Legal Knowledge and Information Systems
A. Wyner and G. Casini (Eds.)

doi:10.3233/978-1-61499-838-9-21

Cloudy with a Chance of Concepts

Suzanne BARDELMEIJER[a], Alexander BOER[b], Radboud WINKELS[b]
[a] *Science Faculty, University of Amsterdam*
[b] *Leibniz Center for Law, University of Amsterdam*

Abstract. In this paper we present the results of a study to see whether automatically generated concept maps help users of legal information systems in understanding the topics of documents they retrieve. A small formative evaluation with novice users is presented. We did not find a significant difference between the ability to connect the correct visualisation to a document between a topic cloud and concept map approach. Topic clouds are probably a little easier to understand quickly in a superficial way.

Keywords: topic clouds, concept maps, LDA, legal recommender system

1. Introduction

Over the past few years, more and more legal documents have become publicly available online. In 2016 for instance, almost 34,000 new court decisions were published on the official Dutch judicial portal[1] [1]. These decisions are often complex due to their lengthy and complicated structure. A clear visualization of the different topics that a document deals with, might help in understanding a legal document quickly, and could help professionals and prevent novices from feeling overwhelmed by the length of the document. In this way it fits in our research on *legal recommender systems* in recent years (cf.[2]).

One way to visualize different topics is through topic clouds. A topic cloud is a visual representation of words (concepts) where the importance of a word in an underlying set of data (text) is expressed by its size [3].

Another way of visualizing knowledge in complex documents is via concept maps. A concept map is a graphical representation of knowledge in which the core concepts and the relations that connect these concepts are structured in a network diagram [4]. These concept maps, constructed by an expert in a certain field, help organize prior and newly acquired knowledge and therefore assist information gathering. In recent years, different algorithms have been created that quickly generate topic clouds automatically. However, the creation of a concept map is still a labour-intensive and time consuming process, mainly because substantial expert knowledge is typically needed.

With previous results and attempts in mind, this paper presents a new approach for the creation of concept maps from court decision documents. We aim to create small and comprehensive concept maps that capture the essence of document topics. We are interested in knowing to what extent it is possible to automatically create such a concept map from Dutch court decisions quickly. Furthermore, a comparison will be

[1] www.rechtspraak.nl

made between the generated concept maps and topic clouds to answer the question whether users have a preference for one or the other as a means of visualisation.

We will start with describing related work. Next, the concept map generation method is discussed, followed by the evaluation and interpretation of experimental results. To conclude this paper, recommendations for future research will be made.

2. Concept Maps & Topic Modelling

Fundamental research in concept mapping was first done by Novak and his researchers. Concept maps are graphical tools for representing knowledge through organizing concepts and their relations, to advance human learning and understanding [4]. The process of constructing a concept map normally requires a substantial amount of time and expert knowledge. A number of studies have focused on providing methods to help automate the process (e.g. [5]).

Most topic models are currently constructed using Latent Dirichlet Allocation, which was first presented by Blei e.a. [6]. The idea behind LDA is that documents are seen as a mixture of different topics, where a topic is formally defined as a distribution over a fixed vocabulary. LDA results in the creation of a topic model where each topic is associated with a document in different proportions and has probabilities of generating various words.

LDA operates under the bag-of-words assumption, meaning that word order is not taken into account. This assumption is plausible for the identification of a topic, but gives a disadvantage when interpreting them. Research showed that topic models based on LDA with a multi-word expression approach provide a better understanding for what a topic is about [7,8].

3. Research Method

We selected a sample of case law from the Dutch portal on immigration law for the period 2015-2017: 250 cases. Most Dutch court decisions have a common structure, starting off with a summary of the case, followed by the actual verdict which consists of the procedure, the considerations, and the decision. Since every section of the decision contains information about the case and could therefore be of interest in identifying underlying topics, all sections are considered relevant. All in all we processed 968 text files from the 250 cases. Every file contains a section of a court decision. A file is named after the case's ECLI number [2], which is a unique identification number, and a section number. To finalize the pre-processing step, we removed punctuation, and all capital letters were converted to lowercase. A list of stop words was created by computing the frequency of every word in the corpus. If a word occurred in more than 20% of all files, the word was added to the list of stop words.

To select promising n-grams, all text files were divided in bi-, tri- and tetra-grams. An n-gram is considered promising when none of its terms consist of one character or

[2] European Case Law Identifier (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:127:0001:0007:EN:PDF)

one digit. E.g. the bi-gram “is_a" is not considered promising. Also, none of the words should be a stop word. All promising n-grams are appended to the text file.

4. Building the Models

We used the open source implementation of LDA in MALLET[3] to build our topic models. We set the target number of topics to 50. Although the number of topics that LDA produces is arbitrary, it has a big influence on the informational value of the created topics. If the target number is reduced, separate topics have to merge; if the number increases, topics have to split. In order to ensure a topic is describing a theme that is well-represented in a set of documents, only documents that have more than 20% similarity with one of the topics are selected. To guarantee a topic is not too specific, the remaining topics are compared based on the number of documents they describe. For the 20 topics that describe most documents a concept map is created.

4.1. Creating Concept Maps

After finishing the steps above, 20 topics with their descriptive terms were extracted from the data set. These terms represent the concept nodes in a concept map of that topic. Since a concept map is created out of concepts and a limited number of most relevant relations between them, the next step is to assign weights to term pairs. If a topic is described by n terms, then there are $n(n-1)/2$ pairs, i.e. 171 pairs for 19 terms. The weight of a term pair is the summation, over the set of documents, of the products of the term frequencies of the term pair in the document. The weighted term pairs are stored in a list in descending order. The most informative pairs, with the highest weight, are selected as links and a concept map is constructed using CmapTools[4].
A comprehensible concept map should not contain concepts with more than three (incoming and outgoing) links. Therefore, once a concept already has three links with other concepts, new links to that concept are skipped. This process will continue until the concept map consists of 15 connected concepts. Figure 1 shows an example for topic 36 (in Dutch).

Although traditional concept maps exist of linking phrases containing a verb, the linking phrases in the created concept maps have a number that refers to a section of a court decision in which the link has the highest score. This approach ensures that the idea of the traditional concept map is maintained while avoiding natural language processing. In the evaluation this link information was not used. When integrated in an application, this information can provide users more insight in the origin of the link between two concepts.

4.2. Creating Topic Clouds

The output of LDA using MALLET is a table that expresses i.a. the importance of a term to a topic. This table gives valuable information; Some terms are more strongly connected to a topic than others.

[3] MAchine Learning for LanguagE Toolkit (http://mallet.cs.umass.edu)

[4] https://cmap.ihmc.us/

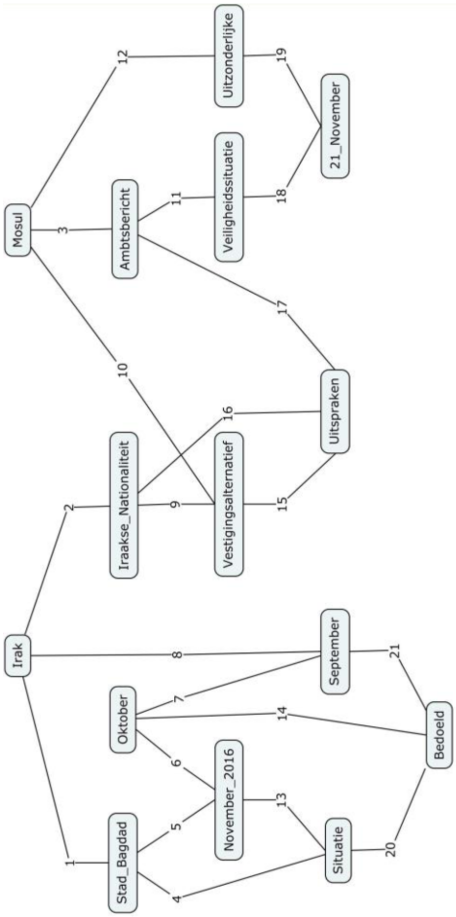

Figure 1: Example Concept Map for Topic 36

Figure 2: Example Topic Cloud for Topic 36

The strength of association of a term to a topic is expressed by the size of a term in a topic cloud. Figure 2 gives an example for topic 36 (in Dutch).

5. Formative Evaluation

We created concept maps and a topic clouds for each of the 20 topics that were extracted from our data. To evaluate these concept maps and topic clouds, six novices in the field of law participated in a study in which they had to complete sorting tasks. The study consisted of four tasks where in each task the participant was given a section of a case law and was asked to:

1. Rank in descending order four concept maps according to their similarity with the given document;
2. Rank in descending order four topic clouds according to their similarity with the given document.

Table 1 presents the documents, identified by their ECLI number, that were used for the study. These documents were selected since their content covers a wide variety of topics which makes them suitable for ranking. The topics beneath the ECLI number fit the document in different degrees and this value decreases gradually. For instance, document ECLI:NL:RBDHA:2017:3176-2 is matched by topic 7 for 51%, topic 21 for 12%, etc.

The Spearman rank-order correlation, which is a non-parametric measure of the degree of association between two variables, was used to evaluate the survey. Table 2 presents the Spearman rank-order correlation coefficient and p-value for every task and in total. Although the number of participants is too small to draw any statistically sound conclusions, the correlation coefficients indicate a positive correlation for both concept maps and topic clouds. The values also suggest that perhaps participants were slightly more capable of adequate ranking using topic clouds than concept maps. Task 4 apparently either was more difficult than the other three tasks or the subjects were a bit tired and payed less attention.

Table 1: The four documents and their topics

ECLI:NL:RBDHA: 2017:3176-2		ECLI:NL:RBDHA: 2017:417-2	
Topic 7	0,513	Topic 48	0,295
Topic 21	0,118	Topic 33	0,113
Topic 33	0,056	Topic 30	0,067
Topic 16	0,014	Topic 12	0,015
ECLI:NL:RBDHA: 2017:2654-2		ECLI:NL:RBDHA: 2017:780-2	
Topic 36	0,489	Topic 10	0,328
Topic 29	0,131	Topic 45	0,106
Topic 3	0,021	Topic 30	0,030
Topic 9	0,009	Topic 12	0,015

Table 2: Spearman rank-order and p-values

	Concept Maps	p-value	Topic Clouds	p-value
1	1.00	0.0	1.00	0.0
2	0.77	1.00 E-05	0.97	0.0
3	0.93	0.0	1.00	0.0
4	0.40	0.053	0.39	0.056
Total	0.78	1.94 E-04	0.84	1.62 E-10

6. Conclusion

The method that is presented in this paper for the automatic creation of comprehensible concept maps from case law documents shows potential since novices in the field of law were able to rank the created concept maps adequately given a section of a case law. However, due to the marginal difference in performance of the two visualizations, we cannot conclude that there exists a distinct preference for either topic clouds or concept maps for the visualization of underlying topics in the case law documents.

While these results suggest that participants were able to make more correct rankings using topic clouds than concept maps, this does not necessary imply that topic clouds have more informational value. Participants could use two main methods to rank the concept maps and topic clouds. One method is aimed at identifying the internal structure of the given document, whereas the other is focused on words itself, as the LDA algorithm does. The first method requires substantial knowledge of the content of the document, while the second is based on superficial resemblance. Therefore, the results of the survey could be misleading if either the second method simply works better on LDA generated topic classifications, or if laymen use term frequency to determine similarity to both types of diagram. Further research could include experts as well to examine whether experts perform better or worse. In addition, obviously, more participants are needed in order to draw statistically sound conclusions.

Although a number of preliminary pre-processing steps were performed in order to establish a clear-cut topic model, this process can be improved. E.g. stemming was not performed since Dutch parsers do not achieve high results on Dutch case law documents due to their complicated structure and use of language. The development of parsers specially made for the analysis of legal documents could lead to better results in future research.

The last issue that needs to be addressed is the appearance of the concept maps. A traditional concept map is composed of a number of propositions in which two concepts are linked by a linking phrase. Although this linking phrase normally contains a verb phrase, the linking phrases in this case contain a whole section of a case law. Moreover, these sections corresponding to the linking phrases were not shown to participants during the evaluation. This deprived them of potentially valuable information about the origin of the link between two concepts and could therefore result in different ranks opposed to when this information was available to them. Further research should perhaps integrate the document sections corresponding to a linking phrase to provide this knowledge.

References

1. Rechtspraak, de (2017). *Jaarverslag 2016*. Retrieved from http://www.jaarverslagrechtspraak.nl ([Online; accessed 05-30-2017]). In Dutch.
2. Winkels, R., Boer, A., Vredebregt, B. and Someren, A. van. Towards a legal recommender system. In *Legal Knowledge and Information Systems: JURIX 2014*, the Twenty-seventh Annual Conference, pages 169-178. IOS Press, 2014.
3. Castano, S., Ferrara, A., & Montanelli, S. (2013). Mining topic clouds from social data. In *Proceedings of the fifth international conference on management of emergent digital ecosystems* (pp. 108{112). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2536146.2536171 doi: 10.1145/2536146.2536171
4. Novak, J., & Cañes, A. (2008). *The theory underlying concept maps and how to construct and use them*. Florida Institute for Human and Machine Cognition.
5. Villalon, J. & Calvo, R. (2009). Concept extraction from student essays, towards concept map mining. *IIEE International Conference on Advanced Learning Technologies*. DOI: 10.1109/ICALT.2009.215
6. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3* (Jan), 993{1022.
7. Blei, D. M., & Lafferty, J. D. (2009). Visualizing topics with multi-word expressions. arXiv preprint arXiv:0907.1013 .
8. Wang, X., McCallum, A., & Wei, X. (2007). Topical n-grams: Phrase and topic discovery, with an application to information retrieval. , 697-702. Retrieved from http://dx.doi.org/10.1109/ICDM.2007.86 doi: 10.1109/ICDM.2007.86.