Carnegie Mellon University TRECVID Automatic and Interactive Search

Mike Christel, Alex Hauptmann, Howard Wactlar, Rong Yan,
Jun Yang, Bob Baron, Bryan Maher, Ming-Yu Chen, Wei-Hao Lin
Carnegie Mellon University
Pittsburgh, USA

November 14, 2006

Talk Overview

- Automatic Search
- CMU Informedia Interactive Search Runs
  - Why these runs?
  - What did we learn?
- Additional “Real Users” Run from late September
- TRECVID Interactive Search and Ecological Validity
- Conclusions

Informedia Acknowledgments

- Support through the Advanced Research and Development Activity under contract number NBCHC040037 and H98230-04-C-0406
- Concept ontology support through NSF IIS-0205219
- Contributions from many researchers – see www.informedia.cs.cmu.edu for more details

Automatic Search

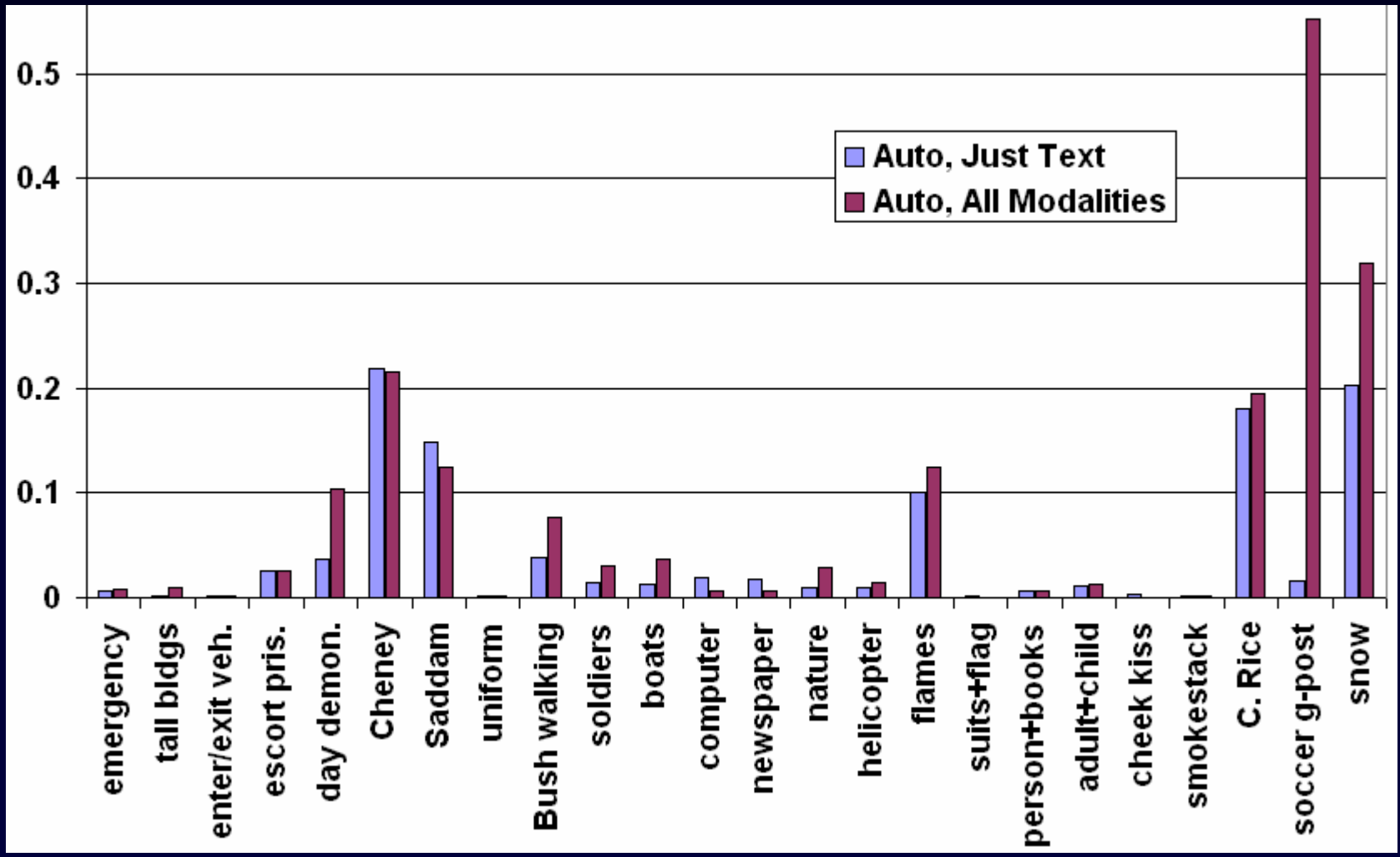
For details, consult both the CMU TRECVID 2006 workshop paper and Rong Yan's just-completed PhD thesis:

Probabilistic Models for Combining Diverse Knowledge Sources in Multimedia Retrieval. Ph.D. thesis, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, 2006

Run Name "Touch": Automatic retrieval based on only transcript text, MAP 0.045

Run Name "Taste": Automatic retrieval based on transcript text and all other modalities, MAP 0.079

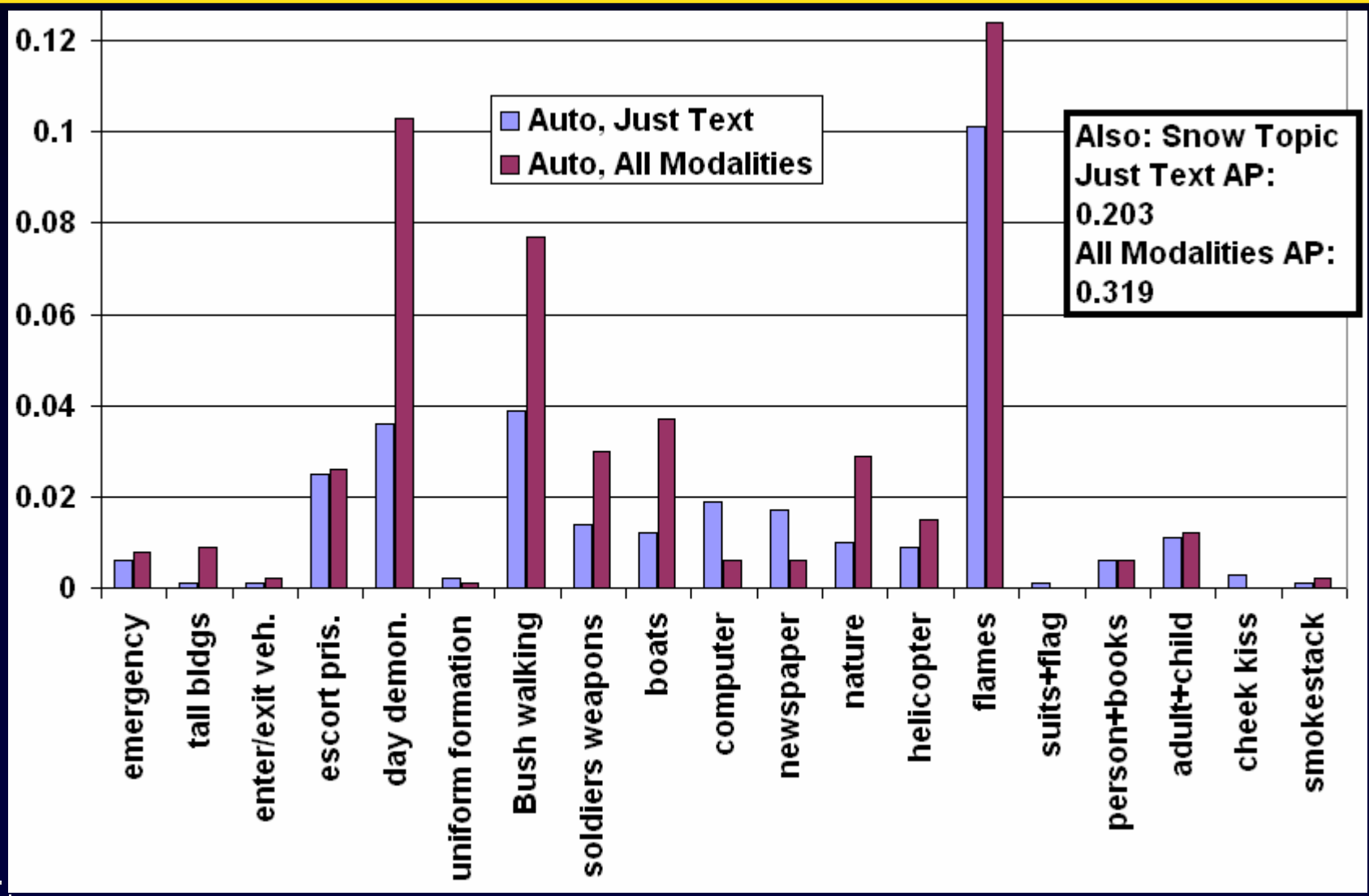
Average Precision, TRECVID 2006 Topics



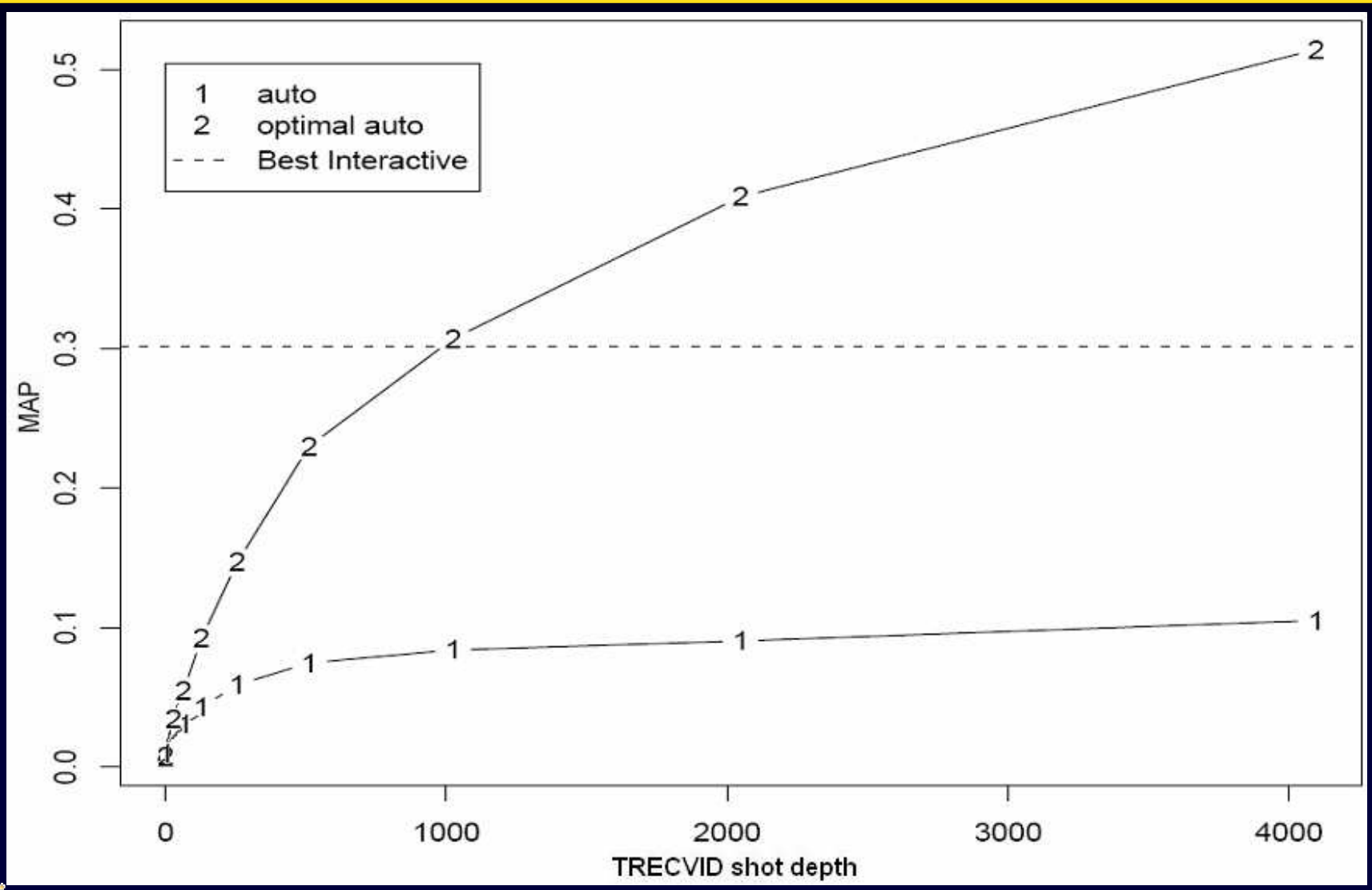
MAP, Automatic Runs, Different Subsets

Topic Set Description	MAP Auto Text	MAP Auto All
All 24 Topics	0.045	0.079
Sports (just 195, soccer goalposts)	0.016	0.552
Non-Sports (all topics except for 195)	0.046	0.058
Specific (named people, 178, 179, 194 about Dick Cheney, Saddam Hussein, Condoleezza Rice)	0.183	0.178
Specific, including Bush walking topic too (181)	0.147	0.153
Generic, non-sports (including topic 181)	0.026	0.041
Generic, non-sports (excluding topic 181)	0.025	0.039

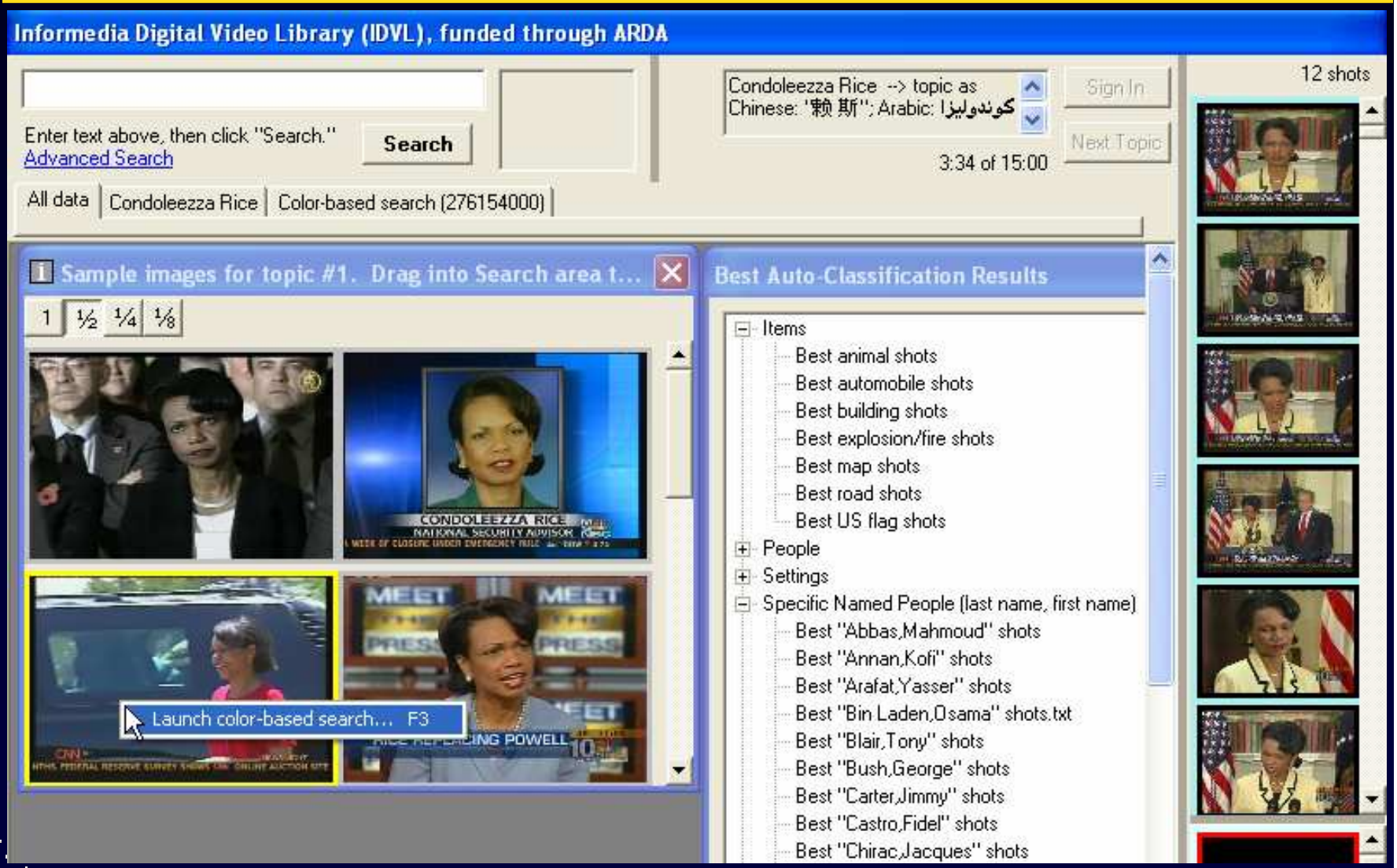
Avg. Precision, Generic Non-Sports Subset



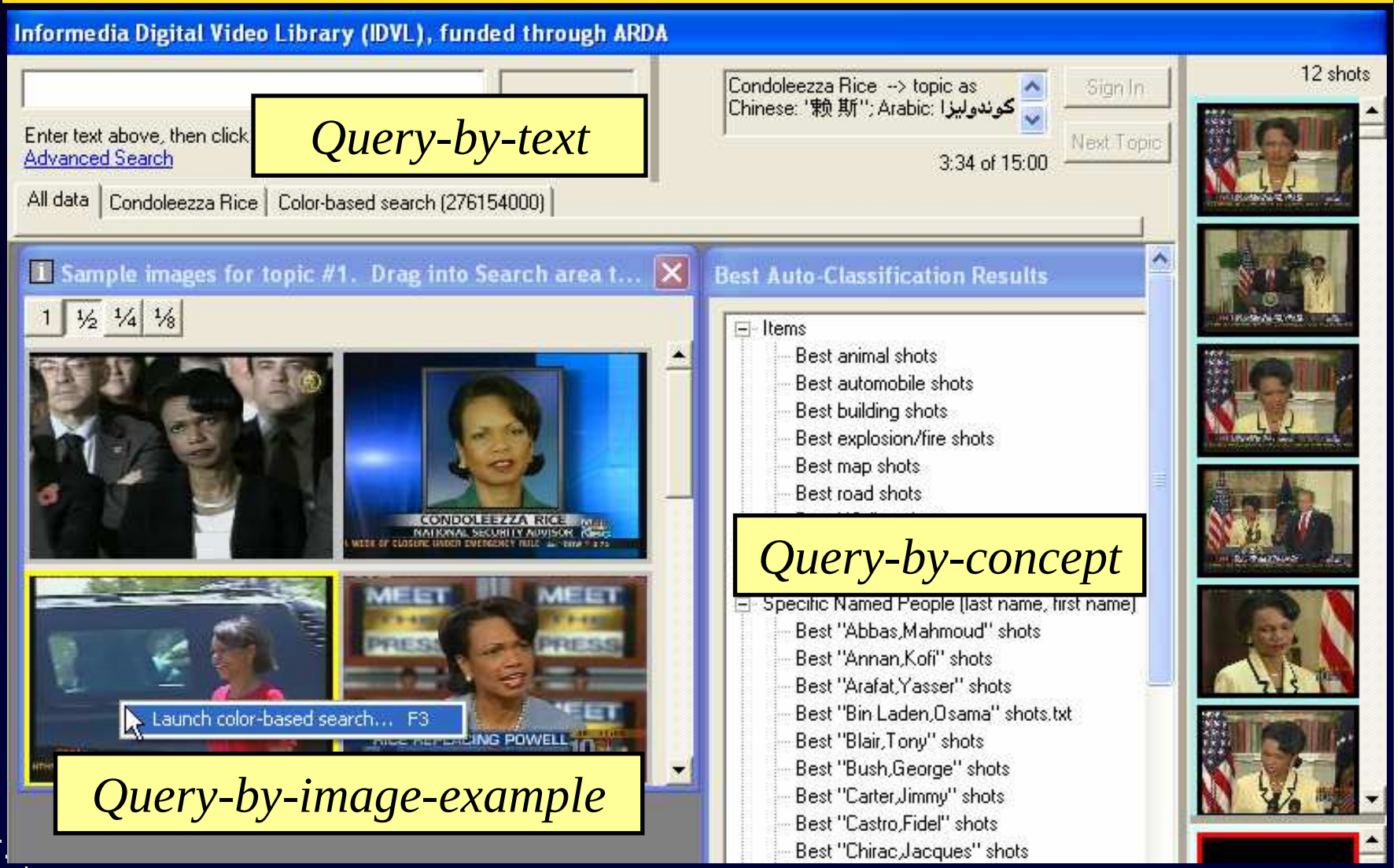
Evidence of Value within the Automatic Run



Looking Back: CMU TRECVID 2005 Interface



TRECVID Interface: 3 Main Access Strategies



Consistent Context Menu for Thumbnails

Other Features, “Classic” Informedia

- Representing both subshot (NRKF) and shot (RKF) from the 79,484 common shot reference (146,328 Informedia shots)
- “Overlooked” and “Captured” shot set bookkeeping to suppress shots already seen and judged (note CIVR 2006 paper about trusting “overlooked” too much as negative set)
- Clever caching of non-anchor, non-commercial shots for increased performance in refreshing storyboards
- Optimized layouts to pack more imagery in screen for user review
- Clustering shots by story segment to better preserve temporal flow
- Navigation mechanisms to move from shot to segment, from shot to neighboring shots, and from segment to neighboring segments

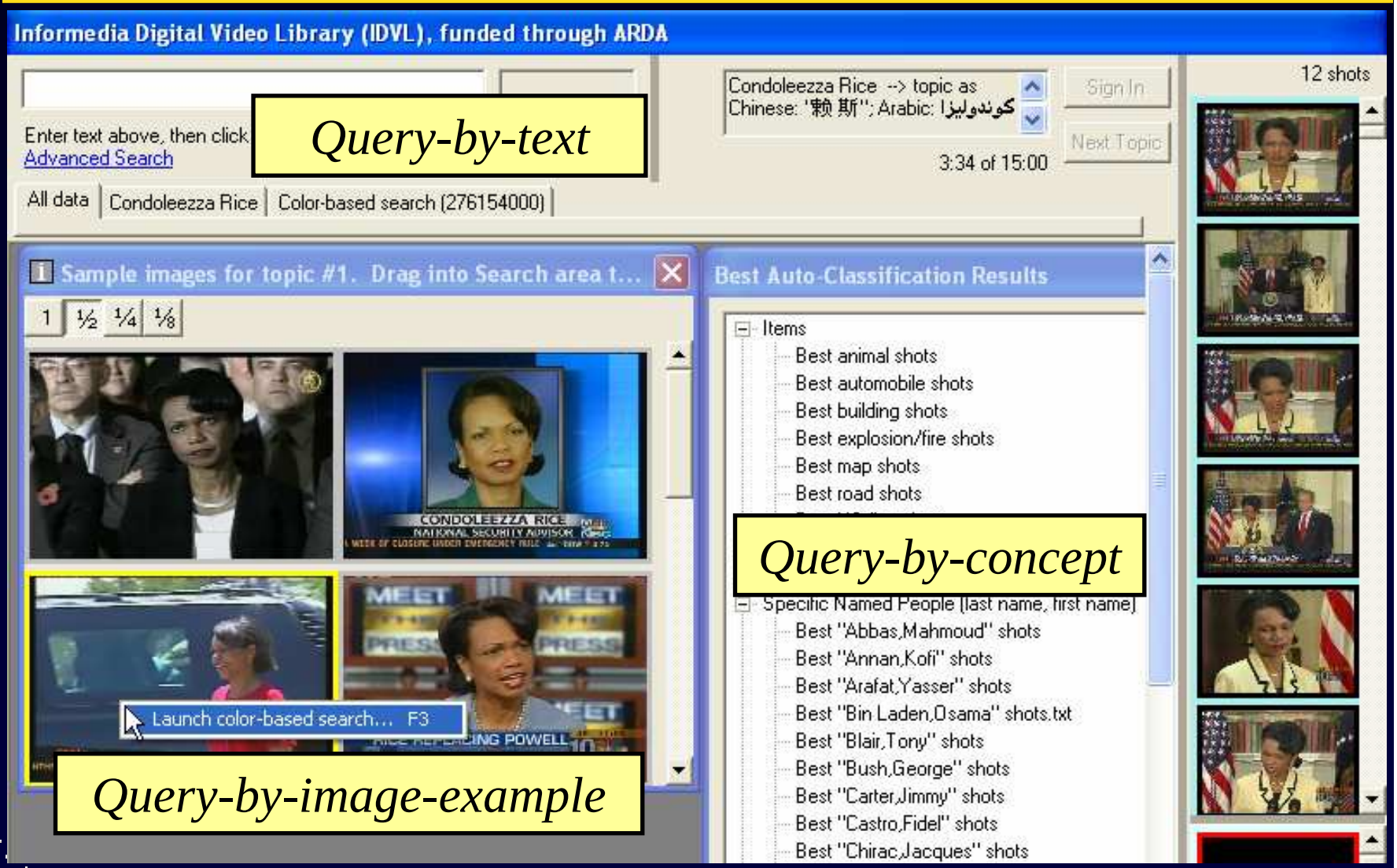
Motivation for CMU Interactive Search Runs

Question: Can the automatic run help the interactive user?

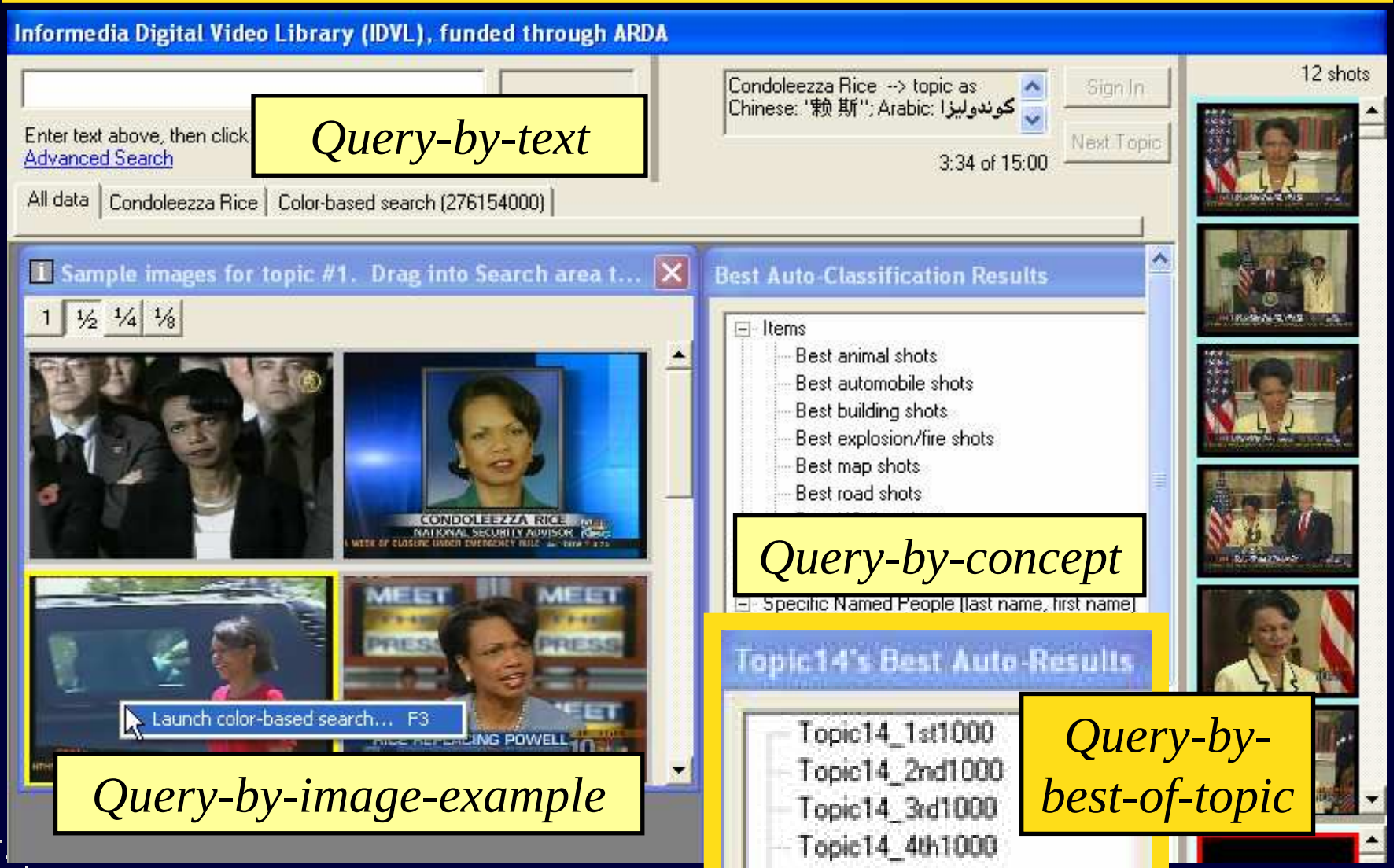
From the success of the CMU Extreme Video Retrieval (XVR) runs of TRECVID 2005, the answer seems to be yes.

Hence, query-by-best-of-topic added into the “classic” interface.

TRECVID 2005: 3 Main Access Strategies

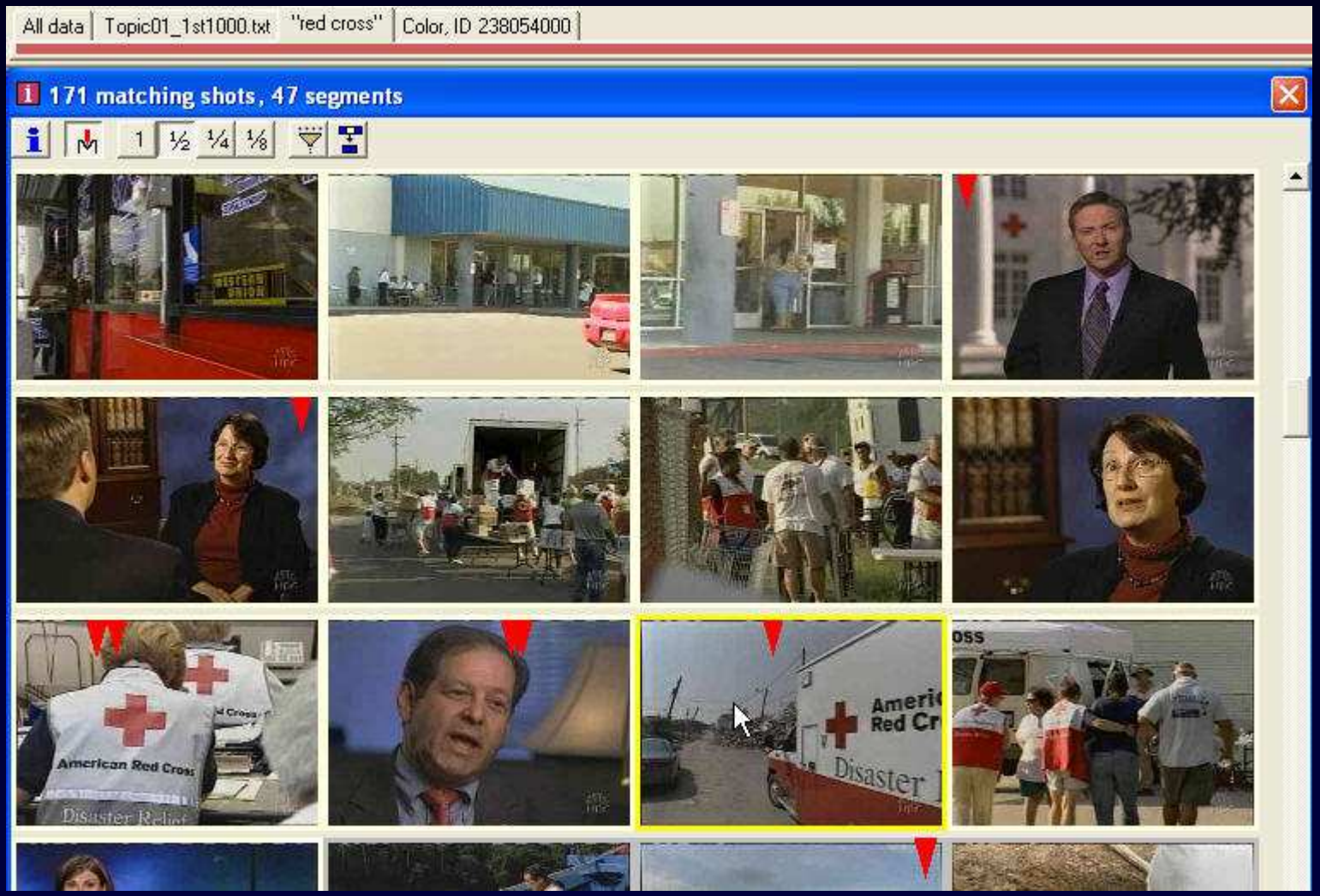


TRECVID 2006 Update: 4 Access Strategies



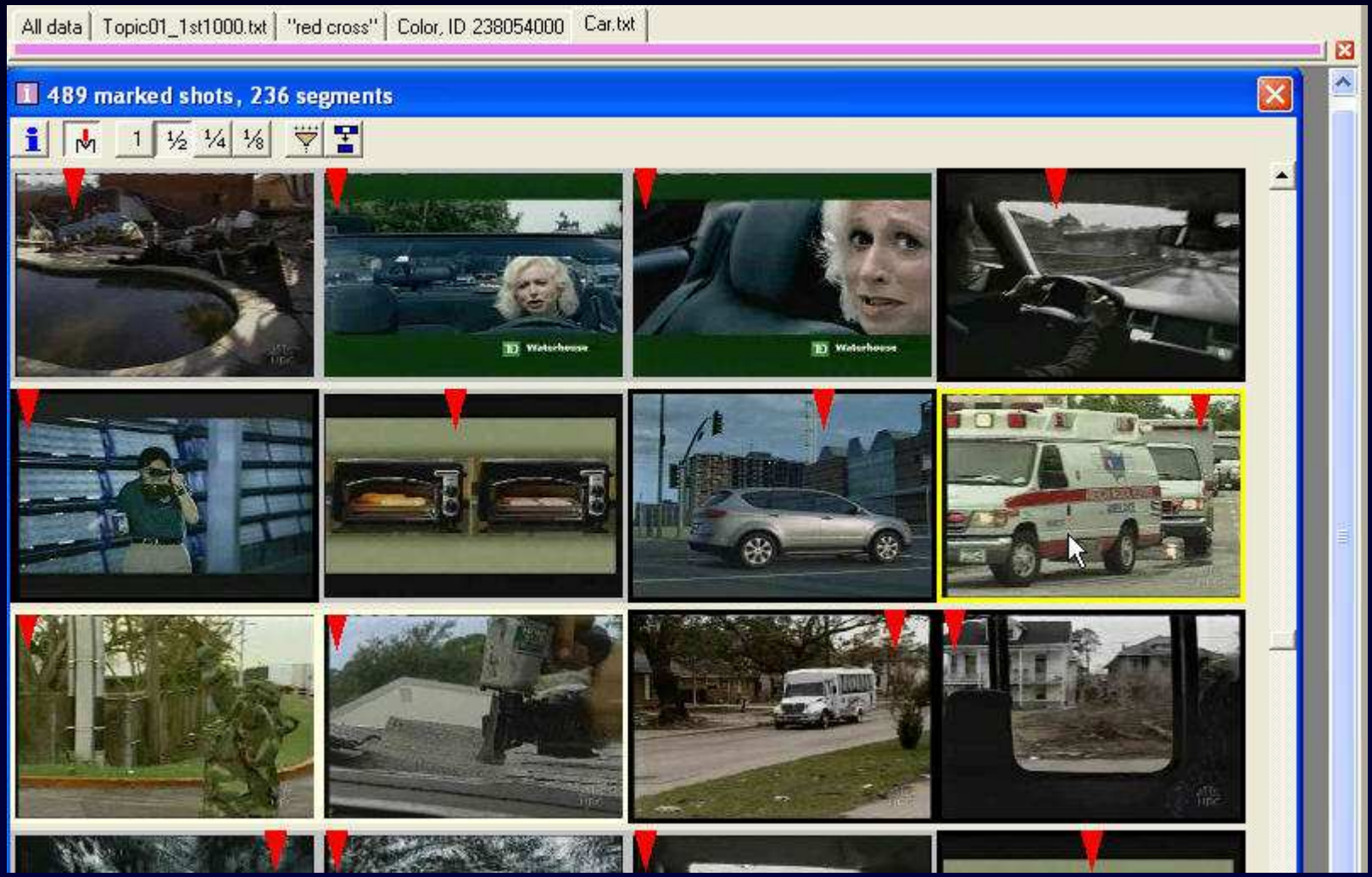
Example: Best-of-Topic (Emergency Vehicles)

Example: Query by Text “Red Cross”



Example: Query by Image Example

Example: Query by Concept (Car)



Motivation for CMU Interactive Search Runs

Question: Can the automatic run help the interactive user?

From the success of the CMU Extreme Video Retrieval (XVR) runs of TRECVID 2005, the answer seems to be yes.

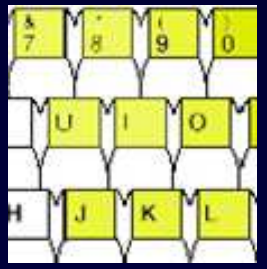
Hence, query-by-best-of-topic added into the “classic” interface.

Extreme Video Retrieval runs kept to confirm the value of the XVR approach:

(i) manual browsing with resizable pages (MBRP)

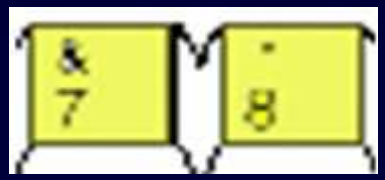
(ii) rapid serial visual presentation (RSVP) with system-controlled presentation intervals

MBRP Interface



Keyhole RSVP (Click when Relevant)

Stereo View in RSVP



Motivation for CMU Interactive Search Runs

Question: Can the automatic run be improved “on the fly” through interactive use?

Based on user input, the positive examples are easily noted (the chosen/marked shots) with precision at very high 90+% levels based on prior TRECVID analysis of user input . Negative examples are less precise, but are the set of “overlooked” shots passed over when selecting relevant ones.

Hence, active learning/relevance feedback from positive and negative user-supplied samples added into the extreme video retrieval runs, and used throughout for auto-expansion.

First 3 Screens of 9 Images, Auto-Ordering

Learning Possible from Marked User Set…

Next 2 Screens of 9 Images, Auto-Ordering

Same “Next 2” Screens, Example Reordering

Example Reordering through Active Learning on the User Input to This Point

Motivation for CMU Interactive Search Runs

Question: Does the interface into the automatic run matter to the interactive user?

In 2005, tested 2 variations of CMU Extreme Video Retrieval: manual browsing with resizable pages (MBRP) and rapid serial visual presentation (RSVP) .

In 2006, added Informedia classic storyboard interface as another window into the automated runs, trying to preserve benefits without requiring the “extreme” stress and keeping more control with user.

Informedia Storyboard Interface

Informedia Storyboard Under User Control

Informedia Storyboard with Concept Filters

TRECVID 2006 CMU Interactive Search Runs

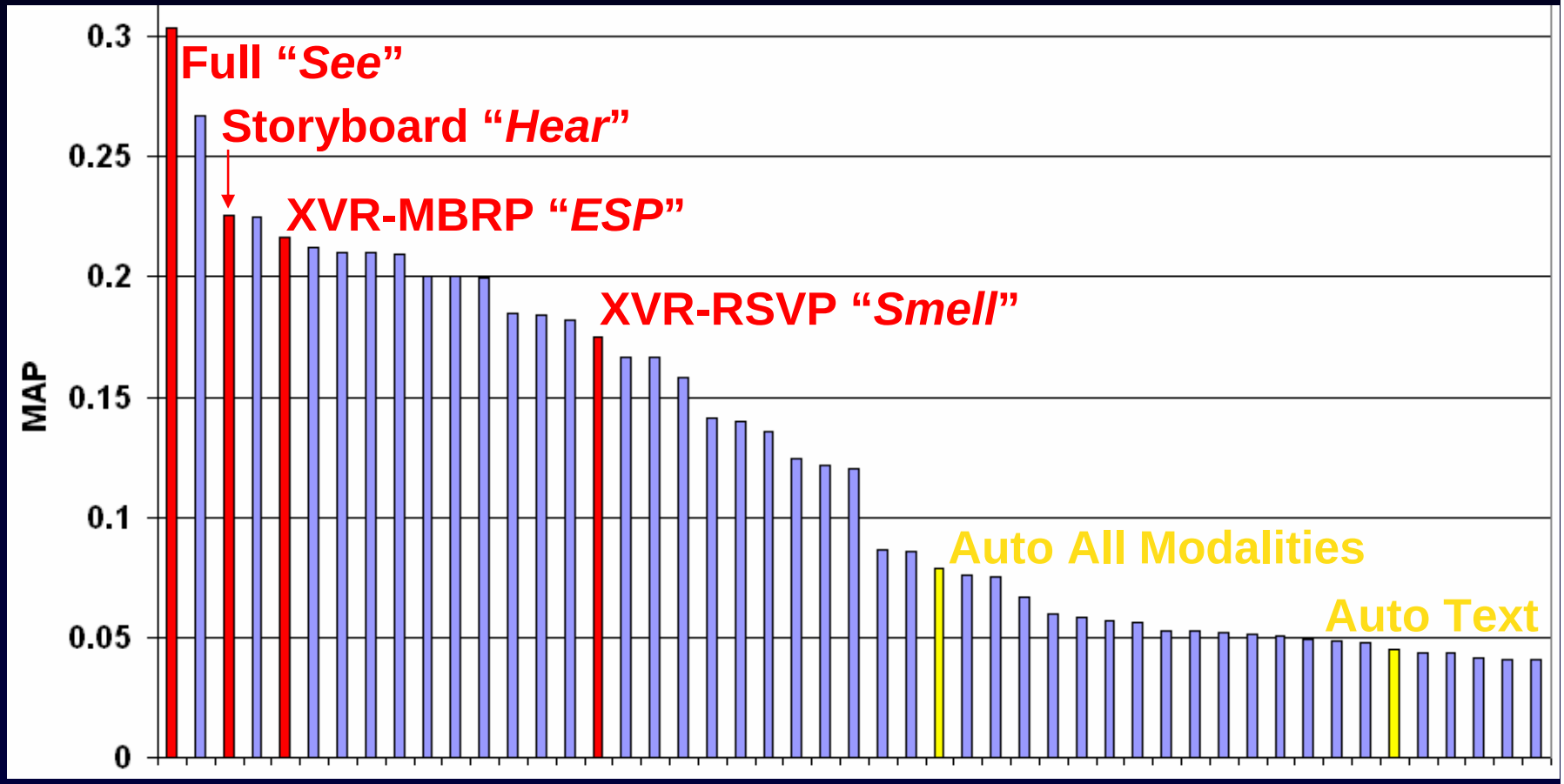
Run	Description
See	Full Informedia interface, expert user, query-by-text, by-image, by-concept, and auto-topic functionality
Hear	Image storyboards working only from shots-by-auto-topic (no query functionality), 2 expert users
ESP	Extreme video retrieval (XVR) using MBRP, relevance feedback, no query functionality
Smell	Extreme video retrieval (XVR) using RSVP with system controlled presentation intervals, relevance feedback, no query functionality

TRECVID 2006 CMU Interactive Search Runs

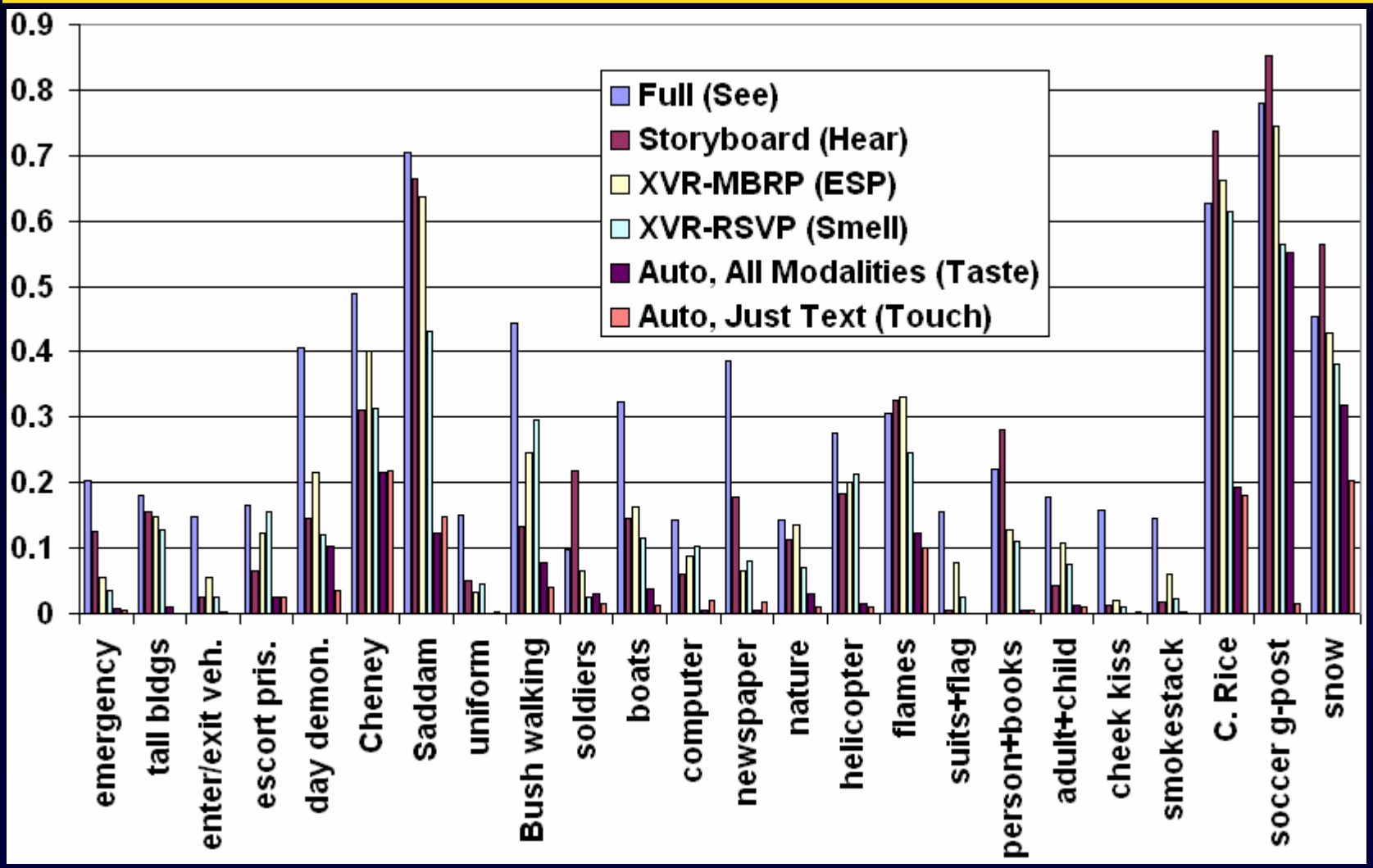
Run	Description	MAP
See	**Full Informedia**	0.303
Hear	**Informedia interface to just best-of-topic**	0.226
ESP	XVR using MBRP	0.216
Smell	XVR using RSVP	0.175

- Automatic output does hold value in interactive users' hands
- Learning strategies confounded in RSVP (2 shots marked per interaction, but 1 was almost always wrong)
- Additional capability (to query by text, image, concept) leads to improved performance with the "See" run

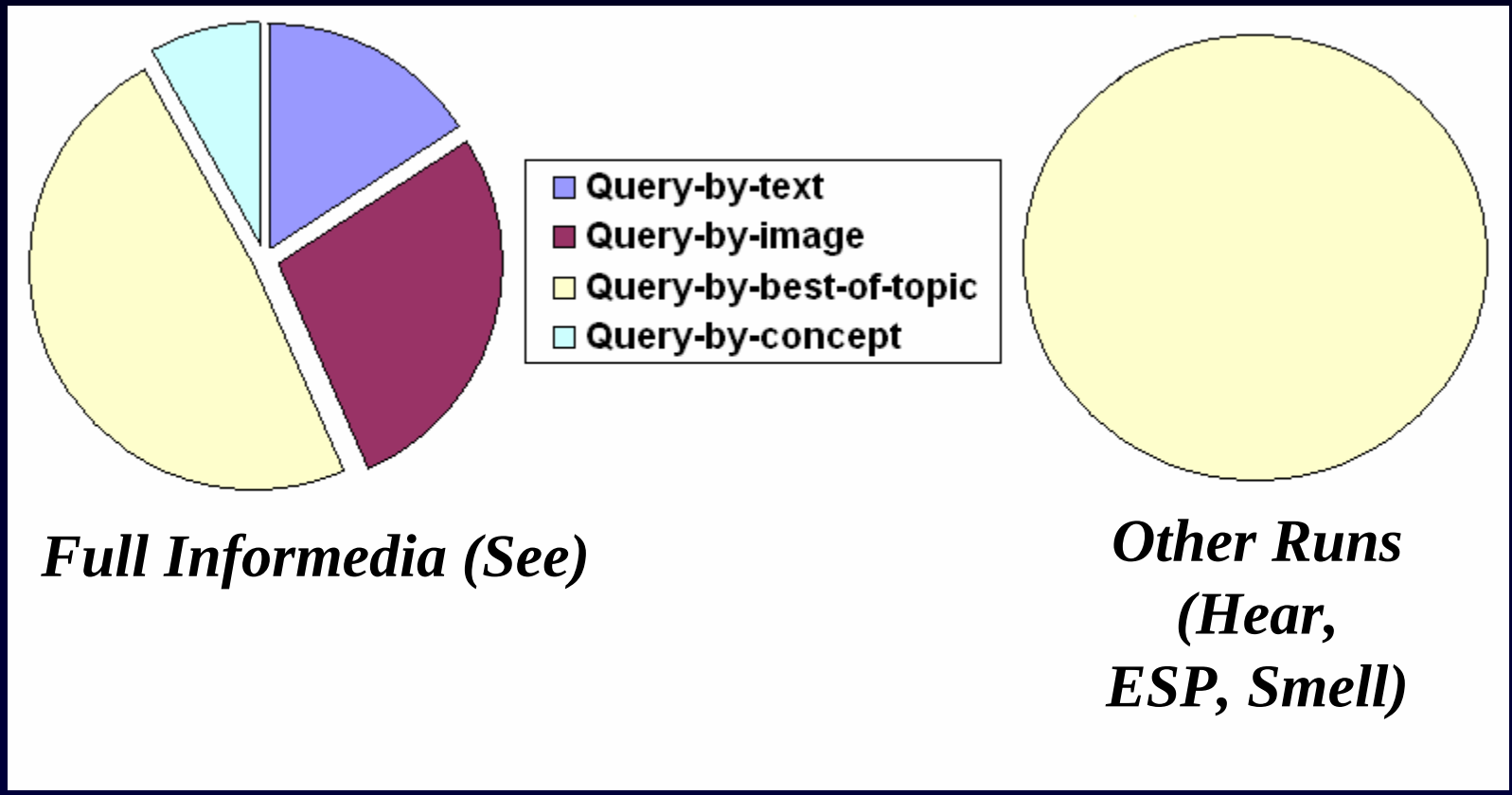
MAP Top 50 Search Runs



Average Precision, CMU Search Runs



System Usage, CMU Interactive Runs



Full Informedia (See)

Other Runs (Hear, ESP, Smell)

What About "Typical" Use? …Ecological Validity

Ecological validity – the extent to which the context of a user study matches the context of actual use of a system, such that

- it is reasonable to suppose that the results of the study are representative of actual usage, and
- the differences in context are unlikely to impact the conclusions drawn.

All factors of how the study is constructed must be considered: how representative are the tasks, the users, the context, and the computer systems?

TRECVID for Interactive Search Evaluation

- TRECVID provides a public corpus with shared metadata to international researchers, allowing for metrics-based evaluations and repeatable experiments
- An evaluation risk with over-relying on TRECVID is tailoring interface work to deal solely with the genre of video in the TRECVID corpus, e.g., international broadcast news
  - This risk is mitigated by varying the TRECVID corpus
- A risk in being closed: test subjects are all developers
- Another risk: topics and corpus drifting from being representative of real user communities and their tasks
- Exploratory browsing interface capabilities supported by video collages and other information visualization techniques not evaluated via IR-influenced TRECVID

TRECVID for Interactive Search Evaluation

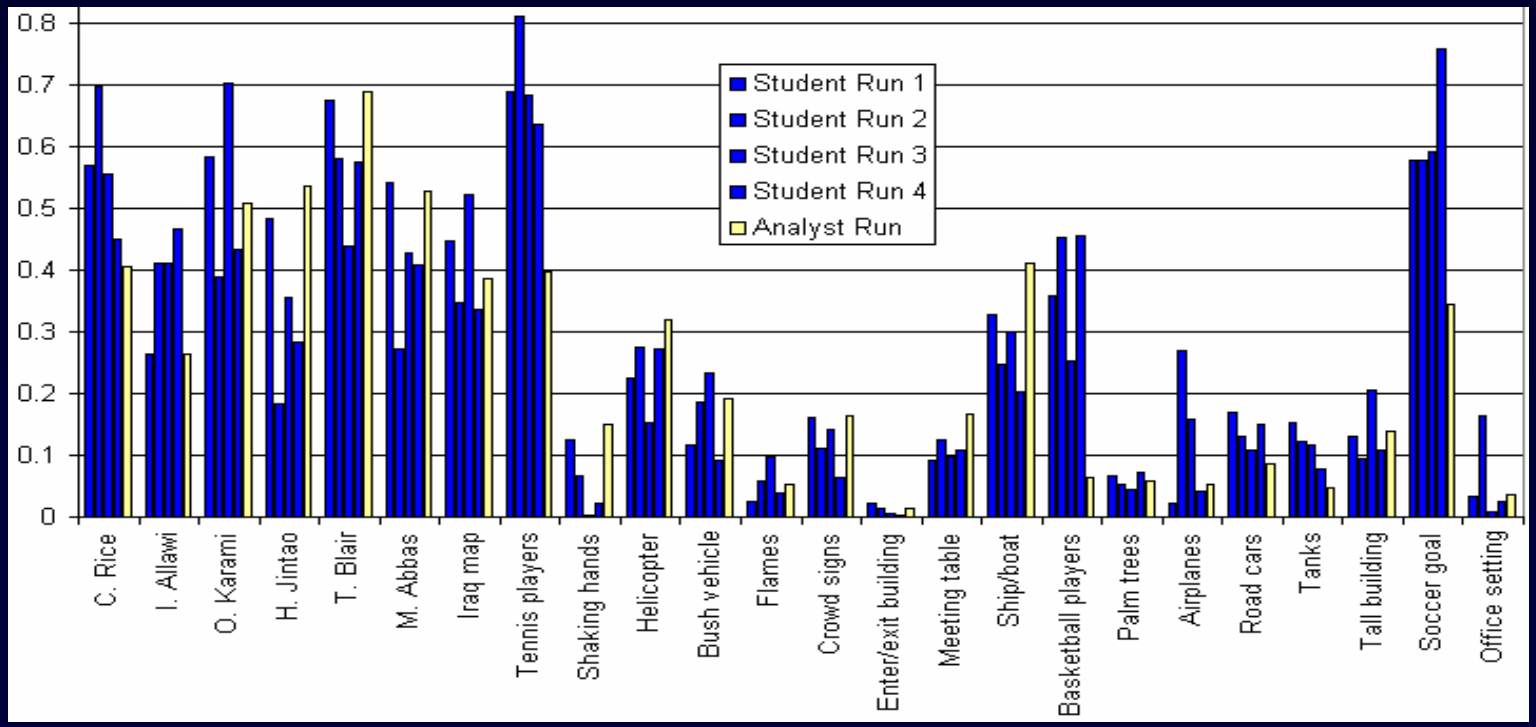
- TRECVID provides a public corpus with shared metadata to international researchers, allowing for metrics-based evaluations and repeatable experiments
- An evaluation risk with over-relying on TRECVID is tailoring interface work to deal solely with the genre of video in the TRECVID corpus, e.g., international broadcast news
  - This risk is mitigated by varying the TRECVID corpus
- A risk in being closed: test subjects are all developers
- Another risk: topics and corpus drifting from being representative of real user communities and their tasks
- Exploratory browsing interface capabilities supported by video collages and other information visualization techniques not evaluated via IR-influenced TRECVID

Analyst Run, TRECVID Tasks

- 6 Analysts, 2-day Informedia Evaluation Workshop
  - TRECVID 2005 under 2 variations, 8 topics each
  - Exploratory tasks
  - TRECVID 2006, 4 topics each, “Informedia Full” system as was used in the “See” submitted run
- Analysts’ profile similar to CMU students, except analysts are more experienced with text search systems, less experienced with video search systems; also an older group

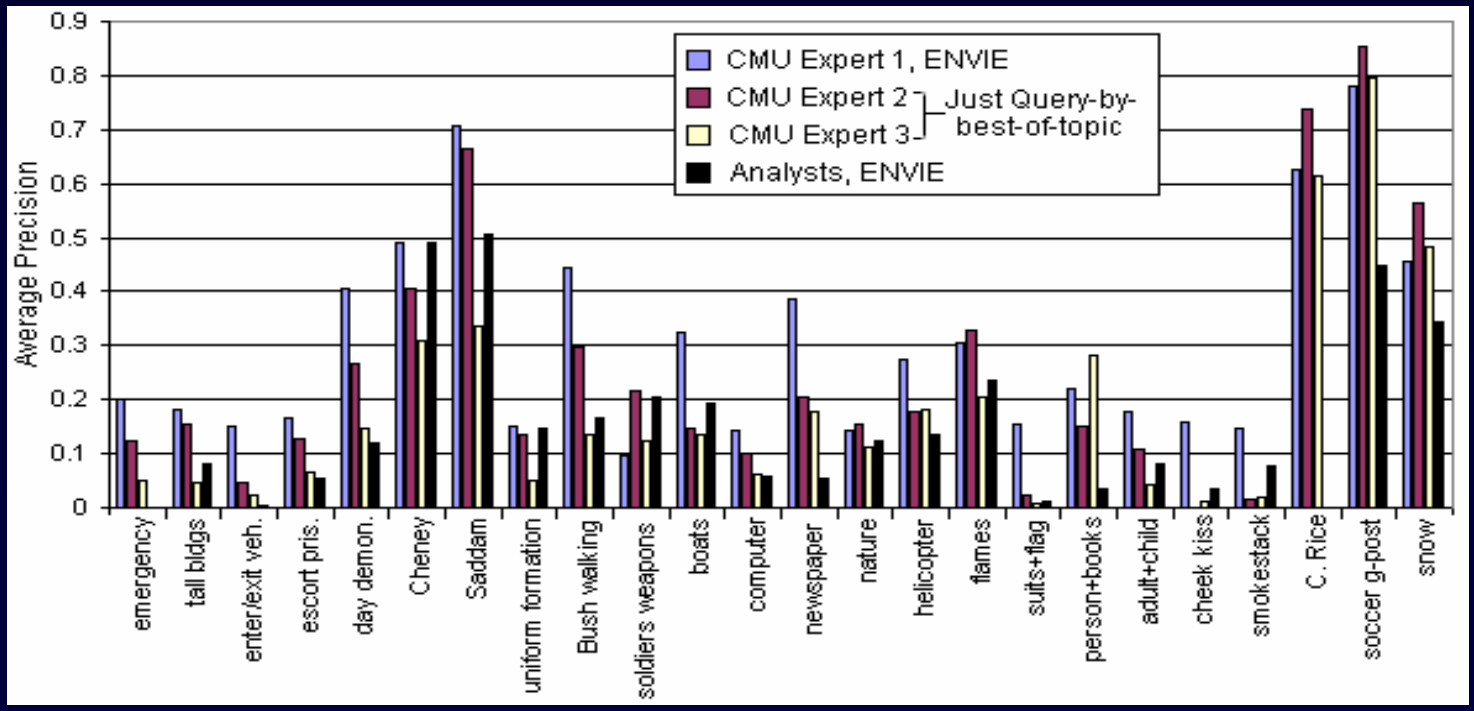
Analysts, Quick Look Back at TRECVID 2005

- MAP of 0.251 correlates well with the 4 student runs' MAP in a TRECVID 2005 study of 0.253 through 0.286 (the best runs from users outside of the system development teams)
- Without underperforming sports topics, MAP is 0.248, vs. student runs of 0.249, 0.228, 0.242, and 0.201



Analysts, TRECVID 2006

- Sports topic again underperformed, one topic (194) skipped
- MAP for 24 topics: 0.150; for the 23 answered: 0.157
- Analysts' goals different, content with much less than 100s (as evidenced from TREC Interactive Track questionnaires, the same ones we used as a group for TRECVID 2004)



Analysts Post-Topic Questionnaire Data

5-point scale, 1="Not at all" with 5="Very much"

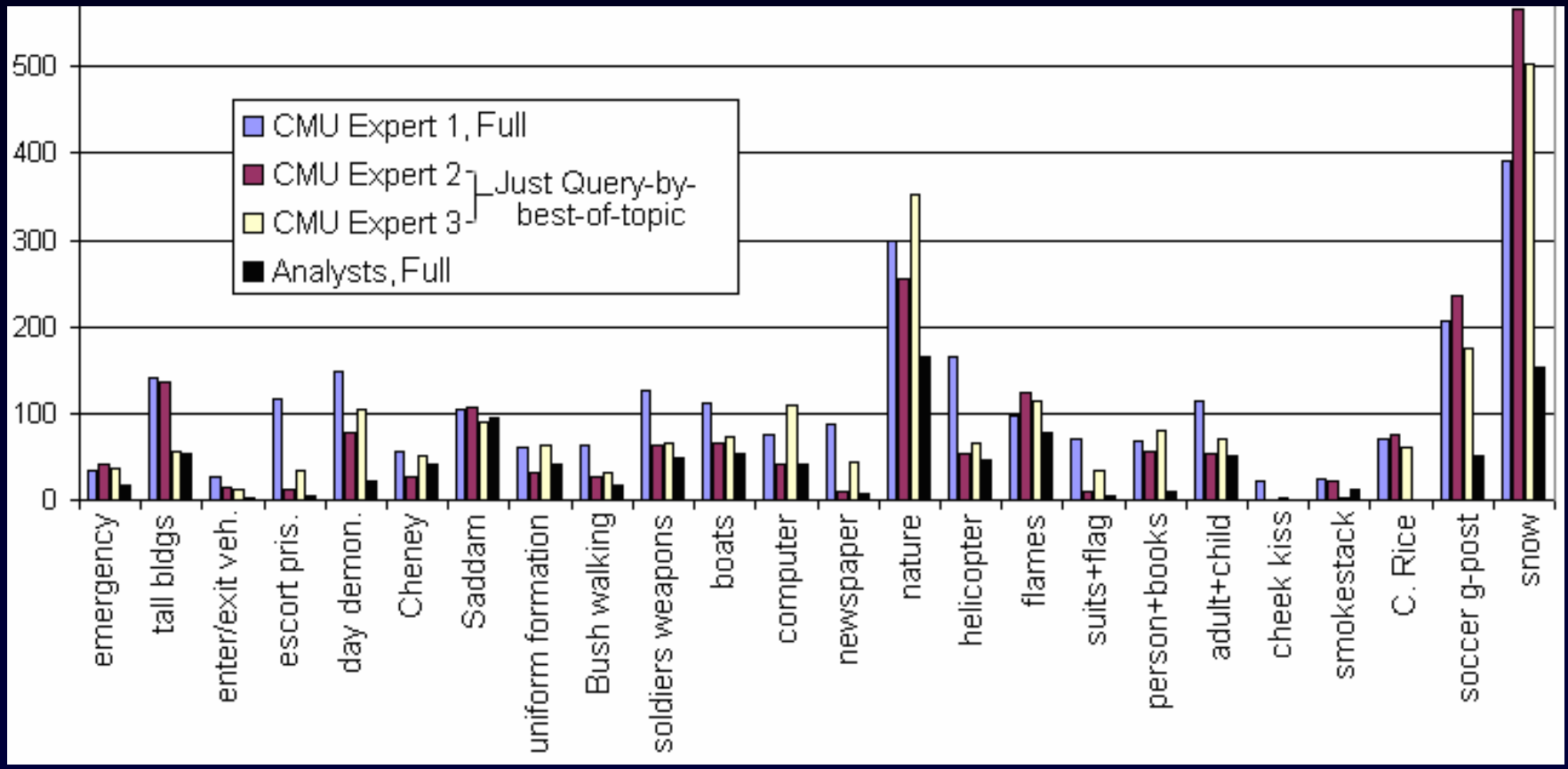
1. I found that it was easy to find shots that are relevant for this topic.
   - CMU Expert: 4.17 (easy to find shots)
   - Analysts: 3.83 (fairly easy to find shots)
2. For this topic I had enough time to find enough answer shots.
   - CMU Expert: 2.46 (not enough time)
   - Analysts: 4.21 (had more than enough time)
3. For this particular topic I was satisfied with the results of my search.
   - CMU Expert: 2.75 (not satisfied with results)
   - Analysts: 4 (satisfied with results)

Analysts Post-Topic Questionnaire Data

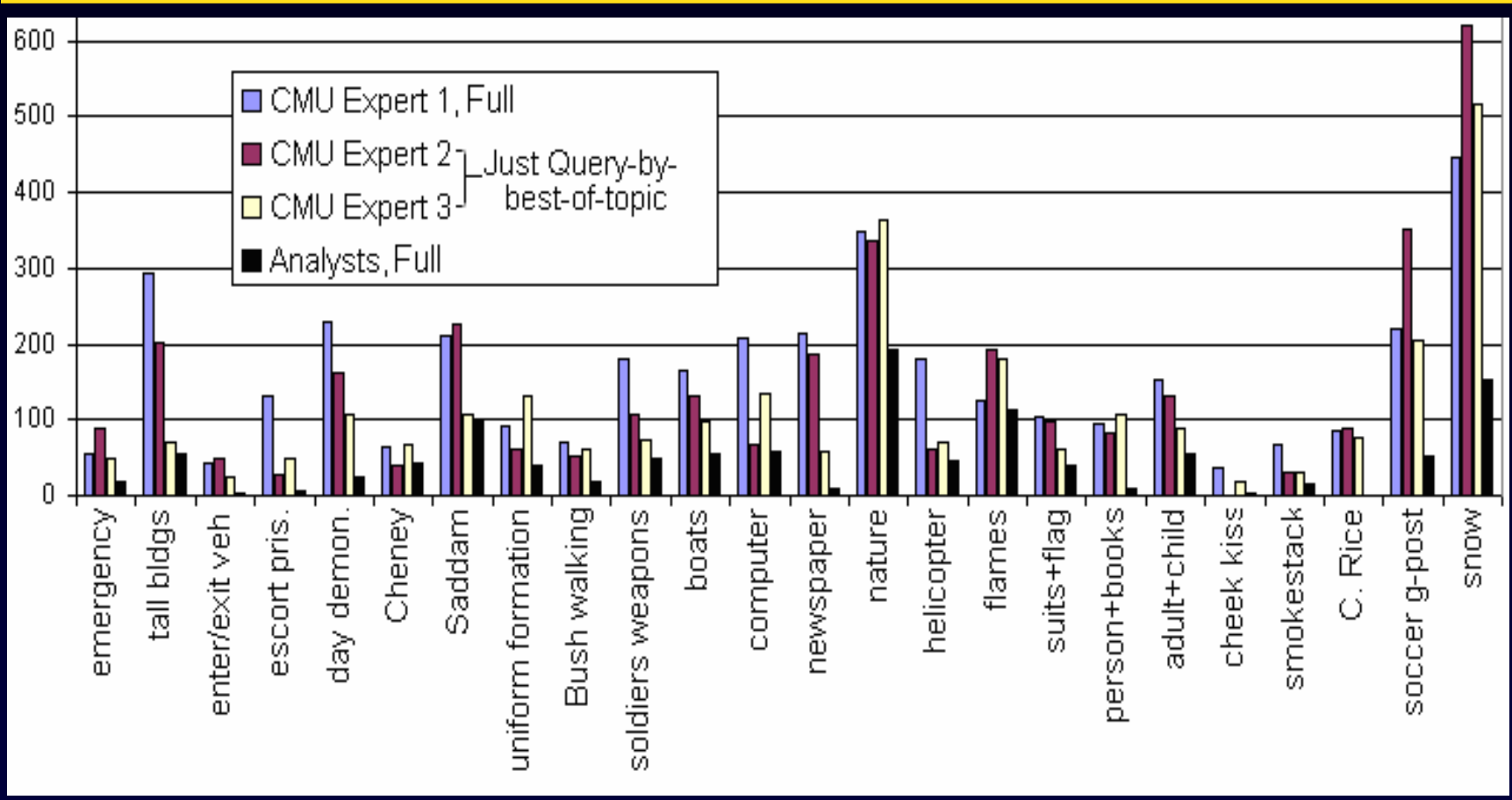
5-point scale, 1="Not at all" with 5="Very much"

1. I found that it was easy to find shots that are relevant for this topic.
   - CMU Expert: 4.17 (easy to find shots)
   - Analysts: 3.83 (fairly easy to find shots)
2. For this topic I had enough time to find enough answer shots.
   - CMU Expert: 2.46 (not enough time)
   - Analysts: 4.21 (had more than enough time)
3. For this particular topic I was satisfied with the results of my search.
   - CMU Expert: 2.75 (not satisfied with results)
   - Analysts: 4 (satisfied with results)

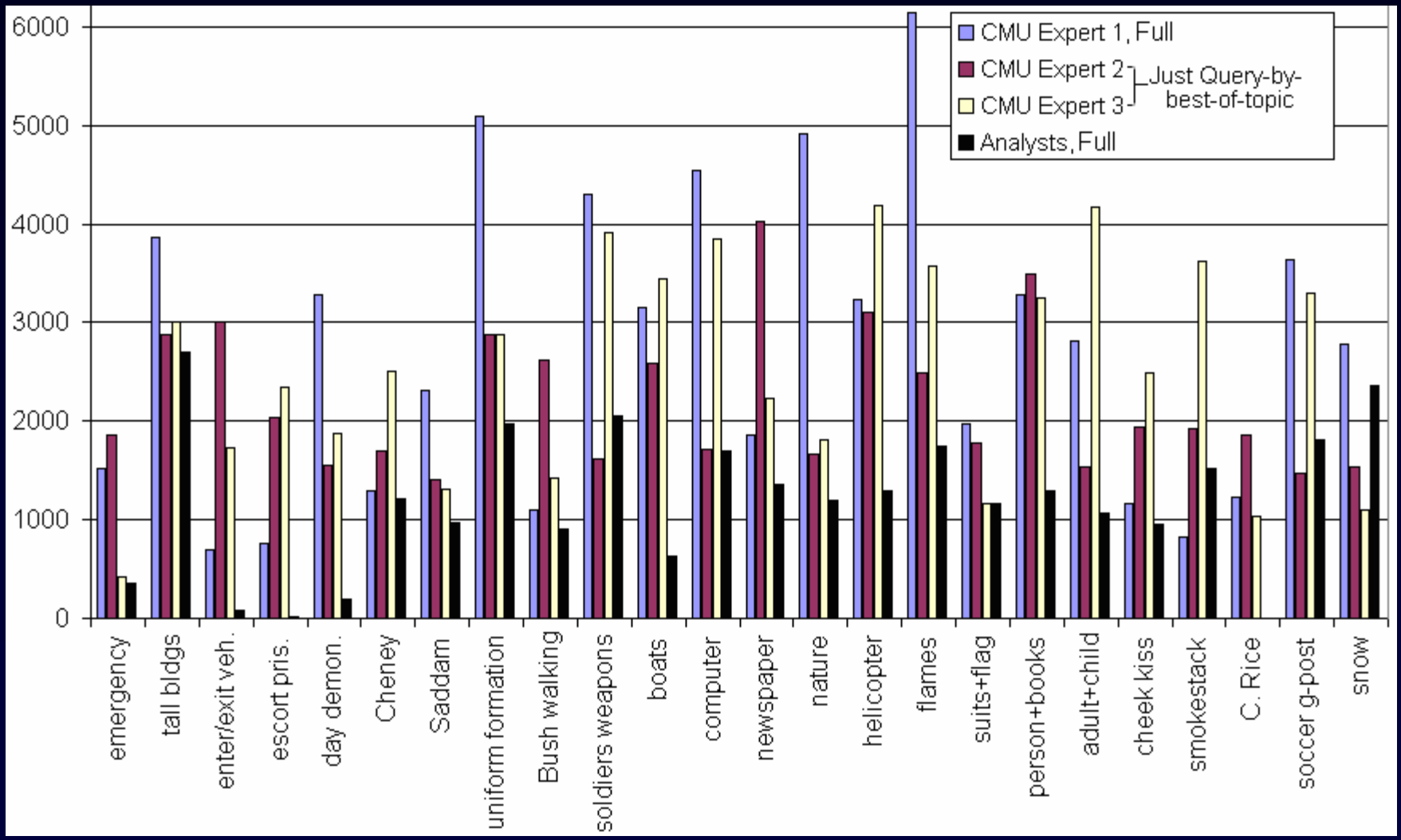
TRECVID "Yes" Shot Count per Topic



TRECVID "Yes" + "Maybe" Shots Per Topic



Reviewed Informedia Shot Count Per Topic

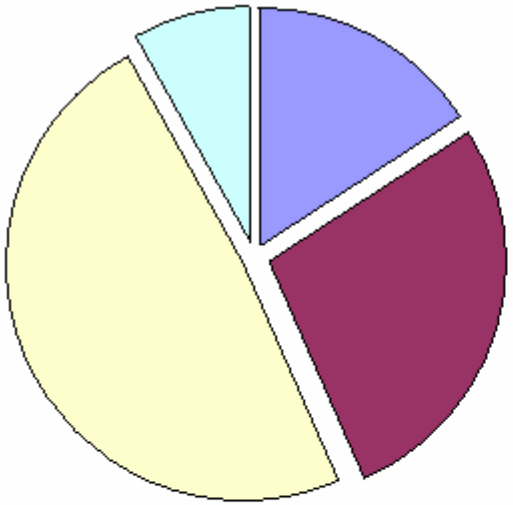


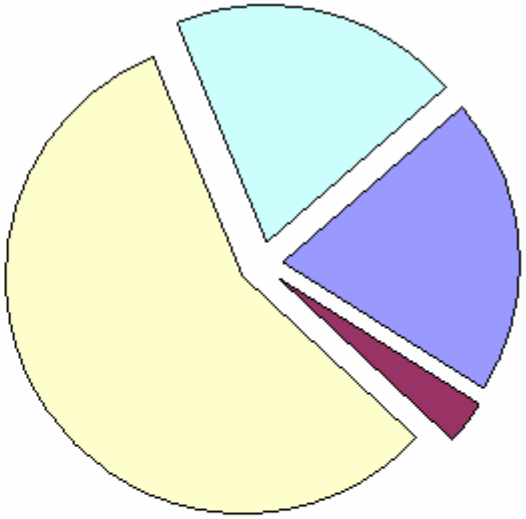
Average Reviewed Informedia Shots/Topic

Average Informedia shots reviewed per topic

Analysts, Full:	1194
XVR-MBRP:	1314
XVR-RSVP:	1364
CMU Expert 2 (“Hear”):	2195
CMU Expert 3 (“Hear”):	2526
CMU Expert 1, Full:	2740

System Usage, Full Informedia System Runs





Full Informedia (the "See" run) with CMU Developer

Full Informedia with Analysts

Conclusions from Analyst TRECVID Runs

- Lots of shots are successfully reviewed within 15 minutes (interface success!)
- Query-by-example, query-by-concept, and query-by-best-of-topic collectively were used much more than query-by-text, despite the analysts' high level of expertise with text retrieval and inexperience with video retrieval (success!)
- Performance is good, with room for growth
- Real users' tasks should be reconsidered.
  - What real-world task asks for great precision at 1000? Is precision at 100 a better metric?
  - Sports topics very different from other topic types.
  - Who are the users? What are the tasks? HCI fundamental questions that TRECVID has addressed by reference to Enser's work, BBC and CNN logs, etc. Is it time to revisit these questions?

CMU Search Run Conclusions – 1 of 2

- Automated search run an excellent starting point for interactive use, with “extreme” interfaces not necessary
- Relevance feedback and active learning approaches have great potential to help performance based on users’ input
- RSVP and system-controlled interface options will decrease precision of user response, and hence need more tuning use with machine learning
- Informedia interface successful in promoting multiple access strategies (image, text, LSCOM-lite concepts) for both system developers and also users new to the system

CMU Search Run Conclusions – 2 of 2

- Interesting future work as concept space grows from 10s to 100s, LSCOM-lite to LSCOM:
  - Will utility of “query-by-concept” also grow?
  - Will impact of relevance feedback to reweight semantic concepts and change shot ordering improve?
  - Will machine learning be useful in thinning concept options to a smaller recommended set for a given topic?
- More results mining to be conducted to determine value of confidence tagging of results (“Yes” and “Maybe” sets), and importance of auto-fill-to-1000 strategies
- Traditional Informedia “let the user drive” and XVR “system controls all” likely to merge in future work: video retrieval with ideal automated presets, plus user option to override

Thanks!

Thank you for your attention, and a special thanks to NIST and all of the evaluators whose collection, organization, management, and pooled truth generation make our work possible.