ISM@FIRE-2013 Information Access in The Legal Domain

Task

Ambedkar Kanapala, Sukomal Pal

Dept. of CSE,

Indian School of Mines, Dhanbad, India.

(ambedkar.kanapala@live.com, sukomalpal@gmail.com)

Abstract

This paper describes the work that we did at Indian School of Mines, Dhanbad for FIRE 2013. This year we participated in one task: Adhoc retrieval from Legal Document(Consumer Law and Hindu Marriage & Diverse Law). Within the adhoc task, we participated using Indri search engine. We submitted a 1000 relevant documents for each query for both Consumer Law, Hindu Marriage &Diverse Law. We submitted one run(Consumer Law and Hindu Marriage & Diverse Law) in the adhoc task.

1 Introduction

At FIRE-2013 we have participated in one tasks. Within Adhoc retrieval from Legal Document, we have participated in both Consumer Law, Hindu Marriage & Diverse Law. We submitted a 1000 relevant documents for each query for both Consumer Law, Hindu Marriage &Diverse Law in the Adhoc Task. This task is done by using Indri search engine.

Our main objective is to return most releveant documents that can answer such queries.

2 Related work

Legal IR is currently divided between text-based boolean retrieval, knowledge engineering (KE) frameworks, epitomized by West's Key Numbers system, and inference network IR. KE ontologies successfully reflect how advocates recall and leverage legal precedent, but at a great expense of human labour. Blair and Maron clearly demonstrate that boolean techniques do not meet expected standards in legal retrieval: in their study, it was found that legal researchers retrieve less than 20% of relevant documents [2]. we try to retrieve most relevant document using indri search engine which is based on the combination of language[8] and inference network[9] models.

3 Our Participation

We participated in one task

1. Adhoc retrieval from Legal Document

(a)Consumer Law

(b) Hindu Marriage & Divorce Law

3.1 Adhoc Retrieval Task

Adhoc Task is been participated in two Domains. The approach for both of these Domain is same.

3.1.1 Approach

The Adhoc Retrieval task is based on two steps. These are:-

1.Indexing.

2.Retrieval.

3.1.1.1 Indexing

The indexing task approach is same for both Consumer Law and Hindu Marriage & Diverse Law.

During indexing documents are prepared for use by an IR system. This means preparing the raw document collection into an easily accessible representation of documents. This transformation from a document text into a representation of text is known as indexing the documents.

The indexing system builds compressed inverted lists for each term and field in memory. The data that is written to disk is self-contained: it contains all information necessary to perform queries on that data. In a sense, an Indri index can be considered a set of smaller indexes. The retrieval system has been written to be able to query many indexes together. A parameter file is needed for indexing the documents. The default smoothing method for Indri is Dirichlet smoothing with mu parameter set to 2500.

For Indexing, we used IndriBulidIndex command of Indri search engine.

3.1.1.2 Retrieval

We have used raw-tf based language modeling and inference network retrieval model.

3.1.2 Consumer Law

A parameter file needed to be constructed using combined operater. The retrival smoothing parameters is Dirichlet smoothing with mu=2500. For retrieval we used language modeling and inference network of Information Retrieval. A total of 20 quries in which 10 quries are from consumer,10 quries are from Hindu Marriage &Diverse Law. We have submited 1000 document for each query .

3.1.2.1 Data

Documents

Documents on which task has to be carried out were received from FIRE 2013 organisers in form

of a corpus in compressed format. The total text files are 385806 .The text file for Consumer Law is 216945 respectively. We also got a set of 10 queries . We run the consumer quries on conusumer court data(CC) and overall corpus(CO).

We retrieved one run after performing above mentioned steps. They are:-

RUN# CSISM_CC_INDRI,CSISM_CO_INDRI

3.1.3 Hindu Marriage &Diverse Law

For Hindu Marriage &Diverse Law, command used is IndriRunQuery of Indri Search engine. Firstly, a parameter file needed to be constructed using combined operator. The retrival smoothing parameters is Dirichlet smoothing with mu=2500. For retrieval we used language modeling and inference network of Information Retrieval.

3.1.3.1 Data

Documents

Documents on which task has to be carried out were received from FIRE 2013 organisers in form of a corpus in compressed format. The text file for Hindu Marriage &Diverse Law is 168861 respectively. We also got a set of 10 queries . We run the Hindu marriage quries on Hindu marriage data (HH) and overall corpus(HO).

We retrieved one run after performing above mentioned steps. They are:-

RUN# CSISM_HH_INDRI,CSISM_HO_INDRI

4 Conclusion

This paper describes our participation in FIRE-2013 Adhoc retrieval from Legal Document(Consumer Law and Hindu Marriage & Diverse Law). In adhoc retrieval task, we participated with a minimum set up based on open source tools. We have used Indri Search Engine on Consumer Law and Hindu Marriage & Diverse Law test collection. We are working on this which will hopefully improve its performance further in future.

5 References

[1]Cristopher D.Manning, Prabhakar Raghawan, Hinrich Schutze- An introduction to information retrieval, Cambridge University press 2008.

[2]K. Tamsin Maxwell and Burkhard Schafer. 2008. Concept and Context in Legal Information Retrieval. In Proceedings of the 2008 conference on Legal Knowledge and Information Systems: JURIX 2008: The Twenty-First Annual Conference, Enrico Francesconi, Giovanni Sartor, and

Daniela Tiscornia (Eds.). IOS Press, Amsterdam, The Netherlands, The Netherlands, 63-72.

[3]www.isical.ac.in/~fire/ (as on 20.11.2013)

[4] http://ciir.cs.umass.edu/~metzler/indriretmodel.html (as on 20.11.2013)

[5]http://en.wikipedia.org/wiki/Information_retrieval (as on 20.11.2013)

[6]www.lemurproject.org. (as on 20.11.2013)

[7]http://sourceforge.net/p/lemur/wiki/Quick%20Start/ (as on 20.11.2013)

[8]Ponte, J. M. and Croft, W. B., "A language modeling approach to information retrieval," Proceedings of the 21st Annual international ACM SIGIR Conference on Research and Development in information Retrieval (SIGIR '98), 275-281, 1998.

[9]Turtle, H. and Croft, W.B.,"Evaluation of an Inference Network-Based Retrieval Model," ACM Transactions on Information System, in 9(3),187-222, 1991.