

# TREC, TAC, takeoffs, tacks, tasks, and titillations for 2009

Ian Soboroff, NIST

[ian.soboroff@nist.gov](mailto:ian.soboroff@nist.gov)



# Agenda

- TREC 2008
- (some) reflections on TREC
- TAC, a new evaluation conference for NLP
- TREC 2009 preview

# TREC Goals

- To increase research in information retrieval based on large-scale collections
- To provide an open forum for exchange of research ideas to increase communication among academia, industry, and government
- To facilitate technology transfer between research labs and commercial products
- To improve evaluation methodologies and measures for information retrieval
- To create a series of test collections covering different aspects of information retrieval

# TREC 2008 Program Committee

Ellen Voorhees, chair

David Lewis

James Allan

John Prager

Chris Buckley

Steve Robertson

Gord Cormack

Mark Sanderson

Sue Dumais

Ian Soboroff

Donna Harman

Richard Tong

Bill Hersh

Ross Wilkinson

# TREC 2008 Participants

**Beijing Univ. of Posts &  
Telecommunications**

**Brown University**

**Carnegie Mellon University**

**Chinese Acad. of Sciences**

**Clearwell Systems, Inc.**

**CNIPA ICT Lab**

**Dalian U. of Technology**

**Dublin City University**

**Fondazione Ugo Bordoni**

**Fudan University**

**H5**

**Heilongjiang Inst. of Tech.**

**Hong Kong Polytechnic U.**

**IBM Research Lab**

**Indian Inst Tech, Kharagpur**

**Indiana University**

**INRIA**

**Kobe University**

**Korea University**

**Max-Planck-Institut Informatik**

**Nat'l Univ. of Ireland, Galway**

**Northeastern University**

**Open Text Corporation**

**Pohang Univ Science & Tech**

**RMIT University**

**Sabir Research**

**SEBIR**

**St. Petersburg State Univ.**

**SUNY Buffalo**

**TNO ICT**

**Tsinghua University**

**Universidade do Porto**

**University College, London**

**Univ. of Alaska, Fairbanks**

**University of Amsterdam (2)**

**Univ. of Arkansas, Little Rock**

**University of Avignon**

**University of Glasgow**

**Univ. of Illinois, Chicago**

**U. Illinois, Urbana-Champaign**

**University of Iowa (2)**

**University of Lugano**

**Univ. Maryland, College Park**

**University of Massachusetts**

**Univ. of Missouri-Kansas City**

**University of Neuchatel**

**University of Pittsburgh**

**University of Texas at Dallas**

**University of Twente**

**University of Waterloo (2)**

**Ursinus College**

**Wuhan University**

**York University**

# Tracks

blog

Craig Macdonald, Iadh Ounis,  
Ian Soboroff

---

enterprise

Peter Bailey, Nick Craswell,  
Arjen de Vries, Ian Soboroff,  
Paul Thomas

---

legal

Jason Baron, Bruce Hedin,  
Doug Oard, Stephen  
Tomlinson

---

million query

James Allan, Jay Aslam

---

relevance feedback

Chris Buckley, Stephen  
Robertson

# TREC 2008

- TREC 2008: November 18-21  
(we are between the conference and the final proceedings)
- But here are some things to look for..

# TREC 2008

- Evaluation challenges
  - continue exploring alternatives to traditional pooling for test collection building
  - sampling methods in MQ, rel fdbk, legal tracks
  - new samples entail new evaluation measure computations
  - revisit impact of variability in relevance judgments
- Contextualizing search
  - enterprise, legal, blog tasks target specific use cases

# Blog Track

## Tasks:

1. Finding blog posts that contain opinions about the topic
2. Ranking positive and negative blog posts
3. (A separate baseline task to just find blog posts relevant to the topic)
4. Finding blogs that have a principal, recurring interest in the topic

# Enterprise Track

- Enterprise: CSIRO
- Topics taken from CSIRO Enquiries (they get the “contact us” emails)
- Tasks:
  1. Find key pages which answer the enquiry
  2. Find people who are topic experts that might help answer the enquiry

# Legal Track

- Legal discovery search task
  - Topics divided among several complaints.
  - Each topic includes a request, a Boolean query (with negotiation), and more...
- Relevance feedback task
- Interactive task
  - Goal: to find as many responsive documents as possible for any of three topics
  - Each group could use 10 hours of time with a domain expert lawyer

# Million Query Track

- 10,000 queries
- Gov2 collection (25M web pages, 425 GB)
- Queries divided among long/short, many/few clicks
- ~800 queries judged by NIST assessors using two sampling strategies
  - “Minimal test collections” method (Carterette et al, SIGIR 2006)
  - “statAP” method (Aslam et al, SIGIR 2006)

# Relevance Feedback Track

- Goal: look again at relevance feedback, in modern collections and with modern methods
- 264 topics run on the Gov2 collection
  - 50 terabyte topics + 214 MQ topics
  - All queries included in this year's MQ set
- A range of feedback conditions

# TREC 2008

- Results are still preliminary...
- So I won't show them here.
- (Think of this as an invitation to participate)
- Final papers due in February.
- Proceedings in the spring (hopefully).

# Reflections

- TREC 2009 will be our 18th year
- 2 GB → 426 GB
- 50 topics → 1,800 topics
- tasks: ad hoc, filtering, novelty, question answering, known-item search ...
- multiple languages, media, document types
- multiple domains: legal, genomics, enterprise

# The TREC Tracks

Personal documents															■	■	■	■	Blog Spam
Retrieval in a domain											■	■	■	■	■	■			Legal Genome
Answers, not docs								■	■	■	■	■	■	■	■	■			Novelty Q&A
Web searching, size								■	■	■	■	■	■	■	■	■	■	■	Enterprise Terabyte Web VLC
Beyond text				■	■			■	■	■	■	■	■	■	■	■	■	■	Video Speech OCR
Beyond just English			■	■	■			■	■	■	■	■	■						$X \rightarrow \{X, Y, Z\}$ Chinese Spanish
Human-in-the-loop			■	■	■	■	■	■	■	■	■	■	■	■				■	Interactive, HARD, fdbk
Streamed text	■	■	■	■	■	■	■	■	■	■	■								Filtering Routing
Static text	■	■	■	■	■	■	■	■	■		■	■	■				■	■	Million query Ad Hoc, Robust
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008		



- The Text Analysis Conference is a new NIST evaluation forum.
- TAC focuses on natural language processing tasks.

# Why TAC?



*SemEval*

TREC



NTCIR

Open MT

ACE

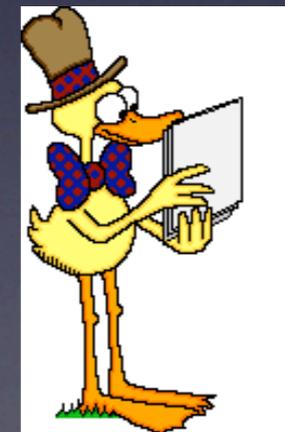


DUC



RTE

CoNLL

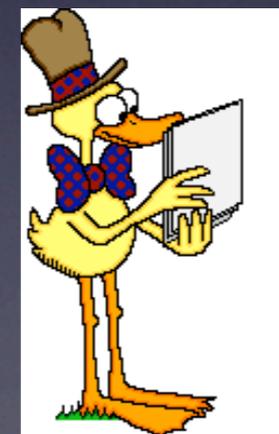


# Why TAC?

TREC - QA



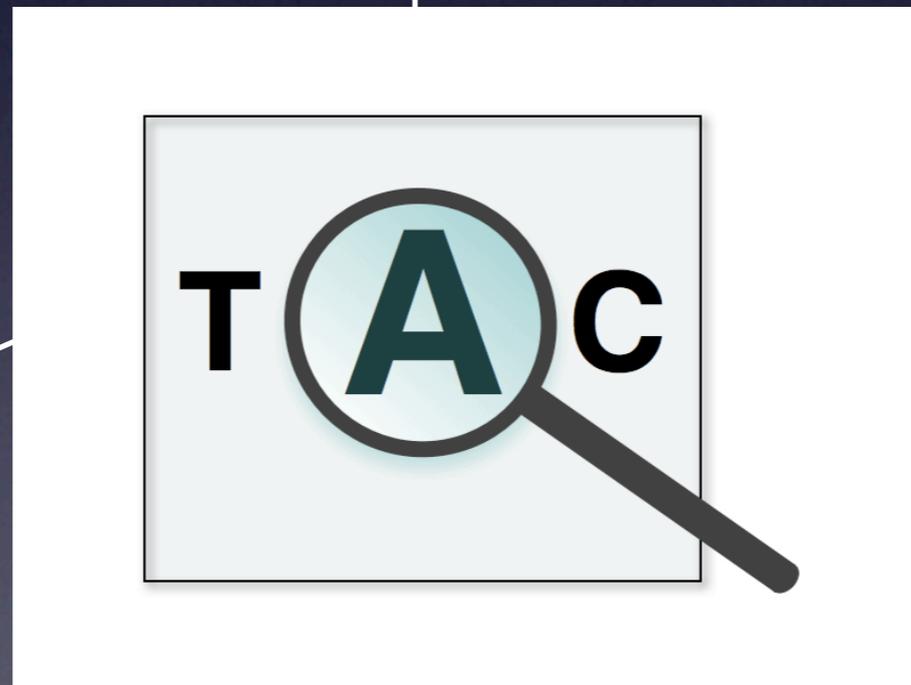
DUC



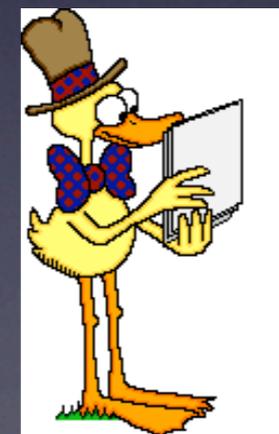
RTE

# Why TAC?

TREC - QA



DUC



RTE

# Features of TAC

- Component evaluations situated within context of end-user tasks (e.g., summarization, QA)
  - opportunity to test components in end-user tasks
- Test common techniques across tracks
- Small number of tracks
  - critical mass of participants per track
  - sufficient resources per track (data, assessing, technical support)
- Leverage shared resources across tracks (organizational infrastructure, data, assessing, tools)

# TAC 2008 Tracks

**RTE:** systems recognize when one piece of text entails or contradicts another

**QA:** systems return a precise answer in response to a question, focusing on opinion questions asked over blogs

**Summarization:** systems return a fluent summary of documents focused by a narrative or set of questions

1. Update: summarize new information in newswire articles for a user who has already read an earlier set of articles
2. Opinion pilot: summarize blog documents containing answers to opinion question(s) -- joint with QA

# Recognizing Textual Entailment (RTE)

- Goal: recognize when one piece of text is entailed by another
- Classification Task: given T(ext) and H(ypothesis)
  - H is entailed by T
  - H is not entailed by T
    - H contradicts T
    - H neither contradicts nor is entailed by T
- T/H pairs from IR, IE, QA, and summarization contexts.

# RTE Pairs from QA Setting

- H: generated from questions and candidate answer terms returned by QA systems searching the Web

*Baldwin is Antigua's Prime Minister.*

- T: candidate answer passages returned by QA systems

The opposition Antigua Labour Party (ALP) has blasted that country's prime minister, Baldwin Spencer, for publicly advocating that Cuba's Fidel Castro be awarded the Order of the Community (OCC) - the Community's highest honour.

# Update Summarization Task

- Given a topic and 2 chronologically ordered clusters of news articles, A and B, where A documents precede B documents
- Create two brief ( $\leq 100$  words), fluent summaries that contribute to satisfying the information need expressed in the topic statement:
  - Initial summary (A): summary of cluster A
  - Update summary (B): summary of cluster B, assuming reader has read cluster A

# Pipelined Opinion QA/Summarization Task

# Pipelined Opinion QA/Summarization Task



Why  
don't people like  
Trader Joe's?

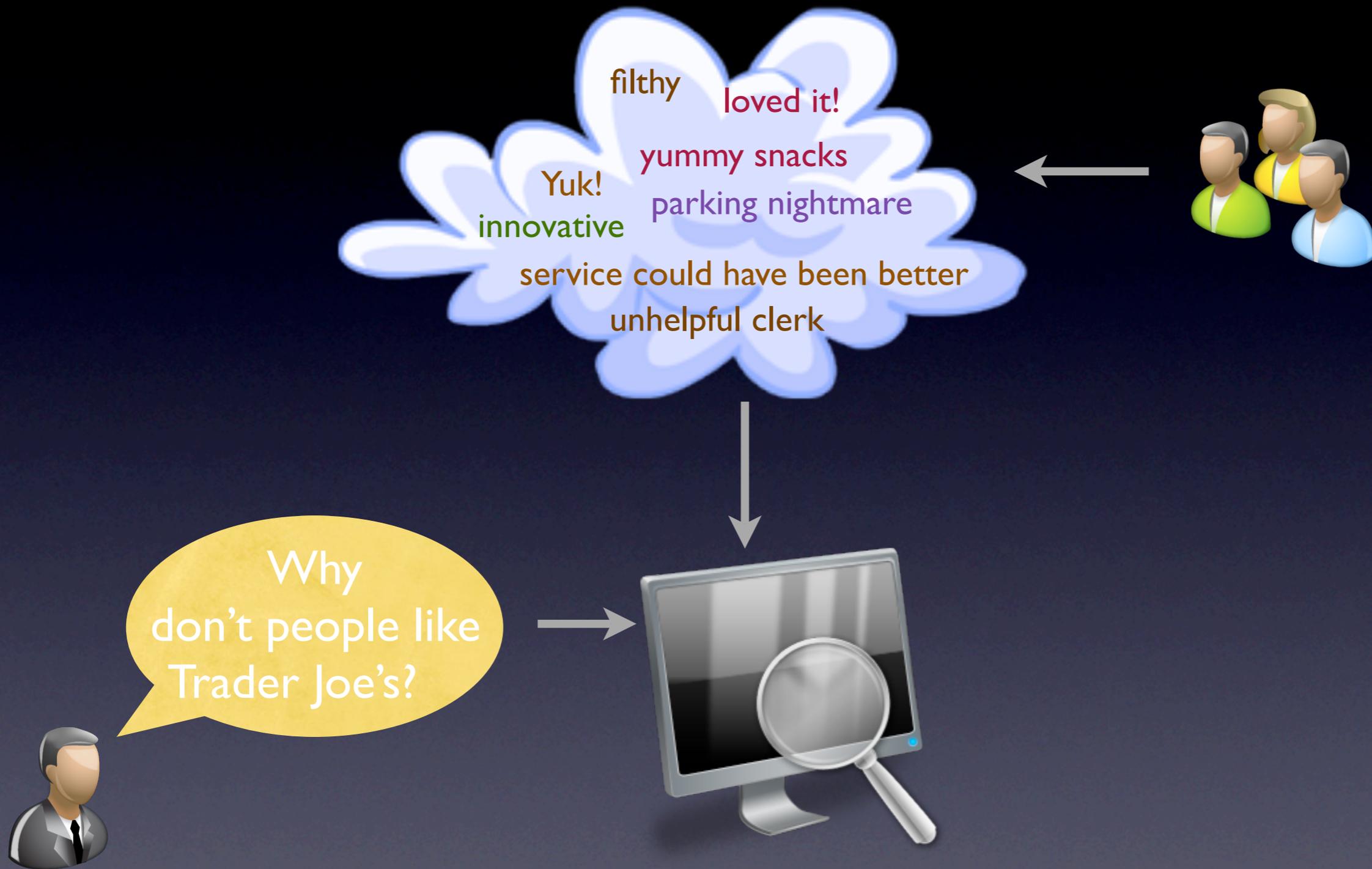
# Pipelined Opinion QA/Summarization Task



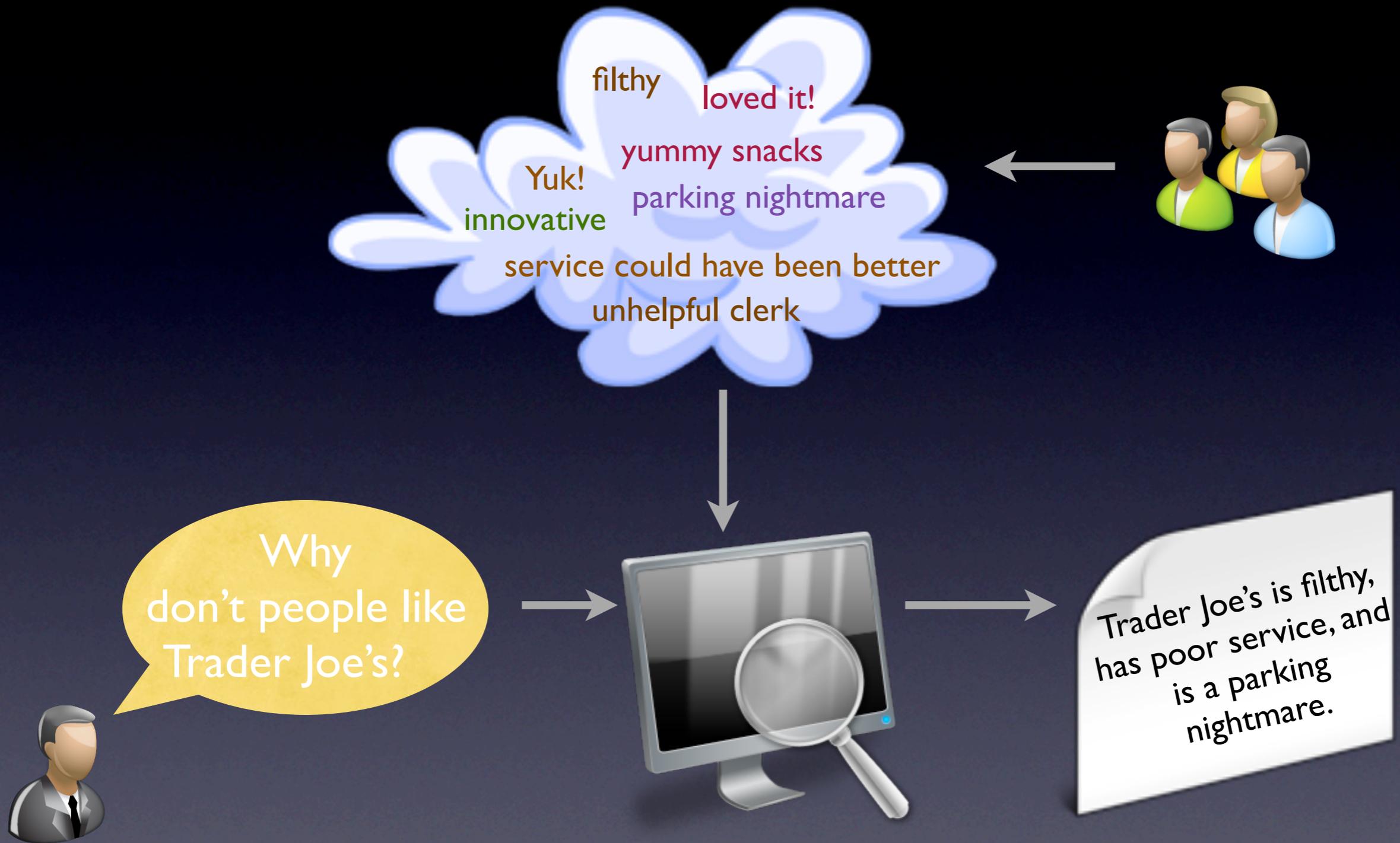
Why  
don't people like  
Trader Joe's?



# Pipelined Opinion QA/Summarization Task



# Pipelined Opinion QA/Summarization Task



# Opinion QA

TARGET: "MythBusters"

1018.1 RIGID LIST Who likes Mythbuster's?

1018.2 SQUISHY LIST Why do people like Mythbuster's?

1018.3 RIGID LIST Who do people like on Mythbuster's?

# Opinion QA

TARGET: "MythBusters"

- |                     |  |
|---------------------|--|
| 1018.1 RIGID LIST   | Who likes Mythbuster's?  |
| BLOG06-3334         | CAPS_CHAMP   |
| BLOG06-8580         | Jon  |
| BLOG06-3982         | Zonk   |
| 1018.2 SQUISHY LIST | Why do people like Mythbuster's?   |
| BLOG06-6706         | The Mythbusters chicas are purdy .   |
| BLOG06-5962         | It's geek, period. And a lot of fun. I like that they have women on their team who are also into mechanical stuff and applied science. |
| 1018.3 RIGID LIST   | Who do people like on Mythbuster's?  |
| BLOG06-3187         | Kari Byron   |
| BLOG06-4849         | scottie  |
| BLOG06-6570         | Jamie Hyneman  |

# Opinion Summarization

- Input
  - Target, 1-2 squishy list questions
  - Documents known to have answers
  - *Optional* answer-snippets in each document
- Output
  - Single fluent summary of the answers to all the questions
- Opinion polarity classification (positive vs negative) may help fluency

# Proposed TAC 2009 Tracks

1. RTE

2. Summarization:

- Update ?
- Opinion ?
- Meeting/Speech ?

3. Information Extraction for Knowledge Base Population

# New for TREC 2009

- Chemical IR track
- New blog track collection
- New web collection supporting four tracks(two new, two old)
- Legal, RF, and MQ tracks will continue
- Enterprise track ending

# Chemical IR track

- Coordinators: John Tait (IRF), Jianhan Zhu (Open U), Jimmy Huang (York U), Mihai Lupu (IRF)
- Document collection (to be provided by IRF)
  - ~100,000 patents (XML formatted)
  - ~45,000 journal articles from the UK Royal Society of Chemistry
- Tasks:
  - Chemical patent claim search: given claims, find the patent references
  - Ad hoc search, with topics developed and judged by domain experts

# Blog track

- Coordinators: Iadh Ounis, Craid Macdonald (U Glasgow), Ian Soboroff (NIST)
- New Blog08 collection
  - spans a full year
  - 600k feeds, 40 million blog posts
  - permalinks, feeds, and blog homepages
- Task changes
  - faceted blog distillation (authority, expertise)
  - news + blog tracking pilot

# New web collection

- 1 billion web pages (~ 25 TB uncompressed)
- Collected by CMU (with advice from major search engines)
- Spans the 10 most prominent languages on the web
- Available on a set of hard drives for nominal cost
- We are exploring making the collection available from one or more clusters
- And establishing subsets of the collection
- RF, MQ, and two new tracks are planning to use this collection

# Web Track

- Coordinators: Nick Craswell (MSR), Charles Clarke (U Waterloo)
- A reborn web track
- Focus on web search tasks
  - navigational, topic distillation
  - diversity ranking
  - failed queries
  - spammed queries
- Driven by query logs and click data

# Entity Track

- An entity is something with a homepage – people, products, organizations...
- Coordinators: Krisztian Balog (UvA), Paul Thomas (CSIRO), Arjen de Vries (CWI), Thijs Westerveld (Teezir)
- Task: related entity finding
  - given a homepage, the entity type, and a narrative
  - return related entity homepages
  - ex: find studios that Tom Cruise worked with
- Other tasks under discussion

# TREC 2009

- Blog
  - Chemical
  - Legal
  - Web
  - Entity
  - Relevance  
Feedback
  - Million Query
- 

- The call for participation is out now:  
<http://trec.nist.gov/>
- CfP includes addresses of track mailing lists  
and other track details



Shameless  
Plug!

# ICWSM 2009

- The International Conference on Weblogging and Social Media has a data challenge
- (chairs: Ian Soboroff and Akshay Java)
- Data: 40 million blog posts covering 8 weeks (Aug - Oct 2008) (provided by Spinn3r.com)
- The data is free to get, free to use
- Conference papers due January 21st
- Data workshop papers due March 1st
- <http://icwsm.org/>