

OLIVEIRA JUNIOR, Carlos Duarte de. **Extração automática de contextos definitórios em textos acadêmicos da ciência da informação**. 2012. 116 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade Brasília.

URL: <http://repositorio.unb.br/handle/10482/10954>

Resumo: O trabalho apresenta estudo sobre o papel da Ciência da Informação, sua interdisciplinaridade e interseção com a Linguística e a Ciência da Computação no que se refere à utilização dos textos como fonte de informação e conhecimento a ser organizado ou reorganizado, nos grandes repositórios de informação já existentes, com a finalidade de recuperação. A ênfase é na extração automática de Contextos Definitórios em textos, o que se entende como qualquer fragmento textual que introduz e associa um termo a uma definição. Cita teorias de Organização da Informação como Classificação Facetada de Ranganathan, a teoria do Conceito de Dahlberg e as teorias da terminologia, tais como a Teoria Geral da Terminologia de Wüster e a Teoria Comunicativa da Terminologia de Cabré. Todas as teorias são abordadas com enfoque na importância do termo e principalmente da definição como elemento primordial para o mapeamento semântico de um documento e de um domínio do conhecimento. Enfatiza a visão da definição como elemento de ligação entre os objetos e seus conceitos, identifica tipos de definições, cita estudos anteriores de identificação e extração automática de enunciados definitórios em inglês, espanhol e francês. Menciona as técnicas de Processamento de Linguagem Natural e Descoberta de Conhecimento em Textos como ferramentas para o processamento e extração de informação em documentos escritos em língua natural. Por fim, propõe um método de extração automática de Contextos Definitórios em textos acadêmicos da Ciência da Informação, a partir de uma gramática de padrões definitórios em língua portuguesa criada no âmbito da pesquisa. Entende-se gramática de padrões definitórios como um conjunto de expressões linguísticas capazes de identificar um CD em um texto. A gramática foi validada comparando uma extração manual com uma automática. O método foi aplicado nas teses e dissertações da Faculdade de Ciência da Informação

Palavras-chave: Contexto definitórios; Extração automática de textos; Organização da informação; Terminologia.

Abstract: The thesis presents a study on the role of Information Science, and its interdisciplinary intersection with Linguistics and Computer Science with regard to the use of texts as a source of information and knowledge to be organized or reorganized, in large information repositories existing, with recovery purposes. The emphasis is on automatic extraction of Definitory Context in texts, which is understood as any fragment of text that introduces and associate a term with a definition. It makes reference to the theories of Information Organization and Faceted Classification of Ranganathan's theory of concept Dahlberg and theories of terminology, such as the General Theory of Terminology of Wüster's and the Communicative Theory of Terminology of Cabré's. All theories are discussed with emphasis on the importance of the term and the definition as a major element for the semantic mapping of a document and a domain of knowledge. It emphasizes the view of the definition as a liaison between the objects and their concepts, identifies types of settings, refer to previous studies of automatic identification and extraction of Definitory Enunciation in English, Spanish and French. It mentions techniques of Natural Language Processing and Knowledge Discovery in Texts as tools for processing and extraction of information in documents written in natural language. Finally, it proposes a method for automatic extraction of Contexts in academic texts of Information Science, from a Definitory grammar patters in Portuguese established within the research. It is understood that Definitories grammar patters as a set of that linguistic expressions can identify a DC in a text. The grammar was validated by comparing an automatic with a manual extraction. The method was applied in thesis and dissertations at the Faculty of Information Science at the University of Brasilia - UNB, available from it repository RIUnB, from 2006 to 2011.

Keywords: Automatic text extraction; Definitory context; Information organization; Terminology.