Knowledge-Based Systems 80 (2015) 3–13

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

25 years at Knowledge-Based Systems: A bibliometric analysis

M.J. Cobo [a,*], M.A. Martínez [b], M. Gutiérrez-Salcedo [c], H. Fujita [d], E. Herrera-Viedma [e]

[a] Dept. Computer Science, University of Cádiz, Spain
[b] Dept. Social Work, UNIR Research, La Rioja, Spain
[c] Dept. Management and Marketing, University of Jaén, Spain
[d] Intelligent Software Systems Laboratory, Iwate Prefectural University, Japan
[e] Dept. Computer Science and Artificial Intelligence, University of Granada, Spain

ARTICLE INFO

Article history:
Received 5 December 2014
Received in revised form 18 December 2014
Accepted 30 December 2014
Available online 8 January 2015

Keywords:
Bibliometrics
Science mapping
Citations
Co-word analysis
h-index

ABSTRACT

In commemoration of the Anniversary 25th of KnoSys we present a bibliometric analysis of the scientific content of the journal during the period 1991–2014. This analysis shows the conceptual evolution of the journal and some of its performance bibliometric indicators based on citations, as the evolution of its impact factor, its h-index, and its most cited authors/documents.

1. Introduction

Bibliometrics is an important tool for assessing and analyzing the academic research output. It contributes to the progress of science in many different ways [1]: allowing assessing progress made, identifying the most reliable sources of scientific publication, laying the academic foundation for the evaluation of new developments, identifying major scientific actors, developing bibliometric indixes to assess academic output, etc.

Bibliometrics provides objective criteria to assess the research developed by researchers, being increasingly valued as a tool for measuring scholarly quality and productivity [2]. It is an important approach to assess and analyze the research developed by different actors: countries, universities, research centers, research groups, journals and, in general, scientists [1,3].

In bibliometrics, there are two main methods for exploring a research field: performance analysis and science mapping [4,5]. While performance analysis aims to evaluate the citation impact of the scientific production of different scientific actors, science mapping aims to display the conceptual, social or intellectual structure of scientific research and its evolution and dynamical aspects.

The main aim of this paper is to carry out a thorough bibliometric analysis of the research conducted by the journal Knowledge-Based Systems (KnoSys) from 1991 to 2014. On the one hand, a performance bibliometric analysis on KnoSys is carried out by showing any data on some important performance indicators, such as, published documents, received citations, impact factor (IF) of journal [6], h-index of journal [7,8], most cited papers [9,3], most cited authors, and data on geographic distribution of publications. On the other hand, using SciMAT[1] [10], a science mapping analysis [11] based on co-word networks is performed in order to discover the most important research themes dealt in the journal and its conceptual evolution across the period of time 1991–2014. This science mapping analysis is based in the approach presented in [12], and it allows us to enrich the analysis with bibliometric performance indicators in order to highlight those themes that have received more attention by the research community.

This article is organized as follows: Section 2 introduces the dataset. In Section 3, the performance bibliometric analysis is carried out. In Section 4 the science mapping analysis of KnoSys is presented. Finally, some conclusions are drawn in Section 5.

2. Dataset

In order to carry out the performance and science mapping analysis, the research documents published by KnoSys must be col-

* Corresponding author.
E-mail addresses: manueljesus.cobo@uca.es (M.J. Cobo), mundodesilencio@hotmail.com (M.A. Martínez), msalcedo@ujaen.es (M. Gutiérrez-Salcedo), HFujita-799@acm.org (H. Fujita), viedma@decsai.ugr.es (E. Herrera-Viedma).

[1] SciMAT (Science Mapping Analysis software Tool) – http://sci2s.ugr.es/scimat.

http://dx.doi.org/10.1016/j.knosys.2014.12.035
0950-7051/

lected and also, preprocessed.

Since ISI Web of Science (ISIWoS) is the most important bibliometric database, the research documents published by KnoSys were downloaded from it using the following advance query: *SO=("KNOWLEDGE-BASED SYSTEMS")*.

This query retrieved a total of 1864 documents from 1991 to 2014. The corpus was further restricted to articles and reviews. Citations of these documents are also used in this study; they were counted up to 17th October, 2014.

The raw data was downloaded from ISIWoS as plain text and entered into SciMAT to build the knowledge base for the science mapping analysis. Thus, it contains the bibliographic information stored by ISIWoS for each research document. To improve the data quality, a de-duplicating process was applied (the authors keywords and the ISI keywords plus were used as unit of analysis). Words representing the same concepts were grouped. Furthermore, some meaningless keywords in this context, such as stop words or words with a very broad and general meaning, e.g. "ALGORITHMS", were removed. In addition, authors and affiliations were preprocessed.

Next, using the SciMAT period manager, two consecutive periods of times were established to show the conceptual evolution of the KnoSys in the science mapping analysis that will be done in Section 4. To avoid data smoothness, the best option would have been to choose one-year periods. However, it was found that not enough data were generated in the span of a single year to obtain good results from science mapping analysis. For this reason, the entire time period (1991–2014) was subdivided into periods of more than 1 year. Additionally, although it is common to use periods covering the same time span, the decision was taken to have the first period span eighteen years (1991–2008) because of the low numbers of researchers and publications in the early years. This achieved a first period of a reasonable size when compared with the subsequent period, which was necessary for a good science mapping analysis and to detect the main research themes. Therefore, the data are divided into two consecutive periods of time: 1991–2008, 2009–2014, with 2425 and 4112 keywords, respectively.

3. Performance bibliometric analysis of the KnoSys

In this section an analysis based on performance bibliometric indicators is carried out. As aforementioned, the following performance bibliometric indicators are used in our analysis: published documents, received citations, impact factor (IF) of journal [6], h-index of journal [7,8], most cited papers [1,3], most cited authors, and data on geographic distribution of publications.

3.1. Publications and citations

The distribution of publications by year is shown in Fig. 1. As we can observe, the number of publications of KnoSys by year has gone increasing. Until 2007 the number of publications is around 50 by year, i.e, the number is low. From 2008 until 2011, that number of publications is increased to twice, i.e, around 100 by year. And, in the last two years, it has almost tripled, i.e., around 300. Therefore, KnoSys is a journal which has attracted an increasing interest in the scientific community. This is corroborated by the steady growth of papers submitted to the journal in the last years: 1300 submissions were received in 2012, 1546 submissions for 2013, and for 2014 KnoSys has received 1665 submissions.

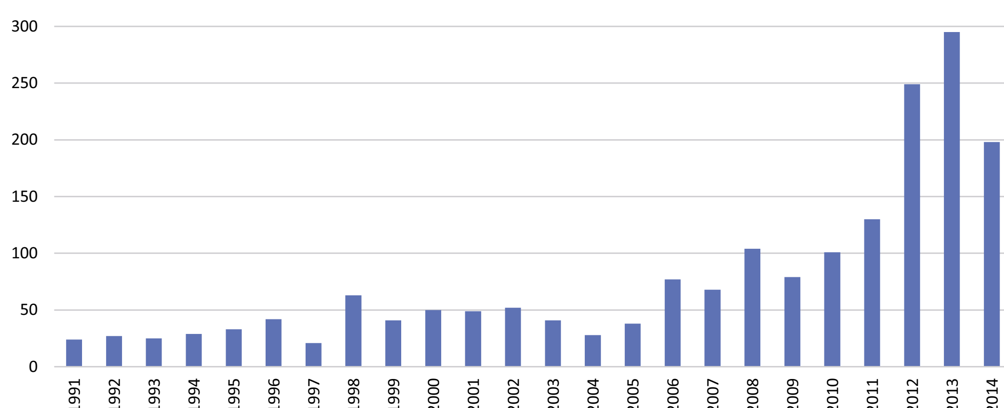


Fig. 1. Distribution of documents by year (1991–2014).

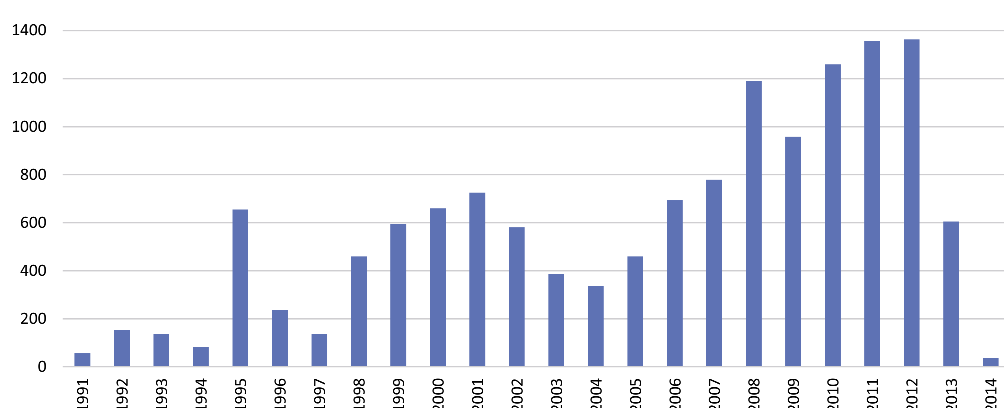


Fig. 2. Distribution of citations by year (1991–2014).

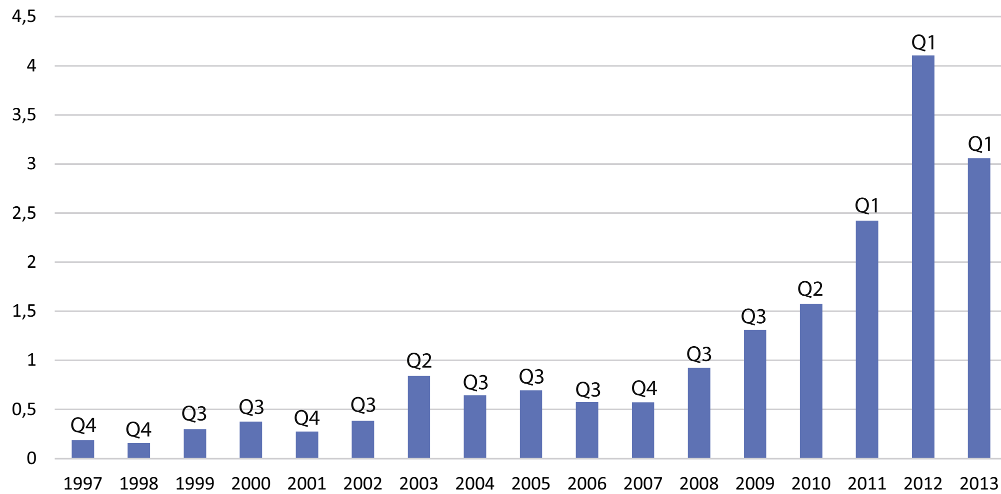


Fig. 3. Impact factor by year (1997–2013).

The distribution of citations count achieved by those documents is shown in Fig. 2. Until 2004 the pattern of citations received by year is similar to a roller coaster, reaching peaks of 600 citations in the years close to 2004. From 2004, where around 300 citations are reached, the citations received has increased constantly until reaching around 1400 citations in 2012. The significant drop in the citations count of the two last years (2013 and 2014) is caused by the citation time-window. Therefore, we could predict that the citations will continue improving in the next years.

3.2. Impact factor

The impact factor of the KnoSys published by Thomson Reuters in its bibliographic database, the Science Journal Citation Reports, from 1997 to 2013 is shown in Fig. 3. We also include information on the position (quartile) inside the ISI category "Computer Science–Artificial Intelligence". The evolution of the impact factor of the KnoSys has been very positive, ranking it in the top positions of the ISI category in the last years, i.e., in the quartile Q1.

3.3. h-index of KnoSys and most cited papers

h-index was first introduced by Hirsch [8], becoming one of the principal bibliometric indexes to assess the scientific performance of a researcher [7]. Its original definition was [8]: "A scientist has index of h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ have $\leqslant h$ citations h".

Although h-index was first developed to measure the scientific performance of researchers, it has been used to measure the performance of different scientific actors [7], such as, journals [13], countries [14], institutes or universities [15].

Thus, in this section the h-index is used to measure the scientific performance of KnoSys. Taking into account the whole period, KnoSys achieves an h-index of 40. In comparison with other journals of the same ISI category, such as, Data & Knowledge Engineering (h-index = 50), Applied Soft Computing (h-index = 49), Information Fusion (h-index = 28), Knowledge and Information System (h-index = 28), KnoSys can be considered as high impact journal.

In addition, the h-index can be used to identify the most prominent output of a scientific output, i.e., the most cited papers [16,17]. To do that, in [9] it was introduced the concept of H-Classics.

Formally, the H-Classics is defined as [9]: "*H-Classics of a research area A could be defined as the H-core of A that is composed of the H highly cited papers with more than H citations received.*"

H-Classics applied on a journal allows us to discover the most cited papers of the journal. In Table 1 is listed the 40 most cited papers that we can identify with the h-index of KnoSys.

The distribution of H-Classics or most cited papers by year is shown in Fig. 4. The most cited papers are usually located in the remotest years due to citation window, i.e., an article need a period of 3 to 7 years to achieve the majority of citations [58]. However, KnoSys presents a good number of most cited documents shown in the most recent years (11 between 2010 and 2012). Clearly, this indicates that the journal is publishing important works that arouse the interest of the scientific community.

Regarding the authors and geographic distribution of the most cited papers, Tables 2 and 3 show the authors and countries, respectively, with more than one most cited paper.

Only three authors have more than one most cited paper, being published in recent years: Dr. Z.S. Xu (2007, and two in 2011), Dr. E. Herrera-Viedma (two in 2010) and G.W. Wei (2010 and 2012).

We must highlight the importance of countries like Peoples Republic of China and United Kingdom, which concentrate the greatest number of most cited papers published by KnoSys.

3.4. Most productive/cited authors and geographic distribution of publications

The most productive/cited authors are shown in Tables 4 and 5, respectively. We should highlight that Drs. Xu, Coonen and Fujita (Editor in Chief of KnoSys) are the most productive authors in the journal. While Drs. Andrews, Diederich, Tickle and Xu are the most cited authors.

Regarding the geographic distribution of publications, in what follows, the most productive countries in different time-windows are analyzed.

First, the whole period 1991–2014 is analyzed. In that sense, the countries with more research documents published in KnoSys are shown in Table 6.

Globally, Peoples Republic of China is the country with the largest published documents. In fact, it doubles the number of documents of the second one. United Kingdom and USA published more than 200 articles.

In order to show the evolution of the geographic distribution of publications, the whole period is divided into two consecutive periods of time: 1991–2008 and 2009–2014.

Regarding the first period (1991–2008), Table 7 shows the most productive countries in that period.

Table 1
H-Classics of KnoSys.

Rank	Title	Authors	Year	#Citations	Refs.
1	Survey and critique of techniques for extracting rules from trained artificial neural networks	Andrews, R, Diederich, J, Tickle, AB	1995	398	[18]
2	Case-based reasoning is a methodology not a technology	Watson, I	1999	143	[19]
3	A resource limited artificial immune system for data analysis	Timmis, J, Neal, M	2001	137	[20]
4	GRA method for multiple attribute decision making with incomplete weight information in intuitionistic fuzzy setting	Wei, GW	2010	98	[21]
5	Induced generalized intuitionistic fuzzy operators	Xu, ZS, Xia, MM	2011	93	[22]
6	Managing the consensus in group decision making in an unbalanced fuzzy linguistic context with incomplete information	Cabrerizo, FJ, Perez, IJ, Herrera-Viedma, E	2010	92	[23]
7	On rule interestingness measures	Freitas, AA	1999	91	[24]
8	Approaches to multiple attribute group decision making based on intuitionistic fuzzy power aggregation operators	Xu, ZS	2011	85	[25]
9	Data analysis approaches of soft sets under incomplete information	Xiao, Z, Zou, Y	2008	79	[26]
10	A rough sets based characteristic relation approach for dynamic attribute generalization in data mining	Li, TR, Xu, Y, Ruan, D, Geert, W, Song, J	2007	78	[27]
11	Agent-based formation of virtual organisations	Jennings, NR, Norman, TJ, Preece, A, Chalmers, S, Luck, M, Dang, VD, Nguyen, TD, Deora, V, Shao, JH, Gray, WA, Fiddian, NJ	2004	73	[28]
12	Hesitant fuzzy prioritized operators and their application to multiple attribute decision making	Wei, GW	2012	68	[29]
13	Intuitionistic fuzzy ordered weighted distance operator	Zeng, SZ, Su, WH	2011	65	[30]
14	Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries	Herrera-Viedma, E, Porcel, C	2010	63	[31]
15	Wizard of oz studies – why and how	Dahlback, N, Jonsson, A, Ahrenberg, L	1993	62	[32]
16	The omnipresence of case-based reasoning in science and application	Aha, DW	1998	61	[33]
17	Creativity, emergence and evolution in design	Gero, JS	1996	60	[34]
18	Decider: A fuzzy multi-criteria group decision support system	Ma, J, Zhang, GQ, Lu, J	2010	59	[35]
19	Ontology-based context synchronization for ad hoc social collaborations	Jung, JJ	2008	58	[36]
20	Collaborative filtering adapted to recommender systems of e-learning	Bobadilla, J, Hernando, A, Serradilla, F	2009	58	[37]
21	A knowledge-based supplier intelligence retrieval system for outsource manufacturing	Choy, KL, Lee, WB, Lau, HCW, Choy, LC	2005	57	[38]
22	Feature selection in bankruptcy prediction	Tsai, CF	2009	56	[39]
23	A method for multiple attribute decision making with incomplete weight information in linguistic setting	Xu, ZS	2007	54	[40]
24	A data mining approach to the prediction of corporate failure	McClean, S, Lin, FY	2001	52	[41]
25	Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting	Hadavandi, E, Ghanbari, A, Shavandi, H	2010	51	[42]
26	Mixed feature selection based on granulation and approximation	Hu, QH, Yu, DR, Liu, JF	2008	50	[43]
27	An extended TOPSIS for determining weights of decision makers with interval numbers	Yue, ZL	2011	48	[44]
28	Design engineering – a need to rethink the solution using knowledge based engineering	Chapman, CB, Pinfold, M	1999	47	[45]
29	A case-based technique for tracking concept drift in spam filtering	Delany, SJ, Cunningham, P, Tsymbal, A, Coyle, L	2005	47	[46]
30	Ranking-order case-based reasoning for financial distress prediction	Sun, J, Li, H	2008	47	[47]
31	Towards the application of case based reasoning to decision-making in concurrent product development (concurrent engineering)	Pawar, KS, Haque, BU, Belecheanu, RA, Barson, RJ	2000	46	[48]
32	An information filtering model on the Web and its application in JobAgent	Li, Y, Zhang, C, Swan, JR	2000	45	[49]
33	Risk and confidence analysis for fuzzy multicriteria decision making	Fenton, N, Wang, W	2006	45	[50]
34	Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm	Shen, W, Guo, XP, Wu, C, Wu, DS	2011	43	[51]
35	Multiple algorithms for fraud detection	Aitken, S, Wheeler, R	2000	42	[52]
36	An improved hybrid genetic algorithm: new results for the quadratic assignment problem	Misevicius, A	2004	41	[53]
37	An early warning system for loan risk assessment using artificial neural networks	Xu, J, Li, LX, Yang, BA	2001	40	[54]
38	Information access in context	Budzik, J, Hammond, kJ, Birnbaum, L	2001	40	[55]
39	Knowledge-based process management – an approach to handling adaptive workflow	Chung, PWH, Cheung, L, Stader, J, Jarvis, P, Moore, J, Macintosh, A	2003	40	[56]
40	Web mining model and its applications for information gathering	Li, YF, Zhong, N	2004	40	[57]

In this first period, United Kingdom is the most productive country, followed by USA and Peoples Republic of China. Each of them published more than one hundred documents.

In the next period (2009–2014), Peoples Republic of China and Spain rise the top positions (see Table 8). Peoples Republic of China almost quadrupled the number of published documents by Spain.

Moreover, Peoples Republic of China, Spain and Taiwan increased sharply the number of published documents (114 vs 390, 29 vs 107 and 30 vs 92, respectively). Also, Republic of Korea increased its number of publications.

4. Science mapping analysis of the KnoSys

Science mapping or bibliometric mapping is a spatial representation of how disciplines, fields, specialities, and documents or

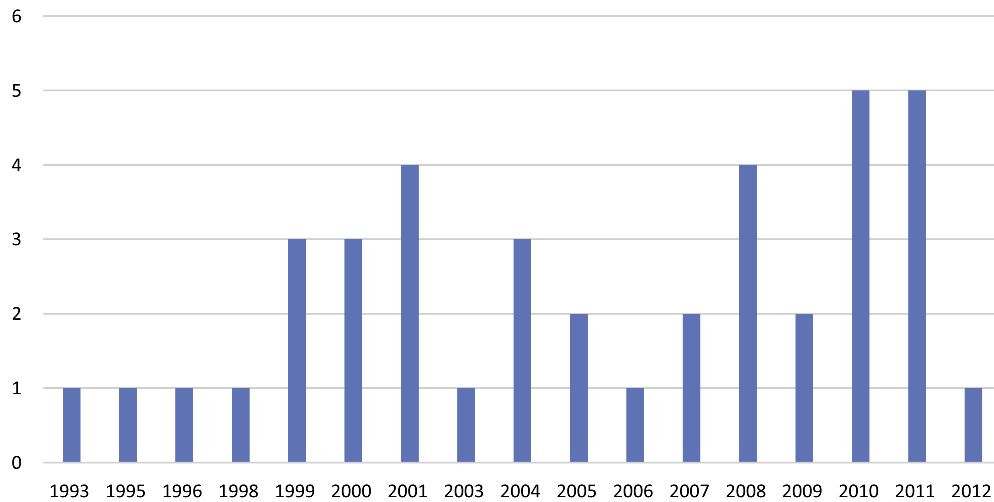


Fig. 4. Distribution of most cited papers by year.

Table 2
Authors with highest number of most cited papers.

Name	Number of documents
XU, ZS	3
HERRERA-VIEDMA, E	2
WEI, GW	2

Table 3
Countries with highest number of most cited papers.

Name	Number of documents
Peoples Republic of China	14
United Kingdom	9
USA	3
Spain	3
Australia	3

Table 4
Most productive authors (1991–2014).

Name	Number of documents
XU, ZS	13
COENEN, F	12
FUJITA, H	12
LIU, J	10
OHSUGA, S	10
TZENG, GH	10
LI, DY	9
SMYTH, B	9
TREUR, J	9
WANG, L	9

Table 5
Most cited authors (1991–2014).

Name	Number of citations
ANDREWS, R	398
DIEDERICH, J	398
TICKLE, AB	398
XU, ZS	327
WEI, GW	214
WATSON, I	170
HERRERA-VIEDMA, E	155
XU, LD	145
PREECE, A	142
XIA, MM	139

Table 6
Most productive countries (1991–2014).

Country	Number of documents
Peoples Republic of China	504
United Kingdom	244
USA	226
Spain	136
Taiwan	122
Canada	85
Australia	76
Japan	72
Republic of Korea	58
Iran	54

Table 7
Most productive countries (1991–2008).

Country	Number of documents
United Kingdom	183
USA	133
Peoples Republic of China	114
Japan	55
Canada	46
Australia	41
Germany	30
Taiwan	30
Spain	29
Republic of Korea	25

Table 8
Most productive countries (2009–2014).

Country	Number of documents
Peoples Republic of China	390
Spain	107
USA	93
Taiwan	92
United Kingdom	61
Iran	46
Canada	39
Australia	35
Republic of Korea	33
Turkey	33

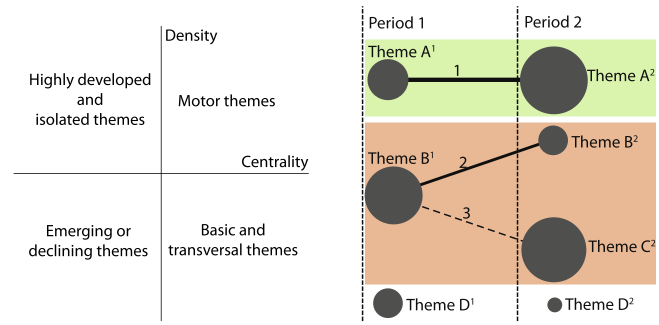
authors are related to one another [59]. It has been widely used to show and uncover the hidden key elements (documents, authors, institutions, topics, etc.) in different research fields [60–68].

Science mapping analysis can be carried out with different software tools [69]. Particularly, SciMAT (Science Mapping Analysis software Tool) was presented in [10] as a powerful science mapping software tool that integrates the majority of the advantages of available science mapping software tools [69]. SciMAT was designed according to the science mapping analysis approach presented in [12]. SciMAT can be freely downloaded, modified and redistributed according to the terms of the GPLv3 license. The executable file, user-guide and source code can be downloaded via the following website (http://sci2s.ugr.es/scimat).

Cobo et al. [12] defined a bibliometric approach that combines both performance analysis tools and science mapping tools to analyze a research field and detect and visualize its conceptual subdomains (particular topics/themes or general thematic areas) and its thematic evolution. It is based on a co-word analysis [70] and the h-index [8].

This approach establishes four stages to analyze a research field in a longitudinal framework:

1. *Detection of the research themes*. In each period of time studied the corresponding research themes are detected by applying a co-word analysis [70] to raw data for all the published documents in the research field, followed by a clustering of keywords to topics/themes using the simple centers algorithm [71], which locates keyword networks that are strongly linked to each other and that correspond to centers of interest or to research problems that are the subject of significant interest among researchers. The similarity between the keywords is assessed using the equivalence index [72]: $e_{ij} = c_{ij}^2/c_ic_j$, where c_{ij} is the number of documents in which two keywords i and j co-occur and c_i and c_j represent the number of documents in which each one appears.
2. *Visualizing research themes and thematic network*. In this phase the detected themes are visualized by means of two different visualization instruments: strategic diagram [73] and thematic network. Each theme can be characterized by two measures [72]: *centrality* and *density*. Centrality measures the degree of interaction of a network with other networks and can be defined as $c = 10 * \sum e_{kh}$, with k a keyword that belongs to the theme and h a keyword that belongs to other themes. The density measures the internal strength of the network and can be defined as $d = 100(\sum e_{ij}/w)$, where i and j are keywords belonging to the theme and w the number of keywords in the theme. Given both measurements, a research field can be visualized as a set of research themes, mapped in a two-dimensional strategic diagram (Fig. 5a) and classified into four groups:
   (a) Themes in the upper-right quadrant are both well developed and important for the structuring of a research field. They are known as the *motor-themes* of the speciality, given that they present strong centrality and high density.
   (b) Themes in the upper-left quadrant have well-developed internal ties but unimportant external ties and so are of only marginal importance for the field. These themes are very *specialized and peripheral*.
   (c) Themes in the lower-left quadrant are both weakly developed and marginal. The themes in this quadrant have low density and low centrality and mainly represent either *emerging or disappearing* themes.
   (d) Themes in the lower-right quadrant are important for a research field but are not developed. This quadrant contains *transversal and general*, basic themes.



(a) The strategic diagram. (b) Thematic evolution.

Fig. 5. Strategic diagram and thematic evolution.

3. *Discovery of thematic areas*. In this phase, the evolution of the research themes over a set of periods of time is first detected and then analyzed to identify the main general areas of evolution in the research field, their origins and their interrelationships. Their evolution over the whole period is then measured as the overlapping of clusters from two consecutive periods. For this purpose, the inclusion index [74] is used to detect conceptual nexuses between research themes in different periods and, in this way, to identify the thematic areas in a research field. A thematic area is defined as a set of themes that have evolved over several periods of time. It is worth noting that interrelationships between research themes could indicate that a particular research theme belongs to a unique thematic area or to more than one thematic area. It could also be that a particular research theme cannot be associated with any of the thematic areas identified and therefore could be interpreted as the origin of a new thematic area in the research field. For example, Fig. 5b shows a bibliometric map of thematic evolution over two time periods. The solid lines (lines 1 and 2) mean that the linked themes share the same name: either the themes are labeled with the same keywords, or the label of one theme is part of the other theme (name of theme $\in$ {thematic nexuses}). A dotted line (line 3) means that the themes share elements that are not the names of the themes (name of theme $\notin$ {thematic nexuses}). The thickness of the lines is proportional to the inclusion index and the volume of the spheres is proportional to the number of published documents associated with each theme. Hence, two different thematic areas can be observed, shaded in different colors, while $ThemeD^1$ is discontinued, and $ThemeD^2$ is considered a new theme. As each theme is associated with a set of documents each thematic area could also have an associated collection of documents, obtained by combining the documents associated with its set of themes.
4. *Performance analysis*. In this phase, the relative contribution of research themes and thematic areas to the whole research field is measured (quantitatively and qualitatively) and used to establish the most prominent, most productive and highest-impact subfields. This performance analysis is developed as a complement to the analysis step of the science mapping workflow. Some of the bibliometric indicators to use are: number of published documents, number of citations, and different types of h-index [7,9,8]. In order to assess performance measures or indicators, first a set of documents must be assigned to each detected theme by means of a document mapper function [10]. In particular, the union document mapper returns the algebraic union of the set of documents associated with the keywords of theme. Thus, since keywords associated with a document could belong to different themes, a document could

be associated with several themes. Moreover, since thematic areas could share some themes, they could also share some documents.

4.1. Analysis of the content of the papers published

In order to analyze the most highlighted themes of the journal for each period, several strategic diagrams are shown. In addition, the sphere size is proportional to the number of published documents associated with each research theme. Furthermore, the number of citations achieved by each theme is shown in brackets.

First period (1991–2008). According to the strategic diagram presented in Fig. 6, during this period of time KnoSys pivots on nineteen research themes, with the following ten major themes (motor themes plus basic themes): *Casual-networks*, *Data-mining*, *Natural-language-preprocessing*, *Optimization*, *Neural-network*, *Classification*, *Bayesian-networks*, *Case-based-reasoning*, *Intelligent-agents* and *Expert-systems*. The performance measures of the period themes are given in Table 9: the number of documents, the citations achieved by those documents, and h-index. According to the performance measures, the following six themes could be highlighted: *Expert-system*, *Data-mining*, *Intelligent-agents*, *Case-based-reasoning*, *Neural-networks* and *Classification*. These research themes get important impact rate, achieving more than one-thousand citations and getting higher h-indexes in comparison with the remaining themes.

The motor theme *Data-mining* gets the highest citations count in that period, and also achieves a great h-index. It is related with general topics of data mining (theory and practice) non-related with classification, such as, clustering, association rules, or frequent itemsets. Particularly it is focused on extracting knowledge from databases. Also, it comprises research conducted on rough sets.

Table 9
Performance of the themes in the 1991–2008 period.

Name	Number of documents	Number of citations	h-index
Expert-system	120	1001	17
Data-mining	110	1536	21
Intelligent-agents	106	1235	19
Case-based-reasoning	96	1362	21
Neural-networks	88	1276	20
Classification	76	1462	22
Natural-language-processing	49	322	9
Optimization	37	538	13
Bayesian-networks	20	190	8
Computational-reasoning	19	104	7
Fault-diagnosis	14	113	6
Logic-programming	14	151	5
Conceptual-graphs	11	57	5
Verification	11	63	6
Causal-networks	9	209	5
Dialog	9	132	6
Formal-concept-analysis	8	145	6
Aggregation-operators	7	189	7
Lyee-methodology	3	7	2

The motor theme *Natural-language-preprocessing* gets a low impact rates. It collects the research conducted on knowledge representation, conversational agents and general aspects related with the language preprocessing in a computational way. Also, it studies topics related with the semantic web and ontologies.

Optimization is a motor theme which is mainly focused on the optimization of decision support systems and other aspects of the process of decision making. Also, it covers topics related with parameters optimization.

The basic and transversal theme *Classification* receives a great number of citations, moreover it gets the best h-index rate in this period. It represents the research conducted on different aspects of

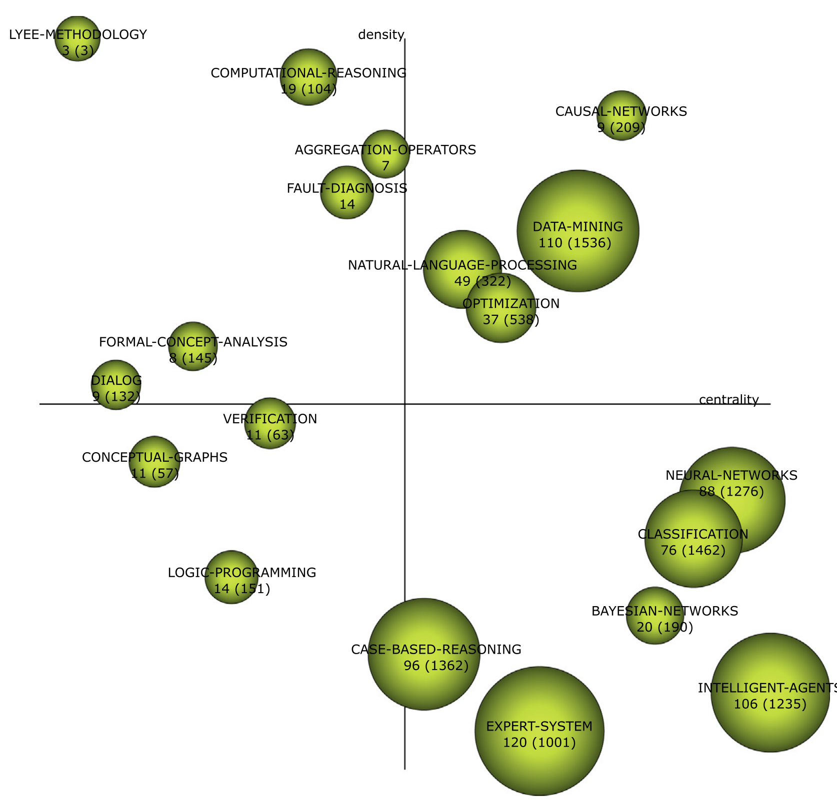


Fig. 6. Strategic diagram for the 1991–2008 period.

machine learning focused in the classification task: machine learning, feature selection, nearest neighbors or support vector machines. Furthermore, this topic has been applied in bank failure in order to predict the bankruptcy.

The basic and transversal theme *Neural-network* specializes in aspects related with classification issues, such as neural network systems, genetic algorithm, decision-tree or identification. In fact, the hybridization of some of this techniques are developed and also compared. For instance, neural-network and genetic algorithm are commonly used together. Regarding its performance indicators, *Neural-network* is very similar to the theme *Classification*, also we could remark that it has fewer impact whereas concentrates a bigger amount of documents.

The basic and transversal theme *Case-based-reasoning* achieves great impact rates. It is related with different aspects and applications of the case based reasoning systems (decision making in product development, engineering, aids assessment, etc.), and their use in information retrieval systems.

The theme *Intelligent-agents* is related with various aspects of agent systems, such as, multi-agent-systems, workflow monitoring, agent collaboration, negotiation, etc. This basic theme gets the highest centrality rank in this period.

The *Expert-systems* covers different aspects related with decision support system based on multicriteria decision making, muticriteria decision aid and decision trees. Also the theme is related with expert systems for different contexts (e.g. enterprise performance, software evaluation, process selection, diagnostic of diseases or ecosystem) and knowledge based and acquisition in expert systems. The theme *Expert-systems* gets the highest number of documents, but its impact rate is low in comparison with the other highlighted themes.

Second period (2009–2014). The research conducted in this period pivots on twenty-five themes. According to the strategic diagram shown in Fig. 7, thirteen major themes can be identified (motor plus basic themes): *Rough-sets*, *Neural-networks*, *Group-decision-making*, *Hybrid-Systems*, *Databases*, *Gene-expression-data*, *Feature-selection*, *Nearest-Neighbor*, *Similarity-measure*, *Multi-attribute-decision-making*, *Uncertainty*, *Clustering* and *Decision-Support-Systems*.

Regarding the performance measures shown in Table 10, two themes stand out due to the citations achieved: *Neural-networks* and *Group-decision-making*. Moreover, the following seven themes should be remarked due to they get an h-index over ten: *Decision-support-systems*, *Similarity-measure*, *Rough-sets*, *Uncertainty*, *Gene-expression-data*, *Feature-selection* and *Multi-attribute-decision-making*.

We should point out that taking into account the structural composition and the impact measures, three themes vertebrate the research carried out by KnoSys in this period: *Rough-sets*, *Neural-networks* and *Group-decision-making*.

The motor theme *Rough-sets* appears in this period as a bifurcation of the theme *Data-mining* from the previous period. It obtains the highest density score, which means that the research conducted in it has great internal cohesion. This research theme covers different aspects related with the application of rough set in data mining. Regarding its performance indicators, it gets a great impact rate with a moderate number of documents.

Neural-networks consolidates as motor theme in this period. Moreover, it is the theme with the highest number of documents and it achieves the highest number of citations count. *Neural-networks* maintains the focus in the topics that it pivoted in the previous period, that is, specific classification techniques (e.g. genetic algorithms, support vector machines, neural networks or particle swarm optimization) and how they can be hybridized with neural networks. Also, it covers themes related with prediction and time series.

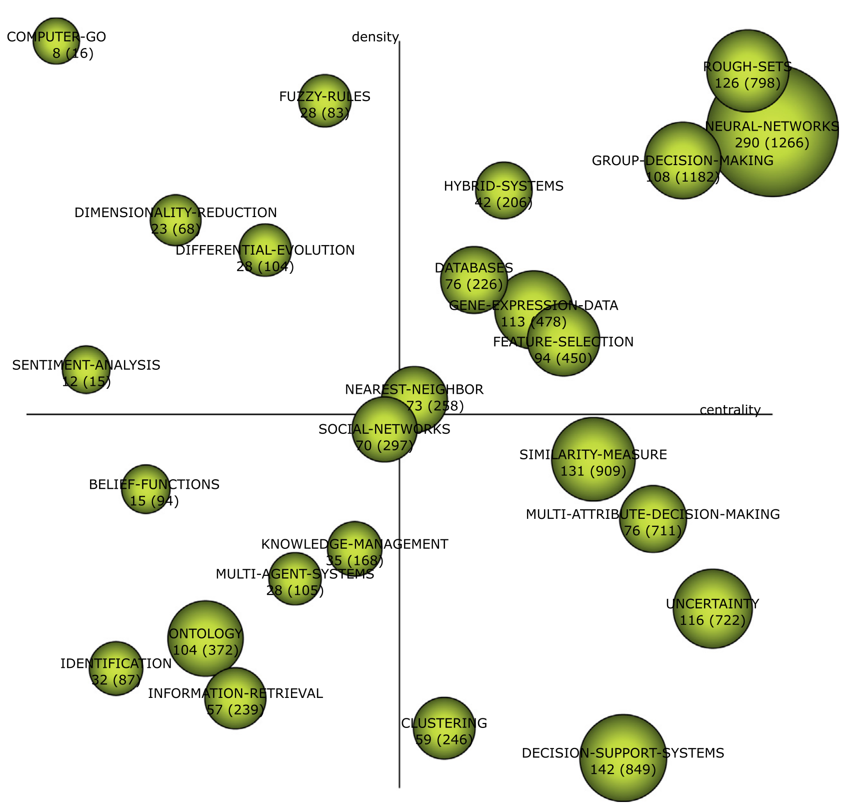


Fig. 7. Strategic diagram for the 2009–2014 period.

Table 10
Performance of the themes in the 2009–2014 period.

Name	Number of documents	Number of citations	h-index
Neural-networks	290	1,266	17
Decision-support-systems	142	849	14
Similarity-measure	131	909	17
Rough-sets	126	798	17
Uncertainty	116	722	15
Gene-expression-data	113	478	11
Group-decision-making	108	1182	18
Ontology	104	372	9
Feature-selection	94	450	12
Multi-attribute-decision-making	76	711	14
Databases	76	226	8
Nearest-neighbor	73	258	9
Social-networks	70	297	9
Clustering	59	246	8
Information-retrieval	57	239	7
Hybrid-systems	42	206	8
Knowledge-management	35	168	7
Identification	32	87	6
Fuzzy-rules	28	83	6
Differential-evolution	28	104	7
Multi-agent-systems	28	105	5
Dimensionality-reduction	23	68	5
Belief-functions	15	94	5
Sentiment-analysis	12	15	2
Computer-go	8	16	2

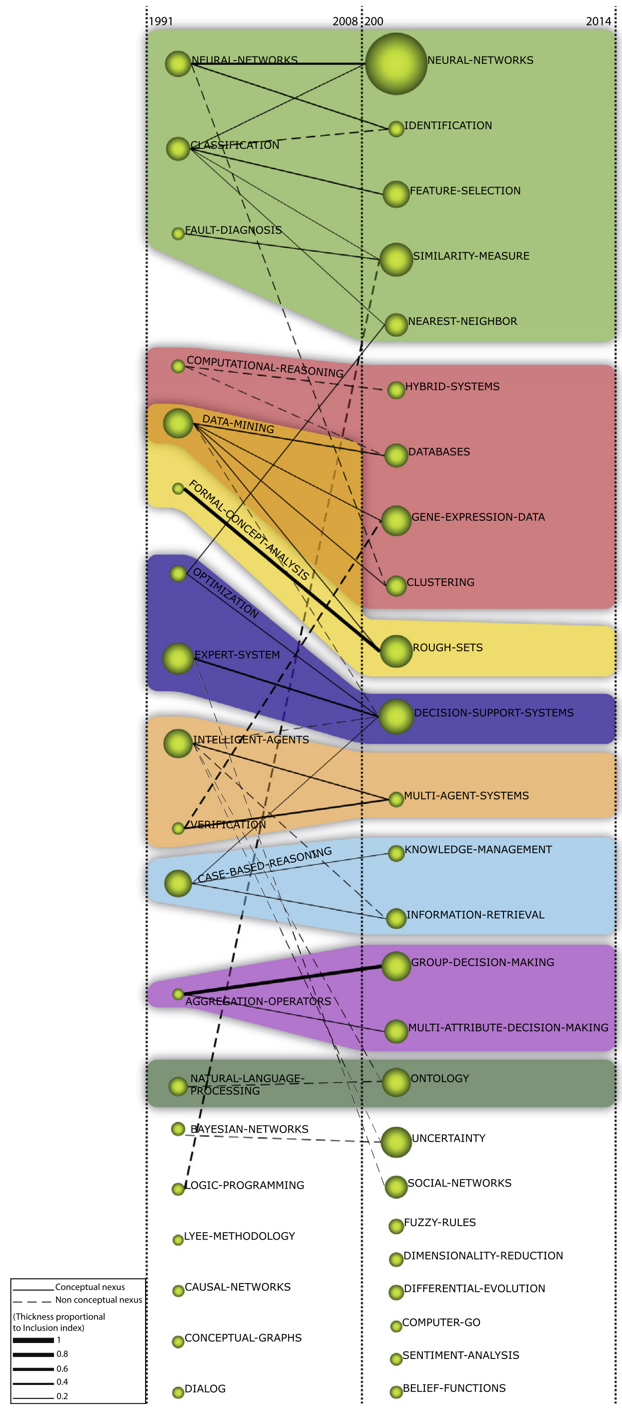


Fig. 8. Thematic evolution (1991–2014).

The motor theme *Group-decision-making* emerges in this period and it could be considered as one of the main topics of KnoSys. In fact, it gets the highest h-index score and achieves a huge number of citations. *Group-decision-making* rises up from the evolution of the theme *Aggregation-operators* of the previous period. The research conducted in this theme is related with the management of incomplete information, aggregation operators (general, intuitionistic, prioritized, weighted, interval-valued, etc.), consensus, consistency or preference relations.

The basic and transversal theme *Multi-attribute-decision-making* is close related with *Group-decision-making*. In that case, it is focused on vague sets, fuzzy numbers and hesitant fuzzy set. Although *Multi-attribute-decision-making* gets a moderate number of documents it achieves large impact rates.

The basic and transversal theme *Uncertainty* conducted the research related with management of the uncertainty on decision making systems. For instance, it covers topics related with Dempster–Shafer theory, analytical hierarchy process or fuzzy sets.

We should highlight the presence of two important emergent themes that presents remarkable impact rates, i.e., *Ontology* and *Social-networks*. *Ontology* is related with the semantic web tools and *Social-networks* is related with relevant topics such as community detection, reputation in social networks, trust propagation and Web 2.0 tools. It is clear that both themes will attract the interest of the scientific community in the future.

4.2. Conceptual evolution map

Using SciMAT, an analysis of the themes detected in each period of time by considering their keywords and evolution across time was developed. In such a way, the thematic areas that concentrate the research conducted in KnoSys were identified. So, eight thematic areas were identified: *Classification*, *Data-mining*, *Rough-sets*, *Decision-support-systems*, *Agent-systems*, *Information-retrieval*, *Group-decision-making* and *Semantic-web*.

The conceptual evolution and the detected thematic areas are shown in Fig. 8. As aforementioned, in this map the solid lines mean a thematic nexus, a dotted line means that the linked themes share keywords different to the name of the themes, the thickness of the edge is proportional to the inclusion index, and the sphere size is proportional to the number of retrieved documents in each theme. Furthermore, the different shadows are used to group the themes that belong to the same thematic area. Moreover, the performance measures of the detected thematic areas are shown in Table 11.

The thematic area *Classification* is the largest one, and gets the best performance indicators. It doubles the documents count and the citations achieved by the second one. Furthermore, thirty-two of its documents were highly cited according to the h-index. Regarding to its structure and theme composition, *Classification* increases the number of themes in the second period. Also, it is

Table 11
Performance of the thematic areas.

Thematic area	Number of documents	Number of citations	h-index
Classification	614	4744	32
Data-mining	360	2582	25
Decision-support-systems	293	2335	24
Rough-sets	242	2453	25
Information-retrieval	186	1765	21
Group-decision-making	160	1704	22
Semantic-web	153	694	13
Agent-systems	143	1397	20

mainly composed of motor and basic themes. *Classification* covers topics related with different techniques of machine learning used in the classification task, such as, genetic algorithms, neural networks, support vector machines, or particle swarm optimization. Also, some hybridizations are developed.

Data-mining is a thematic area covering different aspects of knowledge extraction from databases and specific areas of data mining. We should highlight that in the last period there is an increase on the interest in applying data mining techniques in cancer research. Regarding its thematic composition, *Data-mining* started composed of one motor theme plus one isolated theme. In the second period it consolidated and the majority of its themes are motor themes. According to its performance indicators, it is the second one with best number of documents, number of citations achieved and h-index score.

Rough-sets consolidates as research area in the second period, evolving from the theme *Data-mining* of the previous one. In fact, it could be considered as one of the most important topics covered by the journal in the last years. Also, it gets important impact rates with a moderate number of documents.

Decision-support-systems is devoted to the decision systems and how they are employed in expert systems, and also, knowledge systems. It is mainly composed of basic and transversal themes.

Regarding the structural composition of the research area *Agent-systems* (first basic themes, and finally declining themes), we could point out that the research conducted in the journal has lose part of the interest on this thematic. In fact, this thematic area presents the lowest number of documents and presents a low impact scores in comparison with the other thematic areas.

Information-retrieval is composed in the two studied period of themes with low density and centrality. It could indicates that the research conducted in this area has not attracted the attention of the research community of the journal.

Regarding the thematic area *Group-decision-making*, meanwhile it presents moderate impact rates, it is one of the most important area of the journal in the last years. It emerges from the highly isolated theme of the first period *Aggregation-operators*, and becomes one of the most important theme in the second period.

Finally, the thematic area *Semantic-web* started composed of a motor themes and finalized composed by only one emerging theme.

To sum up, two conclusions could be made:

- The thematic areas *Classification*, *Data-mining*, *Rough-sets* and *Group-decision-making* present a growing pattern and in the second period they get a higher interest from the scientific community.
- The thematic areas *Decision-support-systems*, *Information-retrieval*, *Agent-systems* and *Semantic-web* present a shrinking patterns. We should remark that this issue does not mean that the research conducted on those research areas has less quality. It means that the journal and its research community has been less interested in those topics.

5. Conclusions

In this paper, a bibliometric analysis of KnoSys has been developed. Some important findings are the followings:

- The journal has attracted the interest of the scientific community throughout years, which is observed in the great growth of publications, citations and submissions received.
- The impact factor of KnoSys has increased until to consolidate the journal in the first quartile in the ISI category "Computer Science–Artificial Intelligence" in the last editions of Journal Citation Reports.
- The h-index of KnoSys is 40, therefore according to the concept of H-Classics [9], we can identify 40 most cited papers in KnoSys. The authors Drs. Xu, Herrera-Viedma and Wei contributed with more than one most cited papers, and also, those documents where published in recent years. Moreover, we should point out that Peoples Republic of China is the country with more highly cited documents, publishing an amount of 14 most cited papers.
- We have identified the authors that have played a prominent role in the development of KnoSys. Some of them are: Drs. Andrew, Coonen, Diederich, Fujita, Tickle and Xu.
- During 1991–2008, United Kingdom, USA and Peoples Republic of China are the countries most productive. But, during 2009–2014 these positions changed, being Peoples Republic of China and Spain on the top, as most productive. Also, Taiwan increased its number of published documents.
- In the whole period, Peoples Republic of China is undoubtedly the most productive country, doubling the publications of the second one.
- The publications of KnoSys is focused on eight great thematic areas: *Classification*, *Data-mining*, *Rough-sets*, *Decision-support-systems*, *Agent-systems*, *Information-retrieval*, *Group-decision-making* and *Semantic-web*.
- We have identified two emergent themes that could attract the interest of the scientific community in the future: Ontology and Social networks. Both important topics related with the development of the Web and the problem of Big Data.

Acknowledgments

This work has been supported by the Excellence Andalusian Projects TIC-5299 and TIC-5991, and National Project TIN2010-17876 and TIN2013-40658-P.

References

[1] M.A. Martínez-Sánchez, M.J. Cobo, M. Herrera, E. Herrera-Viedma, Analyzing the scientific evolution of social work using science mapping, Res. Soc. Work Pract. (2014) in press, http://dx.doi.org/10.1177/1049731514522101.
[2] H. Moed, R. De Bruin, T. Van Leeuwen, New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications, Scientometrics 33 (1995) 381–422.
[3] M.A. Martínez, M. Herrera, E. Contreras, A. Ruíz, E. Herrera-Viedma, Characterizing highly cited papers in social work through h-classics, Scientometrics (2014). in press, http://dx.doi.org/10.1007/s11192-014-1460-y.
[4] E.C.M. Noyons, H.F. Moed, M. Luwel, Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study, J. Am. Soc. Inf. Sci. 50 (1999) 115–131.
[5] A.F.J. Van Raan, Measuring science, in: H.F. Moed, W. Glänzel, U. Schmoch (Eds.), Handbook of Quantitative Science and Technology Research, Springer, Netherlands, 2005, pp. 19–50.
[6] E. Garfield, Citation analysis as a tool in journal evaluation, Science 178 (1972) 417–479.
[7] S. Alonso, F. Cabrerizo, E. Herrera-Viedma, F. Herrera, h-index: a review focused in its variants, computation and standardization for different scientific fields, J. Informetr. 3 (2009) 273–289.
[8] J. Hirsch, An index to quantify an individuals scientific research out-put, Proc. Nat. Acad. Sci. 102 (2005) 16569–16572.

[9] M.A. Martínez, M. Herrera, J. López-Gijón, E. Herrera-Viedma, H-classics: characterizing the concept of citation classics through h-index, Scientometrics 98 (2014) 1971–1983.

[10] M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, F. Herrera, Scimat: a new science mapping analysis software tool, J. Am. Soc. Inf. Sci. Technol. 63 (2012) 1609–1630.

[11] K. Börner, C. Chen, K.W. Boyack, Visualizing knowledge domains, Annu. Rev. Inf. Sci. Technol. 37 (2003) 179–255.

[12] M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, F. Herrera, An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field, J. Informetr. 5 (2011) 146–166.

[13] T. Braun, W. Glanzel, A. Schubert, A hirsch-type index for journals, Scientometrics 69 (2006) 169–173.

[14] J. Guan, X. Gao, Comparison and evaluation of chinese research performance in the field of bioinformatics, Scientometrics 75 (2008) 357–379.

[15] A. Schubert, Successive h-indices, Scientometrics 70 (2007) 201–205.

[16] E. Garfield, Introducing citation classics: the human side of scientific reports, Curr. Comments 1 (1977) 5–7.

[17] F. Ponce, A. Lozano, The most cited works in parkinson's disease, Mov. Disord. 26 (2011) 380–390.

[18] R. Andrews, J. Diederich, A. Tickle, Survey and critique of techniques for extracting rules from trained artificial neural networks, Knowl.-based Syst. 8 (1995) 373–389.

[19] I. Watson, Case-based reasoning is a methodology not a technology, Knowl.-based Syst. 12 (1999) 303–308.

[20] J. Timmis, M. Neal, A resource limited artificial immune system for data analysis, Knowl.-based Syst. 14 (2001) 121–130.

[21] G. Wei, GRA method for multiple attribute decision making with incomplete weight information in intuitionistic fuzzy setting, Knowl.-based Syst. 23 (2010) 243–247.

[22] Z. Xu, M. Xia, Induced generalized intuitionistic fuzzy operators, Knowl.-based Syst. 24 (2011) 197–209.

[23] F. Cabrerizo, I. Perez, E. Herrera-Viedma, Managing the consensus in group decision making in an unbalanced fuzzy linguistic context with incomplete information, Knowl.-based Syst. 23 (2010) 169–181.

[24] A. Freitas, On rule interestingness measures, Knowl.-based Syst. 12 (1999) 309–315.

[25] Z. Xu, Approaches to multiple attribute group decision making based on intuitionistic fuzzy power aggregation operators, Knowl.-based Syst. 24 (2011) 749–760.

[26] Y. Zou, Z. Xiao, Data analysis approaches of soft sets under incomplete information, Knowl.-based Syst. 21 (2008) 941–945.

[27] T. Li, D. Ruan, W. Geert, J. Song, Y. Xu, A rough sets based characteristic relation approach for dynamic attribute generalization in data mining, Knowl.-based Syst. 20 (2007) 485–494.

[28] T. Norman, A. Preece, S. Chalmers, N. Jennings, M. Luck, V. Dang, T. Nguyen, V. Deora, J. Shao, W. Gray, N. Fiddian, Agent-based formation of virtual organisations, Knowl.-based Syst. 17 (2004) 103–111.

[29] G. Wei, Hesitant fuzzy prioritized operators and their application to multiple attribute decision making, Knowl.-based Syst. 31 (2012) 176–182.

[30] S. Zeng, W. Su, Intuitionistic fuzzy ordered weighted distance operator, Knowl.-based Syst. 24 (2011) 1224–1232.

[31] C. Porcel, E. Herrera-Viedma, Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries, Knowl.-based Syst. 23 (2010) 32–39.

[32] N. Dahlback, A. Jonsson, L. Ahrenberg, Wizard of oz studies – why and how, Knowl.-based Syst. 6 (1993) 258–266.

[33] D. Aha, The omnipresence of case-based reasoning in science and application, Knowl.-based Syst. 11 (1998) 261–273.

[34] J. Gero, Creativity, emergence and evolution in design, Knowl.-based Syst. 9 (1996) 435–448.

[35] J. Ma, J. Lu, G. Zhang, Decider: a fuzzy multi-criteria group decision support system, Knowl.-based Syst. 23 (2010) 23–31.

[36] J. Jung, Ontology-based context synchronization for ad hoc social collaborations, Knowl.-based Syst. 21 (2008) 573–580.

[37] J. Bobadilla, F. Serradilla, A. Hernando, Collaborative filtering adapted to recommender systems of e-learning, Knowl.-based Syst. 22 (2009) 261–265.

[38] K. Choy, W. Lee, H. Lau, L. Choy, A knowledge-based supplier intelligence retrieval system for outsource manufacturing, Knowl.-based Syst. 18 (2005) 1–17.

[39] C. Tsai, Feature selection in bankruptcy prediction, Knowl.-based Syst. 22 (2009) 120–127.

[40] Z. Xu, A method for multiple attribute decision making with incomplete weight information in linguistic setting, Knowl.-based Syst. 20 (2007) 719–725.

[41] F. Lin, S. McClean, A data mining approach to the prediction of corporate failure, Knowl.-based Syst. 14 (2001) 189–195.

[42] E. Hadavandi, H. Shavandi, A. Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, Knowl.-based Syst. 23 (2010) 800–808.

[43] Q. Hu, J. Liu, D. Yu, Mixed feature selection based on granulation and approximation, Knowl.-based Syst. 21 (2008) 294–304.

[44] Z. Yue, An extended TOPSIS for determining weights of decision makers with interval numbers, Knowl.-based Syst. 24 (2011) 146–153.

[45] C. Chapman, M. Pinfold, Design engineering – a need to rethink the solution using knowledge based engineering, Knowl.-based Syst. 12 (1999) 257–267.

[46] S. Delany, P. Cunningham, A. Tsymbal, L. Coyle, A case-based technique for tracking concept drift in spam filtering, Knowl.-based Syst. 18 (2005) 187–195.

[47] H. Li, J. Sun, Ranking-order case-based reasoning for financial distress prediction, Knowl.-based Syst. 21 (2008) 868–878.

[48] B. Haque, R. Belecheanu, R. Barson, K. Pawar, Towards the application of case based reasoning to decision-making in concurrent product development (concurrent engineering), Knowl.-based Syst. 13 (2000) 101–112.

[49] Y. Li, C. Zhang, J. Swan, An information filtering model on the web and its application in jobagent, Knowl.-based Syst. 13 (2000) 285–296.

[50] N. Fenton, W. Wang, Risk and confidence analysis for fuzzy multicriteria decision making, Knowl.-based Syst. 19 (2006) 430–437.

[51] W. Shen, X. Guo, C. Wu, D. Wu, Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, Knowl.-based Syst. 24 (2011) 378–385.

[52] R. Wheeler, S. Aitken, Multiple algorithms for fraud detection, Knowl.-based Syst. 13 (2000) 93–99.

[53] A. Misevicius, An improved hybrid genetic algorithm: new results for the quadratic assignment problem, Knowl.-based Syst. 17 (2004) 65–73.

[54] B. Yang, L. Li, J. Xu, An early warning system for loan risk assessment using artificial neural networks, Knowl.-based Syst. 14 (2001) 303–306.

[55] J. Budzik, K. Hammond, L. Birnbaum, Information access in context, Knowl.-based Syst. 14 (2001) 37–53.

[56] P. Chung, L. Cheung, J. Stader, P. Jarvis, J. Moore, A. Macintosh, Knowledge-based process management – an approach to handling adaptive workflow, Knowl.-based Syst. 16 (2003) 149–160.

[57] Y. Li, N. Zhong, Web mining model and its applications for information gathering, Knowl.-based Syst. 17 (2004) 207–217.

[58] J. Wang, Citation time window choice for research impact evaluation, Scientometrics 94 (2013) 851–972.

[59] H. Small, Visualizing science by citation mapping, J. Am. Soc. Inf. Sci. 50 (1999) 799–813.

[60] R. Cartes-Velásquez, C. Manterola-Delgado, Bibliometric analysis of articles published in ISI dental journals, 2007–2011, Scientometrics 98 (2014) 2223–2233.

[61] M.J. Cobo, A.G. López-Herrera, F. Herrera, E. Herrera-Viedma, A note on the ITS topic evolution in the period 2000–2009 at T-ITS, IEEE Trans. Intell. Transp. Syst. 13 (2012) 413–420.

[62] M.J. Cobo, F. Chiclana, A. Collop, J. de Oña, E. Herrera-Viedma, A bibliometric analysis of the intelligent transportation systems research based on science mapping, IEEE Trans. Intell. Transp. Syst. 11 (2014) 901–908.

[63] L. Gao-Yong, H. Ji-Ming, W. Hui-Ling, A co-word analysis of digital library field in china, Scientometrics 91 (2012) 203–217.

[64] M.-H. Huang, C.-P. Chang, Detecting research fronts in OLED field using bibliographic coupling with sliding window, Scientometrics 98 (2014).

[65] A.G. López-Herrera, E. Herrera-Viedma, M.J. Cobo, M.A. Martínez, G. Kou, Y. Shi, A conceptual snapshot of the first decade (2002-2011) of the international journal of information technology & decision making, Int. J. Inf. Technol. Decis. Making 11 (2012) 247–270.

[66] E. Murgado-Armenteros, M. Gutiérrez-Salcedo, F. Torres-Ruiz, M.J. Cobo, Analysing the conceptual evolution of qualitative marketing research through science mapping analysis, Scientometrics 102 (2015) 519–557.

[67] A. Rodríguez-Ledesma, M. Cobo, C. López-Pujalte, E. Herrera-Viedma, An overview of animal science research 1945–2011 through science mapping analysis, J. Anim. Breed. Genet. (2014) in press, http://dx.doi.org/10.1111/jbg.12124.

[68] L. Tang, P. Shapira, China-us scientific collaboration in nanotechnology: patterns and dynamics, Scientometrics 88 (2011) 1–16.

[69] M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, F. Herrera, Science mapping software tools: review, analysis and cooperative study among tools, J. Am. Soc. Inf. Sci. Technol. 62 (2011) 1382–1402.

[70] M. Callon, J.P. Courtial, W.A. Turner, S. Bauin, From translations to problematic networks: an introduction to co-word analysis, Soc. Sci. Inf. 22 (1983) 191–235.

[71] N. Coulter, I. Monarch, S. Konda, Software engineering as seen through its research literature: a study in co-word analysis, J. Am. Soc. Inf. Sci. 49 (1998) 1206–1223.

[72] M. Callon, J. Courtial, F. Laville, Co-word analysis as a tool for describing the network of interactions between basic and technological research – the case of polymer chemistry, Scientometrics 22 (1991) 155–205.

[73] Q. He, Knowledge discovery through co-word analysis, Libr. Trends 48 (1999) 133–159.

[74] C. Sternitzke, I. Bergmann, Similarity measures for document mapping: a comparative study on the level of an individual scientist, Scientometrics 78 (2009) 113–130.