

Overview of Information Access in Legal domain @ FIRE 2013

Parth Mehta¹, Madhulika Agrawal¹, Kripabandhu Ghosh² and
Abhik Majumdar³

DA-IICT Gandhinagar¹
ISI Kolkata²
NLUO Odisha³

December 4, 2013

Motivation

Motivation

- Initiate research in Legal informatics in India

Motivation

- Initiate research in Legal informatics in India
- Supplement the project by providing benchmark dataset and relevance judgements

Motivation

- Initiate research in Legal informatics in India
- Supplement the project by providing benchmark dataset and relevance judgements
- Final goal is to make an automated system that can provide guidance in legal matters to the end user.

Motivation

- Initiate research in Legal informatics in India
- Supplement the project by providing benchmark dataset and relevance judgements

- Final goal is to “attempt” to make an automated system that can provide guidance in legal matters to the end user.

Task Description

Task Description

1 Adhoc retrieval from legal documents

Task Description

- 1 Adhoc retrieval from legal documents
- 2 (a) Identification of propositions from legal documents

Task Description

- 1 Adhoc retrieval from legal documents
- 2 (a) Identification of propositions from legal documents
(b) Classification of propositions into one of the nine categories.

Corpus Details

- **Adhoc Retrieval**

High court and Supreme court verdicts: 180,000

Consumer court verdicts: 170,000

20 topics (10+10) taken from various online forums,
representing the actual queries related to *Consumer Rights*
and Hindu Marriage Act

- **Classification and Identification of Propositions**

10 manually segmented and annotated documents

900 unannotated documents

Evaluation criteria

- ▶ **Task 1:** Mean Average Precision

- ▶ **Task 2(a):** $A_{ik} = \max_j \frac{\{P_{ik} \cap Q_{jk}\}}{\{P_{ik} \cup Q_{jk}\}}$

P_{ik} : is the proposition number i of k^{th} original text

Q_j : is the proposition number j of k^{th} text in submitted run

$$A_{overall} = \frac{\sum_{k=1}^N \sum_{i=1}^{M_k} A_{ik}}{\sum_{k=1}^N M_k}$$

M_k is number of propositions in in K_{th} document.

- ▶ **Task 2(b):** Macro and Micro Precision

Participation Overview

Team	Task Participated	Runs Submitted
EVORA	Ad-hoc Retrieval	2
ISM	Ad-hoc Retrieval	1
TRDDC	Proposition Classification	1

Results

Adhoc Retrieval

Team	Runs Number	Mean Average Precision	
		Focused Corpus	Overall Corpus
EVORA	Run 1	0.1627	0.1489
EVORA	Run 2	0.2186	0.2159
ISM	Run 1	0.1995	0.1413

Results

Proposition Classification Participating team: TRDDC

Category	Total	Correct	Precision
Intrinsic Facts	60	51	0.85
Extrinsic Facts	18	0	0
Issue	2	0	0
Arguments	40	0	0
Ruling by lower court	12	1	0.08
Statute	38	9	0.24
Precedent	40	4	0.1
Other general standards	6	0	0
Ruling by present court	174	139	0.79

Macro Precision: 0.52

Micro Precision: 0.23

There will be a next time

Possible problems

- Focused passage level retrieval
- Classification of propositions depending on their positive/negative impact on the decision

Questions ??