

# TREC 2006 Genomics Track Overview

William Hersh<sup>1</sup>  
Aaron M. Cohen<sup>1</sup>  
Phoebe Roberts<sup>2</sup>  
Hari Krishna Rekapalli<sup>1</sup>

<sup>1</sup>Department of Medical Informatics & Clinical Epidemiology  
Oregon Health & Science University

<sup>2</sup>Biogen Idec Corp.  
Boston, MA, USA



1

# Acknowledgements

- Data development, management, and scoring – Aaron Cohen, Hari Krishna Rekapalli, Jianji Yang
- Relevance judgment process – Phoebe Roberts
- Data providers
  - Highwire Press
  - National Library of Medicine
- Funder
  - National Science Foundation Grant ITR-0325160
- NIST – Ellen Voorhees, Ian Soboroff, Lori Buckland
- Track steering committee



2

# Overview

- Motivation
- Task
- Evaluation measures
- Document collection
- Topics
- Relevance judgments
- Results
- Future Directions



3

# Motivation

- Information retrieval (IR) systems aim to retrieve documents that a user might find relevant to his or her information need
- In contrast, information extraction (IE) or text mining (TM) systems aim to process document text to provide the user with one or more specific “answers” to a question or information need
- What many information seekers, especially users of the biomedical literature, may desire is something in the middle, i.e., a system that attempts to provide short, specific answers to questions and puts them in context while providing supporting information and linking to original sources



4

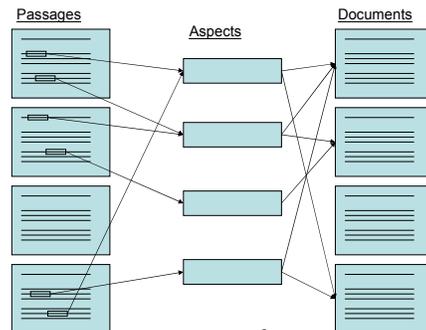
# Task (or use case)

- A scientist searching the biomedical literature, aided by a system that provides
  - Retrieval of passages – portions of text that contain an answer to the question
  - Grouped by aspects – to show which passages provide similar information; a complete answer may require several different aspects
  - Linked to documents – that the user ultimately wants to retrieve
- Used mean average precision (MAP)-like measures for these three levels of retrieval



5

# From passages to aspects to documents



6

## Evaluation measures

- Based on passage spans submitted in runs
  - Passages could not cross `<P...>` or `</P...>` boundaries
- Passage-level MAP
  - Character-based, modeled on measure used for TREC 2003-2004 HARD Track (Allan, 2004)
- Aspect-level MAP
  - Similar to approach used in TREC 6-9 Interactive Track (Hersh, 2001)
- Document-level MAP
  - “Rolled up” from passages to document level

7



## Document collection

- Full text HTML journal articles that preserved formatting, structure, table and figure legends, etc.
- 162,259 documents from 49 journals published by Highwire Press ([www.highwire.org](http://www.highwire.org))
  - Documents contained 12,641,127 maximum-length legal spans, which consisted of all text delimited by HTML paragraph tags and were listed in file [legalspans.txt](#)
- Corresponding MEDLINE record provided by National Library of Medicine (NLM), including Medical Subject Headings (MeSH) terms
  - Full-text file name was PMID provided by Highwire, with link to actual article in file [metadata.txt](#)

8



## Some issues with document collection

- Ugly HTML – generated automatically from SGML by Highwire Press, though renders well in a browser
- Not all documents from all issues of all journals present – ok, due to creation of fixed collection
- About 1% of Highwire PMIDs incorrect
  - Due to errors in Highwire-generated links
- A number of empty or very small files due to Highwire process
  - Empty files discarded but small ones kept, though not likely to have relevant passages

9



## Topics

- Phrased as questions
- Based on four of the five generic topic types (GTTs) developed for 2005
- Adapted from 2005 topics, though sometimes changed in substantial ways
  - Source of controversy due to concern over systems from 2005 being tuned for those topics

10



## Example topics with GTT and question pattern

GTT	Question Pattern	Example
Find articles describing the role of a gene involved in a given disease.	What is the role of gene in disease?	What is the role of DRD4 in alcoholism?
Find articles describing the role of a gene in a specific biological process.	What effect does gene have on biological process?	What effect does the insulin receptor gene have on tumorigenesis?
Find articles describing interactions (e.g., promote, suppress, inhibit, etc.) between two or more genes in the function of an organ or in a disease.	How do genes interact in organ function?	How do HMG and HMGB1 interact in hepatitis?
Find articles describing one or more mutations of a given gene and its biological impact.	How does a mutation in gene influence biological process?	How does a mutation in Ret influence thyroid function?

11



## Relevance judgments

- Pooling
- Judging
- Processing

12



## Relevance judgments – pooling

- Collected ranked passages in round robin manner from each submission until had 1000 per topic
- Text presented to judge was entire maximum-length legal span in which passage appeared, with HTML tags removed for better readability
- Passages prepared in Excel spreadsheets and sent to judges

13



## Relevance judgments – judging

- Performed by nine biology experts (mostly PhD)
- Developed documentation and training session for judges
- Judges instructed to
  - Select passages (from maximum-length legal spans provided) that were definitely or possibly relevant
  - Group relevant passages into aspects, designated by one or more MeSH terms assigned by judge
- Work reviewed by another biological expert (PR) before accepted

14



## Relevance judgments – processing

- Used BLAST-like algorithm to map relevant passages back to spans in original HTML file
- Created trec\_eval-like file with PMID, span start, span length, and MeSH aspects
- Developed Python programs to calculate passage, aspect, and document MAP from submission and gold standard files

15



## Relevance judgment results

- Two topics had no relevant passages and were discarded
- Remaining 26 topics had
  - 3451 relevant passages
  - 781 distinct aspects
  - 1450 relevant documents

16



## Relevance judgment results – statistics

Measure	Relevant passages per topic	Distinct aspects per topic	Mean relevant passage length
Min	3	7	27
Mean	35	22	400
Median	133	30	229
Max	593	96	6928

17



## Relevance judgment results – interjudge consistency

All 6 topics	Relevant	Not Relevant	Minus topic 181	Relevant	Not Relevant
Relevant	253	789	Relevant	234	228
Not Relevant	53	4905	Not Relevant	53	4485

Cohen's kappa = 0.32

Cohen's kappa = 0.60

- Six topics judged in duplicate (so far, at least one more coming)
- Above passage results show maximum legal passage relevance-based overlap
  - Consistency "good" for 5 topics, very poor for topic 181
  - Judges apparently interpreted topic 181 very differently
- Aspect (MeSH)-based consistency harder to assess, but appears to be less consistent
  - Similar to above, consistency "good" for 5, very poor for topic 181

18



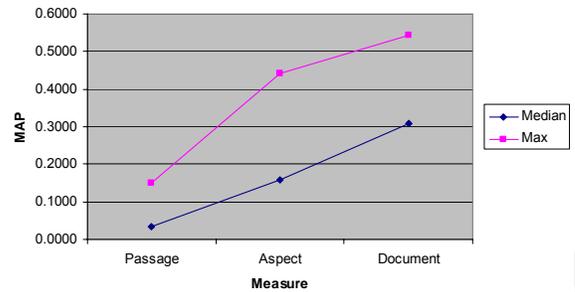
## Results

- Median and best results overall and by run category
- Run results sorted by passage, aspect, and document MAP

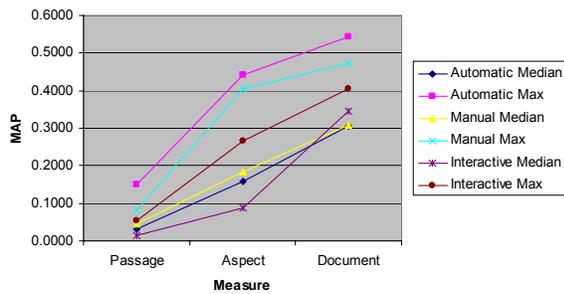
19



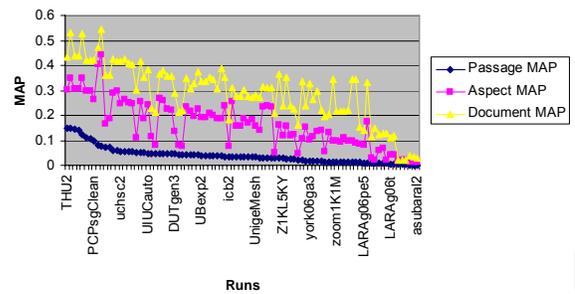
## Median and best results overall



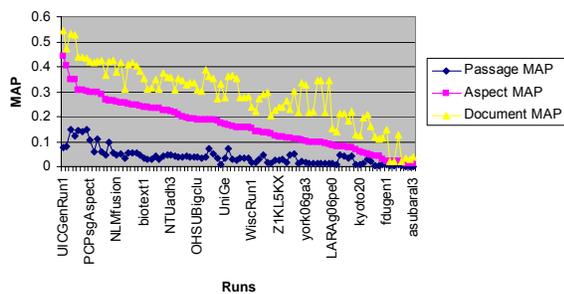
## Median and best results by run type



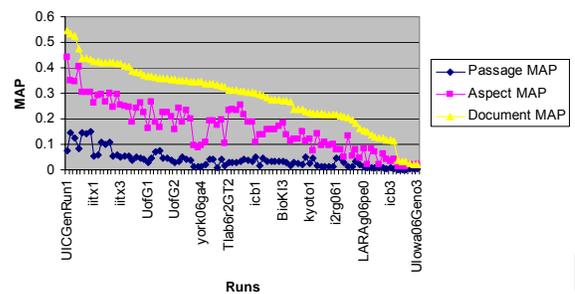
## Results sorted by passage MAP



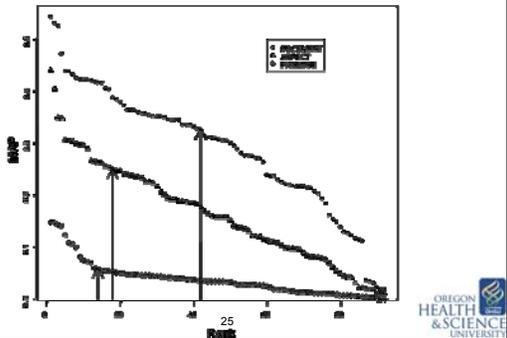
## Results sorted by aspect MAP



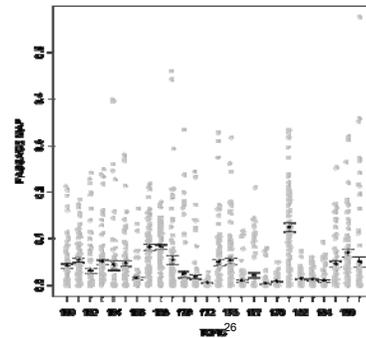
## Results sorted by document MAP



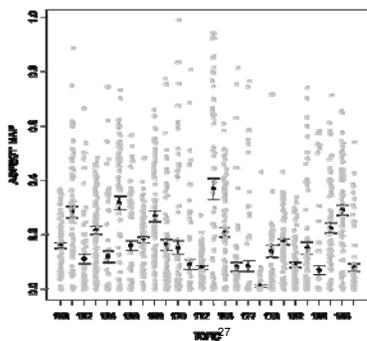
### Ranked MAP showing first run statistically different from best



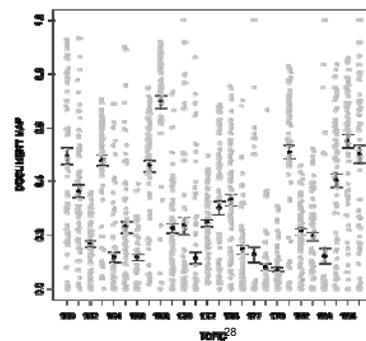
### Passage MAP by topic



### Aspect MAP by topic



### Document MAP by topic



### Preliminary analysis of results

- Based on very preliminary assessment from papers
- Baseline experiments lacking, so value of specific features unknown
- Best results from UIC and Tsinghua show common approaches
  - Query expansion using MeSH and UMLS for synonyms, hyper/hyponyms, and related terms
  - Paragraph retrieval → sentence selection → passage designation based on high scoring
- What did not work well (from OHSU middle-of-the-pack run)
  - Clustering passages and selecting in "round robin" manner



### Future directions

- With NSF grant ending and desire for new tracks in TREC, 2007 will be last year for TREC Genomics Track
- Plan to continue same task with new topics
  - May make tweaks around the edges
  - 2006 data will serve as good training data
- Beyond 2007? Stay tuned...

