

A Framework for Information Extraction, Storage and Retrieval

Samhaa R. El-Beltagy, Mohammed Said, and Khaled Shaalan

ABSTRACT

This paper presents a set of tools that were developed in order to facilitate and speed up the process of building information extraction and retrieval systems for documents that exhibit a set of predefined characteristics. Specifically, the work presents a simple framework for extracting information found in publications or documents that are issued in large volumes and which cover similar concepts or issues within a given domain. The paper presents a simple model for defining background knowledge and for using that to automatically augment segments of input documents with metadata in order to assist users in easily locating information within these documents through a structured front end. The model presented makes use of both document structure as well as dynamically acquired background knowledge to achieve its goals.