

# DEFT Open and End-to-End Evaluations

Ian Soboroff

NIST

# Open Evaluations

- Examples: TREC, TAC, OpenMT.
- Includes performers but is open to anyone.
- Aimed at the component level, but focused on a user task.
- Evaluation data is the same for all participants.
- Some evaluations share training data, too.
  
- Payoff: a wider community helping advance the state of the art in DEFT problems.

# End-to-end evaluations

- More common in the DARPA context.
- Can be intrinsic or extrinsic. The DEFT E2E evaluations will be intrinsic.
- Systems measured in isolation.
- Measures may not be strictly comparable.
- Payoff: see how well all the components work together in a prototype scenario.

## Technical Area 1: Implemented Algorithms\*

- Smart Filtering Example Algorithm Topics:
  - Redundancy Detection
  - Semantic Equivalence Detection
- Relational Analysis Example Algorithm Topics:
  - Event and Entity Coreference Resolution
  - Semantic Relation Detection
- Anomaly Analysis Example Algorithm Topics:
  - Disfluency Detection
  - Novelty Detection
- Alerts: Document clusters with Summary Bullets

•\*Not exhaustive or exclusive



# Smart Filtering (e.g., Redundancy Detection)

Input: Find information on Company X and their merger with Company Y



Redundancy Detection

Paraphrase and Event Coref across Dialogue turns



Output:

Retrieved Documents Determined to be Redundant

Company X and Company Y have merged in a 1.9 billion dollar deal.

The merger between Company X and Company Y had a transaction value of \$1.9 billion.

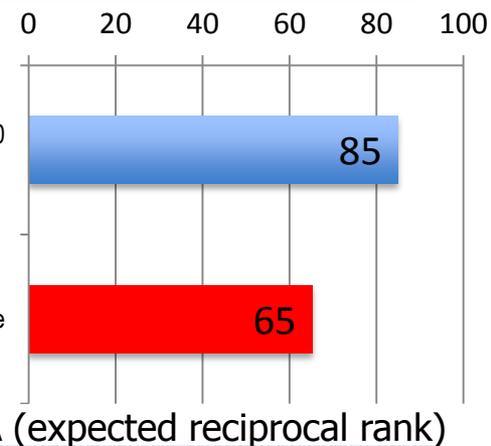
Retrieved Documents Determined to be Non-Redundant

This merger is a cash plus stock reorganization.

The total merger consideration is \$29.394 per Company X share.

Company X and Company Y have completed their merger deal worth about \$1.9 billion.

Top Score on Clean Data (NIST 2010 Web Diversity Track)



Report:

... **returning from** a certain corridor just **outside the capitol**

... **fleeing** the region

Informal Communication:

We need a **new place to meet** ...tonight.

You know, **the one where we met last time.**



# Smart Filtering (e.g., Semantic Equivalence Detection)

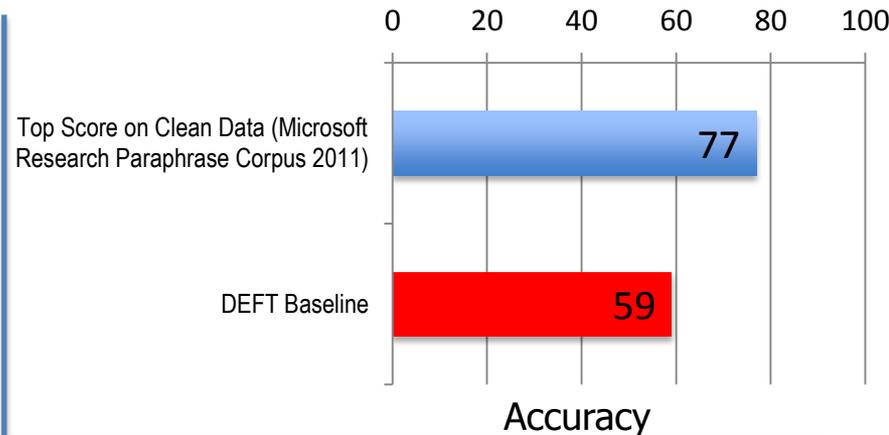
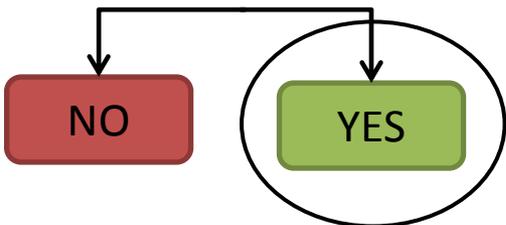
About 120 potential jurors were being asked to complete a lengthy questionnaire.

The jurors were taken into the court room in groups of 40 and asked to fill out a questionnaire.



Passage Alignment

Cross-Sentential Entity & Event Coref and Paraphrase



## Report:

...**sneak** a **car** laden with explosives **into** ...  
... **get** their **vehicle inside** ...

## Informal Communication:

A: How about the...uh...uh... **house**?  
You know, the **one** where we met last time.

B: You mean **your uncle's**?



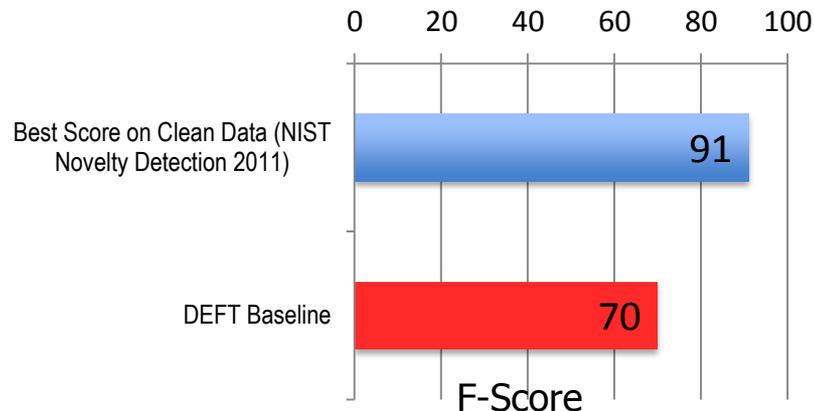
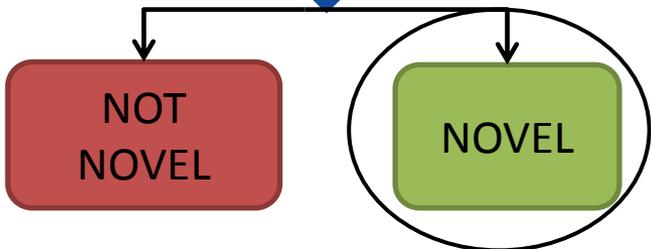
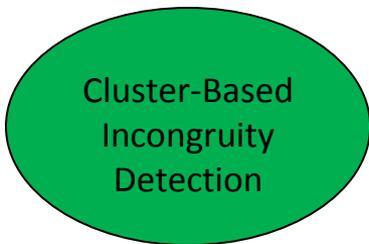
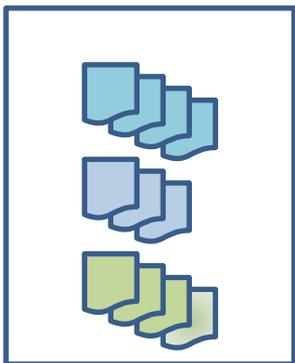
# Anomaly Analysis (e.g., Novelty Detection)

News stories (input)



Hypothesis: Thousands of people were forced to leave their pets behind when they evacuated New Orleans

“Memory of events”



Report: Terrorist organization X managed to **sneak a car laden with explosives into** the central intelligence **building**...

- Informal Communication: You know we need the stuff for the...**the party**
- ...You had better not **\$%\*! this up again.**

# Technical Area 2: End-to-End DEFT Alert System

