The use of an indexing language in the catalogues of university libraries: a method for its evaluation by users based on a sociocognitive approach

Vera Regina Casari Boccato

Department of Information Science, Universidade Federal de São Carlos – UFSCar, Brazil

Mariângela Spotti Lopes Fujita

Department of Information Science, Universidade Estadual Paulista – UNESP, Brazil

Abstract

The use of an indexing language in the catalogues of university libraries was evaluated from the perspective of the users, using a sociocognitive approach. The methodology consisted of a diagnostic study of the processing and use of indexes of nine libraries of the UNESP Network, specifically those covering civil engineering, education and dentistry, using the Introspective Verbal Protocol technique in both Individual and Group forms. The purpose of the study was to examine the process of bibliographic search as well as views concerning the performance of the indexing languages in dealing with information retrieval. The individuals participating in the Group Verbal Protocols consisted of heads of libraries, cataloguers, reference librarians, faculty staff (teachers) and students. Those participating in the Individual Verbal Protocols consisted of students in their first and last years of the relevant courses. The study reported on statements issued by the 63 participating individuals who obtained unsatisfactory results using the Subject Headings List of the BIBLIODATA Network, this being the indexing language used by the Brazilian UNESP Libraries Network in retrieving information from the ATHENA Catalogue. Problems were reported with the following aspects of the language: a lack of specialized vocabulary and updating, inconsistencies in the syntactic-semantic structure of headings, and ambiguous headings. The evaluation shows that to enable effective information retrieval in a specialized scientific area, an indexing language needs updating, appropriate specificity and consistency.

Introduction

Systems of knowledge organization are identified by indexing languages – lists of subject headings and thesauri. These are controlled languages created according to the guidelines of international standards, whose vocabularies are based on meanings stemming from terms pertaining to both

specialized and natural languages (commonly spoken languages). Indexing languages are therefore the medium through which the user accesses an information system to retrieve information.

However, when an indexing language is inconsistent with the user's retrieval language, the quality of the search and the credibility of the information system are uncertain, compromising the researcher's ability to extract the indexed information he needs.

Thus, the purpose of this study is to evaluate the use of indexing languages of online catalogues, focusing on the information retrieval technology used by university libraries and on the user's perspective, based on a sociocognitive approach.

The Verbal Protocol Technique employed in the evaluation of an indexing language, based on the sociocognitive approach

The Introspective Verbal Protocol or 'think-aloud' technique is employed in qualitative studies in which the subject expresses his thoughts verbally while performing a task. These statements are recorded, and the subject's behaviour, e.g. facial expressions, are also recorded (gestures and eye movements). Thus, the language of thought performs many cognitive processes, such as perception and reasoning.

From a sociocognitive standpoint, the verbal protocol technique focuses on the subject performing a particular activity and his perception of the execution of the task. From the user's perspective, the focus is on information retrieval by subject, using language to search for information and to interact with the search environment. According to Hjørland (2002), the sociocognitive context involves the retrieval of information from the perspective of the individual in his sociocultural and historical context, and is based on the assumption that the user's information needs are constructed from the outside to the inside. In other words, the user's external environment influences and contributes to the development of his internal need, which is represented by the mental models he associates with his conception of the world and reflects his real interest in the information he seeks.

Methodology

Data were collected from a sample of nine libraries of the UNESP network (Brazil) and involved three areas of knowledge: civil engineering, education, and dentistry. The objective of the empirical study was to examine the Subject Headings List of the BIBLIODATA Network (LCARB) used in the process of retrieving information from the online catalogue ATHENA.

According to Boccato and Fujita (2011), the methodology consisted of a two-part diagnostic study: 1) evaluation of the operation and procedures for processing indexed information in the libraries of the

UNESP Network, based on the perspective of the General Coordinator of Libraries (CGB – Coordenadoria Geral de Bibliotecas); and 2) the user's evaluation of the access to, and retrieval of, online information.

In the first part, the heads of university libraries were asked to fill out an organizational diagnostic questionnaire. In the second part, the cognitive approach was employed to evaluate the user's opinion about the operation and procedures for retrieving information from the libraries of the UNESP Network, based on the Introspective Verbal Protocol technique, as follows:

- Individual Verbal Protocol (IVP): involved first and last year undergraduate students of civil engineering, education and dentistry who access and retrieve online information, aiming to identify the difficulties and limitations of using the LCARB.
- Group Verbal Protocol (GVP): involved heads of libraries, cataloguers, reference librarians, faculty (professors), head researchers, members of research groups, and students to determine their knowledge about indexing and retrieval of information from university libraries as a source of qualified data for the diagnosis.

The different verbal protocols were applied in similar procedures, as follows:

1) Activities preceding the application of the Verbal Protocol

   a) Definition of the universe of the research:

      GVP and IVP: university libraries of the UNESP Network;

   b) Selection of copy-text:

      GVP: passage between pages 205 and 208 of the following paper: Dias, E. W., Naves, M.M.L. & Moura, M.A. (2001). O usuário-pesquisador e a análise de assunto. *Perspectivas em Ciência da Informação, 6,* pp. 205-21. Available at: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/429/237.

      IVP: the bibliographic record itself retrieved from the ATHENA catalogue of the UNESP Library Network.

   c) Definition of the task:

      GVP: discussion of the copy-text previously referenced

      IVP: the users' retrieval of information on dentistry, civil engineering and education-related topics from the ATHENA catalogue, by subject field, using the LCARB indexing language.

   d) Selection of the Individuals:

      GVP: three librarians (chief, cataloguer and reference), one professor and one student.

IVP: first and last year undergraduate students of dentistry, civil engineering and education who use nine libraries of the UNESP Network, making a total of eighteen participants;

e) Informal conversations with individuals:

In these conversations, the researchers contacted individuals through the Office of the General Coordinator of Libraries to explain the objectives and methodology of the research, and to set a date for collecting data. The anonymity of all the participants was ensured.

2) <u>Procedures performed during the application of the Verbal Protocol</u>

GVP: after the participants read the copy-text, a discussion was held during which the researcher acted as a mediator to prompt the participants. The entire discussion was recorded and transcribed verbatim.

IVP: this involved recording every thought expressed aloud by the user during his bibliographic search, using an MP3 recorder.

3) <u>Procedures following the application of the Verbal Protocol</u>

a) Transcription of the recordings:

GVP: after recording the individuals' discussion of the text, a literal transcription was made, identifying the sources of individual statements. The sources were identified as heads of libraries, cataloguers, reference librarians, faculty (professors), and students.

IVP: after recording the thoughts expressed aloud during information retrieval, the recordings of the individual statements were transcribed literally.

b) Analysis of the transcripts: the transcripts were scrutinized to identify significant and recurring phenomena, which would serve as the basis for the construction of categories. Then, having constructed the categories, the transcripts were examined once more to extract passages that exemplified each category.

Forty-five samples of data were collected from nine groups of verbal protocols (involving heads of libraries, cataloguers, reference librarians, faculty (professors) and students, as well as eighteen individual verbal protocols with students). A total of sixty-three individuals participated in this research. Data sampling started in May 2006 and was concluded in September 2007, and was conducted in the individuals' study environments, i.e. at the nine libraries of the UNESP Network that were selected for participation in this study.

To analyse the collected data and to meet the proposals of this study, five categories were determined based on theoretical frameworks, research objectives and statements of the participants who contributed to exemplify each phenomenon. The categories were systematized as follows:

1) Knowledge and Importance of language;
2) Search strategy;
3) Performance of language in the information retrieval process;

4) Accuracy and retrievability of the system;

5) Evaluation of the information retrieval system.

Example: <u>Category 3 – Performance of language in the information retrieval process</u>

IVP – Dentistry – first year undergraduates *[incompatibility between the user's language and that of the system]*

'[...] I'm going to try ((SS)) INCISOR TEETH. Look, it's got this paper by ((RA)) Celso Eduardo de Moraes Barbosa, in 1973, under the subject topic ((RS)) INCISOR TEETH). The way you type it is different from what they have on the computer [...].((SS)) CENTRAL INCISORS. Nothing, again. It only shows ((SS)) INCISOR TEETH. Nobody's going to search for that. Nobody talks about incisor teeth. It is "incisors". The computer subjects should be more like what people are looking for [...].'

Results

The indexing language was widely discussed by users of the three knowledge areas in group and individual verbal protocols. These discussions gave rise to many questions and concerns about the performance and importance of this instrument of mediation and communication for information retrieval from online catalogues of specialized scientific areas in university libraries.

The users' sociocognitive context, shaped by prior knowledge about the university, by research work developed at scientific research level, by the curriculum of scientific-academic disciplines, and by the ATHENA catalogue, revealed relevant results about the use of LCARB for information retrieval.

The performance of LCARB as an instrument of mediation between user and system was considered poor, demonstrating the incompatibility of the system's indexing language with the user's search.

The system's overall results were one of the most important issues for users, who stated that the search by topic is incompatible with their investigative needs, leading to a high recall rate in information retrieval. The lack of knowledge of the language used by the system and about how to employ this language in search strategies also contributed to the inadequate performance of the indexing language in the data retrieval process. To solve these problems, users suggest the implementation of more specific headings to represent the indexed contents and the construction of a controlled vocabulary based on the language used by specialists and students.

Users also pointed out the need to align the LCARB's indexing language with their own search language. To this end, they offered suggestions about accessibility to the search interface of See and Also See links as guiding links between the searched headings and retrievals by the language. In addition, they emphasized the need for a user manual and accessible language to serve as guidelines and meet the information retrieval needs of local and remote users.

From the user's perspective, the ATHENA catalogue should work analogously to a database, allowing for retrieval by subject of analytical documents such as book chapters and journal articles, and the ordering of retrieved records in descending order, by date of publication, as well as by other aspects.

Conclusions

The sociocognitive qualitative approach associated with the verbal protocol used as a method of evaluation enabled us to analyse the use of indexing languages in online catalogues of specialized scientific areas. With regard to information retrieval by the user, the method allowed us to evaluate the use of language in developing search strategies and the user's perception of the performance of language in his search by subject. It also shed light on what the user considers to be satisfactory in terms of performance and indexing language.

University libraries contain numerous knowledge areas covering an extremely wide range of subjects, thus requiring adequate products and tools to represent their content. Because the indexing language is a component of library catalogues, it should be representative of this highly specialized scientific content, providing optimal mediation and communication between the information retrieval system and its local and remote user communities. This language should be based on a multidimensional concept that takes into account the syntactic-semantic relationships stemming from specialized scientific areas and from the user's language.

Note: Some of the content of this article was taken from a research paper by Boccato & Fujita (2011).

References

Boccato, V.R.C. (2009). *Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo sociocognitivo com protocolo verbal*. Thesis, (PhD). Universidade Estadual Paulista, Brazil. Retrieved from http://www.athena.biblioteca.unesp.br/exlibris/bd/bma/33004110043P4/2009/boccato_ vrc_dr_mar.pdf.

Boccato, V.R.C. & Fujita, M.S.L. (2011). Evaluation of indexed language used in collective catalogues of university libraries: a socio-cognitive study using verbal protocol. *International Journal of Library and Information Science.* [In press].

Boccato, V.R.C. & Fujita, M.S.L. (2006). Evaluación del vocabulario controlado DeCS en logopedia desde la perspectiva del usuário: un estudio de observación de la recuperación de información con protocolo verbal. *Scire (Zaragoza), 12*, pp. 179-95.

Hjørland, B. (2002). Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology, 53*, 257- 70.

Corresponding author:
Vera Regina Casari Boccato can be contacted at vboccato@yahoo.com.br