Experiments in finding relevant case law

Winkels, R.G.F.

Link to publication

Open Data Workshop

3RD INTERNATIONAL WORKSHOP ON "NETWORK ANALYSIS IN LAW"

WEDNESDAY DECEMBER 9TH 2015
IN CONJUNCTION WITH JURIX 2015:
28TH INTERNATIONAL CONFERENCE ON LEGAL KNOWLEDGE AND
INFORMATION SYSTEMS, BRAGA, ORTUGAL

Table of Contents

Experiments in finding relevant case law

Within the OpenLaws.eu project, we attempt to suggest relevant new sources of law to users of legal portals based on the documents they are focussing on at a certain moment in time, or those they have selected. In the future we attempt to do this both based on 'objective' features of the documents themselves and on 'subjective' information gathered from other users ('crowd sourcing'). At this moment we concentrate on the first method.

In Winkels e.a. (2014) we have described results of experiments using network analysis to suggest these new documents. Based on the current document in focus, i.e. an article in a Dutch law, we calculated a local network of sources of law based on the references from and to this focus document. This network contained both case law and legislation. Now we describe two experiments where we mix the use of a network of references with similarity based on the comparison of the text of documents.

The first experiment concerns case law within the Dutch tax domain. Bag-of-words combined with TFIDF weighting and cosine similarity has been used to find the textual similarity between two case law documents; this serves as a baseline for evaluating. The same algorithms are used to calculate the reference structure similarity between two documents. The references are extracted from the documents by a parser that has a precision of 0.55 in this domain. Experts evaluated the recommendations done by the baseline and the text similarity combined with the reference similarity algorithm.

The evaluation resulted in the conclusion that adding a similarity measure on reference structures is not performing as well as text similarity alone if not all references are identified by the parser.

The second experiment concerns case law in the Dutch immigration law domain. Suggestions for new documents are generated by comparing the (pre-processed) texts of the case laws using a topic model. The topic model represents the case laws as mixtures of topics, after which the most similar case laws are found by calculating the similarity between the topic mixtures.

The topic similarity based suggestions that this project generated, were evaluated by legal experts and novices, who ranked three suggestions from best to worst, and stated whether the suggestion is good enough for a recommender system. One of the three suggestions was based on topic similarity, another was based on references to legislation and the other was based on a combination of these two methods. Both legal experts and novices showed significant preference for the suggestions based on topic similarity. The legal experts wanted to see 85% of the suggestions based on topic similarity in a recommender system, for the novices this was 86.7%. The legal experts ranked the suggestions based on topic similarity as best suggestion 80% of the time, while the novices always ranked the suggestion based on topic similarity as best suggestion. This suggests that suggestions based on topic similarity can give useful suggestions within Dutch case law.

References
Winkels, R.G.F., Boer, A., Vredebregt, B. & Someren, A. van (2014). Towards a Legal Recommender System. In R. Hoekstra (ed). *Proceedings of JURIX 2014*, IOS Press, Amsterdam, pp. 169-178.