

A Retrieval System of Greek Legal Documents

Aggeliki Plati, Theodore Kalamboukis
Department of Informatics
Athens University of Economics and Business
tzk@aueb.gr

An information retrieval system has been implemented, that serves the need for convenient, simple and fast information concerning legal documents. The retrieval of legal documents usually requires that the person concerned, should have specialized knowledge on the law and on the structure of the statutes, regulations, judicial decisions etc. Not every person has this knowledge on the law matters. The proposed system allows the user to search for legal information by asking simpler, non-standardized and free formed questions-queries.

Lawyers would like the law documents to have the same form as the original document. In addition they need more information to accompany the legal documents, such as law categorization, links to relative laws and to laws that the current law modifies, so that they can have access to laws back in the past. The relation between the modified law and the law that modifies the first one should be bidirectional.

The problems that we came across are mainly due to

- the unstable and loose format of the texts, that does not facilitate their automatic processing
- the nature of the Greek language. One of these problems, for example, is often the inability of reading a PDF file in order to convert it to a TXT file, suitable for processing by the system. Additionally, the peculiarities of the legal language in Greek raise difficulties and are a challenge in handling various problems in Natural Language Processing.

In this study, a requirements analysis is done so as to create a system that can actually be useful by simulating the way that the lawyers and legals search out for information.

An other interesting and innovating approach that has been done under this work is the segmentation of the laws for the improvement of the performance of the document retrieval. For this purpose, an information retrieval system is implemented with the Apache Lucene search engine.

At present the system is based on passage retrieval where passages are defined equal to the size of the pages in the original law. This transformation, splitting into pages and conversion to txt format are applied automatically. These pages are pre-processed in order to transform to HTML format with exactly the same appearance as the original document. Furthermore links to the next and previous pages have been added automatically to support browsing at searching time.

In concluding, this work deals with the design and implementation of a pilot legal information retrieval system, from both a theoretical and a practical point of view. It is certainly the first attempt, as far as we know to deal with law documents in hypertext form for the Greek language that gave us the motivation for new research problems which are currently under investigation.

References

1. Aggeliki Plati, *Nomothetes, A Law Information Retrieval System, BSc Thesis 2011, Athens University of Economics and Business.*
2. *Passage Retrieval in Legal Texts, Paolo Rossoa, Santiago Correea, Davide Buscaldib, Journal of Logic and Algebraic Programming 80, 3-5 (2011) 139-153*