

Information organization and visualization in cyberspace: interdisciplinary study based in conceptual maps

Mônica de Fátima Loureiro

Information Science Doctorate – Universidade de São Paulo (USP), São Paulo, SP, Brasil.
monicaeros@hotmail.com

Abstract

The paper argues the theoretical viability of the use of conceptual maps for information organization and visualization in virtual environment. Starting on the assumption that, in the Internet it is possible to develop catalogues with more attractive visual format to users, using conceptual maps. Being accompanied by a paradigm change, with the advent of the virtual environment, the text argues about of rhizome as a new model for information organization in this environment. The information is understood, here, as symbolic forms that are disconnected from material register. It is understood that this boarding applies to cyberspace, considering hypertext as a form of representation for informational mobility in the net, and the metadata as important tools to describe and to organize information. About metadata, some of its main aspects are argued, such as functionality, the importance of the context and the necessity of existence of interlocutors, that is, the possibility of interpretation by a man such as by a machine. For the assignment of an informational object in the Internet, it is necessary to retake the document concept throughout history, understanding itself as the lesser unit in the process of information organization. Entering into an alliance between the Information Science and Computer Sciences studies, it is understood that the process of information organization, in the virtual environment, will be “empowered”, favouring the users and taking them to acquire a greater knowledge about as research domain, as much as about the proper information search process.

Keywords: Conceptual maps, Cyberspace, Information organization, Information visualization.

Resumen

La comunicación trata la viabilidad teórica del uso de los mapas conceptuales para la organización y la visualización de la información en el entorno virtual. Se parte del presupuesto de que en Internet es posible desarrollar catálogos con un formato visual más

atractivo para los usuarios utilizando mapas conceptuales. El texto propone el rizoma, como nuevo modelo para la organización de la información en el entorno virtual. La información se entiende, aquí, como formas simbólicas que se desconectan del registro material. En el ciberespacio el hipertexto es el mecanismo de representación de la movilidad informativa en la red, y los metadatos las herramientas fundamentales para describir y organizar esa información. En relación con los metadatos, se discuten algunos de sus aspectos principales, por ejemplo su funcionalidad, la importancia del contexto y la necesidad de la existencia de interlocutores, es decir, la posibilidad de interpretación por un hombre al igual que por una máquina. Para la asignación de un objeto informativo en Internet, es necesario retomar el concepto de documento a través de la historia, entendiéndose como la unidad mínima en el proceso de organización de la información. Consiguiendo una alianza entre la Ciencia de la Información y la Informática, el proceso de la organización de la información, en el entorno virtual, se verá fortalecido favoreciendo a los usuarios y llevándolos a adquirir un mayor conocimiento acerca del dominio de investigación, al igual que sobre el proceso mismo de la búsqueda de información.

Palabras clave: Ciberespacio, Mapas conceptuales, Organización de la información, Visualización de la información.

1 Introduction

The first idea when regarding the Information and Knowledge Era, is that the technological advance basically highlights it. A general analysis of the society in Modernity shows a wide range of cultural representations, allowing people, even in temporarily, to adopt varied identities. Thus, it is possible to “flow” through cult and popular, traditional and modern groups (García Canclini, 1997). Therefore, it is understood that these connection possibilities are “potentialized” by the power of communication offered by technologies such as, for instance, the Internet.

Globalization may be considered as the icon of change in the Second half of the XX century, causing impacts even over cultural identity, as the possibility of overcoming the “unity” of national cultures, aiming at a global cultural homogenization. If, on the one hand, the place is characterized as different, on the other hand, the homogenization process is higher and higher, and national cultures are more and more connected, their communities are steadily changing, updating and standardizing culturally (Hall, 2003).

Hall (2003) reminds that traditional societies have been characterized as worshipping the past and their symbols, which have been given value because they contain and perpetuate ancient generation’s experience. Regarding information retrieving systems, such as libraries, it can be seen that, in the past, they were developed to aim at serving minorities, and many times cultural information, linked to people’s knowledge, was not contemplated into certain systems, thus laying aside and taken as simply popular one (García Canclini, 1997).

The new paradigm placed herein is ideologically based on globalization, interconnection, by democratization project that holds currently excluded sectors from access to information (García Canclini, 1997). On the other hand, it is also characterized by human identity fragmentation; that leads to a somewhat social disarrangement (Hall, 2003), among other

consequences. This critical look, however, requires all research on access to information to be more and more developed and motivated.

This present study places itself within the Knowledge Organization (KO) field, approaching, on a multidisciplinary manner, new theoretical possibilities to enhance information access within virtual environments. Therefore, in a nutshell, it is understood that organizing knowledge means creating forms to represent such knowledge, and up to a point, viewing its future retrieval. In this paper, information is taken as a basic need, as raw material for labor; it is information that goes through the avail of a scientific community, a research office, government or a socially accredited society, i.e., such information “which can be systematized and represented from specific contexts” (Campos, 2004, p. 25).

According to Campos (2004, p. 23), based on Le Moigne’s modeling theory, “knowing is modeling, i.e., the process of knowing is equivalent to building models of the world/domain to be constituted and which allow us to describe and provide explanations about the phenomena we observe”. The author also points out that in order for the knowledge domain modeling process to be possible, four principles must be used, as follows: a) the thought method used for knowledge organization within a domain; b) how the representation object is defined, i.e., what knowledge unit must be represented; c) the relationship among objects, aiming at verifying semantic link/separation possibilities among the concepts within a given domain; d) the graphic representation forms which a model may adopt.

Regarding the two levels pointed out by Svenonius (2001) as central ones for information organization: 1) representing the document; 2) representing the information (message, content); it is noticed that some information organization actions need to be recalled by their importance, such as the MARC (*MAchine Readable Catalog*) catalog and document representation format, the *Dublin Core*, and, more currently, the XML (*eXtensible Mark-up Language*) data markup language; and, within information representation (message/content), automatic indexing practices and ontologies.

According to Svenonius (2001, p. 31) ontology is “the science or study of being. More particularly, it is ‘a theory regarding the entities, especially abstract entities to be admitted into a language of description’”. That way, how can one organize information available at the web, when they own have varied characteristics? How can one understand WEB’s informational unity concept? One of the initial problems in this sense is the web informational diversity. In order to try to solve such issues, one may roughly think, the book is the informational unity in a library. On the Internet we notice that, for journals, the informational unity is the article. Concerning videos, what would the unit be? A single image? The video itself? Whatever owns autonomous meaning? Whatever can, summarizing, take up the informational content of registered information, for instance, in a video?

2 The registered information: concept of document

Ranganathan apud Svenonius (2001, p. 9), who understands that a document may be considered as the “embodiment of information”, i.e., the information registered within some sort of support: it may be a book, painting, recording and even a computer *chip*. Thus, a video is equivalent to a document. Robredo and Cunha (1994, p. 3), had long before understood that the term “document” was widened in its meanings and has incorporated “every sort of physical support to information which allows storage”, then including, mainly, puncture cards

used early in computing age, recorded magnetic tapes and computer programs. Shera (1980, p. 95) points out that Documentation – area which started as Otlet and La Fontaine’s initiative, characterized as the “process which enables gathering, classifying and publicize all documents of any sort, related to all sectors of human activity” – had early considered a document as informational unit, aiming at facilitating organization and future retrieval.

According to Buckland (1991), the term document or documentary unit was used as a general term that inferred informative things. Briet apud Buckland (1991, p. 355), defined document as “any concrete or symbolic indication, preserved or recorded, for reconstructing or for proving a phenomenon, whether physical or mental”. Therefore, a document could be anything from handwriting to a piece from a museum. In the same work, Buckland (1991) treats one information aspect as *thing*, referring to objects, data or documents which contain some information and aims at spreading knowledge or communicate information. Lima (2004, p. 26) quotes such representation object as “the lesser unit of manipulation/representation of a given context”.

Méndez Rodríguez (2002) discusses the importance of metadata within the Internet universe as facilitating tools for information organization, and, from a statement by Dempsey and Heery, associates metadata to objects, as shown below:

“Los metadatos son asociados con objetos [DLO] que liberan a sus usuarios potenciales [personas o programas] de la necesidad de tener un conocimiento avanzado completo sobre su existencia o características [...]. Los metadatos son conocimiento que permite a los usuarios, humanos o automatizados, comportarse de manera inteligente (Dempsey e Heery apud Méndez Rodríguez, 2002, p. 40)”.

The term used in the latter quotation, DLO, stands for *Document Like Object*, and, according to Mendez Rodríguez (2002), Internet information are documents taken as information objects, and then metadata will be used for describing and organizing such information. In conclusion, one may think metadata, according to aforementioned literature, are the most suitable tools for organizing the information within Internet virtual environment. In order to facilitate this text exposure, as of now, any sort of material which owns any informative function shall be named, in an indistinct way, “document”.

Metadata have three main aspects: their functionality, both for identifying and for describing, the Internet *context*, as the “venue” to be organized, and the *interlocutor*, i.e., the possibility of being interpreted by machines – linking to the current context, totally influenced by the Information and Communication Technologies (ICT). When facing the several existing definitions for metadata, a remark is deemed needed: there is no better definition than another, yet there is one that is more suitable for each area of study and specialization. Then, for this present study, metadata is understood as “significant data which represent other district data objects [...] structured descriptions of a computerized object”, i.e., DLO (Gill quoted by Méndez Rodríguez, 2002, p. 45).

Méndez Rodríguez (2002, p. 46) characterizes an information object as “an element or a group of elements which constitute an informative unit that a computer may handle, regardless the file format and the kind of information it refers to”. It is useful to highlight that the concept of informational object, even though it has already been concerned in IC literature, such as in Buckland’s (1991) studies; he looked into information in three levels: process, knowledge and thing, receives now a new direction towards the virtual environment

– towards the Internet organizational universe –. We understand, however, that what currently changes is the context influenced by the ICT; this context interferes in attitudes both from the users who need such information and from the professionals who will organize information viewing their future retrieval.

It is relevant to quote Méndez Rodríguez's (2002, p. 46) statement, who says, "concept abstraction is the key to electronic information management, and, from our point of view, the fundamental of metadata". We need a sort of information organization that handles the essence of each document. From that point of view, one may soon think about indexation and documentary representation processes, which are not new and aim at providing the essence of any bibliographic material, to ensure efficient retrieval to the user. On the other hand, we must also consider that, lately, the idea is that more and more automated processes will be used to perform indexing operations¹; even automatic search engines, such as the *Google* as well as other well-known ones, have proved to be not enough a position for sorting out the problem.

3 Information Retrieval in virtual environment: users' behavior

The information processed by the human brain is assimilated and produces knowledge. Generally speaking, people search for information based on a certain need. In this sense, Barreto (1994) used the Maslow's Pyramid to understand informational demands and detected the existence of three groups: Group 1 – people who own *basic needs* of information: those who need *utilitary information*, i.e., useful information to ensure their safety; Group 2 – people who own *participation needs*: those who take part in any existing social groups, such as the workplace, study place or any other social aspect. These people need *contextual information*, according to the interests of the group they belong to; Group 3 – people who own *self-achievement needs*: those whose previously mentioned needs have already been met, and who look for professional or personal achievements. Due to this, these people need *selective information*, for they will use it for reflection and creativity.

Regarding users' behavior facing the Information Retrieval process, Ingwersen (2002) reports by presenting four basic kinds of searchers, according to their experience levels in using Information Retrieval or intermediate systems: 1) Expert; 2) Subject specialist; 3) Information Retrieval Specialist; 4) Non-specialist.

The first kind of searcher is called Expert; Ingwersen (2002, p. 143) considers it as the "person who possesses both types of knowledge at event of retrieval. His state of knowledge and problem space match the actual 'information space'. As pointed out above, such a person may lose his 'expert' status in other IR situations".

The subject specialist is "a person who possesses conceptual knowledge within that domain in which he is performing retrieval at a given moment" (Ingwersen, 2002, p.142). The Information Retrieval Specialist "is a person who may be regarded as a subject generalist, except within the domain of IR in which he supposedly is an 'expert'" (Ingwersen, 2002, p. 143).

¹ For several reasons, they are not the focus of this study, and therefore they will not be deepened, such as different visions from the indexing professionals, amount of information to be organized, target public of such information, etc .

Finally, the fourth kind of searcher indicated by Ingwersen (2002, p. 142) is the Non-Specialist, the one “who, in his actual state of knowledge and problem space at event of IR, possesses insufficient knowledge of both types to perform retrieval effectively in a given ‘information space’”. It is important to remind that Ingwersen (2002) established the four types of searchers mentioned based on their knowledge characterization, not in their possible roles in Information retrieval. The root for such determination came out of statistical groupings made in investigating online population in several countries.

In conclusion, one may infer that one of the most important to be concerned by the information organization process within the virtual environment is information retrieval by users, reminding that a large number of them do not own deep knowledge of information retrieval processes, and that leads to problems regarding their searching behavior. Even bearing in mind that several recorders are, mostly, targeted to a general public profile, what happens is that they must be ready for assisting different experience levels, not only in information retrieval but also in database.

4 Information Organization

Talking about information organization means talking about some sort of classification, as the latter is inherent to the act of organizing. Roughly, in order for something to be classified, it is needed to know its concept, and then a *ranking* is established, arising from a number of the most relevant characteristics which infer some sort of category. Dahlberg apud Himalata (1995, p. 40), comprehends that a concept can be defined as “a knowledge unit, comprising necessary and verifiable statements about a referent, being represented by a designation”. Concepts provide a global view of a certain thing, differing it from an impression noticed when something is only heard or seen. A concept may be basically represented by the *referent*, that is the thing; by the *verbal expression*, which represents such thing and by the *characteristics* of such thing (Himalata, 1995). For instance, one may think of the referent CAT, of its verbal expression Cat and, of a cat’s general characteristics, thus forming a concept of what a cat is.

Starting from concept establishment, one proceeds to the categorization phase, which is roughly category establishment. Jacob and Shaw, quoted by Lima (2003, p. 82) understand categorization as “a cognitive process of sharing the world experiences in groups of entities or categories, so as to build a physical and social world order”. Lima (2003, p. 92) points out that categorization has grown larger than “an individual cognitive process to a cultural, social process of building reality, which organizes concepts based on thought psychology”. Himalata (1995) points out that categorization is based on similarity, and the equivalence among several concepts can be determined by common properties in these concepts, and they may vary when based on objectives; physical characteristics or attributes; based on the environment; or based on function. Afterwards, structured hierarchical groups containing common properties are developed. Lima (2003) adds that there is also the context influence: this is an important component in the categorization process.

The human brain, early from its cognitive activities, performs a series of classifications and connections between concepts and referents. According to San Segundo Manuel (1996, p. 31), every classification is the “result of a reality structuring artificially elaborated and contrasted with the one around us”. Ingwersen’s (2002) studies, based on human cognitive processes, have presented A. R. Luria’s object classification elements, which were developed in the

1930's with nomads in Central Asia; they had, as contribution, two kinds of classification: the Categorical and the Situational.

The Categorical classification indicates "that individuals sort out an abstract concept and choose the objects which can be included under this concept" (INGWERSEN, 2002, p. 129). The Situational classification implies that "individuals involve the objects in different concrete situations, thereby grouping objects which belong together" (INGWERSEN, 2002, p. 129).

One thinks that, for the existence of information organization on the Internet that facilitates retrieval and considering the different kinds and levels of knowledge by the users, several classifications and categorizations have to be made. However, Ingwersen's (2002) studies on the cognitive activities targeted to developing information retrieval systems state that humans quite often, according to their knowledge level on a domain and depending on their educational and cultural level, make connections between concepts and terms, many times due to situations they have been through, influenced by individual emotional factors.

5 A change of paradigm: from physical to virtual

Early in this text, a discussion on what an informational unit represents was brought up, and it was understood that, from literature data (Shera, 1980; Buckland, 1991; Robredo and Cunha, 1994; Svenonius, 2001), the document, also called documentary unit, was appointed as the informational unit for the information organization process.

All these informational units, the documents, throughout History, have always had a physical support: manuscripts, books, journals, museum items, file documents, recordings, videos, CD ROMs, etc –, even depending on such support for being organized, and for the information they contain no to be lost as time passes by. Although all these documents carry information, they themselves, from their supports, owned a special way of organization, which were based on these types of supports. As for the Internet and its huge informational collection, despite the fact that many types of information are kept in two formats: physical and virtual, some of them are only virtual, i.e., they only exist in the Net. This factor causes a real change of paradigm for information organization, since it dislocates the idea of organizing physical collections or information storages, into virtual, untouchable collections.

Thus, Monteiro (2003) presents a philosophical discussion on knowledge virtual organization on the net, i.e., on the cyberspace. The writer highlights the virtual as a reality which yet influences the way the objects are, different from the languages and the works, once it sets itself apart from materiality and, therefore, from the physical control and organization. The term virtual comes from the Latin word *virtualis* rooted *virtus*, which designates strength and power. So, "the virtual exists in power, not in act; that is why it has the current pole, not the actual one, commonly associated to the term" (Monteiro, 2003, p. 2). This is the reason why the updating process is inherent to the virtual environment: we change the reading space, the writing space, etc. The hypertext comes as a language and a mean, putting the text in motion and setting aside a whole format of linear reading, as well as being accessible everywhere. Therefore, the text is discussed about not owning printed, static, linear characteristics and not big stuck to "the object materiality" (Monteiro, 2003, p. 6).

Initially, the information organization process was characterized by classifications that mostly followed hierarchization, i.e., they owned a logical sequence that divided and combined concepts and ideas. Burke (2003) reminds that the analogy with the tree had major influence on knowledge classification throughout History, starting with the knowledge tree proposed by Raymond Lullius in circa 1300 A. D., in which there was ascending order, from the roots to the leaves and fruit, inferring a clear distinction between the dominant and the dominated sides.

Monteiro (2003) points out that, with dematerialization in the cyberspace, the classic knowledge organization (cataloging and classification) is deemed limited, for cataloging, for instance, deals with description of document format. The author says that the rhizome would be the new knowledge organization model currently present in the virtual environment, thus opposing to the tree model previously present (Deleuze and Guattari, 1995 apud Monteiro, 2003). Moreover, the rhizome is directly related to the hypertext as the new form of representation in the virtual environment.

The rhizome own characteristics as follows: 1) connection: this would be the possibility of connection to any point of the system, not according to a hierarchical order; 2) heterogeneity: the ability of a search result to come up in several languages, images, texts, songs and so on; 3) multiplicity: In the cyberspace there is no structured, hierarchical, linear significance center with a single sense and a fixed reference; 4) a-significant rupture: there is no logical-linear sequence, for example, from genders to species, from general to specific, as it takes place in hierarchical classifications, there is no meaning center, and yet leaving all possible meanings for retrieval; 5) cartography: the idea of map is perfect for illustrating the rhizome, since it does not have a center or hierarchy; it has several entries and moving forms (Deleuze and Guattari, 2004).

In addition, Herrero-Solana and Morales-del-Castillo (2004) state that map construction may graphically reflect actual relations established among varied entities. The authors add that these representations may be important tools to perform analysis of reality itself, besides serving as a valuable source of information. It is interesting to bring back Van Raan's (2003) vision, who states that publication grouping procedures, based on the similarity of informational elements – the concepts –, are the most suitable ones for mapping science, bearing in mind that bibliometric mapping processes are promising, as they are characterized as “unique instruments to discover patterns in the structure of scientific fields, to identify processes of knowledge dissemination, and to visualize the dynamics of scientific developments”. They may also potentially work as “unraveling interdisciplinary developments and interfaces between science and technology” (Van Raan, 2003, p. 20).

Thinking of the Internet as an informational recorder to be organized, using maps can be a feasible proposal. Monteiro's (2003) study confirms such possibility by pointing out that the virtual environment or cyberspace does not own the same characteristics for knowledge organization as the real (physical) world. It is emphasized that this is an initial approach to some ongoing research. In the future we shall proceed to the study empirical phase, following the line stated by Herrero-Solana and Morales-del-Castillo (2004) regarding the use of graphic representations –maps – as means of access to information. These authors understand that knowledge or science map studying, usually constituted by quantitative methods, allows the identification and representation of relations between knowledge field investigators, research trends and also dominant paradigms.

6 Final Considerations

We consider that Internet, in a wide view, cannot be considered as a “domain” for information organization. It does not represent totality, i.e.; it represents the global context. The virtual environment brings a change of paradigm to the information organization field, “dstructuring” classic knowledge classification bases established for such purpose. Besides, it has also influenced document format, which have lost the ever-present linearity in written or printed documents.

Hypertext is pointed out both as a language to the Internet and as a means to it. On the other hand, the metadata is understood as one of the most important tools for aiding the information organization process within the virtual environment. The use of conceptual maps in this environment has turned out to be a theoretically proven possibility. Aligning Information Science and Computer Science studies, the information organization process in the virtual environment will be potentialized, pleasing users and leading them to acquire larger knowledge both of a research domain and regarding their information search process.

References

- BARRETO, A. A. A questão da informação. *São Paulo em Perspectiva*, 1994, vol. 8, n. 4, p. 3-8.
- BUCKLAND, M. K. Information as thing. *Journal of the American Society for Information Science*, 1991, vol. 42, n. 5, p. 351-150.
- BURKE, P. *Uma história social do conhecimento: de Gutenberg a Diderot*. Rio de Janeiro: Jorge Zahar, 2003.
- CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ciência da Informação*, 2004, vol. 33, n. 1, p. 22-32.
- DELEUZE, G.; GUATTARI, F. *Mil platôs: capitalismo e esquizofrenia*. Vol. 1. Rio de Janeiro: Ed. 34, 2004.
- GARCÍA CANCLINI, N. *Culturas híbridas: estratégias para entrar e sair da modernidade*. São Paulo: EdUSP, 1997.
- HALL, S. *A identidade cultural na pós-modernidade*. Rio de Janeiro: DP&A, 2003.
- HERRERO-SOLANA, V.; MORALES-DEL-CASTILLO, J. Mapas “geopolíticos” de internet: aplicación de las nuevas técnicas de representación de la información. *Ciência da Informação*, 2004, vol. 33, n. 3, p. 69-75.
- HIMALATA, I. Cognition and categories. In: HIMALATA, I. *Classificatory structures: concepts, relations and representation*. Frankfurt: Indeks Verlag, 1995. p. 40-59.
- INGWERSEN, P. *Information retrieval interaction* [electronic resource]. London: Taylor Graham Publishing, 2002. <www.db.dk/pi/iri>. [Consulted: 11 mar. 2003].
- LIMA, G. A. B. Interfaces entre a ciência da informação e a ciência cognitiva. *Ciência da Informação*, 2003, vol. 32, n. 1, p. 77-87.
- MÉNDEZ RODRÍGUEZ, E. *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón: Trea, 2002.

- MONTEIRO, S. D. A organização virtual do conhecimento no ciberespaço. *DataGramaZero – Revista de Ciência da Informação*, 2003, vol. 4, n. 6, p. 1-20.
- ROBREDO, J. CUNHA, M. B. *Documentação de hoje e de amanhã: uma abordagem informatizada da biblioteconomia e dos sistemas de informação*. São Paulo: Global, 1994.
- SAN SEGUNDO MANUEL, R. *Sistemas de organización del conocimiento: la organización del conocimiento en las bibliotecas españolas*. Madrid: Universidad Carlos III de Madrid, 1996.
- SHERA, J. H. Sobre biblioteconomia, documentação e ciência da informação. In: GOMES, H. E. (org.). *Ciência da informação ou informática?*. Rio de Janeiro: Calunga, 1980. p. 91-105.
- SVENONIUS, E. *Intellectual foundation of information organization*. Cambridge: MIT Press, 2001.
- VAN RAAN, A .F. J. The use of bibliometric análisis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technikfolgenabschätzung: Theorie und Praxis*, 2003, vol. 1, n. 12, p. 20-29.