The TREC Bio/Medical Tracks

William Hersh
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
School of Medicine
Oregon Health & Science University
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: http://informaticsprofessor.blogspot.com
Twitter: @williamhersh

References

Amini, I, Martinez, D, et al. (2016). Improving patient record search: a meta-data based approach. *Information Processing & Management*. 52: 258-272.
Anonymous (2012). From Screen to Script: The Doctor's Digital Path to Treatment. New York, NY, Manhattan Research; Google. http://www.thinkwithgoogle.com/insights/library/studies/the-doctors-digital-path-to-treatment/
Baker, M (2016). 1,500 scientists lift the lid on reproducibility. *Nature*. 533: 452-454.
Bastian, H, Glasziou, P, et al. (2010). Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Medicine*. 7(9): e1000326. http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000326
Blumenthal, D (2011). Implementation of the federal health information technology initiative. *New England Journal of Medicine*. 365: 2426-2431.
Blumenthal, D (2011). Wiring the health system--origins and provisions of a new federal program. *New England Journal of Medicine*. 365: 2323-2329.
Demner-Fushman, D, Abhyankar, S, et al. (2012). NLM at TREC 2012 Medical Records Track. *The Twenty-First Text REtrieval Conference Proceedings (TREC 2012)*, Gaithersburg, MD. National Institute for Standards and Technology http://trec.nist.gov/pubs/trec21/papers/NLM.medical.final.pdf
Demner-Fushman, D, Abhyankar, S, et al. (2011). A knowledge-based approach to medical records retrieval. *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, Gaithersburg, MD. National Institute for Standards and Technology
Edinger, T, Cohen, AM, et al. (2012). Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC Medical Records Track. *AMIA 2012 Annual Symposium*, Chicago, IL. 180-188.
Fox, S (2011). Health Topics. Washington, DC, Pew Internet & American Life Project. http://www.pewinternet.org/Reports/2011/HealthTopics.aspx
Hersh, W, Müller, H, et al. (2009). The ImageCLEFmed medical image retrieval task test collection. *Journal of Digital Imaging*. 22: 648-655.
Hersh, W, Turpin, A, et al. (2001). Challenging conventional assumptions of automated information retrieval with real users: Boolean searching and batch retrieval evaluations. *Information Processing and Management*. 37: 383-402.
Hersh, W and Voorhees, E (2009). TREC genomics special issue overview. *Information Retrieval*. 12: 1-15.
Hersh, WR (2001). Interactivity at the Text Retrieval Conference (TREC). *Information Processing and Management*. 37: 365-366.

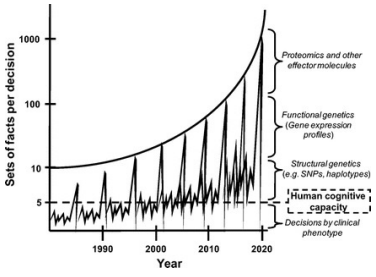
Hersh, WR, Crabtree, MK, et al. (2002). Factors associated with success for searching MEDLINE and applying evidence to answer clinical questions. *Journal of the American Medical Informatics Association*. 9: 283-293.
Hersh, WR and Hickam, DH (1995). An evaluation of interactive Boolean and natural language searching with an on-line medical textbook. *Journal of the American Society for Information Science*. 46: 478-489.
Hersh, WR, Müller, H, et al. (2006). Advancing biomedical image retrieval: development and analysis of a test collection. *Journal of the American Medical Informatics Association*. 13: 488-496.
Hersh, WR, Pentecost, J, et al. (1996). A task-oriented approach to information retrieval evaluation. *Journal of the American Society for Information Science*. 47: 50-56.
Ide, NC, Loane, RF, et al. (2007). Essie: a concept-based search engine for structured biomedical text. *Journal of the American Medical Informatics Association*. 14: 253-263.
King, B, Wang, L, et al. (2011). Cengage Learning at TREC 2011 Medical Track. *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, Gaithersburg, MD. National Institute for Standards and Technology
Martinez, D, Otegi, A, et al. (2014). Improving search over electronic health records using UMLS-based query expansion through random walks. *Journal of Biomedical Informatics*. 51: 100-106.
Müller, H, Clough, P, et al., Eds. (2010). ImageCLEF: Experimental Evaluation in Visual Information Retrieval. Heidelberg, Germany, Springer.
Roberts, K, Simpson, M, et al. (2016). State-of-the-art in biomedical literature retrieval for clinical cases: a survey of the TREC 2014 CDS track. *Information Retrieval Journal*. 19: 113-148.
Safran, C, Bloomrosen, M, et al. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *Journal of the American Medical Informatics Association*. 14: 1-9.
Stead, WW, Searle, JR, et al. (2011). Biomedical informatics: changing what physicians need to know and how they learn. *Academic Medicine*. 86: 429-434.
Tenenbaum, JD, Avillach, P, et al. (2016). An informatics research agenda to support precision medicine: seven key areas. *Journal of the American Medical Informatics Association*: Epub ahead of print.
Voorhees, E and Hersh, W (2012). Overview of the TREC 2012 Medical Records Track. *The Twenty-First Text REtrieval Conference Proceedings (TREC 2012)*, Gaithersburg, MD. National Institute of Standards and Technology http://trec.nist.gov/pubs/trec21/papers/MED12OVERVIEW.pdf
Voorhees, EM and Tong, RM (2011). Overview of the TREC 2011 Medical Records Track. *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, Gaithersburg, MD. National Institute of Standards and Technology

The TREC Bio/Medical Tracks

William Hersh
Professor and Chair
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: http://informaticsprofessor.blogspot.com
Twitter: @williamhersh

Why is IR important in health and biomedical domain?

- Clinicians cannot keep up – average of 75 clinical trials and 11 systematic reviews published each day (Bastian, 2010)
- Data points per clinical decision increasing (Stead, 2011)
- Search for health information by clinicians, researchers, and patients/ consumers is ubiquitous (Fox, 2011; Google/Manhattan Research, 2012)
- Concerns about reproducibility of science (Baker, 2016)
- "Precision medicine" requires retrieval not only of documents but also data (Tenenbaum, 2016)



The TREC Bio/Medical Tracks

- TREC Genomics Track
- ImageCLEFmed
- TREC Medical Records Track
- TREC Clinical Decision Support Track
- TREC Precision Medicine Track

TREC Genomics Track (Hersh, 2009)

- Motivated by completion of Human Genome Project, exploding research in genomics, and inability to biologists to know all that might impact work
- First TREC track devoted to "domain-specific" retrieval, with focus on IR systems for genomics researchers
  - Supported by NSF Information Technology Research (ITR) grant
- History
  - 2004-2005 – ad hoc retrieval and document categorization with bibliographic (MEDLINE) records
  - 2006-2007 – passage retrieval and question-answering (QA) with full-text journal articles

Lessons learned (Hersh, 2009)

- Ad hoc retrieval
  - Modest benefit for techniques known to work well in general IR, e.g., stop word removal, stemming, weighting
  - Query term expansion, especially domain-specific and/or done by humans, helped most
- QA
  - Most consistent benefit from synonym query expansion and paragraph-length or larger passage retrieval
- For all experiments (and papers describing them), major problems were
  - Lack of detailed description of systems
  - Use of low-performing baselines

Image retrieval – ImageCLEF medical image retrieval task

- Biomedical professionals increasingly use images for research, clinical care, and education, yet we know very little about how to best retrieve them
- Developed test collection and exploration of information needs motivating use of image retrieval systems (Hersh, 2006; Hersh, 2009; Müller, 2010)
  - Funded with supplement to NSF ITR grant
- Started with ad hoc retrieval and added tasks
  - Modality detection
  - Case finding
- In general, text retrieval yielded most consistent results with image features providing variable value
- Continues on with highly defined tasks

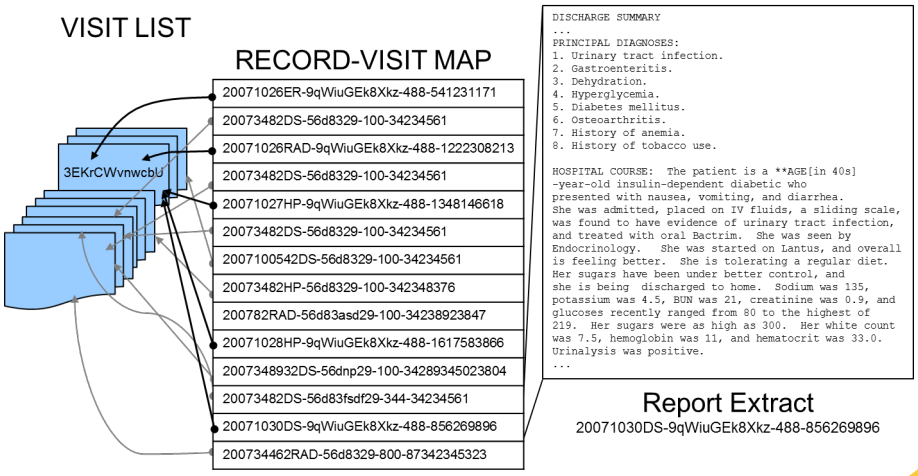
TREC Medical Records Track (Voorhees, 2011; Voorhees, 2012)

- Adapting IR techniques to electronic health records (EHRs)
- Use case somewhat different – want to retrieve records and data within them to identify patients who might be candidates for clinical studies
- Motivated by larger desire for "re-use" of clinical data (Safran, 2007)
- Opportunities facilitated by incentives for "meaningful use" of EHRs in the HITECH Act (Blumenthal, 2011; Blumenthal, 2011)

Challenges for informatics research with medical records

- Has always been easier with knowledge-based content than patient-specific data due to a variety of reasons
  - Privacy issues
  - Task issues
- Facilitated with development of large-scale, de-identified data set from University of Pittsburgh Medical Center (UPMC)
- Launched in 2011, repeated in 2012

Test collection



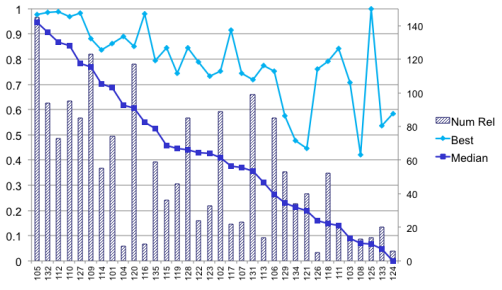
17,265 visits

101,712 reports (93,552 mapped to visits)

(Courtesy, Ellen Voorhees, NIST)

TREC Medical Records Track results

- Highly variable across different topics
  - Easiest – consistently best results
    - 105: Patients with dementia
  - Hardest – consistently worst results
    - 108: Patients treated for vascular claudication surgically
  - Large differences between best and worst results
    - 125: Patients co-infected with Hepatitis C and HIV
- Overall results show substantial room for improvement
  - Best results involve manual modification of queries



(Voorhees, 2011; Voorhees, 2012)

Which approaches worked?

- Best results in 2011 and 2012 obtained from NLM group (Demner-Fushman, 2011; Demner-Fushman, 2012)
  - Top results from manually constructed queries using Essie domain-specific search engine (Ide, 2007)
- Many approaches known to work in general IR fared less well, e.g., term expansion, document focusing, etc.
  - Other domain-specific approaches also did not show benefit, e.g., creation of PICO frames, negation
- Some success with
  - Results filtered by age, race, gender, admission status; terms expanded by UMLS Metathesaurus (King, 2011)
  - Expansion by concepts and relationships in UMLS Metathesaurus (Martinez, 2014)
  - Pseudorelevance feedback using ICD-9 codes (Amini, 2016)

OHSU

Failure analysis for 2011 topics (Edinger, 2012)

Reasons for Incorrect Retrieval	Number of Visits	Number of Topics
Visits Judged Not Relevant		
Topic terms mentioned as future possibility	16	9
Topic symptom/condition/procedure done in the past	22	9
All topic criteria present but not in the time/sequence specified by the topic description	19	6
Most, but not all, required topic criteria present	17	8
Topic terms denied or ruled out	19	10
Notes contain very similar term confused with topic term	13	11
Non-relevant reference in record to topic terms	37	18
Topic terms not present—unclear why record was ranked highly	14	8
Topic present—record is relevant—disagree with expert judgment	25	11
Visits Judged Relevant		
Topic not present—record is not relevant—disagree with expert judgment	44	21
Topic present in record but overlooked in search	103	27
Visit notes used a synonym or lexical variant for topic terms	22	10
Topic terms not named in notes and must be inferred	3	2
Topic terms present in diagnosis list but not visit notes	5	5

OHSU

TREC Clinical Decision Support Track (Roberts, 2016)

- www.trec-cds.org
- Ad hoc search of biomedical literature (PubMed Central Open Access Subset – 1.25M articles)
- Topics are patient descriptions in three information need categories
  - Diagnosis
  - Test
  - Treatment
- Currently in third year of operation
- Transitioning to Precision Medicine Track

13

TREC biomedical tracks have inspired other challenge evaluations

- i2b2
  - https://www.i2b2.org/NLP
  - Various NLP-related tasks, including extraction and de-identification
- CLEF eHealth
  - https://sites.google.com/site/clefehealth/home
  - Information extraction and patient-centered IR
- bioCADDIE
  - https://biocaddie.org/biocaddie-2016-dataset-retrieval-challenge-registration
  - Data set retrieval

14

System-oriented retrieval is not enough

- How well do IR systems enable users to perform their tasks?
  - Major task in medicine: answering clinical questions
- Undertook parallel approaches assessing Boolean vs. natural language searching in
  - Medicine – using
    - Electronic textbook – Scientific American Medicine (Hersh, 1995)
    - Bibliographic database – MEDLINE (Hersh, 1996)
  - General news – TREC Interactive Track (Hersh, 2001)
- Largest study employed 45 medical and 21 NP students answering tripartite questions using MEDLINE, with results showing (Hersh, 2002)
  - Medical students improved from chance (33%) to 50% correct, whereas NP students showed little benefit from using IR system
  - Recall and precision of searches showed no association with correct answers

Conclusions

- TREC challenges have provided insights and enduring test collections for continued research
- Importance of IR in biomedicine will continue, especially as volume, variety, and velocity of science continue to expand
  - Varying benefits for different use cases, but in general, medical vocabulary resources offer most value via query expansion
- While ad hoc IR for general information needs relatively understood, still challenges with
  - Novel types of data, e.g., EHRs and other structured data
  - High-recall tasks, e.g., systematic reviews
- Research confounded by larger issues, e.g.,
  - Private data
  - Proprietary data