

Knowledge Extraction System from Unstructured Documents

Rodriguez T.

Aguilar J.

In this paper, we design and implement a knowledge extraction system from unstructured documents (unstructured documents are documents where the information is in natural language, and require natural language processing techniques for processing) in HTML format. Basically, the system allows to transform the content of a text that is in natural language, into structured and organized knowledge, semantically described (a Semantic Ontology). Therefore, it is proposed to generate semantic knowledge based on the extraction of entities and relationships, where entities are anything about which something can be said, and relations the interactions between entities. From the generated semantic knowledge model, it is possible to infer new knowledge, such as lexicons, taxonomies and specialized terminological bases. The system can be used by any semantic processing application, in its processes of enriching its information and knowledge. ©

2003-2012 IEEE.

Knowledge Extraction

Natural Language Processing

Semantic Knowledge

Semantic Ontologies

Extraction

Knowledge based systems

Ontology

Semantics

Design and implements

Knowledge extraction

Natural languages

Semantic knowledge

Semantic ontology

Semantic processing

Unstructured documents

Natural language processing systems