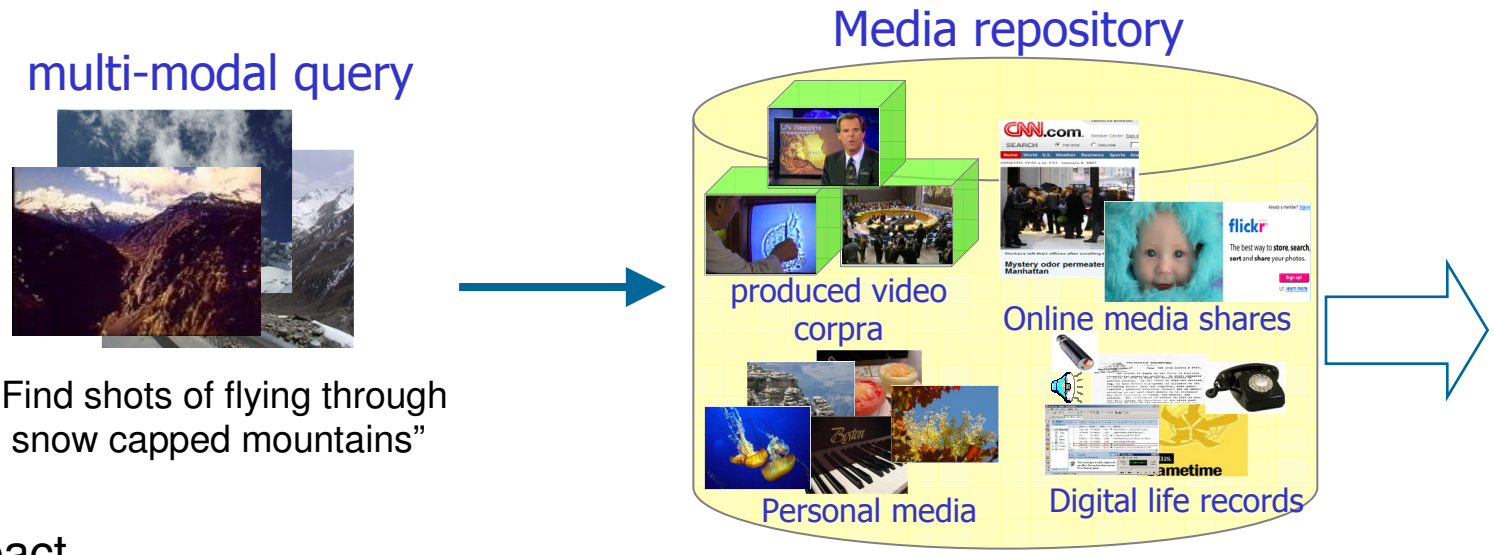
Dynamic Multimodal Fusion in Video Search

Lexing Xie
IBM T J Watson Research Center
joint work with Apostol Natsev, Jelena Tesic, Rong Yan, John R Smith

WOCC-2007, April 28th, 2007

The Multimedia Search Problem



- Impact
  - Widely applicable: consumer, media, enterprise, web, science …
  - Bridging Traditional Search and Multimedia Knowledge Extraction
- Multiple Tipping Points: multimedia semantics learning, multimedia ontology, training and evaluation, …
- NIST TRECVID benchmark
  - Validation for emerging technologies
  - Scaling video retrieval technologies to large-scale applications

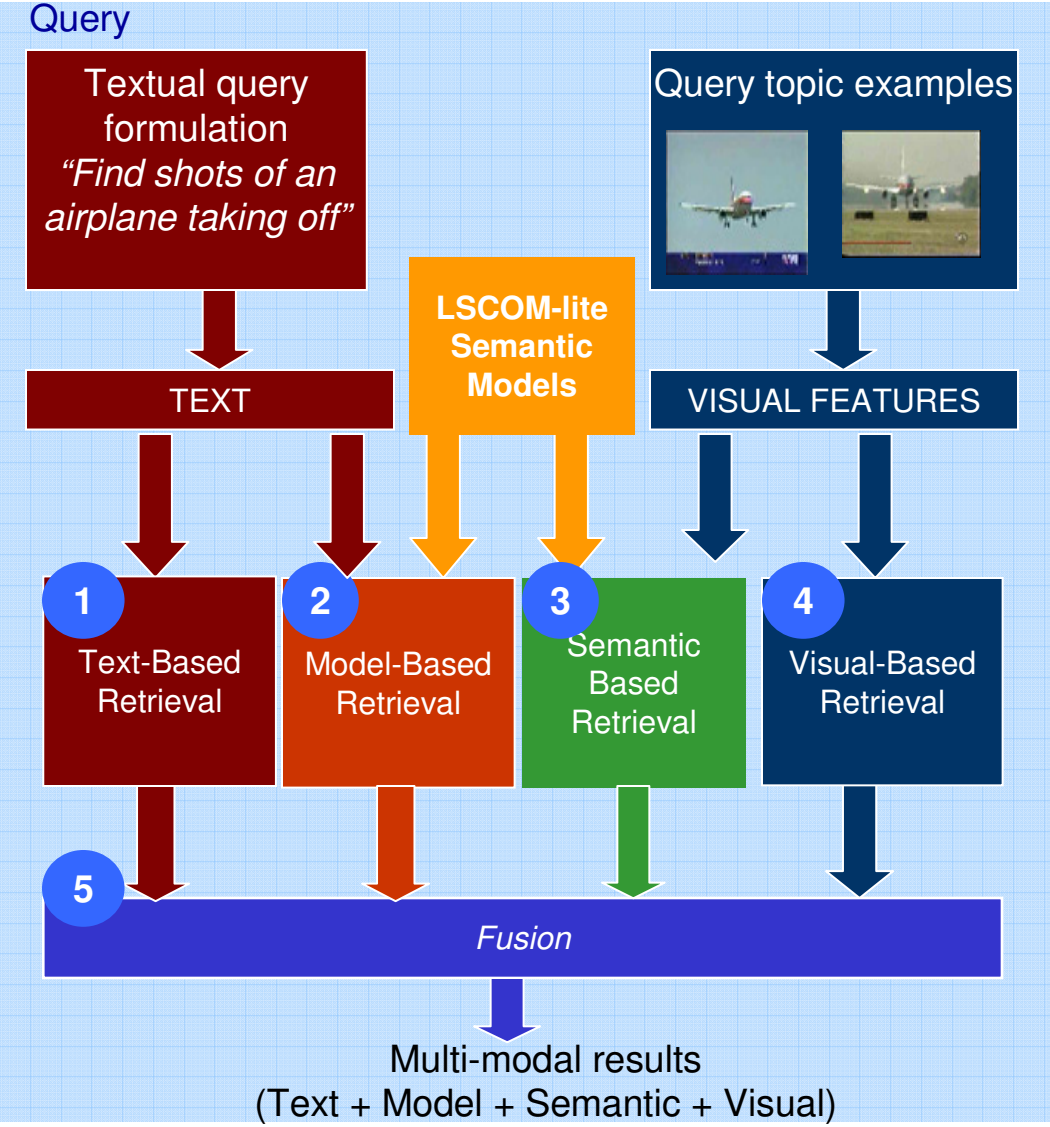
Multimedia Search: Query Topics overview

Query Types	Query topic	Query Specificity	
		Generic	Specific/Named
Find objects	 *Topic 13. Speaker talking in front of the US flag*	Palm trees, tanks, ship/boat	
Find people	 *Topic 4. Scenes of snow capped mountains*	People shaking hands, people with banners, people in a meeting, people entering or leaving a building	George W. Bush, Condoleeza Rice, Tony Blair, Iyad Allawi, Omar Karami, Hu Jintao, Mahmoud Abbas
Find events	 *Topic 48. Other examples of overhead zooming-in views of canyons in Western United States.*	Soccer score, basketball, tennis, airplane take-off, helicopter in flight	
Find sites		Tall buildings, office setting, road with cars, fire	Map of Iraq showing Baghdad

Outline

- The multimedia search challenge
- A birds-eye's view of IBM Multimedia Search System
- Query-dependent multimodal fusion
- Evaluation on TRECVID benchmark
- Summary

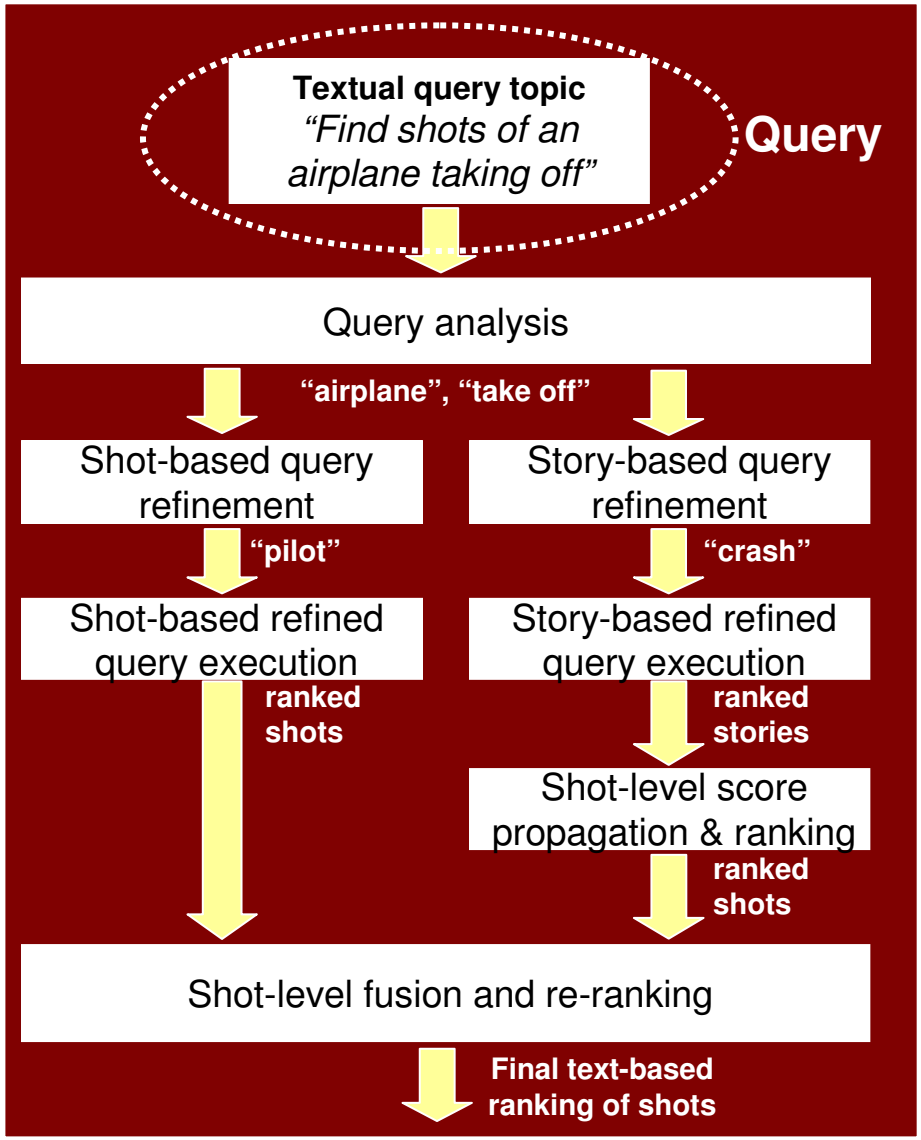
IBM Multimedia Search System Overview



Approaches:

1. Text-based: story-based retrieval with automatic query refinement/re-ranking
2. Model-based: automatic query-to-model mapping based on query topic text
3. Semantic-based: cluster-based semantic space modeling and data selection
4. Visual-based: light-weight learning (discriminative and nearest neighbor modeling) with smart sampling
5. **Fusion:**
   - Query independent
   - **Query-class-dependent with soft, hard or dynamic class membership**

1 IBM Text Retrieval System



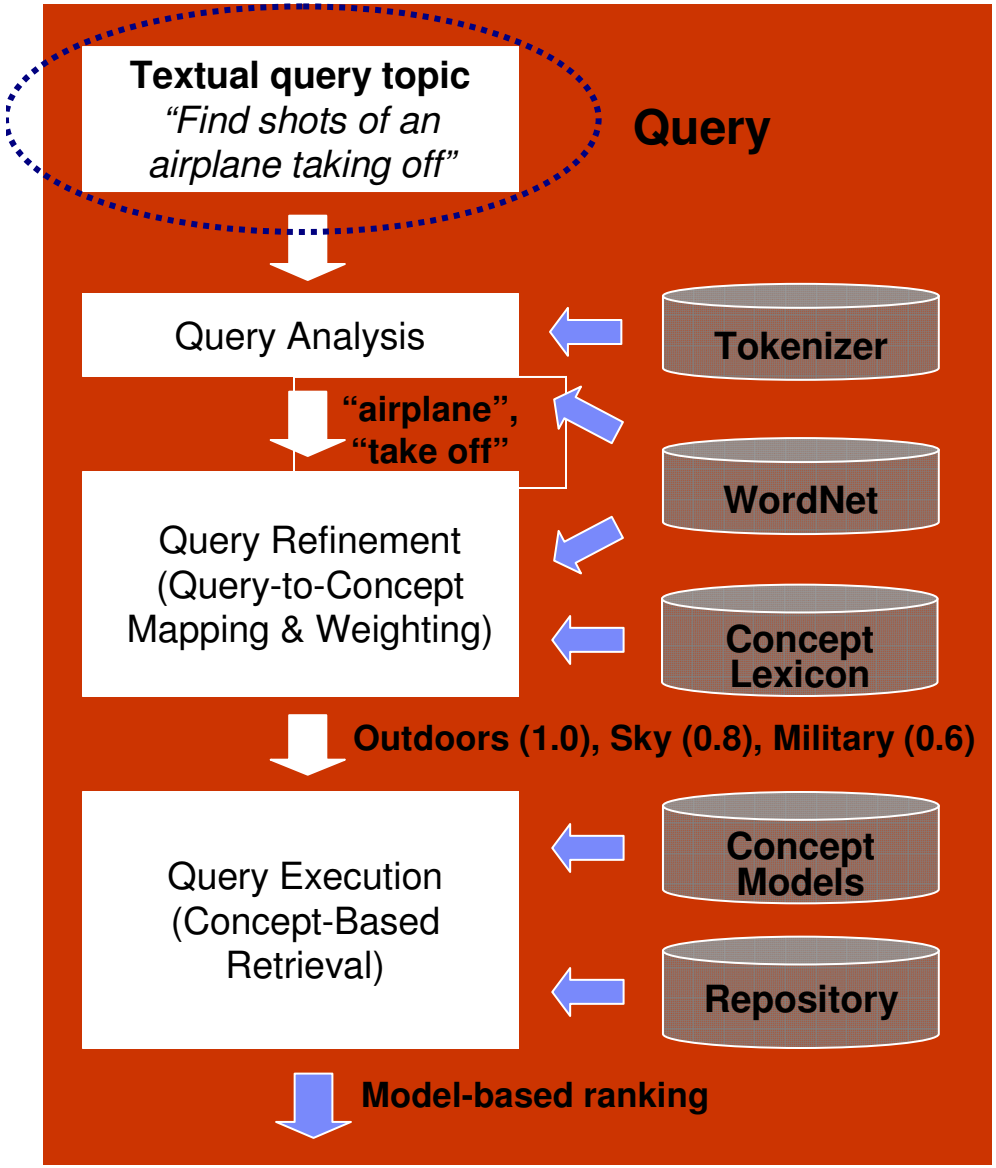
- Corpus Indexing
  - Shot-level ASR/MT documents
  - Story-level ASR/MT documents
  - Both aligned at phrase level
- Query analysis:
  - Tokenization, phrasing, stemming
  - Part-of-speech tagging & filtering
- Query refinement:
  - Pseudo-Relevance Feedback
- Query execution and fusion
  - IBM Semantic Search Engine (Juru)
  - Using TF*IDF-based retrieval
  - Fusion of shot- and story-based results

Performance (MAP):

- 2005: 0.09873
- 2006: 0.05169

T. Volkmer et al, ICME 2006

2 IBM Model-based Retrieval System



Lexical (WordNet) Approach

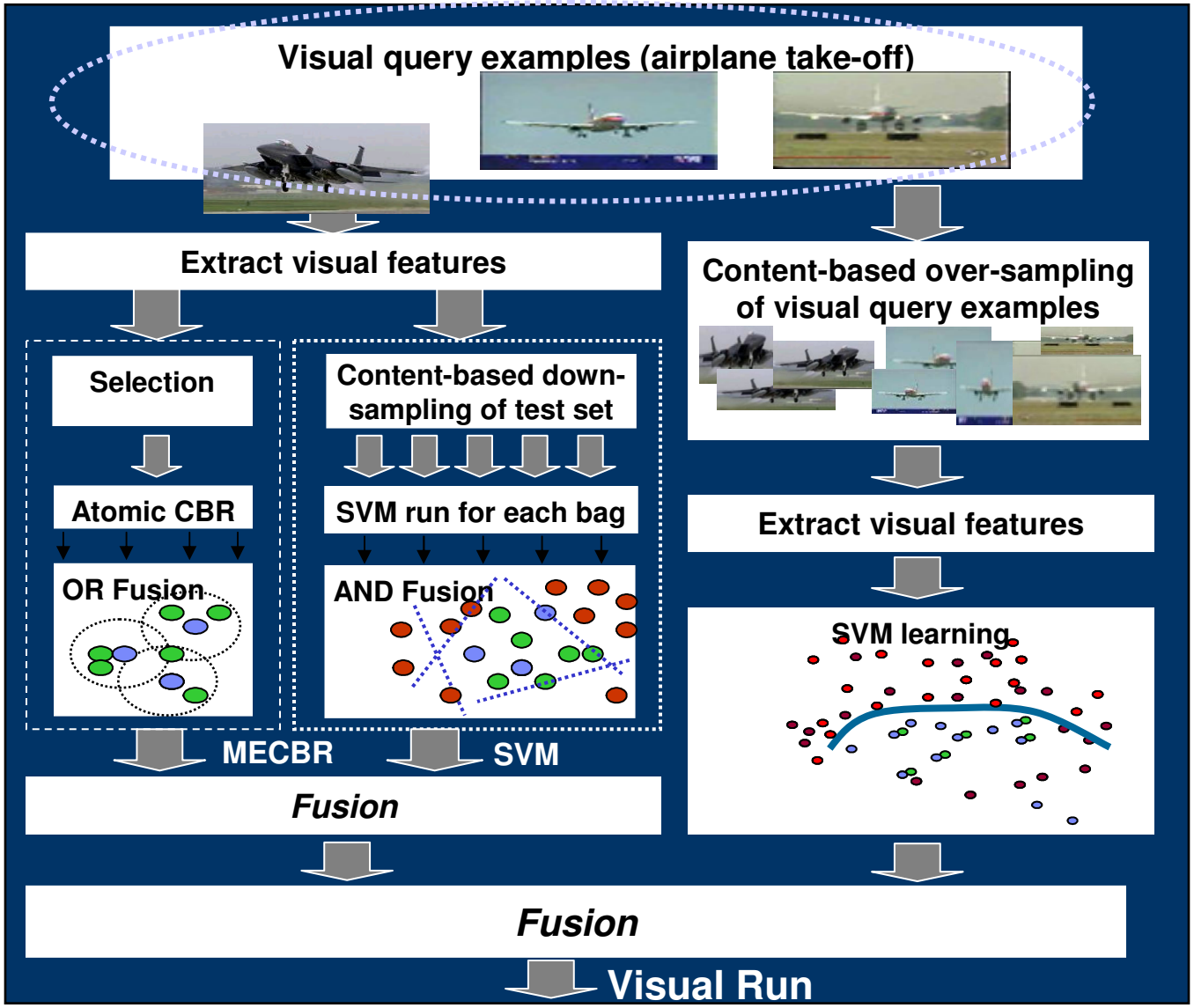
- Query analysis
  - Tokens, stems, phrases
  - Stop word removal
- Query refinement
  - Automatic mapping of query text to concept models & weights
  - Lexical approach based on WordNet Lesk similarity
- Query execution
  - Concept-based retrieval using statistical concept models
  - Weighted averaging model fusion

Performance (MAP)

- 2006: 0.029

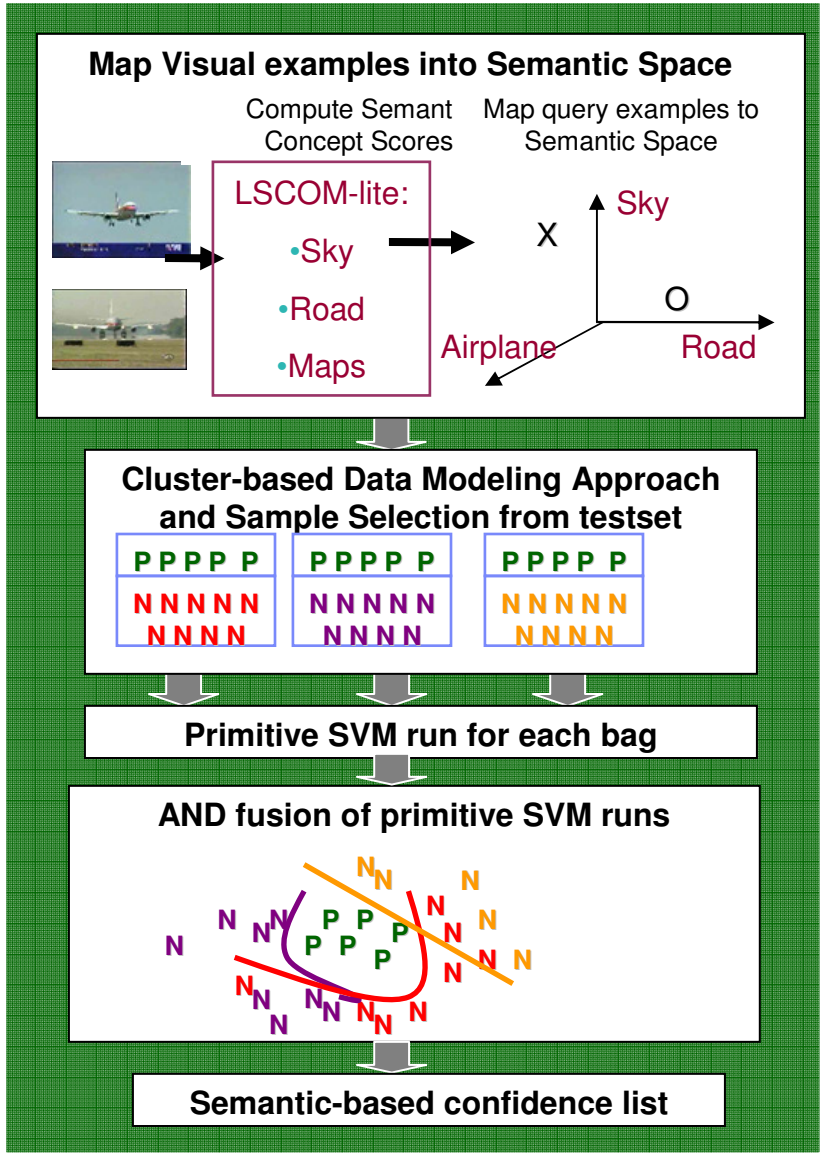
Haubold et al. (ICME 2006)

4 IBM Visual Retrieval System



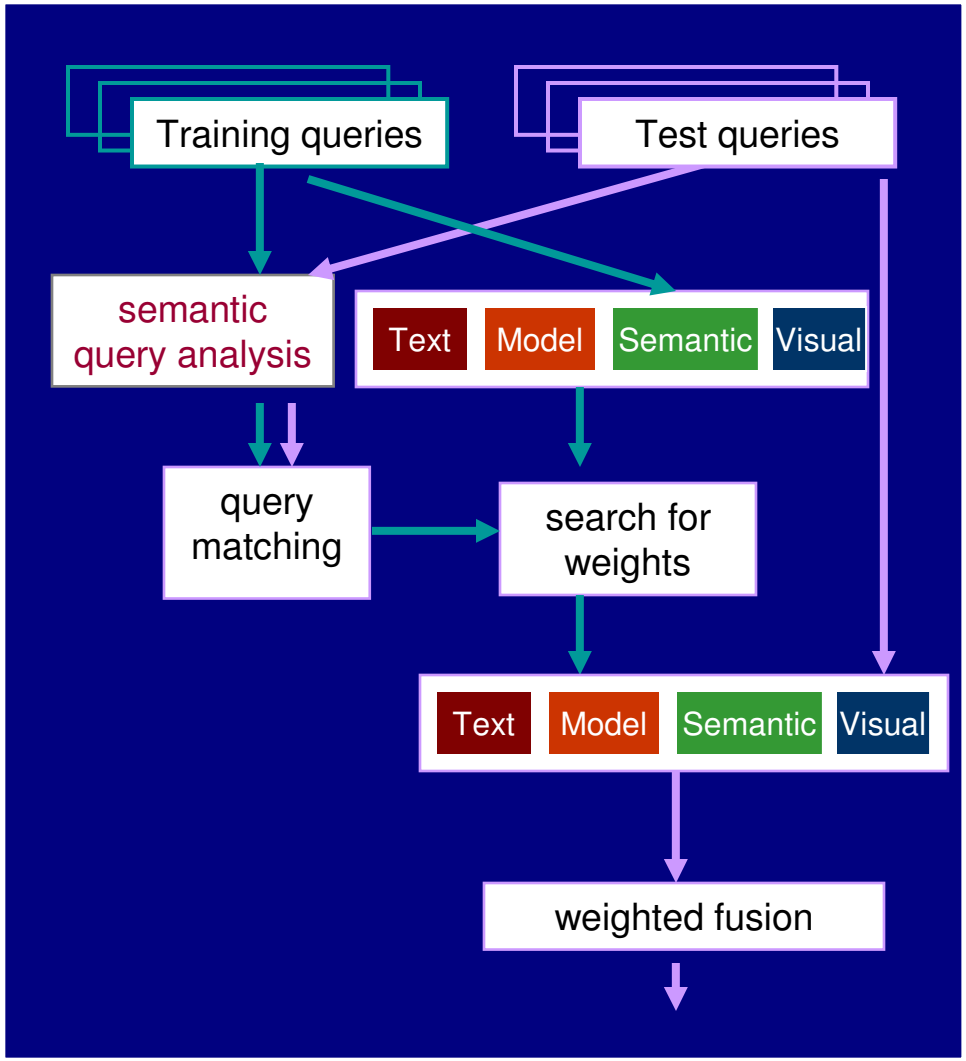
- Visual concepts can help refine query topics
- Data modeling for discriminative learning:
  - Content-based over-sampling of visual query to create positive examples
  - Content –based pseudo-negative sampling in the development and test set.

3 IBM Visual-Semantic Retrieval



- Observations
  - Visual context can disambiguate word senses
- Idea
  - Leverage automatic visual concept detectors for semantic model-based query refinement
- Semantic Concept Lexicon
  - Hierarchy of 39 LSCOM-lite concepts
  - Statistical models based on visual features and machine learning
  - Dimensions correspond to LSCOM-lite
- We model query topic examples in semantic space
  - Any descriptor space

5 Multi-modal Fusion



- Each modality is good in finding certain things
  - Text: named people, other named entity
  - Visual: semantic scenes consistent in color or layout, e.g., sports, weather
  - Concept Models: non-specific queries, e.g. protest, boat, fire
- Averaging the retrieval models helps in any case
- Query-class or query-cluster dependent fusion helps more [CMU, NUS, Columbia]
- Query soft/hard/dynamic class approach
  - Semantic query analysis for query matching
  - Matching based on PIQUANT-II Q&A text features)
  - Weighted linear combination for fusion
  - Training Queries: TRECVID2005

Performance MAP

- 2006: 0.0756 -> 0.087 -> 0.0937

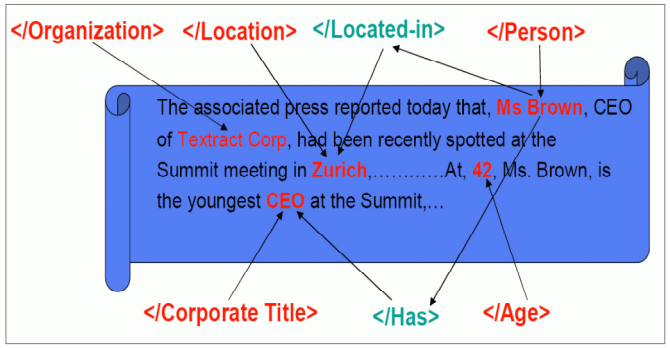
Extract Semantic Query Features

Input Query Text → Semantic Tagging → Semantic Categories → Query Feature Vector

"people with" computer display

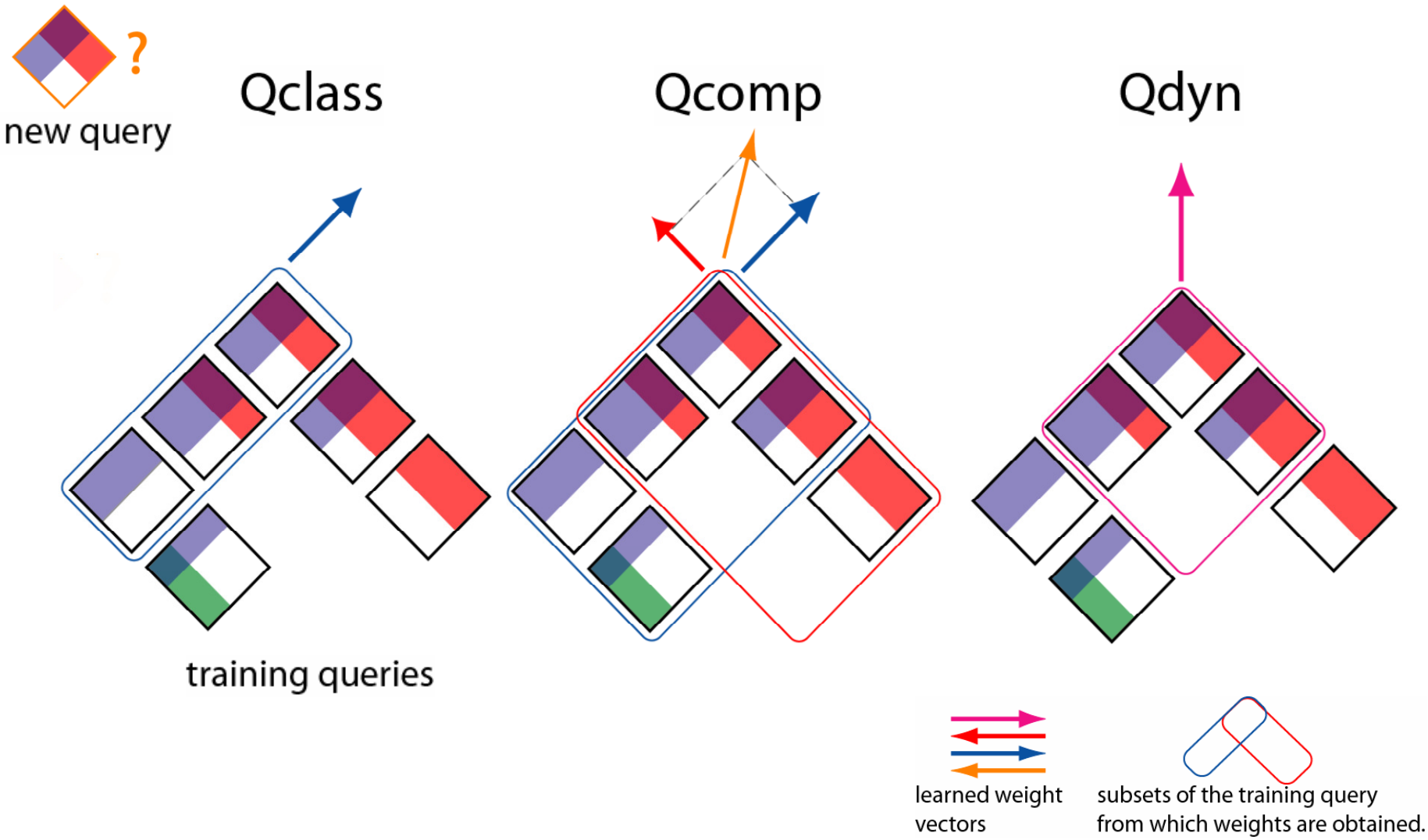
Person:CATEGORY
BodyPart:UNKNOWN
Furniture:UNKNOWN

...



[IBM UIMA and PIQUANT Analysis Engine]

Three Query-dependent Fusion Strategies



Outline

- The multimedia search challenge
- A birds-eye's view of IBM Multimedia Search System
- Query-dependent multimodal fusion
- Evaluation and demo on TRECVID benchmark
- Summary

NIST TRECVID Benchmark at a Glance

- NIST benchmark for evaluating state-of-the-art in video retrieval
- Benchmark tasks:
  - Semantic Concept Detection; Search (2003-); Rushes (2005-)
  - Shot detection and story segmentation

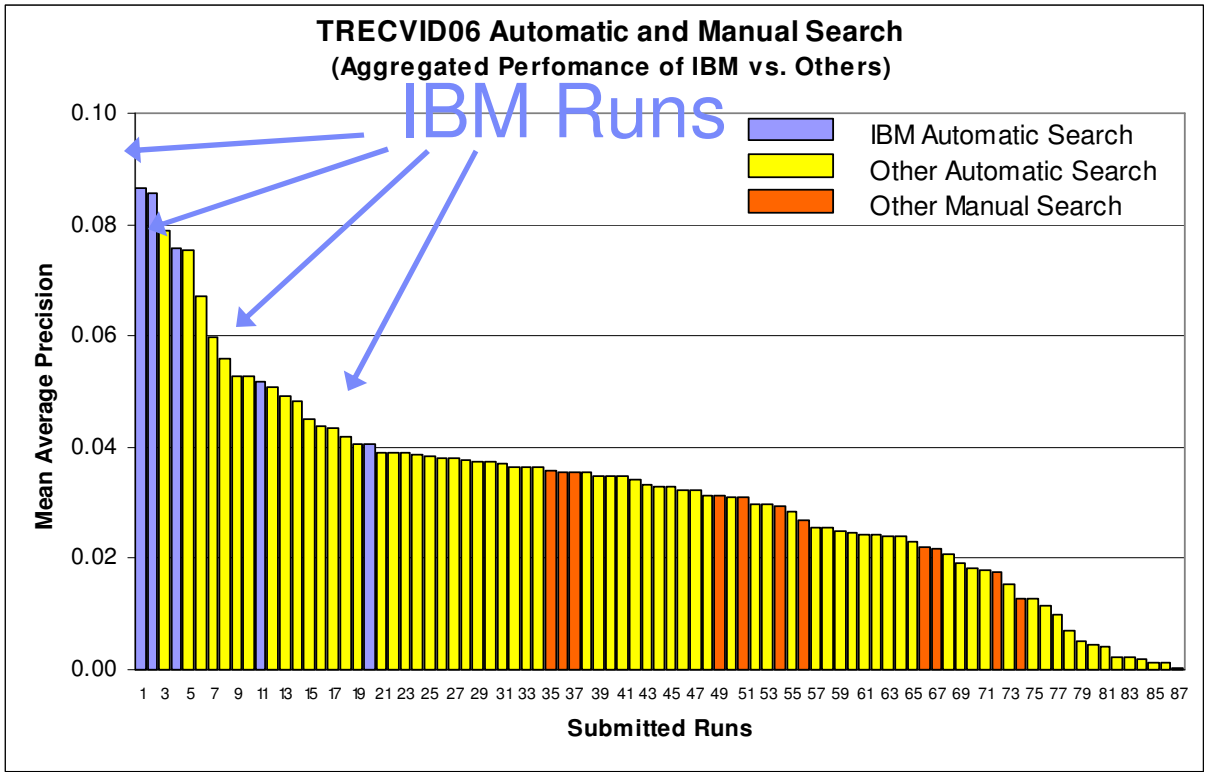
Growing Participation

TRECVID 2001	TRECVID 2002	TRECVID 2003	TRECVID 2004	TRECVID 2005	TRECVID 2006
12* Participants	17 Participants	24 Participants	38 Participants	42* Participants	54* Participants
11 Hours NIST video	73 Hours Video from Prelinger archives	133 Hours 1998 ABC, CNN news & C-SPAN	173 Hours 1998 ABC, CNN news & C-SPAN	220 Hours of 2004 news from U.S., Arabic, Chinese sources 50 hours BBC stock shots (vacation spots)	380 Hours of 2004 and 2005 news from U.S., Arabic, Chinese source 100 hours BBC rushes (interviews)

Growing Data Sets

* Number of participants that completed at least one task

Automatic/Manual Search Overall Performance (Mean AP)



IBM Official Runs:

Text (baseline):	**0.041**
Text (story-based):	**0.052**
Multimodal Fusion:	
Query independent:	**0.076**
Query classes (soft):	**0.086**
Query classes (hard):	**0.087**

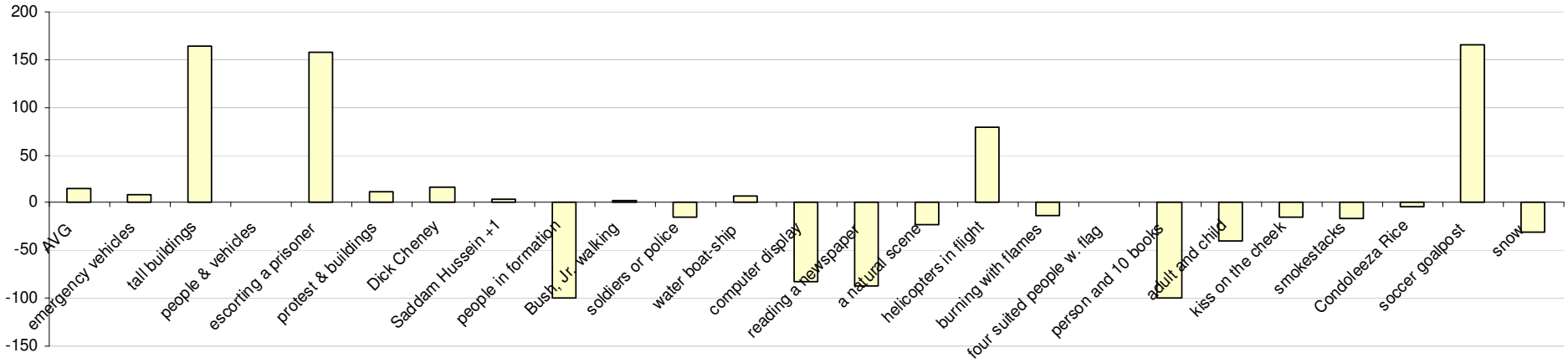
IBM Modal Runs:

Semantic-based Run:	**0.037**
Visual-based Run:	**0.0696**

- Multi-modal fusion doubles baseline performance!
- Visual and semantic retrieval have a significant impact over a text baseline

TRECVID'06 Fusion Results (Relative Performance)

- Relative improvement (%)
  - query-class fusion vs. query-independent fusion
- Observations Qind→Qclass
  - Concept-related queries improved the most:
    "tall building", "prisoner", "helicopters in flight", "soccer"
  - Named-entity queries improved slightly:
    "Dick Cheney", "Saddam Hussein", "Bush Jr."
  - Generic people category deteriorated the most:
    "people in formation", "at computer display", "w/ newspaper", "w/ books"

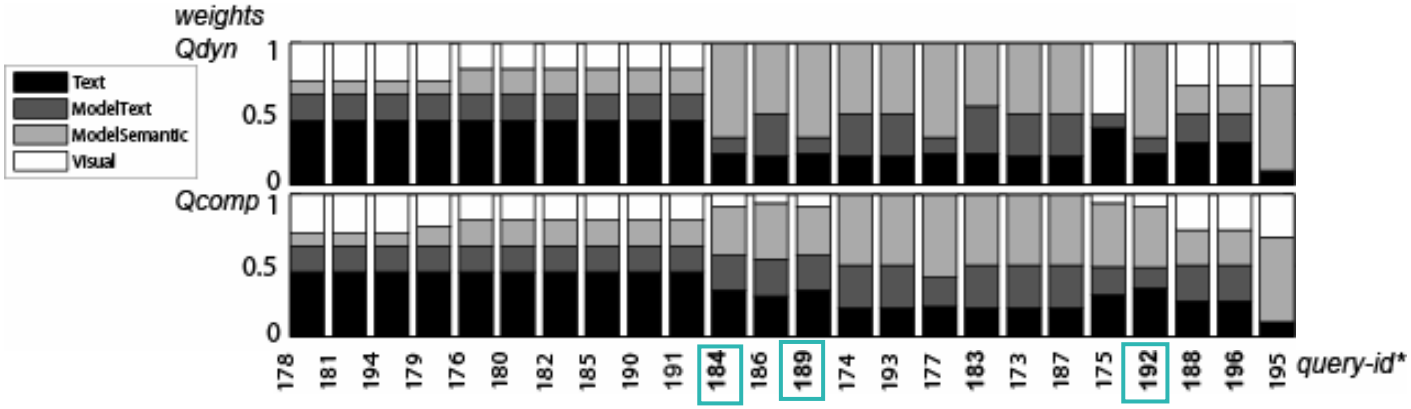


Improvement over TREC06' and TREC'05

Qdyn and *Qcomp* > *Qclass* >> *Qind*

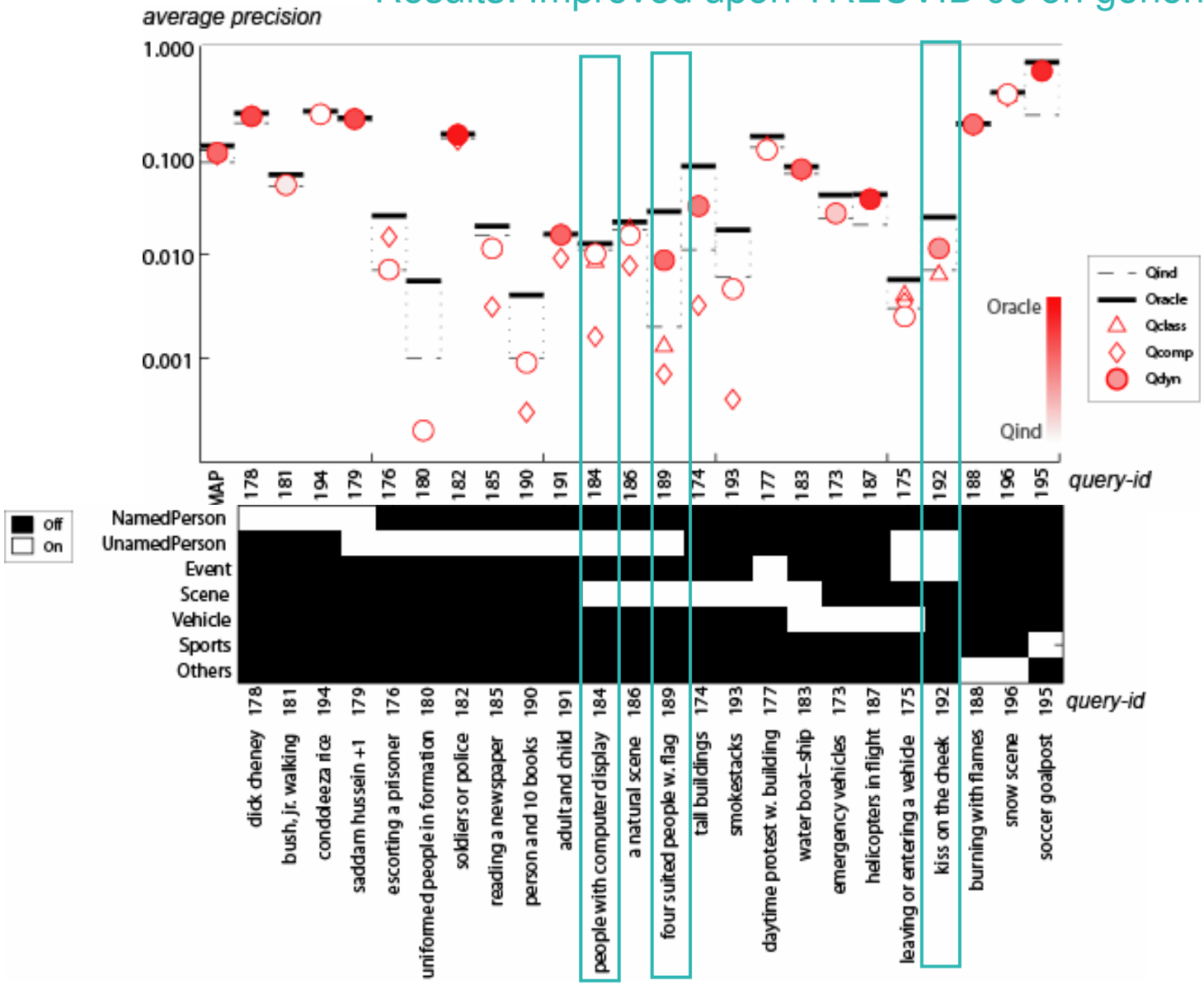
dataset	method	Qind	Oracle	Qclass	Qcomp	Qdyn
trec06	all-queries	0.0756	0.1089	0.0897	0.0929	0.0937
	non-sports	0.0695	0.0835	0.0686	0.0720	0.0728
trec05	all-queries	0.1681	0.2184	0.1591	0.1621	0.1604
	non-sports	0.1263	0.1681	0.1291	0.1326	0.1307

New improvements come from the increased weights on concepts for object queries (next slide).

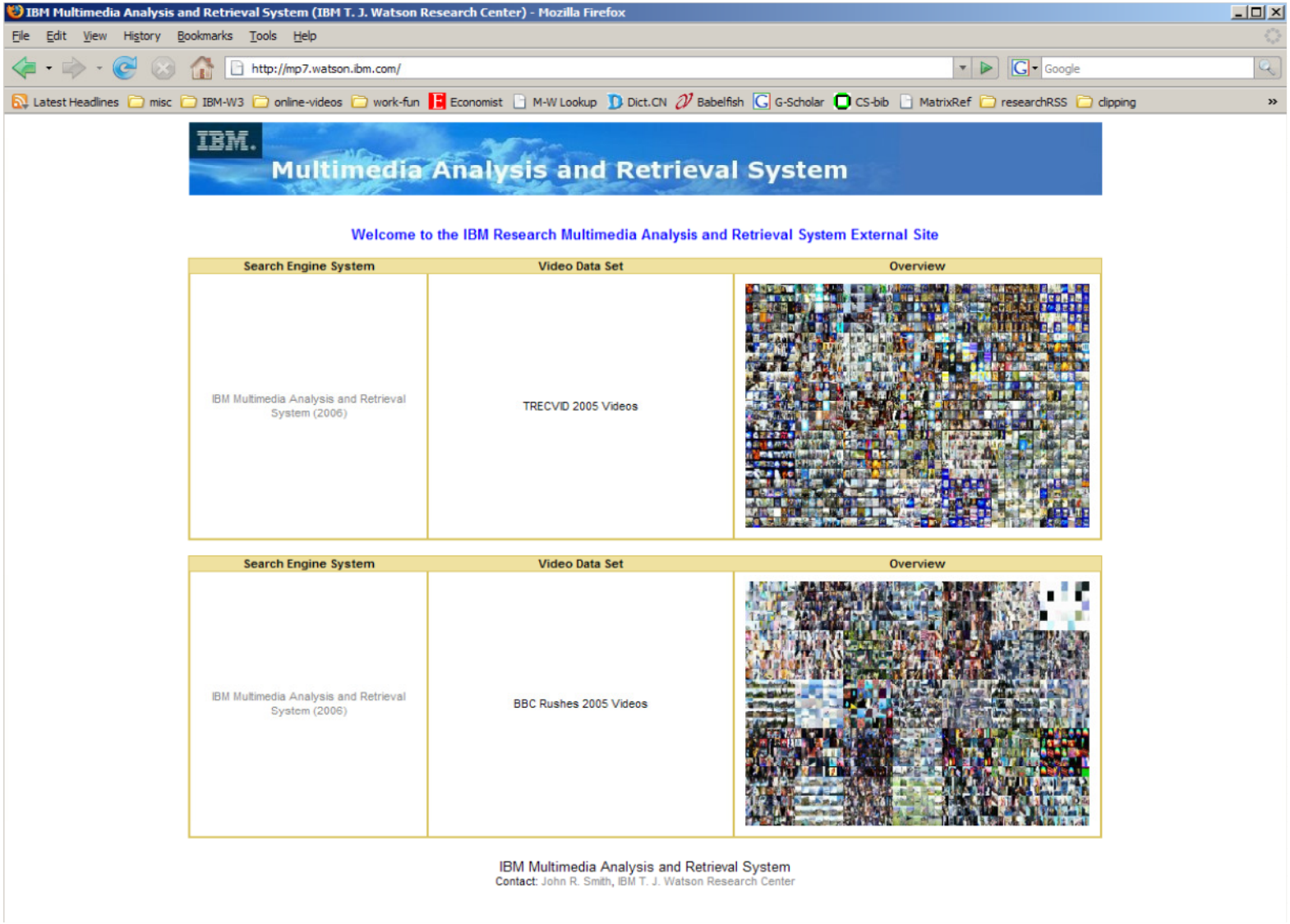


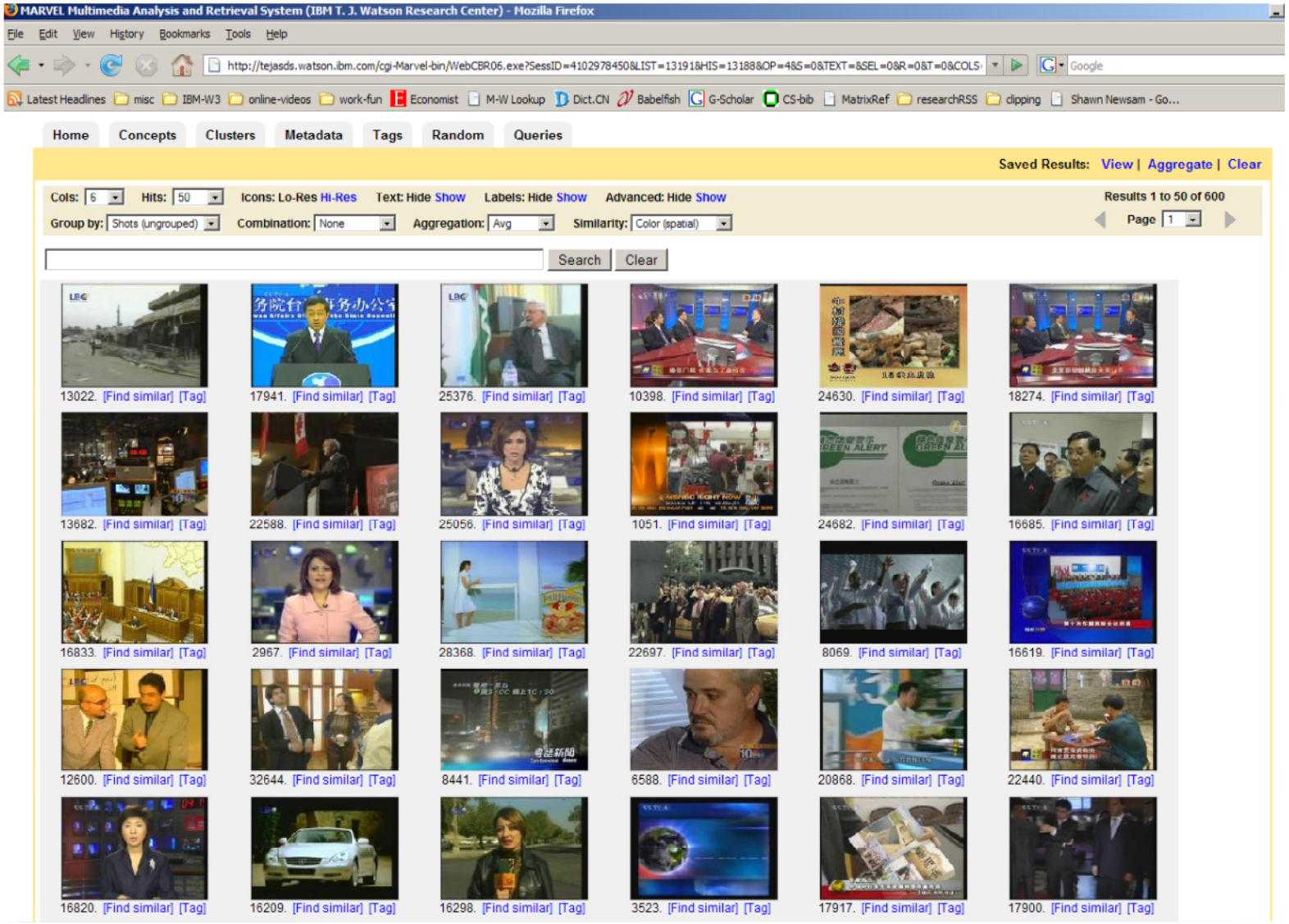
Improved queries …

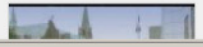
Results: Improved upon TRECVID'06 on generic people queries.



Demo







Cols: 6 Hits: 50 Icons: Lo-Res Hi-Res Text: Hide Show Labels: Hide Show Advanced: Hide Show

Results 1 to 50 of 112

Group by: Shots (ungrouped) Combination: None Aggregation: Avg Similarity: Color (spatial)

Page 1

soccer worldcup Search Clear Query terms (add all): soccer | worldcup

Refine with: FACE | ANIMAL | MILITARY | OFFICE | SNOW

Filter by VIDEOS: TRECVID2005 28 (26) | TRECVID2005 48 (10) | TRECVID2005 13 (7) | TRECVID2005 97 (4) | TRECVID2005 140 (3)

32459. [More like this]
China's National Football Team in World Cup soccer qualifying competition being eliminated after Chinese officials that they should all the blame to coach Arica artists body but also confirmed that officials in the...

32458. [More like this]
China's National Football Team in World Cup soccer qualifying competition being eliminated after Chinese officials that they should all the blame to coach Arica artists body but also confirmed that officials in the...

32457. [More like this]
China's National Football Team in World Cup soccer qualifying competition being eliminated after Chinese officials that they should all the blame to coach Arica artists body but also confirmed that officials in the...

32460. [More like this]
China's National Football Team in World Cup soccer qualifying competition being eliminated after Chinese officials that they should all the blame to coach Arica artists body but also confirmed that officials in the...

[0.979] [More like this]
China's national soccer team to the World Cup soccer qualifying competition regret that... ...Football Team in World Cup soccer qualifying competition being eliminated after Chinese officials that they should all the blame to coach...

[0.979] [More like this]
China's national soccer team to the World Cup soccer qualifying competition regret that... ...Football Team in World Cup soccer qualifying competition being eliminated after Chinese officials that they should all the blame to coach...

[0.979] [More like this]
China's national soccer team to the World Cup soccer qualifying competition regret that... ...Football Team in World Cup soccer qualifying competition being eliminated after Chinese officials that they should all the blame to coach...

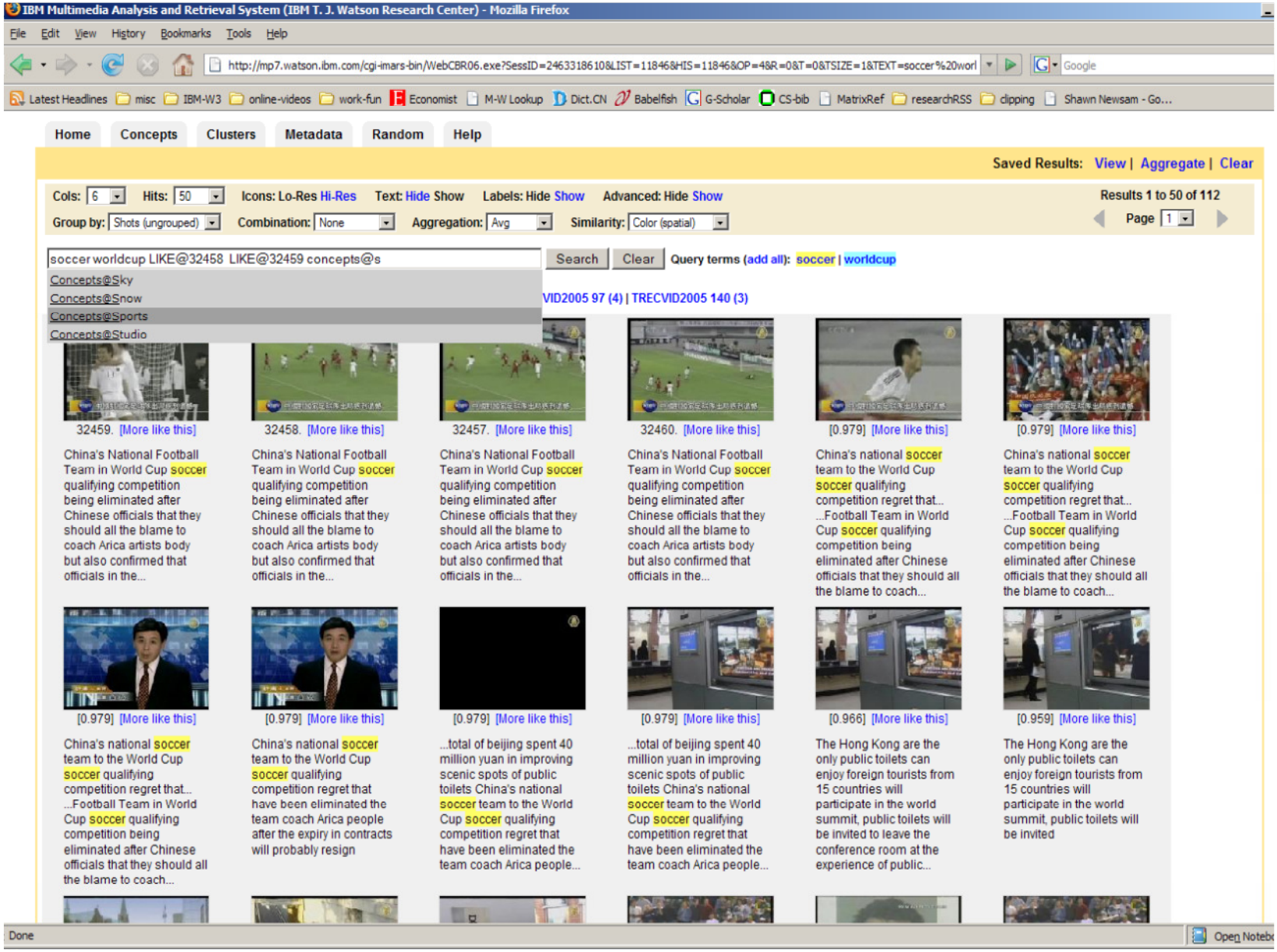
[0.979] [More like this]
China's national soccer team to the World Cup soccer qualifying competition regret that have been eliminated the team coach Arica people after the expiry in contracts will probably resign

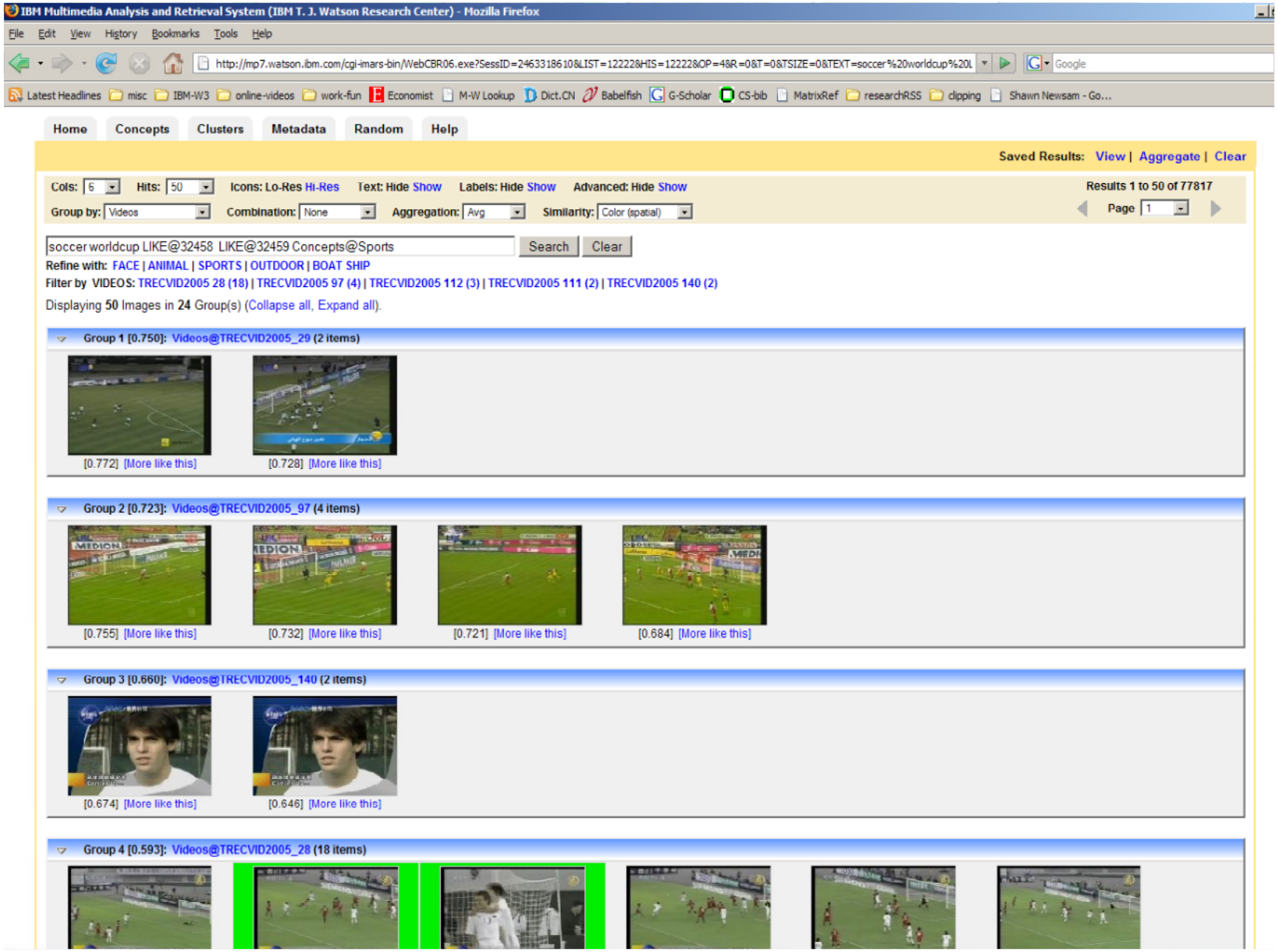
[0.979] [More like this]
...total of beijing spent 40 million yuan in improving scenic spots of public toilets China's national soccer team to the World Cup soccer qualifying competition regret that have been eliminated the team coach Arica people...

[0.979] [More like this]
...total of beijing spent 40 million yuan in improving scenic spots of public toilets China's national soccer team to the World Cup soccer qualifying competition regret that have been eliminated the team coach Arica people...

[0.966] [More like this]
The Hong Kong are the only public toilets can enjoy foreign tourists from 15 countries will participate in the world summit, public toilets will be invited to leave the conference room at the experience of public...

[0.959] [More like this]
The Hong Kong are the only public toilets can enjoy foreign tourists from 15 countries will participate in the world summit, public toilets will be invited





Thank You

For more information:

- IBM Multimedia Retrieval Demo
  http://mp7.watson.ibm.com/
- "Dynamic Multimodal Fusion for Video Search", by L. Xie, A. Natsev, J. Tesic, ICME 2007, to appear
- "IBM research TRECVID-2006 video retrieval system", by M. Campbell, S. Ebadollahi, M. Naphade, A. P. Natsev, J. R. Smith, J. Tesic, L. Xie, K. Scheinberg, J. Seidl, and A. Haubold, in *NIST TRECVID Workshop*, November 2006.