

The logo for 'informedia' is displayed in a stylized, lowercase font. Below it, the tagline 'digital video understanding' is written in a smaller, sans-serif font. The background of the header features a blurred image of a person's face and some abstract digital patterns.

**informedia**  
digital video understanding

SEARCH

summarize

visualize

retrieve

# **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 Infromedia 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](http://www.informedia.cs.cmu.edu) for more details



The screenshot shows the Informedia Project website. The header features the 'informedia' logo and the tagline 'digital video understanding'. Navigation links include 'SEARCH', 'summarize', 'visualize', 'retrieve', and 'digital video under'. A left sidebar lists various research areas: 'Current Research' (Informedia-II, Aquaint, CareMedia, Collaborations, Knowledge Discovery, YACE, Video Ontology), 'Past Research', 'Research Timeline', 'Publications', 'Demos/Downloads', 'Project Team' (highlighted), 'What's New', 'Site Map', and 'Home'. The main content area is titled 'The Informedia Project' and 'Automated digital video understanding research at Carnegie Mellon'. It lists the 'Project Team' and 'Principal Investigators, Faculty & Researchers', including Howard Wactlar, Michael Christel, Alex Hauptmann, Takeo Kanade, Ashok J. Bharucha, Christopher Atkeson, Mark Derthick, Christos Faloutsos, John Lafferty, Dorbin Ng, Henry Schneiderman, Scott Stevens, and Jie Yang. A 'Research Staff' section lists Robert Baron as the Principal Research Programmer.

**informedia** digital video understanding

SEARCH summarize visualize retrieve digital video under

**The Informedia Project**  
Automated digital video understanding research at Carnegie Mellon

**Project Team**

**Principal Investigators, Faculty & Researchers**  
[Howard Wactlar](#), Project Director and PI. Also Vice President, CMU  
[Michael Christel](#), Senior Systems Scientist, CSD  
[Alex Hauptmann](#), Senior Systems Scientist, CSD  
[Takeo Kanade](#), Director of Robotics Institute, RI  
[Ashok J. Bharucha](#), MD, Assistant Professor, University of Pittsburgh  
[Christopher Atkeson](#), Professor, Robotics Institute, CMU  
[Mark Derthick](#), Research Scientist, HCI  
[Christos Faloutsos](#), Associate Professor, CSD  
[John Lafferty](#), Associate Professor, CSD  
[Dorbin Ng](#), Systems Scientist, CSD  
[Henry Schneiderman](#), Research Scientist, RI  
[Scott Stevens](#), Senior Systems Scientist, HCI  
[Jie Yang](#), Senior Systems Scientist, HCI

**Research Staff**  
[Robert Baron](#), Principal Research Programmer



# Automatic Search

---

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

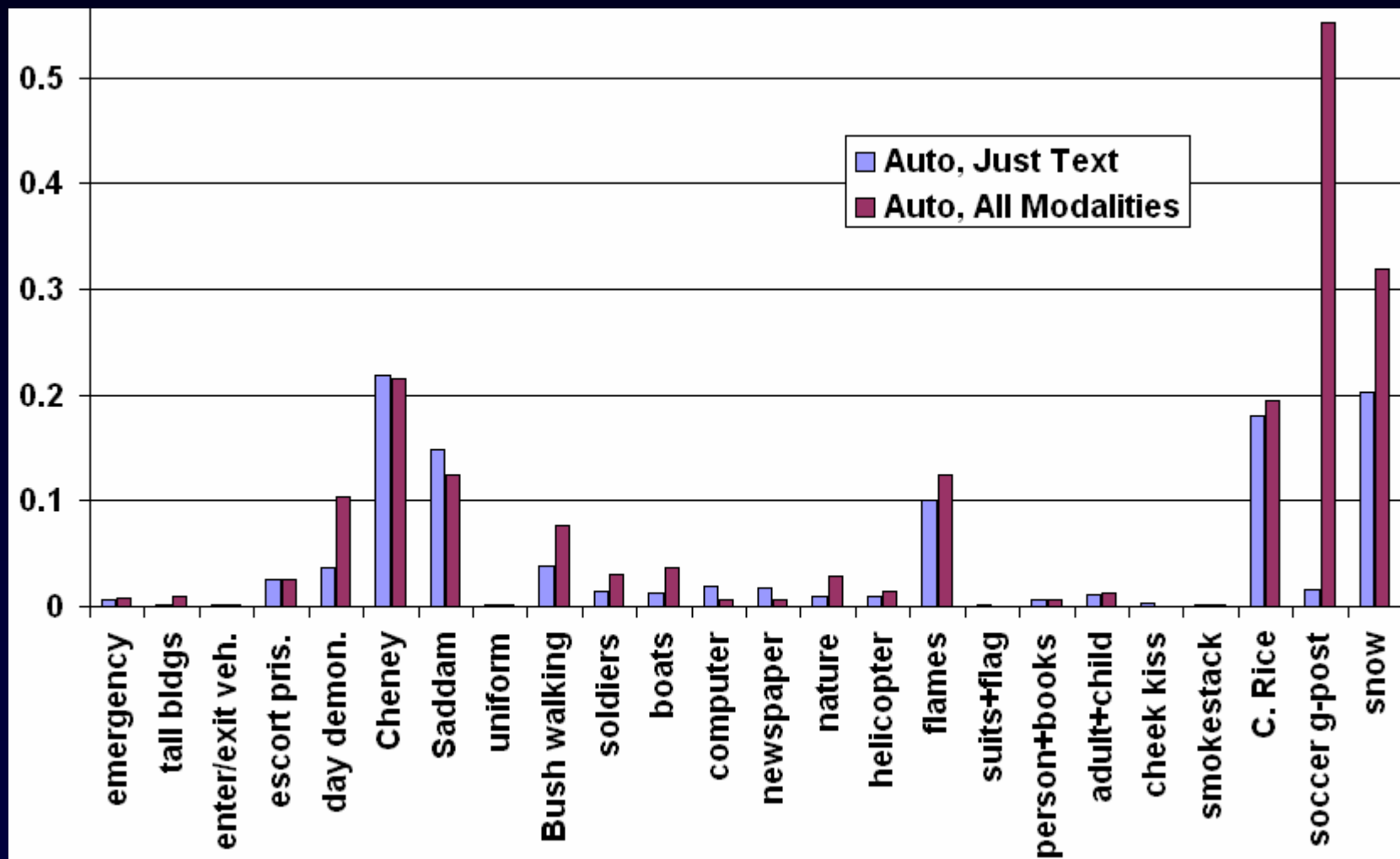
*Probabilistic Models for Combining Diverse Knowledge Sources in Multimedia Retrieval.* Ph.D. thesis, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, 2006

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

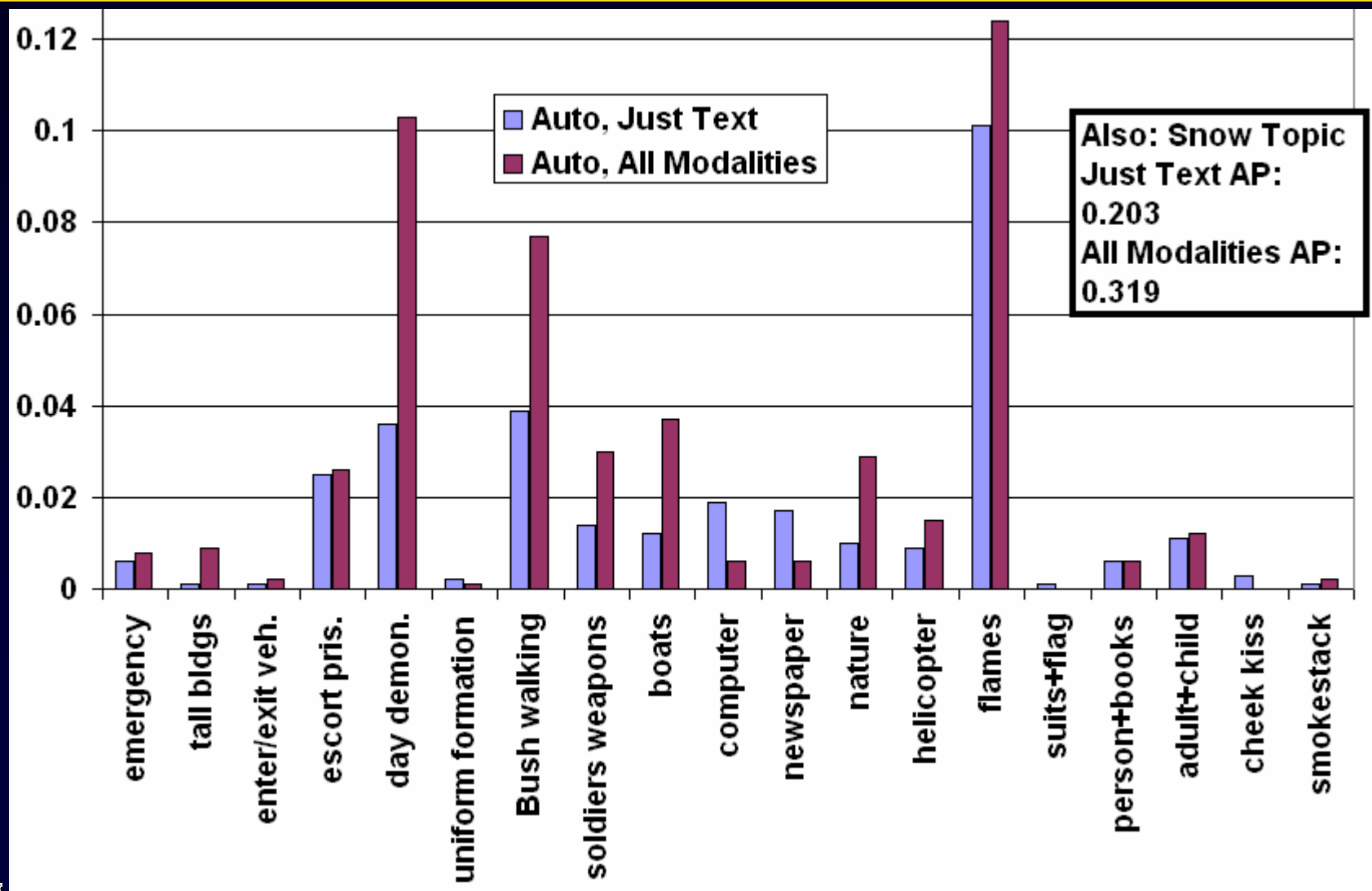


# MAP, Automatic Runs, Different Subsets

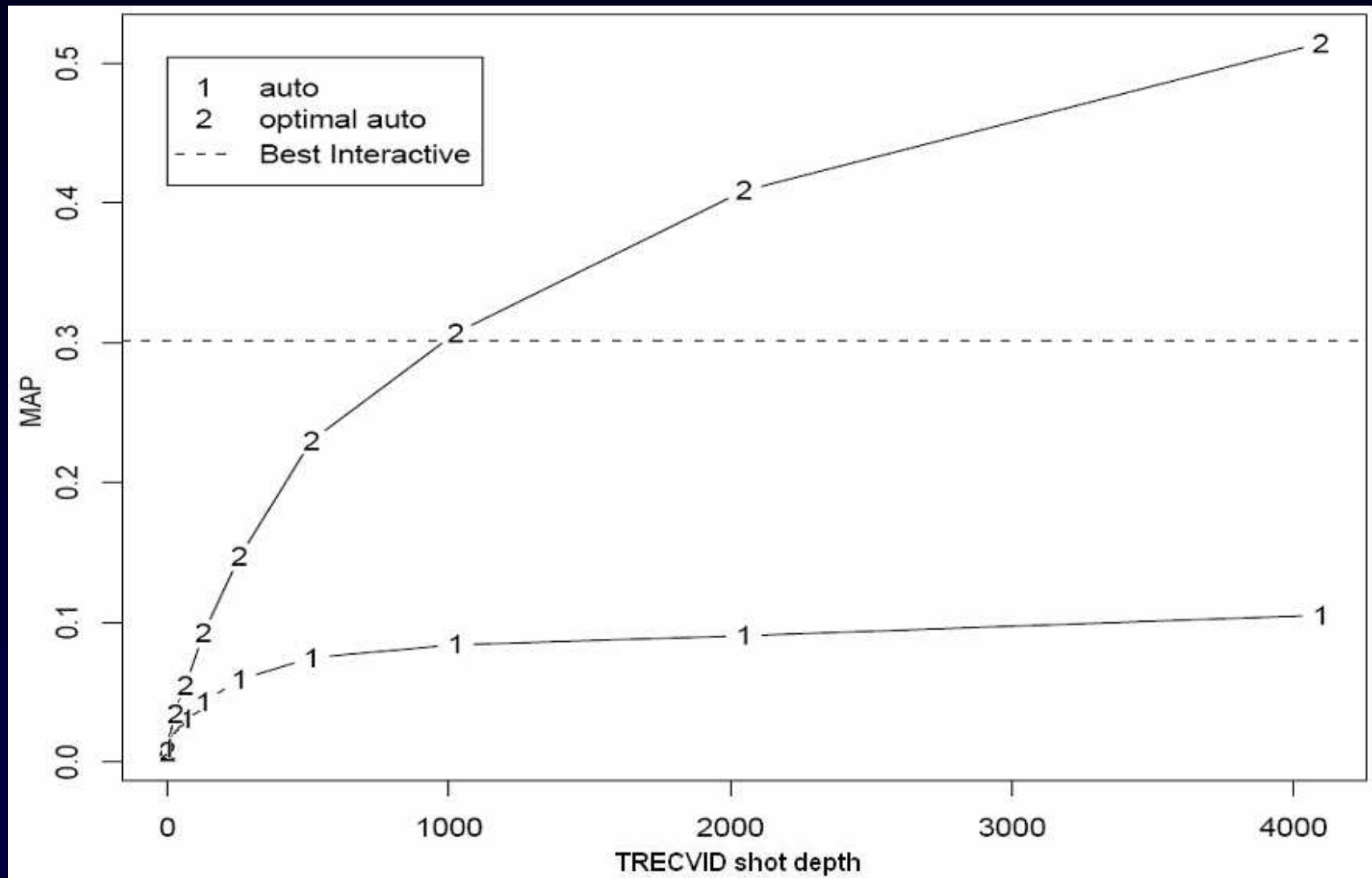
Topic Set Description	MAP Auto Text	MAP Auto All
All 24 Topics	0.045	0.079
Sports (just 195, soccer goalposts)	0.016	0.552
Non-Sports (all topics except for 195)	0.046	0.058
Specific (named people, 178, 179, 194 about Dick Cheney, Saddam Hussein, Condoleezza Rice)	0.183	0.178
Specific, including Bush 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

Informedia Digital Video Library (IDVL), funded through ARDA

Enter text above, then click "Search."

[Advanced Search](#)

Search

Condoleezza Rice --> topic as  
Chinese: '赖斯'; Arabic: كوندوليزا

Sign In

Next Topic

3:34 of 15:00

All data | Condoleezza Rice | Color-based search (276154000)

Sample images for topic #1. Drag into Search area t...

1 1/2 1/4 1/8



Launch color-based search... F3

Best Auto-Classification Results

Items

- Best animal shots
- Best automobile shots
- Best building shots
- Best explosion/fire shots
- Best map shots
- Best road shots
- Best US flag shots

People

Settings

Specific Named People (last name, first name)

- Best "Abbas,Mahmoud" shots
- Best "Annan,Kofi" shots
- Best "Arafat,Yasser" shots
- Best "Bin Laden,Osama" shots.txt
- Best "Blair,Tony" shots
- Best "Bush,George" shots
- Best "Carter,Jimmy" shots
- Best "Castro,Fidel" shots
- Best "Chirac,Jacques" shots

12 shots



Carnegie Mellon

# TRECVID Interface: 3 Main Access Strategies

Informedia Digital Video Library (IDVL), funded through ARDA

Enter text above, then click  
[Advanced Search](#)

*Query-by-text*

Condoleezza Rice --> topic as  
Chinese: '赖斯'; Arabic: كوندوليزا

Sign In

Next Topic

3:34 of 15:00

All data | Condoleezza Rice | Color-based search (276154000)

Sample images for topic #1. Drag into Search area t...

1 1/2 1/4 1/8



Launch color-based search... F3

*Query-by-image-example*

Best Auto-Classification Results

Items

- Best animal shots
- Best automobile shots
- Best building shots
- Best explosion/fire shots
- Best map shots
- Best road shots

*Query-by-concept*

Specific Named People (last name, first name)

- Best "Abbas,Mahmoud" shots
- Best "Annan,Kofi" shots
- Best "Arafat,Yasser" shots
- Best "Bin Laden,Osama" shots.txt
- Best "Blair,Tony" shots
- Best "Bush,George" shots
- Best "Carter,Jimmy" shots
- Best "Castro,Fidel" shots
- Best "Chirac,Jacques" shots

12 shots



# 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

Informedia Digital Video Library (IDVL), funded through ARDA

Enter text above, then click  
[Advanced Search](#)

*Query-by-text*

Condoleezza Rice --> topic as  
Chinese: '赖斯'; Arabic: كوندوليزا

Sign In

Next Topic

3:34 of 15:00

All data | Condoleezza Rice | Color-based search (276154000)

Sample images for topic #1. Drag into Search area t...

1 1/2 1/4 1/8



Launch color-based search... F3

*Query-by-image-example*

Best Auto-Classification Results

Items

- Best animal shots
- Best automobile shots
- Best building shots
- Best explosion/fire shots
- Best map shots
- Best road shots

*Query-by-concept*

Specific Named People (last name, first name)

- Best "Abbas,Mahmoud" shots
- Best "Annan,Kofi" shots
- Best "Arafat,Yasser" shots
- Best "Bin Laden,Osama" shots.txt
- Best "Blair,Tony" shots
- Best "Bush,George" shots
- Best "Carter,Jimmy" shots
- Best "Castro,Fidel" shots
- Best "Chirac,Jacques" shots

12 shots



# TRECVID 2006 Update: 4 Access Strategies

Informedia Digital Video Library (IDVL), funded through ARDA

Enter text above, then click  
[Advanced Search](#)

*Query-by-text*

Condoleezza Rice --> topic as  
Chinese: '赖斯'; Arabic: كوندوليزا

Sign In

Next Topic

3:34 of 15:00

12 shots

All data | Condoleezza Rice | Color-based search (276154000)

Sample images for topic #1. Drag into Search area t...

1 1/2 1/4 1/8



*Query-by-image-example*

Best Auto-Classification Results

Items

- Best animal shots
- Best automobile shots
- Best building shots
- Best explosion/fire shots
- Best map shots
- Best road shots

*Query-by-concept*

Specific Named People (last name, first name)

Topic14's Best Auto-Results

- Topic14\_1st1000
- Topic14\_2nd1000
- Topic14\_3rd1000
- Topic14\_4th1000

*Query-by-best-of-topic*



Carnegie Mellon



# Example: Best-of-Topic (Emergency Vehicles)





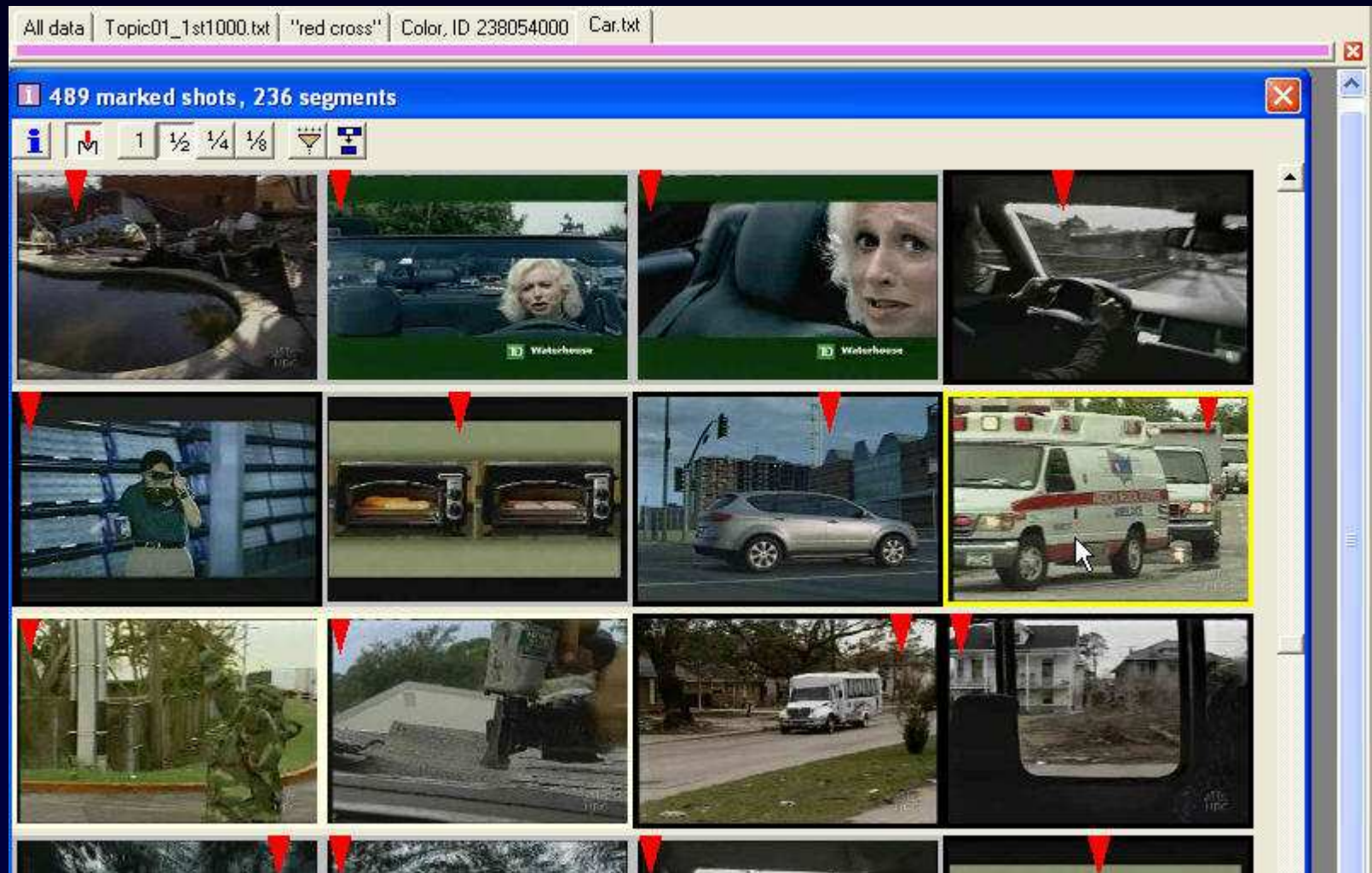
# Example: Query by Text “Red Cross”



[illegible]



# 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.

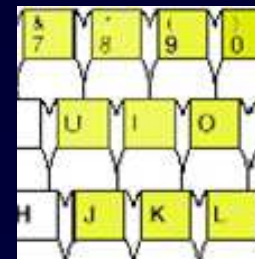
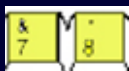
Hence, query-by-best-of-topic added into the “classic” interface.

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

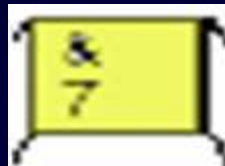
- (i) manual browsing with resizable pages (MBRP)
- (ii) rapid serial visual presentation (RSVP) with system-controlled presentation intervals



# 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?

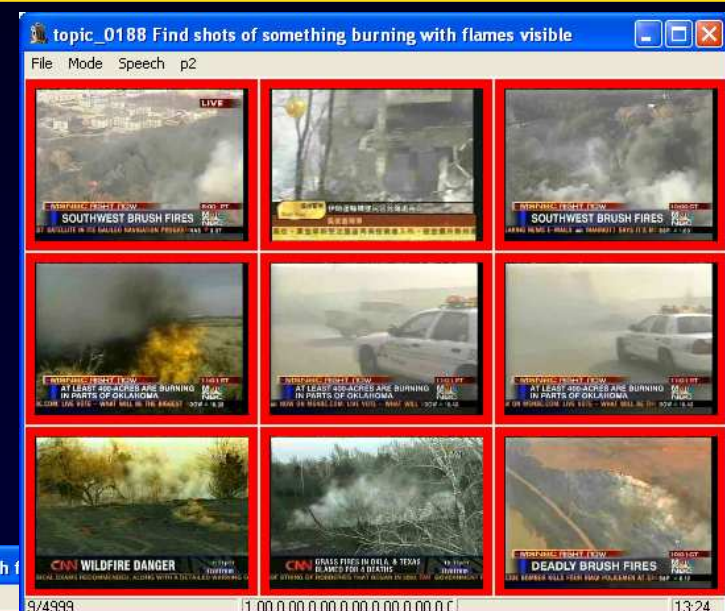
Based on user input, the positive examples are easily noted (the chosen/marked shots) with precision at very high 90+% levels based on prior TRECVID analysis of user input . Negative examples are less precise, but are the set of “overlooked” shots passed over when selecting relevant ones.

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...

### topic\_0188 Find shots of something burning with flames visible

File Mode Speech p2

 OKLAHOMA FIRE A THREAT TO SURVIVIT NEXT WEEK'S CANNON PROCEEDINGS... 1:00 PT	 DEADLY BRUSH FIRES A NEWER BRUSH FIRES... 1:00 PT	 WESTERN FLOODING A NEWER OF IN CANNON... 1:00 PT
 MUSTANG, OK A NEWER PERSONAL... 1:00 PT	 CNN WILDFIRE DANGER A NEWER PERSONAL... 1:00 PT	 TARRANT COUNTY, TX A NEWER PERSONAL... 1:00 PT
 TARRANT COUNTY, TX A NEWER PERSONAL... 1:00 PT	 NEAR SHAWNEE, OK A NEWER PERSONAL... 1:00 PT	 NEAR SHAWNEE, OK A NEWER PERSONAL... 1:00 PT

0/4999 14:09

### topic\_0188 Find shots of something burning with flames visible

File Mode Speech p2

 SOUTHWEST BRUSH FIRES A NEWER PERSONAL... 1:00 PT	 SOUTHWEST BRUSH FIRES A NEWER PERSONAL... 1:00 PT	 SOUTHWEST BRUSH FIRES A NEWER PERSONAL... 1:00 PT
 AT LEAST 100 ACRES ARE BURNING IN PARTS OF OKLAHOMA... 1:00 PT	 AT LEAST 100 ACRES ARE BURNING IN PARTS OF OKLAHOMA... 1:00 PT	 AT LEAST 100 ACRES ARE BURNING IN PARTS OF OKLAHOMA... 1:00 PT
 CNN WILDFIRE DANGER A NEWER PERSONAL... 1:00 PT	 GRASS FIRES IN OKLAHOMA & TEXAS A NEWER PERSONAL... 1:00 PT	 DEADLY BRUSH FIRES A NEWER PERSONAL... 1:00 PT

3/4999 13:24

 SOUTHWEST BRUSH FIRES A NEWER PERSONAL... 1:00 PT	 AT LEAST 100 ACRES ARE BURNING IN PARTS OF OKLAHOMA... 1:00 PT	 AT LEAST 100 ACRES ARE BURNING IN PARTS OF OKLAHOMA... 1:00 PT
 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT
 SOUTHWEST BRUSH FIRES A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT

18/4999 13:10

 SOUTHWEST BRUSH FIRES A NEWER PERSONAL... 1:00 PT	 AT LEAST 100 ACRES ARE BURNING IN PARTS OF OKLAHOMA... 1:00 PT	 AT LEAST 100 ACRES ARE BURNING IN PARTS OF OKLAHOMA... 1:00 PT
 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT
 SOUTHWEST BRUSH FIRES A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT	 WHITE-HOT ATTACKS? A NEWER PERSONAL... 1:00 PT

13:10



## 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

**This Topic's Best Auto-Results**

- Topic16\_1st1000
- Topic16\_2nd1000
- Topic16\_3rd1000
- Topic16\_4th1000
- Topic16\_5th1000

 Go

[illegible]



# Informedia Storyboard Under User Control



# Informedia Storyboard with Concept Filters





# **TRECVID 2006 CMU Interactive Search Runs**

---

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



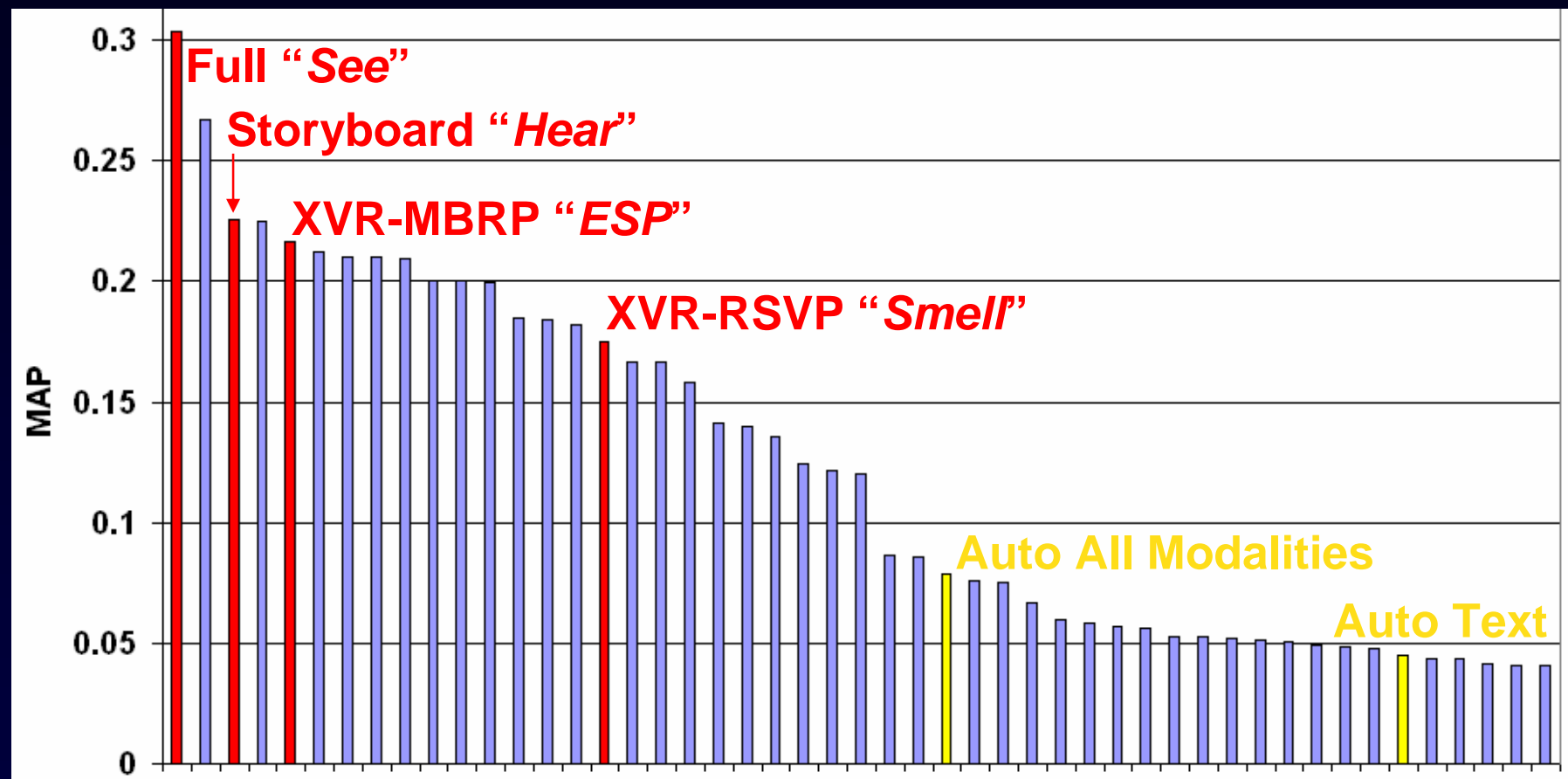
# TRECVID 2006 CMU Interactive Search Runs

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

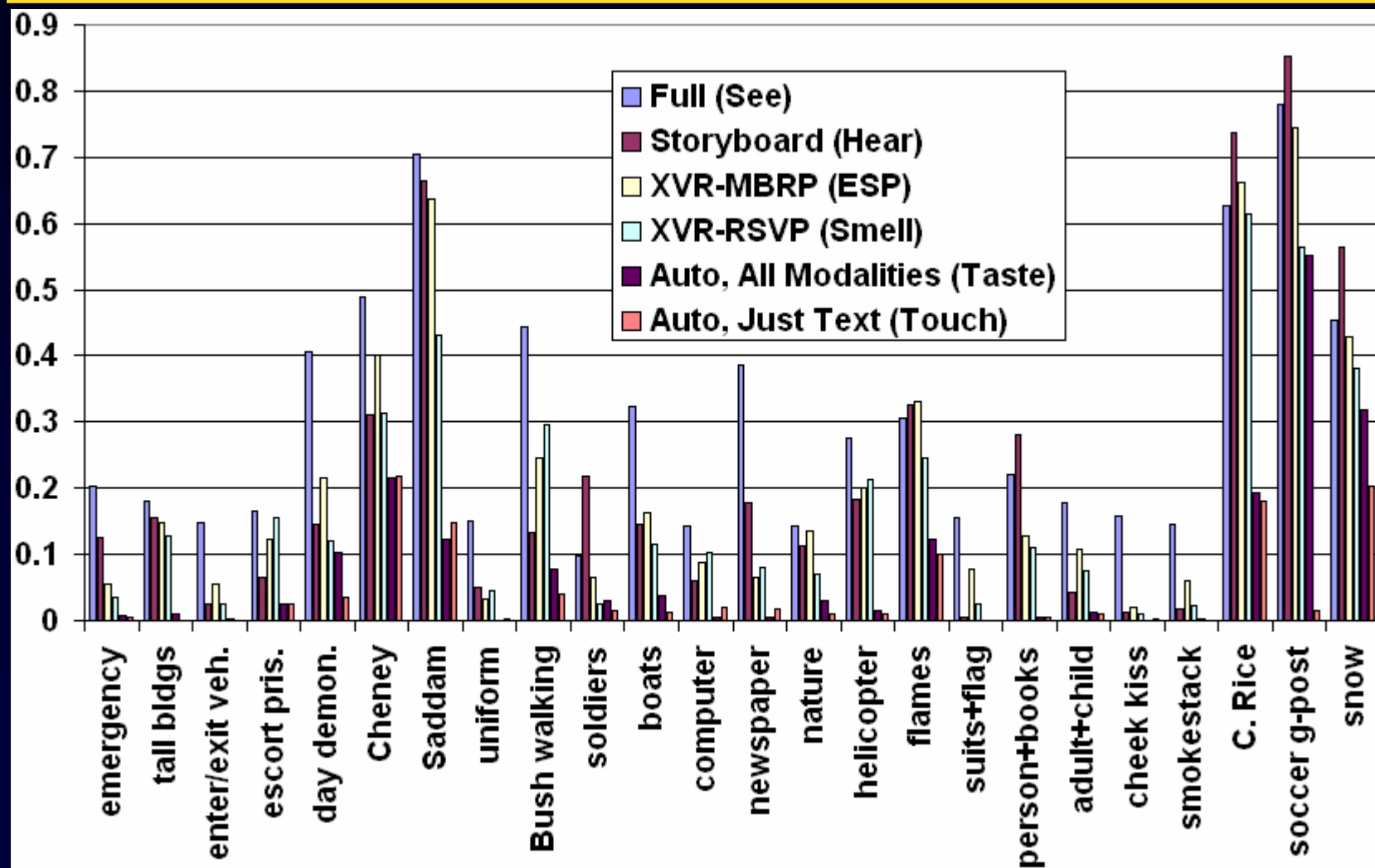
- Automatic output does hold value in interactive users' hands
- Learning strategies confounded in RSVP (2 shots marked per interaction, but 1 was almost always wrong)
- Additional capability (to query by text, image, concept) leads to improved performance with the "See" run



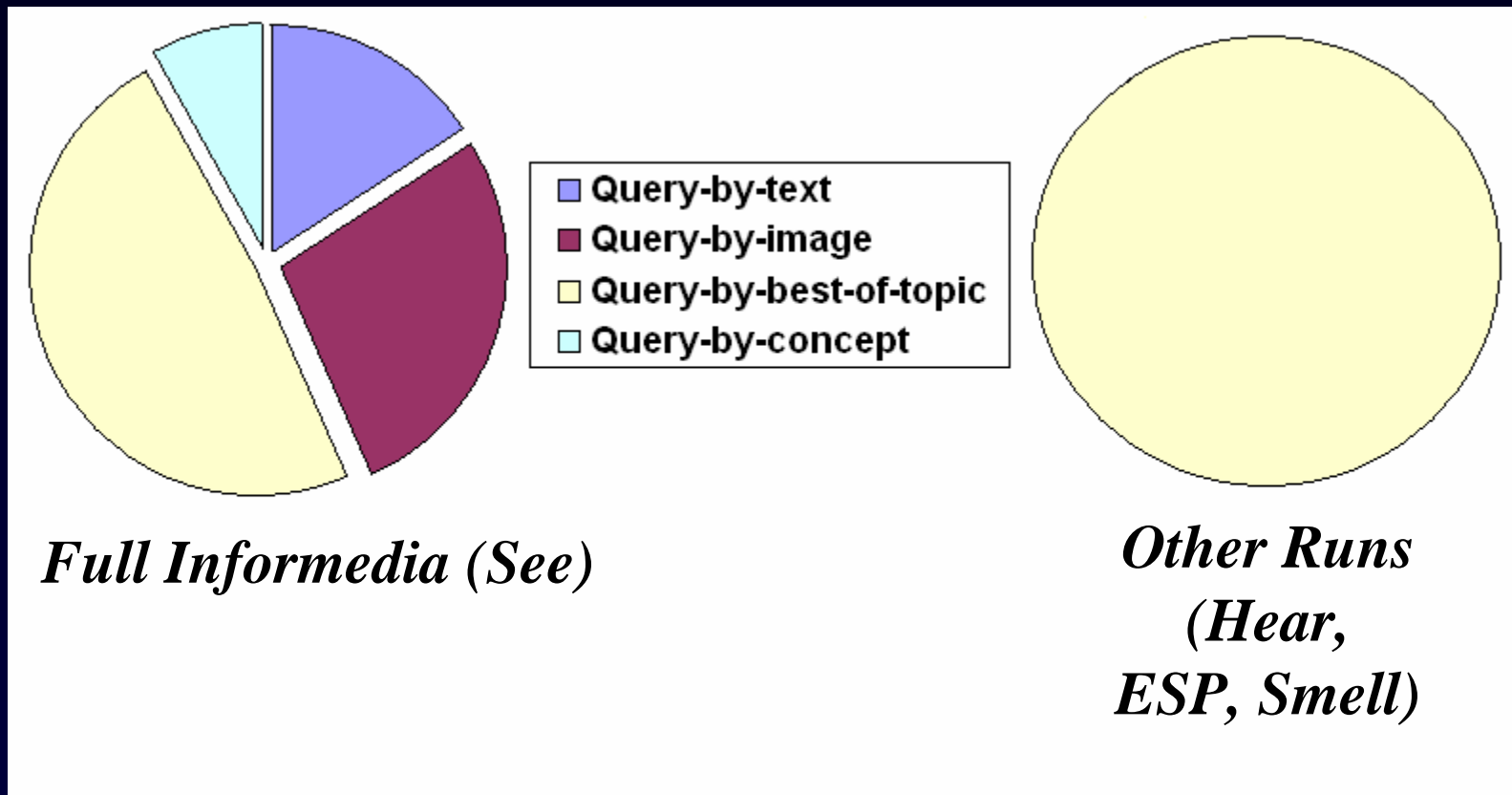
# MAP Top 50 Search Runs



# Average Precision, CMU Search Runs



# System Usage, CMU Interactive Runs



## 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

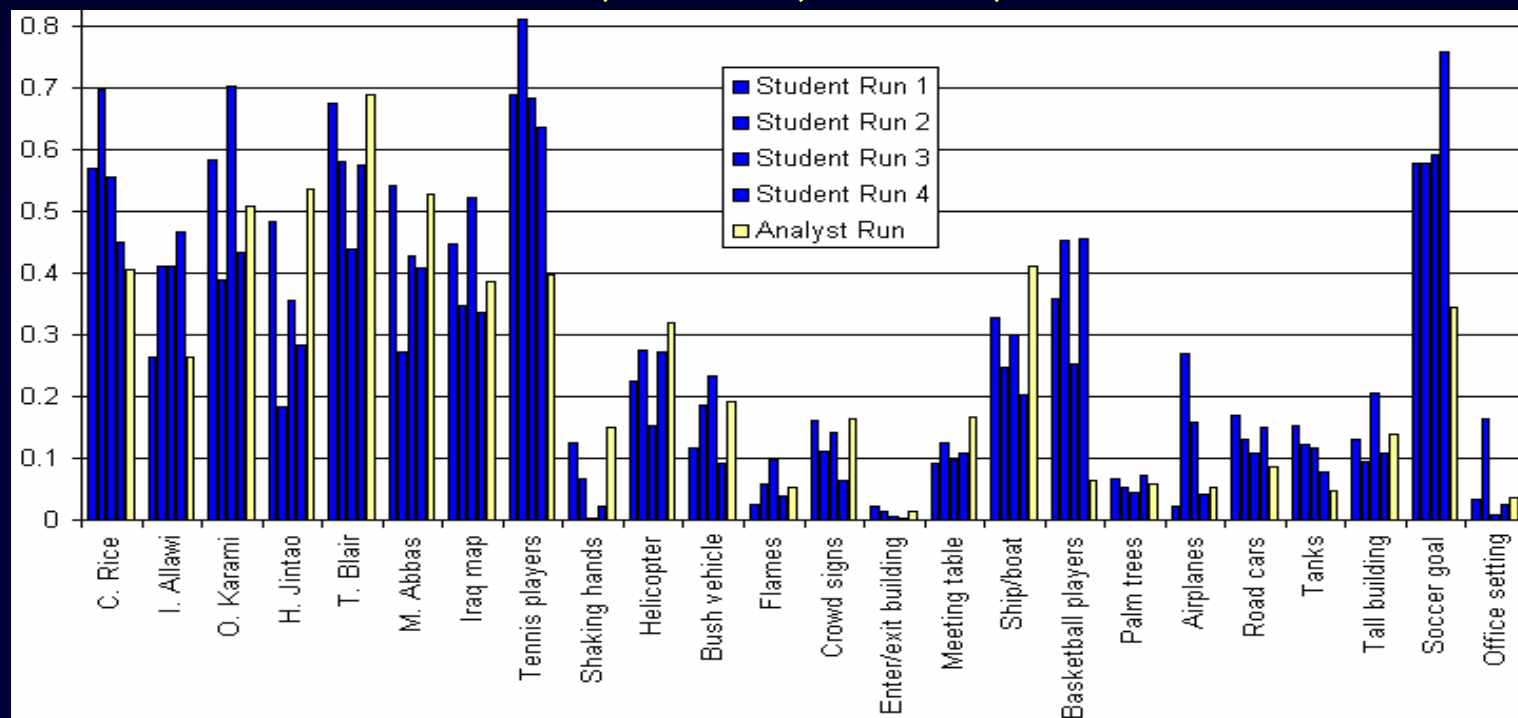
---

- 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



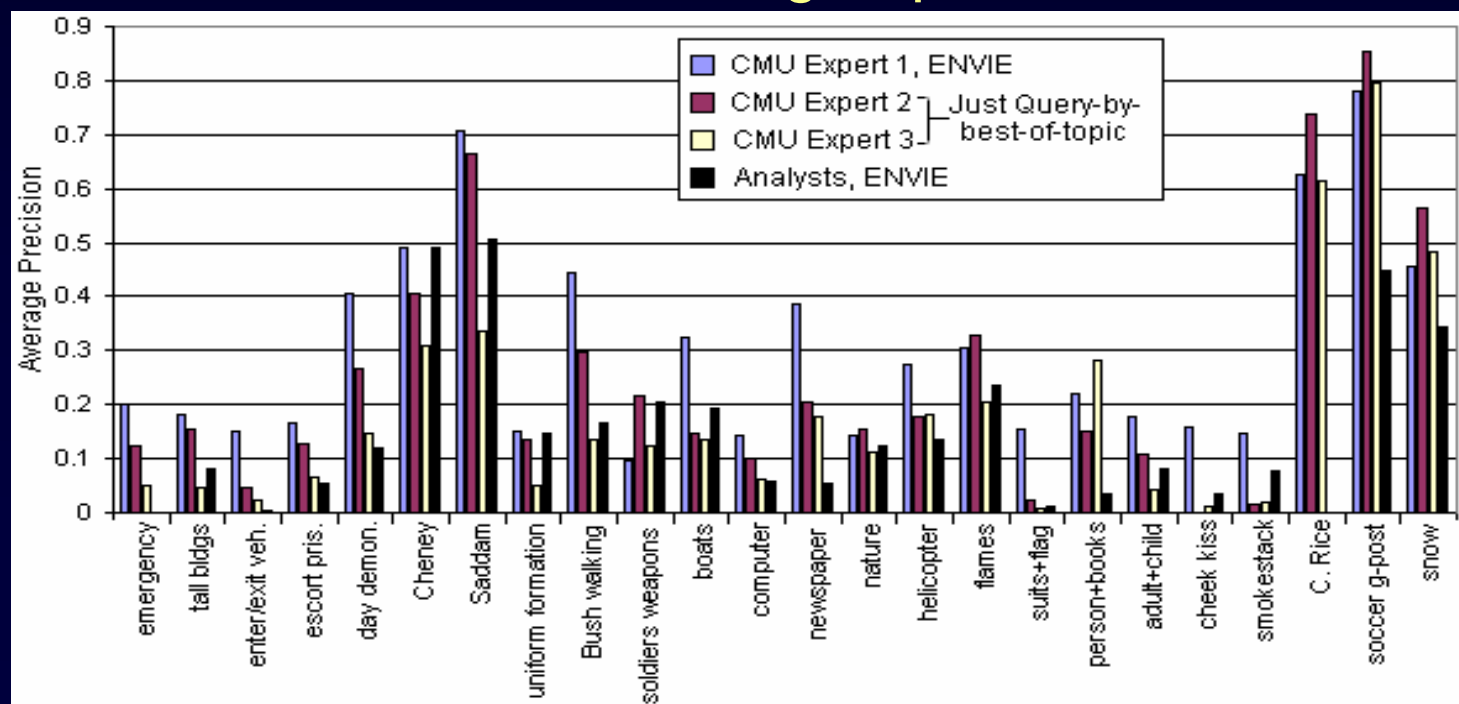
# Analysts, Quick Look Back at TRECVID 2005

- 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)

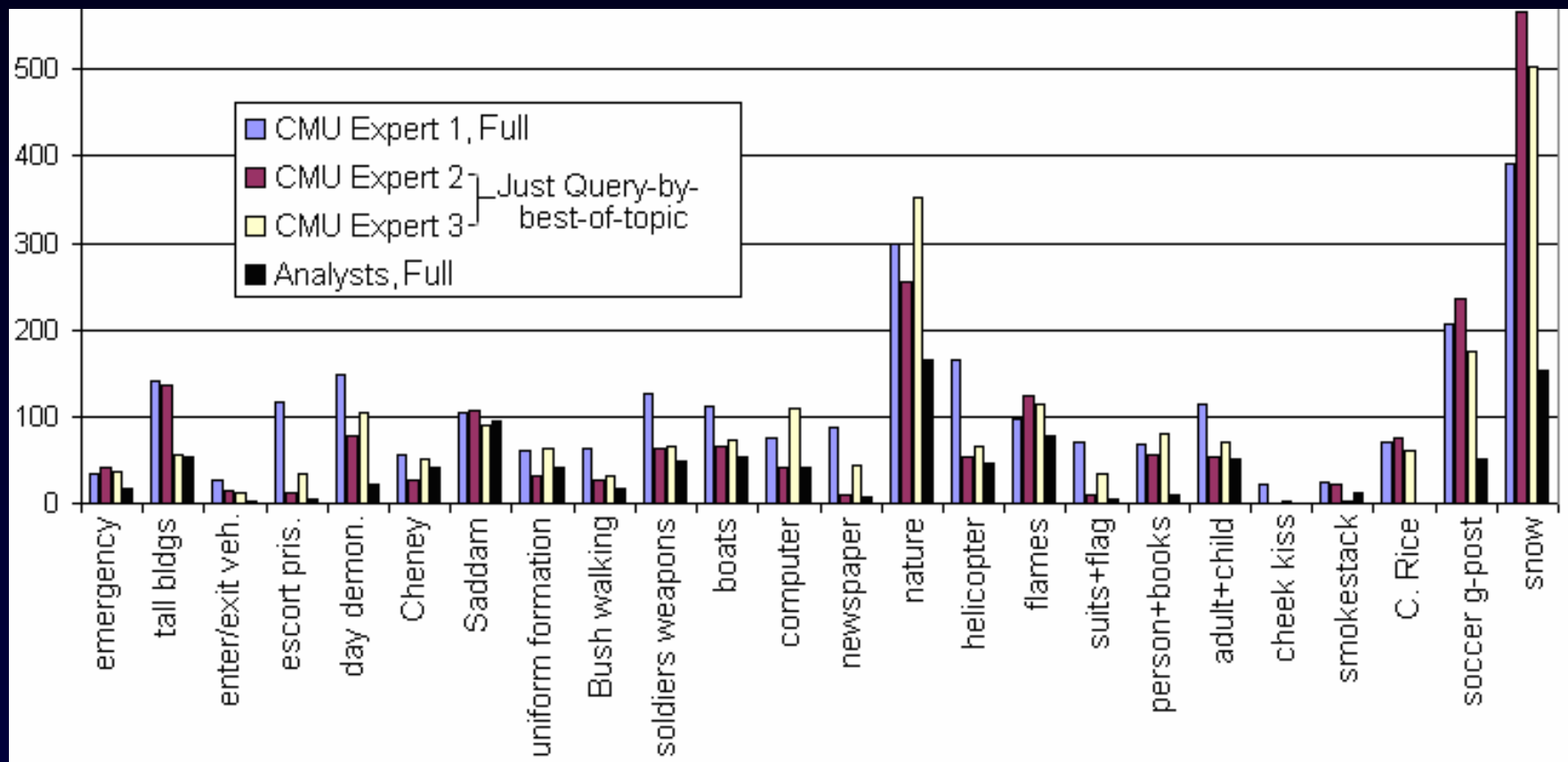
3. For this particular topic I was satisfied with the results of my search.

- CMU Expert: 2.75 (not satisfied with results)
- Analysts: 4 (satisfied with results)

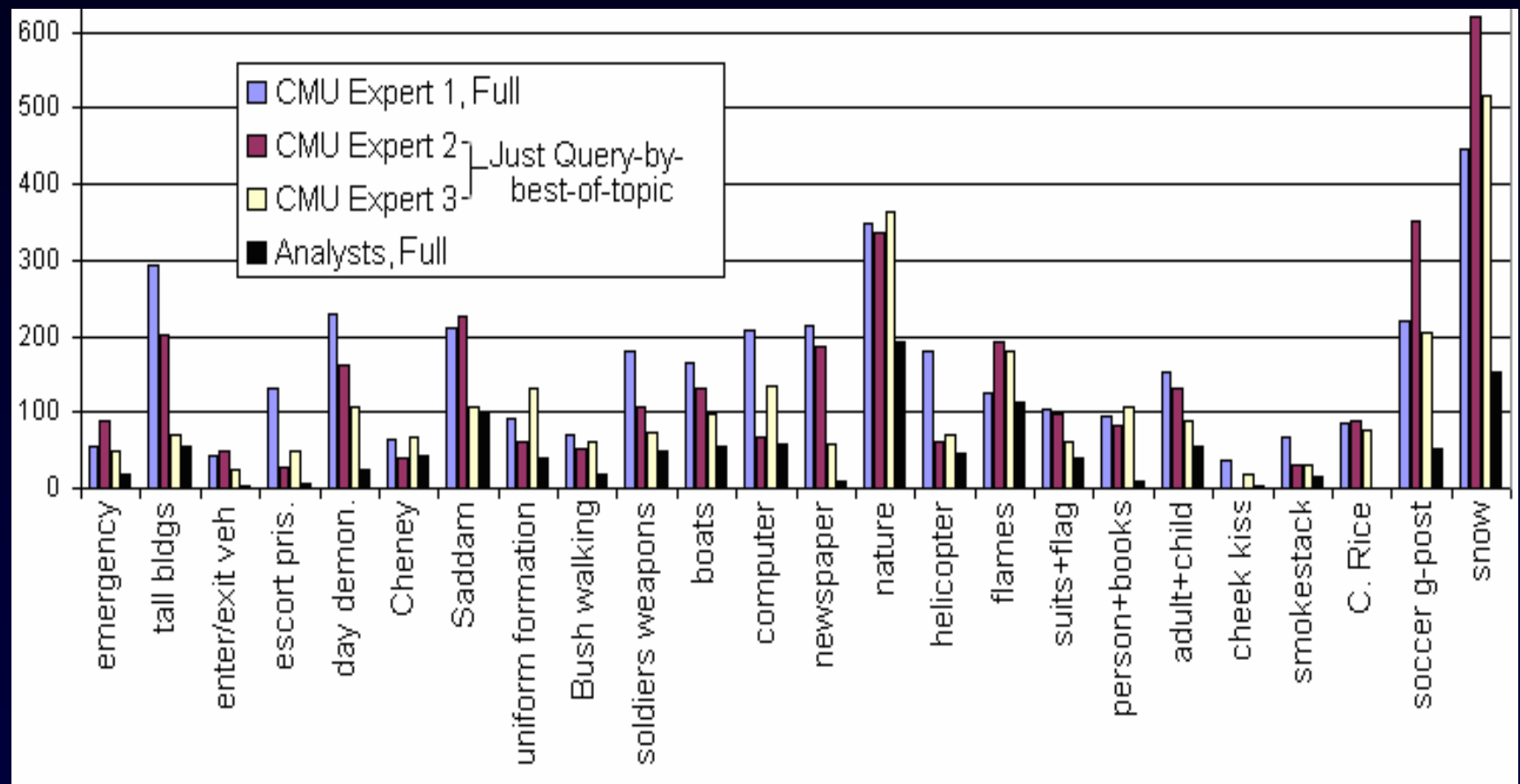




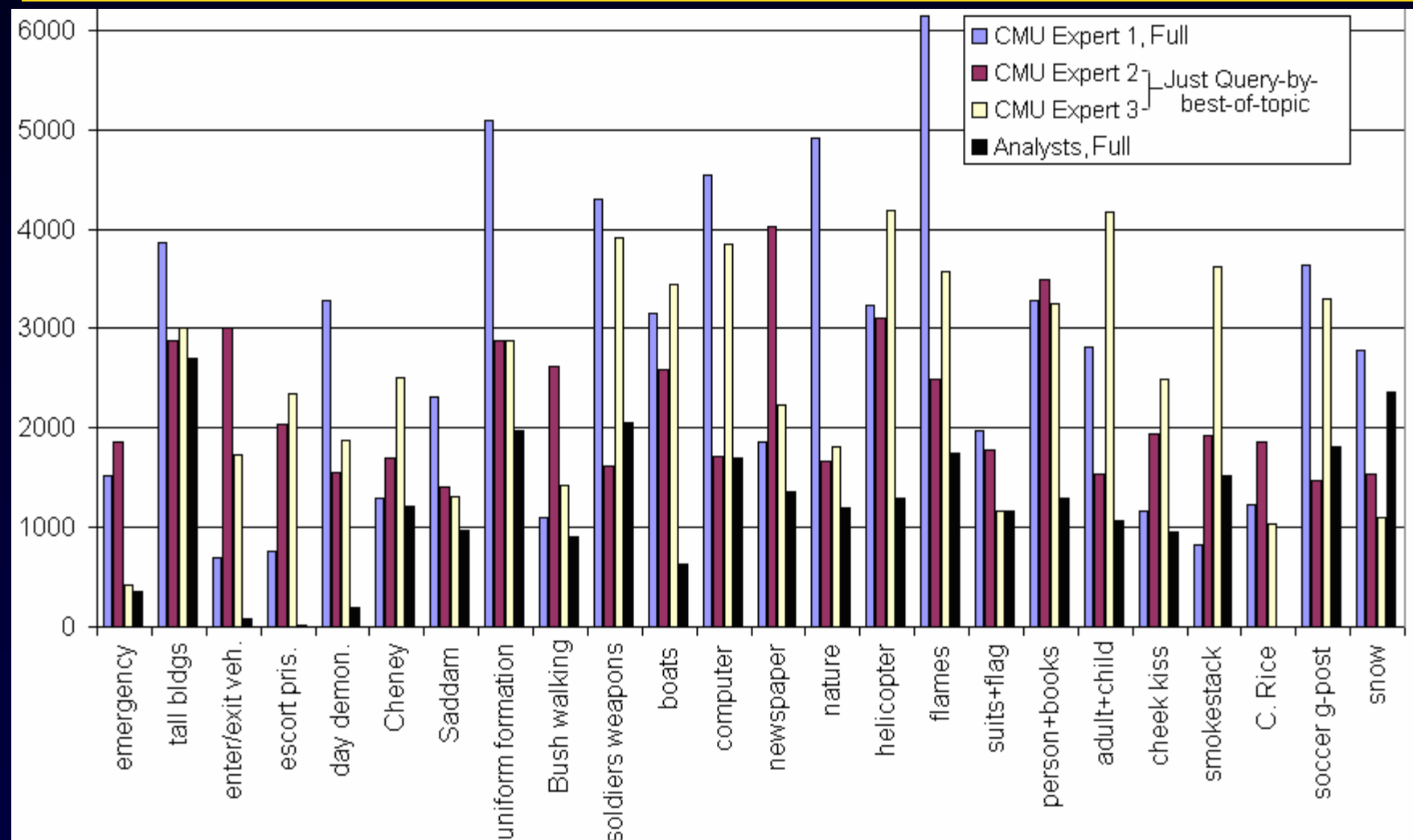
# TRECVID “Yes” Shot Count per Topic



# TRECVID “Yes” + “Maybe” Shots Per Topic



# Reviewed Informedia Shot Count Per Topic



# Average Reviewed Informedia Shots/Topic

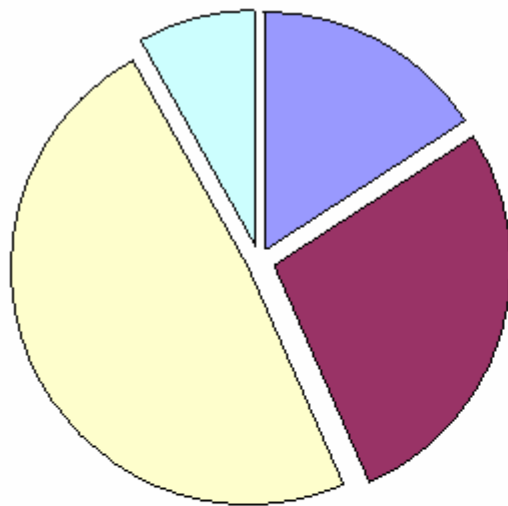
Average Informedia shots reviewed per topic

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

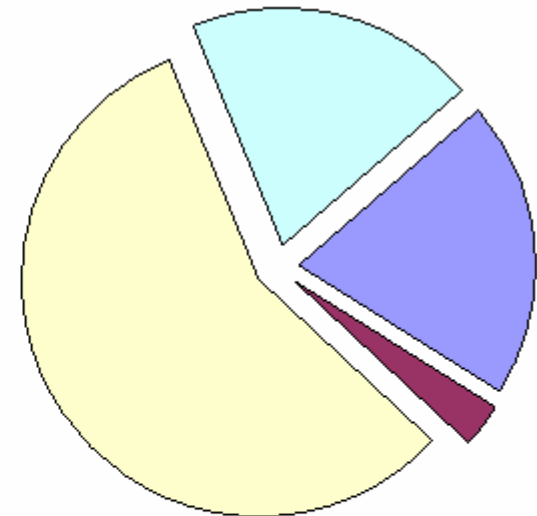
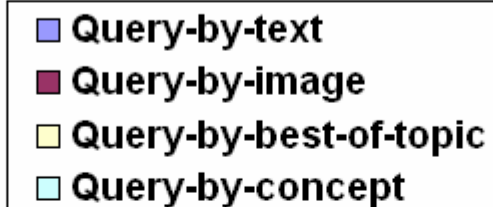




# System Usage, Full Informedia System Runs



*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.

