Abstract

It is a challenge to organize the contents of a news portal, in which the most evident aspect is the encyclopedic character of the information produced and aired daily. Usually, the users of such services increasingly use the portal as a source of daily information, replacing printed media. In this context, news production includes not only its preparation itself, but also the creation of significant links between texts, images and sounds. Therefore, in this environment, the journalist/editor simultaneously assumes the roles as portal’s producer, organizer and user. The process of creating news in an electronic portal leads to dispersion of related issues and to improper relations between themes for different reasons: intense work pace, significant number of professionals involved, lack of policies and shared devices to treat and retrieve information, among others. These issues cause breaks in. The relation between adequate and consistent subjects is, thus, an important factor to retain users. This is the problem that motivated the development of this research. Theoretical and methodological aspects of the research are presented here, as well as a set of parameters of creation and maintenance of controlled vocabularies. It is also presented a manual of good practices of indexing pages in the portal. The study was done based on: Domain Analysis and Guidelines for the construction of languages knowledge Organization.

Theoretical and Methodological References

Domain analysis was proposed in the 1990s to study domains of knowledge . (Hjørland; Albrechtsen,1995 p.400). From this perspective, we should not treat the knowledge fields as if they were all fundamentally similar. In Domain analysis, several applications are proposed (Hjørland, 2002):

1. Producing literature guides or subject access;
2. Construction of special classifications and thesauri;
3. Specificity and specialty of indexing and information retrieval;
4. Empirical users studies involving investigation of the information needs of different communities.
5. Bibliometric studies, to analyze links between documents.
6. Historical studies, considered as fundamental methods to provide deeper perspectives and relevant context of the studied domain.
7. Documents and genres studies, to support studies of Information Architecture.
8. Epistemological and critical studies, to substantiate critical analysis of the knowledge domains.
9. **Terminological studies, special purposes languages (SPL), database semantic and speeches studies.**

10. **Structures and institutions in scientific communication** to support the characteristics, to identify the actors and map the information flows in this domain. The structures studies of the internal divisions inside the working areas and all the information exchange between domains provides useful information for understanding functions of specific types of documents and information services.

11. **Scientific cognition expertise and knowledge about artificial intelligence.** This approach can provide useful techniques to supplement other approaches to domain analysis in information science. As noted by the author, Domain Analysis has long been used in computer science, but for other purposes. Even so, relies on useful techniques for our area, which should be adapted.

In this research information and language issues are considered as part of a cultural system that influences information organization and retrieval (Hjørland and Albrechtsen, 1995, p. 407). In this sense, taxonomy and folksonomy or social information classification concepts were discussed. Quintarelli (2005) defines folksonomy as a new approach to classification of distributed digital resources. Likewise, it is seen as an approach to classification (Hammond, 2005). For these authors, it is an unstructured classification made by digital resources users. Can the publishing systems for news portals be considered social tools? While its use is collective, is there interaction? The only feature that could bring the language of the portal system close to folksonomy is to use natural language to create and assign descriptors to each tag which, according to Lancaster (1993), is the language commonly used in writing and speech, and that is the opposite of 'controlled vocabulary'. For semantic and cognitive aspects of classification, Golder and Huberman (2006) say that labeling systems need to attack problems that are inherent in the process of creating semantic relationships between concepts. To create the controlled vocabulary for the portal, the three main problems of language - polysemy, synonymy and basic level variation - were considered.

**Results and Provisional Conclusions**

To solve the mentioned problems we considered the need of a moderating (mediating) process for creating terms, which was based on methodological theories on information and knowledge organization and guidelines such as ANSI / NISOZ39.19 (2005) of creation and maintenance of controlled vocabularies. The following conclusions resulted: a) information organization and retrieval from portals requires a combination of automatic processing and indexing tools, given the quantity and diversity of news published each day; b) it is necessary that the tool has clear syntax, easy understanding and fast processing, to any Java-based application. The tool should allow the management of texts and images publication (photos and infographics) and the integration of text and related images, in the form of an album and finally, a set of good indexing and updating of the system vocabulary.

**References**


