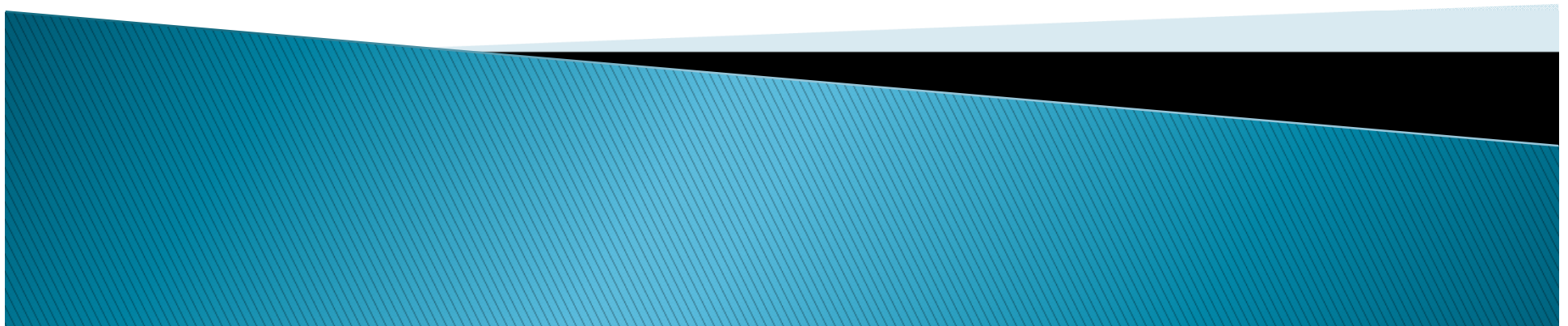


# TREC Session Track

Ben Carterette  
Paul Clough  
Evangelos Kanoulas  
Mark Sanderson



# Vision



# Vision



## Cranfield Paradigm

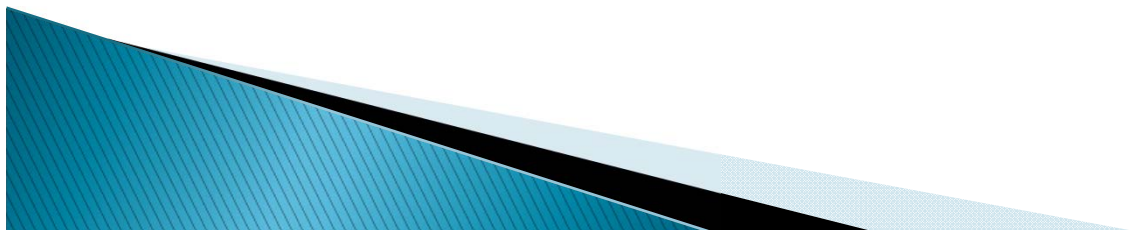
- Simple user model
- Controlled experiments
- Reusable but static test collections

## Session Track

## Interactive Evaluation/ Live Labs

- Full user participation
- Many degrees of freedom
- Unrepeatable experiments

- ▶ Put user in the evaluation loop



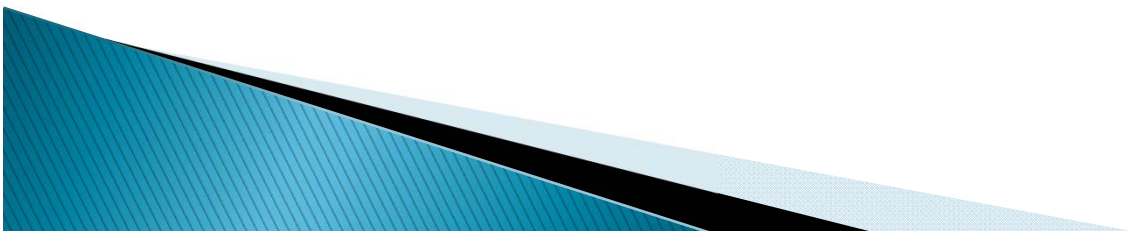
# Vision

- ▶ Simulate an interactive experiment
  - Compare IIR systems
    - by controlling user interactions
  - Build a collection that is **portable and reusable**
  - Devise measures to evaluate the utility a user obtains throughout a session
- ▶ Test systems in a large variety of cases that lead to sessions



# Sessions

- ▶ Categorization on “who’s to blame” basis?
  - Corpus
    - There is no composite document that fulfills the users need
  - User
    - Users cannot express their need by the appropriate query
    - Users learn about their need throughout the session
  - System



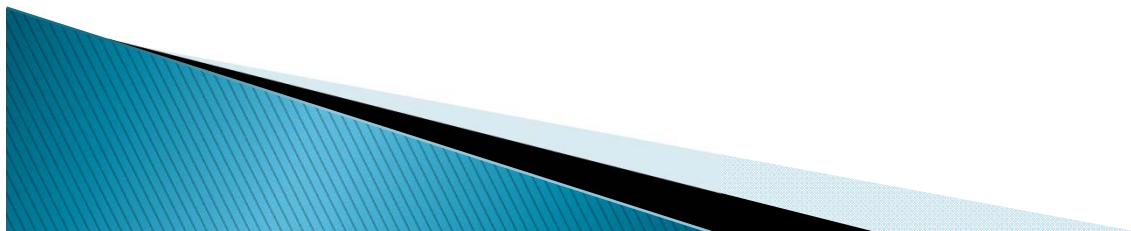
# IIR systems for sessions

- ▶ Accumulate information through the session
  - Learn from the past
    - Improve on the current query
  - Look into the future



# Goals

- (G1) Test the ability of retrieval systems to use the history of user interactions (past queries,...) to improve performance on the current query
- (G2) Evaluate system performance over an entire session instead of a single query



# Reality





# Task 1: If history can teach us anything...

- ▶ All participants should be provided with a fixed history of user interactions

$Q_1$



$RL_1$

[TREC](#)  
Forensic Video Enhancement, Video Enhancement, Authentication, Law Enforcement, Industrial Security, Courts, Surveillance Video, Digital Magnification, Depositions, Expert, Witness ...  
[trec.com](#) · [Cached page](#)

[TREC - Training Resources for the Environmental Community](#)  
TREC's mission is to catalyze the habitat conservation and wilderness protection community in Western North America to increase its effectiveness in promoting sustainable ...  
[trec.org](#) · [Cached page](#)

[Text Retrieval Conference \(TREC\) Home Page](#)  
TREC Economic Impact Study The TREC Conference series is co-sponsored by the NIST Information Technology Laboratory's (ITL) Retrieval Group of the ...  
[trec.nist.gov](#) · [Cached page](#)

[TREC - Agency Info Main Page](#)  
contact, Agency Information: Address, email, e-mail, Fax, Phone, Hours, Staff  
[www.trec.state.tx.us/agencyinfo/default.asp](#) · [Cached page](#)

[Text Retrieval Conference - Wikipedia, the free encyclopedia](#)  
Tracks · Conference Contributions · Participation  
The **Text Retrieval Conference (TREC)** is an on-going series of workshops focusing on a list of different information retrieval (IR) research areas, or tracks.  
[en.wikipedia.org/wiki/Text\\_Retrieval\\_Conference](#) · [Cached page](#)

[Text Retrieval Conference \(TREC\) Overview](#)  
The **Text Retrieval Conference (TREC)**, co-sponsored by the National Institute of Standards and Technology (NIST) and U.S. Department of Defense, was started in 1992 as part of the ...  
[trec.nist.gov/overview.html](#) · [Cached page](#)

[TREC RENTAL: Rent photography equipment, cameras, digital backs...](#)  
TREC RENTAL: Rent photo equipment, cameras, digital backs, digital assistants, lighting, grip, profoto, and more in NYC and beyond  
[trecrental.com](#) · [Cached page](#)

[TREC - Wikipedia, the free encyclopedia](#)  
TREC may refer to: **Text Retrieval Conference**, an on-going series of workshops co-sponsored by the National Institute of Standards and Technology (NIST) and the U.S. Department of ...  
[en.wikipedia.org/wiki/Trec](#) · [Cached page](#)

[Tennessee Real Estate Commission - Home](#)  
The TREC CORE course has increased from four (4) hours to six (6) hours continuing education credit beginning January 1, 2011. TREC NEWS-JOURNAL  
[state.tn.us/commerce/boards/trec](#) · [Cached page](#)

$Q_n$



$RL_n$

[Text Retrieval Conference \(TREC\) Home Page](#)  
TREC Economic Impact Study The TREC Conference series is co-sponsored by the NIST Information Technology Laboratory's (ITL) Retrieval Group of the ...  
[trec.nist.gov](#) · [Cached page](#)

[Text Retrieval Conference \(TREC\) Publications](#)  
The TREC Conference series is co-sponsored by the NIST Information Technology Laboratory's (ITL) Retrieval Group of the Information Access Division (IAD) and the ...  
[trec.nist.gov/pubs.html](#) · [Cached page](#)

[TREC Video Retrieval Evaluation Home Page](#)  
The TREC conference series is sponsored by the National Institute of Standards and Technology with additional support from other U.S. government agencies.  
[www.nipir.nist.gov/projects/trecvid](#) · [Cached page](#)

[Terabyte TREC Homepage](#)  
Introduction. Early retrieval test collections were small, allowing relevance judgments to be based on an exhaustive examination of the documents, but limiting the general ...  
[www.nipir.nist.gov/projects/terabyte](#) · [Cached page](#)

[Text Retrieval Conference - Wikipedia, the free encyclopedia](#)  
Tracks · Conference Contributions · Participation  
The **Text Retrieval Conference (TREC)** is an on-going series of workshops focusing on a list of ... It is co-sponsored by the National Institute of Standards and Technology (NIST) ...  
[en.wikipedia.org/wiki/Text\\_Retrieval\\_Conference](#) · [Cached page](#)

[TREC-BLOG - Information Retrieval Wiki](#)  
The TREC 2006 Blog track 'Overview paper' has appeared in the Proceedings of TREC 2006, and is available from the TREC Web site at <http://trec.nist.gov/pubs/trec15/papers/BLOG06...>  
[ir.dcs.gla.ac.uk/wiki/TREC-BLOG](#) · [Cached page](#)

[Table of Contents](#)  
For more information regarding Text Retrieval Conferences (TREC), please visit the TREC web site at <http://trec.nist.gov>. There have been six Text Retrieval Conferences (TREC s ...  
[itl.nist.gov/iaui/894.02/related\\_projects/tipster/trec.htm](#) · [Cached page](#)

[TRUCHIRAPPALLI REGIONAL ENGINEERING COLLEGE](#)  
Shri. P. Chidambaram Former Finance Minister, Government of India Home Minister, Government of India Inaugurated TREC-STEP's New Technology Incubator Building  
[www.trecstep.com](#) · [Cached page](#)

[TRECVID Data Availability](#)  
Buy NIST videos from the NIST Standard Reference Data site. search topics ... These are available in the protected Past Results section of the TREC website under "TREC ...  
[trecvid.nist.gov/trecvid.data.html](#) · [Cached page](#)

$Q_{n+1}$



# Participants Task

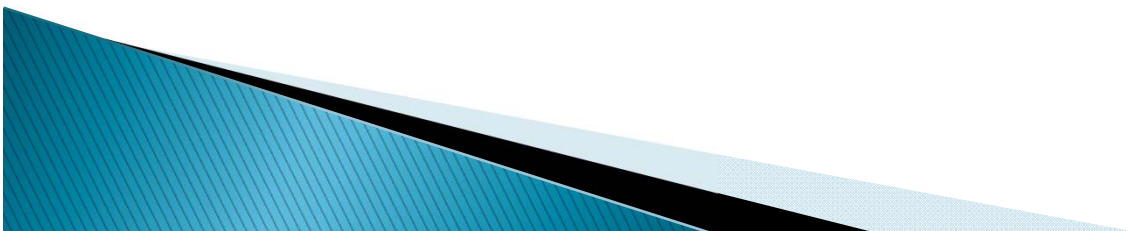
- ▶ For each session participants submitted 4 ranked lists (RLs)

(RL1) current query

(RL2) current query & past queries in the session

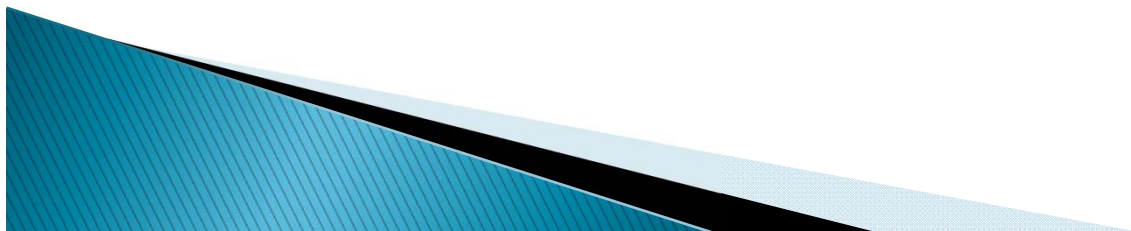
(RL3) current query & past queries in the session and ranked URLs

(RL4) current query & past queries in the session, ranked URLs, clicks and dwell times



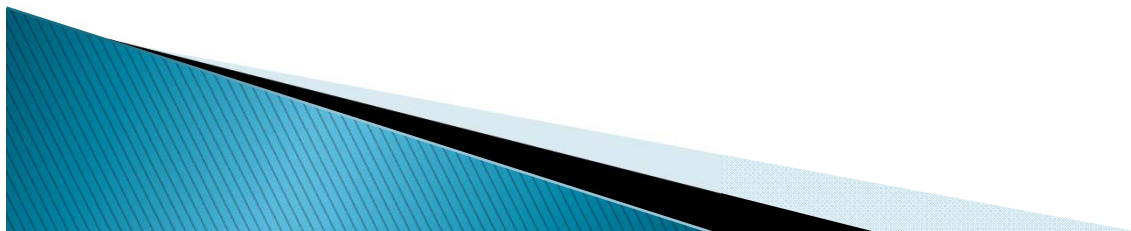
# Challenges

- ▶ Come up with topics that can lead to sessions with more than one query.
  - In 2011 we generated about 100 topics, collected about 1000 sessions:
    - 90% were single session queries.
  - In 2012 we got about the same percentage of single-query sessions



# Challenges

- ▶ Come up with topics that can lead to different search task types.
  - In 2011 we generated factual tasks with specific goal(s)
  - In 2012 we generated a variety of task types
    - factual / intellectual tasks
    - specific / amorphous goals



# More challenges

- ▶ Evaluating retrieval systems
  - The notion of relevance may change throughout the session
  - Novelty needs to be considered
  - Different levels of diversification throughout the session should be applied



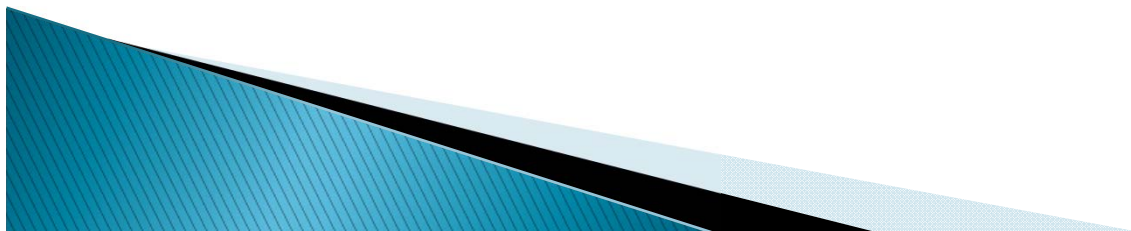
# Task 2 : Looking into the future ...

- ▶ Whole-session evaluation



# Task 2 : Looking into the future ...

- ▶ Whole-session evaluation
- ▶ Simulate user interactions
  - Given a topic, a ranked list, and relevance information
    - Simulate browsing behavior
    - Simulate clicks, dwell times, etc.
    - Simulate query reformulation
- ▶ Quantify utility



# Conclusions

- ▶ Built static test collection
- ▶ Evaluate retrieval performance when session information is available
- ▶ Still far from a whole-session evaluation
  - Is it possible to build reusable and portable but dynamic test collections?
  - Is it worth it?

