Legal Statutes Retrieval: A Comparative Approach on Performance of Title and Statutes Descriptive Text

Moemedi Lefoane, Tshepho Koboyatshwene, Goaletsa Rammidi, and V. Lakshmi Narasimham

University of Botswana, Gaborone, Botswana {moemedi.lefoane, tshepho.koboyatshwene, goaletsa.rammidi, lakshmi.narasimhan}@mopipi.ub.bw

Abstract. Legal Statutes play a crucial role in the Justice system. For countries that adopt the common law system they are often cited in court decisions to argue cases of interest. AILA 2019 track presented two tasks; precedents retrieval task and statutes retrieval task¹. Our team participated in the latter. The statutes provided consisted of two components namely; Title and Statute description. In this study we first conduct the experiment to determine the best term weighting model for this task. After determining the best term weighting model, the second set of experiments which aimed to determine the extent to which these components (title and description of statutes) contribute to retrieval effectiveness. To find out how retrieval effectiveness is affected by different components three experiments were conducted; the first involved indexing title and description of each statute as a document, retrieval using IFB2 is performed generating the first run (Baseline), the second experiment involved indexing only title disregarding description of the statutes, this generate the second run. For the final experiment, only description of statutes are indexed disregarding title, again indexing, retrieval performed to generate the third run. The three runs were then sent to organisers for evaluation. The evaluation results shows our team came second, furthermore results suggest that indexing with title only and disregarding description of statutes is sufficient enough for retrieval of statutes.

Keywords: Legal Statutes Retrieval \cdot Legal Text Mining \cdot Information Retrieval .

1 Introduction

Information retrieval(IR) is concerned with finding documents of unstructured text that are relevant to the information need from a collection of documents or from other material provided. Material or a document is relevant if it has information of value to satisfy the information need [7]. As indicated earlier

Copyright ©2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). FIRE 2019, 12-15 December 2019, Kolkata, India.

¹ https://sites.google.com/view/fire-2019-aila/track-description

2 M. Lefoane et al.

Artificial Intelligence for Legal Assistance (AILA 2019) track was divided into two tasks [1]. Our team participated in Legal Statutes Retrieval task, the goal was to generate a ranked list of relevant statutes for each object query provided in the dataset [1].

Experiments conducted by Tamrakar et al [5] on FIRE 2011 datasets using different probabilistic models in Terrier 3.5 such as BM25, BB2, IFB2, In_expB2 , In_expC2 , InL2, DFR_BM25 , DFI0 and PL2 yielded promising results. The datasets used consisted of various documents from newspapers and websites. MeanAveragePrecision(MAP) and R-precision were used for measuring the performance of the different models. The results indicated the highest MAP value of 0.7846 for the IFB2 model with the usage of a sample or few of news corpus dataset. IFB2 is one of the DFR models implemented in Terrier [4].

Another study conducted by Diana [6] used two variants from DFR models namely, PL2 and DLH13 for the CHiC 2013 Lab using a collection of textual cultural heritage objects based on the English and/or Italian languages. The best performance was obtained using DLH13 for the monolingual experiments with two of the collections which were made available.

Divergence from Randomness (DFR) is a probabilistic keyword indexing model, which was proposed by Amati et al [2] and was then incorporated in Terrier as one of the IR models. In DRF, a term weight is computed by measuring the divergence between a term distribution produced by a random process within the collection and the actual term distribution within a document. The assumption is that some words are not equally important when describing the content of the documents. Considering the entire document collection C, there is a random distribution of words (such as stop words) that carry little information or are deemed as less important across all documents. Another assumption is that there is an elite set of documents that contain speciality words or terms that are more informative following Poisson distribution [2].

The rest of the paper is organised as follows; Section 2 outlines our proposed approach detailing dataset description and experimental setup, 3 and 4 discuss results and conclusion respectively.

2 Methodology

We submitted 3 runs for this task; The first run formed the baseline. To generate the second and final run we relied on field base indexing model to index; first title only without description of the statutes and finally indexing only description without title of the statutes. The rest of this section provides more details on how the runs were generated. For all the three runs we used IFB2 retrieval model.

2.1 Data set Description

The dataset for this study consist of 50 object queries, of which the first 10 formed part of the training data. The remaining queries (11 - 50) formed part

of the Test data for which 3 runs were generated and submitted to Forum for Information Retrieval Evaluation (FIRE) for evaluation. For the training data, relevance assessments were provided and the document collection for training data consisted of statutes document collection of 197 states. The 197 statutes also formed document collection for the Testdata set².

2.2 Experimental Setup

The first part of the experiment was to address the question; which of the term weighting model performed best for retrieval of statutes, so the experiment was set for training data. In order to perform experiments the data set provided was transformed into TREC Style format, that is for both object queries as well as statutes documents. The parsed documents follows TREC format and shell scripting was used for parsing. Section 2.3 and Section 2.4 illustrate the object query as well as document/statute in TREC format.

We used Terrier 4.2^3 [3] to perform all our experiments for indexing, retrieval. For evaluation we used trec_eval 9.0^4 . The platform has been used successfully for ad-hoc retrieval tasks. Preprocessing performed for all experiments are; stemming using Potter's stemmer, stopwords were removed using Terrier stopword list. We then performed retrieval using different term weighting models as implemented in Terrier and the results are shown in Table 1. *MeanAveragePrecision* results revealed that Divergence from Randomness *IFB2* overall performance was better than the other models. We therefore chose *IFB2* for the next set of experiments to investigate retrieval effectiveness of each of the statutes component.

To generate the first run, we first separated the given queries (qurey 1 - 50) into training and test queries. Queries 1- 10 form training queries for our training data, and queries 11 - 50 form test queries. The first experiment was conducted to investigating different weighting models as implemented in terrier in order to find which one performs the best on the training data. We observe that the *IFB2* gives the best performance followed by *LemurTFIDF* and finally *InExpB2*. We therefore generate the first run (UBLTM1) using *IFB2*.

For the second and third run we transform into TREC style format but this time with two fields namely: Title and Description. We then index the statutes using title only and retrieve using test queries as well as IFB2 to generate our second run (UBLTM2). For the final run we indexed using description only and retrieve using IFB2 to generate final run (UBLTM3). The idea is so that we can investigate the effect of title and description only on the retrieval effectiveness of each of the two fields.

² https://sites.google.com/view/fire-2019-aila/dataset-evaluation-plan

³ http://terrier.org/docs/v4.2/

⁴ https://trec.nist.gov/trec_eval/

4 M. Lefoane et al.

2.3 Sample AILA Query transformed in to a format that can be used as query

Because the aim of the study was to compare the extent to which components of the statutes contributed to retrieval effectiveness, the statutes were transformed into two types of TREC Style document collection; one where the entire content of the statutes i.e. the title and description of the statutes was transformed as shown below:

Below is a sample of part of AILA_Q1 parsed into TREC TOPIC format: $<\!TOP\!>$

<NUM> AILA_Q1 </NUM>

<DESC>Description:

The appellant on February 9, 1961 was appointed as an Officer in Grade III in the respondent Bank (for short 'the Bank'). He was promoted on April 1, 1968 to the Grade officer in the Foreign Exchange Department in the Head Office of the Bank. Sometime in 1964,...[TEXT OMITTED]

</DESC> </NARR>Narrative: </TOP>

2.4 Sample Transformed Statute

Below is a sample of part of prior_case_0001 parsed into TREC DOCUMENT format:

<DOC> <DOCNO> S103</DOCNO> <TITLE> Freedom to manage religious affairs

... </TITLE> <TEXT>

Subject to public order, morality and health, every religious denomination or any section thereof shall have the right- (a) to establish and maintain institutions for religious and charitable purposes; (b) to manage its own affairs in matters of religion; (c) to own and acquire movable and immovable property; and (d) to administer such property in accordance with law.

3 Results

Table 1 shows results for experiments performed to determine the best term weighting model for Legal Statutes retrieval task. Results shows that IFB2 is the

best model in terms of MAP. Table 2 shows top 9 results for runs submitted to AILA 2019 organisers for evaluation. Our Team name in the table is UBLTM. In the table, P@10 refers to Precision@10, MAP refers to Mean Average Precision, BPREF refers to binary preference-based measure and $Recip_Rank$ refers to Reciprocal Rank.

Table 1: Mean Average Precision for performance of different measures for TrainingSet.

Query	TFIDF	LemurTFIDF	BM25	DFRBM25	PL2	IFB2	In Exp B2
AILA_Q1	0.0200	0.0206	0.0217	0.0215	0.0213	0.0382	0.0206
AILA_Q2	0.0601	0.0523	0.0608	0.0608	0.088	0.0624	0.0606
AILA_Q3	0.1367	0.3986	0.1624	0.1504	0.1564	0.5169	0.3919
AILA_Q4	0.0565	0.0864	0.0735	0.0718	0.0722	0.0517	0.0663
$AILA_Q5$	0.0205	0.0214	0.0226	0.0227	0.0223	0.0634	0.0211
AILA_Q6	0.0795	0.0886	0.0852	0.0878	0.0932	0.0615	0.0856
AILA_Q7	0.1693	0.3270	0.3059	0.3028	0.1628	0.2895	0.3030
AILA_Q8	0.0294	0.0538	0.0368	0.0365	0.0368	0.1007	0.0361
AILA_Q9	0.0282	0.0324	0.0369	0.0362	0.0355	0.0475	0.0306
AILA_Q10	0.3155	0.3130	0.3176	0.3384	0.3696	0.3871	0.3353
all Queries	0.0916	0.1394	0.1123	0.1129	0.1058	0.1619	0.1351

Table 2: Top 9 Results of the AILA runs for Task 2 – Statute Retrieval form AILA 2019 organisers.

Team name	P@10	MAP	BPREF	$Recip_Rank$
Yunqiu Shao thuir_le	gal 0.0975	0.1566	0.0961	0.281
Yunqiu Shao thuir_le	gal 0.0900	0.1318	0.0742	0.247
Yunqiu Shao thuir_le	gal 0.0650	0.1115	0.0653	0.230
UBLTM	0.0725	0.1023	0.0571	0.211
UBLTM	0.0725	0.1023	0.0571	0.211
UBLTM	0.0725	0.1022	0.0571	0.214
Sara Renjit - CUSAT_N	NLP 0.055	0.0967	0.0377	0.199
Soumil Mandal JU_SI	RM 0.06	0.0918	0.0402	0.201
Sara Renjit - CUSAT_N	NLP 0.055	0.0866	0.0412	0.202

4 Conclusion

Our experiments set out to investigate the extent to which different parts of statutes contribute to retrieval effectiveness, results reveal that titles of the 6 M. Lefoane et al.

statutes contain sufficient information to aid retrieval. For future work, the nature of statutes could be investigated further to understand characteristics better and inform the direction to take.

References

- P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, P. Mehta, A. Bhattacharya., P. Majumder, Overview of the Fire 2019 AILA track: Artificial Intelligence for Legal Assistance. In Proc. of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019.
- Gianni Amati and Cornelis Joost Van Rijsbergen. Proba- bilistic Models of Information Retrieval Based on Measuring the Divergence from Randomness. ACM Trans. Inf. Syst. 20, 4 (Oct. 2002), 357389. DOI:http://dx.doi.org/10.1145/582415.582416 Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. 2008. Introduction to Information Retrieval. Cambridge University Press, Cambridge, UK. http://nlp.stanford.edu/IR- book/information-retrieval-book.html
- 3. I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and C. Lioma. Terrier: A High Performance and Scalable In- formation Retrieval Platform. In Proceedings of ACM SIGIR06 Workshop on Open Source Information Retrieval (OSIR 2006).
- I. Ounis, G. Amati, Plachouras V., B. He, C. Macdonald, and Johnson. Terrier Information Retrieval Platform. In Pro- ceedings of the 27th European Conference on IR Research (ECIR 2005) (Lecture Notes in Computer Science), Vol. 3408. Springer, 517519 (2005).
- A. Tamrakar and S. K. Vishwakarma. Analysis of Proba- bilistic Model for Document Retrieval in Information Retrieval. In 2015 International Conference on Computational Intelli- gence and Communication Networks (CICN). 760765. (2015), DOI: http://dx.doi.org/10.1109/CICN.2015.155
- 6. D. Tanase. Using the divergence framework for randomness: CHiC 2013 lab report. CEUR Workshop Proceedings 1179 (2013).
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. 2008. Introduction to Information Retrieval. Cambridge University Press, Cambridge, UK. http://nlp.stanford.edu/IR- book/information-retrieval-book.html