Evaluation of Query Formulations in the Negotiated Query Refinement Process of Legal e-Discovery
UMKC at TREC 2007 Legal Track

F.C. Zhao   Y. Lee   D. Medhi

School of Computing and Engineering
University of Missouri - Kansas City

Nov.6 2006 Gaithersburg, MD

Problem

Negotiated Query Refinement Process

- RequestText
- ProposalByDefendant
- RejoinderByPlaintiff
- FinalQuery

Questions

- Will negotiations definitely improve recall?
- If not, how to improve?

Measures & Methods

Measures

- MAP
- est_R
- est_P

Methods

- Different query formulations in VSM & LM
- Evaluating relative performance

System & Runs

- A modified Lucene (VSM & LM)
- Indexed only the OCR text
- Submitted six runs

Runs	Query Source	Retrieval Model
$UMKC_1$	ProposalByDefendant	LM
$UMKC_2$	RequestText	LM
$UMKC_3$	FinalQuery	LM
$UMKC_4$	ProposalByDefendant	VSM
$UMKC_5$	RequestText	VSM
$UMKC_6$	FinalQuery	VSM

Queries

[RequestText] {Please produce any and all documents concerning soil water management as it pertains to commercial irrigation. }

[Final Query:]
irrig (0.3084472), soil (0.25898176), water (0.2516427), pertain (0.20087002), commerci (0.1465618), tobacco (0.08702624), cigarett (0.03782016), plant (0.037246022), product (0.031287868), smoke (0.029167147) ...

Results

Measures	UMKC$_1$	UMKC$_2$	UMKC$_3$	UMKC$_4$	UMKC$_5$	UMKC$_6$
MAP	0.094098	0.090642	0.084286	0.102965	0.098791	0.105095
est_RB	0.13513	0.100381	0.106542	0.157151	0.125995	0.137151

T-TEST

Measures	(1,2)	(2,3)	(1,3)	(4,5)
MAP Diff %	0.036726	0.07012	0.104271	0.040542
MAP P-value	0.680553	0.462849	0.36522	0.670065
est_RB P-value	*0.015904*	0.728471	0.081248	0.055975

(5,6)	(4,6)	(1,4)	(2,5)	(3,6)
-0.06382	-0.02069	-0.08612	-0.08249	-0.198
0.604629	0.865816	0.389588	0.380572	*0.015177*
0.563488	0.307274	0.201638	*0.029731*	*0.032912*

Findings

Findings

No statistical significant improvement at the 0.05 level among different runs regardless of the query sources.

Why?

- special purposed Boolean query refinement
- enriching the query with synonym-like terms
- relaxing Boolean constraints

Findings

Findings

No statistical significant improvement at the 0.05 level among different runs regardless of the query sources.

Why?

- special purposed Boolean query refinement
- enriching the query with synonym-like terms
- relaxing Boolean constraints

Future Work for Negotiation

- Discard the assumption of "standard Boolean retrieval system"
- Identify the overall query intention
- Locate a solid list of core concepts
- Discover new facets