The semantic retrieval of information in the context of external control

Márcia Martins de Araújo Altounian

is an employee of the Federal Court of Accounts – Brazil. She has a B.A. in Library Science and Documentation from the University of Brasília (UnB) and a specialization degree in Knowledge Management from the Blaise Pascal Institute and in Information Architecture and Organization from UFMG.

Beatriz Pinheiro de Melo Gomes

is an employee of the Federal Court of Accounts – Brazil. She has B.A. in in Library Science and Documentation from the University of Brasília (UnB) and a specialization degree in Strategic Knowledge Management and Business Intelligence from the Catholic University in the State of Paraná (PUC).

SUMMARY

In the information retrieval systems (IRS), syntactic base techniques are being supplanted by the increasing exploitation of semantic retrieval techniques, which enable the understanding of the concepts in their context and purpose. Some technologies have contributed to this reality, such as the data semantic markup, used in the semantic web, natural language processing and neural networks. The thesaurus is also a semantic component that affects IRS performance. Thesauruses are artificial language tools in a specific domain, formed by a system of related concepts. This article presents the application of the TCU thesaurus, called External Control Vocabulary, in some corporate information systems.

Keywords: Semantics. Information retrieval. Thesaurus.

1.   INTRODUCTION

Information retrieval (IR) is a field of common interest in computer science and information science that is concerned with developing and studying aspects of efficiency and effectiveness of searches in an information system, so that results are coherent with its search expression and, above all, relevant to the user of the system.

Searching in large repositories of unstructured and non-standardized information becomes an arduous task, of-

ten leading to ambiguities, out-of-context content, and even failure to retrieve desired information. The search tools, when using the natural language, need knowledge about the meaning of the expressions used and the relations between them, which allows contextualization and treatment of linguistic phenomena that affect the quality of the retrieval.

For the process of classification and retrieval of information, several automatic methods have been developed in order to obtain relevant answers for any research carried out. To date, most of them are based on syntactic and statistical aspects, taking into account frequency and distribution of words present in documents. Based on the syntactic aspects of the information, the classification performed by these methods is still very distant, in terms of quality, from the classification based on indexation of the subjects of the documents carried out by specialists.

Recently, new methods of automatic document classification have been explored based on semantic aspects of language that refer to the meaning of the terms in their context and purpose. One of these initiatives consists in the use of instruments of representation of semantic and conceptual relationships such as thesauruses, which are characterized by the domain of knowledge and aim to solve problems of ambiguity inherent in everyday words.

The External Control Vocabulary, TCU thesaurus, was developed with the objective of standardizing the treatment of specialized contents and contributing to semantic recovery in the context of external control, in addition to making searches in the Court's information systems more agile and precise.

2. SEMANTICS

The term "semantics", traditionally studied in linguistics and philosophy, has been widely used in a wide range of fields, especially in information technology. The so-called semantic web (SW), which originates in the expansion of the web and in the limitations of the search tools based on syntax, is certainly one of the reasons for this association.

But what is semantics? The word originates in the Greek term semantiké and many meanings can be found according to the perspective used, since there are, among others, textual, cognitive, lexical, formal and argumentative semantics. In general, all converge on the same point: they study meaning, or significance. The wide range of possibilities testifies that the study of meaning can be done from various angles.

Three basic properties stand out in semantics. They are synonymy, antonyms, and polysemy. Synonymy is the division of semantics that studies the relation between linguistic expressions that have the same meaning. For example, "girl" and "lassie" are nouns that have the same meaning, referring to the figure of a young woman; so it is with verbs like "renounce," "refuse," and "reject," which convey the idea of revulsion. Except for the rare occurrence of perfect synonymy, one can be affirm that synonymy is the relation between words and expressions that have a common meaning and significance.

If on the one hand synonymy studies the words with similar meanings in the language, antonyms deal with the study of words that indicate opposite meanings. In the same line as the previous example, we can cite the nouns "girl" and "lady", as well as the verbs "renounce" and "accept" - all are terms with opposite meanings.

It is also possible for the same word to assume different meanings - in which case, the context in which it is inserted will dictate its meaning. A good example of daily life is the word "orange", from which we can quickly think of two very different senses: a citrus fruit and a reddish-yellow color. This is a common example of polysemy, a term which, formed by the prefix "poly" ("many") and by the suffix "semy" ("meanings"), is the part of semantics that studies the significations that a word assumes in a particular linguistic context.

There are still two other semantic properties, in the same line of study of signification, that deserve to be mentioned: homonymy and paronymy. The homonymy studies the relation of two or more words that, although they have different meanings, have the same form and the same sound - it is the case of terms like ad/add and to/too (preposition/adverb), among many others. Paronymy studies the particularities of words that, on the contrary, have similar spelling and pronunciation, but different meanings - are examples: eminent/imminent and elicit/illicit.

Finally, semantics also studies the properties of denotation and connotation of words. Denotation is the property that a word has to be limited to its own concept - for example, the term "stars" in "the stars of the sky". Connotation is the property that a word has to be expanded in its semantic field, within a context, generating several interpretative possibilities - for example, the same term "stars" in "movie stars".

When we transpose the concepts explored here to what we now know as SW, it is easy to see that we are talking about promoting improvements in the processes of representation and retrieval of information on the web. Since 1990, the web is characterized by the use of markup languages that aim at the presentation and reading by people and search engines based on algorithms with syntax orientation. The use of semantics can increase the possibility of document associations to their meanings by means of descriptive metadata. Therefore, the question of meaning in the semantics is fundamental to SW.

Of the several lines of study related to semantics, the one of formal semantics seems to be the most relevant to information technology. Three main aspects permeate the studies on formal semantics:

1. The principle of compositionality, which states that the meaning of sentences depends on the meaning of the words that compose them - that is, the meaning of the whole is a function of the meaning of the parts and the syntactic combination between them. To know the meaning of a sentence, it is necessary to know the meaning of its parts, as well as the rules that define its combination.

2. The condition of truth, which determines the conditions under which such a sentence is true. In this

context, knowing the meaning of a sentence is equivalent to knowing its truth conditions, which is not the same as knowing its truth value, that is, whether the fact is true or false.

3. The models in semantics, in which simple systems are constructed, in relation to the complex systems that are to be studied. A logical theory is constructed for the model and, if the results are reasonably positive with respect to the complex system, it is said that the simple system is a good model; otherwise it is abandoned.

Formal semantics also considers the fact that natural languages are used to talk about objects, individuals, facts, events and properties, described as external to the language itself - thus, referentiality is one of its fundamental aspects. For this reason, in formal semantics meaning is understood, on the one hand, as a relation to language, and, on the other, to what language speaks about.

Formal semantics seeks to answer the following questions: what do linguistic expressions represent or denote? how do we calculate the meaning of complex expressions from the meanings of their parts?

3. SEMANTIC RECOVERY OF INFORMATION

The Information Retrieval Systems (IRS) seek to represent the content of documents and present them to the user to meet their information needs quickly.

To do so, IR researches techniques for treatment, organization and content search based on the use of standards. IR tools generally work with indexing techniques that can quickly indicate and access documents from a textual database.

There are three main types of indexing:

- traditional indexing, in which the descriptive or characterizing terms of the documents are determined;

- full-text indexing (or indexing of the whole text), in which all the terms that make up the document are part of the index;

- indexing by tags (by parts of the text), in which only some parts of the text are chosen to generate the entries in the index (only those considered more important or more characterizing).

Searches are usually carried out using terms provided by the user or chosen by him/her among those presented.

These terms can mean the subject or class to which the desired documents belong (in traditional indexing) or the terms that must be present in the desired documents (in full-text indexes and by tags indexes).

In conventional search systems, the techniques used are syntactically based. However, when the user search involves information whose relevance cannot be given by keywords, this model is not efficient. Therefore, we note the increasing exploitation of semantic information, which enables the understanding of the concepts in their context and purpose.

Semantic marking of data at source is an example of the new technologies used for IR. SW has used this strategy to implement metadata standards that add meaningful information to the data about their contexts, marking them semantically.

The exploration of the intrinsic semantics of the data seeks the foundations of linguistics and information science to expand the universe of information retrieved and the measurement of contexts, through the use of natural language structures, such as verbal and nominal phrases, and tools for representing semantic and conceptual relationships.

There is increasing demand for Natural Language Processing (NLP) techniques that allow the construction of algorithms to search for relevant information in a large number of documents. The foundation for efficient NLP methods should be basic knowledge about language properties and especially about the semantics of concepts. This context arises from the idea of semantic memory, a theme that has been the object of psycholinguistic theories and is a rich source for the development of computational

models that intend to approach the mental processes used by the human brain to understand language. In addition to algorithms, computational models require data representing knowledge about language and common sense associations between lexical concepts and their properties. Currently, this is very difficult because this task can only be produced and verified by humans. Semantic memory works with a mental lexicon - that is, concepts and units of knowledge - and contains information about the relations between concepts, forming a conceptual network of elements connected to each other by different types of associations.

Neural networks, in turn, are a representation that have many characteristics common to human memory: they can handle incomplete or distorted information, allow automatic generalizations and display content based on context. These functions allow several applications in the IR stored in human memory, essential in precise contexts. These applications are we intend to reproduce in the computational environment.

A semantic component that influences the performance of an IRS is the thesaurus, an artificial language tool of a known domain, built by experts to represent the informational content through concepts, specifying the relations between them. It is a system of concepts that relate to each other and are represented by terms.

Each term must be linked to another, and it is this link that forms the structure of the thesaurus. Terms are used by indexers at the time of indexing and should be made available to the user at the time of IR.

Silveira and Ribeiro-Neto (2004) study the automatic use of thesaurus to improve web search results, through a ranking based on concepts that have been studied for IR in specific domains. In an experiment conducted by the authors, the search terms used in an IRS on the web were compared to the concepts of a thesaurus, used to find related concepts. Each related concept was interpreted independently and processed separately, and then combined into a Bayesian network[1], to allow a final, concept-based ranking. The aim was to see if increased concept information would increase the average accuracy of the search results.

Six sources of information were used in the study: key word, concept, specific term, general term, related term and synonym. The authors verified, among other things, that the use of a specific thesaurus for a specific domain is fundamental to improving search performance.

The experiment demonstrated a 30% increase in the average accuracy of the search results. Thus, it proposes the view that thesauruses improve both recall and accuracy in an IRS.

4. THESAURUS

According to Moreira, Alvarenga and Oliveira (2004), the term "thesaurus" originates from the Greek *thesaurós* and means treasure or repository. This term came about with the publication of the analogical dictionary of Peter Mark Roget, London, in 1852, Thesaurus of English words and phrases. The term also designates vocabulary, dictionary or lexicon, but Roget's dictionary
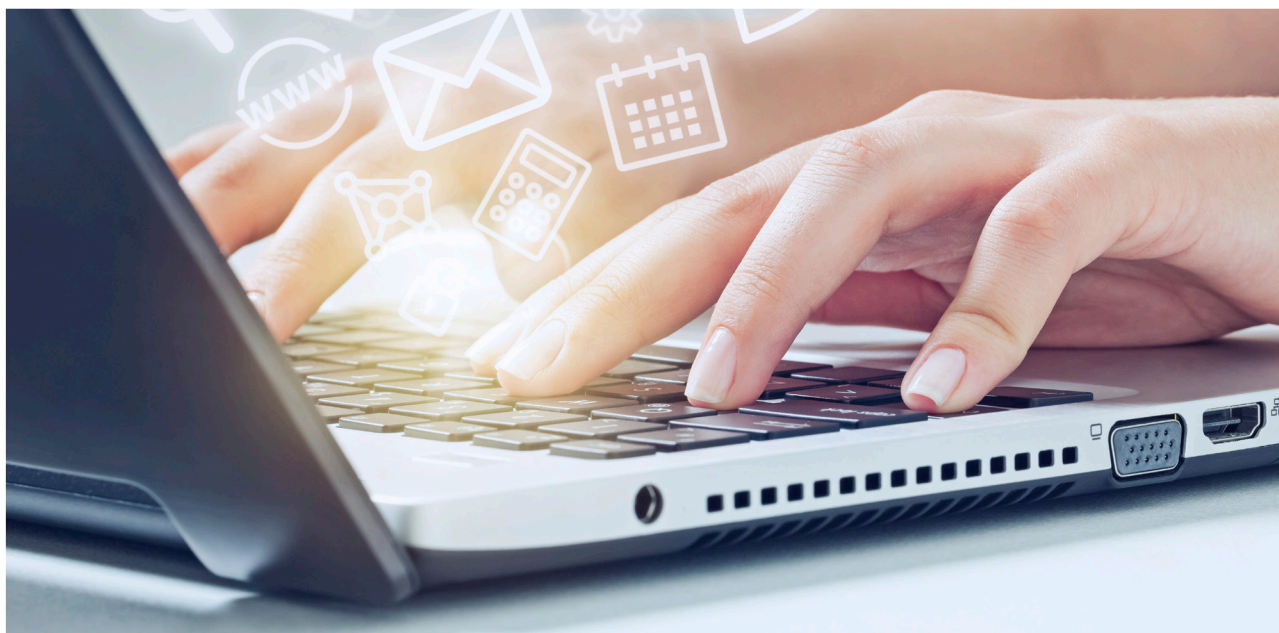
differs from others because it is a vocabulary organized according to its meaning, not in alphabetical order. The work had the merit of establishing the denomination for vocabularies that relate their terms through some kind of relation of meaning.

In his introduction, Roget defines his dictionary as a "classification of ideas" and explains that, unlike the others, his allows one to arrive at the most adequate word or the one that best fits the needs of the writer without him/her knowing, at first, what it is (GOMES, 1996).

In the 1960s, information scientist Brian Campbell Vickery presented four meanings for the term "thesaurus" in the literature of his area, the most common meaning being that of an alphabetical list of words in which each word is followed by others related to it (VICKERY, 1980 apud MOREIRA; ALVARENGA; OLIVEIRA, 2004).

Currás (1995, p.88) defines thesaurus as "a specialized, normalized, post-coordinated language, used for documentary purposes, where the linguistic elements that compose it - simple or compound terms - are syntactically and semantically related to each other. "

Tristan (2004, p. 167) defines it as "a vocabulary of terms, which is nothing more than a selection of terms, based on concept analysis, in which is defined the general greater scope term and its relation to more specific terms, which represent the minor concepts". The National Organization of Information Standards specifies:

A thesaurus is a controlled vocabulary organized in a pre-established order and structured so that the relationships of equivalence, homography, hierarchy, and association between terms are clearly indicated and identified by standardized relationship indicators used reciprocally. The primary purposes of a thesaurus are (a) to facilitate retrieval of documents and (b) to achieve consistency in indexing of written or otherwise recorded documents and other types, particularly post-coordinated information storage and retrieval systems. (ANSI / NISO Z39.19, 2003 apud SALES; CAFÉ, 2008).

According to Campos and Gomes (2006), the evolution in the construction of thesauruses is based on two perspectives, the American and the European. The thesaurus elaborated in the United States from the 1950s onwards was the result of the development that took place from the heading of subjects to the uniterm, moving from a pre-coordinated system to post-coordinated systems.

Silva (2008) states that at the same time, in England, the Classification Research Group (CRG) – based on the Faceted Classification Theory developed by the Indian mathematician and librarian Shiyali Ramamrita Ranganathan - expanded the categories of personality, matter, energy, space and time (PMEST) and developed several classification tables. This gave rise to a technique called Thesaurofacet, which allowed better posi-

tioning of the concept in the system of concepts in a given subject area, through the use of its categories. In addition, the terminological thesauruses are also based on Faceted Classification Theory, Concept Theory and some terminological principles. These instruments find in the characteristics of the concept an essential element to show the relations between the concepts and their positioning in the system, besides defining it.

The aforementioned Concept Theory, focused on the referent and originally called Analytical Theory of Concept, was launched in the late 1970s by information scientist Ingetraut Dahlberg, adding terminological principles related to the conceptual content and its definition. For Campos and Gomes (2006), this is a consolidated theory to determine what would be understood as the smaller unit in a thesaurus: the concept represented by a term. In addition, Moreira (2003) points out as innovation the use of definitions to position the concept in the system.

Bräscher (2010) points out as a function of thesaurus the translation of the language of documents, indexers and researchers into a controlled language, used in indexing and in IR in information systems. According to Sales and Café (2008), the ANSI / NISO Z39.19, 2003, emphasizes that thesauruses are not only used by information specialists at the time of indexing, but also by information users when searching for documents.

According to Café and Bräscher (2011), in thesauruses the "semantic relations are established through the analysis of the characteristics or properties of the concepts, which allow identification of differences and similarities that evidence certain types of relationships". A term present in a thesaurus can be characterized in different ways

depending on the subject in question and also the type of system one wishes to construct. The thesaurus structure comprises three main types of semantic relations to relate the terms: hierarchy, equivalence, and association.

In hierarchical relations the terms are organized into genus/species. The equivalence relations are of synonymy, that is, there are synonymous terms in the thesaurus and it should be indicated which one is the appropriate term to represent a certain concept. The association relations, in turn, present associations between terms, without specifying what kind of relation actually exists - they are just terms that relate in some way.

Thesauruses also deal with cases of ambiguity (the possibility that a linguistic communication is open to more than one interpretation) and polysemy (the possibility that a word admits more than one meaning).

5. EXTERNAL CONTROL VOCABULARY

The TCU thesaurus, called *Vocabulário de Controle Externo* - VCE (External Control Vocabulary - VCE), was launched in 2015 and aims to be a terminological control instrument that allows standardization of technical treatment and greater accuracy in the retrieval of the contents in the TCU information systems.

The interrelationship of the concepts in the VCE was expressed through relations of three types: equivalence, hierarchical and associative. The relations aim to present the descriptors in their semantic context.

- **Equivalence relation:** if it is considered synonymous or almost synonymous, , represent

the same concept for vocabulary purposes, one of them is chosen as descriptor and the others are prohibited.

- **Hierarchical relation**: relationship that expresses degrees or levels of superordination and subordination of terms; the superordinate term represents the genus of which the subordinate term is type or species.

- **Associative relation**: a gathering of related concepts that deserve to be related, but which are not bound by equivalence or hierarchical relationships.

In the VCE, each term corresponds to a concept and all terms have relationships. The relationship is determined by the meaning of the term. The relations between terms help to understand the specific concepts of external control and related areas that make up the thesaurus.

Beyond a hierarchical list of words, the VCE is composed of three distinct but interrelated parts. The first one is formed by key words, related to the fields of performance of the Court. They are accompanied by definitions and synonyms. The second part corresponds to the TCU clientele and the Supreme Audit Institutions (SAIs) associated with the International Organization of Supreme Audit Institutions (Intosai), and includes in-

formation such as history, alternative names, CNPJ and similar institutions. The third part considers the national toponymy formed by the regions, mesoregions, federative units and Brazilian municipalities.

6.   VCE APPLICATIONS

6.1   E-JURIS

The E-Juris is a corporate tool that is part of the e-TCU, which aggregates all the Court's systems related to issuing opinions and controlling court cases, containing the same logic, structure and presentation of the other court cases systems. In addition, it is integrated with the TCU portal, the corporate search system and other TCU systems such as Sagas and VCE. The premises adopted for the new system were selectivity, quality, relevance, timeliness and simplicity.

The main purpose of the e-Juris is to disseminate the relevant theses, from the point of view of precedents, which based TCU sentences. This is done through periodical publications (bulletins and newsletters) and creation of a database containing the Court precedents and making it available for research and consultations. This way, the new system unifies work processes that were previously done separately and without integration.

The relevant precedents are represented by statements, in the form of a summary. The statements repre-

sent jurisprudential precedents, not the "understanding" or prevailing Court precedent on a given issue.

The adoption of indexing by e-Juris allows greater precision in the retrieval of statements. In addition to predicting the search for synonyms, the VCE adds to the system the suggestion functionality of correlated subjects to be searched, since the tool is structured in a system of interrelated concepts by hierarchy, equivalence and association.

6.2   TCU PORTAL DIGITAL LIBRARY

Corporate repositories of knowledge are used to disseminate the information produced internally, to allow access to the organizational culture and to subsidize the information transformed into knowledge. Since they are more than a depository of documents, repositories can act decisively, supporting the development of new products and services and capacity building and training of the organization's workforce. In addition, they usually serve as an official source of information to partners and collaborators, support daily activities and help the decision-making process.

At TCU, the digital library is one of easiest to use and most accessible corporate repositories of knowledge. Developed to organize, handle and disseminate information that can generate new knowledge, the digital library allows the insertion of material with several documentary typologies. Thus, in the same environment it is possible to find books, academic papers, presentations, booklets, periodicals, agreements, contracts, official correspondence and norms, as well as various images and other types of resources. The tool allows two levels of access to the deposited contents to be established: it is possible to allow free access to documents of a public nature, in the same way that it is possible to limit access to documents of an internal nature or that have some type of restriction.

The insertion of documents into the environment is done in a decentralized way and there are content managers responsible for approving what will be available in the repository. The digital library also has a data entry form that can be used throughout the Court. This form is composed of controlled metadata and allows the description of the content with elements such as title, authorship and date. Moreover, it requires that the documents be classified in a thematic tree and that the subjects treated be translated by keywords derived from a controlled vocabulary.

In other words, TCU digital library is a corporate repository of knowledge that requires information to be classified and indexed. In order to index content, the environment is integrated with the VCE. In addition, since the library is also integrated with the textual search tool of the corporate portal, it is possible to directly search the portal and retrieve the content deposited in the library environment.

6.3   GUIDANCE SYSTEM

The Guidance System was designed as a tool to guide, manage and disseminate knowledge about external

control. It allows any Court employee to forward questions about pre-selected topics of external control, such as auditing, planning, annual accounts, special accounts, representations, complaints, requests, quality evaluation, executive debt collection, external control standards and other process procedures, as well to ask questions about the Fiscalis system

After selecting the theme, the system directs the question to the unit responsible for the area. Based on the questions and answers collected, each respondent unit automatically creates a database of frequently asked questions (FAQ), which is stored in the system and available for queries and searches by all servers.

Several TCU units have already been registered as respondent units and the system is integrated with the VCE. Adoption of the controlled vocabulary allows greater accuracy in the retrieval of the desired information, since index terms that represent the subjects treated are assigned to both questions and answers.

6.4   EXTERNAL CONTROL WIKI

Among the various possibilities of storing the knowledge of an institution, one in particular deserves to be highlighted in an era in which we produce knowledge through various collaborators and partners: wikis - a technology tool known as "social software" and designed with a set of characteristics which allow the creation and organization of knowledge in the collaborative world. The use of wikis has proved to be a low-cost solution with a high degree of efficiency, to foster the cooperative creation of knowledge within organizations.

The wiki environment is the evolution of the concept of computer supported cooperative work (CSCW), which arose from the need for organizations to have people working in different physical locations and, at the same time, needing to achieve quick results together; that is; it has emerged to facilitate the communication and productivity of remote groups.

Since 2009, the TCU uses the free software Mediawiki to manage an external control wiki and restricted access to its employees. It is an important collaborative space of knowledge construction that brings together informal tutorials and specialized entries coming from the VCE.

The Wiki can be accessed and edited by all employees; they aggregate information to entries and tutorials on topics covering norm, law, and doctrine. The Wiki is an important knowledge management tool insofar as it provides information and documents useful to the daily work of the auditors, according to their area of activity. It is con-

figured as a genuine collective intelligence environment of the organization.

6.5   FUTURE APPLICATIONS

The VCE also has potential applications in various information systems of the Court. The integration of terminology into the TCU portal search tool, through the adoption of controlled vocabulary for both treatment and search, is a key factor for increasing precision and speed in IR.

Another possibility of using VCE is as a dictionary in software for automatic indexing of large volumes of documents, serving as a terminological parameter in the area of external control. This functionality has already been tested during the process of automatic indexing of statements, which was part of a conceptual test to acquire data mining and semantic data analysis software.

In an ideal scenario, we envisage that the Court will adopt standards such as subject metadata and the terminological control tool to ensure the improvement of its of IRS performance.

NOTES

1   Bayesian networks constitute a graphical model that simply represents the causal relations of the variables of a system. In summary, Bayesian networks, also known as networks of opinion, causal networks and probabilistic dependency graphs, are graphical models for reasoning (conclusions) based on uncertainty, in which the nodes represent the variables (discrete or continuous), and the arcs represent the direct connection between them (SILVEIRA; RIBEIRO NETO, 2004).

REFERENCES

ALMEIDA, M. B.; SOUZA, R. R. Avaliação do espectro semântico de instrumentos para organização da informação (Evaluation of the semantic spectrum of instruments for the organization of information). Encontros Bibli – Revista Eletrônica de Biblioteconomia e Ciência da Informação (Meetings Bibli - Electronic Journal of Library Science and Information Science), Florianópolis, v. 16, n. 31, p. 25 50, 2011. Available in: <http://mba.eci.ufmg.br/downloads/11963-60907-1-PB.pdf>. Access in: Nov 23 2016.

ALTOUNIAN, M.; ZAULI, A. A semântica na recuperação da informação na web: novas tendências (Semantics in the retrieval of information on the web: new trends). 2013. Work submitted as a partial requirement for approval in the discipline Recuperação da Informação (Information Retrieval), Escola de Ciência da Informação (School of Information Science), Universidade Federal de Minas Gerais (Federal University of Minas Gerais), Belo Horizonte, 2013.

BRÄSCHER, M. Elaboração de tesauros (Elaboration of thesaurus). Brasília, 2010.

CAFÉ, L.; BRÄSCHER, M. Organização do conhecimento: teorias semânticas como base para estudo e representação de conceitos (Organization of knowledge: semantic theories as a basis for study and representation of concepts). Informação & Informação (Information & Information), Londrina, v. 16, n. 3, p. 25 51, Jan./jun. 2011.

CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador (Conceptual thesaurus elaboration methodology: categorization as guiding principle). Perspectivas em Ciência da Informação (Perspectives in Information Science), Belo Horizonte, v. 11, n. 3, p. 348 359, Sept./dec. 2006.

CAPRI, D.; GARRIDO, I.; DUARTE, R. Recuperação semântica da informação (Semantic information Retrieval). 2009. Work submitted as a partial requirement for approval in the discipline Recuperação da Informação (Information Retrieval), Centro de Ciências da Educação (Center of Educational Sciences), Universidade Federal de Santa Catarina (Federal University of Santa Catarina), Florianópolis, 2009. Available in: <http://pt.slideshare.net/doritchka/angel-recuperao-semntica-da-informao>. Access in: 23 Nov. 2016.

CURRÁS, E. Tesauros: linguagens terminológicas (Thesaurus: terminological languages). Brasília, DF: CNPq; Ibict, 1995.

GOMES, H.E. Classificação, tesauro e terminologia; fundamentos comuns (Classification, thesaurus and terminology; common fundamentals). Rio de Janeiro: UNIRIO, 1996.

MOREIRA, A. Tesauros e ontologias: estudo de definições presentes na literatura das áreas das ciências da computação e da informação, utilizando-se o método analítico-sintético (Thesaurus and ontology: a study of the definitions found in the Computer and Information Science Literature, by means of an analytical-synthetic method). Perspectivas em Ciência da Informação (Perspectives in Information Science), Belo Horizonte, v. 8, n. 2, p. 216-226, Jul / dec.2003.

_____; ALVARENGA, L.; OLIVEIRA, A. P. Thesaurus and ontology: a study of the definitions found in the Computer and Information Science Literature, by means of an analytical-synthetic method. Knowledge Organization, v. 31, n. 4, p. 231-244, 2004.

SALES, R., CAFÉ, Lígia. Semelhanças e Diferenças entre Tesauros e Ontologias (Similarities and Differences between Thesaurus and Ontologies). DataGramaZero - Revista de Ciência da Informação (DataGramaZero - Revista de Ciência da Informação), v.9 n.4 Aug 2008.

SILVA, A. Análise das relações semânticas em tesauros jurídicos brasileiros: orientações das normas e aplicação prática (Analysis of semantic relations in Brazilian legal thesaurus: guidelines of the norms and practical application). 2013. Completion of course work Bacharel em Biblioteconomia (Bachelor of Librarianship) – Centro de Ciências da Educação (Center for Educational Sciences), Universidade Federal de Santa Catarina (Federal University of Santa Catarina), Florianópolis, 2013. Available in: <https://repositorio.ufsc.br/bitstream/handle/123456789/103801/TCC_Aline_da_Silva_20131PDFA.pdf?sequence=1>. Access in: 23 Nov. 2016.

SILVA, D., SOUZA, R., ALMEIDA, M.B. Ontologias e vocabulários controlados: comparação de metodologias para construção (Ontologies and controlled vocabularies: comparison of methodologies for construction). Ciência da Informação (Ciência da Informação), v. 37, n. 3, p. 60-75, Sept / dec. 2008.

SILVEIRA, M. L.; RIBEIRO-NETO, B. Concept-based ranking: a case study in the juridical domain. Information Processing & Management, Doha, v. 4, n. 5, p. 791 805, Sept. 2004.

SOUZA, A. et al. Recuperação semântica de objetos de aprendizagem: uma abordagem baseada em tesauros de propósito genérico (Semantic retrieval of learning objects: an approach based on thesaurus of general purpose). In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (BRAZILIAN SYMPOSIUM ON COMPUTER EDUCATION), 19., 2008, Uberlândia. Anais… Uberlândia: SBIE, 2008. p. 603 612.

SOUZA, A.; ROCHA, R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências (Information retrieval systems and web search engines: current outlook and trends). Perspectivas em Ciência da Informação (Perspectives in Information Science), Belo Horizonte, v. 11, n. 2, p. 161 173, May / Aug. 2006. Available in: <http://www.scielo.br/pdf/pci/v11n2/v11n2a02.pdf>. Access in: 23 Nov. 2016.

SZYMANSKI, J.; DUCH, W. Information retrieval with semantic memory model. Cognitive Systems Research, Kalamazoo, v. 14, n. 1, p. 84 100, Apr. 2012. Available in: <http://ethologie.unige.ch/etho5.10/themes/semantic.memory/szymanski.duch.2011.information.retrieval.in.sematic.memory.neural.network.models.pdf>. Access in: 23 Nov. 2016.