (2014).
39. Landauer, T.K., Foltz, P.W., and Laham, D. "An introduction to latent semantic analysis", *Discourse Processes*, **25**, pp. 259–284 (1998).

40. Hofmann, T. "Probabilistic latent semantic indexing", in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 50–57 (1999).
41. Blei, D.M., Ng, Andrew, Y., and Jordan, M.I. "Latent Dirichlet allocation", *Journal of Machine Learning Research*, **3**, pp. 993–1022 (2003).
42. Griffiths, T.L., Jordan, M.I., Tenenbaum, J.B., and Blei, D.M. "Hierarchical topic models and the nested Chinese restaurant process", in *Advances in Neural Information Processing Systems*, pp. 17–24 (2004).
43. Lafferty, J.D. and Blei, D.M. "Correlated topic models", in *Advances in Neural Information Processing Systems*, pp. 147–154 (2006).
44. Chang, J. and Blei, D.M. "Relational topic models for document networks", in *International Conference on Artificial Intelligence and Statistics*, pp. 81–88 (2009).
45. Bosman, J., Mourik, I.V., Rasch, M., Sieverts, E., and Verhoeff, H. "Scopus reviewed and compared: The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar", Universiteitsbibliotheek, pp. 31–63 (2006).
46. Qin, J. "Semantic similarities between a keyword database and a controlled vocabulary database: An investigation in the antibiotic resistance literature", *Journal of the Association for Information Science and Technology*, **51**, pp. 166–180 (2000).
47. Chaudhry, S.S. and Luo, W. "Application of genetic algorithms in production and operations management: a review", *International Journal of Production Research*, **43**, pp. 4083–4101 (2005).
48. Gen, M. and Cheng, R., *Genetic Algorithms and Engineering Optimization*, **7**, John Wiley & Sons (2000).
49. Dasgupta, D. and Michalewicz, Z., *Evolutionary Algorithms in Engineering Applications*, Springer Science & Business Media (2013).
50. Zhang, Y., Porter, A.L., Hu, Z., Guo, Y., and Newman, N.C. "'Term clumping' for technical intelligence: A case study on dye-sensitized solar cells", *Technological Forecasting and Social Change*, **85**, pp. 26–39 (2014).
51. Thukral, A., Jain, A., Aggarwal, M., and Sharma, M., *Semi-automatic Ontology Builder Based on Relation Extraction from Textual Data*, Singapore, pp. 343–350 (2018).
52. Mohapatra, P., Deng, Y., Gupta, A., "Domain knowledge driven key term extraction for IT services", *Cham*, **11236**, pp. 489–504 (2018).
53. Duari, S. and Bhatnagar, V. "sCAKE: semantic connectivity aware keyword extraction", *Information Sciences,* **477**, pp. 100–117 (2019).
54. Huh, J. "Big data analysis for personalized health activities: machine learning processing for automatic keyword extraction approach", *Symmetry*, **10**, p. 93 (2018).
55. Momtazi, S. and Moradiannasab, O. "A statistical approach to knowledge discovery: Bootstrap analysis of language models for knowledge base population from unstructured text", *Scientia Iranica*, **26**, pp. 26–39 (2019).
56. Hall, D., Jurafsky, D., and Manning, C.D. "Studying the history of ideas using topic models", in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 363–371 (2008).
57. Hornik, K. and Grün, B. "Topicmodels: An R package for fitting topic models", *Journal of Statistical Software,* **40**, pp. 1–30 (2011).
58. Weston, S. and Calaway, R. "Getting Started with doParallel and foreach", *Date of Access*, **30** (2017).
59. Klavans, R. and Boyack, K.W. "Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?", *Journal of the Association for Information Science and Technology*, **68**, pp. 984–998 (2017).
60. Singhal, A. "Modern information retrieval: A brief overview", *IEEE Data Eng. Bull.*, **24**, pp. 35–43 (2001).
61. Gronsbell, J., Minnier, J., Yu, S., Liao, K., and Cai, T. "Automated feature selection of predictors in electronic medical records data", *Biometrics*, **75**, pp. 268–277 (2019).
62. Hong, C., Liao, K.P., and Cai, T. "Semi-supervised validation of multiple surrogate outcomes with application to electronic medical records phenotyping", *Biometrics*, **75**, pp. 78–89 (2019).
63. Izadi, N., Aminian, O., and Esmaeili, B. "Occupational accidents in Iran: risk factors and long term trend (2007–2016)", *Journal of Research in Health Sciences*, **19**, pp. 1–6 (2019).
64. Kogi, K. "Work improvement and occupational safety and health management systems: common features and research needs", *Industrial Health*, **40**, pp. 121–133 (2002).
65. Ohniwa, R.L., Hibino, A., and Takeyasu, K. "Trends in research foci in life science fields over the last 30 years monitored by emerging topics", *Scientometrics*, **85**, pp. 111–127 (2010).
66. Bornmann, L., Leydesdorff, L., and Wang, J. "How to improve the prediction based on citation impact percentiles for years shortly after the publication date?", *Journal of Informetrics*, **8**, pp. 175–180 (2014).
67. Wang, J. "Citation time window choice for research impact evaluation", *Scientometrics*, **94**, pp. 851–872 (2013).
68. Wei, X., Xue, H., and Zhang, J. "Partnership in supply Chain risk management research", *Journal of Applied Science and Engineering Innovation*, **2**, pp. 55–60 (2015).
69. Lee, J., Lapira, E., Bagheri, B., and Kao, H. "Recent advances and trends in predictive manufacturing systems in big data environment", *Manufacturing Letters*, **1**, pp. 38–41 (2013).

70. Zhang, L., Luo, Y., Tao, F., et al. "Cloud manufacturing: a new manufacturing paradigm", *Enterprise Information Systems*, **8**, pp. 167–187 (2014).

71. Herrmann, C., Schmidt, C., Kurle, D., Blume, S., and Thiede, S. "Sustainability in manufacturing and factories of the future", *International Journal of Precision Engineering and Manufacturing-Green Technology*, **1**, pp. 283–292 (2014).

72. Chai, J., Liu, J.N.K., and Ngai, E.W.T. "Application of decision-making techniques in supplier selection: A systematic review of literature", *Expert Systems with Applications*, **40**, pp. 3872–3885 (2013).

73. Wu, F., Yeniyurt, S., Kim, D., and Cavusgil, S.T. "The impact of information technology on supply chain capabilities and firm performance: A resource-based view", *Industrial Marketing Management*, **35**, pp. 493–504 (2006).

74. Akbari, O.Z. "A survey of agent-oriented software engineering paradigm: Towards its industrial acceptance", *International Journal of Computer Engineering Research*, **1**, pp. 14–28 (2010).

75. Amezquita-Sanchez, J.P., Valtierra-Rodriguez, M., and Adeli, H. "Wireless smart sensors for monitoring the health condition of civil infrastructure", *Scientia Iranica*, **25**, pp. 2913–2925 (2018).

76. Entezami, A., Shariatmadar, H., and Karamodin, A. "An improvement on feature extraction via time series modeling for structural health monitoring based on unsupervised learning methods", *Scientia Iranica*, **27**(3), pp. 1001–1018 (2020).

77. Vazirizade, S.M., Bakhshi, A., Bahar, O., and Nozhati, S. "Online nonlinear structural damage detection using Hilbert Huang transform and artificial neural networks", *Scientia Iranica*, **26**, pp. 1266–1279 (2019).

78. Kontos, E., Blake, K.D., Chou, W.Y.S., and Prestin, A. "Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012", *Journal of Medical Internet Research*, **16**, pp. 1–16 (2014).

79. Holzinger, A., Dehmer, M., and Jurisica, I. "Knowledge discovery and interactive data mining in Bioinformatics-state-of-the-art, future challenges and research directions", *BMC Bioinformatics*, **15**, p. I1 (2014).

80. Fan, W. and Bifet, A. "Mining big data: current status, and forecast to the future", *ACM sIGKDD Explorations Newsletter*, **14**, pp. 1–5 (2013).

81. Hashem, I., Abaker Targio, Y., Ibrar, A., Nor Badrul, M.S., Gani, A., and Khan, S.U. "The rise of "big data" on cloud computing: Review and open research issues", *Information Systems*, **47**, pp. 98–115 (2015).

82. Cheung, L.T.O., Fok, L., Tsang, E.P.K., Fang, W., and Tsang, H.Y. "Understanding residents' environmental knowledge in a metropolitan city of Hong Kong, China", *Environmental Education Research*, **21**, pp. 507–524 (2015).

83. Cornell, S., Berkhout, F., Tuinstra, W., et al. "Opening up knowledge systems for better responses to global environmental change", *Environmental Science & Policy*, **28**, pp. 60–70 (2013).

84. Bala, S. and Kumar, K. "A literature review on kidney disease prediction using data mining classification technique", *International Journal of Computer Science and Mobile Computing*, **3**, pp. 960–967 (2014).

85. Chaurasia, V. and Pal, S. "Early prediction of heart diseases using data mining techniques", *Caribbean Journal of Science and Technology*, **1**, pp. 208–217 (2013).

86. Masethe, H.D. and Masethe, M.A. "Prediction of heart disease using classification algorithms", in *Proceedings of the World Congress on Engineering and Computer Science*, p. 2224 (2014).

87. Wager, K.A., Lee, F.W., and Glaser, J.P., *Health Care Information Systems: A Practical Approach for Health Care Management*, John Wiley & Sons (2017).

88. Wickens, C.D., Hollands, J.G., Banbury, S., and Parasuraman, R., *Engineering Psychology & Human Performance*, Psychology Press (2015).

89. Landy, F.J. and Conte, J.M. *Work in the 21st Century, Binder Ready Version: An Introduction to Industrial and Organizational Psychology*, John Wiley & Sons (2016).

Biographies

Mohammad Rabiei received his BS degree in Computer Science from Shahid Bahonar University of Kerman, Iran, in 2007, and his MS degree in IT Engineering (E-Commerce) from K.N.T University of Technology, Iran, in 2009. He is currently a PhD student in IT Engineering (E-Commerce) at Iran University of Science and Technology. He became a lecturer in 2006 and joined the Iranian Research Institute for Information Science and Technology (IranDoc) as a faculty member in 2010. He is also an instructor of an E-Business Group. His research interests include text mining, Natural Language Processing (NLP), and user information behavior analysis.

Seyyed-Mahdi Hosseini-Motlagh obtained his BS degree in Industrial Engineering from Iran University of Science and Technology, Iran, and his MS and PhD degrees in Industrial Engineering, in 2003 and 2008, respectively, from Tarbiat Modares University, Iran. He is currently Associate Professor of Industrial Engineering at Iran University of Science and Technology. His research interests include healthcare operation management, supply chain network design, stochastic and robust optimization, channel coordination and routing problems.

Abdorrahman Haeri obtained his BS degree in Industrial Engineering from Isfahan University of Technology, Iran, his MS degree in Industrial Engineering, in 2008, from Sharif University of Technology, Iran, and his PhD in Industrial Engineering, in 2012, from Tehran University, Iran. He is currently Assistant Professor of Industrial Engineering at Iran University of Science and Technology. His research interests include data mining, text mining, and Data Envelopment Analysis (DEA).

Behrouz Minaei Bidgoli obtained his PhD from Michigan State University, Michigan, USA, in Computer Science and Engineering, and is currently Associate Professor in the School of Computer Engineering at Iran University of Science & Technology, Tehran, Iran. He is the head of a research group in video game development, as well as a laboratory in Data Mining. His research interests include soft computing, machine learning, video game development, text mining, and natural language processing.