

CAMARA JUNIOR, Auto Tavares da. **Processamento de linguagem natural para indexação automática semântico-ontológica**. Brasília, 2013. 180 f., il. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília.

URL: <http://repositorio.unb.br/handle/10482/13768>

Resumo: A pesquisa propõe uma arquitetura de indexação automática de documentos utilizando mecanismos de processamento de linguagem natural em nível semântico. Por meio do arranjo de ferramentas e recursos existentes, agregado ao desenvolvimento de software para integração, é construído um sistema de indexação automática que utiliza conhecimento modelado em uma ontologia para análise semântica. A aplicação da arquitetura é exemplificada e posta à prova em um conjunto de laudos periciais de crimes cibernéticos produzidos pela Perícia Criminal da Polícia Federal. Os resultados de pesquisa apontam para a melhoria que o aprofundamento à análise semântica do processamento de linguagem natural oferece para a seleção automática de descritores e, por extensão, para a recuperação da informação. Dentre as contribuições inéditas desta tese incluem-se a arquitetura de sistema de informação para indexação automática proposta, a ontologia construída para a análise semântica e as implementações em linguagem de programação Python. Por fim, o trabalho insere-se em uma área de pesquisa que tem sido cada vez mais investigada, no entanto ainda carece de resultados amadurecidos, qual seja o processamento de linguagem natural em língua portuguesa.

Palavras-chave: Indexação automática; Linguística - processamento de dados; Ontologia; Semântica.

Abstract: The research proposes an automatic indexing architecture of documents using natural language processing mechanisms on semantic level. By organizing existing tools and resources, added to software development for integration, an automatic indexing system that uses knowledge modeled by ontology to semantic analysis is built. The applicability of the architecture is exemplified and put into proof on forensics reports of cybercrimes produced by Federal Police Forensics Experts. The research results point to the benefits that semantic analysis on natural language processing offers to automatic descriptor selection and information retrieval. As contributions of this thesis there are the proposed automatic indexing information system architecture, the ontology built to semantic analysis and the implementations on Python programming language. Concluding, the work inserts itself in a research area that has been increasingly more investigated, however still demanding mature results, which is natural language processing on Portuguese language.

Keywords: Automatic indexing; Linguistics – processing of data; Ontology; Semantics.