Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 131 (2018) 739–745

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Congress of Information and Communication Technology (ICICT-2018)

Exploring the Metaphor of Information under the Context of Information Science Research: A Co-words Analysis

Aiping Guan[a], Chong Gao[a], Tongtong Zhang[a], Wenjie Zhou[a]*

[a]*Business School of Northwest Normal University, 967# Anning east road, Lanzhou, 730070, China*

Abstract

Present study intends to explore the various definitions of information via co-words analysis procedure based on the Natural Language Process (NLP). 3593 academic papers from an academic database, LISTA, are covered. In short, present study proposes four types of information definitions based on co-words for full spectrum analysis on abstracts of scientific papers under the context of Information Science research as: information as records, information as individual experience, information as the update of data or as the foundation of knowledge and information as ability of individuals.

Keywords: Information definition; information science; co-words analysis

1. Introduction

Information has been widely regarded as one of the central concepts for Information Science, not least because the field itself is named after it and presumably takes it as its object of study. However, until now, information has not earned an agreed definition within the research community of Information Science. The first attempt to conceptualize the term specifically for information science dates back to the 1960s (Wellisch, 1972); the latest are still being reported (Budd, 2011; Gnoli and Ridi, 2014; Ma, 2012; Robinson and Bawden, 2014). In between these decades, many different definitions emerged. These definitions are not without impact, but far from to dominate Information Science understanding of information. For this reason, Present study intends to explore the various definitions of information via co-words analysis procedure based on the Natural Language Process (NLP). Through this research, we expect to describe the landscape of various information definitions among the Information Science and related research community. Instead of the attempt to develop a universal definition of information, this research

* Corresponding author. Wenjie Zhou，Tel.: +86-931-7975677; fax: +86-931-7971781.
*E-mail address:*wj_lp@sina.com

10.1016/j.procs.2018.04.319

endeavors to scaffold further researches aiming to develop the information theory under the context of Information Science.

2. Literature review

A number of reasons focus on the definition of information but far from receive an agreement in Information Science research community. The first and most obvious lies in the fact that information is a term widely used within and beyond academic disciplines (Derr, 1985; Robinson and Bawden, 2014). Many researchers attempt to define information as uncertainty or the reduction of uncertainty (e.g.Wersig, cited in Ingwersen and Jarvelin, 2005), for instance, demonstrates clearly the influence of Shannon and Weaver's information theory; the attempt to define information as the pattern or complexity of organization (e.g. Bates, 2006; Bawden, 2007; Madden, 2004) reveals the influence from theories of evolution; and the attempt to define information as a communicated message or the content of a message (e.g. Shenton, 2004; Svenonius, 2000) shows arguably the influence of the term's usage in ordinary discourse. When different sources of influence fail to find common ground for dialogue, disagreements inevitably arise. The second reason lies probably in the fact that the concept of information is closely associated with such epistemological concepts as knowledge, knowing, and truth, and like these latter concepts, is susceptible to the influence of fundamental philosophical disagreements. This kind of influence is most clearly revealed in the debate between Bates (2008) and Hjørland (2007, 2009) regarding the subjective/objective nature of information, and in the critiques of the DIKW (Data, Information, Knowledge, and Wisdom) hierarchy by Frické (2009), Ma (2012), Nielsen and Hjørland (2014), and like-minded scholars. Besides these two reasons, it can be argued that the lack of a solid methodology in defining information has also contributed to the divergences of its denotation. For an elusive phenomenon such as information, it is difficult if not impossible to base its conceptualization on empirical observation, for to define it in this way, we would have to know what to observe first.

3. Methodology

3.1. Literature retrieval and filtering

LISTA (Library, Information Science and Technology Abstract) literature database is used for academic artical retrieval. LISTA is huge academic paper collections developed by EBSCO, which nearly covered all the academic papers from Information Science and related research community. Search strategy of present study is that publish time limit to 2016 or 2017, document type is set as "article" for retrieving journal papers to meet our research target. A total of nearly 20,000 bibliographical references are obtained from LISTA database. Among the 20,000 bibliographical references, only 3593 papers meet the requirement (contain the abstract and not a book review or related non-academic works) of this research. Thus, the NLP is conducted based on these 3593 academic papers.

3.2. High-frequency-word revision and standardization

A natural language process (NLP) is conducted to absorb the high frequency words from abstract via Python, a computer language which knowing as a universal tool for textual data mining. Due to different words/terms may be selected for describing the same concept, it is necessary to standardize words/terms that are used to express the similar concept. For example, 1) singular or plural words are standardized to singular form; 2) IR is standardized to information retrieval; and 3) behavior and behaviour are standardized to behavior, etc.

3.3. VOSviewer and Visualization of high-frequency-word network

"VOSviewer" is a computer program that used co-word matrix based on the corpus selected from the full-text database. This program is used for the purpose of mapping co-words structures (Callon et al. 1986). Recently, many researchers tend to explain the co-word clusters in terms of graphs due to the development of graph theory and matrix operation (Courtial 1986; Xavier 2005).

3.4. Co-word analysis and key indexes

Co-occurrence of words, or so called 'co-words', has been considered as the carrier of meaning across different fields in science and technology studies (Callon et al., 1983). As a kind of content analysis methods, the co-word analysis is main analysis procedure includes: firstly, counting the times of term occurrence or co-occurrence in the same articles; Secondly, selecting terms according to the research topics; Thirdly, making similarity matrix; Fourthly, clustering the terms; Last but not least is to interpret the structure of topics represented by clusters (Ying, 2006).

Density and Centrality are two most important indexes to measure the co-words matrices. The strength of relations that make terms into a cluster is measured by density. Generally speaking, density of co-word network shows the total times that each pair of terms/works appears in an identified article. Based on the counting of the times of terms/words, a cluster analysis procedure is conducted and a ranked topic is formulated according to internal coherence of each cluster. The density can be measured via the average sum of internal links, the sum of the squares of the value of internal links, or the median value of internal links (Qin, 1999). In this research, we defined the density as following:

$$D(k) = \frac{\sum_{i=0}^{N} \sum_{j=0, j\neq 1}^{N} r_{ij}}{N-1}$$

where D(k) is the density of cluster k, N is the number of words in cluster k, and rij is the relation value between word i and word j which are both within the cluster k.

Centrality is an index to measure the connectedness of the clusters and it is represented by all external links within a cluster.(Courtial and Sigogneau, 2009). Due to the centrality shows the extent to which topics are central within a co-words network, we can rank topics according to their centrality. Based on related theory, we define the centrality as following.

$$C(k) = \frac{\sum_{i=0}^{N} \sum_{j=0}^{M-N} r_{ij}}{(M-N) * N}$$

where C(k) is the centrality of the cluster k, M is the number of all terms which are selected for clustering, N is the number of terms in cluster k, and r_{ij} is the relation value between word i within the cluster k and word j without the cluster k.

4. Results and discussion

4.1. Landscape of information science related research community

Density and Centrality of the whole co-words network was computed. Based on these key indexes, the cluster analysis procedure is performed and 4 clusters of research topics were separated out to describe the landscape of Information Science research community on whole via VOSviewer as Fig.1.

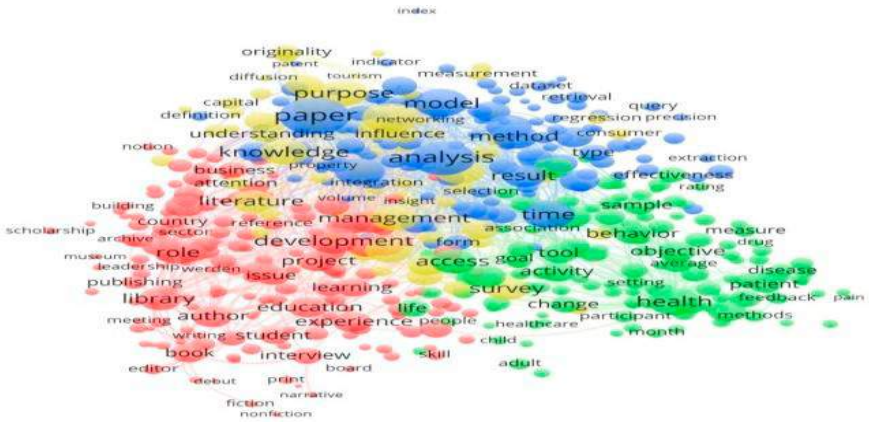

Fig. 1. Landscape of Information Science research community

In Fig.1, each bubble represents a research topic with different colors for different clusters and the size of the bubble represents the centrality of the research topic. For the purpose of exploring the various metaphor of information definitions among Information Science research community, present research then performed further analysis based on the landscape and 4 types of definitions of information were issued as following.

4.2. Information as records

As Hughes(2014) pointed out that spoken and written words, gestures, smells, pictures, memories, music, clothing and sounds all constitute information. Obviously, this first kind of the metaphor of information definition root in the idea that information is a kind of records. Furthermore, Brookes(1981) identified that "information is recorded in artifacts" based on the interpretation of Poppers' three world theory. From this point, Brookes wish that information science to be a science. Fig. 2 shows more details of this idea. From the perspective of defining information as records, various topics were covered as collection, content, literature and so on.

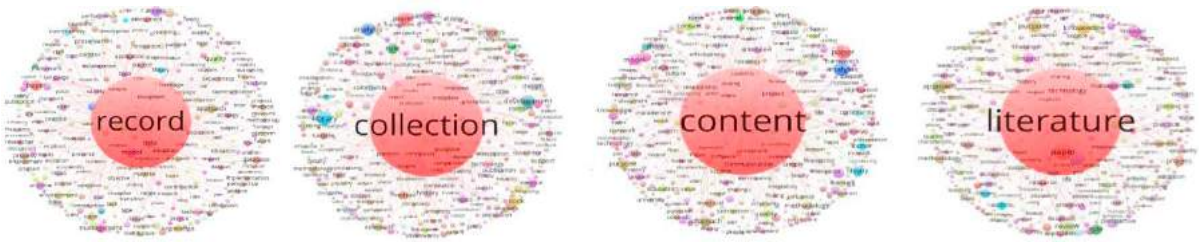

Fig. 2. information as records

4.3. Information as individual's experience

Dervin (1985), an American information scientist, used to point out that information can be whatever an individual finds 'informing'. According to the theory of Poppers' World Three, information is "Products of human thoughts" and is conceptualized as "objective entities—contents—embodied in physical forms". These ideas of information definitions also get a feedback from the research community of Information Science. As Fig. 3 shows that experience, engagement, understanding, originality and related research topic involved in this type of information definition.

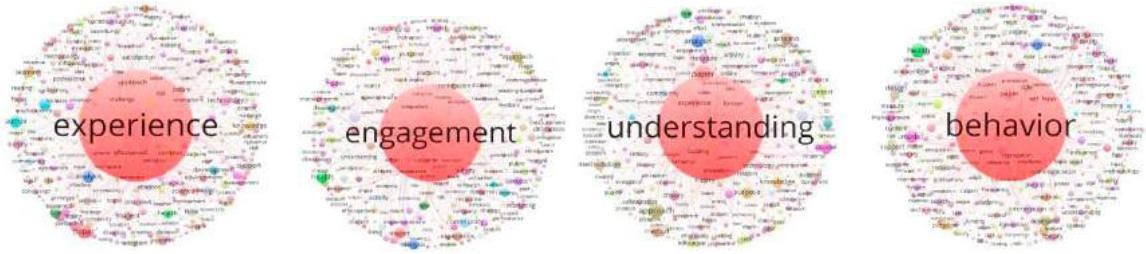

Fig. 3. Information as individual's experience

4.4. Information as the update of data or as the foundation of knowledge

According to Oxford Online Dictionary, "data that have been transformed through analysis and interpretation into a form which is useful for drawing conclusions and making decisions, that is the formulating process of information". In another words, information is update of data which comes from data but transform the data into a specific context. Furthermore, information regards as links to knowledge, but from different perspectives. Based on the consuming process of information, Knowledge is refined over time and becomes part of a world view or set of principles. Fig. 4 shows the updating process from data to information and knowledge according to DIKW theory of information developed by Fricke (2009), Middleton (2002) et al.

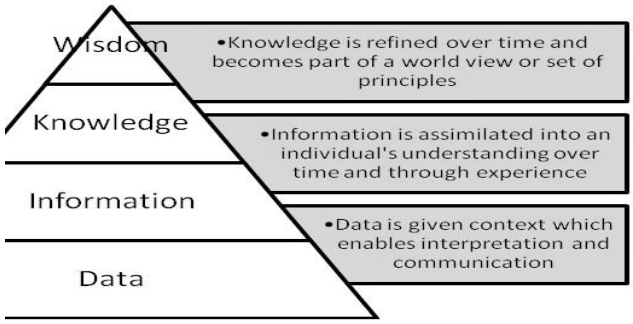

Fig. 4. Data-Information-Knowledge-Wisdom hierarchy (adapted from Fricke, 2009; Middleton, 2002)

The logic of DIKW theory also gets a reflection from Information Science research community. As Fig. 5 shows that ontology represent that how data is updated to information as well as how innovation and related works help to creative original knowledge.

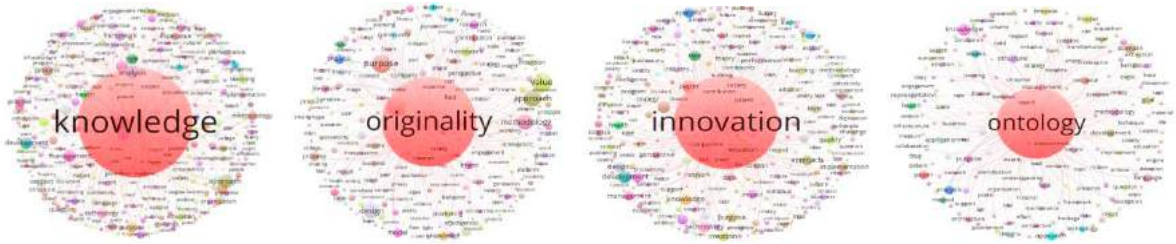

Fig. 5. Information as the update of data or as the foundation of knowledge

4.5. Information as ability of individuals

The definition of information, if not limiting to the Information Science horizon, is rooted in the fundamental theory developed by Shannon's Mathematical Theory of Communication (or Information Theory). According to Shannon, all about the definition of information is ability to transmit and receive a message. Due to Information Science mainly focus on the people's information behavior, Shannon's theory seems to make sense on the perspective of enable people to meet their information need base on specific information seeking ability. Fig. 6 shows the feedback of identify information as the individuals' ability.

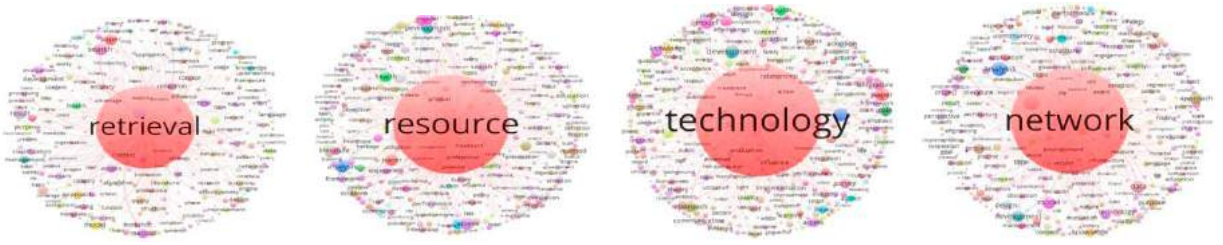

Fig. 6. Information as ability of individuals

5. Summary and conclusion

In summary, present study proposes four types of information definitions based on co-words for full spectrum analysis on abstracts of scientific articles. These reflect metaphor of information definitions on different perspective. A total of 3593 abstracts of academic papers have been covered in this study. A whole network of information definitions from Information Science research community and four-type maps are created to describe the metaphor of information definitions in this field. As results, present study summarized the definitions of information into four types under the context of Information Science research: information as records, information as individual experience, information as the update of data or as the foundation of knowledge and information as ability of individuals.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (project number:71563042) "The reliability and validity of co-words analysis in scientometrics". Wenjie Zhou is the correspondence author of this paper.

References

1. An, X. & Wu, Q., (2011), "Co-word analysis of the trends in stem cells field based on subject heading weighting", Scientometrics, Vol. 88 , pp. 133-144.
2. Bates, M.J. (2006), "Fundamental forms of information", Journal of the American Society for Information Science and Technology, Vol. 57 No. 8, pp. 1033-1045.
3. Bauin, S., Michelet, B., Schweighoffer, M. G., & Vermeulin, P. (1991). 'Using bibliometrics in strategic analysis: ''Understanding chemical reactions'' at the CNRS'. Scientometrics, 22(1), 113–137.
4. Bawden, D. (2007), "Organised complexity, meaning and understanding: an approach to a unified view of information for information science", Aslib Proceedings: New Information Perspectives, Vol. 59 Nos 4/5, pp. 307-327.
5. Budd, J.M. (2011), "Meaning, truth, and information: Prolegomena to a theory", Journal of Documentation, Vol. 67 No. 1, pp. 56-74.
6. Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). "From translations to problematic networks: An introduction to co-word analysis". Social Science Information, 22(2), 191–235.
7. Courtial, J. P., & Sigogneau, A. (2009). "How to use scientific and technological information to reveal strategic technologies". International Journal of Technology Management, 10(1), 31–44.

8. Derr, R.L. (1985), “The concept of information in ordinary discourse”, Information Processing and Management, Vol. 21 No. 6, pp. 489-499.
9. Frické, M. (2009), “The knowledge pyramid: a critique of the DIKW hierarchy”, Journal of Information Science, Vol. 35 No. 2, pp. 131-142.
10. Gnoli, C. and Ridi, R. (2014), “Unified theory of information, hypertextuality and levels of reality”, Journal of Documentation, Vol. 70 No. 3, pp. 443-460.
11. Hjørland, B. (2007), “Information: objective or subjective/situational?”, Journal of the American Society for Information Science and Technology, Vol. 58 No. 10, pp. 1448-1456
12. Hjørland, B. (2009), “The controversy over the concept of ‘information’: a rejoinder to professor Bates”, Journal of the American Society for Information Science and Technology, Vol. 60 No. 3, p. 643.
13. Hjørland, B. (2014), “Information science and its core concepts: levels of disagreement”, in Ibekwe-SanJuan, F. and Dousa, T.M. (Eds), Theories of Information, Communication and Knowledge, Studies in History and Philosophy of Science 34, Springer-Verlag, Berlin Heidelberg, pp. 205-235.
14. Ingwersen, P. and Järvelin, K. (2005), “The sense of information: understanding the cognitive conditional information concept in relation to information acquisition”, in Crestani, F. and Ruthven, I. (Eds), CoLIS 2005, LNCS 3507, Springer-Verlag, Berlin Heidelberg, pp. 7-19.
15. Lee, P. & Su, H., (2011), “Quantitative mapping of scientific research—The case of electrical conducting polymer nanocomposite”, Technological Forecasting & Social Change, Vol. 78 , pp. 132-151.
16. Madden, A.D. (2004), “Evolution and information”, Journal of Documentation, Vol. 60 No. 1, pp. 9-23.
17. Ma, L. (2012), “Meanings of information: the assumptions and research consequences of three foundational LIS theories”, Journal of the American Society for Information Science and Technology, Vol. 63 No. 4, pp. 716-723.
18. Qin, H. (1999). “Knowledge discovery through co-word analysis”. Library Trends, 48(1), 133–159.
19. Robinson, L. and Bawden, D. (2014), “Mind the gap: transition between concepts of information in varied domains”, in Ibekwe-SanJuan, F. and Dousa, T.M. (Eds), Theories of Information, Communication andKnowledge, Studies in History and Philosophy of Science 34, Springer-Verlag, Berlin Heidelberg, pp. 121-141.
20. Svenonius, E. (2000), The Intellectual Foundation of Information Organization, The MIT Press, Cambridge, MA.
21. Shenton, A.K. (2004), “Operationalizing the concept of ‘information’ for research into information behaviour”, Aslib Proceedings: New Information Perspectives, Vol. 56 No. 6, pp. 367-372.
22. Wellisch, H. (1972), “From information science to informatics: a terminological investigation”, Journal of Librarianship and Information Science, Vol. 4 No. 3, pp. 157-187.
23. Xavier, P. (2005). “Co-word analysis revisited: Modeling co-word clusters in terms of graph theory”. Proceeding or ISSI, 2, 662–663.
24. Ying, J. (2006). “A co-word analysis of bibliometrics in 1995-2004”. Journal of the China Society for Scientific and Technical Information, 25(4), 503–512
25. Yu, L., (2015) "Back to the fundamentals again: A redefinition of information and associated LIS concepts following a deductive approach", Journal of Documentation, Vol. 71 Issue: 4, pp.795-816.