TRECVID 2005 Workshop

Columbia University TRECVID 2005 Search Task

Shih-Fu Chang, Winston Hsu, Lyndon Kennedy,
Akira Yanagawa, Eric Zavesky, Dong-Qing Zhang

Digital Video and Multimedia Lab
Columbia University

Nov. 14 2005
http://www.ee.columbia.edu/dvmm

Columbia University in the City of New York

dvmm DIGITAL VIDEO • MULTIMEDIA LAB

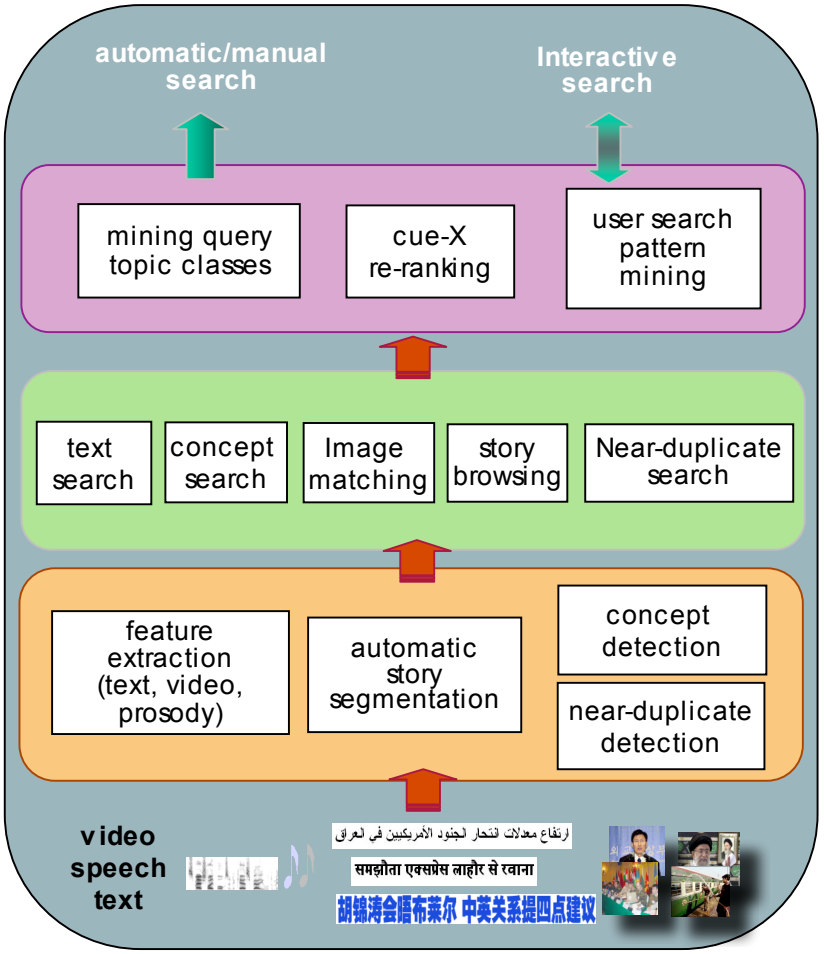
Columbia Video Search System Overview

http://www.ee.columbia.edu/cuvidsearch

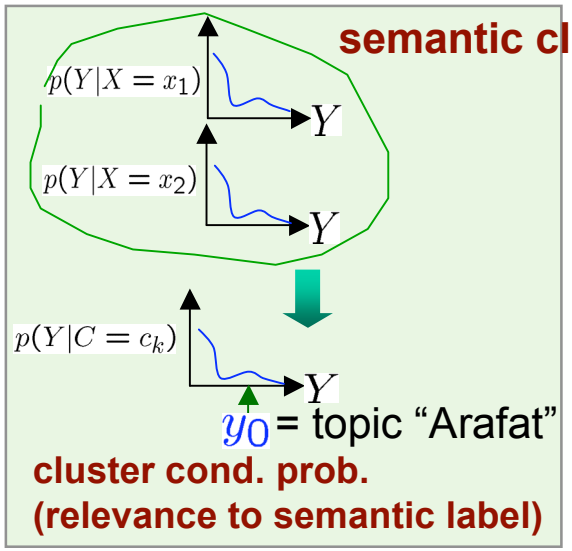
User Level Search Objects
- Query topic class mining
- Cue-X reranking
- Interactive activity log

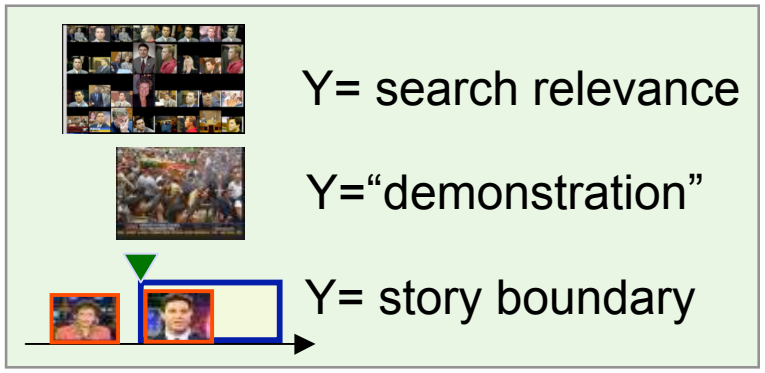
Multi-modal Search Tools
- combined text-concept search
- story-based browsing
- near-duplicate browsing

Content Exploitation
- multi-modal feature extraction
- story segmentation
- semantic concept detection



Cue-X Information-theoretic Framework





Y

semantic label

Information Bottleneck principle

$$C^* = \underset{C|R}{\operatorname{argmin}}\{I(X;Y) - I(C;Y)\}$$

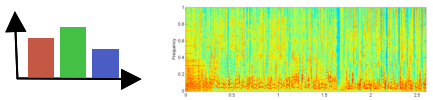
C_1 C_2 ... C_k ... C_{K-1} C_K

↑ **cue-X clusters automatically discovered via**

Information Bottleneck principle & Kernel Density Estimation (KDE)

X

low-level features



胡锦涛会晤布莱尔 中英关系提四点建议

ارتفاع معدلات انتحار الجنود الأمريكيين في العراق

नाहार से रवाना

News Story Segmentation in TRECVID 2005

- Cue-X framework effectively applied to discover salient features and achieve accurate story segmentation
  - Focus on visual and audio (prosody) features only
  - Without a priori manual selection of features
  - High accuracy across multi-lingual data sources
- TRECVID 2005
  - Dataset
    - 277 videos, 3 languages (ARB, CHN, and ENG),
    - 7 channels, 10+ different programs
    - Poor or missing ASR/MT transcripts
  - Accuracy on the validation set
    - Cue-X features + prosody features (no text features!)
    - ARB-0.87, CHN-0.84, and ENG-0.52 (F1 measure)
  - Results donated to whole TRECIVD 2005 community
- Story boundary results available for download at http://www.ee.columbia.edu/dvmm/downloads/cuex_story.htm

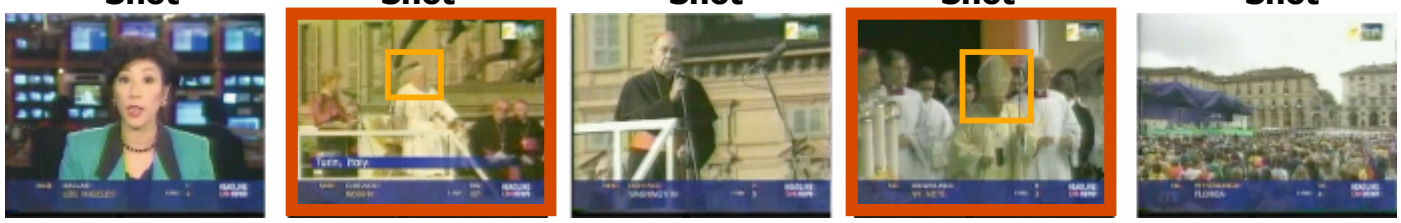
Enhancing Interactive Search Using Story Boundaries

Query

Find shots of Pope John Paul second

Story

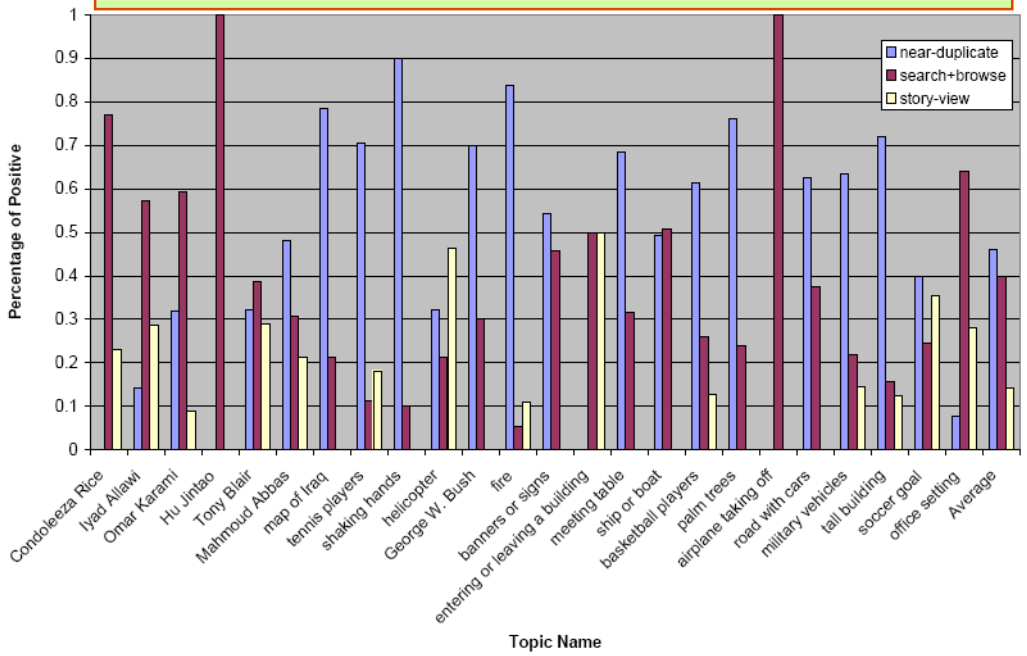
Shot **Shot** **Shot** **Shot** **Shot**



in other news **pope john paul** the **second** will get his first look at the shroud of turin today that's the piece of linen many believe was the burial cloth of jesus the round is on public display for the first time in twenty years it has already drawn up million visitors the pope's visit to northwest italy has also included beatification services for three people the vatican says **john paul** is now the longest serving pope this century he has surpassed **pope** pious the twelfth who served for nineteen years seven months and seven days

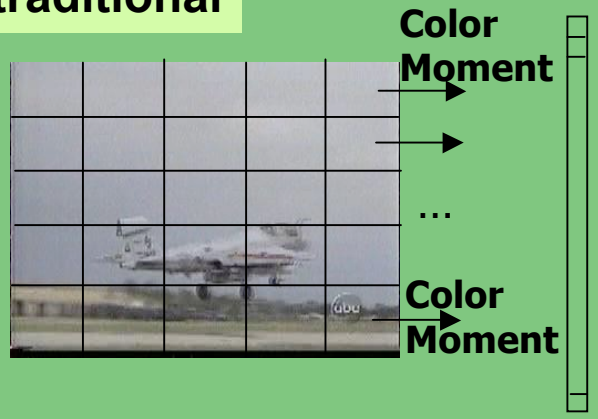
- **Stories define an intuitive unit with coherent semantics**
- **Story boundaries are effectively detected by Cue-X using audio-visual features**
- **Improves text search by more than 100% in TRECVID 2005 automatic search**
- **Major contributor to good performance of interactive video search**

Relative contributions from different search tools



Enhancing Semantic Concept Detection Performance Using Local Features and Spatial Context

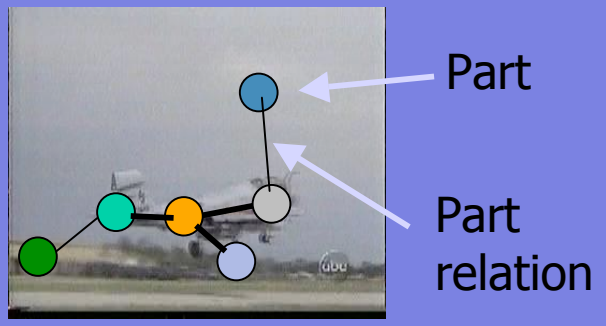
traditional



Global or block-based features:

- Difficult to achieve robustness against background clutter
- Difficult to model object appearance variations

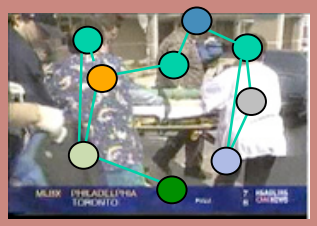
enhanced



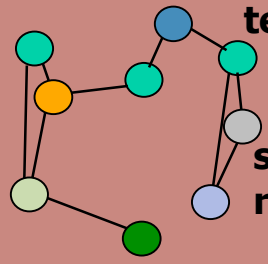
Part-based model:

- Eliminate background clutter
- Model part appearance more accurately
- Model part relation more accurately

Extracting Graphical Representations of Visual Content and Learning Statistical Models of Content Classes



Individual images
→ Salient points, high entropy regions

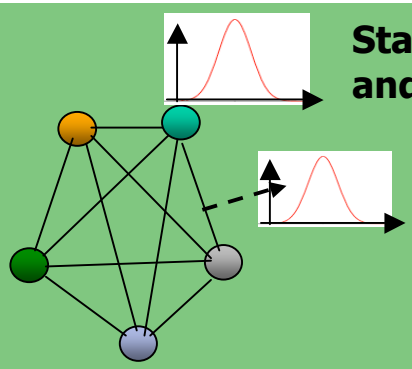


Attributed Relational Graph (ARG)

Graph Representation of Visual Content

Collection of training images

machine learning

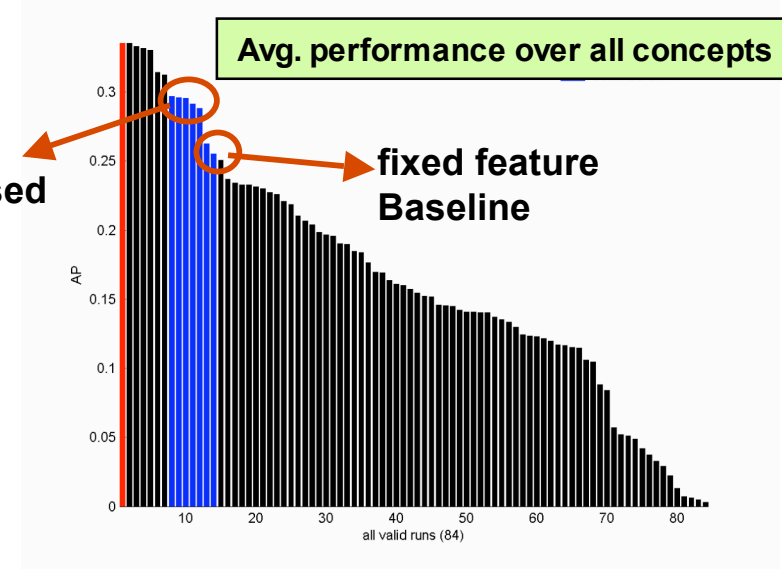


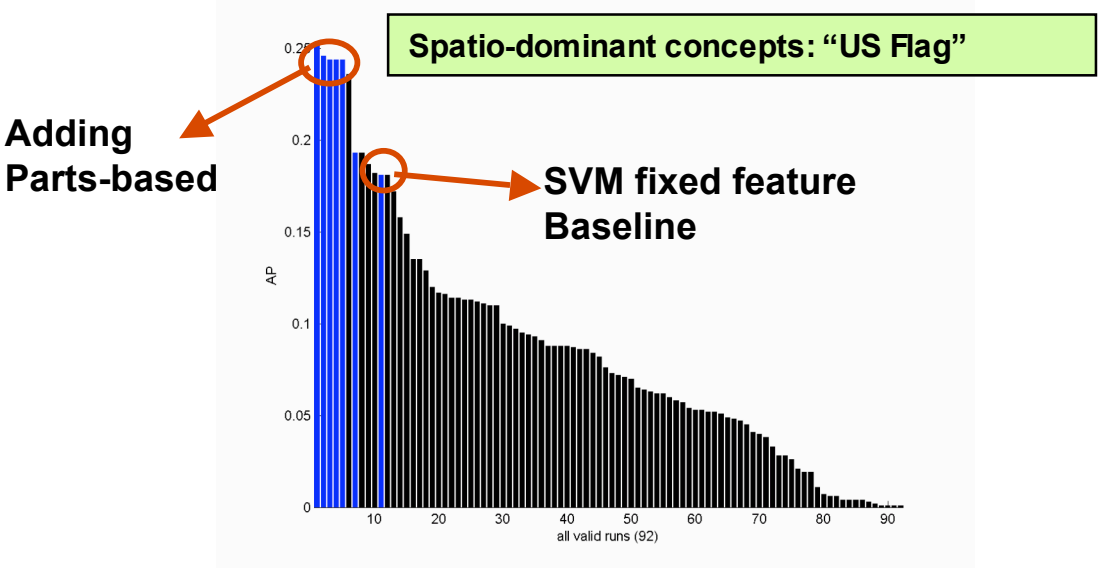
Random Attributed Relational Graph (R-ARG)

Statistical Graph Representation of Model

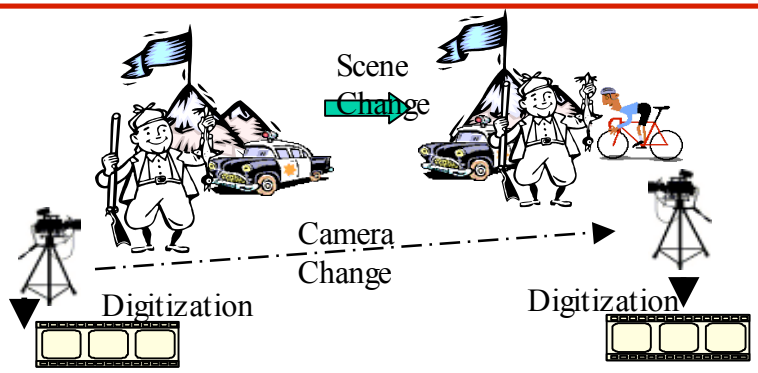
Parts-based detector performance in TRECVID 2005

- Parts-based detector consistently improves by more than 10% for all concepts
- It performs best for spatio-dominant concepts such as “US flag”.
- It complements nicely with the discriminant classifiers using fixed features.





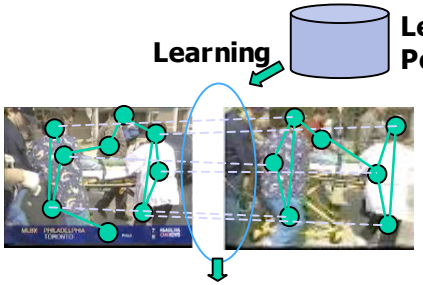
Search Components:
Detecting Image Near Duplicates (IND)

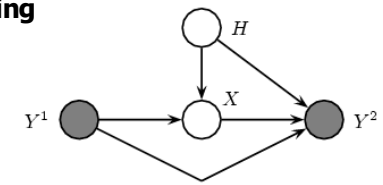


- **Near duplicates occur frequently in multi-channel broadcast**
- **But difficult to detect due to diverse variations**
- **Problem Complexity**

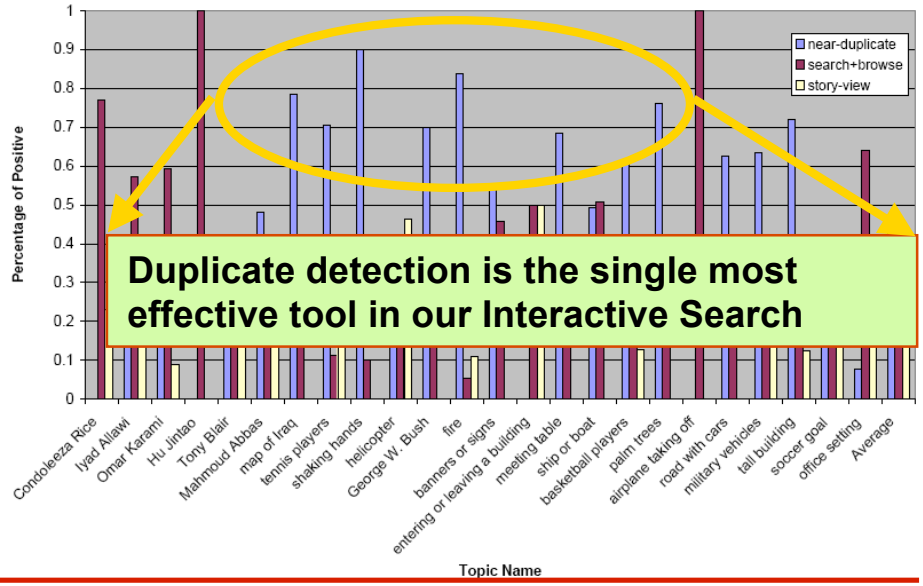
Similarity matching < IND detection < object recognition

Parts-based Stochastic Attribute Relational Graph Learning

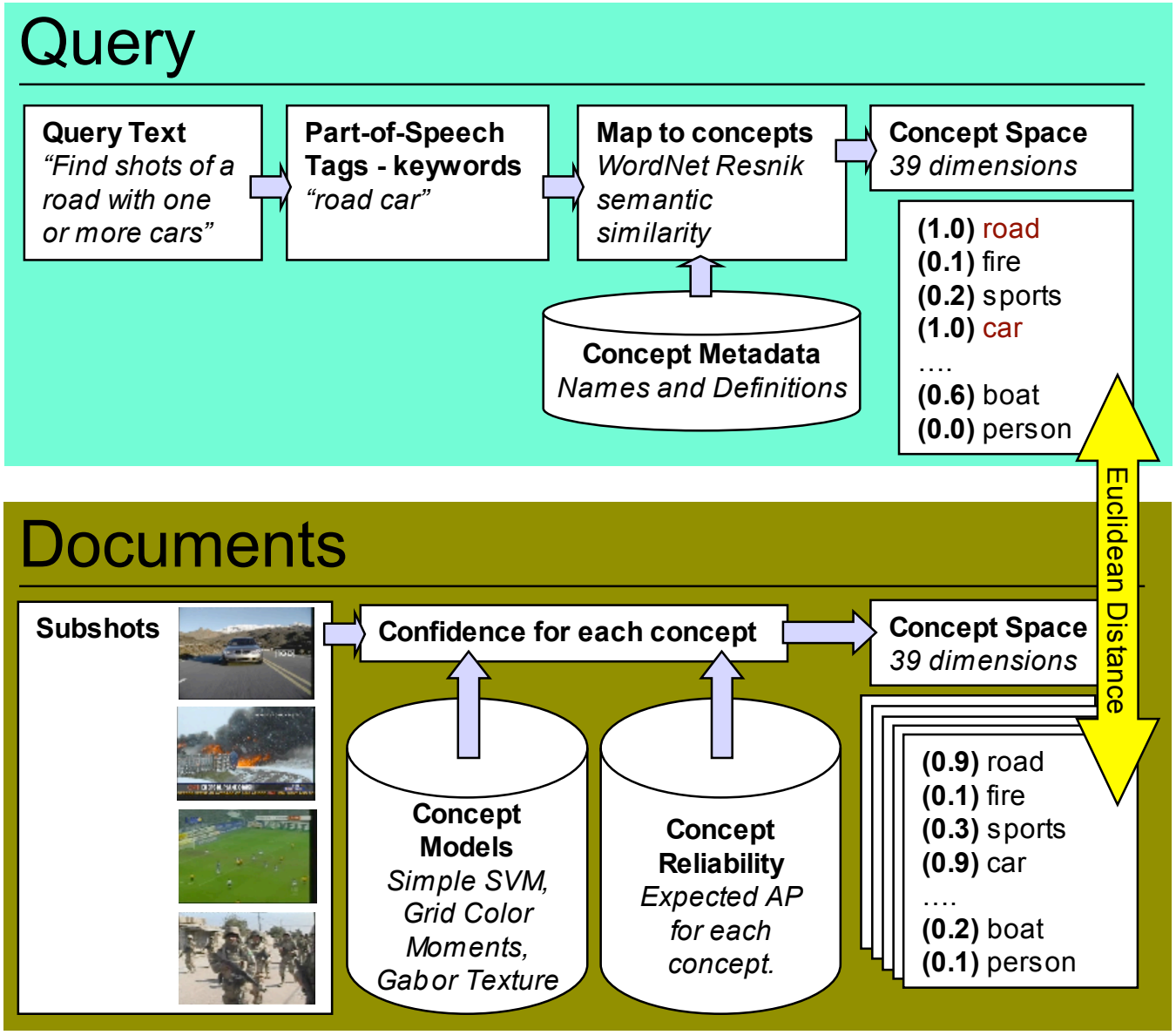




Stochastic graph models the physics of scene transformation



Concept Search



- Map text queries to high-level feature detection
- Use human-defined keywords from concept definitions
- Measure semantic distance between query and concept
- Use detection and reliability for subshot documents

Concept Search

Automatic - Can help for queries with related concepts

"Find shots of boats."

Method	AP
Story Text	.169
CBIR	.002
Concept	.115
Fused	**.195**

"Find shots of a road with one or more cars."

Method	AP
Story Text	.053
CBIR	.009
Concept	.090
Fused	**.095**

Manual / Interactive

Manual keyword selection allows more relationships to be found.

Query Text
"Find shots of an office setting, i.e., one or more desks/tables and one or more computers and one or more people"

Concepts
Office

Query Text
"Find shots of a graphic map of Iraq, location of Bagdhad marked - not a weather map"

Concepts
Map

Query Text
"Find shots of one or more people entering or leaving a building"

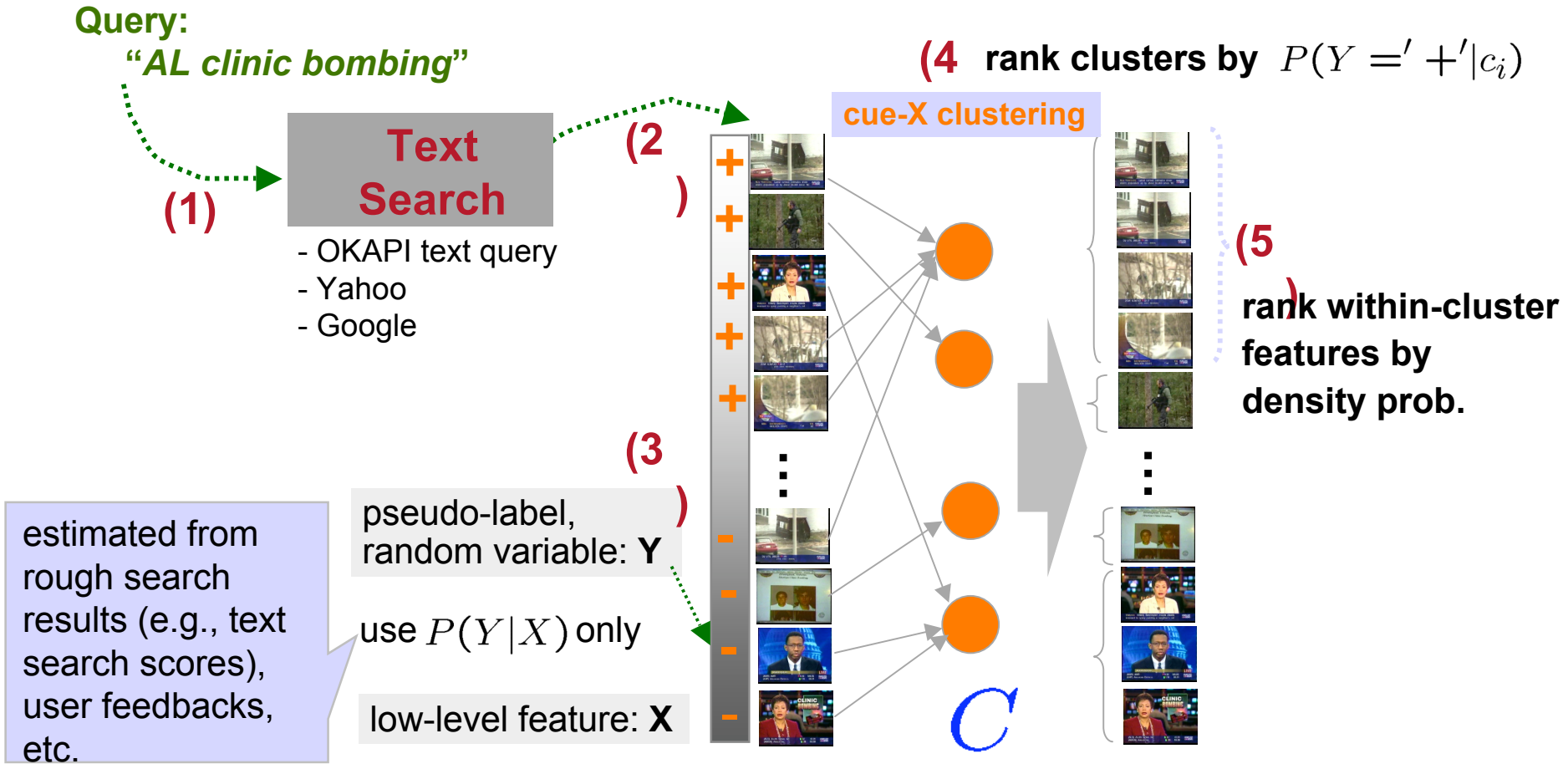
Concepts
Person, Building, Urban

Query Text
Find shots of people with banners or signs

Concepts
March or protest

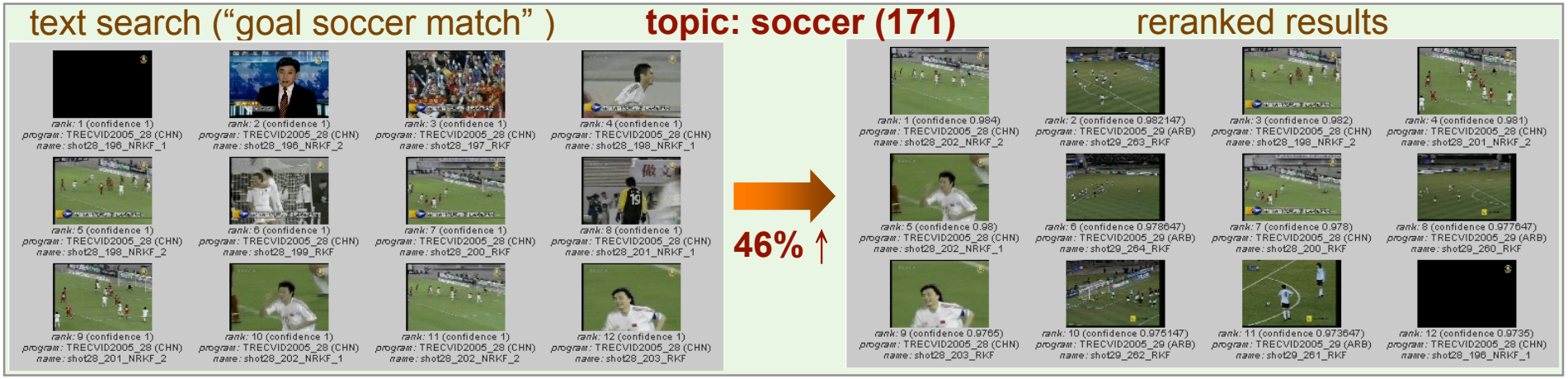
Cue-X Reranking by Pseudo-Labeling

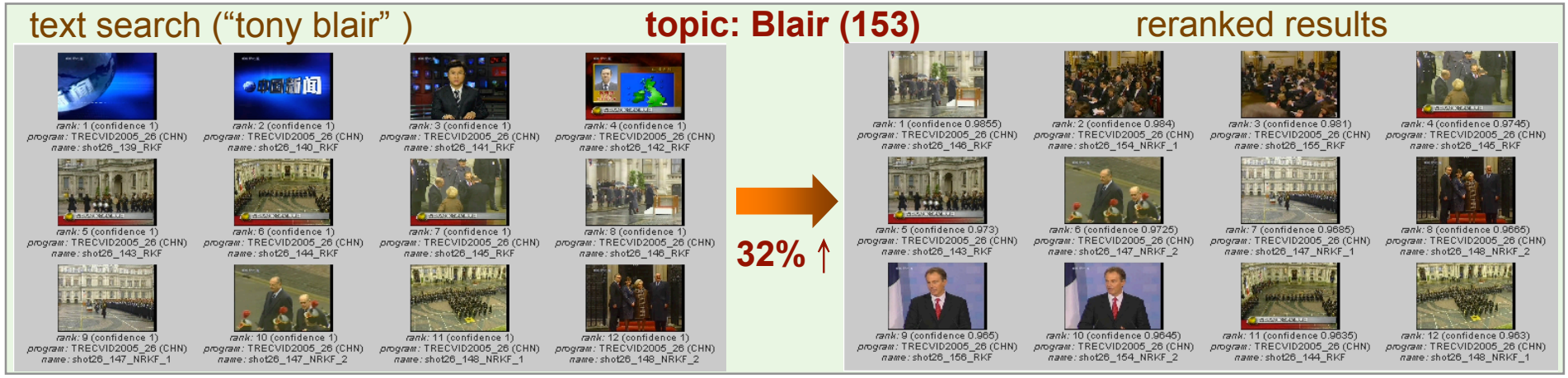
- Learn the recurrent relevant and irrelevant low-level patterns from the estimated pseudo-labels
- Reorder shots by the smoothed cluster relevance



Effect of Cue-X Reranking in Video Search

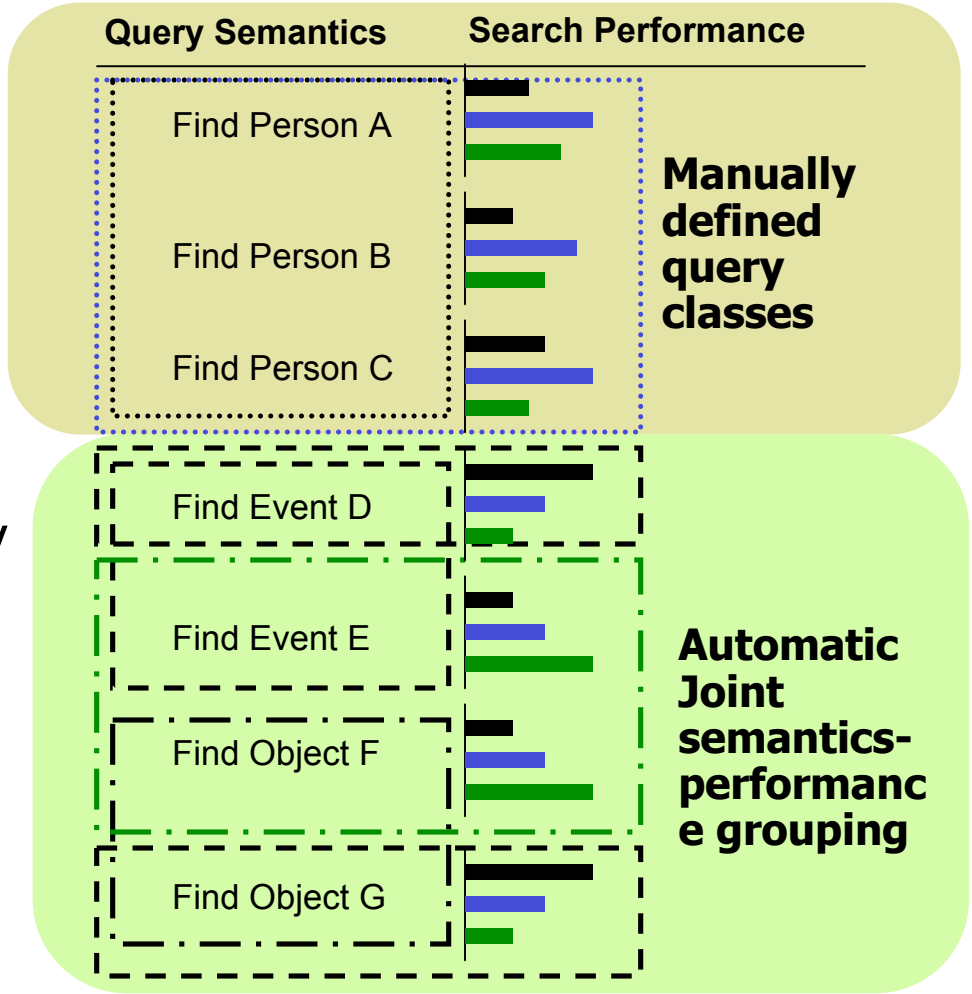
- **Improvement over story-based text search (in automatic search TRECVID 2005)**
  - **17% in MAP, 46% in soccer (171), 36% in helicopter (158), 32% in Blair (153), 28% in Abbas (154), etc.**
  - **No external search examples provided but discovered automatically**





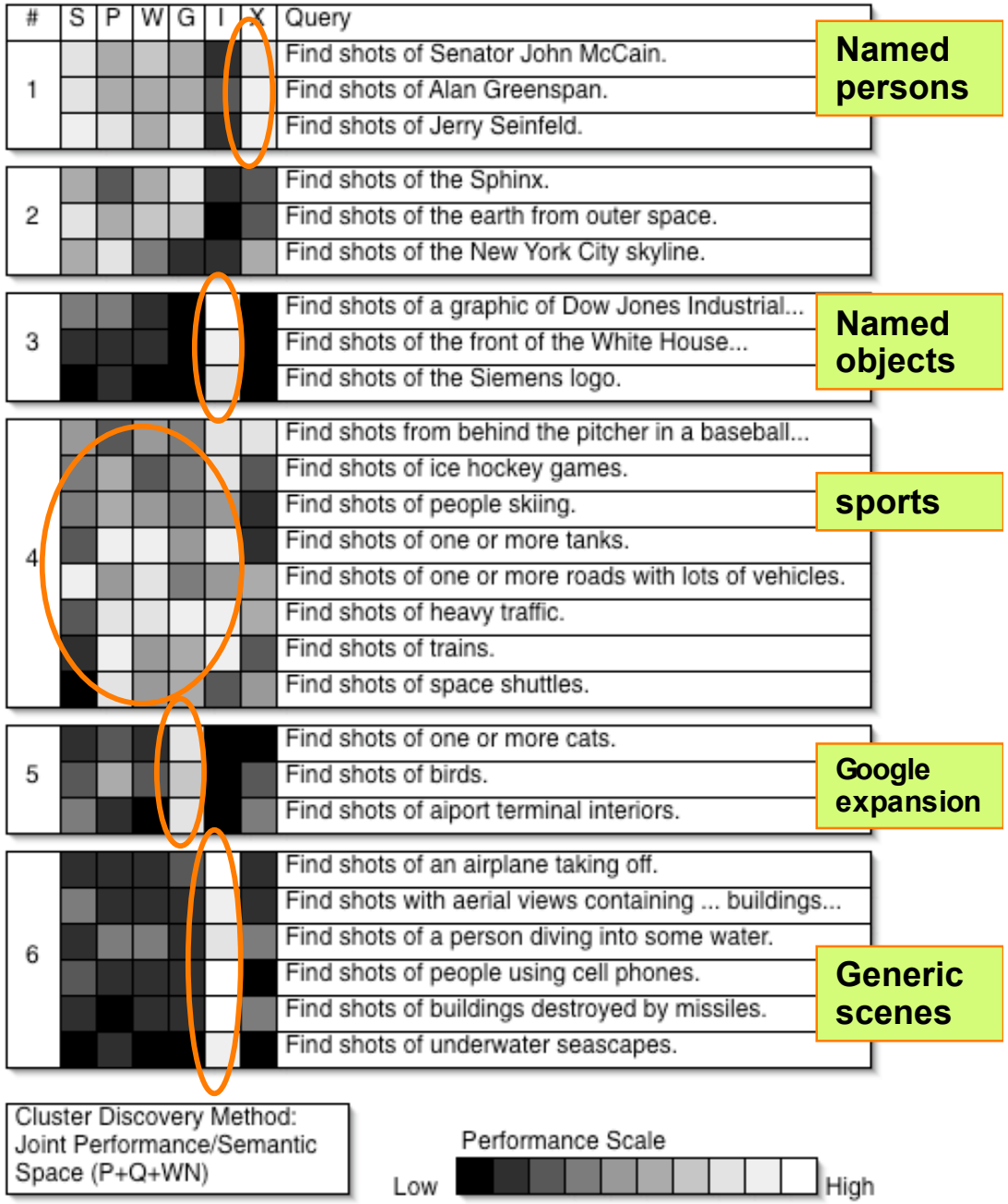
Automatic Discovery of Multimodal Query Classes

- Distinct query classes use customized fusion strategies
- How to automatically discover query classes?
- When and how does each modality help for each query?
- Existing methods: define query classes using human knowledge.
- New method: discover queries according to performance and semantics of searches.



Auto. Discovered Query Clusters

- Learned over a large query topic pool
- Text search and person-X
  - *named persons*
- Image search
  - *named objects*,
  - *sports*, and
  - *generic scene classes*
- Automated term expansion
  - *Google* class for cats, birds and airport terminals.



Interactive Activity Logging

Example Log Detail

```
- <videoSearchPerformances>
  - <videoSearchPerformance sysId="CU_FU" priority="1">
    - <searchResultSummary tNum="0167" searcherId="1" timeStart="Mon, 19 Sep 2005 17:17:54
      -0400" timeEnd="Mon, 19 Sep 2005 17:32:54 -0400" cuSearcherName="eric"
      cuMainSearch="Find shots of an airplane taking off" cuSelfId="648" elapsedTime="15"
      cuNistTruthId="982">
      - <raw>
          <new_search t_total="4" t_15_minutes="0" t_14_minutes="0" t_10_minutes="0"
          t_5_minutes="3" t_3_minutes="0" t_1_minute="0" t_30_seconds="1" t_0_seconds="0"/>
          <label_positive t_total="16" t_15_minutes="0" t_14_minutes="2" t_10_minutes="3"
          t_5_minutes="2" t_3_minutes="2" t_1_minute="7" t_30_seconds="0" t_0_seconds="0"/>
          <page_forward t_total="65" t_15_minutes="3" t_14_minutes="7" t_10_minutes="11"
          t_5_minutes="22" t_3_minutes="12" t_1_minute="10" t_30_seconds="0" t_0_seconds="0"/>
          <true_positive t_total="1" t_15_minutes="0" t_14_minutes="0" t_10_minutes="0"
          t_5_minutes="0" t_3_minutes="0" t_1_minute="1" t_30_seconds="0" t_0_seconds="0"/>
          <mode_change t_total="13" t_15_minutes="3" t_14_minutes="4" t_10_minutes="4"
          t_5_minutes="0" t_3_minutes="2" t_1_minute="0" t_30_seconds="0" t_0_seconds="0"/>
          <story_expand t_total="5" t_15_minutes="0" t_14_minutes="1" t_10_minutes="
          t_5_minutes="2" t_3_minutes="1" t_1_minute="0" t_30_seconds="0" t_0_seconds
          <query_remove t_total="3" t_15_minutes="0" t_14_minutes="0" t_10_minutes="0"
          t_5_minutes="3" t_3_minutes="0" t_1_minute="0" t_30_seconds="0" t_0_seconds
          <query_add t_total="4" t_15_minutes="0" t_14_minutes="0" t_10_minutes="0"
          t_5_minutes="4" t_3_minutes="0" t_1_minute="0" t_30_seconds="0" t_0_seconds="0"/>
          <label_negative t_total="3" t_15_minutes="0" t_14_minutes="0" t_10_minutes="0"
          t_5_minutes="3" t_3_minutes="0" t_1_minute="0" t_30_seconds="0" t_0_seconds="0"/>
          <page_backward t_total="3" t_15_minutes="0" t_14_minutes="2" t_10_minutes="1"
          t_5_minutes="0" t_3_minutes="0" t_1_minute="0" t_30_seconds="0" t_0_seconds="0"/>
        </raw>
```

Detailed search and topic criterion

Ground truth included in label actions

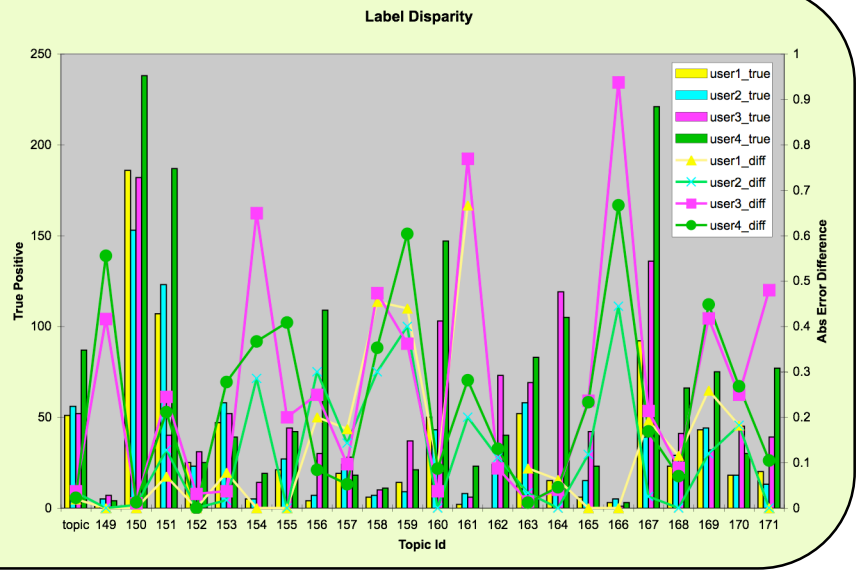
Aggregate tool actions by search time

```
- <dupall>
    <mode_change t_total="0.30769230769231" t_15_minutes="0.15384615384615"
    t_14_minutes="0" t_10_minutes="0.076923076923077" t_5_minutes="0"
    t_3_minutes="0.076923076923077" t_1_minute="0" t_30_seconds="0" t_0_seconds="0"/>
    <label_positive t_total="0.125" t_15_minutes="0" t_14_minutes="0" t_10_minutes="0.0625"
    t_5_minutes="0" t_3_minutes="0.0625" t_1_minute="0" t_30_seconds="0" t_0_seconds="0"/>
    <page_forward t_total="0.10769230769231" t_15_minutes="0.015384615384615"
    t_14_minutes="0" t_10_minutes="0.046153846153846" t_5_minutes="0"
    t_3_minutes="0.046153846153846" t_1_minute="0" t_30_seconds="0" t_0_seconds="0"/>
  </dupall>
- <story>
    <mode_change t_total="0.69230769230769" t_15_minutes="0.076923076
    t_14_minutes="0.30769230769231" t_10_minutes="0.23076923076923" t_
    t_3_minutes="0.076923076923077" t_1_minute="0" t_30_seconds="0" t_
    <story_expand t_total="1" t_15_minutes="0" t_14_minutes="0.2" t_10_m
    t_5_minutes="0.4" t_3_minutes="0.2" t_1_minute="0" t_30_seconds="0"
```

Monitor labeling to understand interface usage

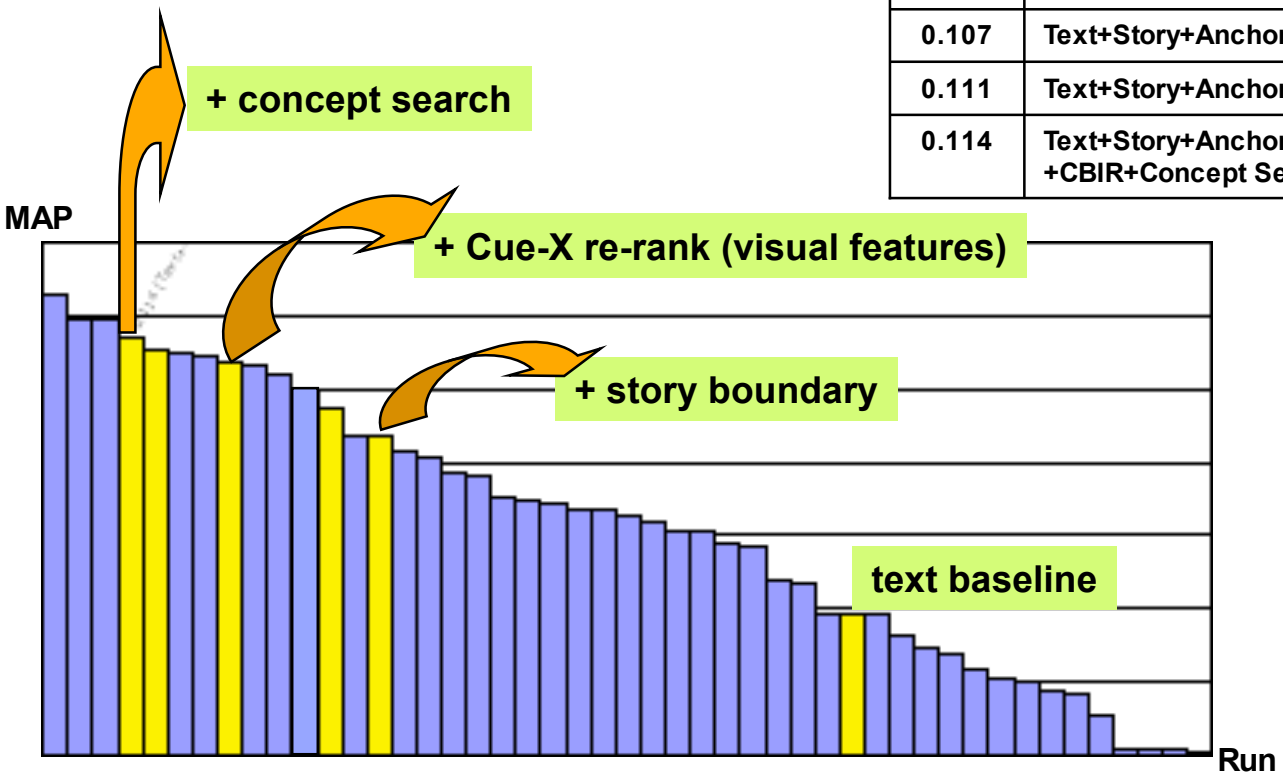
Post-Mortem Analysis

- Analyze inter-labeler disparity
- Find difficult search topics by high common error rate
- Discover where certain tools failed
- In the future, use actions as passive relevance feedback rounds



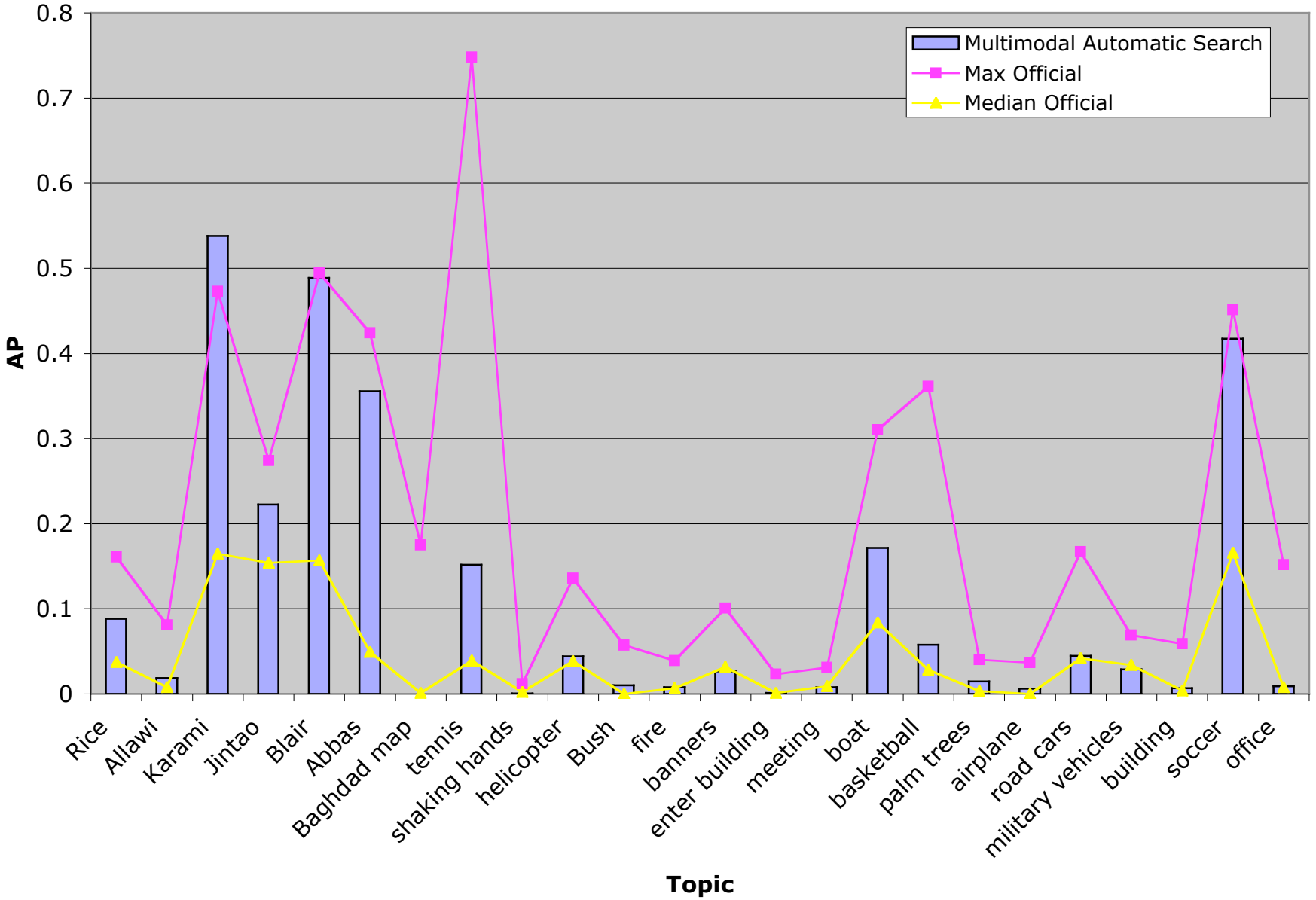
Automatic Search (Performance Breakdown)

MAP	Components
0.039	Text
0.087	Text+Story
0.095	Text+Story+Anchor Removal
0.107	Text+Story+Anchor Removal +CueX Re-rank
0.111	Text+Story+Anchor Removal +CueX Re-rank +CBIR
0.114	Text+Story+Anchor Removal +CueX Re-rank +CBIR+Concept Search



- Largest improvement from story segmentation
- Noticeable improvements from other components
  - especially cue-x rerank and concept search

Automatic Search



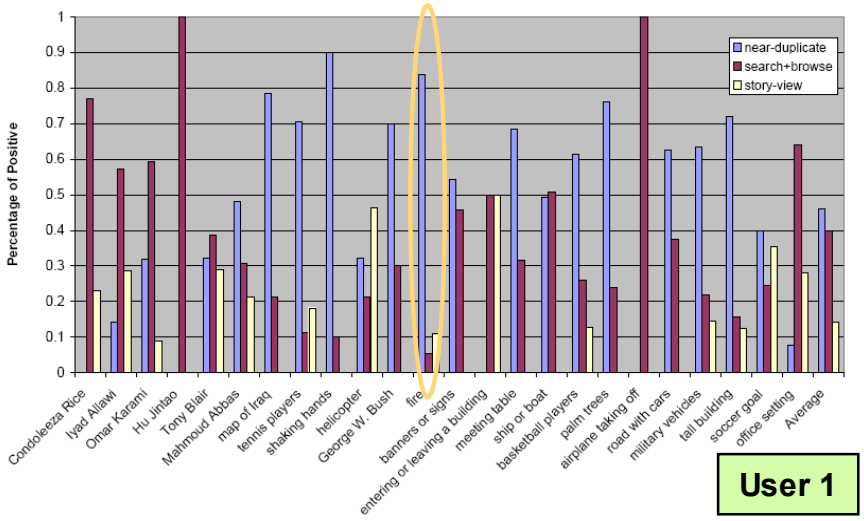
Interactive Tool Contribution

Varied search strategies

- User 1: prefers story browsing, duplicate and traditional search
- User 2: no story discovery, use lots of duplicate browsing

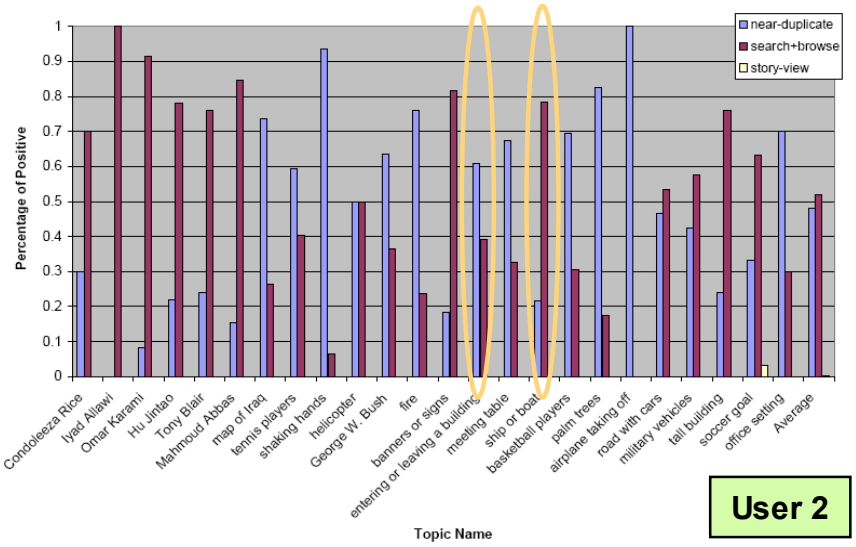
Strategy dynamic for each topic

- Common visual concepts good candidates for duplicates
- Temporal events best suited for discovery by story browsing
- Named entities or specific actions usually best in traditional search methods



User 1

Top-ranking interactive searches



User 2

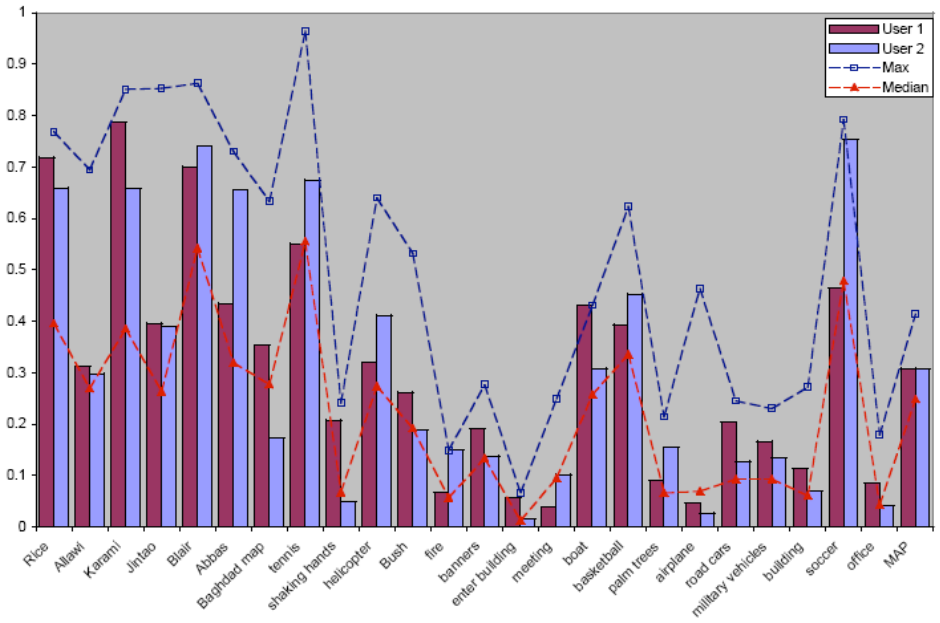
Interactive Search

Best Overall Performance

160 (fire), 164 (boat), and
162 (entering building)

Close to Best

149 (Rice), 151 (Karami),
153 (Blair), 154 (Abbas),
157 (shaking hands),
161 (banners),
166 (palm trees),
168 (roads/cars),
169 (military vehicles), and
171 (soccer)



Formula for Success:

1. Find positives through any search method
2. Iteratively browse through the near-duplicates or story browsing