

Collective construction of Information Science through a Web 2.0 environment

Lizete OLIVEIRA¹

lee7@ufrgs.br

Rafael PORT DA ROCHA

rafael.rocha@ufrgs.br

Maria Manuela GOMES
DE AZEVEDO PINTO²

mmpinto@letras.up.pt

Armando MALHEIRO
DA SILVA

malheiro@letras.up.pt

1 Faculdade de Biblioteconomia e Comunicação (FABICO), Universidade Federal do Rio Grande do Sul (UFRGS), Brazil

2 Faculdade de Letras FLUP, Universidade do Porto / CETAC.Media, Portugal

RÉSUMÉ : Ce document décrit la portée et les objectifs de la "CIFrag projet, qui consiste à la construction collective de la science de l'information et de ses concepts, à travers la combinaison de Web sémantique et Web 2.0 techniques. Cette construction est réalisée par une communauté virtuelle (VC), impliquant des enseignants et des étudiants de la Faculté de Bibliothéconomie et de la Communication (FABICO), Université Fédérale de Rio Grande do Sul (UFRGS), au Brésil, et de la Faculté ès Lettres, Université de Porto (UP), au Portugal. Le méthode quadripolaire est utilisée et appliquée, avec ses pôles – épistémologique, théorique, technique et morphologique Le contenu en rapport avec l'information et des sciences de la communication sont collectivement construit à partir de l'extraction et l'organisation des fragments de textes web. Ces fragments sont organisées par les métadonnées produites par les membres de la communauté et selon une ontologie, caractérisant un environnement collectif à l'intention du Web sémantique. Cet environnement répond à la proposition de Gruber Collective Knowledge Systems.

MOTS-CLÉS : information Science ; Semantic Web ; Web 2.0. ; Collective Knowledge Systems

ABSTRACT : This paper describes the "CIFrag" Project, which consists in the collective construction of Information Science and its associated concepts, through the combination of Semantic Web and Web 2.0 techniques. This construction is achieved by a virtual community (VC) involving lecturers and students from the Faculty of Librarianship and Communication (FABICO), Federal University of Rio Grande do Sul (UFRGS), Brazil, and the Faculty of Arts, University of Porto (UP), Portugal, and Quadripolar Method is used and applied, with its epistemological assumptions and methodological procedures. The contents related with Information and Communication Sciences are collectively built from the extraction and organization of fragments of web texts. These fragments are organized through metadata produced by the community members and according to an ontology, characterizing a collective environment directed at the Semantic Web.

KEYWORDS : Information Science ; Semantic Web ; Web 2.0. ; Collective Knowledge Systems

1. Introduction

The "CIFrag" Project consists in the collective construction of Information Science and its associated concepts, achieved by a virtual community (VC) involving lecturers and students from the Faculty of Librarianship and Communication (FABICO), Federal University of Rio Grande do Sul (UFRGS), Brazil, and the Faculty of Arts, University of Porto (UP), Portugal. The VC is implemented

based on the environment provided by the Fabico/UFRGS experience, which is a VC for an undergraduate training environment, involving Semantic Web and Web 2.0 techniques, and, with regard specifically to the Information Science dimension of the project, the Quadripolar Method is used and applied, with its epistemological assumptions and methodological procedures.

The contents related with Information and Communication Sciences are collectively built from the extraction and organization of fragments of web texts. These fragments are organized through metadata produced by the community members and according to an ontology, characterizing a collective environment directed at the Semantic Web. The project is generating rich scientific experiences related with: the formation of VCs ; experimenting with the Semantic Web in an environment of collective construction which focuses on Web 2.0 ; the combination of techniques from the Semantic Web and Web 2.0 ; the elaboration of a thesaurus for the Semantic Web, resulting from the generalization of existing standards of metadata and ontologies ; the usage of folksonomies ; the usage of collective writing (wiki) ; and manifestations of the descriptive and thematic representations of Information Science, in Web 2.0 and Semantic Web. This environment meets with Gruber's proposal for Collective Knowledge Systems.

2. The paradigmatic transition and the Information Science Quadripolar Method

Our epistemological proposal for Information Science presents some important differences in relation to the dominant perspective and its variants, particularly the approach which insists on, and is channelled by IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia), the distinction between the original paradigm and the evolution of the Librarianship, Documentation and Information Science disciplines. In counterpoint to this supposed multidisciplinary, we propose the fundamentals of a still emerging Information Science (IS), which we define as *being a social science which investigates the problems, themes and cases related to the info-communicational phenomenon, understandable and cognoscible through the confirmation, or not, of the properties inherent to the genesis of informational flows, organization and behaviour* (Silva, 2006: 140-141). IS will be consolidated when, through more widespread debate and consensus, the intersection or fusion of professionalizing and practical disciplines such as Archivistics, Librarianship, Documentation and Information Science (focused on the processing, storage and retrieval of automatized or digital information), are brought together in a transdisciplinary matrix, leading to the genesis of a new scientific discipline, as a result of new internal and external conditions, which can be placed in the historical process. This qualitative jump, designated as transdisciplinary, results from an ongoing shift from a still dominant paradigm, consolidated throughout the 19th century, which we call custodial, patrimonialist, historicist and technicist (Silva, 2006: 158), to an emergent, post-custodial, informational and scientific paradigm (Silva, 2006: 158-159).

This new paradigm is intrinsically related to the Information Era which developed after World War II, one of whose master stalwarts is the revolutionary expansion of Information and Communication Technologies and the construction and growing hegemony of the "internet galaxy" (Castells, 2004). This emerging paradigm is essentially characterized by : 1) a change with regard to the object of study, which is no longer the document, but is rather more focused on concepts such as info-communicational phenomenon, information (confronted with other correlatives : communication, document, culture and knowledge), information properties and method ; 2) a change in terms of approach, which has to be scientific, rooting this scientificity in the epistemology of the Social Sciences and the application of a Method which has already been tested in qualitative research

projects, such as the Quadripolar Method, and not a simple, equivocal and vague formal exercise, based on standards and empirical considerations saturated with common sense ; 3) and, finally, a change with regard to the professional matrix supporting this area, in the sense that a radical distinction is established between professional practices (the functional activity of the librarian, archivist and documentalist) and the research agenda in Information Science, where the former should not commandeer the latter, but research (done and redone) should be converted into a factor of renewal and improvement of the professions related with the info-communicational process.

Within the proposal we have been developing in an academic and research environment, the operational concept of information is crucial and has direct implications on all the theoretical-practical work which is performed in IS. The following definition is used as a boundary for our field of study and work : information is a scientific object which entails the structured set of codified mental and emotional representations (signs and symbols) which are modelled with/through social interaction, able to be recorded on any information storage material (paper, film, magnetic tape, compact disk, etc.) and, thus, communicated in an asynchronous and multidirectional manner (Silva, 2006 : 150).

As for the Quadripolar Method, it should be noted that it emerged as an alternative answer to positivism and to the reductionist dichotomy between “quantitative” and “qualitative”. This method was proposed in 1974 by P. De Bruyne and other authors as the operational instrument of a research dynamic brought on by a new paradigm in the Social and Human Sciences, with particular emphasis for the interdisciplinary field of Education Sciences. In 2002, it was adopted and suggested as a global and methodological structure for IS (Silva; Ribeiro, 2002 : 86-91), and has since been applied and practiced in both the IS Undergraduate Course of the University of Porto, and in Masters dissertations, Doctoral theses and research projects in Portugal and in Brazil. The quadripolar research dynamic results from an interaction between the epistemological, theoretical, technical and morphological poles. At the epistemological pole, the subject-researcher debates the paradigmatic shift and is clearly positioned in the face of existing paradigms. At the theoretical pole, the problem to be studied is set, the hypothesis is formulated, there is a search for and the creation of a theory, and the model of intervention is built. At the technical pole, crucial operations, such as the observation of cases and variables, analysis/evaluation, infometry and even mitigated experimentation, including also instrumental techniques such as the questionnaire, the interview, the application of tests and protocols, etc., are applied. At the morphological pole, the results of the research conducted and operationalized by the conjugation of the preceding poles are formalized (Silva, 2006 : 154-155).

3. Collectively Building Information Science

The environment created in the Fabrico/UFRGS experiment is made operational in a Web 2.0 environment, in which lecturers/researchers and students of both institutions learn to collectively develop Information Science from fragments of texts published in several formats. These fragments are organized through metadata produced by the community members according to ontologies, characterizing an environment directed at the semantic web. The research is developed according to the Quadripolar Method.

3.1. Technical Pole

At the technical pole we have the environment itself, which comprises the following tools : a text

- **Characterizes** : indicates that the fragment presents characteristics of the term informed as the value of this property.
- **Categorizes** : indicates that the fragment presents categories of the term informed with value of this property.
- **Composes** : indicates that the fragment presents components of the terms informed with value of this property.

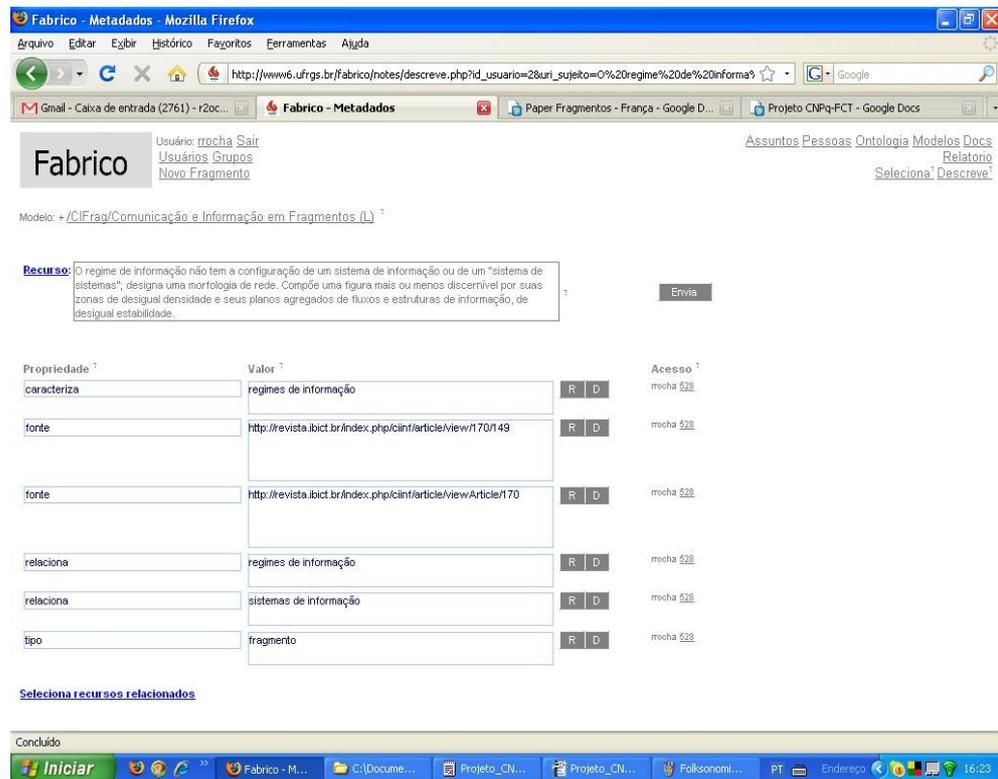


Figure 2 - Extractor and fragments descriptor

For example, Figure 2 presents the descriptions of a fragment (indicated in the property **recurso/resource**) extracted from the document whose link is indicated in the property **fonte/source**. These descriptions indicate that the text fragment characterizes **regimes de informação/information regimes** (property **characterizes**) and establishes a connection between **regimes de informação/information regimes** and **sistemas de informação/information systems** (property **relates**).

Figure 3 presents the fragments related to the term **regimes de informação/information regimes**. These fragments are organized according to the properties whose values were **regimes de informação/information regimes**. In this way, an account of this term is obtained, done through fragments that indicate its definitions, statements, characteristics, compositions, categories and relations.

The screenshot shows the Fabrico web application interface. At the top left, there is a navigation menu with 'Fabrico' and user options: 'Usuário: trocha Sair', 'Usuários Grupos', and 'Novo Fragmento'. On the top right, there are navigation links: 'Assuntos Pessoas Ontologia Modelos Docs', 'Relatório', and 'Seleciona? Descreve?'. Below these is a link for 'Imprimir Detalhes'. The main content area is titled 'regimes de informação' and is divided into several sections: 'Definições', 'Características', 'Categorias', and 'Relacionamentos'. Each section contains text and source references. The 'Relacionamentos' section is a table with two rows, each containing a definition and a link to 'regimes de informação'.

Figure 3 - Fragments related to the term regimes de informação/information regimes

The project is an example of an environment that explores the techniques presented by Gruber (2007) to enable Collective Knowledge Systems. It focuses on the increment and dissemination of information/knowledge related with an Undergraduate Course through a VC directed at the Semantic Web, in which the former is collaboratively constructed, recorded, and formally described and disseminated.

The environment is called Fabrico (www.ufrgs.br/fabrico) and it comprises a collaborative writing tool (wiki), a system of folksonomies, a documents repository, a web texts fragment extractor and an annotator based on a thesaurus. Through these tools, members of the community (lecturers and students) collectively produce wiki texts, store the produced documents, make web page links (bookmarks) available to the community, describe contents (wikis, documents, text fragments extracted from the web pages and link) according to the thesaurus and define descriptors (keywords or labels) for these contents (*folksonomy*).

In the project, VC members use the annotator to describe the meaning of the contents in a similar way to that used in Semantic Web projects, such as Annotea (Kahan, J. et al, 2001), Cyclades (Grass, 2003), Ontoshare (Davies, Duke and Sure, 2003) and Sek/Bt (Thurlow, I.; Duke, A.; Davies, J, 2006), differing from these in the type of community involved (academic) and in the ontology which had to be, as basic requirements, simple, extensible and include a thesaurus and existing metadata standards (eg.: MODS, EAD, Dublin Core, Foaf, Proton), having in mind reuse and interoperability. Web 2.0 requirements also had to be incorporated through the wiki and the folksonomies, characterizing the combination between the Semantic Web and Web 2.0, in which wiki texts are semantically described and the values of the properties of the thesaurus are organized by folksonomies.

3.2 Epistemological, Theoretical and Morphological Poles

At the epistemological and theoretical poles, the clearly assumed paradigm is the Post-custodial, already defined by lecturers from the Faculty of Arts of the University of Porto and discussed in several publications. Theoretically, the Constellation Metaphor is used as a reference, an image created by Walter Benjamin (1984), which considers each idea about a certain theme as being the light of a star. According to this author, “the ideas relate to things such as constellations to stars”. It should be noted that the term constellation (cons-stellation) stems from the idea of *stella*. Throughout time, several narratives have been created about star groupings, resulting from observations (*sidera*), i.e., from different readings of stars. In the same way that in our metaphor the light from several stars illuminate the others and create, overall, a new image, resulting from the several lights created by these ideas, in the case of the research proposed, each textual quotation, which we call fragment, highlights different configurations for IS. It is up to the reader of the several “observed” texts, as is the case of observers of stars, to pick “which of the elements stands out and which connections could be made between those points. If we return to the point that constellations are not natural formations, but rather “cultural images”, which differ according to different eras, that were projected on an arrangement of stars located relatively close together, the reading of the constellate text is characterized by the freedom to establish connections between dispersed parts.” (Volpe).

The morphological pole foresees the publication of results from the research conducted in the web environment and the writing of articles that will be published in scientific journals. At the same time, the knowledge generated by the research will be spread to the academic community through subjects taught in courses of both institutions. Results are also produced in a collective and continued way through wiki texts, which never reach a final version (“the perpetual beta”).

4. Current Results

The project is in process of formation of the VC, in which initially a small group of specialists is producing content for the purpose of giving to the environment a "critical mass" to motivate the formation of a larger group that is capable of evolving in a natural way. Based on this initial experiment, we observed that the use of a Web 2.0 environment is conducting to the construction of a scientific object which is focused on the union of various and conflicting scientific discourses, where the search of the limits of the problem is occurring on a collective and inclusive way. In this case, the paradigms and its contradictions and changes are collectively identified by the extraction and description of fragments, and the analysis of those descriptions and the terms given in these descriptions via cloud of terms and related terms.

4. Final Considerations

The community built through this project and the contents produced constitute tools that enable mutual and continuous growth involving both institutions. Through them, these institutions act collaboratively and in a continuous way, not only in the research and teaching fields, but also in the improvement of their educational offer. They also have a mutual interest in the development of techniques and knowledge related with the collective production of the mentioned contents, with a view to strengthening the ties between Information Science, Web 2.0 (participative architecture and collective construction), semantic web (ontologies) and ICT.

Web 2.0 has raised new challenges to ICT development. Data, collectively built, constitutes the

core of the environments that use Web 2.0. In this context, the innovation differential of Web 2.0 enterprises is to know how to take advantage of the aggregate value resulting from users contributions, i.e., from the collectively built data through a participative architecture. This matter challenges and brings Computation Sciences and Information and Communication Sciences (ICS) closer together, because both areas have accumulated a wealth of valuable experience in research related with information production and organization, whether in selective and directed communication, or in the treatment of user communities. Bearing this in mind, research on ICT, from the perspective of Web 2.0 and its implications on the ICS field is, in fact, a scientific need.

Bibliographic references

- BENJAMIN, Walter, *Origem do drama barroco alemão*, Trad. Sérgio Rouanet. São Paulo: Brasiliense, 1984
- CASTELLS, Manuel, *A Galáxia internet: reflexões sobre internet, negócios e sociedade*, Lisboa : Serviço de Educação e Bolsas-Fundação Calouste Gulbenkian, 2004. ISBN 972-31-1065-2
- GRUBER, T. Collective Knowledge Systems : Where the Social Web meets the Semantic Web. *Web Semantics : Science, Services and Agents on the World Wide Web*. Volume 6, Issue 1, 2007, Available at : <http://tomgruber.org/writing/CollectiveKnowledgeSystems.htm> [Accessed 15 Dec 2002]
- KAHAN, J. et al., Annotea: an open RDF infrastructure for shared web annotations. In: INTERNATIONAL CONFERENCE ON WIDE WORLD WEB, 10, 2001, Hong Kong. Proceedings..., [S.l.:s.n.], 2001, Available at: <http://www10.org/cdrom/papers/pdf/p488.pdf>. [Accessed 10 May 2008]
- GRASS, T. CYCLADES : A distributed system for virtual community support based on open archives. In. EUROMICRO CONFERENCE ON PARALLEL, DISTRIBUTED AND NETWORK-BASED PROCESSING, 11. Proceedings..., [S.l.]:IEEE, 2003, pp. 5-30. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1183629 [Accessed 10 May 2008]
- DAVIES, J., DUKE, A., SURE, Y. OntoShare – A knowledge management environment for virtual communities of practice. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE CAPTURE, 2, 2003, Sanibel Island, FL. Proceedings... New York:ACM. 2003. pp.20-27. Available at: <http://doi.acm.org/10.1145/945645.945652>. [Accessed 10 May 2008]
- THURLOW, I.; DUKE, A.; DAVIES, J., Applying semantic web technology in a digital library. In: EUROPEAN SEMANTIC WEB CONFERENCE, 3, 2006, Budva, Montenegro. Proceedings... [S.l.:s.n.], 2006. Available at: <http://www.eswc2006.org/poster-papers/FP44-Thurlow.pdf>. [Accessed 10 May 2008]
- SILVA, Armando Malheiro da, *A Informação: da compreensão do fenómeno e construção do objecto científico*. Porto: Edições Afrontamento ; CETAC.COM, 2006. 176 p. (Comunicação, Arte, Informação; 1). ISBN 978-972-36-0859-5
- SILVA, Armando Malheiro da, RIBEIRO, Fernanda, *Das "Ciências" Documentais à Ciência da Informação : ensaio epistemológico para um novo modelo curricular*. Porto : Edições Afrontamento, 2002. 174 p. (Biblioteca das Ciências do Homem. Plural; 4). ISBN 972-36-0622-4
- VOLPE, Miriam, Um olhar constelar sobre o pensamento de Walter Benjamin. Available at : <http://www.mestradoletras.ufjf.br/profs/textos/textomi2.pdf> [Accessed 26 Nov 2008]