

Run Descriptions: TREC 2008 Legal Track, Relevance Feedback Task

Run IowaSL08RF1A (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields:

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 2

This feedback run corresponds to baseline run IowaSL08RF1B.

Description of this run:

A collection of features is calculated for all documents submitted in the source runs for IowaSL08RF1B, judged and unjudged. The features describe the position of the document in the contributing runs, as well as the relevance characteristics of its “neighborhood”. We train a classifier using the calculated features of the judged documents. The classifier, implemented using Weka’s SMO package, fits a logit regression to generate probabilities for relevance and nonrelevance for each of the test documents. It is then applied to the unjudged, which are then ranked by estimated probability of relevance.

Run IowaSL08RF1B (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields:

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 6

This run is a baseline (i.e., no relevance feedback)

Description of this run:

Each submitted run for the previous ad hoc tasks were pooled to build a collection of judged and unjudged documents. The unjudged documents are merged, ranked using a score reflecting a summation of the ranks across the runs and a multiplier reflecting the number of runs to select the document (CombMNZ). For the 2007 topics, the 101000 documents with the best scores are selected.

Run IowaSL08RF2A (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields:

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 3

This feedback run corresponds to baseline run IowaSL08RF2B.

Description of this run:

This run is processed in the same manner IowaSL08RF1A. The documents pooled for IowaSL08RF2B are

split into training and evaluation pools in the same manner as described for IowaSL08RF1A.

Run IowaSL08RF2B (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields:

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 7

This run is a baseline (i.e., no relevance feedback)

Description of this run:

This run is processed in the same manner IowaSL08RF1B, the difference laying in the runs contributing to the pool of judged and unjudged documents. The seven runs submitted by the IowaS group for the TREC Legal 2007 Ad Hoc task are used for the subset of 2007 topics. Results generated Post TREC 2007 by applying 2007 strategies against the 2006 topics are used as sources for the 2006 topics.

Run IowaSL08RF2C (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields:

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 5

This feedback run corresponds to baseline run IowaSL08RF2B.

Description of this run:

This run is also processed in the same manner IowaSL08RF2A. The documents pooled for IowaSL08RF2B are split into training and evaluation pools in the same manner as described for IowaSL08RF1A. For this run, the training pool is discarded, and the classifiers built using all submitted runs, for run IowaSL08RF1A, are used for classifying and ranking the evaluation pool.

Run IowaSL08RF3A (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields:

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 4

This feedback run corresponds to baseline run IowaSL08RF3B.

Description of this run:

This run is also processed in the same manner IowaSL08RF1A. The documents pooled for IowaSL08RF3B are split into training and evaluation pools in the same manner as described for IowaSL08RF1A.

Run IowaSL08RF3B (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields:

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 8

This run is a baseline (i.e., no relevance feedback)

Description of this run:

This run is processed in the same manner IowaSL08RF1B, the difference laying in the runs contributing to the pool of judged and unjudged documents. For 2006 topics, four runs with leading performance on the R-Prec measure across all topics were selected, and four runs with poorer performance were selected. A similar technique was used to identify contributing runs for the 2007 topics, using the Est_Recall@B measure.

Run IowaSL08RFTr (University of Iowa (Srinivasan))

Task: feedback

Run type: automatic

Fields: request text, initial proposal, rejoinder, final negotiated Boolean query

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 1

This feedback run does not have a corresponding baseline run.

Description of this run:

This run is run in the same manner as our ad hoc run IowaSL0804 with the following modifications to take advantage of relevance feedback. Documents that have not been judged to be relevant are deemed ineligible for providing terms for query expansion. Training has allowed us to expand the number of query expansion terms to 12 from up to 10 relevant documents. The boolean reference run boost has not been applied. K and Kh values are from R and R/3 respectively.

Run UMKCTL08RF1 (University of Missouri-Kansas City)

Task: feedback

Run type: manual

Description of manual approach:

No

Fields: final negotiated Boolean query

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 1

This feedback run does not have a corresponding baseline run.

Description of this run:

TL08RF VSM Expand15 Ratio1 SPDL719-719_0.8 101k

Run UMKCTL08RF2 (University of Missouri-Kansas City)

Task: feedback

Run type: manual

Description of manual approach:

No

Fields: final negotiated Boolean query

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 2

This feedback run does not have a corresponding baseline run.

Description of this run:

TL08RF BM25 Expand15 Ratio1 s0.9k3 101k K1000

Run UMKCTL08RF3 (University of Missouri-Kansas City)

Task: feedback

Run type: manual

Description of manual approach:

No

Fields: final negotiated Boolean query
Document fields indexed: OCR
Boolean reference run used? no
Judging order: 3
This feedback run does not have a corresponding baseline run.
Description of this run:
UMKC TL08RF LM_0.2-1.0 Expand15 Ratio1 101k K1000

Run UMKCTL08RF5 (University of Missouri-Kansas City)
Task: feedback
Run type: manual
Description of manual approach:
No
Fields: final negotiated Boolean query
Document fields indexed: OCR
Boolean reference run used? yes
Judging order: 5
This feedback run does not have a corresponding baseline run.
Description of this run:
CombMNZ Rank Fusion Normalized Score refRF08B.append

Run UMKCTL08RF6 (University of Missouri-Kansas City)
Task: feedback
Run type: manual
Description of manual approach:
No
Fields: final negotiated Boolean query
Document fields indexed: OCR
Boolean reference run used? yes
Judging order: 6
This feedback run does not have a corresponding baseline run.
Description of this run:
CombMNZ Rank Fusion Normalized Score with refRF08B refRF08B.append

Run UCBM25T10Th5 (Ursinus College)
Task: feedback
Run type: automatic
Fields: request text
Document fields indexed: OCR
Boolean reference run used? no
Judging order: 2
This feedback run corresponds to baseline run UCRFBM25BL.
Description of this run:
BM25 with 10 terms from judged rel docs added to each query (as long as the term weight was at least 5)

Run UCBM25T5Th5 (Ursinus College)
Task: feedback
Run type: automatic
Fields: request text
Document fields indexed: OCR
Boolean reference run used? no

Judging order: 1

This feedback run corresponds to baseline run UCRFBM25BL.

Description of this run:

BM25 with automatic rel feedback using top 5 terms from all relevant docs (with term weight i 5)

Run UCPwrT10Th5 (Ursinus College)

Task: feedback

Run type: automatic

Fields: request text

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 4

This feedback run corresponds to baseline run UCRFPwrBL.

Description of this run:

Power Norm with auto rel feedback, the top 10 terms from any document judged relevant.

Run UCPwrT5Th5 (Ursinus College)

Task: feedback

Run type: automatic

Fields: request text

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 3

This feedback run corresponds to baseline run UCRFPwrBL.

Description of this run:

Power Norm with 5 terms added to each query. Top 5 weighted terms from any relevant document.

Run UCRFBM25BL (Ursinus College)

Task: feedback

Run type: automatic

Fields: request text

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 5

This run is a baseline (i.e., no relevance feedback)

Description of this run:

BM25 baseline for comparison with BM25T5Th5 and BM25T10Th5

Run UCRFPwrBL (Ursinus College)

Task: feedback

Run type: automatic

Fields: request text

Document fields indexed: OCR

Boolean reference run used? no

Judging order: 6

This run is a baseline (i.e., no relevance feedback)

Description of this run:

Power Norm baseline for UCPwrT5Th5 and UCPwrT10Th5

Run otRF08F (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields:

Document fields indexed: OCR, metadata

Boolean reference run used? no

Judging order: 2

This feedback run corresponds to baseline run otRF08fv.

Description of this run:

pure relevance feedback run based on forming a query from the known relevant documents (excluding those larger than 1 million bytes); no topic fields were used; K includes rsv of 200 or more, Kh includes rsv of 225 or more

Run otRF08FR (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields:

Document fields indexed: OCR, metadata

Boolean reference run used? no

Judging order: 7

This feedback run corresponds to baseline run otRF08fv.

Description of this run:

same as otRF08F except that K is based on estimated R from humL06tvz sample for first 15 topics and L07 est_rel for latter 25 topics; Kh same as K

Run otRF08fb (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields: final negotiated Boolean query, B value

Document fields indexed: OCR, metadata

Boolean reference run used? yes

Judging order: 1

This run is a baseline (i.e., no relevance feedback)

Description of this run:

Boolean run based on final negotiated query (same as reference Boolean run refRF08B except re-ordered by relevance)

Run otRF08fbF (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields: final negotiated Boolean query

Document fields indexed: OCR, metadata

Boolean reference run used? yes

Judging order: 3

This feedback run corresponds to baseline run otRF08fb.

Description of this run:
ranked-based fusion of Boolean (otRF08fb) and feedback (otRF08F) runs, equal weights; K and Kh are the same as for otRF08F

Run otRF08fbFR (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields: final negotiated Boolean query

Document fields indexed: OCR, metadata

Boolean reference run used? yes

Judging order: 8

This feedback run corresponds to baseline run otRF08fb.

Description of this run:

same documents as otRF08fbF and same K and Kh as otRF08FR

Run otRF08frw (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields: request text, final negotiated Boolean query

Document fields indexed: OCR, metadata

Boolean reference run used? yes

Judging order: 6

This run is a baseline (i.e., no relevance feedback)

Description of this run:

weighted rank-based fusion of Boolean and vector runs (same approach as otL08frw, i.e. weight 3 on otRF08fb, weight 2 on otRF08rvl, weight 1 on otRF08fv; K includes weight sum of 3 or more, Kh includes weight sum of 4 or more)

Run otRF08fv (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields: final negotiated Boolean query

Document fields indexed: OCR, metadata

Boolean reference run used? no

Judging order: 5

This run is a baseline (i.e., no relevance feedback)

Description of this run:

vector run based on final negotiated query terms (same approach as otL08fv)

Run otRF08rvl (Open Text Corporation)

Task: feedback

Run type: manual

Description of manual approach:

0

Fields: request text

Document fields indexed: OCR, metadata
Boolean reference run used? no
Judging order: 4
This run is a baseline (i.e., no relevance feedback)
Description of this run:
vector run based on request text terms (same approach as otL08rv1)

Run SabL08rf1 (Sabir Research, Inc)
Task: feedback
Run type: automatic
Fields: request text, initial proposal, rejoinder, other negotiation history, final negotiated Boolean query
Document fields indexed: OCR, metadata
Boolean reference run used? no
Judging order: 1
This feedback run corresponds to baseline run SabL08rfbase.
Description of this run:
Basic rocchio feedback, all judged docs, add 40terms. SMART ltu_Lnu vector run. default constant K. Used all boolean query negotiation terms

Run SabL08rfbase (Sabir Research, Inc)
Task: feedback
Run type: automatic
Fields: request text, initial proposal, rejoinder, other negotiation history, final negotiated Boolean query
Document fields indexed: OCR, metadata
Boolean reference run used? no
Judging order: 2
This run is a baseline (i.e., no relevance feedback)
Description of this run:
Baseline for Rocchiofeedback, SMART ltu_Lnu vector run. default constant K. Used all boolean query negotiation terms