Aslib Journal of Information Management

Global ontology research progress: a bibliometric analysis
Qiaoli Zhu, Xuesong Kong, Song Hong, Junli Li, Zongyi He,

Global ontology research progress: a bibliometric analysis

Qiaoli Zhu, Xuesong Kong, Song Hong, Junli Li and Zongyi He
School of Resource and Environmental Science, Wuhan University, Wuhan, China

Received 15 May 2014
Revised 19 September 2014
Accepted 3 November 2014

Abstract

Purpose – The purpose of this paper is to analyse the global scientific outputs of ontology research, an important emerging discipline that has huge potential to improve information understanding, organization, and management.

Design/methodology/approach – This study collected literature published during 1900-2012 from the Web of Science database. The bibliometric analysis was performed from authorial, institutional, national, spatiotemporal, and topical aspects. Basic statistical analysis, visualization of geographic distribution, co-word analysis, and a new index were applied to the selected data.

Findings – Characteristics of publication outputs suggested that ontology research has entered into the soaring stage, along with increased participation and collaboration. The authors identified the leading authors, institutions, nations, and articles in ontology research. Authors were more from North America, Europe, and East Asia. The USA took the lead, while China grew fastest. Four major categories of frequently used keywords were identified: applications in Semantic Web, applications in bioinformatics, philosophy theories, and common supporting technology. Semantic Web research played a core role, and gene ontology study was well-developed. The study focus of ontology has shifted from philosophy to information science.

Originality/value – This is the first study to quantify global research patterns and trends in ontology, which might provide a potential guide for the future research. The new index provides an alternative way to evaluate the multidisciplinary influence of researchers.

Keywords Research trend, Bibliometrics, Ontology, Disciplinary incidence index (DII), Scientific outputs

Paper type Research paper

Introduction

Ontology is a notion originated from philosophy, mainly defined as a theory describing the construction of the world. Even as a pure philosophical concept, it was elected for scientific research as the theoretical base (Chidamber and Kemerer, 1994). Artificial intelligence (AI), basically, can be regarded as the entrance to information science community for ontology. Although the initial use of ontology in AI was merely as an ordinary term (McCarthy, 1980) without scientific description, from then on, the term ontology obtained its orientation in information science that the study of ontology could be of potential in information and knowledge domains (Guarino, 1995, 1998). Ontology was an emerging theory in information science which was responsible for long period of arguments before its maturity, such as the various definitions (Neches *et al.*, 1991; Gruber, 1993; Uschold and Gruninger, 1996; Studer *et al.*, 1998). Among these, Gruber's short definition "An ontology is an explicit specification of a conceptualization" (Gruber, 1993), was commonly accepted and sometimes enriched by subsequent researchers (Borst, 1997; Studer *et al.*, 1998). The rules of ontology construction were also discussed in that stage (Gruber, 1995; Guarino and Welty, 2000). Along with these debates, research on ontology was becoming increasingly prevalent in practice,

This study was funded by National Natural Science Foundation of China (No. 41071290).

Aslib Journal of Information Management
Vol. 67 No. 1, 2015
pp. 27-54
© Emerald Group Publishing Limited
2050-3806
DOI 10.1108/AJIM-05-2014-0061

e.g. knowledge engineering (Guarino, 1997; vanHeijst *et al.*, 1997), natural language processing (Lang, 1991; Bateman, 1995; Bateman *et al.*, 2010), knowledge representation (Guarino, 1995; Sowa, 2000), information retrieval (McGuinness, 1998; Li *et al.*, 2008), and knowledge management (Kietz *et al.*, 2000; Holsapple and Joshi, 2004; Brandt *et al.*, 2008). Over the recent two decades, scientific articles on ontology research have demonstrated an expeditious increase in quantity, and the reason for ontology research obtaining increasing scientific attention is the need of effective methods for communication both from human and computer perspectives (Gruber, 1995; Studer *et al.*, 1998; Ding, 2001). Our world is becoming more connected especially with the development of the Internet, but also accompanying with an information deluge (Bawden and Robinson, 2008). Ontology has the huge potential to help people and computers to extract the information they need and to communicate effectively with each other (Ding and Foo, 2002; Myrgioti *et al.*, 2009). It has been crucial since the Semantic Web (Berners-Lee *et al.*, 2001) was proposed as the next wave of web transformation, because of enabling content-based access, interoperability and communication across the Web (Halpin and Presutti, 2011). In addition, with the exponential growth of accessible biological information, ontology research in bioinformatics has aroused great interest of scholars (Ashburner *et al.*, 2001). The gene ontology (GO) is an extremely important tool for the unification of biology, including three independent ontologies: biological process, molecular function, and cellular component (Ashburner *et al.*, 2000, 2001).

Despite growing popularity, there have been few attempts to gather systematic data on the global scientific production of ontology research. Although ontology-related research as an addition to the Semantic Web (SW) has been studied with bibliometric techniques (Ding, 2010), this analysis was confined by the most productive authors, highly cited journals, authors, and papers in the SW field over 1960-2009. The number of citations was the only evaluating indicator used, and it is not comprehensive to provide the specific patterns of ontology research. Likewise, synonymy and homonymy problems about the authors were not expressed. Ontology research is burgeoning and widespread in the discipline of information, which has particularly close ties to current society. Besides statistical analysis of professional aspects, it is essential to show readers the basic development situation, providing an easy access to understanding ontology research. Recently, the bibliometric method, as a common and effective tool, has already been widely applied for scientific production and research trend analysis in various fields (van Raan *et al.*, 2003; Leydesdorff and Wagner, 2009; Lariviere *et al.*, 2012). The conventional bibliometric methods generally evaluate research trends by analysing the publication outputs, citation times, and keyword frequencies (van Raan, 2008; Bornmann *et al.*, 2012; Costas and Bordons, 2008). In the meantime, a number of innovative techniques have been excogitated, such as citation structure analysis (Small and Upham, 2009), base maps and overlay techniques (Leydesdorff *et al.*, 2012), and network analysis (Waltman *et al.*, 2010). Moreover, in addition to straightforward counting, various indicators have been established to measure the scientific performance from different perspectives, such as the *h*-index (Hirsch, 2005), the Eigenfactor (Bergstrom *et al.*, 2008), the Audience factor (Zitt and Small, 2008), and the SNIP (Moed, 2010). As an important emerging discipline and a complex multidisciplinary field, we think the interdisciplinary research of ontology could be interesting and meaningful. Interdisciplinary research has variant name forms (multi, cross, transdisciplinary) and can be approached from different perspectives (Bordons *et al.*, 2005). Hence, various bibliometric indicators have been proposed to facilitate it, based on different units of analysis. For example, diversity and network coherence indicators based on references patterns (Rafols and Meyer, 2010), an indicator for interdisciplinary on the basis of aggregated journal-journal citations (Leydesdorff *et al.*, 2013), and "diffusion score" in terms of citing patterns to publications (Carley and Porter, 2012). Knowledge integration and/or diffusion can be two key concepts of these investigations. Because in many works, the central point of "interdisciplinary" is "integration" (Rafols and Meyer, 2010), we use "multidisciplinary" to indicate knowledge diffusion in our study. To a particular paper, the number of research areas covered by its citations is also an alternative indicator, which can be used to reflect the multidisciplinary influence. The number may vary according to different papers and authors, that is, the multidisciplinary impact of researchers may be distinct. Hence, for pilot exploration, we established a new index, the disciplinary incidence index (DII), to measure the multidisciplinary influence of researchers during a certain period. DII is an indicator to reflect how many research areas on average cite per article of the author, on the other a relatively simple metric can aid in multidisciplinary research assessment. It will be further explained in the "Materials and methods" section.

In this study, we conducted a comprehensive bibliometric analysis of published ontology-related research from 1900 to 2012 by combining the traditional bibliometric methods and the new index. To provide an overall background, the basic situation of ontology research in scientific community was first statistically analysed. For professional knowledge, we supplied previous bibliometric studies of ontology research with more detailed research patterns, by incorporating the author keywords as smaller unit of analysis. The purpose of this study was to reveal the research progress and patterns in ontology research from time to space, micro to macro, general to particular, and provide potential directions for future development of ontology research.

Materials and methods

Data sources

The Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) bibliographic databases were chosen as the data source in this study. We collected literatures related to ontology research from the online version of the SCI and SSCI accessed via Web of Science (WOS). Based on previous related researches (Ding, 2010), the term "ontolog*" (including any word that begins with "ontolog", such as "ontology", "ontologies", "ontologic", "ontological", "ontologically", and "ontologist") was searched in the bibliographic field of "Topic". This searching strategy allowed us to locate publications that contain these words in their titles, abstracts, or keywords (including author keywords and keywords plus). The time span was set to 1900-2012 so that our analysis could cover almost all the ontology-related publications in the databases. The bibliographic search and information retrieving were performed on 1 July 2013. We obtained a total of 20,185 publications and corresponding document information from the SCI and SSCI databases.

Statistical indicators and bibliometric methods

We conducted the data analyses from two aspects: basic situation and professional knowledge. We identified basic situation as properties irrelevant to textual content of papers, including document types and publishing languages, characteristics of publication outputs, author performance, institutional and international collaborations. Professional knowledge analysis comprised research emphasis and trends, and frequently cited articles. The data were mainly organized and analysed by Microsoft Excel.

Some basic statistical indicators were ordinarily calculated in our analysis, such as annual outputs of papers, average number of authors, citations, references, and pages in a paper. Table I is a summary list of statistical units and bibliometric methods used in our study, as well as corresponding goals.

Authors were identified with their institutes to tackle synonymy and homonymy problems. We manually checked the publications of cognominal authors, and then removed the ones with different author full names or institutes. For accurate evaluations, an individual's role in his or her publications can be reflected through the articles published by him or her as first or corresponding author (FCA). The local citation score (LCS) (Garfield *et al.*, 2003) and *h*-index were used to indicate the author's academic influence in ontology field. The new index DII was applied to reflect the author's academic influence in multifarious research areas. These descriptors were directly or indirectly related to citations and easy access. CiteSpace (Chen, 2004) was used to geocode the affiliations of authors and plot the geographic distribution of authors. Normally, publications originating from England, Scotland, North Ireland, and Wales were uniformly labelled as documents from the UK, and publications from Mainland China, Hong Kong, and Taiwan were treated separately (Chiu and Ho, 2007; Liu *et al.*, 2012).

Collaboration analysis comprised authorial, institutional, and national levels based on the complete count strategy, implying each signatory on individual publications was treated equally. Accordingly, publications with only one signatory author, institution, and country/territory were grouped under the "single" heading. "Inter-institution collaboration" and "international collaboration" were defined as publications by more than one institution, and country/territory, respectively.

Author keywords were provided by articles' authors as fundamental parts of the articles, through which the readers could roughly know about the referred fields and

Analytical perspectives	Analytical units	Major bibliometric methods	Goals
Basic situation	Document type Publishing language	Straightforward counting	To identify conventional document types and languages of ontology-related publications
	Annual publication output Scientific productivity descriptors of article	Time-trend analysis	To study the overall evolvement of publication outputs and detailed characteristics changes of articles over time
	Author	Academic evaluation and spatial analysis	To identify top authors and global geographic distribution of authors
	Institution	Research output analysis	To study the collaborative modes in institution level
	Country/territory	Research output and time-trend analysis	To study the collaborative modes in country level and temporal variation of countries outputs
Professional knowledge	Author keyword	Co-word analysis and time-trend analysis	To investigate the research emphasis and patterns, as well as topics changes over time
	Frequently cited article	Academic evaluation	To identify and introduce the influential papers

Table I.
Bibliometric data analysis schema

contents of the articles. In our study, these words are more precise than keywords plus, which are derived from the titles of articles cited in the documents being indexed (definition in WOS). Hence we employed the former to gain insights about the emphasis and trends of ontological research. Hot issues were the focus of this part, so we chose the tops as the target of analysis. Procedures of co-word analysis in our study included keyword clustering and strategic analysis (Callon *et al.*, 1991), which were used to reveal the ontology research patterns and trends. The co-word matrix was built based on the word co-occurrence, as the original processing object for further analysis. The cell value of the matrix is equal to the times two words both appear in the same paper. High co-occurrence frequency of two words suggests a close connection between them (Ding *et al.*, 2001). Original co-word matrix was exported in network file format by UCINet that can be read by VOSviewer (van Eck and Waltman, 2010). VOSviewer is free software combining visualization and clustering techniques. We chose the density view map, one of the four options offered by VOSviewer, to display the 50 hot issues. This type of visualization permits quick and easy identification of the clustering pattern of hot issues. In it, keywords are located in areas with different background colour, which represents the density (depends on the total number of occurrences or co-occurrences) of keywords. The distance between two labels is inversely proportional to the number of co-occurrence between individual keywords. For strategic analysis, a more exact classification of clusters is prerequisite. In line with the clustering method used by Ding *et al.* (2001), the original co-word matrix was transformed into a Pearson-normalized correlation matrix and then clustered by using hierarchical clustering techniques with Ward's method in SPSS. Strategic coordinate and graph (Callon *et al.*, 1991) were obtained on the basis of clustering results. Additionally, we divided the whole period and presented the top 25 author keywords within each interval, to trace the dynamic changes of research focuses in ontology field.

After the general analysis, we transferred our focus to especial individuals, which were defined as highly cited articles. Citations per publication can be used to evaluate the influence of one paper to the entire field relatively reliably (Herbertz and MullerHill, 1995; Riikonen and Vihinen, 2008).

The DII

In WOS, research area terms[1] are article-based and assigned to a record by the Institute for Scientific Information (ISI). Based on these predefined categories, the citing research areas (CRA) are registered by all research area terms of the sets that cite a given article. The CRA can be obtained very easily by using the results analysis tool in the ISI WOS database. The DII is defined as the average number of CRA per article of a given researcher, and it can be represented as:

$$\mathrm{DII} = \frac{\sum_{i=1}^{\mathrm{TP}} \mathrm{CRA}_i}{\mathrm{TP}}$$

where CRA_i is the number of CRA of the ith article for a given researcher, TP is the total published articles of the researcher. The maximal value of DII is the total number of ISI defined research areas, and the DII is equal to 1 when each article was cited on average by one research area. Regarding the degree of general applicability, the variables of DII (e.g. CRA and TP) are non-special and easily retrievable, thus it could be utilized for other disciplines.

Results and discussion

Document types and publishing languages

According to the classification of document type identified by ISI, the total 20,185 publications distributed in 18 document types[2]. Naturally, the most common document type was peer-reviewed journal articles (14,709), which comprised 72.87 per cent of the total publications. Proceedings papers (3,565; 17.66 per cent), reviews (768; 3.80 per cent), editorial materials (412; 2.04 per cent), and book reviews (399; 1.98 per cent) also accounted for a relatively higher proportion of the total. Others showing less significance were meeting abstracts (155), letters (48), discussions (32), notes (27), corrections (22), software reviews (21), database reviews (nine), biographical-items (five), reprints (five), corrections/additions (three), news items (three), hardware reviews (one), and items about individuals (one). The number in parentheses represents the quantity of each individual document type. Original and peer-reviewed articles were still the focus of conventional bibliometric analysis, but considering the significant contribution of proceedings papers in our analysis, which comprised nearly the same amount with journal articles during the period 2002-2006 (see Figure 1(a)). Therefore we focused our further analysis on peer-reviewed journal articles and proceedings papers, which were uniformly denoted by articles or papers, and publications of all other types were excluded from further analysis.

In terms of publishing language, English was the predominant language accounting for 17,789 or 97.35 per cent of the 18,274 articles, which reflected the fact that English is the dominant academic language, but also might be due to the preference of the SCI and SSCI databases for English language journals (Seglen, 1997). Other minor publication languages in the list were Spanish (96), German (95), Czech (85), French (62), Russian (52), Portuguese (38), Chinese (12), Turkish (seven), Italian (six), Slovak (five), Lithuanian (five), Norwegian (three), Swedish (three), Japanese (two), Croatian (two), Slovene (two), Korean (two), Polish (two), Afrikaans (two), Danish (two), Rumanian (one), and Serbian (one).

Characteristics of publication outputs

Although the earliest ontology-related publication in the SCI and SSCI databases was published in German in 1909, a notable growth did not appear until the 1990s (see Figure 1(a)). As displayed in Figure 1(a), according to the annual change of publication productivity, the evolvement of ontology-related research can be divided into three gradations: first, the enlightenment stage (1909-1990). In this stage, the research productivity kept a very low level, with the average annual number of SCI/SSCI publications being only 7.51. We also defined this stage as the transition period during which the philosophy concept was initially introduced into information science. Second, the growth stage (1991-2000). The related research started to go up significantly in this decade, with the annual publications increasing from 58 in 1991 to 291 in 2000. As a newly emerging concept in information science deriving from philosophy, various definitions of the term "ontology" appeared in this stage, after the first definition of ontology in information science given by Neches *et al.* (1991). Ontology is the backbone of the Semantic Web. Late this period, the European Union and the USA secured these important innovations with significant funding (Shadbolt *et al.*, 2006; Ding, 2010). Along with the groundbreaking progress in Semantic Web, ontological research began to attract more attentions from various fields. Finally, the soaring stage (2001-2006, 2007-2012). Based on the previous work, the ontology-related research rocketed in this stage, except for a decrease of publications from 1,717 in 2006 to 1,420 in 2007.

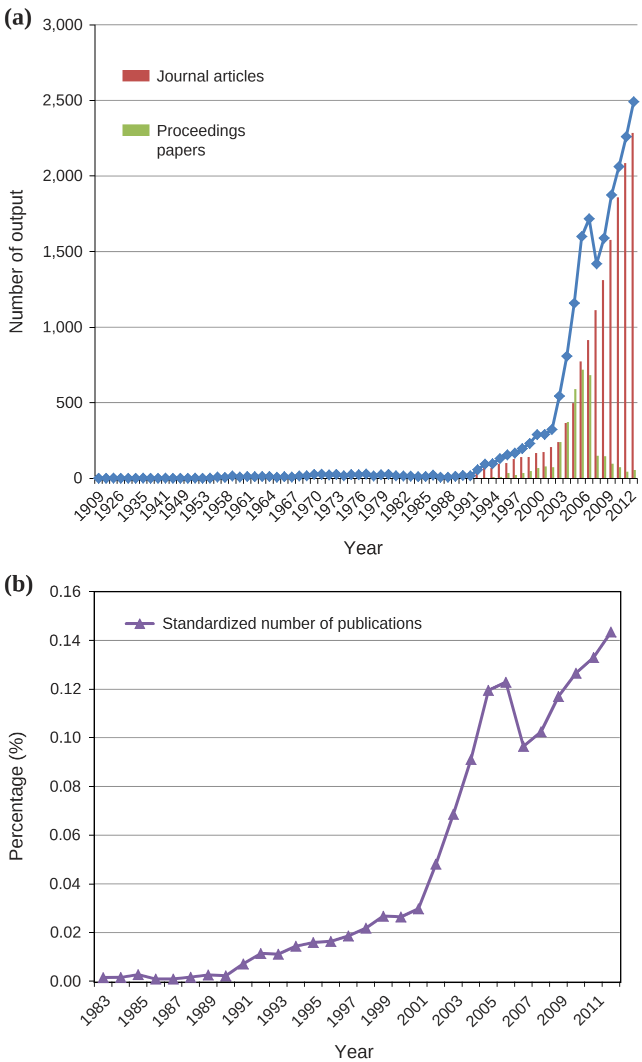

Figure 1.
Annual number and standardized number of publication outputs on ontology

We ascribed this decline to the sharp decrease of proceedings papers in 2007 (see Figure 1(a)). Since 2007, all major computer science conference proceedings have been transferred from WOS to the ISI proceedings which are not part of WOS anymore (Ding, 2010). Consequently, this stage was divided into two periods, 2001-2006 and 2007-2012. The global publication outputs in ontological research increased at the annual average growth of approximately 278.6 and 214.4 for these two periods, respectively. In order to evaluate the research interest in ontological studies more objectively, we similarly counted the standardized number of publications (Wang *et al.*, 2012; Zhuang *et al.*, 2013) on ontology. It was defined as the ratio of the annual number of ontology-related publications to the annual number of publications in the SCI and SSCI databases, and displayed the result (see Figure 1(b)).

The overall growth trend of the standardized number also suggested a pure research interest in ontological studies.

Specially, the temporal evolution of ontology-related research papers for the period of 1991-2012 is shown in Table II. The annual number of articles increased from 45 in 1991 to 2,343 in 2012, but the average length of the individual articles basically presented downtrend with small fluctuations, which decreased 2.74 pages from 1991 to 2012. In general, the expanding accumulation of knowledge with the growth of time existed in all scientific fields, which could be indicated by the increasing number of references per article (Liu *et al.*, 2011; Wang *et al.*, 2012). In ontological research field, the average number of references grew from 31.58 in 1991 to 45.31 in 2012. The collaboration index, which was defined as the mean number of authors per article, grew from 1.49 in 1991 to 4.87 in 2012. This growing collaboration index indicated an increasing cooperation in ontological research field. It also can be observed from other bibliometric studies (Xie *et al.*, 2008; Wang *et al.*, 2012; Liu *et al.*, 2012) that the scientific community has becoming more collaborative. In addition, the number of countries and institutions participating in the ontological research increased from 13 in 1991 to 76 in 2012 and 48 in 1991 to 2,376 in 2012, respectively, which partly reflected an expanding global interest in ontological field.

Author performance

The author performance in our study comprised the productivity and geographic distribution of authors. Among the 18,274 articles for further analysis, four articles were anonymous, thus the remaining 18,270 articles were signed by 44,652 authors. The result of author productivity analysis was consistent with normal performance that, productive authors account for a small part of total authors but have contributed a substantial number of articles on ontology. Of the 44,652 authors, 33,187 or 74.32 per cent (co)authored only one ontological paper, 42,766 or 95.78 per cent contributed less than five papers, while the top 1,886 or 4.22 per cent authors with at least five papers produced 7,764 or 42.49 per cent of the total articles (TA). We listed 20 the most productive authors with their research institutes in Table III, together with the TA, collaborated articles (CA), FCA, and some descriptors used to measure the academic impact of authors such as the total citations (TC), average citations (TC/TA), total local citation score (TLCS), the *h*-index, and DII.

The most productive authors in ontology research were Smith B from State University of New York (SUNY) Buffalo and Musen MA from Stanford University with 59 and 47 articles, respectively. Other prolific authors included Blake JA from Jackson Laboratory with 42 papers, Jung JJ from Yeungnam University with 40, and Staab S from University of Koblenz-Landau with 47. Especially, Jung JJ was the one who produced the most papers as first author or corresponding author, though he ranked fourth in the total number of articles. Out of his 40 papers, only 13 were CA, while the cooperation rates of other prolific authors were all more than 80 per cent. As for academic impact, Blake JA ranked first in TC (3,665) and second in *h*-index (22), with the average citations of up to 88.26, which indicated Blake JA had more high quality articles in ontology research, as well as Chou KC from Gordon Life Science Institute, whose *h*-index ranked first (29). Mungall C from University of California Berkeley got the highest TLCS (1,095), which meant his articles were cited by the most ontology-related papers. By analysing the CRAs, the DII of Lewis S from Lawrence Berkeley National Laboratory (LBNL) (23.90) was the largest, which indicated Lewis S carried a greater multidisciplinary influence than other authors,

Table II. Characteristics of scientific articles on ontology from 1991 to 2012

PY	TA	AU	AU/TA	NR	NR/TA	PG	PG/TA	TC	TC/TA	Ins	Cou
1991	45	67	1.49	1,421	31.58	755	16.78	619	13.76	48	13
1992	75	102	1.36	2,592	34.56	1,185	15.80	961	12.81	77	20
1993	76	110	1.45	2,608	34.32	1,382	18.18	3,980	52.37	81	21
1994	105	164	1.56	3,513	33.46	1,669	15.90	1,867	17.78	103	23
1995	136	221	1.63	4,503	33.11	2,228	16.38	3,369	24.77	140	28
1996	151	261	1.73	5,599	37.08	2,892	19.15	3,222	21.34	149	27
1997	176	277	1.57	5,672	32.23	2,972	16.89	2,793	15.87	172	29
1998	189	342	1.81	6,577	34.80	3,255	17.22	3,185	16.85	188	33
1999	238	482	2.03	8,509	35.75	4,063	17.07	3,939	16.55	246	32
2000	253	569	2.25	8,473	33.49	4,216	16.66	3,871	15.30	289	37
2001	280	617	2.20	8,974	32.05	4,364	15.59	4,830	17.25	300	41
2002	481	1,573	3.27	14,062	29.23	7,079	14.72	8,734	18.16	492	43
2003	740	2,475	3.34	19,349	26.15	10,339	13.97	18,295	24.72	693	50
2004	1,087	3,922	3.61	28,694	26.40	14,040	12.92	22,187	20.41	984	51
2005	1,493	5,660	3.79	42,181	28.25	19,534	13.08	28,080	18.81	1,272	60
2006	1,597	6,292	3.94	47,318	29.63	21,072	13.19	24,637	15.43	1,431	64
2007	1,263	5,476	4.34	49,095	38.87	17,708	14.02	24,281	19.22	1,372	64
2008	1,458	6,333	4.34	58,350	40.02	21,185	14.53	21,744	14.91	1,520	65
2009	1,676	7,298	4.35	68,899	41.11	23,613	14.09	20,522	12.24	1,711	64
2010	1,932	8,653	4.48	83,194	43.06	27,792	14.39	13,490	6.98	2,001	66
2011	2,131	9,920	4.66	96,749	45.40	29,503	13.84	9,117	4.28	2,242	79
2012	2,343	11,414	4.87	106,167	45.31	32,881	14.03	3,074	1.31	2,376	76
Total	17,925	72,228		672,499		253,727		226,797			

Notes: TA, number of articles; AU, number of authors; NR, cited reference count; PG, page count; TC, total citations; Ins, number of institutes; Cou, number of countries; AU/TA, NR/TA, PG/TA, and TC/TA, average of authors, references, pages, and citations in an article

Table III.
Top 20 most productive authors in ontology research

Author/research institute	TA	CA(%)	FCA(R)	TC(R)	Descriptors of academic impact TC/TA(R)	TLCS(R)	*h*-index(R)	DII(R)
Smith, B./State University of New York (SUNY) Buffalo	59	54(91.53)	23(2)	1,578(6)	26.75(8)	577(4)	20(3)	11.08(6)
Musen, M.A./Stanford University	47	43(91.49)	9(10)	1,319(7)	28.06(7)	419(5)	19(4)	10.28(7)
Blake, J.A./Jackson Laboratory	42	42(100.0)	5(15)	3,707(1)	88.26(3)	883(3)	22(2)	17.86(4)
Jung, J.J./Yeungnam University	40	13(32.50)	37(1)	554(14)	13.85(14)	133(15)	15(8)	3.25(20)
Staab, S./University of Koblenz-Landau	39	39(100.0)	4(17)	1,002(9)	25.69(9)	343(8)	15(8)	6.15(11)
Stevens, R./University of Manchester	38	37(97.37)	6(12)	942(10)	24.79(10)	382(7)	13(11)	9.95(8)
Mizoguchi, R./Osaka University	37	33(89.19)	5(15)	424(15)	11.46(15)	155(13)	10(17)	4.57(15)
Horrocks, I./University of Oxford	37	30(81.08)	13(8)	1,065(8)	28.78(6)	414(6)	15(8)	7.30(9)
Chou, K.C./Gordon Life Science Institute	35	35(100.0)	20(3)	2,605(4)	74.43(4)	288(9)	29(1)	14.49(5)
Fernandez-Breis, J.T./University of Murcia	35	35(100.0)	14(6)	242(17)	6.91(18)	17(19)	11(14)	4.14(16)
Mungall, C./University of California Berkeley	35	34(97.14)	6(12)	3,673(2)	104.94(2)	1,095(1)	18(6)	20.34(2)
Gomez-Perez, A./Polytechnic University of Madrid	35	33(94.29)	7(11)	572(13)	16.34(13)	181(12)	11(14)	5.09(14)
Studer, R./Karlsruhe Institute of Technology	32	32(100.0)	1(19)	651(11)	20.34(12)	233(10)	13(11)	6.28(10)
Dopazo, J./Centro de Investigación Príncipe Felipe	32	31(96.88)	16(5)	2,046(5)	63.94(5)	152(14)	16(7)	18.56(3)
Motta, E/Open University – UK	31	31(100.0)	3(18)	352(16)	11.35(16)	99(16)	11(14)	4.03(17)
Lewis, S./Lawrence Berkeley National Laboratory (LBNL)	30	30(100.0)	1(19)	3,659(3)	121.97(1)	1,071(2)	19(4)	23.90(1)
Sure, Y./Karlsruhe Institute of Technology	29	29(100.0)	6(12)	618(12)	21.31(11)	208(11)	13(11)	5.83(12)
Valencia-Garcia, R./University of Murcia	29	29(100.0)	12(9)	173(20)	5.97(20)	5(20)	7(19)	3.69(19)
Sanchez, D./Universitat Rovira i Virgili	28	27(96.43)	19(4)	196(18)	7.00(17)	94(17)	9(18)	3.86(18)
Schulz, S./University of Freiburg	28	27(96.43)	14(6)	183(19)	6.54(19)	85(18)	7(19)	5.11(13)

Notes: TA, total articles; CA(%), collaborated articles (percentage of collaborated articles to total articles); FCA, articles published as the first author of the corresponding author; TC, total citations; TC/TA, average citations; TLCS, total local citation score; *h*-index, *h*-index in ontology field; DII, disciplinary incidence index; R, rank in the list followed by Mungall C with 20.34, and Dopazo J from Centro de Investigación Principe Felipe with 18.56.

The global geographic distribution of authors was plotted according to the author affiliations (see Figure 2). We could identify the major spatial clusters of authors in North America, Europe, and East Asia and several minor clusters in other areas. As the background, we took a choropleth map based on the research and development (R&D) Expenditure of different countries for the year 2009, the latest data from the World Bank WDI Database by the time we conducted our research. The distribution of denser authors was consistent with the R&D expenditure that the major clusters overlapped with high investment regions.

Institutional and international collaborations

Collaboration plays an ever growing role in ontology-related research, which can be reflected not only by the rising average number of authors per article, but also by the cooperation between institutions or countries. We analysed the institutional and international collaborations based on the full affiliation information of authors. There were 350 articles without any author address information in our database, and the remaining 17,924 papers covered 8,339 different institutions and 112 different countries/territories. Of the 17,924 papers, 14,381 or 80.23 per cent were nationally independent while only 9,490 or 52.95 per cent were single-institution papers, which meant that 4,891 of all single-country articles were nationally inter-institutional collaborative works. Although both independent and collaborative articles showed an ascending trend along with the temporal evolution, the percentages of single-institution articles and single-country articles to the total number of papers on ontology decreased from 80.00 per cent in 1991 to 43.70 per cent in 2012 and 91.11 per cent in 1991 to 75.25 per cent in 2012, respectively. In contrast, the proportions of collaborative papers increased steadily (see Figures 3 and 4). The greater change in the proportion of inter-institutional collaborative articles suggested that inter-institutional collaboration was more prevalent than international collaboration. Especially, the number of inter-institutional collaborative papers exceeded the amount of single-institution articles in 2007 (see Figure 3).

Among the referred 8,339 different institutions that contributed to ontology-related papers, 6,696 or 80.30 per cent published more inter-institutional collaborative articles

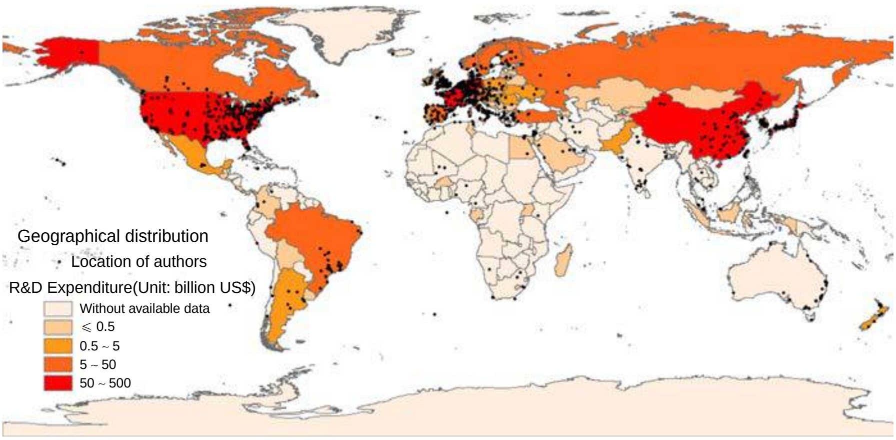

Figure 2. Compare global geographical distribution of authors with research and development expenditure of individual countries in 2009

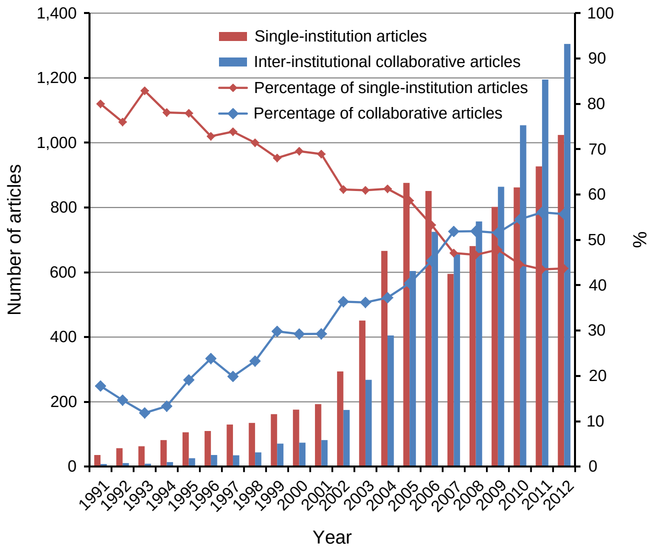

Figure 3.
Trend of inter-institutional collaborative and single-institution articles in ontology research

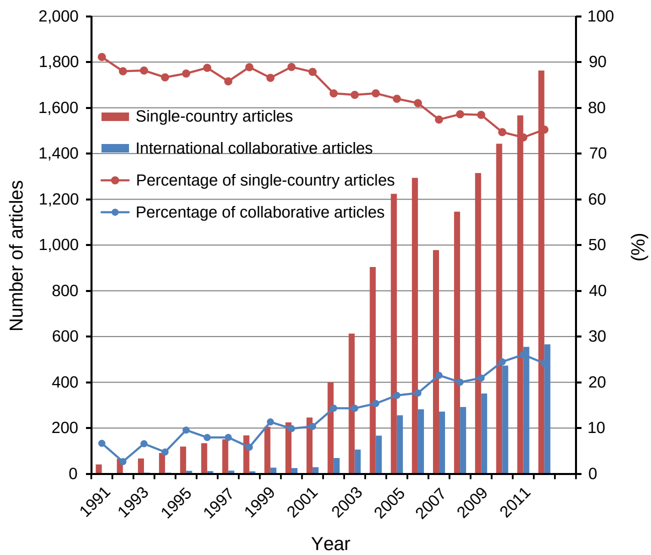

Figure 4.
Trend of international collaborative and single-country articles in ontology research

than single-institution ones. Table IV summarizes the top 30 most productive institutes, in terms of the number of articles and citations. The USA, with the most institutes (13) in the list, again demonstrated its dominance in ontology research. Six of these 30 institutes were located in the UK and the University of Manchester in UK led institutional productivity with 254 articles. Among the remaining 11 institutes, three were in China, two were in Canada, and other six institutes were from Australia, Germany, Italy, the Netherlands, South Korea, and Singapore, respectively. The Stanford University in USA and the Chinese Academy of Sciences in China, the

Table IV. Top 30 most productive institutes in ontology research

Institution	TA	Av.TC	SI	Single-institution Av.TC	%SI	CI	Inter-institutional Av.TC	%CI
Univ Manchester, UK	254	19.20	102	17.33	40.16	152	20.45	59.84
Stanford Univ, USA	246	39.56	83	40.42	33.74	163	39.12	66.26
Chinese Acad Sci, China	232	8.00	53	7.98	22.84	179	8.00	77.16
Harvard Univ, USA	192	33.24	24	19.50	12.50	168	35.21	87.50
Univ Cambridge, UK	167	35.28	33	12.42	19.76	134	40.91	80.24
Univ Toronto, Canada	161	24.71	48	11.46	29.81	113	30.34	70.19
Univ Washington, USA	142	28.42	44	20.64	30.99	98	31.92	69.01
Univ Queensland, Australia	126	27.01	38	10.03	30.16	88	34.34	69.84
Univ British Columbia, Canada	122	22.52	38	11.03	31.15	84	27.73	68.85
CNR, Italy	121	15.40	43	19.28	35.54	78	13.27	64.46
Univ Karlsruhe, Germany	120	18.47	52	18.25	43.33	68	18.63	56.67
Univ Edinburgh, UK	118	18.64	37	5.73	31.36	81	24.53	68.64
Shanghai Jiao Tong Univ, China	117	14.46	54	3.65	46.15	63	23.73	53.85
Univ Calif San Diego, USA	114	37.12	41	19.98	35.96	73	46.75	64.04
Univ Calif Berkeley, USA	114	38.56	27	15.00	23.68	87	45.87	76.32
European Bioinformat Inst, UK	111	48.26	14	46.86	12.61	97	48.46	87.39
Zhejiang Univ, China	109	6.94	45	5.47	41.28	64	7.97	58.72
Univ Illinois, USA	108	11.39	44	9.48	40.74	64	12.70	59.26
SUNY Buffalo, USA	104	16.72	37	13.05	35.58	67	18.75	64.42
Univ Oxford, UK	103	24.40	21	8.48	20.39	82	28.48	79.61
Univ N Carolina, USA	103	10.89	33	13.09	32.04	70	9.86	67.96
Univ Michigan, USA	102	29.06	34	13.85	33.33	68	36.66	66.67
Univ Amsterdam, The Netherlands	99	20.06	37	12.81	37.37	62	24.39	62.63
Univ Maryland, USA	94	18.05	28	22.04	29.79	66	16.36	70.21
Univ Calif Los Angeles, USA	94	25.31	28	20.11	29.79	66	27.52	70.21
Seoul Natl Univ, South Korea	94	6.00	22	3.68	23.40	72	6.71	76.60
Indiana Univ, USA	92	23.45	38	20.97	41.30	54	25.19	58.70
Natl Univ Singapore, Singapore	89	14.13	20	3.80	22.47	69	17.13	77.53
UCL, UK	87	28.86	28	30.11	32.18	59	28.27	67.82
Columbia Univ, USA	86	22.37	24	15.58	27.91	62	25.00	72.09

Note: TA, total articles; Av.TC, average citations per article; SI, single-institution articles; CI, inter-institutional collaborated articles; %SI and %CI, percentage of single-institution articles and inter-institutional collaborated articles to total articles

Harvard University in USA, and the University of Cambridge in the UK, ranked second to fifth, contributing 246, 232, 192, and 167 papers, respectively. The European Bioinformatics Institute in the UK, contributed a correspondingly small number of articles (ranked 15th), had the most average citations per paper (48.26). We conjecture that this highly cited proportion was correlated with the research fields of the institute, because the European Bioinformatics Institute focus on bioinformatics services such as GO, and chemical entities of biological interest, which belong to dynamic research fields having a correspondingly high proportion of citations to recent publications (Seglen, 1997). The same as we observed from internationally collaborative articles, collaborations were positively associated with the citation rate that inter-institutionally CA obtained more citations than those produced by single institutions.

Among the 112 countries/territories participating in ontology research, 23 had no single-country paper and 16 contributed only one single-country paper, while ten countries/territories did not have any internationally CA and 25 produced only one internationally collaborated paper. We further analysed the collaborative situation of the 30 most productive countries/territories, providing the number of TAs and TC for single-country articles and internationally CA, respectively, in Table V. Of these 30 countries, 17 were from Europe, eight were from Asia, two were from North America, two were from Oceania, and one was from South America, which suggested a geographic inequality in ontology research.

The USA headed the productivity ranking of countries/territories, with the most nationally independent (4,126) and internationally collaborative (1,601) articles. The UK and Mainland China ranked the second and the third with 2,477 and 1,483 articles, respectively, followed by Germany (1,381), and Canada (1,007). As revealed by other bibliometric analyses (Tarkowski, 2007; Xie *et al.*, 2008), that the academic outputs were positively correlated with economic developments and academic investment: the seven major industrial countries (G7: Canada, France, Germany, Italy, Japan, the UK, and the USA) were all ranked in our list of top 30 countries/territories, and four developing countries ("BRIC": Brazil, Russia, India, and China) were also among the top list but with much less productivity than G7. We also observed that internationally collaborative articles generally drew more citations than single-country articles, and the average citations per article of developing countries were fewer. In addition, we ranked the 112 countries/territories in terms of their productivity for the periods 1966-1990, 1991-1995, 1996-2000, 2001-2006, and 2007-2012, respectively, and revealed temporal variation of top ten countries/territories (see Figure 5). China emerged to be the fastest growing country, especially in the soaring stage of ontology-related research, rocketed to the third place. The USA and the UK maintained in the top over the five periods, suggesting the leading position in ontology field.

Research emphasis and trends

Since ISI database gathered the author keywords from 1990, there were 9,923 (54.30 per cent) of the total 18,274 papers containing author keywords in the period 1990-2012. Examination of author keywords in our study revealed that the 9,923 articles had 25,411 unique keywords, which appeared 51,927 times. However, among these 25,411 keywords, 20,022 (78.79 per cent) appeared only once, and 24,055 (94.66 per cent) keywords appeared in less than five papers. The large number of once-only author keywords probably indicated a lack of continuity in research and a wide disparity in research focuses (Chuang *et al.*, 2007), or these keywords might not be

Table V. Top 30 most productive countries/territories in ontology research

Country/territory	TA	Av.TC	SA	Single-country Av.TC	%SA	CA	International collaboration Av.TC	%CA
USA	5,727	20.05	4,126	18.65	72.04	1,601	23.65	27.96
UK	2,477	17.39	1,551	12.37	62.62	926	25.80	37.38
Mainland China	1,486	7.41	1,045	4.88	70.32	441	13.39	29.68
Germany	1,381	15.14	791	11.91	57.28	590	19.46	42.72
Canada	1,007	14.37	584	10.01	57.99	423	20.39	42.01
Spain	959	10.13	607	9.95	63.30	352	10.43	36.70
Italy	883	11.01	561	8.38	63.53	322	15.60	36.47
Australia	836	13.63	511	8.54	61.12	325	21.63	38.88
France	762	13.14	437	9.24	57.35	325	18.39	42.65
The Netherlands	613	15.82	328	13.07	53.51	285	18.98	46.49
South Korea	601	7.46	462	4.46	76.87	139	17.42	23.13
Japan	539	17.04	341	11.26	63.27	198	26.98	36.73
Taiwan	377	7.16	309	6.27	81.96	68	11.21	18.04
Brazil	313	6.96	193	4.89	61.66	120	10.31	38.34
Greece	302	6.51	203	5.60	67.22	99	8.38	32.78
Switzerland	268	18.12	96	6.15	35.82	172	24.80	64.18
Sweden	265	16.24	132	12.20	49.81	133	20.25	50.19
Belgium	259	14.53	140	13.69	54.05	119	15.51	45.95
Austria	232	11.70	97	9.59	41.81	135	13.21	58.19
Finland	216	9.29	144	8.13	66.67	72	11.61	33.33
Israel	197	13.12	130	9.15	65.99	67	20.82	34.01
Norway	183	12.35	91	5.81	49.73	92	18.82	50.27
Ireland	182	10.88	68	8.41	37.36	114	12.35	62.64
Denmark	175	19.25	93	10.82	53.14	82	28.82	46.86
Singapore	173	17.51	75	9.51	43.35	98	23.64	56.65
India	167	5.87	109	4.09	65.27	58	9.21	34.73
Poland	156	9.64	93	6.94	59.62	63	13.63	40.38
Hong Kong	146	5.13	55	3.51	37.67	91	6.11	62.33
New Zealand	130	11.35	68	10.74	52.31	62	12.03	47.69
Russia	125	11.03	82	2.63	65.60	43	27.05	34.40

Notes: TA, total articles; Av.TC, average citations per article; SA, single-country articles; CA, internationally collaborated articles; %SA and %CA, percentage of single-country articles and internationally collaborated articles to total articles

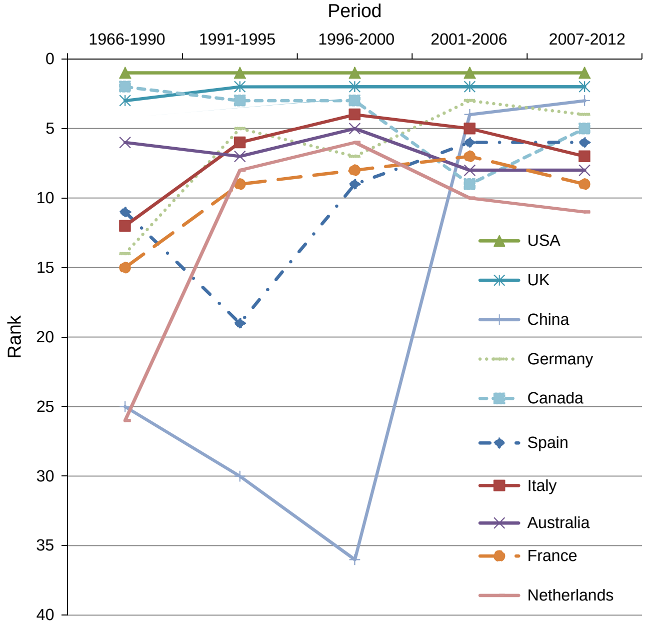

Figure 5.
Ranking variation of top ten countries for five periods

widely accepted by researchers (Ugolini *et al.*, 2001). According to the ranking of author keywords based on occurrence number, the 50th keyword appeared 49 times among the 9,923 papers, which was comparatively low. Therefore, these top 50 author keywords were adequate to reflect the hotspots in ontology research. In this paper, we counted and analysed the co-occurrences of the 50 most frequently used author keywords, to reveal patterns and trends of ontology-related research. Based on the available co-occurrence matrix of 50 × 50 keywords, a density view map was presented (see Figure 6). We identified four major clusters of author keywords from the general structure of the map, which could be summarized as four categories: applications in Semantic Web, applications in bioinformatics, philosophy theories, and common supporting technology. As can be seen, the densest area was labelled “ontology”, which was responsible for the most co-occurrence times (1,169) with the other 49 keywords as one of our search terms. As a philosophy concept being introduced into information science field, “ontology” has two definitions, philosophic and information scientific. Therefore, it co-occurred with almost every keyword else and located in the middle part of the map, and we did not relegate it to the four categories. The second densest area was predominated by “semantic web” and “ontologies”, which were the centralities of the cluster of applications in Semantic Web. The applications in bioinformatics and philosophy theories pivoted around “genomics” and “epistemology”, respectively. Common supporting technologies including “clustering”, “annotation”, “semantic similarity”, and “data mining” were relatively dispersed, because these common techniques were used in both Semantic Web and bioinformatics.

For further analysis, the universal keyword “ontology” and some keywords such as “development” and “knowledge”, whose meanings were too generalized to reflect the research hotspots, were excluded. Finally, we got nine clusters and corresponding strategic coordinates (see Table VI), as well as the strategic graph (see Figure 7). Only

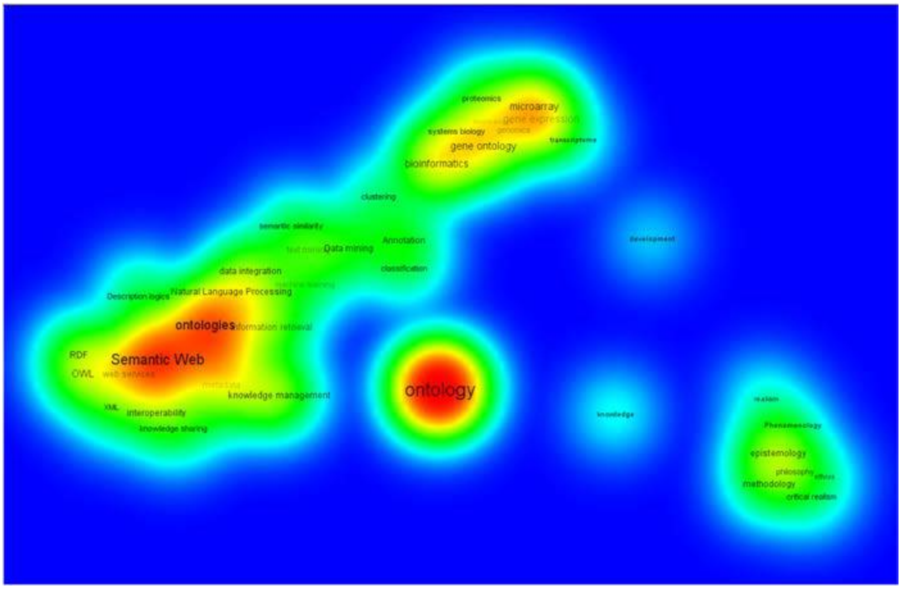

Figure 6. Clustering pattern of the 50 most frequently used author keywords in ontology research during 1990-2012

Cluster	Author keywords	Centrality	Density	Strategic coordinate
A	Gene ontology; gene expression; microarray; microRNA; transcriptome; genomics; microarrays	36.86	79.14	(−42.4,146.27)
B	Bioinformatics; proteomics; systems biology	80.67	29.33	(1.40,−3.54)
C	Epistemology; methodology; critical realism; ethics; philosophy; phenomenology; realism	37.14	26.86	(−42.12,−6.02)
D	Information extraction; natural language processing; knowledge acquisition; knowledge representation; knowledge engineering	81.60	21.60	(2.33,−11.28)
E	Machine learning; text mining; data mining	73.67	12.00	(−5.60,−20.88)
F	Classification; clustering; semantic similarity; annotation; database	67.60	4.80	(−11.67,−28.08)
G	Interoperability; metadata; XML; information retrieval; semantics; knowledge management; knowledge sharing; semantic interoperability	83.38	18.50	(4.11,−14.38)
H	OWL; RDF; data integration	108.33	18.67	(29.07,−14.21)
I	Multi-agent systems; Semantic Web services; Ontologies; Semantic Web; Web services; description logics	144.17	85.00	(64.90,52.12)

Table VI. Strategic coordinates of the nine clusters of frequent keywords in ontology research

the cluster I located in quadrant 1, the cluster B, D, G, H were in quadrant 2, the cluster A was in quadrant 3, and C, E, F were in quadrant 4. The distribution of these clusters represented that the Semantic Web research got a high degree of connection to other clusters and development. It played a core role in the ontological field, which could be interpreted by that ontology is bound up with the Semantic Web. The

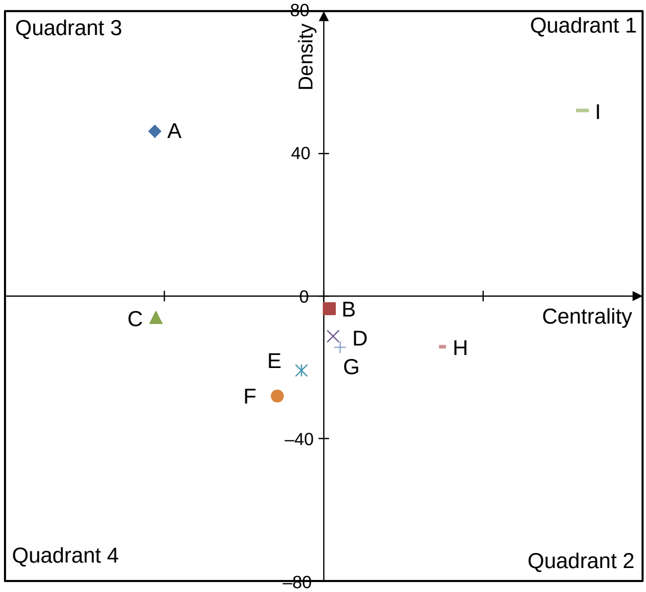

Figure 7.
Strategic graph of the nine clusters

researches about information and knowledge, and the bioinformatics, proteomics, systems biology were also important to the ontological field. Nevertheless, they were in the development phase, and might draw more attention in the future. The ontological applications in gene study were mature and well-developed, but they had few connections with other subjects. The philosophy theories and data mining, clustering were peripheral and little developed, only a dynamic analysis or a comparative one allowed us to determine their contribution to the ontological field.

The top 25 most frequently used author keywords were listed in Table VII within the periods 1990-1995, 1996-2000, 2001-2006, and 2007-2012, respectively. The occurrence frequency was used instead of the occurrence number for analysis, to avoid the bias caused by unequal time intervals (Liu *et al.*, 2013).

"Ontology", as one of our search terms, has kept the highest occurrence frequency during the four periods, and the possible reason for this result is that "ontology" has double meaning as mentioned previously. In contrast, the other searching term "ontologies" was not active until the second period when ontology was preliminarily applied in the domain of information science. Besides the search terms, "semantic web", "GO", and "microarray" were listed in the top five most frequently author keywords during the last two periods, even though these three keywords did not occurred in the years from 1990 to 2000. Especially, the occurrence number of "semantic web" increased from 140 in 2001-2006 to 430 in 2007-2012, and exceeded "ontologies" (393). The same as "GO" and "microarray", some keywords relating to biology such as "gene expression", "bioinformatics", "proteomics", "transcriptome", "systems biology", and "microRNA", appeared on our list till the last two periods. The extremely high increasing rates of these keywords indicated that the SW and bioinformatics were identified as current ontology-related research hotspots. On the contrary, the major component parts of top 25 most frequently used keywords in the first two periods, such as "epistemology", "ethics", "realism", and "philosophy", were related to philosophy

Table VII. The temporal evolution of frequent author keywords in ontology research

	1990-1995			1996-2000			2001-2006			2007-2012		
No	DE	TA	%	DE	TA	%	DE	TA	%	DE	TA	%
1	Ontology	22	19.47	Ontology	101	26.72↑	Ontology	480	21.52↓	Ontology	1,334	18.52↓
2	Epistemology	7	6.19	Knowledge representation	20	5.29↑	Ontologies	171	7.67↑	Semantic Web	430	5.97↓
3	Methodology	3	2.65	Epistemology	16	4.23↓	Semantic Web	140	6.28↑	Ontologies	393	5.46↓
4	Knowledge acquisition	3	2.65	Ontologies	12	3.17↑	Gene ontology	108	4.84↑	Microarray	378	5.25↑
5	Philosophy	3	2.65	Ethics	8	2.12↑	Microarray	95	4.26↑	Gene ontology	296	4.11↓
6	Artificial intelligence	3	2.65	Realism	8	2.12↑	Gene expression	67	3.00↑	Gene expression	267	3.71↑
7	Evolution	3	2.65	Philosophy	7	1.85↓	Knowledge representation	61	2.74↓	Bioinformatics	122	1.69↓
8	Knowledge representation	2	1.77	Communication	7	1.85↑	Bioinformatics	55	2.47↑	OWL	114	1.58↑
9	Realism	2	1.77	Temporal reasoning	7	1.85↑	Knowledge management	46	2.06↑	Knowledge management	113	1.57↓
10	Conceptual modelling	2	1.77	Methodology	6	1.59↓	Epistemology	45	2.02↓	Proteomics	103	1.43↑
11	Knowledge-based systems	2	1.77	Knowledge acquisition	6	1.59↓	Information retrieval	34	1.52↑	Knowledge representation	102	1.42↓
12	Representation	2	1.77	Knowledge sharing	6	1.59↑	Web services	33	1.48↑	Epistemology	97	1.35↓
13	Complexity	2	1.77	Knowledge-based systems	6	1.59↓	Semantics	28	1.26↑	Semantics	90	1.25↓
14	Constructivism	2	1.77	Design	6	1.59↑	Knowledge acquisition	27	1.21↓	Information retrieval	86	1.19↓
15	Explanation	2	1.77	Representation	6	1.59↓	Natural language processing	24	1.08↑	Web services	74	1.03↓
16	Ethnography	2	1.77	Hermeneutics	6	1.59↑	XML	24	1.08↑	Data mining	74	1.03↑
17	Cognitive science	2	1.77	Cybernetics	6	1.59↑	Database	23	1.03↑	Data integration	71	0.99↑
18	Mind	2	1.77	Phenomenology	5	1.32↑	Artificial intelligence	23	1.03↓	Transcriptome	70	0.97↑
19	Metrics	2	1.77	Heidegger	5	1.32↑	Interoperability	22	0.99↓	Interoperability	65	0.90↓
20	Control	2	1.77	Terminology	5	1.32↑	Metadata	22	0.99↑	Description logics	65	0.90↑

(*continued*)

Table VII.

No	1990-1995 DE	TA	%	1996-2000 DE	TA	%	2001-2006 DE	TA	%	2007-2012 DE	TA	%
21	Knowledge bases	2	1.77	Reductionism	5	1.32↑	Realism	22	0.99↓	Methodology	60	0.83↓
22	Critical systems thinking	2	1.77	Information	5	1.32↑	Methodology	21	0.94↓	Database	60	0.83↓
23	Ontogeny	2	1.77	Semantics	4	1.06↑	Annotation	21	0.94↑	Systems biology	57	0.79↑
24	Complementarity	2	1.77	Interoperability	4	1.06↑	Description logics	19	0.85↑	MicroRNA	54	0.75↑
25	Human rights	2	1.77	Critical realism	4	1.06↑	Text mining	19	0.85↑	Natural language processing	50	0.69↓

Notes: DE, author keyword; TA, total articles; %, occurrence frequency (the ratio of articles with the keyword to all articles with keywords in prescriptive period); ↑↓, occurrence frequency rose or dropped (compared with previous stage) theories and became gradually less absorbing during the last two periods, that only "epistemology" was still in the top list. This change in author keywords reflected the transformation of ontology research from philosophy to information science, theories to applications. As for the technical aspect, "OWL", "data mining", "data integration", and "description logics" all showed an incremental trend in the period of 2007-2012, and they might attract more attention in the future. OWL, Web Ontology Language developed from former languages such as XML, RDF, and DAML + OIL, was currently the W3C standard and a frequent language of choice to describe ontologies (Waloszek, 2012). Description logics constituted the logical basis of the language to provide reasoning services and support ontology design (Horrocks, 2002; Zuo and Zhou, 2003). Data mining and integration were important techniques for generating automatic ontologies, but also the major blocks (Ding and Foo, 2002).

Frequently cited articles

Table VIII listed the top 20 most-frequently cited articles from our database, and the TC count of each paper was obtained from WOS on the day we conducted our search process. LCS and CRA were also presented as descriptors of each paper.

"A translation approach to portable ontology specifications" written by Gruber, T,R. was ranked as the most cited paper with 2,861 citations in all, the LCS (953) and CRA (78) of which were both the highest. This paper could be deemed to play an important role in promoting ontology research, not only because the definition of ontology proposed in this paper was commonly accepted, but also because it addressed the portability problem for ontologies (Gruber, 1993). It has provided general and significant knowledge about ontology for multifarious ontology-related research areas. We noticed the paper "Toward principles for the design of ontologies used for knowledge sharing" published in 1995 by Gruber, T,R. had the third most citations (1,024), the second highly LCS (373), and a relatively high CRA (69). Thus, we conclude that Gruber, T,R. has made great contributions to the establishment of ontology theory in scientific fields. "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources" by Huang, DW. was cited 2,264 times in all and ranked as the second highly cited paper, which was a very young article published in 2009. This paper had a high CRA (77) and growth rate of citations that it drew 877 citations on the third year since its publication, suggesting an extensive multidisciplinary influence and a tremendous propagation velocity in scientific community.

Out of these 20 most cited articles, four articles related to research in ontology engineering, such as the definition and construction, ranked the first, third, fourth, and 20th, respectively. The remaining 16 were from bioinformatics such as GO, genomics, and protein. Through examining the highly cited articles in this field, it is worth mentioning that papers related to bioinformatics usually draw more citations than many other scientific fields. Nevertheless, the descriptors such as TC, LCS, and CRA can be biased by the fact that older articles are probably to more citations (Marx and Cardona, 2003). As can be seen, the bioinformatics papers are younger than the four ontology engineering ones, because specific applications are based on theoretical researches, and bioinformatics is one of the leading fields that applies ontology and achieves actual results.

Conclusions

In this study, we provided a panorama of global ontology research during 1900-2012, as well as some significant points on the research performance throughout the period.

Table VIII. Top 20 most-frequently cited articles in ontology research

No	Title	TC	PY	FAU	LCS(R)	CRA(R)
1	A translation approach to portable ontology specifications	2,861	1993	Gruber, T.R.	953(1)	78(1)
2	Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources	2,264	2009	Huang, D.W.	246(4)	77(3)
3	Toward principles for the design of ontologies used for knowledge sharing	1,024	1995	Gruber, T.R.	373(2)	69(4)
4	A metrics suite for object-oriented design	1,024	1994	Chidamber, S.R.	10(15)	23(20)
5	Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research	1,012	2005	Conesa, A.	210(5)	49(15)
6	The gene ontology (GO) database and informatics resource	911	2004	Harris, M.A.	304(3)	78(1)
7	KEGG for linking genomes to life and the environment	860	2008	Kanehisa, M.	77(10)	66(8)
8	Identifying biological themes within lists of genes with EASE	774	2003	Hosack, D.A.	0(17)	68(5)
9	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	760	2002	Okazaki,Y.	26(13)	65(10)
10	GoMiner: a resource for biological interpretation of genomic and proteomic data	667	2003	Zeeberg, B.R.	0(17)	66(8)
11	OrthoMCL: identification of ortholog groups for eukaryotic genomes	662	2003	Li, L.	38(12)	29(19)
12	PANTHER: a library of protein families and subfamilies indexed by function	653	2003	Thomas, P.D.	40(11)	68(5)
13	BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks	651	2005	Maere, S.	152(8)	56(14)
14	Prediction and functional analysis of native disorder in proteins from the three kingdoms of life	634	2004	Ward, J.J.	11(14)	36(18)
15	GOstat: find statistically overrepresented gene ontologies within a group of genes	595	2004	Beissbarth, T.	151(9)	61(12)
16	FatiGO: a web tool for finding significant associations of gene ontology terms with groups of genes	560	2004	Al-Shahrour, F.	0(17)	67(7)
17	MAPPFinder: using gene ontology and GenMAPP to create a global gene-expression profile from microarray data	557	2003	Doniger, S.W.	0(17)	64(11)
18	Creating the gene ontology resource: design and implementation	543	2001	Ashburner, M.	191(7)	61(12)
19	I-TASSER: a unified platform for automated protein structure and function prediction	516	2010	Roy, A.	3(16)	48(16)
20	Ontologies: principles, methods and applications	512	1996	Uschold, M.	209(6)	45(17)

Notes: TC, total citations; LCS, local citation score; FAU, first author; CRA, citing research areas; R, rank in the list

Basic situation helped people get a common understanding about ontology research development, from document types, publishing languages, publication outputs, authors, institutes, and countries. Author keywords and frequently cited articles were used to express professional knowledge, i.e. research emphasis and patterns, as well as influential studies. The new index, DII, was applied to measure the multidisciplinary influence of researchers.

Peer-review journal article was the most common document type, and English was still the leading scientific language.

In the aspect of publication outputs, since the first ontology-related SCI/SSCI publication appeared in 1909, the development of ontology research has gone through three stages: the enlightenment stage (1909-1990), the growth stage (1991-2000), and the soaring stage (2001-2012). Significant growth of scientific outputs was observed, particularly in the last ten years. Meanwhile, more and more authors, countries and institutes engaged in ontology research over years.

Authors, institutes and countries covered three levels of publication ownership. At the micro (authors) level, Smith, B. from SUNY Buffalo was the most prolific author, Chou, K.C. from Gordon Life Science Institute produced more high quality articles, and Lewis, S. from LBNL carried a greater multidisciplinary influence than other authors, among the top 20 productive authors. The spatial distribution of authors was also visualized, and the major spatial clusters were in North America, Europe, and East Asia, with several minor clusters in other parts of the world. Additionally, scientific outputs were positively correlated with the R&D expenditure. We were, of course, aware of the fact that the indicator *h*-index had many flaws and inconsistency. Both *h*-index and DII were citation-based indicators, and this might be biased due to different research fields. Dynamic research fields such as bioinformatics usually got more citations, so the scholars in these areas might have higher scores. A field-based weight was suggested to integrate into citation-based indicators in future work.

According to evaluation of top authors, DII was successfully applied to evaluate the multidisciplinary influence of authors. DII measures in how many research areas on average cited per article of the author. The generality of the formulation and easily retrievable parameters allow its application among other disciplines. Since there is a trade-off between accuracy and simplicity of the indicator, we should mention that there are enriching perspectives. First, investigations employing larger bibliometric sets are needed to check practicability. Second, the diversity and coherence of research areas should be considered when calculating DII.

When it comes to the meso (institutions) and macro (countries) levels, 8,339 institutions and 112 countries/territories participated in ontology research. Institutional as well as international collaborations played an ever growing role in ontology-related research. Moreover, inter-institutional collaboration was more prevalent than international collaboration. At institutional level, the most productive institution was the University of Manchester in UK, followed by the Stanford University in USA and the Chinese Academy of Sciences in China. At country level, the USA attained a leading position in ontology research by contributing the most independent and international collaborative articles. China emerged to be the fastest growing country, especially in the soaring stage of ontology research.

For professional knowledge, author keywords analysis offered insights into the emphasis and trends of ontological research. Four major categories of ontology research were summarized: applications in Semantic Web, applications in bioinformatics, philosophy theories, and common supporting technology. Furthermore, the Semantic Web research played a core role. The ontological applications in gene study were well-developed. "Semantic Web", "GO" and "microarray" were growing research focuses in ontology research, while "epistemology", "ethics", "realism", and "philosophy" were becoming gradually less absorbing, which reflected the transformation from philosophy to information science, theories to applications.

Furthermore, the top 20 most-frequently cited articles were analysed as special cases. "A translation approach to portable ontology specifications" and "Toward principles for the design of ontologies used for knowledge sharing", both written by Gruber, T,R. were the most influential papers in ontology field. Ontology papers related to bioinformatics were comparatively younger and spread faster, suggesting ontology has been successfully applied in bioinformatics.

In short, the results put forth here comprehensively reveal the research progress and patterns in ontology research. Ontology research is an emerging and multidisciplinary field. We suggest the interdisciplinary research of ontology to be an interesting and meaningful future analysis, from the aspects of knowledge integration and diffusion. Effective indicators should also be established to aid in interdisciplinary research assessment.

Notes

1. A total of 151 research areas were defined by ISI. A list of all research areas can be found on ISI web sites.
2. Records in the SCI/SSCI databases were categorized as one of the 38 ISI document types. A list of all document types can be found on ISI web sites.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. (2000), "Gene ontology: tool for the unification of biology", *Nature Genetics*, Vol. 25 No. 1, pp. 25-29.

Ashburner, M., Ball, C.A., Blake, J.A., Butler, H., Cherry, J.M., Corradi, J., Dolinski, K., Eppig, J.T., Harris, M., Hill, D.P., Lewis, S., Marshall, B., Mungall, C., Reiser, L., Rhee, S., Richardson, J.E., Richter, J., Ringwald, M., Rubin, G.M., Sherlock, G. and Yoon, J. (2001), "Creating the gene ontology resource: design and implementation", *Genome Res*, Vol. 11 No. 8, pp. 1425-1433.

Bateman, J.A. (1995), "On the relationship between ontology construction and natural language: a socio-semiotic view", *International Journal Of Human-Computer Studies*, Vol. 43 Nos 5-6, pp. 929-944.

Bateman, J.A., Hois, J., Ross, R. and Tenbrink, T. (2010), "A linguistic ontology of space for natural language processing", *Artificial Intelligence*, Vol. 174 No. 14, pp. 1027-1071.

Bawden, D. and Robinson, L. (2008), "The dark side of information: overload, anxiety and other paradoxes and pathologies", *Journal of Information Science*, Vol. 35 No. 2, pp. 180-191.

Bergstrom, C.T., West, J.D. and Wiseman, M.A. (2008), "The eigenfactor (TM) metrics", *Journal of Neuroscience*, Vol. 28 No. 45, pp. 11433-11434.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001), "The Semantic Web – a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", *Scientific American*, Vol. 284 No. 5, pp. 34-43.

Bordons, M., Morillo, F. and Gómez, I. (2005), "Analysis of cross-disciplinary research through bibliometric tools", in Henk, F.M., Wolfgang, G. and Ulrich, S. (Eds), *Handbook of Quantitative Science and Technology Research*, Springer, pp. 437-456.

Bornmann, L., Schier, H., Marx, W. and Daniel, H.D. (2012), "What factors determine citation counts of publications in chemistry besides their quality?", *Journal of Informetrics*, Vol. 6 No. 1, pp. 11-18.

Borst, W.N. (1997), "Construction of engineering ontologies for knowledge sharing and reuse", Phd thesis, Universiteit Twente, Enschede.

Brandt, S.C., Morbach, J., Miatidis, M., Theißen, M., Jarke, M. and Marquardt, W. (2008), "An ontology-based approach to knowledge management in design processes", *Computers & Chemical Engineering*, Vol. 32 Nos 1-2, pp. 320-342.

Callon, M., Courtial, J.P. and Laville, F. (1991), "Co-word analysis as a tool for describing the network of interactions between basic and technological research – the case of polymer chemistry", *Scientometrics*, Vol. 22 No. 1, pp. 155-205.

Carley, S. and Porter, A.L. (2012), "A forward diversity index", *Scientometrics*, Vol. 90 No. 2, pp. 407-427.

Chen, C. (2004), "Searching for intellectual turning points: progressive knowledge domain visualization", *Proc Natl Acad Sci U S A*, Vol. 101 No. 1, pp. 5303-5310.

Chidamber, S.R. and Kemerer, C.F. (1994), "A metrics suite for object oriented design", *IEEE Transactions on Software Engineering*, Vol. 20 No. 6, pp. 476-493.

Chiu, W.T. and Ho, Y.S. (2007), "Bibliometric analysis of tsunami research", *Scientometrics*, Vol. 73 No. 1, pp. 3-17.

Chuang, K.Y., Huang, Y.L. and Ho, Y.S. (2007), "A bibliometric and citation analysis of stroke-related research in Taiwan", *Scientometrics*, Vol. 72 No. 2, pp. 201-212.

Costas, R. and Bordons, M. (2008), "Development of a thematic filter for the bibliometric delimitation on interdisciplinary area: the case of Marine Science", *Revista Espanola De Documentacion Cientifica*, Vol. 31 No. 2, pp. 261-272.

Ding, Y. (2001), "A review of ontologies with the Semantic Web in view", *Journal of Information Science*, Vol. 27 No. 6, pp. 377-384.

Ding, Y. (2010), "Semantic Web: who is who in the field – a bibliometric analysis", *Journal of Information Science*, Vol. 36 No. 3, pp. 335-356.

Ding, Y. and Foo, S. (2002), "Ontology research and development. Part 1 – a review of ontology generation", *Journal of Information Science*, Vol. 28 No. 2, pp. 123-136.

Ding, Y., Chowdhury, G.G. and Foo, S. (2001), "Bibliometric cartography of information retrieval research by using co-word analysis", *Information Processing & Management*, Vol. 37 No. 6, pp. 817-842.

Garfield, E., Pudovkin, A.I. and Istomin, V.S. (2003), "Why do we need algorithmic historiography?", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 5, pp. 400-412.

Gruber, T.R. (1993), "A translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol. 5 No. 2, pp. 199-220.

Gruber, T.R. (1995), "Toward principles for the design of ontologies used for knowledge sharing", *International Journal Of Human-Computer Studies*, Vol. 43 Nos 5-6, pp. 907-928.

Guarino, N. (1995), "Formal ontology, conceptual analysis and knowledge representation", *International Journal of Human-Computer Studies*, Vol. 43 Nos 5-6, pp. 625-640.

Guarino, N. (1997), "Understanding, building and using ontologies", *International Journal of Human-Computer Studies*, Vol. 46 Nos 2-3, pp. 293-310.

Guarino, N. (1998), "Formal ontology and information systems", in Guarino, N. (Ed.), *Formal Ontology in Information Systems*, I O S Press, Amsterdam, pp. 3-15.

Guarino, N. and Welty, C. (2000), "A formal ontology of properties", in Dieng, R. and Corby, O. (Eds), *Knowledge Engineering And Knowledge Management, Proceedings: Methods, Models, And Tools*, Springer-Verlag, Berlin, pp. 97-112.

Halpin, H. and Presutti, V. (2011), "The identity of resources on the Web: an ontology for Web architecture", *Applied Ontology*, Vol. 6 No. 3, pp. 263-293.

Herbertz, H. and MullerHill, B. (1995), "Quality and efficiency of basic research in molecular biology: a bibliometric analysis of thirteen excellent research institutes", *Research Policy*, Vol. 24 No. 6, pp. 959-979.

Hirsch, J.E. (2005), "An index to quantify an individual's scientific research output", *Proc Natl Acad Sci U S A*, Vol. 102 No. 46, pp. 16569-16572.

Holsapple, C.W. and Joshi, K.D. (2004), "Formal knowledge management ontology: conduct, activities, resources, and influences", *Journal of the American Society for Information Science and Technology*, Vol. 55 No. 7, pp. 593-612.

Horrocks, I. (2002), "DAML+OIL: a reason-able Web ontology language", in Jensen, C.S., Jeffery, K.G., Pokorny, J., Saltenis, S., Bertino, E., Bohm, K. and Jarke, M. (Eds), *Advances in Database Technology - Edbt 2002*, Springer-Verlag, Berlin, pp. 2-13.

Kietz, J.-U., Volz, R. and Maedche, A. (2000), "Extracting a domain-specific ontology from a corporate intranet", *Proceedings of the 2nd Workshop on Learning Language in Logic and the 4th Conference on Computational Natural Language Learning-Vol. 7, Association for Computational Linguistics*, pp. 167-175.

Lang, E. (1991), "The LILOG ontology from a linguistic point of view", in Herzog, O. and Rollinger, C.-R. (Eds), *Text Understanding in LILOG*, Springer-Verlag, Berlin, pp. 464-481.

Lariviere, V., Sugimoto, C.R. and Cronin, B. (2012), "A bibliometric chronicling of library and information science's first hundred years", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 5, pp. 997-1016.

Leydesdorff, L. and Wagner, C. (2009), "Is the United States losing ground in science? A global perspective on the world science system", *Scientometrics*, Vol. 78 No. 1, pp. 23-36.

Leydesdorff, L., Rafols, I. and Chen, C.M. (2013), "Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 12, pp. 2573-2586.

Leydesdorff, L., Rotolo, D. and Rafols, I. (2012), "Bibliometric perspectives on medical innovation using the medical subject headings of pubmed", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 11, pp. 2239-2253.

Li, Z., Raskin, V. and Ramani, K. (2008), "Developing engineering ontology for information retrieval", *Journal of Computing and Information Science in Engineering*, Vol. 8 No. 1, pp. 011003-0110013.

Liu, X.J., Zhan, F.B., Hong, S., Niu, B.B. and Liu, Y.L. (2012), "A bibliometric study of earthquake research: 1900-2010", *Scientometrics*, Vol. 92 No. 3, pp. 747-765.

Liu, X.J., Zhang, L.A. and Hong, S. (2011), "Global biodiversity research during 1900-2009: a bibliometric analysis", *Biodiversity and Conservation*, Vol. 20 No. 4, pp. 807-826.

Liu, Z.Q., Liu, Y.L., Guo, Y.J. and Wang, H. (2013), "Progress in global parallel computing research: a bibliometric approach", *Scientometrics*, Vol. 95 No. 3, pp. 967-983.

McCarthy, J. (1980), "Circumscription – a form of non-monotonic reasoning", *Artificial Intelligence*, Vol. 13 Nos 1-2, pp. 27-39.

McGuinness, D.L. (1998), "Ontological issues for knowledge-enhanced search", in Guarino, N. (Ed.), *Formal Ontology in Information Systems*, I O S Press, Amsterdam, pp. 302-316.

Marx, W. and Cardona, M. (2003), "The impact of solid state communications in view of the ISI citation data", *Solid State Communications*, Vol. 127 No. 5, pp. 323-336.

Moed, H.F. (2010), "Measuring contextual citation impact of scientific journals", *Journal of Informetrics*, Vol. 4 No. 3, pp. 265-277.

Myrgioti, E.V., Chouvardas, V.G. and Miliou, A.N. (2009), "Ontological representation of tactile information for software development", *Applied Ontology*, Vol. 4 No. 2, pp. 139-167.

Neches, R., Fikes, R.E., Finin, T., Gruber, T., Patil, R., Senator, T. and Swartout, W.R. (1991), "Enabling technology for knowledge sharing", *AI Magazine*, Vol. 12 No. 3, p. 36.

Rafols, I. and Meyer, M. (2010), "Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience", *Scientometrics*, Vol. 82 No. 2, pp. 263-287.

Riikonen, P. and Vihinen, M. (2008), "National research contributions: a case study on Finnish biomedical research", *Scientometrics*, Vol. 77 No. 2, pp. 207-222.

Seglen, P.O. (1997), "Why the impact factor of journals should not be used for evaluating research", *British Medical Journal*, Vol. 314 No. 7079, pp. 498-502.

Shadbolt, N., Hall, W. and Berners-Lee, T. (2006), "The Semantic Web revisited", *Ieee Intelligent Systems*, Vol. 21 No. 3, pp. 96-101.

Small, H. and Upham, P. (2009), "Citation structure of an emerging research area on the verge of application", *Scientometrics*, Vol. 79 No. 2, pp. 365-375.

Sowa, J.F. (2000), *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks/Cole, New York, NY.

Studer, R., Benjamins, V.R. and Fensel, D. (1998), "Knowledge engineering: principles and methods", *Data & Knowledge Engineering*, Vol. 25 No. 1, pp. 161-197.

Tarkowski, S.M. (2007), "Environmental health research in Europe – bibliometric analysis", *European Journal of Public Health*, Vol. 17 No. S1, pp. 14-18.

Ugolini, D., Cimmino, M.A., Casilli, C. and Mela, G.S. (2001), "How the European union writes about ophthalmology", *Scientometrics*, Vol. 52 No. 1, pp. 45-58.

Uschold, M. and Gruninger, M. (1996), "Ontologies: principles, methods and applications", *Knowledge Engineering Review*, Vol. 11 No. 2, pp. 93-136.

van Eck, N.J. and Waltman, L. (2010), "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Vol. 84 No. 2, pp. 523-538.

van Raan, A.F.J. (2008), "Bibliometric statistical properties of the 100 largest European research universities: prevalent scaling rules in the science system", *Journal of the American Society for Information Science and Technology*, Vol. 59 No. 3, pp. 461-475.

van Raan, A.F.J., Visser, M.S., Van Leeuwen, T.N. and van Wijk, E. (2003), "Bibliometric analysis of psychotherapy research: performance assessment and position in the journal landscape", *Psychotherapy Research*, Vol. 13 No. 4, pp. 511-528.

vanHeijst, G., Schreiber, A.T. and Wielinga, B.J. (1997), "Using explicit ontologies in KBS development", *International Journal of Human-Computer Studies*, Vol. 46 Nos 2-3, pp. 183-292.

Waloszek, W. (2012), "Measures for evaluation of structure and semantics of ontologies", *Metrology and Measurement Systems*, Vol. 19 No. 2, pp. 343-354.

Waltman, L., van Eck, N.J. and Noyons, E.C.M. (2010), "A unified approach to mapping and clustering of bibliometric networks", *Journal of Informetrics*, Vol. 4 No. 4, pp. 629-635.

Wang, H., He, Q., Liu, X., Zhuang, Y. and Hong, S. (2012), "Global urbanization research from 1991 to 2009: a systematic research review", *Landscape and Urban Planning*, Vol. 104 Nos 3-4, pp. 299-309.

Xie, S., Zhang, J. and Ho, Y.-S. (2008), "Assessment of world aerosol research trends by bibliometric analysis", *Scientometrics*, Vol. 77 No. 1, pp. 113-130.

Zhuang, Y.H., Liu, X.J., Nguyen, T., He, Q.Q. and Hong, S. (2013), "Global remote sensing research trends during 1991-2010: a bibliometric analysis", *Scientometrics*, Vol. 96 No. 1, pp. 203-219.

Zitt, M. and Small, H. (2008), "Modifying the journal impact factor by fractional citation weighting: the audience factor", *Journal of the American Society for Information Science and Technology*, Vol. 59 No. 11, pp. 1856-1860.

Zuo, Z.H. and Zhou, M.T. (2003), "Web ontology language OWL and its description logic foundation", *Proceedings of the 4th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT 2003), IEEE, New York, NY*, pp. 157-160.

Corresponding author
Professor Zongyi He can be contacted at: hezongyi2013@163.com

This article has been cited by:

1. WangQing, Qing Wang. 2018. Distribution features and intellectual structures of digital humanities. *Journal of Documentation* **74**:1, 223-246. [Abstract] [Full Text] [PDF]
2. LvHong, Hong Lv. 2017. Assessment of global law and psychiatry research in the period of 1993-2012. *The Electronic Library* **35**:3, 559-572. [Abstract] [Full Text] [PDF]
3. Delcea Camelia. 2015. Grey systems theory in economics – bibliometric analysis and applications' overview. *Grey Systems: Theory and Application* **5**:2, 244-262. [Abstract] [Full Text] [PDF]