

<https://dx.doi.org/10.4314/ijjkm.v10i1.3>

Information Impact:

Journal of Information and Knowledge Management

Vol. 10 (1), June, 2019

ISSN: 2141-4297 (Print) ISSN: 2360-994X (Online)

Big Data Moment and Faceted Theory: A Structured Review

Patrícia de Almeida

Faculty of Arts and Humanities | University of Coimbra

Largo da Porta Férrea, 3000-370 Coimbra, Portugal.

mebpatricia@gmail.com

Abstract

Over the last years, information growth has been the source of a large volume and variety of data, therefore new processes to represent, organize, seek and retrieve the information are needed. The background theory that supports the big data settings is diverse, however, one theory has deserved special attention in the literature, not only as a model for structuring data but also for locating information: the faceted theory. In this paper, a brief review of the literature about big data and faceted theory is presented. According to literature, web system architecture based on the faceted analysis appears to be highly common even in the big data settings. Regarding description, navigation, seek and search, and retrieval of the information, the authors write up the faceted theory, methods, and models in a digital environment. Nevertheless, more theoretical and empirical work is necessary to firm up this standpoint.

Keywords: Faceted theory, Big Data, Information seeking, Information retrieval, Web system architecture,

Introduction

Nowadays, one of the biggest challenges of humanity in a globalized world is to find individuality and personality. Regarding Information Science, the equivalent challenge for the information professionals is to find singularity in a big data world (or to make it possible for others). Using the analogy *Data Lake vs. Data Swamp*, it is crucial not fall in the so-called data swamp, a swamp where the data is unusable for analysis and finding until it is processed (Brackenbury, Liu,

Mondal, Elmore, Ur, Chard, & Franklin, 2018), that is why the big data issues are growing in interest and popularity in the research community, both academics, and public and private sectors.

Nanal (2018) states that there is no exact definition of how big a dataset must be to stay referred to as *big data*. This expression may not have a unanimous definition but is commonly used to describe an amount of data, structured or not, that can no longer be stored in the

static and traditional repositories (Davenport, 2014). This data cannot be processed without specific technical and computational structures, because it exceeds the traditional informatic systems capacities (Souza, Almeida, & Baracho, 2015; Nanal, 2018).

Over the last years, the universe of data is increasing exponentially. Lately, it was estimated that 2.5 quintillion bytes of data were daily produced (Wu, Zhu, Wu, & Ding, 2014). Until 2003, five exabytes of data were created, but nowadays this same volume of data is produced every two days. Consequently, it is predictable that data stored on the internet doubles every two years (Victorino, Shiessl, Oliveira, Ishikawa, Holanda, & Hokama, 2017).

Therefore, in the present and future time, in this world of huge quantity and a wide range of data, it is a real privilege to *get* information. The difficulties related to big data moment are associated with information embody capture, storage, seek and search, sharing, analytics, and visualizing (Nanal, 2018). This fact affects directly the work of governments, companies or researchers and, on the other hand, reaffirms the importance of

the analysis and representation of the information (Jiménez, & García, 2018). Whatever the sector in question, one always looks for the veracity, value and, velocity in the volume and variety of data – here are the well-known five V's of the big data (Gandomi, & Haider, 2015).

The developments in modern society make logical the data growing in different sectors. The big data settings are current in day-to-day life, not only regarding the seek and search, the customization, and the advertising of information, but also the recognition of information patterns, the predictive keyboard, or the development of research (Jiménez, & García, 2018). Consequently, it is necessary (new) standpoints to describe, to organize and to retrieval the information produced over time. And this is precisely what Information Science is about, a truly dynamic and challenging science with fluid and ubiquitous objects to study and work with (Souza, Almeida, & Baracho, 2015).

To Information Science and especially to knowledge organization, big data moment present itself as one more challenge. In other words, this means a provocative moment but also a real stirring thing up

for the information professionals. However, it is a shared challenge with Computer Science, where big data computing is perhaps the major innovation within the last decade (Nanal, 2018). Though, until now, it seems that the focus is more in the engineering of the big data and less in the meaningful structures and in the knowledge that it can make available. It is possible that the knowledge organization community has not been involved enough in the organization of the big data settings, but it is true that big data moment can benefit from a greater application and involvement of knowledge organization principles (Martínez-Ávila, Smiraglia, Szostak, Scharnhorst, Beek, Siebes, Ridenour, & Schlais, 2018).

The growth of new technologies in the XX and XXI centuries came to agitate this issue in different analytical perspectives, for example, create, describe, store, organize and use data. Therefore, the work of information

professionals is more and more important for society. That is why the organization, the representation and the visualization of the information in the big data settings are (and will continue to be) among the emerging areas that researchers will have to deal with (Shiri, 2014). In that way, to talk about big data is also to remember that one has never needed so much an Information Science like today, in particular, its theoretical and practical models for data architecture. In the literature, the concept of data architecture is a little bit ambiguous, but it is possible to simply understand it as a guide to structuring and locating any kind of information (Victorino et al., 2017).

The theoretical background that supports academic research about big data is diversified. However, one specific approach seems to be highlighted by some authors, as a model for organizing and retrieving information: the faceted theory.

The Faceted Theory

The faceted theory is a structured approach applicable to dynamic and flexible settings. A faceted analysis

recognizes many aspects in a certain item and that will allow its description as well as a synthesis process.

The faceted theory rises in the bibliographical classification systems, namely the ones that have their basis on the faceted analysis of the documents. Shiyali Ramamrita Ranganathan (1892-1972), labeled as the father of the Library Science in India, has the credit for introducing the method of knowledge analysis straight through the facets. That can be verified by reading his theoretical works and especially the Colon Classification (Ranganathan, 1933; Ranganathan 1937).

Garfield (1984) says that Ranganathan was the first to explain the faceted theory in a complete manner, but that Sayers, Bliss, and Richardson works contribute in somehow to that (Sales, Martínez-Ávila, & Guimarães, 2018). Likewise, Broughton (2006) states that other theorists before Ranganathan adopted similar ideas and concepts but in a limited way. Even though this is true, there is no doubt that the major landmarks of the faceted analysis of knowledge are Ranganathan's work and, some years after, the Research Classification Group studies. This group was founded in England, in the 50s of the XX century, and intended to develop theoretical and

practical works on knowledge classification (Lima, 2004).

In the faceted classification schemes, it is possible to split a given subject into its homogeneous features (or, in other words, into basic categories and intrinsic facets) and this procedure brings a lot of advantages regarding knowledge organization. In theory, in the analysis moment, the faceted systems allow the recognition of several details in only one subject; in the synthesis moment, the schemes try to encapsulate these same details in a way that enable a better description of the subject. In this way, it is feasible to clarify and explain the multidimensionality and the paths of knowledge (Lima, 2004). That's why previous research on facet analysis and information retrieval has usually focused on the construction of ontologies for specific domains (for example music domain - Madalli, Balaji, & Sarangi, 2015).

Among many other benefits mentioned by Broughton (2006), there is the fact that faceted systems provide tools for visualization and searching for the different points of the subject. In this way, it becomes an important method for the

organization and the visualization of data, as well as for web navigation. The faceted arrangements are common on the web. This is possible because of the logic and predictable structure of the faceted schemes that make them compatible and adaptable to the requests of the computer software. On one hand, the Information Science professionals, namely librarians, think that facets are a good path to organize things but are insecure about how to use it; on the other hand, the Computer Science professionals, who are already using it, are frequently unaware of the work of Ranganathan and of the Classification Research Group, and all the history behind facets (Denton, 2003).

Thus, the faceted theory and models came out suitable to be used in the description, the navigation, the seek and search, and the retrieval of the information in a digital environment. This is possible because it allows multiple paths – hypertext links. Consequently, it is understandable that, over the last years, several researchers had explicitly written up the faceted theory and models, especially in the web system architecture.

Early, Vickery (1965) considers Ranganathan's methods as a real model

for those who work with mechanized systems. More recently, Kashyap (2001) declares that the faceted analysis technique is, in some points, superior to the other techniques of information retrieval, namely in the online data basis of several information systems. Hudon's work (2006) shows that the basic principles of the faceted theory serve well not only the subject description but also the information retrieval. As an example, she explains the administrative documents management in the Government of Quebec (Canada) as a case in point. Duarte and Cerqueira (2007) state that the faceted model pops up as a helpful tool in the description of ideal concepts in hypertext systems.

Lately, Castro, Cruz, and Oddone (2013) affirm that Ranganathan's theory has influence in the modeling of the information systems and that the recent computer advances are subsidiary of a manual accomplishment. Gomes (2017) says that the fundamental elements to structure taxonomies and navigation interfaces remain in "Prolegomena to Library Classification" (Ranganathan, 1937), and, very significant, that the faceted method is almost unanimous in the semantic organization of the web.

Satija (2017) lists some research that proves that data directory and search engines are using Ranganathan's approach, with good results and efficiency in the information retrieval. He points out that is almost as if Ranganathan had previewed the web settings, for the simple reason that much of what is on the web depends on a faceted analysis.

Conclusion

Regarding the big data moment, the faceted theory has the potential to respond to the Information Science present challenge. In accordance with this brief review of the literature, the description, navigation, seek and search, and retrieval of the information in a digital environment can benefit from the faceted theory, methods, and models.

Despite the view of these authors, there are still remaining questions and uncertainty. Will faceted theory be an effective answer to the real quest for efficient solutions (Victorino et al., 2017)? According to Jiménez and García (2018), it is necessary more empirical work, in order to get a critical perspective that helps the understanding of all these issues.

Nevertheless, more theoretical and empirical work is necessary to firm up (or not) this standpoint. The results of further studies about faceted theory and related issues may question some topics or provide answers to others that will help information professionals, both Information Science and Computer Science fields.

References

Brackenbury, W., Liu, R., Mondal, M., Elmore, A.J., Ur, B., Chard, K., & Franklin, M.J. (2018). Draining the Data Swamp: A Similarity-based Approach. *Proceedings of the*

Workshop on Human-In-the-Loop Data Analytics, 13. Retrieved from <https://dl.acm.org/citation.cfm?id=3209911&dl=ACM&coll=DL>

Broughton, V. (2006). The need for a faceted classification as the basis of all

- methods of information retrieval. *Aslib Proceedings New Information Perspectives*, 58(1/2), pp. 49-72. Retrieved from <https://doi.org/10.1108/00012530610648671>
- Castro, F., Cruz, F., & Oddone, N. (2013). O paradigma da orientação a objetos, a linguagem unificada de modelagem (uml) e a organização e representação do conhecimento: um estudo de caso de um sistema para bibliotecas. *Informação & Informação*, 18(1), pp. 82-105. Retrieved from <https://doi.org/10.5433/1981-8920.2013v18n1p82>
- Davenport, T. H. (2014). Big data no trabalho: derrubando mitos e descobrindo oportunidades. Rio de Janeiro: Elsevier.
- Denton, W. (2003). *How to Make a Faceted Classification and Put It on the Web*. Retrieved from <https://www.miskatonic.org/library/facet-web-howto.html>
- Duarte, E. A., & Cerqueira, R. (2007). Análise Facetada: Um olhar Face a Modelagem conceitual. *Revista Digital de Biblioteconomia e Ciência da Informação*, 4(2), pp. 39-52. Retrieved from <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2020>
- Gandomi, A., & Haider, M. (2015) Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), pp. 137-144. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0268401214001066?via%3Dihub>
- Garfield, E. (1984). A Tribute to S. R. Ranganathan, the Father of Indian Library Science. Part 1. Life and Works. *Essays of an Information Scientist*, 7, pp. 37-44. Retrieved from <http://garfield.library.upenn.edu/essay/s/v7p037y1984.pdf>
- Gomes, H. E. (2017). Marcos históricos e teóricos da organização do conhecimento. *Informação & Informação*, 22(2), pp. 33-66. Retrieved from <https://doi.org/10.5433/1981-8920.2017v22n2p33>
- Hudon, M. (2006). Le passage au XXIe siècle des grandes classifications documentaires. *Documentation et*

- Bibliothèques*, 52(2), pp. 85-97. Retrieved from <https://www.erudit.org/fr/revues/documentation/2006-v52-n2-documentation01812/1030012ar/>
- Jiménez, A. G., & García, B. C. (2018). Una perspectiva documental y bibliotecológica sobre el *big data* y el periodismo de datos. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 74(32), pp. 77-99. Retrieved from <http://dx.doi.org/10.22201/iibi.24488321xe.2018.74.57910>
- Kashyap, M. M. (2001). Similarity Between the Ranganathan's Postulates for Designing a Scheme for Library Classification and Peter Pin-Sen Chen's Entity Relationship Approach to Data Modelling and Analysis. *DESIDOC Bulletin of Information Technology*, 21(3), pp. 3-16. Retrieved from <http://publications.drdo.gov.in/gsdli/collect/dbit/index/assoc/HASH7eef.dir/dbit2103003.pdf>
- Lima, G. (2004). O modelo simplificado para análise facetada de Spiteri a partir de Ranganathan e do Classification Research Group (CRG). *Información, cultura y sociedad*, 11, pp. 57-72. Retrieved from <http://www.scielo.org.ar/pdf/ics/n11/n11a03.pdf>
- Madalli, Devika P., Balaji, B. Preedip, & Sarangi, Amit Kumar. (2015). Faceted Ontological Representation for a Music Domain. *Knowledge Organization*. 42(1), 8-24. Retrieved from https://www.academia.edu/17512938/Faceted_Ontological_Representation_for_a_Music_Domain
- Martínez-Ávila, D., Smiraglia, R., Szostak, R., Scharnhorst, A., Beek, W., Siebes, R., Ridenour, L., & Schlais, V. (2018). Classifying the lod cloud: digging into the knowledge graph. *Brazilian Journal of Information Studies: Research Trends*, 12(4), 6-10. Retrieved from <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/8328>
- Nanal, A. (2018). Big Data Analytics Framework in Cloud Computing. *International Journal of Emerging Technology and Computer Science*, 3(4), pp. 92-96.

- Ranganathan, S. R. (1933). *Colon Classification*. [First edition]. Madras: Madras Library Association.
- Ranganathan, S. R. (1937). *Prolegomena to Library Classification*. Madras: Madras Library Association.
- Sales, R., Martínez-Ávila, D., & Guimarães, J.A. (2018). Dialogical elements in Harris, Dewey, Cutter, Otlet, Kaiser, and Ranganathan: Theoretical convergences in the history of Knowledge Organization. *Transinformação*, 30(3), 348-362. Retrieved from <http://dx.doi.org/10.1590/2318-08892018000300007>
- Satija, M. P. (2017). Colon Classification (CC). *Knowledge Organization* 44(4), pp. 291-307. Retrieved from http://www.isko.org/cyclo/colon_classification#ref
- Shiri, A. (2014). Linked Data Meets Big Data: A Knowledge Organization Systems Perspective. *Advances in Classification Research Online*, 24(1), pp. 16-20. Retrieved from <http://journals.lib.washington.edu/index.php/acro/article/view/14672/12312>
- Souza, R., Almeida, M., & Baracho, R. (2015). Ciência da informação em transformação: *Big Data*, nuvens, redes sociais e *Web Semântica*. *Ciência da Informação*, 42(2), pp. 159-173. Retrieved from <http://revista.ibict.br/ciinf/article/view/1379>
- Vickery, B. C. (1965). Ranganathan's work on classification. *Library Science Today*, 1. Kaula, P. N. (Ed.). Bombay: Asia Publishing House.
- Victorino, M., Shiessl, M., Oliveira, E., Ishikawa, E., Holanda, M., & Hokama, M. (2017). Uma proposta de ecossistema de *Big Data* para a análise de dados abertos governamentais conectados. *Informação & Sociedade: Estudos*, 27(1), pp. 225-242. Retrieved from <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/29299>
- X. Wu, X. Zhu, G.-Q. Wu, & W. Ding (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26, pp. 97-107.