Something Interesting in TREC 2007

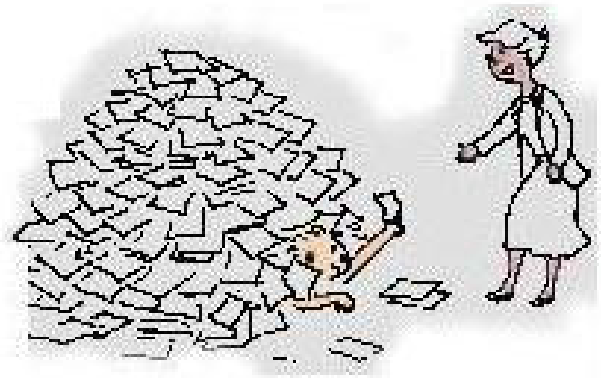
Yiqun Liu

07.11.6–07.11.9

Gaithersberg, U.S.

Outline

- Overview of TREC 2007
- Brief Introduction to Some Tracks
- Interesting Techniques in Enterprise Track



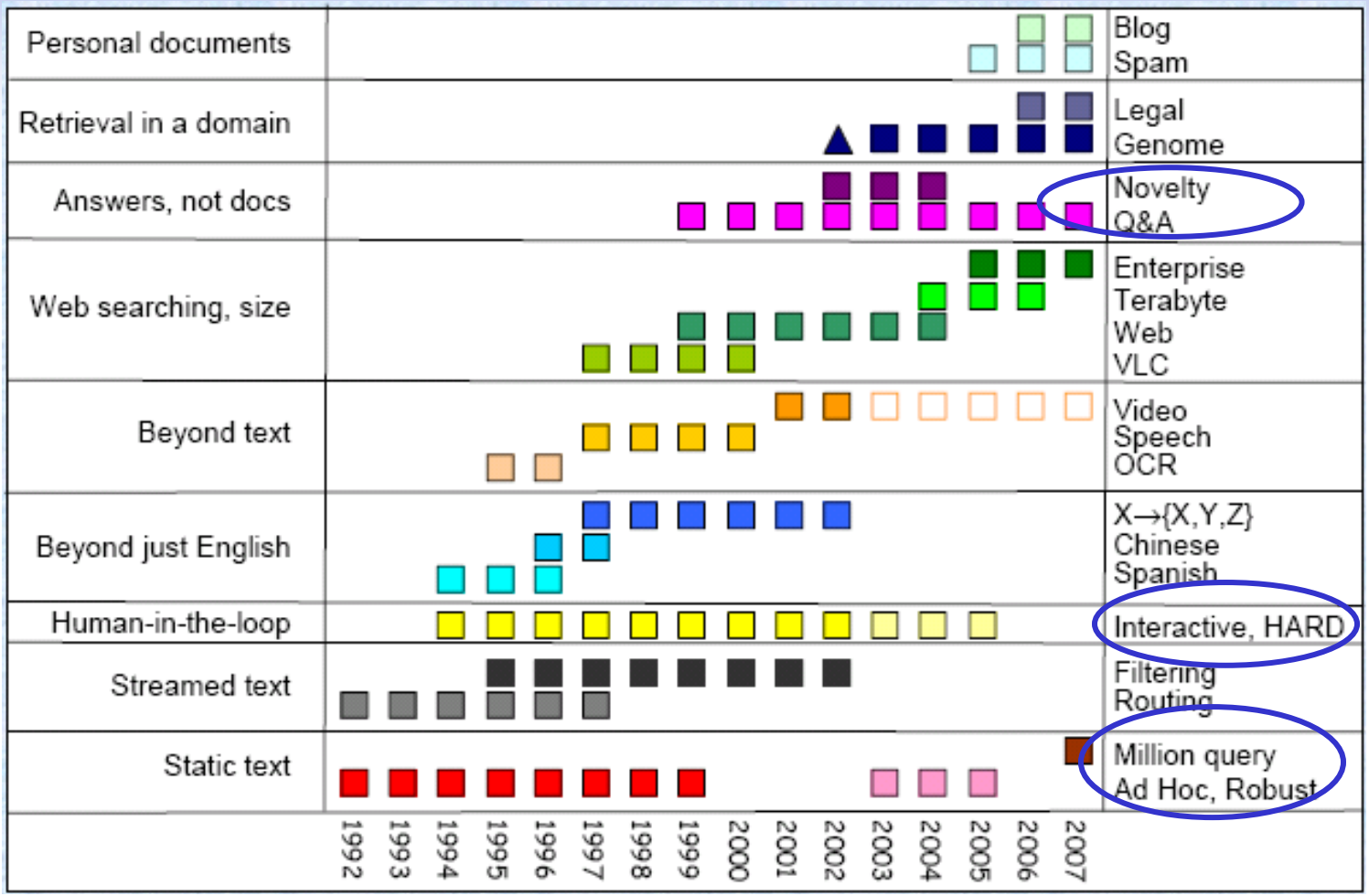
Overview of TREC 2007

- Co-sponsored by NIST and IARPA
  - IARPA: Intelligence Advanced Research Projects Activity, under the Director of National Intelligence's responsibility.
- 95 participating groups from 18 countries
- Seven tracks
  - Six old tracks: QA, Genomics, Enterprise, Spam, Legal, Blog
  - One new track: Million Query

Overview of TREC 2007

- Nine Groups from Mainland China
  - Beijing U. of Posts & Telecommunications
  - Chinese Academy of Sciences
  - DaLian University of Technology
  - Fudan University
  - Peking University
  - Shanghai Jiao Tong University
  - South China University of Technology
  - Tsinghua University
  - Wuhan University

Overview of TREC 2007



Overview of TREC 2007

Track	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'06
Ad Hoc	18	24	26	23	28	31	42	41								
Routing	16	25	25	15	16	21										
Interactive			3	11	2	9	8	7	6	6	6					
Spanish			4	10	7											
Confusion				4	5											
Merging				3	3											
Filtering				4	7	10	12	14	15	19	21					
Chinese					9	12										
NLP					4	2										
Speech						13	10	10	3							
XLingual						13	9	13	16	10	9					
High Prec						5	4									
VLC							7	6								
Query							2	5	6							
QA								20	28	36	34	33	28	33	31	28
Web								17	23	30	23	27	18			
Video										12	19					
Novelty											13	14	14			
Genomics												29	33	41	30	25
HARD												14	16	16		
Robust												16	14	17		
Terabyte													17	19	21	
Enterprise														23	25	20
Spam														13	9	12
Legal															6	14
Blog															16	24
Million Q																11
Participants	22	31	33	36	38	51	56	66	69	87	93	93	103	117	107	95

Brief Introduction to Some Tracks

- Blog Track
  - Collection: 148GB: 88.8GB permalink documents, 38.6GB is feeds, 28.8GB homepages. 3.2 Million.
  - Opinion task (20 participants)
    - locate blog posts that expressed an opinion about a given target
    - Relevance judgment: irrelevant (not on-topic), relevant but not opinionated (adhoc relevant), relevant with opinion
    - Evaluation: MAP, R-prec, bPref, P@10

Brief Introduction to Some Tracks

- Blog Track
  - Polarity subtask (11 participants)
    - Predict judgment: positive, negative, mixed
    - Evaluation problem: not the same document set
    - R-accuracy: the fraction of retrieved documents above rank R that are classified correctly
    - A@10 and A@1000
    - TREC 2006 data can be used for training

Brief Introduction to Some Tracks

- Blog Track
  - Blog distillation (9 participants)
    - Provide the users with the key blogs about a certain topic X (which they may add to their RSS reader)
    - Focused on feed retrieval
    - Participants should also make relevance judgment
    - Evaluation: MAP, R-prec, bpref, P@10

Brief Introduction to Some Tracks

- Legal Track
  - E-discovery: effective discovery of digital documents
  - Boolean expression => documents => examined by humans to determine which are responsive to a discovery request.
  - Changes in the U.S. Federal Rules of Civil Procedure (2006/12/01)
    - A new category of evidence, namely, “Electronically Stored Information” (ESI) in “any medium,”

Brief Introduction to Some Tracks

- Legal Track
  - IIT collection: approximately 7 million documents drawn from the Legacy Tobacco Document Library hosted by UCSF.
  - Contains OCR errors
  - An emphasis on recall over precision
  - Purpose: Whether IR systems can outperform current Boolean search systems?

Brief Introduction to Some Tracks

- Legal Track
  - Main task: ad hoc search task
    - Topics (production requests) are developed by lawyers

```
- <ProductionRequest>
    <RequestNumber>52</RequestNumber>
    <RequestText>Please produce any and all documents that discuss the use or introduction of high-phosphate fertilizers
      (HPF) for the specific purpose of boosting crop yield in commercial agriculture.</RequestText>
  - <BooleanQuery>
      <FinalQuery>(("high-phosphat! fertiliz!" OR hpf) OR ((phosphat! OR phosphorus) w/15 (fertiliz! OR soil))) AND (boost!
        OR increas! OR rais! OR augment! OR affect! OR effect! OR multipl! OR doubl! OR tripl! OR high! OR greater) AND
        (yield! OR output OR produc! OR crop OR crops)</FinalQuery>
    - <NegotiationHistory>
        <ProposalByDefendant>"high-phosphate fertilizer!" AND (boost! w/5 "crop yield") AND (commercial w/5 agricultur!)
          </ProposalByDefendant>
        <RejoinderByPlaintiff>(phosphat! OR hpf OR phosphorus OR fertiliz!) AND (yield! OR output OR produc! OR crop OR
          crops)</RejoinderByPlaintiff>
      </NegotiationHistory>
    </BooleanQuery>
    <FinalB>3078</FinalB>
    <RequestSource>2007-A-1</RequestSource>
  </ProductionRequest>
```

Brief Introduction to Some Tracks

- Legal track
  - Resources that can be used
    - the negotiated Boolean query
    - the set of documents that matched the Boolean query
    - the size of the retrieved set of the Boolean query (B)
  - Results:
    - 25000 documents or B documents (if B>25000)
  - Evaluation
    - By law students
    - Recall@B: unranked measure

Brief Introduction to Some Tracks

- Million-query Track
  - Hypothesis:
    - A test collections built using many topics with few, shallow judgments is a better evaluation tool than a test collection built from fewer topics with relatively thorough judgments.
  - Ben Carterette, James Allan, and Ramesh Sitaraman. SIGIR 2006
  - Collection:
    - .GOV2 collection, 25 million docs

Brief Introduction to Some Tracks

- Million Query Track
  - Topics:
    - 10k queries (some from previous Terabyte Tracks)
    - Only title fields
  - Relevance judgment
    1. The assessment system presented the judge with 5 queries randomly selected from test set.
    2. The judge selected one of the queries; the others were returned to the query pool.
    3. The judge wrote a description and narrative for this query, creating a standard TREC topic statement for it.
    4. The system presented a GOV2 document to the judge and obtained a 3-way judgment (highly relevant, relevant, not relevant) for it.
    5. The process continued until at least 40 documents were judged. The judge could continue past 40 documents if he or she wanted to.

Brief Introduction to Some Tracks

- Million Query Track
  - Document selection
    - Expected AP method
      - Quantify how much a document will contribute to AP calculation if it turns out to be relevant or non-relevant
    - Statistical evaluation method
      - Random sample of documents from the given ranked lists and produces unbiased, low-variance estimates of metrics
    - ½ queries: 20 documents selected with both methods
    - ¼ queries: 40 documents selected with one method
    - 11 groups participated, 1800 queries are judged

Brief Introduction to Some Tracks

- Million Query Track
  - Kendall's tau rank correlation
    - between UMass and NEU : 0.860
    - between UMass and Terabyte : 0.778
    - between NEU and Terabyte : 0.831
  - None of the three agreed about the best system
    - indriDM: 1st by Terabyte, 5th by UMass, 6th by NEU
    - JuruSynE: 1st by UMass, 3rd by Terabyte, 4th by NEU
    - LucSyn0: 1st by NEU, 4th by UMass, 5th by Terabyte

Interesting Tech. in Enterprise Track

- Overview
  - CSIRO collection
    - 0.37 million document, 4 gigabytes
  - Tasks
    - Meet the needs of “science communicators”
    - find a set of key pages and people in that area as a first step in creating an overview page
    - Document search task (16 groups)
    - Expert search task (16 groups)

Interesting Tech. in Enterprise Track

- Document search task
  - Not effective
    - relevance feedback (using example documents)
  - Effective
    - page authority measures, PageRank, anchor text, VIPS, fielded weighting
  - Relevance judgment
    - 59 key page per topic, and 66 maybe-key (maybe too many) (Possible rejudging by SCs)

Interesting Tech. in Enterprise Track

- Expert search task
  - Supporting documents returned, but not used for judgement
  - Effective:
    - Expert identification and mapping to email address
    - Identifying homepages of candidates experts
    - Down-weighting complicated pages
    - Balancing sources of evidence.
  - False alarms:
    - Examine 37 non-experts (5 other science communicators, 2 were knowledgeable, 4 were clearly not knowledgeable, 28 were not known to be knowledgeable or not)

Interesting Tech. in Enterprise Track

- Open University
  - Two stage model
    - Document relevance and co-occurrence
  - Integrate multiple document features
    - Anchor texts
    - URL length (useful)
    - in-links/PageRank (useful) / out-links (not useful)
    - multiple levels of association
      Vechtomova et al 2003
      (http://ovecht2.uwaterloo.ca/publications1.html)

Interesting Tech. in Enterprise Track

- Open University
  - Query Operations
    - remove stop words
    - narrative and title parts are both used, with different weights
    - manual modification to the narrative part
  - Results:
    - Automatic use of narrative field decrease MAP value

Interesting Tech. in Enterprise Track

- Shanghai Jiaotong University
  - Document search
    - Static Ranking Approaches
    - Sparse links
      - Pagerank, Topic-sensitive PageRank
      - HostRank (calculating each host's importance, applying PageRank algorithm on host graph)
    - Data preprocessing
      - Title extraction (extract title for the pages which do not have titles)
      - Body Detection

Interesting Tech. in Enterprise Track

- Shanghai Jiaotong University
  - what works:
    - hostrank
    - title extraction
    - body detection
    - position weighting
  - what doesn't work
    - pagerank / topic sensitive pagerank
    - Query Expansion

Interesting Tech. in Enterprise Track

- Shanghai Jiaotong University
  - ExpertRank & Topic sensitive ExpertRank
  - Data Preprocessing
    - Deal with pages that are too simple or too complicated
    - Homepage detection
  - What works:
    - Homepage Detection
  - What don't work:
    - ExpertRank & Topic sensitive ExpertRank

Interesting Tech. in Enterprise Track

- University of Glasgow
  - Document search
    - Combination from various document priors (inlinks, URL length, Peng et al, RIAO 2007)
    - Click distance from feed back documents
      - Document adjacent to known relevant documents are likely to be relevant
      - A breadth-first search for the shortest path in the hyper link graph
    - Combination of Click-distance Evidence with retrieval score (Craswell, sigir 2005)

Interesting Tech. in Enterprise Track

- University of Glasgow
  - Expert search
    - voting model
      - A voting from a ranking of relevant document to the candidates
      - Each document retrieved which is associated with a candidate is a vote for that candidate
      - different voting techniques (expCombMNZ...)
    - QE : Candidate Centric QE (ECIR 2007)

Interesting Tech. in Enterprise Track

- Next year's proposal
  - Sharing extracted email addresses?
  - Repeat the same tasks on CERC
  - Improve judgment
    - make science communicators to judge
    - use million-query track techniques

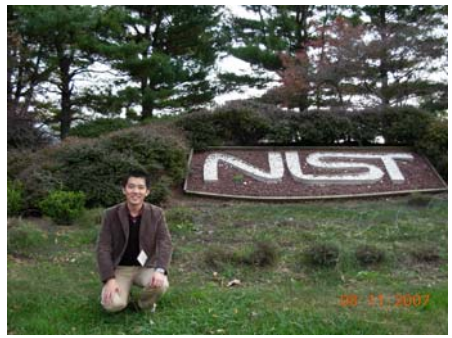
Feelings and the Future

- What's coming in TREC 2008
  - Sunset for genomics and spam
  - Move QA to TAC (expanded of what was DUC)
  - Blog, enterprise, legal, million query continue
  - New: relevance feedback track
    - goal: create evaluation methodology/data that will allow separating effects of different variables so as to improve effectiveness. use Q0 field in TREC submission...
  - New: track on utility of tags under development

Feelings and the Future

- Several feelings:
  - TREC is official

  - TREC people are nice and friendly



  - TREC is a benchmark and a startpoint

Thank You !