Dynamic Multimodal Fusion in Video Search

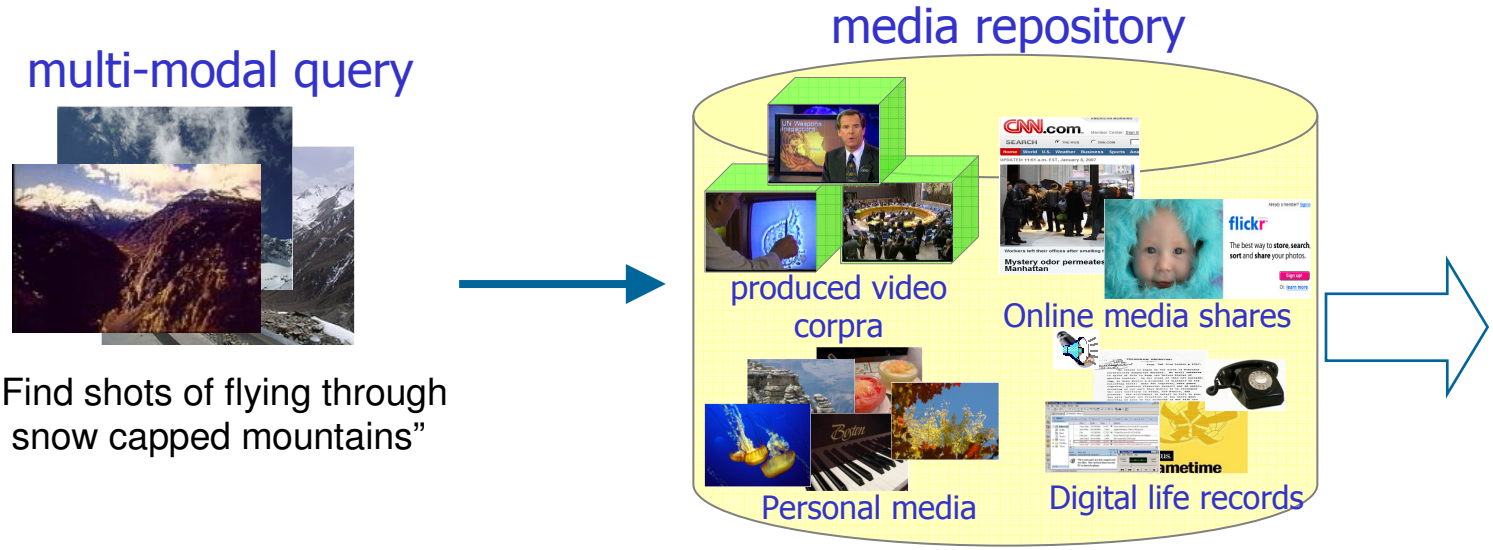
Lexing Xie, Apostol Natsev, Jelena Tesic

IBM T. J. Watson Research Center

July 5th, ICME 2007, Beijing

© 2007 IBM Research

The Multimedia Search Problem



- Wide impact
  - consumer, media, enterprise, web, science …
- Multiple Tipping Points:
  - Multimedia semantics learning,
  - Multimedia ontology,
  - Systematic benchmark and evaluations

© 2007 IBM Research

The Challenge of Choices in Multimodal Search

queries / retrieval experts	*Find shots of Condoleezza Rice*	*Find shots of soccer game with goalpost visible*	*Find scenes of snow capped mountains*
Text	...which he was accompanied by his wife are trying to light and Foreign Minister who resigned as successor Condoleezza Rice, Bush was greeted at the airport Governor-General of	?	??	??
Image similarity	~	??	?	??
Semantic concepts		??	??	?

© 2007 IBM Research

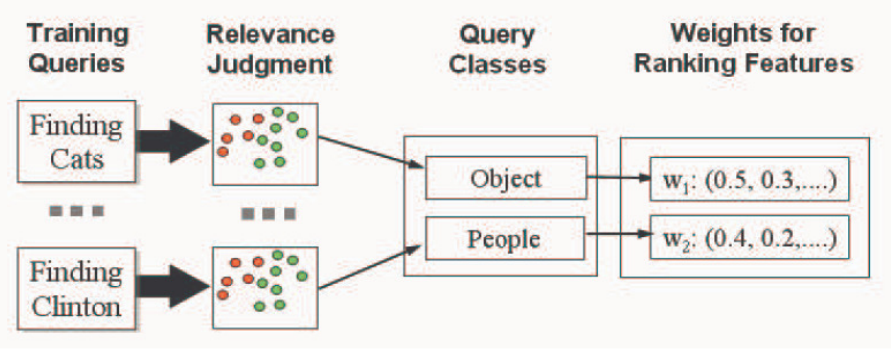
Outline

- The problem
- Dynamic multimodal fusion
  - Query feature extraction
  - Query matching
- Experiments and evaluation
- Conclusion

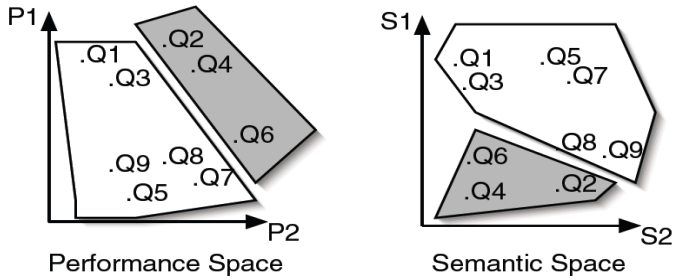
© 2007 IBM Research

Prior Work on Query Dependent Fusion

- Query-classes: manually defined
  - [Chua et. a., NUS]
  - [Yan et. al., CMU]



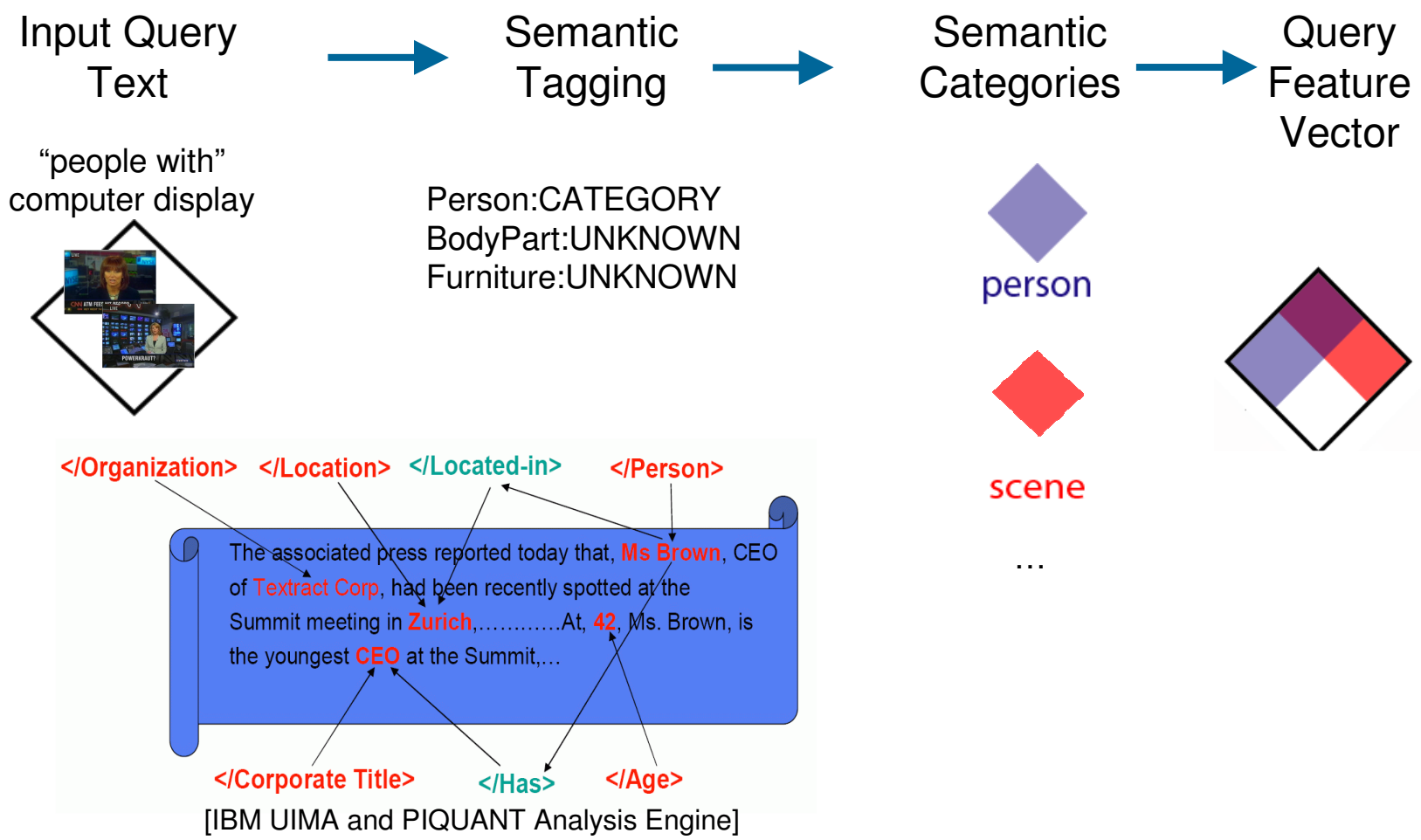
- Joint clustering of queries in feature + performance space [Kennedy et. al.]



- Our approach
  - Can we fuse without pre-defined “query classes” or “query clusters”?
  - Deep text analysis for query features and query matching

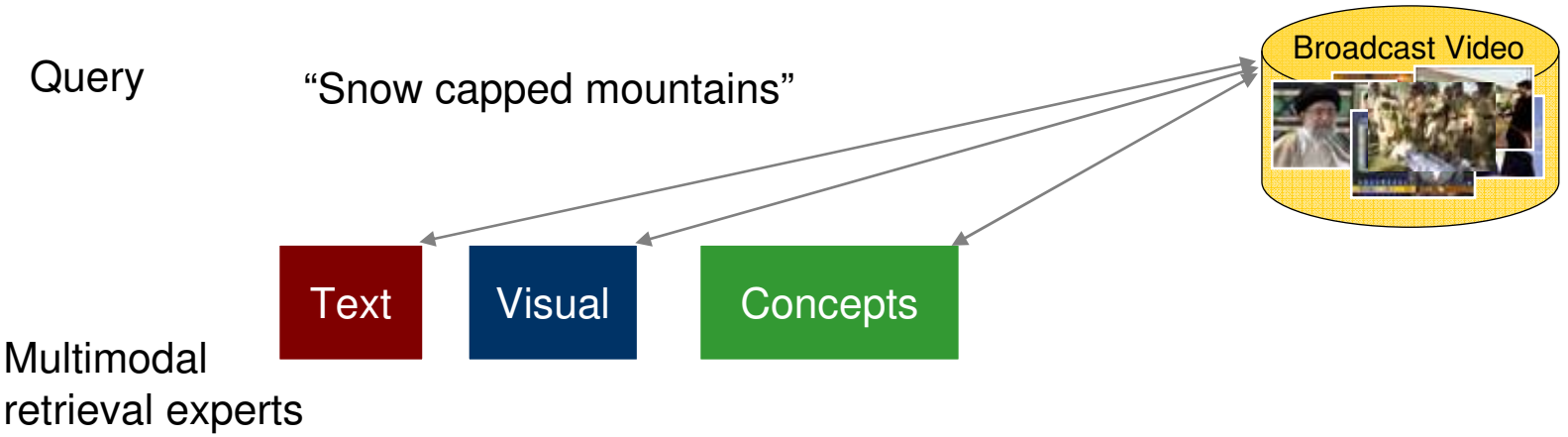
© 2007 IBM Research

Extract Semantic Query Features



© 2007 IBM Research

Weighted linear fusion

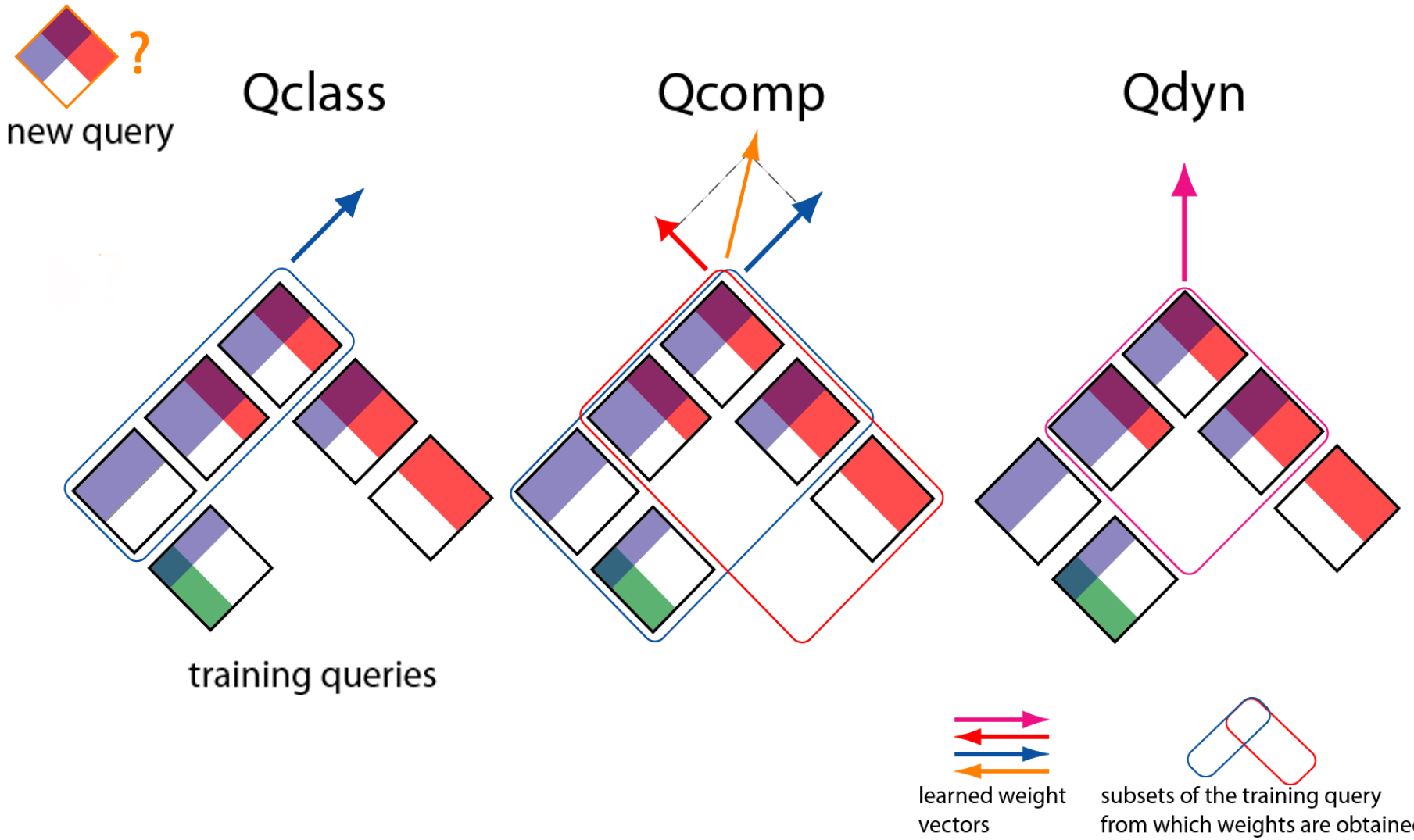


$$r_t \qquad r_v \qquad r_c$$

$$r = w_t r_t + w_v r_v + w_c r_c \qquad r, r_t, r_v, r_c \in \mathcal{R}$$

© 2007 IBM Research

Three Query-dependent Fusion Strategies



© 2007 IBM Research

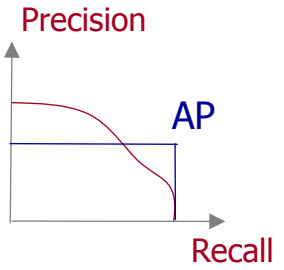
Experimental setup

- TRECVID benchmark corpora
  - Multi-lingual news from 6 channels
  - Training/test split:
    - 2005 24 queries, 110 hours
    - 2006 24 queries, 160 hours

news videos

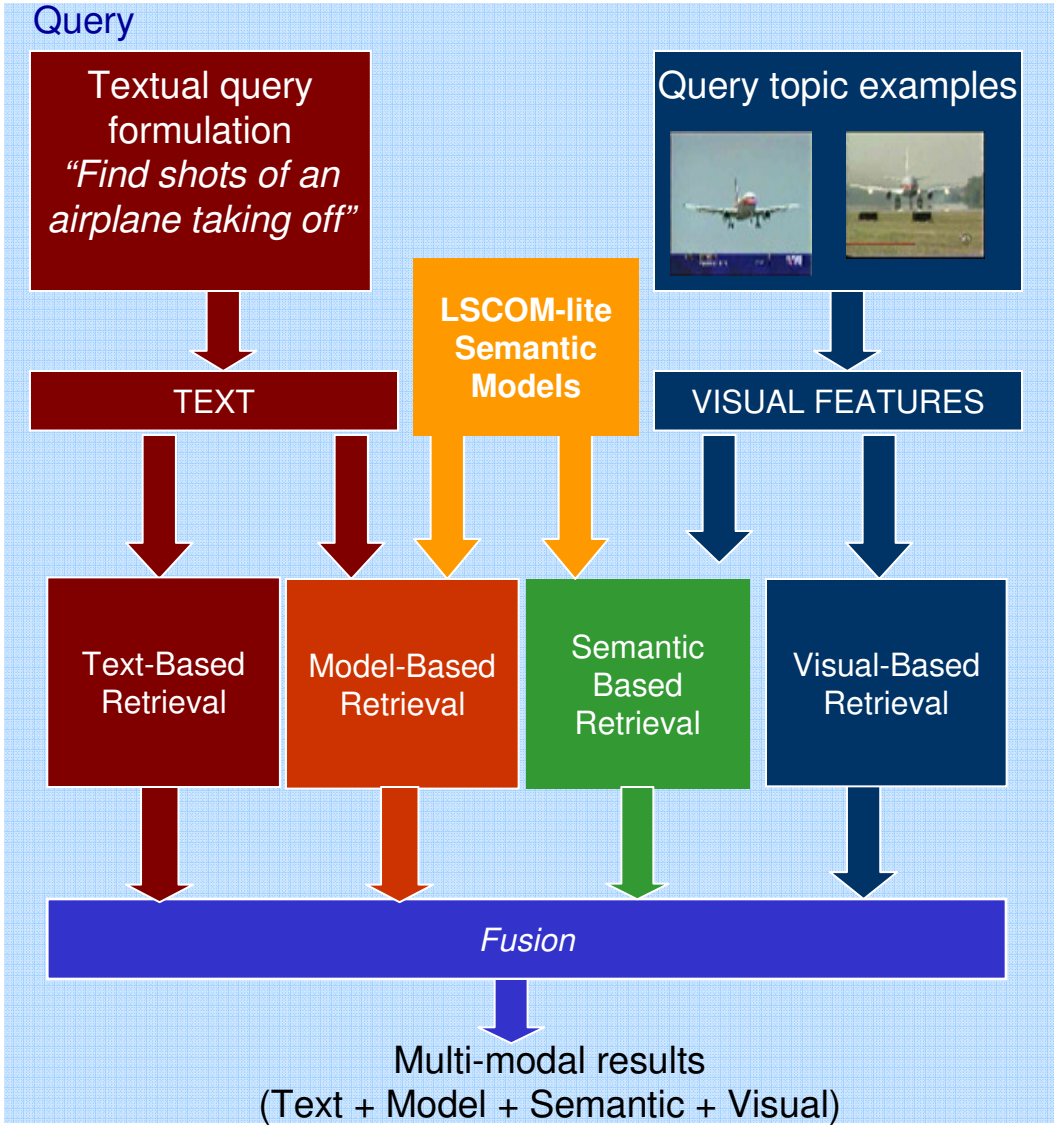
- Measure AP (average precision) on each query

Average precision



© 2007 IBM Research

IBM Multimedia Search System Overview

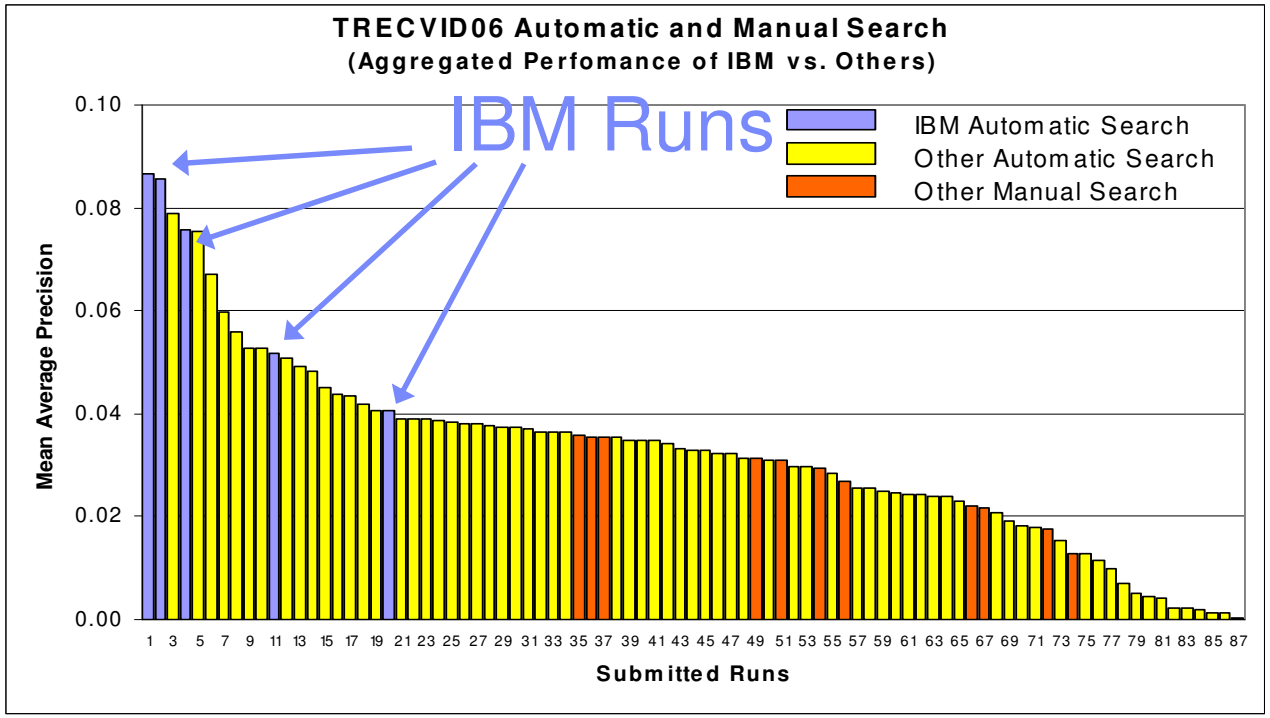


Approaches:

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

© 2007 IBM Research

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**

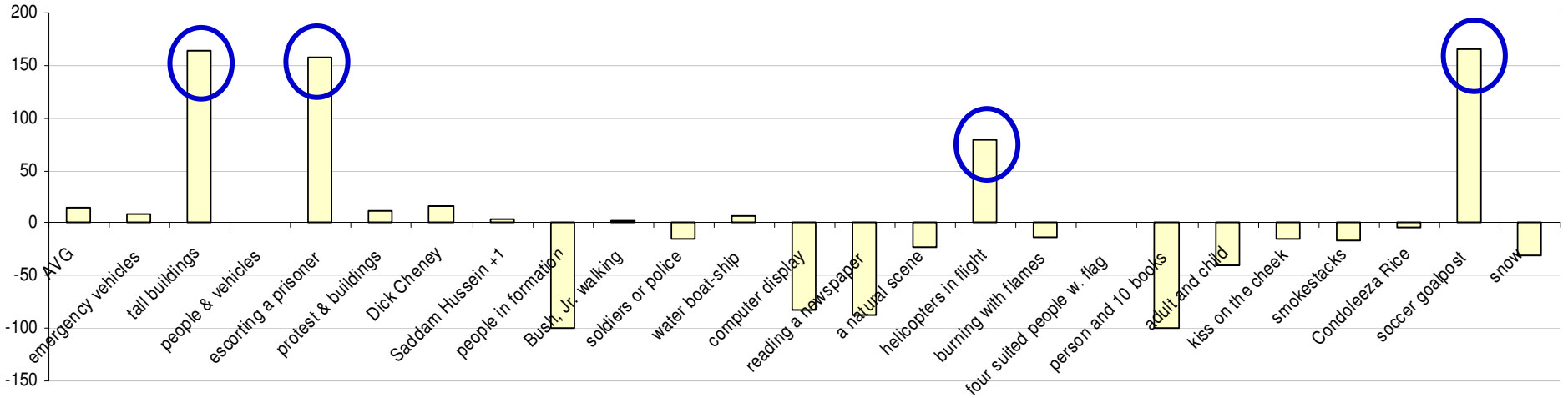
- Multi-modal fusion doubles baseline performance.
- Text-based query matching has the best performance among TRECVID runs
- Improves upon query-independent fusion by 14%

© 2007 IBM Research

Query-dependent Fusion Breakdown

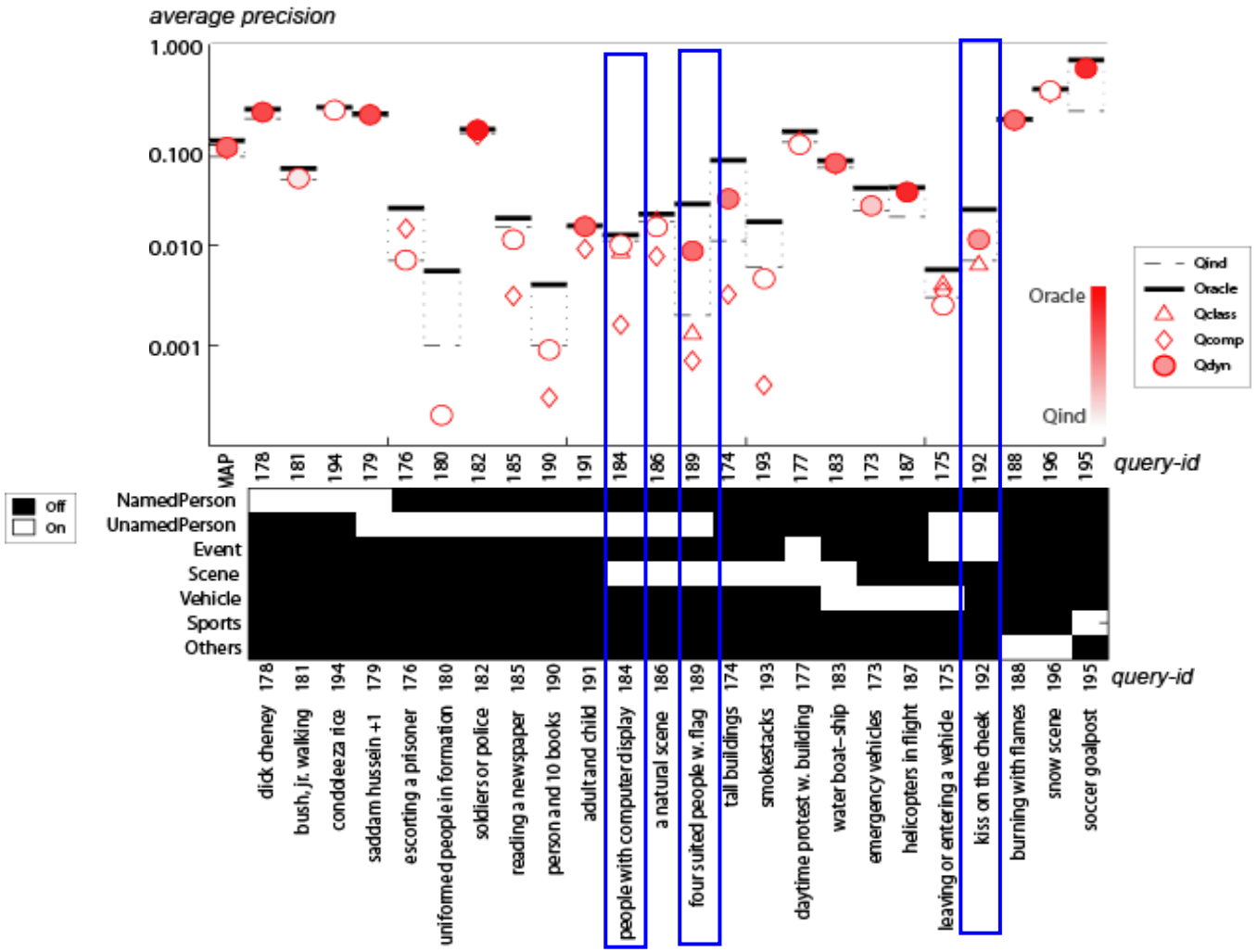
- Observation:
  - Concept-related queries improved the most:
    "tall building", "prisoner", "helicopters in flight", "soccer"
  - Named-entity queries improved slightly:
    "Dick Cheney", "Saddam Hussein", "Bush Jr."
  - Generic people category deteriorated the most:
    "people in formation", "at computer display", "w/ newspaper", "w/ books"

Relative improvement (%): Qcomp vs. Qind



© 2007 IBM Research

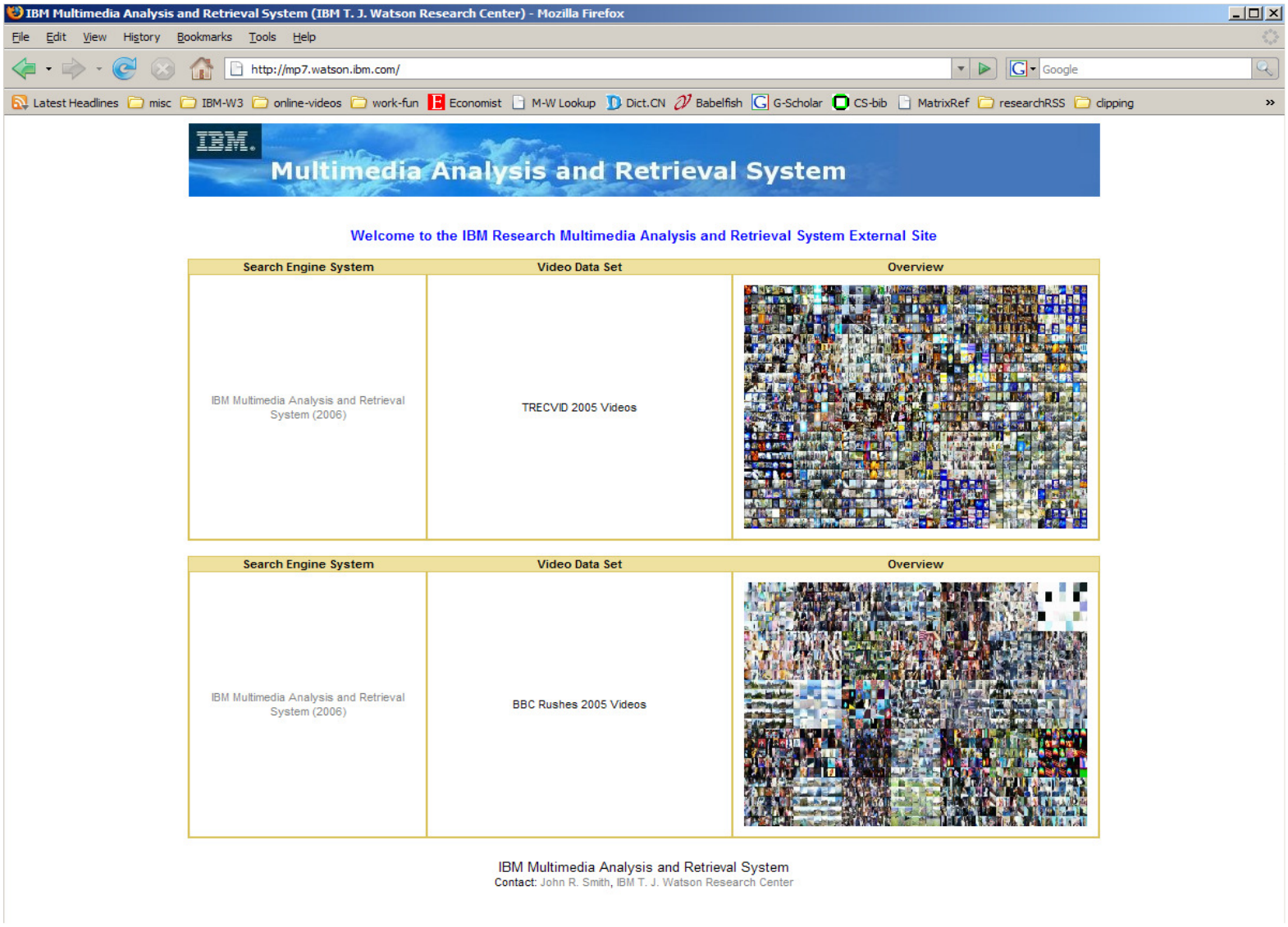
Dynamic Query Fusion Results



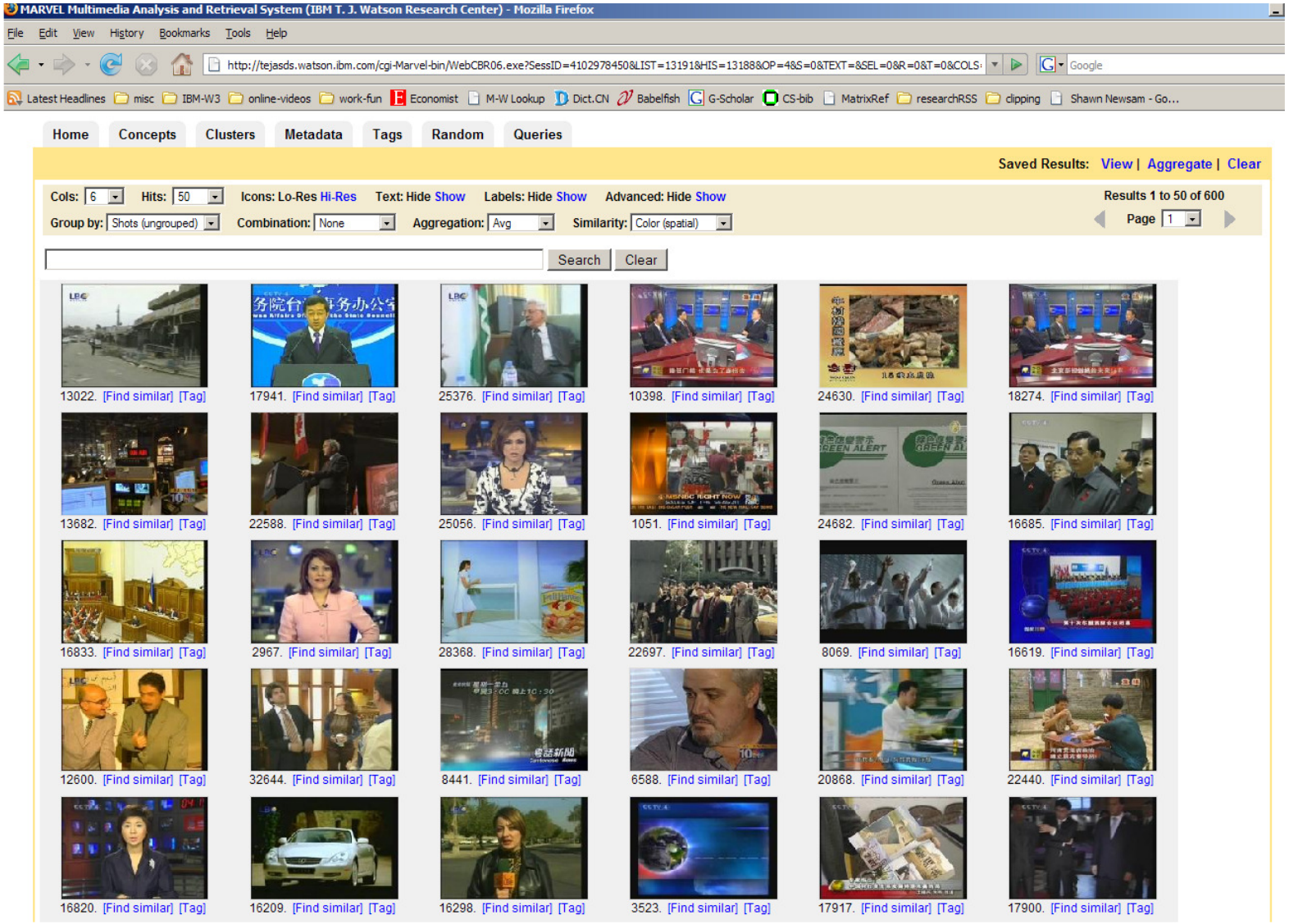
- Qdyn further improves MAP of Qclass by 8%, to 0.094
- The improvement is mainly in generic people queries → where Qclass failed previously

© 2007 IBM Research

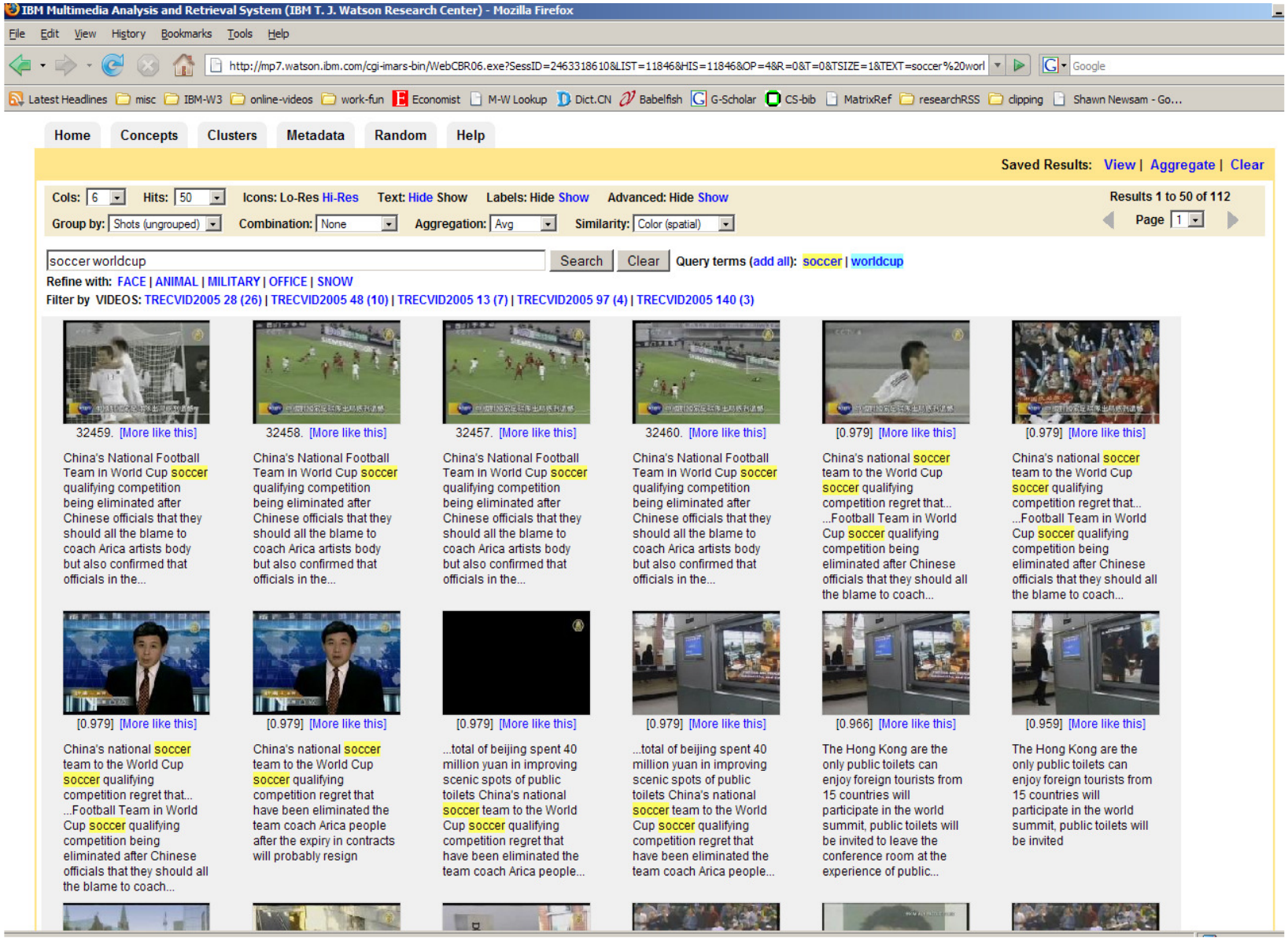
Demo: multimodal interactive search



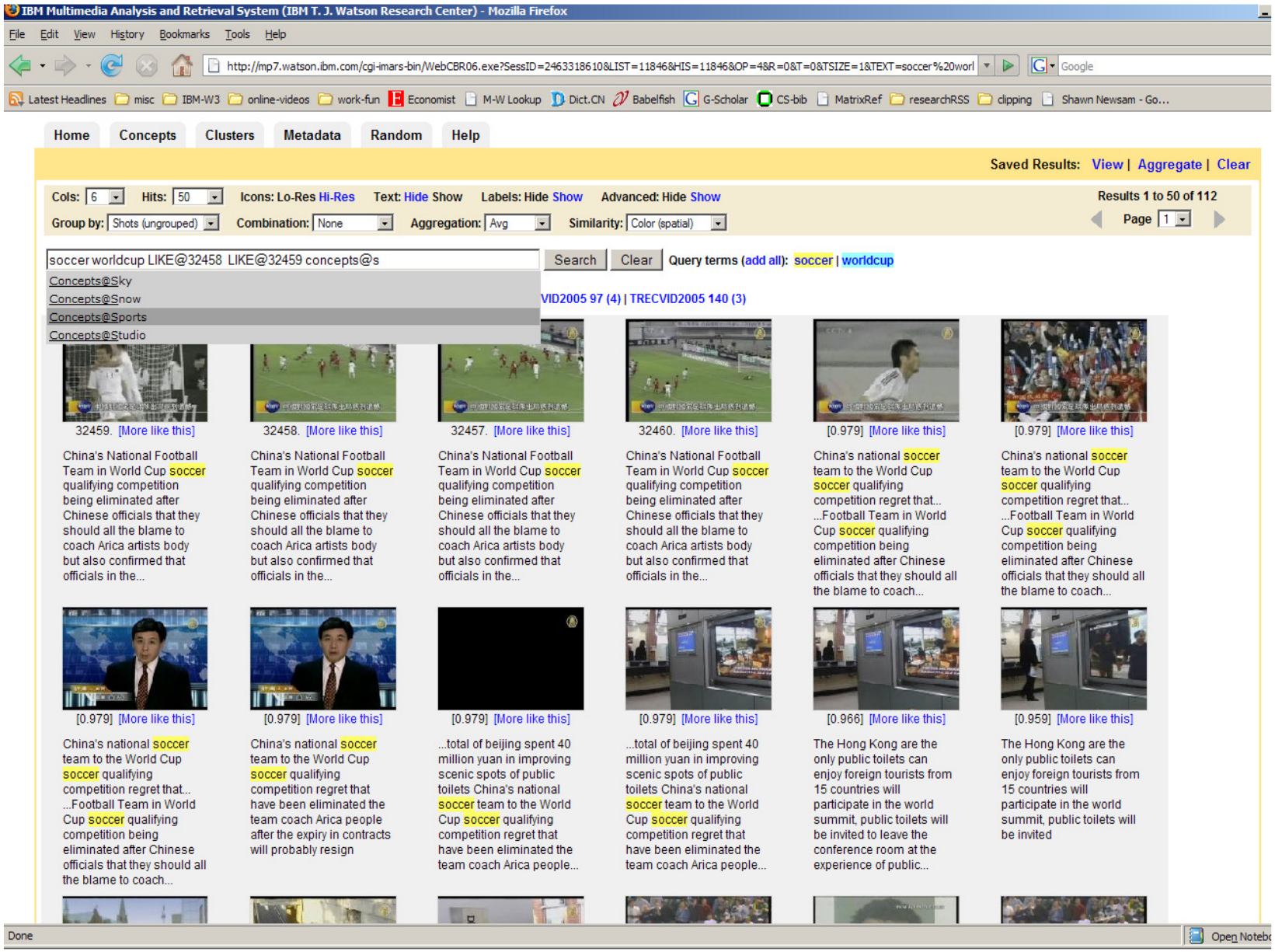
© 2007 IBM Research



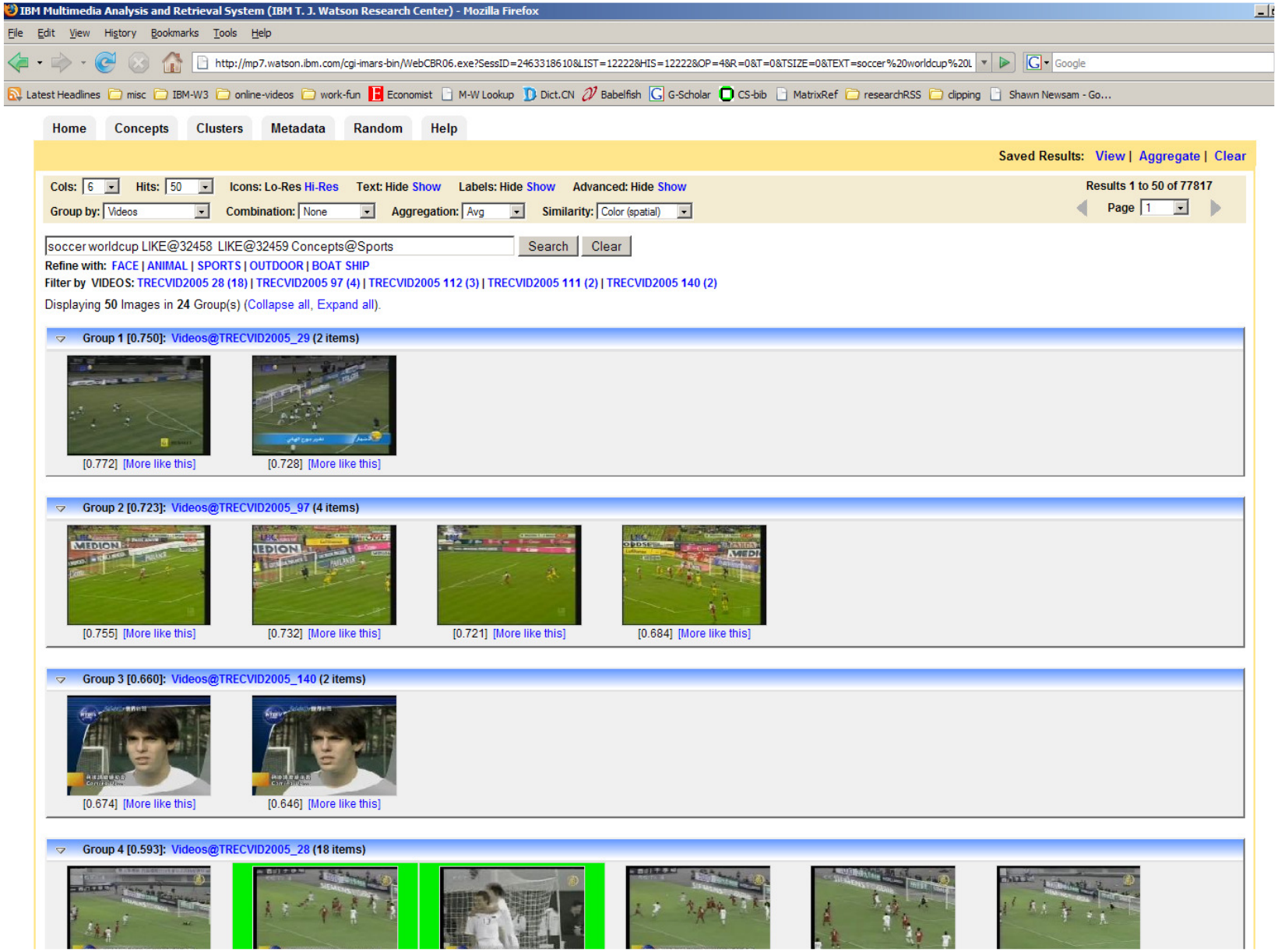
© 2007 IBM Research



© 2007 IBM Research



© 2007 IBM Research



© 2007 IBM Research

Conclusion

- Presented a solution to query-dependent multimodal fusion problem for video search
  - Query features based on deep semantic parsing
  - Dynamic query matching
  - Improvement on TRECVID queries
- Future work
  - More retrieval modality, more diverse queries

© 2007 IBM Research

Thank You

For more information:

- IBM Multimedia Retrieval Demo
  http://mp7.watson.ibm.com/

- "Data modeling strategies for imbalanced learning in video search", J. Tesic, A. Natsev, L. Xie, J. Smith, ICME'07 → next poster session

- "Semantic Concept-Based Query Expansion and Re-ranking for Multimedia Retrieval: A Comparative Review and New Approaches", A. P. Natsev, A. Haubold, J Tesic, L. Xie, R. Yan, ACM Multimedia 2007, to appear

© 2007 IBM Research