CSCanada
Management Science and Engineering
Vol. 10, No. 4, 2016, pp. 30-35
DOI:10.3968/8963

ISSN 1913-0341 [Print]
ISSN 1913-035X [Online]
www.cscanada.net
www.cscanada.org

The Visualization Analysis of Cloud Computing Research Based on Mapping Knowledge

LIU Wanjun[a],*

[a]School of Business Administration, South China University of Technology, Guangzhou, China.
*Corresponding author.

Received 10 September 2016; accepted 15 November 2016
Published online 26 December 2016

Abstract

This paper analyzed the current research on the field of cloud computing by adopting the bibliometric analysis and co-word network analysis based on the knowledge mapping theory and methods with software tools SPSS 19.0, Ucinet6.0 and NetDraw. In this paper, we conducted a visual analysis on the research status of cloud computing for the past five years, including the knowledge structure of cloud computing, research hot spots and trends from macro and micro perspective. The result shows that the research of cloud computing can be divided into eight knowledge groups: (a) technical research related Cloud computing; (b) cloud computing security research; (c) cloud computing performance research; (d) characteristics research of cloud computing; (e) big data processing research based on cloud computing; (f) research in the form of cloud computing services; (g) infrastructure research of cloud computing; (h) application research of cloud computing in chemistry and physics. The paper may provide valuable suggestions for better understanding on the research status of cloud computing.

Key words: Cloud computing; Mapping knowledge; Vector dynamic model; Co-word analysis; Social network analysis

Liu, W. J. (2016). The Visualization Analysis of Cloud Computing Research Based on Mapping Knowledge. *Management Science and Engineering*, *10*(4), 30-35. Available from: URL: http://www.cscanada.net/index.php/mse/article/view/8963
DOI: http://dx.doi.org/10.3968/8963

INTRODUCTION

In recent years, cloud computing has become a hot research field and attracts many researchers. From January 2010 to May 2015, the literature quantity related to cloud computing has reached 52,202 (Source: WEB OF SCIENCE). In order to analyze the research status in the field of cloud computing accurately, methods of visualization analysis based on knowledge mapping are used in this paper to demonstrate the present situation of cloud computing research, including literature analysis, analysis of high-frequency words, major field analysis, research focus analysis and future trends analysis.

1. DATA RESOURCES AND SOFTWARE TOOLS

The data used in this paper was collected from the WEB OF SCIENCE database with the retrieval theme as “CLOUD COMPUTING”, time span as 2010-2015, and literature type as “Article”. Then a total of 4,479 relevant articles were retrieved.

Main software tools used in this paper include Excel, SPSS19.0, Ucinet6 and NetDraw. SPSS is used to get the hierarchical cluster analysis, with “Between Groups Linkage” as clustering method and “squared Euclidean distance” as its distance. The betweenness centrality analysis, nodes centrality degree analysis, subgroup analysis and the knowledge map were also conducted through Ucinet6 and NetDraw tools.

2. RESEARCH STATUS AND TREND ANALYSIS

2.1 Annual Distribution

Figure 1 shows the quantity of publications from January 2010 to May 2015. As it can be seen, in recent years, the amount of papers related to cloud computing has been gradually growing since 2010 (Literature of 2015 only counted the first five months). It indicates that "cloud computing" is gradually becoming the focus of researchers in these five years.

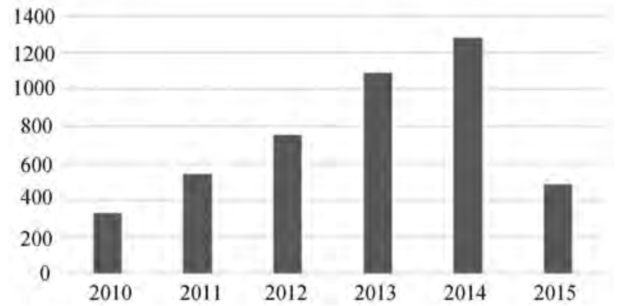

Figure 1
The Annual-Published Figure of Cloud Computing Literature

2.2 Research Status Analysis Based on Vector Dynamic Model

To analyze the research status, the method of vector dynamic model (Zhao & Zhang, 2012) is adopted, in which the annual quantity of published papers, the number of journals and authors are involved. The result is shown as Figure 2 and Table 1.

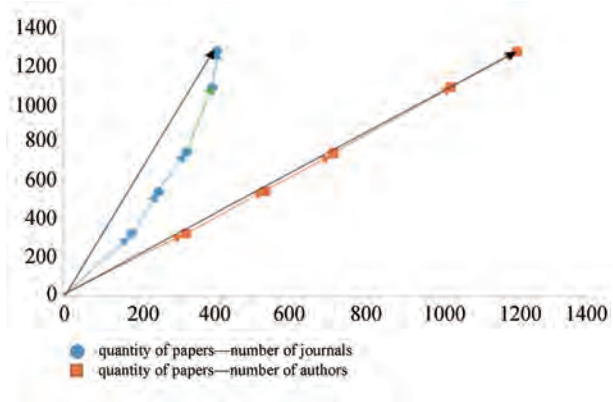

Figure 2
Dynamic Vector Model of Cloud Computing Research

Table 1
Evaluation Results of Vector Dynamic Model Method

Vector and quadrant	Research activity	The level of participation in research activities	Status of development
I	Increase	Increase	Development stage
II	Decrease / stable	Increase / stable	Increase or completion stage
III	Decrease / stable	Decrease	Fully mature or declining stage
IV	Increase	Decrease / stable	Studies or completion stage

As can be seen from Figure 2, statistical vector sum of "quantity of papers- number of journals" and statistical vector sum of "quantity of papers-number of authors" are both in the first quadrant, showing that cloud computing research is in a stable stage of development.

3. RESEARCH HOT SPOTS AND FRONTIERS ANALYSIS

3.1 Word Frequency Statistics and High-Frequency Keywords Analysis

Firstly, the keywords need to be standardized, for example, "Software-as-a-Service", "Software as a Service" and "SaaS" unified as "SaaS"; "web services" and "web service" unified as "web service". Then according to convention which suggests the threshold of high-frequency words is usually set at 30%-40% of the total amount of word frequency (Xie, Liang, & Wang, 2005; Ma, Xu, & Nan, 2010), this paper set threshold as 11, and 131 words are selected. Table 2 lists the top 40 high frequency words, Generally the high frequency words implies the hot spots in the field, and from Table 2 we can see that the hottest words includes: cloud computing, virtualization, cloud, MapReduce, security, Qos, etc..

Table 2
The High-Frequency Keywords in Cloud Computing Field (Part)

Seq.	Keyword	Freq.	Seq.	Keyword	Freq.
1	Cloud computing	1,569	21	SLA	47
2	Virtualization	145	22	Resource management	45
3	Cloud	115	23	Radiative transfer	42
4	Mapreduce	102	24	Web service	41
5	Security	93	25	Distributed system	40
6	QoS	88	26	Point cloud	40

To be continued

Continued

Seq.	Keyword	Freq.	Seq.	Keyword	Freq.
7	Grid computing	74	27	Fault tolerance	38
8	Data center	73	28	Internet of things	37
9	Performance	71	29	Access control	35
10	Privacy	64	30	Distributed computing	34
11	Mobile cloud computing	63	31	Load balance	34
12	SaaS	62	32	Algorithm	33
13	Energy efficiency	58	33	Scientific workflow	33
14	Resource allocation	56	34	Parallel computing	32
15	Design	51	35	Service-oriented architecture	32
16	Scheduling	51	36	Cloud storage	31
17	Virtual machine	51	37	Genetic algorithm	31
18	Big data	49	38	Hadoop	31
19	HPC	48	39	Interoperability	31
20	IaaS	47	40	Astrochemistry	30

3.2 Construction of Co-word Matrix and Dissimilarity Matrix

131*4479 keyword article matrix of high-frequency keywords and 131*131 Co-word matrix of high-frequency keywords were produced via Matlab. As Keywords co-word matrix cannot fully express the relative strength of the co-occurrence between keywords (Ma, Xu, & Nan, 2010), Salton index is used to indicate the associated strength between keywords, calculated as follows:

$$S = n_{ij} / \left(n_i \times n_j\right)^{1/2} \; . \qquad (1)$$

Then the matrix of the degree of dissimilarity between keywords were conducted to indicate the co-word exponential distances (the table is omitted in the paper). And then the co-occurrence network can be generated by UCINET6.0 (Figure 3).

3.3 Analysis on the Importance and the Relations of Hot Spots

As can be seen from Figure 3, "cloud computing" is in the most central location on the network. Distance between "cloud computing" and other keywords shows the importance of the keyword. The shorter the distance, the higher the degree of importance. For example, "security", "MapReduce" and "virtualization" are sub-core themes. In order to analyze the core level of high-frequency keywords accurately, the result of centrality analysis to co-word matrix is calculated and shown in Figure 4. In Figure 4, the high-centrality words may not have high frequencies, e.g., the frequency of "performance" is less than that of "cloud", but the centrality degree of "performance" is higher. That is to say, compared to "cloud", researchers paid more attention to "performance". The degree of "H II regions" and "Magellanic Cloud" is the lowest, indicating that they belong to edge topics.

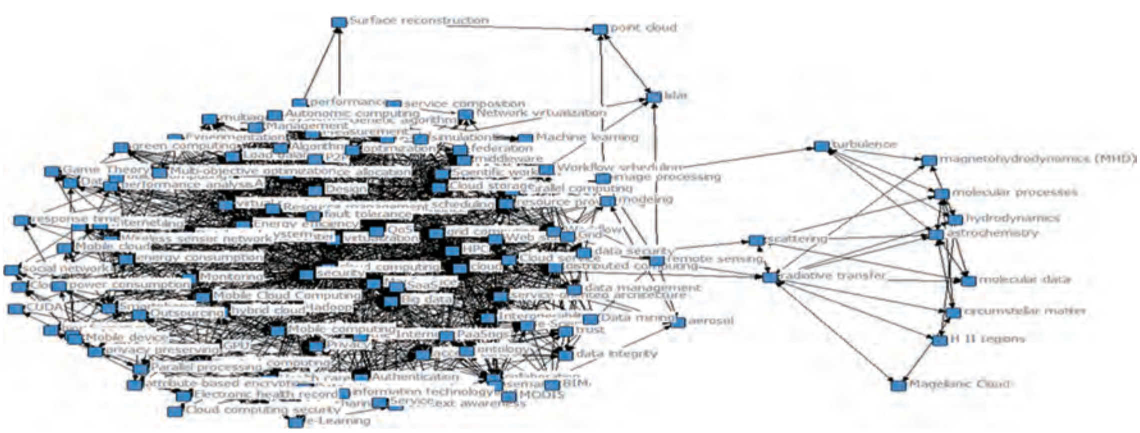

Figure 3
Co-Occurrence Network of High Frequency Words

		1 Degree	2 NrmDegree	3 Share
1	cloud computing	1819.000	14.885	0.224
2	virtualization	316.000	2.586	0.039
9	Performance	221.000	1.809	0.027
15	Design	196.000	1.604	0.024
5	security	181.000	1.481	0.022
4	MapReduce	170.000	1.391	0.021
6	QoS	167.000	1.367	0.021
3	cloud	156.000	1.277	0.019
8	Data center	151.000	1.236	0.019
7	grid computing	133.000	1.088	0.016
12	SaaS	125.000	1.023	0.015
10	Privacy	118.000	0.966	0.015
20	IaaS	114.000	0.933	0.014
13	Energy efficiency	111.000	0.908	0.014
47	Experimentation	111.000	0.908	0.014
16	scheduling	111.000	0.908	0.014
32	Algorithm	107.000	0.876	0.013
18	Big data	101.000	0.827	0.012
22	Resource management	99.000	0.810	0.012
48	Measurement	95.000	0.777	0.012
14	resource allocation	89.000	0.728	0.011
53	Management	88.000	0.720	0.011
21	SLA	87.000	0.712	0.011
24	Web service	86.000	0.704	0.011
17	virtual machine	85.000	0.696	0.010
19	HPC	77.000	0.630	0.009
27	fault tolerance	72.000	0.589	0.009
38	Hadoop	72.000	0.589	0.009
25	Distributed system	66.000	0.540	0.008
35	service-oriented architecture	64.000	0.524	0.008
51	PaaS	64.000	0.524	0.008
31	Load balance	63.000	0.516	0.008
45	reliability	62.000	0.507	0.008
33	Scientific workflow	61.000	0.499	0.008

Figure 4
The Centrality Analysis (Part)

4. ANALYSIS OF RESEARCH FIELDS OF CLOUD COMPUTING

4.1 Keywords Clustering Analysis

Based on the above dissimilarity matrix, we get clustering tree diagram of high-frequency keywords through SPSS19.0 (Figure 5), which reveals that the research of cloud computing includes 8 knowledge clusters (named by the author):

(a) Technical research related cloud computing. Relevant keywords as follows: cloud, QoS, grid computing, mobile cloud computing, resource allocation, scheduling, virtual machine, HPC, SLA, etc.. In the last five years, technical research related cloud computing, whose high-frequency keywords accounts for more than half of all high-frequency keywords, is hot research topic.

(b) Cloud computing security research. Relevant keywords as follows: security, privacy, access control, cloud storage, cloud security, etc..

(c) Cloud computing performance research. Relevant keywords as follows: performance, design, algorithm, reliability, experimentation, measurement, management.

(d) Characteristics of research of cloud computing. Relevant keywords as follows: cloud computing, virtualization, data center, energy efficiency, resource management, green computing.

(e) Big data processing research based on cloud computing. Relevant keywords as follows: MapReduce, big data, Hadoop, data-intensive computing, parallel processing.

(f) Research in the form of cloud computing services. Relevant keywords as follows: SaaS, IaaS, PaaS.

(g) Infrastructure research of cloud computing. Relevant keywords as follows: GPU, CUDA.

(h) Application research of cloud computing in chemistry and physics. Relevant keywords as follows: astrochemistry, molecular processes, molecular data, scattering.

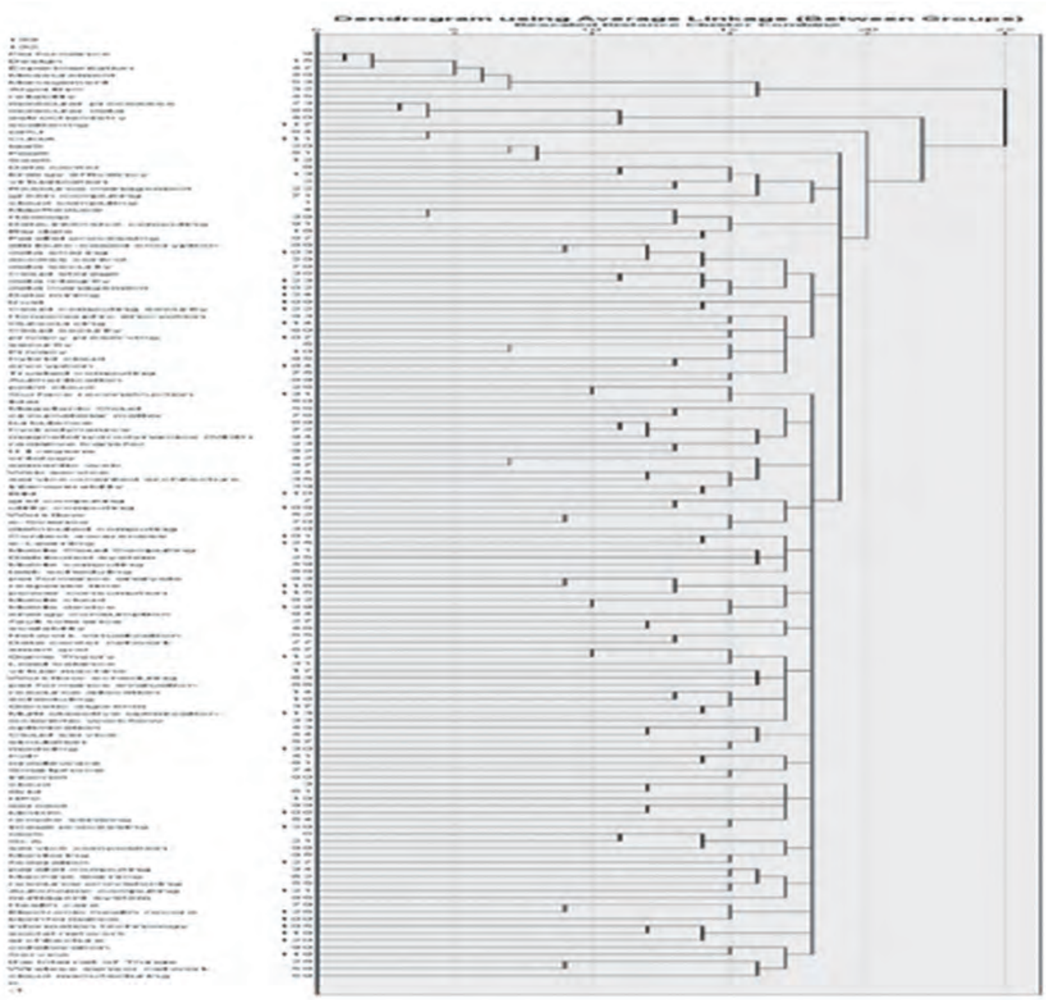

Figure 5
High Frequency Keywords Clustering Map of Cloud Computing

4.2 Cohesive Subgroups Analysis

Cohesive subgroup is a subset of actors (Liu, 2004, 2009). The actors in the subset have relatively strong, cohesive, direct, regular or positive relationship. Cohesive Subgroup analysis could reflect the degree of tightness inside subgroups and the connection between each subgroup.

Cohesive subgroup analysis to co-word network of high-frequency keywords in the field of cloud computing is shown in Figure 6.

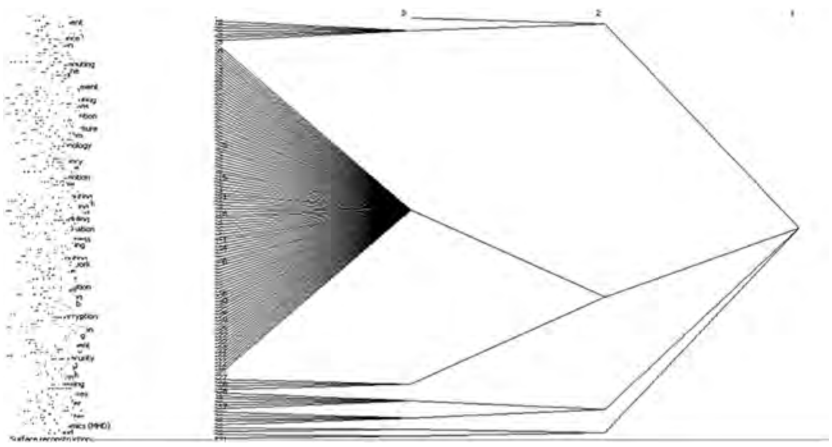

Figure 6
The Subgroups Distribution

As shown in Figure 6, the result of cohesive subgroup analysis is that high-frequency keywords were divided into eight subgroups and the high-frequency keywords within the same subgroup form a small group. It is similar to the result of cluster analysis. Derived from Figure 7, the density of the entire network is 0.493. The density of the second sub-group, namely cloud computing performance research (11.476), is the greatest. It indicates that the internal connection degree of high-frequency keywords within the second sub-group is the highest.

As for the connection degree between one subgroup and other subgroups, the highest is 16.221. It reveals that the keywords within subgroup 1 and the keywords within subgroup 3 contact most. Meanwhile, connection degree between subgroup 1 and subgroup 2 has reached 16.000.

Density Matrix

	1	2	3	4	5	6	7	8
1		16. 000	16. 221	5. 000	0. 000	0. 000	0. 000	0. 000
2	16. 000	11. 476	0. 380	0. 286	0. 000	0. 000	0. 000	0. 071
3	16. 221	0. 380	0. 271	0. 358	0. 002	0. 000	0. 019	0. 019
4	5. 000	0. 286	0. 358	1. 833	0. 208	0. 000	1. 000	0. 000
5	0. 000	0. 000	0. 002	0. 208	3. 000	0. 700	0. 000	0. 000
6	0. 000	0. 000	0. 000	0. 000	0. 700	0. 500	0. 000	0. 000
7	0. 000	0. 000	0. 019	1. 000	0. 000	0. 000	0. 000	2. 000
8	0. 000	0. 071	0. 019	0. 000	0. 000	0. 000	2. 000	0. 000

R-squared = 0. 493

First order actor-by-actor correlation matrix saved as dataset Concor1stCorr
Partition-by-actor indicator matrix saved as dataset ConcorCCPart
Permutation vector saved as dataset ConcorCCPerm

Running time: 00:00:01
Output generated: 02 六月 15 11:47:31
Copyright (c) 1999-2008 Analytic Technologies

Figure 7
The Subgroups Density Matrix

CONCLUSIONS AND SUGGESTIONS

The paper analyzes the status of cloud computing research based on knowledge mapping method, including annual document distribution analysis, authors analysis, word frequency statistics, vector dynamic model analysis, keywords clustering analysis, social network analysis and cohesive subgroups analysis.

(a) The analysis of annual document distribution reveals that the heat of cloud computing research rises year by year for the past five years.

(b) The authors statistics show that there are some relatively active scholars in the field of cloud computing. Such as Buyya Rajkumar, Chen Jinjun, Jin Hai and Li Jin in the past five years.

(c) The result of word frequency statistics reveals that the current hotspots of cloud computing research focus on virtualization, cloud, parallel programming techniques, safety, etc..

(d) We can figure out that research in cloud computing is in a stable stage of development via vector dynamic model analysis in the past five years.

(e) Cluster analysis reveals that the research fields include eight knowledge clusters, namely: technical research related Cloud computing, cloud computing security research, cloud computing performance research, characteristics research of cloud computing, big data processing research based on cloud computing, research in the form of cloud computing services, infrastructure research of cloud computing, application research of cloud computing in chemistry and physics. This classification basically reflects the main areas and content of cloud computing research in the past five years.

(f) Social network analysis illustrates the core contents of the current study about cloud computing include virtualization, performance and security.

(g) Cohesive subgroup analysis indicates the extent of tightness inside the 8 subgroups and connectivity between different subgroups.

Overall, in the past five years, research in the field of cloud computing mainly focus on cloud computing technology. In addition, performance and safety research in cloud computing have become hot spots in recent research. However, the application research of cloud computing is relatively few and further promotions may be conducted in the future.

REFERENCES

Liu, J. (2004). *Social network analysis*. Beijing: Social Sciences Academic Press.

Liu, J. (2009). *Lectures on whole network approach*. Shanghai: Shanghai People's Publishing House.

Ma, B.-Y., Xu, H.-Y., & Nan, Q.-M. (2010). Tendency of geological survey technology development and co-keyword clustering analysis. *Scientific and Technological Management of Land and Resources*, *27*(3), 103-110.

Xie, C. X., Liang, L. M., & Wang, W. H. (2005). Keywords co-occurrence analysis of nanotechnology paper in China. *Journal of Information*, (3), 69-72.

Zhao, L. M., & Zhang, Q. P. (2012). Mapping knowledge domains of research frontier transferring in knowledge management field in China. *Science of Science and Management of S. & T.*, *33*(01), 90-98.

