

A Greek Retrieval System in the Legal Domain

Apostolis Apostolou, Theodore Kalamboukis, Lilian Mitrou,
Athens University of Economics and Business,
Department of Informatics

Extended Abstract

Information retrieval in the law domain has a long history in the research community and is one of the first and even today most challenging research activity. Legal documents are highly structured and usually there is a relation between content and structure. This explains the development of an independent research field in Artificial intelligence and Law.

In this paper, we describe the development of a new Greek Legal Information Retrieval System. The main objective of this system was the implementation of some modern practices, such as the passage based retrieval, the navigation from document to document via hyperlinks, the simultaneous usage of structured and unstructured search and the presentation of the documents using the exact layout of their actual printed counterparts. The system is available to the users through web address *195.251.252.22:8084/laws/*

To increase the precision of retrieval a combination of field-based and passage retrieval was applied. Each legislation document contains two fields: a title and body. The title is automatically extracted from texts and the documents are split into pages which form the basic passage unit. The system initially evaluates and retrieves pages independently, and calculates the total document relevance-score by adding the weighted scores of the title and the body. This scoring technique proved to enhance retrieval performance in the evaluation section.

Two features may be considered novel of this prototype: Documents are presented with their exact physical (printed) layout, and navigation from document to document is possible using automatically assigned hyperlinks in the pdf files. The interface is simple and friendly and offers useful extra information to users when browsing through the results.

Our system gives administrators all the necessary tools to create and manage a legal document collection from scratch. Every procedure, from the initial download of the documents to the reference detection and conversion to links is automated and demands no human inspection.