

ABSTRACT [EN]

Each domain, alongside its knowledge base, changes over time and each period is centered on specific topics that emerge from different ongoing research projects. A researcher's daily activities usually involve the study of new papers and using the information for building solutions and observing how the domain evolves. Since the retrieval of documents from the Internet can lead to large data flows, it is important to consider other approaches for a more comprehensive analysis of the domain. In this context, the *Semantic Meta Annotations* focus on building a scalable paper annotation system that automatically retrieves papers on a given topic and tags them, to make the exploration phase of the research literature substantially easier.

Evolution is based on leveraging existing knowledge, researches and tools to test other ideas. A researcher needs to read many textual materials, which are many times cluttered with irrelevant information. Thus, the focus of our research is shifted towards understanding the way in which humans comprehend texts. Reading is a complex cognitive process which has been the subject of many studies throughout the years. It is one of the oldest ways for learners to acquire new information and to consolidate existing knowledge, representing a key evolutionary element. Each textual material contains facts and topics that activate existing concepts from the reader's prior knowledge (memory). The *Comprehension Model* describes an automated method that analyzes the way in which readers potentially assimilate and conceptualize new text information, which is a novel alternative for indexing and meta-annotating textual corpora. Creating such a method is a challenge, as it requires using a computational knowledge base, parsing unstructured textual materials and linking concepts using various heuristics and semantic similarity measures.

Our research focuses on the semantic analysis of unstructured textual materials by using Natural Language Processing techniques and models such as Latent Semantic Analysis, Latent Dirichlet Allocation, Word2Vec or semantic distances within lexicalized ontologies, i.e., WordNet. Within the experiments focused on semantic meta annotations, these distances are combined with other metrics such as co-citation analysis or co-authorship, thus creating the basis of several interactive and exploratory visual graphs that offer a better domain overview within a scalable infrastructure. In the second experiment, our focus is shifted towards describing an automatic comprehension modelling technique that analyzes using computational representations and algorithms the reading process. Our goal was to create a set of methods and tools to help researchers in their daily work to easily retrieve and understand textual materials.

Keywords: *Discourse Analysis; Semantic Similarity; Extraction of Domain Key Concepts; Text Analysis; Comprehension Model; Text-Based Inferences; Inferred Knowledge.*