

The TREC Track Structure

Ellen Voorhees

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce



Text REtrieval Conference (TREC)

What is TREC?

- A workshop series that provides the infrastructure for large-scale testing of (text) retrieval technology
 - realistic test collections
 - uniform, appropriate scoring procedures
 - a forum for the exchange of research ideas and for the discussion of research methodology

TREC Impacts

- Test collections
- Incubator for new research areas
- Common evaluation methodology and improved measures for text retrieval
- Open forum for exchange of research
- Technology transfer

TREC Tracks

- Task that focuses on a particular subproblem of text retrieval
- Tracks invigorate TREC & keep TREC ahead of the state-of-the-art
 - specialized collections support research in new areas
 - first large-scale experiments debug what the task really is
 - provide evidence of technology's robustness

CLIR Track

- Solidified the CLIR research community
- Built the first large-scale test collections for CLIR
 - pioneered methodology that enabled building such collections:
 - distributed topic development and assessing, recognizing importance of native speakers as judges
- Fostered establishment of related evaluation conferences

QA Track

- Motivated by desire for a common task for IR and IE communities
- Reinvigorated stalled research community
 - explosion of QA workshops, journal issues, books
- Pioneered several new evaluation methodologies
 - factoid questions
 - definition questions
 - "complex" QA with pyramids
- Resulted in significant progress in NLP

Genomics Track

- First domain-specific TREC track
 - investigate use of domain-specific information to improve retrieval performance
 - create effective retrieval engines to support genomics research
- Research methodology informed by domain
 - tasks drawn from specific needs of community
 - required assessors/annotators skilled in the domain

Track Definition

- Paramount is the task definition: what functionality is to be tested?
 - broad enough to attract participants
 - narrow enough to be amenable to evaluation
- Track implementation
 - what data is required; is it available? what intellectual property rights are involved?
 - how can performance on the task be measured? what is needed to obtain those measurements?