

Information retrieval in institutional repositories: Proposal of an ontology for Historical Organizational Memory

Janete Saldanha Bach Estevão and Faimara do Rocio Strauhs

Federal Technological University of Paraná – UTFPR/PPGTE

Abstract. The Historical Organizational Memory consists of the main documents that register the trajectory of an institution. The ontologies aim to make sense of the content of these repositories, using the semantic relationship of the terms that represent the area. This work establishes the basis for proposing an ontology that serves as a reference model in the field of Historical Memory Organization, promoting accurate responses from search engines and information retrieval.

Keywords. Organizational Historical Memory. Organizational Knowledge Management. Ontologies. Representation of Information

1. Introduction

The concept of Organizational Memory (OM) is addressed, in this article, under the perspective of the historical trajectory of organizations, and for this reason acceptance was made to include the qualifying term Historical, to make up the expression Historical Organizational Memory (HOM). This differentiation becomes necessary because the term Organizational Memory is used in literature as foundation in projects of management, storage and retrieval of organizational information in specific areas of the company, like the Accounting, Legal and Technical departments, among others (ESTEVÃO; STRAUHS, 2011). However, according to Almeida (2006, p. 82), “to make an OM feasible, it should be considered in a wider context, as an information system that enables an organization to store and retrieve comprehensively the disperse knowledge in the environment”.

Organizational memory focused on historical retrieval has been used as a tool for management support in institutions of all sizes and from different sectors of the economy in Brazil. There are many possibilities of products developed based on actions of Historical Organizational Memory, like institutional publications, exhibitions and museums, and Documentation and Memory Centers, among others. These products have various possibilities of application, and based on their

Received: 8.4.2015 / Accepted: 9.11.2015
© ISAST

ISSN 2241-1925



characteristics internal and external communication can be reinforced, promoting the institutional marketing of products and services, supporting endomarketing, and for the strategic management of organizations. If the information of the business models, designed and applied, are preserved, the changes in the form of performance of organizations and consequent results, strategic plans of the past, registers of hits and misses, to be avoided or repeated in the future, can be retrieved.

However, the fact that there is an HOM initiative in a certain institution does not mean that it is used optimally and strategically (BROOKING, 1999, p.155). An example of this is the strategic use of historical information that is contained in publicity documents, streamed in magazines and newspapers used as judicial foundation, or data and information related to the development of products and services, from the first prototypes to the first batch produced, transformed into briefings and business cases.

These documents record important information about the trajectory, culture, processes and memory of the Organization. However, according to the results of the research presented in this article, normally the documents and historical institutional content, stored in repositories, do not follow a standardized representation of information, which adds effective value to it, or favors its optimized use and retrieval. This study proposes an ontology that serves as a reference model in the field of Historical Organizational Memory, elaborating a map of concepts and their semantic relationships, which is the basis of development of search engine and information retrieval tools that return more precise and effective responses.

Various authors have attested to the application of ontologies to improve information retrieval processes (BERNERS-LEE; HENDLER; LASSILA, 2001; JASPER; USCHOLD, 1999; GUARINO, 1997; MCGUINNESS, 1998; SOWA, 2000; W3C, 2011a; WELTY, 1998). However, the studies about ontologies related to Historical Organizational Memory in Information Science are still incipient. A bibliographical research returned only one article for the period from 1977 to March 2006 (NASCIMENTO et al., 2007, p. 35). This same research was replicated in March 2012, on bases like Library & Information Science Abstracts (LISA), Documents in Information Science (DoIS), E-prints in Library and Information Science (E-LIS) Information Science & Technology Abstracts (ISTA), besides Google Scholar, retrieving only three published articles from 2006 to 2012.

This article is based on the assumption of an ontology that may contribute towards information retrieval, because it promotes the alignment of concepts and relationship in a certain field, identifies the documents/contents of the repository related to the field, strengthens and maximizes the results of information retrieval, and supports organizations in the retrieval of information and management of the registered knowledge.

2. Concepts and consequences of organizational memory

In organizations, the memory is corporate and originates from individual memories, despite being a space for multidisciplinary investigation and work. The knowledge registered by individuals interacts with their own knowledge of the roles and

responsibilities they assume in the organization (REMOR; RADOS; REMOR, 2009, p.88).

According to Walsh and Ungson (1991), organizational memory is directly related to the information contained in the trajectory of an organization. This information, fruit of decisions already taken, of cognitive processes of individuals and mutual interpretations, are liable to be retrieved and used in current decisions (WALSH; UNGSON, 1991). Organizational Memory in this perspective is a joint construction, at individual and organizational level. For such, Walsh and Ungson (1991) highlight the processes of acquisition, retention and retrieval of information inside the organizational environment. Such processes, under a systematic view, remit to what is conventionally known as Knowledge Management (KM).

The objective of Historical Organizational Memory is to collect, preserve and disseminate systemically the resources of knowledge on the trajectory of an organization, as an action of KM. The forms of applying the historical organizational memory go from the retrieval and restoration of founding documents and photographic files to the structuring of Memory and Reference Centers, where the memory is applied as knowledge, says Nassar (2009). In this context, tacit and explicit knowledge apply, two dimensions widely discussed and conceptualized in the book by Nonaka and Takeuchi (1997, p. 113). The authors, based on this differentiation of tacit and explicit knowledge initially proposed by Michel Polany in 1966, define the first as being “personal, depending on the context, and, therefore, difficult to be formalized and communicated”. Explicit knowledge, however, can be transmitted in a formal and systemic language, and is more objective and associated to reason and theory. However, it represents a small part of the total existing knowledge (NONAKA; TAKEUCHI, 1995, p. 59-60).

Based on the understanding that explicit knowledge is constituted by data that has somehow been coded, these can be processed by information technology and communication systems, enabling their storage and retrieval on database (SPILLER, 2005, p. 29). Taking the above into account, it can be understood that an HOM action used as a tool of Organizational Knowledge Management should include ways of extracting tacit knowledge, so that it may become explicit, coded, registered, preserved and made accessible through computer tools, so that, once accessed and processed mentally by other individuals, it becomes tacit knowledge once again (CHOO, 2003, p.28).

Due to having more accentuated external visibility, mainly if the access is extended beyond the borders of the company, the HOM has become one of the “sponsors” of Knowledge Management. However, HOM products focused on historical retrieval like “show rooms and history books, publications, exhibitions and events” (MUSEU DA PESSOA, 2005), besides its institutional document collection like “objects, packages, photographs, campaigns, documents and registers with a permanent keep-safe value” (DAMANTE, 2004, p.30) need to be used effectively to foment Knowledge Management. Thus, it is important to systemize the coded information and organize the registered knowledge that the company has on itself, in a linear form. And also promote the oral retrieval from employees of the memory that is not

formally registered in manuals or documents. This would be, without any doubt, a form of extracting tacit knowledge.

Highlighted, with the corroboration of Tonini and Gagete (2004), as the main HOM product is the Documentation and Memory Center (DMC), due to having a continuous process of historical preservation of the Institution. The DMCs have similarities in relation to functions of custody of documents, as found in museums, files and in libraries. However, they have particularities in relation to their specific functions, as to how their collections are constituted, concerning the characteristics of their documents, technical processing, information products and services, and the profile of their audience.

3. Information retrieval and the ontologies

The availability of the informational contents by the methods and techniques of biblioteconomy is made through representations about the knowledge contained in the books of a constructed collection, independently of its support being physical or remote. These representations are just as descriptive as they are thematic. They have the purpose of describing the resources of a document in order to facilitate and make the search and retrieval of information more effective. When the library of the American Congress started using computers, in the 1960s, a standard of cataloging was created giving origin to the format MARC 21 (Machine-Readable Cataloging Record) , “a system that uses short numbers, letters and symbols in a register for cataloging, which identifies different types of information” (CONGRESS LIBRARY, 2009).

These efforts were made precisely so that that this data could be read and operated by machines, enabling the automatic exchange of metadata of a certain publication between applications of different libraries. It was perceived that, in the 1960s, with the arrival of the MARC standard, the concept of inoperability of systems already applied to biblioteconomy, i.e. the “exchange of information between systems” (CONGRESS LIBRARY, 2009).

It is also important to state the initiative of the Online Computer Library Center (OCLC) in the creation of the WorldCat , which is a global network providing services to libraries. It was built cooperatively based on the development of metadata databases for bibliographical and institutional registers, for the exchange of these registers between participants of the network. Currently, almost 1.7 billion registers are exchanged (ONLINE..., 2011).

The other form of description of a document appointed by Cesarino and Pinto (1980), intellectual or subjective, refers to the identification of its informational content. Cesarino and Pinto (1980) describe the process of establishing the subject of a document in three stages. Initially, the need to understand the document as a whole. In the second stage, a selection is made of the concepts that best describe the subject of the document. The third stage consists of refining the selection of the second stage, selecting only terms that are actually valid for indexation. These terms are validated descriptors by a type of standardizing instrument. These instruments, called thesauri, controlled vocabularies and taxonomies, can be acquired,

customized or constructed based on methodologies already consecrated in literature (DAHLBERG, 1978; INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 1984; GOMES, 1990).

The whole process of indexation is made based on implicit knowledge, grounded on specific training for qualification in the understanding and use of these instruments, in the analysis of the subjects and in the selection of the descriptors that best represent these subjects, considering issues of polysemy and synonymy .

In this sense, the ontologies arise as a proposal that allows the partial automation of this process, without losing the richness of the tacit knowledge contained in people, but expliciting it so that it is structured, standardized, legible or can be machine interpreted, allowing the interoperability (exchange) of information through systems, gathering documents by area of interest that can be shared and reused (USCHOLD, 1996).

The studies in ontologies applied to the retrieval of information began in the area of Artificial Intelligence (AI), branch of Computer Science (CS), in the 1990s. The concept of Guarino (1998, p.2) considers ontology in AI as an artifact of engineering:

in its most prevalent use in AI, an ontology refers to an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words. This set of assumptions has usually the form of a first-order logical theory, where vocabulary words appear as unary or binary predicate names, respectively called concepts and relations. In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships; in more sophisticated cases, suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation. (GUARINO, 1998, p.2).

A mark in the researches of how to improve the responses of the search engines on the internet was the creation of the World Wide Web Consortium (W3C) by Tim Berners-Lee and other researchers in 1994 (W3C, 2011). W3C makes use of principles like interoperability, evolution and decentralization to identify new Technologies and standardization for the web.

The first step in attributing semantics to the web, envisaging better search and retrieval of information, is the construction of the ontologies (BERNERS-LEE; HENDLER; LASSILA, 2001). Ontologies have been used by W3C as instruments of terminological standardization to enable these results. They are composed of taxonomy and of inference rules. Inferences are the deductions that the machine makes automatically based on the data supplied. This is evidence, which is considered important, from the contribution to the advance of knowledge in this field.

Due to the structure of an ontology being composed of terms, definitions and relations, it is characterized as a documentary language, but with one advantage: it has functions that allow the automated processing of the rationale by the machine, grounded on rules and inferences (SALES; CAMPOS; GOMES, 2008, p. 63). In relation to inference, Dziekaniak (2010, p. 53) states that “this property makes ontologies the best solution currently for information retrieval systems, considering that they can be processed automatically making use of logical and axiomatic relationships”. Kobashi states that (2007) “the ontologies intend to go beyond the documentary thesauri. These approaches expand the possibilities of organizing and having access to information on systems”.

The similarities of ontologies with documentary languages, in the perception of Campos (2004), Sales (2006), Boccato, Ramalho and Fujita (2008), Ramalho (2010), is sufficient to consider them as an instrument of representation and of information retrieval. Thus, ontologies can also be addressed under the optics of Biblioteconomy and Information Science, also concerned with the development of information retrieval systems.

4. Methodology procedures of the research

The research was carried out in 11 institutions that have already applied initiatives in HOM and which have DMCs, participants of the development program for the memory of organizations, the Permanent Forum on Knowledge Management, Communication and Memory of the Brazilian Association of Business Communication (ABERJE). The Forum allows the exchange of experience between the associates and external guests. Among the attributions of the Forum is to “disseminate information about history and corporate memory, promote KM, stimulate and recognize practices of excellence, gathering professionals with common interests and challenges” (ASSOCIAÇÃO BRASILEIRA DE COMUNICAÇÃO EMPRESARIAL, 2008).

The object of the study was investigated based on a field research, applying interviews in the semi-structured modality, which allows the respondents to have the same set of standard questions. After the selection of analysis techniques of the responses, the interpretation of the interviews was based on the analysis of content by Bardin (1995). This process is given through the “dismemberment of the discourse into categories, in which the criteria of selection and of delimitation are guided by the dimension of investigation of the themes related to the object of research, identified in the discourses of the subjects researched” (BARDIN, 1995, p. 80-81). Based on a group of techniques of communication analysis, enabled through the systematic procedures and description objectives of the content of the messages, obtain quantitative indicators or not, which allow the inference of knowledge related to the conditions of production and reception of the messages.

The companies researched are among the ranking list of the “500 biggest and best companies in Brazil” in the following areas of expertise: Corporate Communication, Capital Markets, Agribusiness and Foods, Engineering and Civil Construction, Editing and Communication, Aircraft Industry, Commercial Aviation, Pulp and Paper, Fuel Distribution, Cement, Mining and Metallurgy.

5. Presentation of the results

The results refer to the layer of definition of the ontology, i.e. at terminology level, in which the basic group of concepts and relationships among them are described in the ontology. The terms and their relationships were obtained through the content analysis of the interviews, grounded on a matrix of responses constructed with the information collected from each interviewee. The most repeated terms in the responses were selected. Other sources for the selection of terms were the Collection guides of the DMCs researched, besides information contained on the companies' sites.

A large part of the collection is composed by pieces of communication, materials elaborated by the communication and marketing area of the companies. These items have a nomenclature of the area in which they are produced, unknown by the team of professionals responsible for registering this material. The delay or non-location of items is one of the consequences of the lack of alignment of concepts and adaptation of the nomenclature of the materials.

The terms identified refer to the documental type of the collections, subjects of these documents, target audience and channel of dissemination of this information. The documents can still be considered as individual items or can be part of a collection, in which various items are gathered according to their purpose or theme. However, for delimitation reasons, only the results related to the subjects of the DMC documents will be addressed in this article – Institutional Dimension.

The subjects most mentioned in the interviews and which were grouped into 4 thematic dimensions, are all present in at least 60% of the sample (not considering the processes dimension, the one less stated, all the other three dimensions together are present in 80% of the sample). The four thematic dimensions are: 1) Institutional dimension; 2) Products or services dimension; 3) People dimension; 4) Processes dimension.

The research also reveals that 73% of the sample (8 companies) do not have these same subjects described in a standardized form to execute the indexation of documents, only 27% (3 companies) use at least lists of controlled vocabulary. Thus, it was observed that the sample researched stores documents practically of the same subjects, however, does not index them in a standardized form. Each DMC attributes different nomenclatures to a same subject, according to the professional perspective that is indexing the document, based on the natural language.

The greatest contribution of this study is in this sense: based on the information collected on the DMCs themselves, propose an ontology which covers the terms for documental type, collections, subjects, target audience and dissemination channel, including their subclasses and instances, the ontology of which should later be assessed, validated or extended by the community itself to which it is destined.

Figure 1 represents the general view of the ontology, with the concepts and relations of a DMC object, which is an item of the collection, and also a documental type that

may or may not compose a collection, which can refer to more than a subject, however only has one main subject, is destined to a target audience and enabled through a help desk.

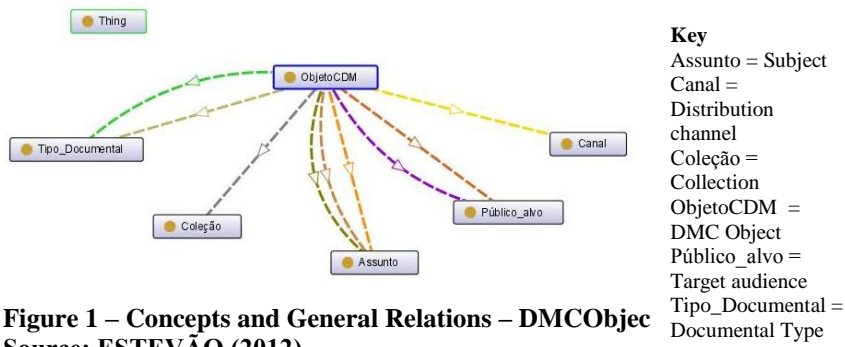
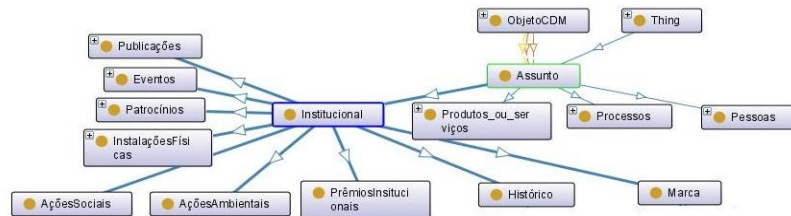


Figure 1 – Concepts and General Relations – DMCObjec
 Source: ESTEVÃO (2012).

Figure 2 represents a part of the ontology proposal, the field subjects – Institutional classification, with its deployment into classes, subclasses. Class refers to a division of the subject, and subclass refers to a subdivision of this greater division, which can even be divided into more sub-subclasses. The instance represents an example inside the subclass or sub-subclass. The terms of this field, as well as all the other terms proposed in the ontology, were constructed based on the content analysis of the interviews carried out and in the analysis of documents, like procedure manuals and collection guides of the DMCs, content of the sites available on the internet and other documents collected during the visits.



Key
 Ações Ambientais = Environmental Actions
 Ações Sociais = Social Actions
 Assunto = Subject
 Eventos = Events
 Histórico = Historical Record
 Instalações Físicas = Physical Installations
 Institucional = Institutional
 Marca = Brand
 Objeto CDM = DMC Object
 Patrocínios = Sponsorships
 Pessoas = People
 Prêmios Institucionais = Institutional Awards
 Processos = Processes
 Produtos ou Serviços = Products or Services
 Publicações = Publications

Figure 2 – Subclass of the Subject Class – Institutional Field
 Source: ESTEVÃO (2012).

6. Final considerations

It was concluded in this research that the HOM initiatives serve, in fact, as disseminating channels of historical organizational knowledge and tools of strategic use by the Institution, not limited to just being guardians of internal registers. Even so, many organizations only register the most important moments in the form of books launched on commemorative dates, like anniversaries of the company, for example. Others structure DMCs so that the preservation of information is not only “celebrative”, but information that can be used strategically if it were stored, retrieved and reused across the organization, however in a non-optimized form.

This information refers to the subjects covered by the informational content of a DMC, which translates the culture, identity, institutional rhetoric and the reputation perceived by a brand, having full potential to be used in favor of the present objectives and future of the organization.

In this sense, the ontology is a structure capable of representing the content of the repositories in the field of historical organizational memory, thus promoting the search and retrieval of information. The concepts are predefined, related to other concepts of the field established and the restrictions declared. Therefore, ontology allows the standardization of the concepts used to index the subjects, i.e. any effective or third party employee who works with the indexation of the items of the collection, will have this conceptual model available allowing that person to understand and make use of the same concepts.

The proposal of an ontology as a model of reference of a field allows, based on the creation of an explicit conceptual model of the base of knowledge of this field, subsidies for the development of search engines and information retrieval based on the context, which guarantees the precision and speed of recovery of information, support in decisions to be made, high recall with precision (large part of the responses used), generating reliability and credibility to the corporate databases. The ability of this search engine operating supported by a conceptual structure, based on ontological descriptions and semantic relationships, will act as an instrument of categorization and classification of historical organizational knowledge, using concepts and their inferences as search criteria, and not only the repetition of words without context, which significantly improves its performance.

References

- ALMEIDA, Mauricio Barcellos. Roteiro para construção de uma ontologia bibliográfica através de ferramenta automatizada. *Perspectivas em Ciências da Informação*, Belo Horizonte, v. 8, n. 2, p.164-179, jul./dez. 2003.
- BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web. *Scientific American*, United States, v. 284, n. 5, p. 34-43, Maio 2001.
- BROOKING, Annie. *Corporate Memory: strategies for Knowledge Management*. London: International Thomas Business Press, 1999.
- CHOO, Chun Wei. *A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões*. São Paulo: SENAC São Paulo, 2003.
- DAMANTE, Nara. Conhece-te a ti mesmo: entender a história como elemento de responsabilidade social conduz empresas e organizações à auto-análise. *Revista Comunicação Empresarial*, São Paulo: Aberje, 2004, v. 14, n. 52, p. 28-35.

ESTEVÃO, Janete Saldanha Bach; STRAUHS, Faimara do Rocio. Proposta de um modelo de referência flexível de ontologia no domínio da Memória Organizacional Histórica. In: CONGRESO LATINO-IBEROAMERICANO DE GESTIÓN TECNOLÓGICA – ALTEC, 14. 2011, Lima - Peru. Anais... XIV Congreso Latino-Iberoamericano de Gestión Tecnológica - ALTEC 2011, 2011.

ESTEVÃO, Janete Saldanha Bach. Proposta de uma ontologia como modelo de referência no domínio da Memória Organizacional Histórica. 2012. 157 f. Dissertação (Mestrado em Tecnologia) - Programa de Pós-graduação em Tecnologia, Universidade Tecnológica Federal do Paraná, Curitiba, 2012.

GUARINO, Nicola. Understanding, Building, and Using Ontologies. *International Journal of Human-Computer Studies*, London, v. 46, n.2-3, p. 293-310, fev./mar. 1997. Duluth: Academic Press, Inc., 1997.

JASPER, Robert; USCHOLD, Mike. A framework for understanding and classifying ontology Applications. 1999

MCGUINNESS, Deborah. Ontological Issues for Knowledge - Enhanced Search. In: FIRST INTERNATIONAL CONFERENCE FORMAL ONTOLOGY IN INFORMATION SYSTEMS - FOIS'98, 1., p. 302-316, 6-8 jun. 1998, Trento, Italy. Proceedings... IOS Press: Amsterdam, 1998.

MUSEU DA PESSOA. Workshop Celebrações: como Utilizar a Memória da sua empresa em grandes eventos comemorativos. São Paulo: Museu da Pessoa, 2005.

NASSAR, Paulo. Entrevista concedida à Camila Amorin. *Questões de Memória. One Voice*. 01 jun. 2009.

NASCIMENTO, Marta Sianes Oliveira do et al. A ontologia na ciência da informação. *Revista Digital de Biblioteconomia e Ciência da Informação, Belo Horizonte*, v.5, n.1, p. 13-39, jul./dez. 2007.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. Criação de conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação. Rio de Janeiro: Campus, 1997.

REMOR, Lourdes de Costa; RADOS, Gregório Varvakis; REMOR, Carlos Augusto Monguilhott. A construção da memória organizacional utilizando o gerenciamento de processo nas pactuações da comissão intergestores bipartite do sistema único de saúde. Enc. Bibli: *Revista Eletrônica Biblioteconomia e Ciência da Informação. Florianópolis*, v. 14, n.27, 2009. 13p.

SOWA, John F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. California: Brooks Cole, Pacific Grove, 2000.

TONINI, Beth; GAGETE, Élide. Como transformar a memória Empresarial em ferramenta de gestão. São Paulo: Aberje, 2004.

WALSH, James P.; UNGSON, Gerardo Rivera. Organizational memory. *The Academy of Management Review*, United States, v. 16, n. 1, p. 57-91, 1991.

WELTY, Christopher A. The Ontological Nature of Subject Taxonomies. In: FIRST INTERNATIONAL CONFERENCE FORMAL ONTOLOGY IN INFORMATION SYSTEMS - FOIS'98, 1., p. 317-324, 6-8 jun. 1998, Trento, Italy. Proceedings...IOS Press: Amsterdam, 1998.

WORLD WIDE WEB CONSORTIUM - W3C. About W3C. 2011. Available at: <<http://www.w3.org/Consortium>>. Access: 1 may 2011.