
TRECVID-2007: Search Task

Alan Smeaton
Dublin City University
&
Paul Over
NIST

Search Task Definition

- ❑ Goal: promote progress in content-based retrieval from digital video via open, metrics-based evaluation
- ❑ Given a test collection, a topic and a common shot boundary reference, return a ranked list of at most 1,000 shots which best satisfy the need
- ❑ Test and training videos were viewed by NIST personnel, notes taken on content, topic candidates chosen, examples added from development set and Web
- ❑ Different in 2007
 - more topics asking for **generic** (vs. **specific, named**) targets
 - Almost 1/2 topics ask for video of an event – encouraging exploration beyond *one-keyframe-per-shot*

Search Task Definition

- ❑ **Per-search** measures: average precision, elapsed time
- ❑ **Per-run** measure: mean average precision (MAP)
- ❑ Interactive search participants were asked to have their subjects complete pre, post-topic and post-search questionnaires;
- ❑ Each result for a topic can come from only 1 user search; same searcher does not need to be used for all topics.
 - EXCEPT: experimental **collaborative runs** from FX Palo Alto Labs.

2007 data different from 2003-6 data

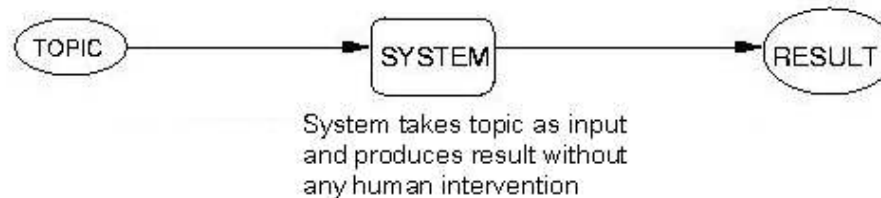
- ☐ Educational, cultural, youth-oriented programming, news magazines, historical footage, etc.
- ☐ Primarily in Dutch
- ☐ Much less repetition
 - No commercials
 - No repeated stock news footage
 - Greater variety of subject matter

2007: Search task participants

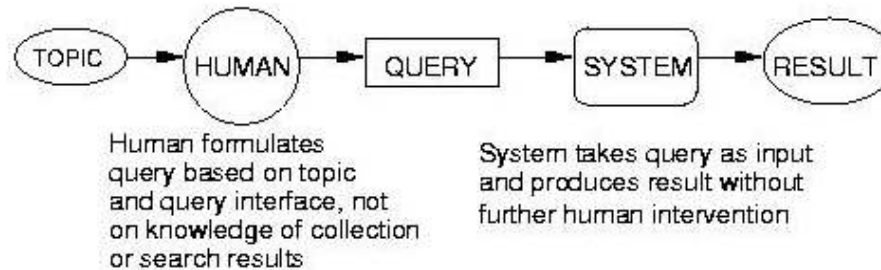
Beijing Jiaotong University	-- -- SE --
Bilkent University	** FE SE **
City University of Hong Kong (CityU)	-- FE SE SU
COST292 Team	SB FE SE SU
CWI-CTIT-U.Twente	-- ** SE --
Dublin City University	-- -- SE SU
Etter Solutions Research Group	-- -- SE --
Fudan University	-- FE SE -
FX Palo Alto Laboratory Inc.	** ** SE SU
Helsinki University of Technology	** FE SE SU
IBM T. J. Watson Research Center	** FE SE **
INESC-Porto	-- ** SE --
K-Space	-- FE SE --
Microsoft Research Asia	-- FE SE **
Multimedia Computing Group (CAS) / Natl. U. of Singapore	-- FE SE **
Oxford University	-- FE SE --
Tsinghua University / Intel China Research Center	SB FE SE SU
Universidad de Jaén (SINAI)	-- -- SE --
University of Amsterdam (MediaMill Team)	-- FE SE --
University of California, Santa Barbara	-- FE SE SU
University of Central Florida	-- FE SE **
University of Glasgow	-- -- SE SU
University of Iowa	** FE SE --
University of Queensland	-- -- SE --

Search Types: Automatic, Manual and Interactive

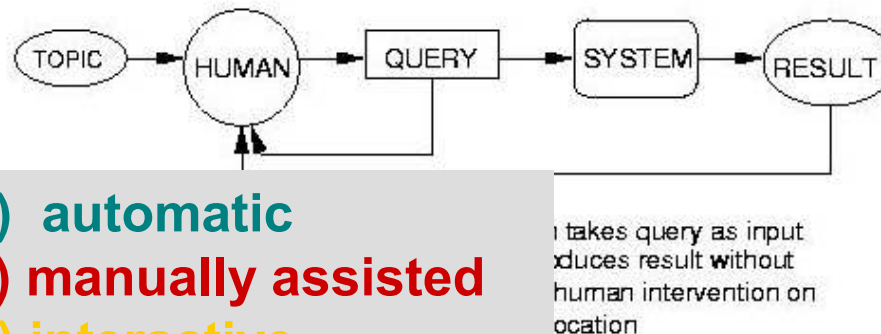
AUTOMATIC :



MANUAL :

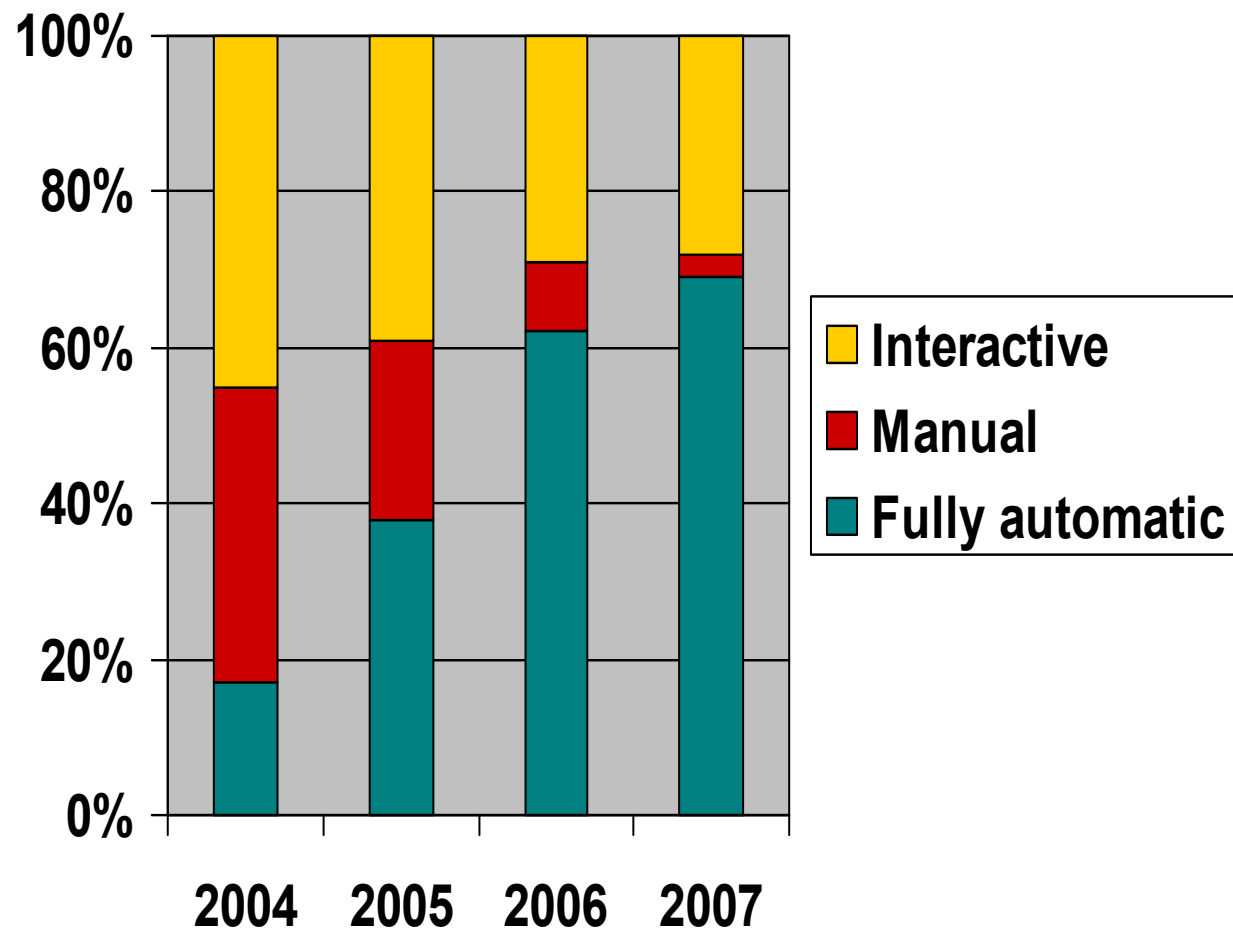


INTERACTIVE :



Number of runs: **81 (76) automatic**
4 (11) manually assisted
33 (36) interactive

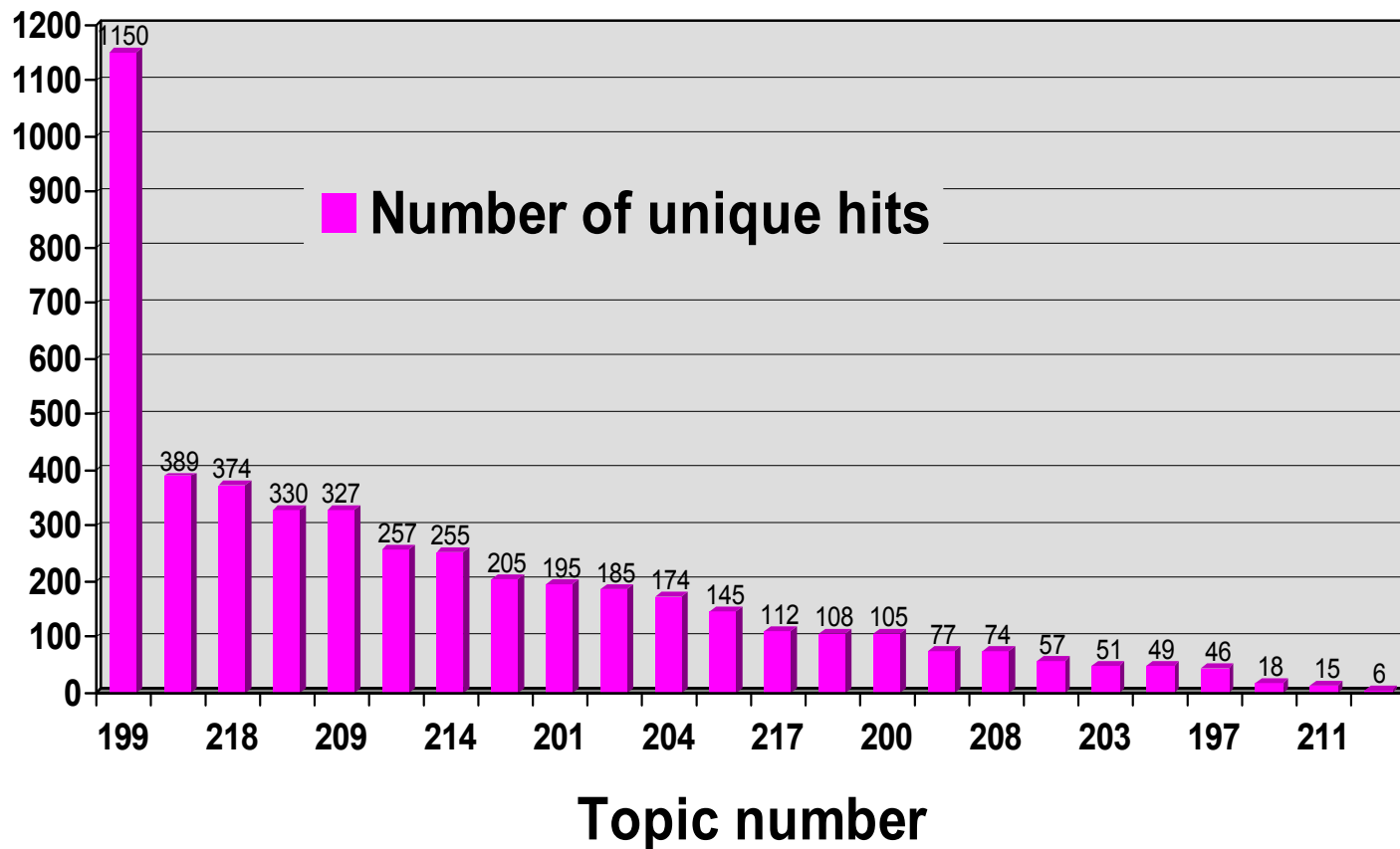
Trends continue



24 Topics (■ events)

- | | |
|--|--|
| ■ Find shots of one or more people walking up stairs. | □ Find shots with 3 or more people sitting at a table. |
| ■ Find shots of a door being opened. | ■ Find shots with one or more people walking with one or more dogs. |
| ■ Find shots of a person walking or riding a bicycle. | □ Find shots with sheep or goats. |
| ■ Find shots of hands at a keyboard typing or using a mouse. | ■ Find shots in which a boat moves past. |
| □ Find shots of a canal, river, or stream with some of both banks visible. | ■ Find shots of a woman talking toward the camera in an interview - no other people visible. |
| ■ Find shots of a person talking on a telephone. | □ Find shots of a very large crowd of people (fills more than half of field of view). |
| □ Find shots of a street market scene. | □ Find shots of a classroom scene with one or more students. |
| □ Find shots of a street protest or parade. | □ Find shots of a bridge. |
| ■ Find shots of a train in motion. | ■ Find shots of a road taken from a moving vehicle through the front windshield. |
| □ Find shots with hills or mountains visible. | ■ Find shots of one or more people playing musical instruments such as drums, guitar, flute, keyboard, piano, etc. |
| □ Find shots of waterfront with water and buildings. | □ Find shots that contain the Cook character in the Klokhuis series. |
| □ Find shots of a street at night. | □ Find grayscale shots of a street with one or more buildings and one or more people. |

Distribution of hits for each topic

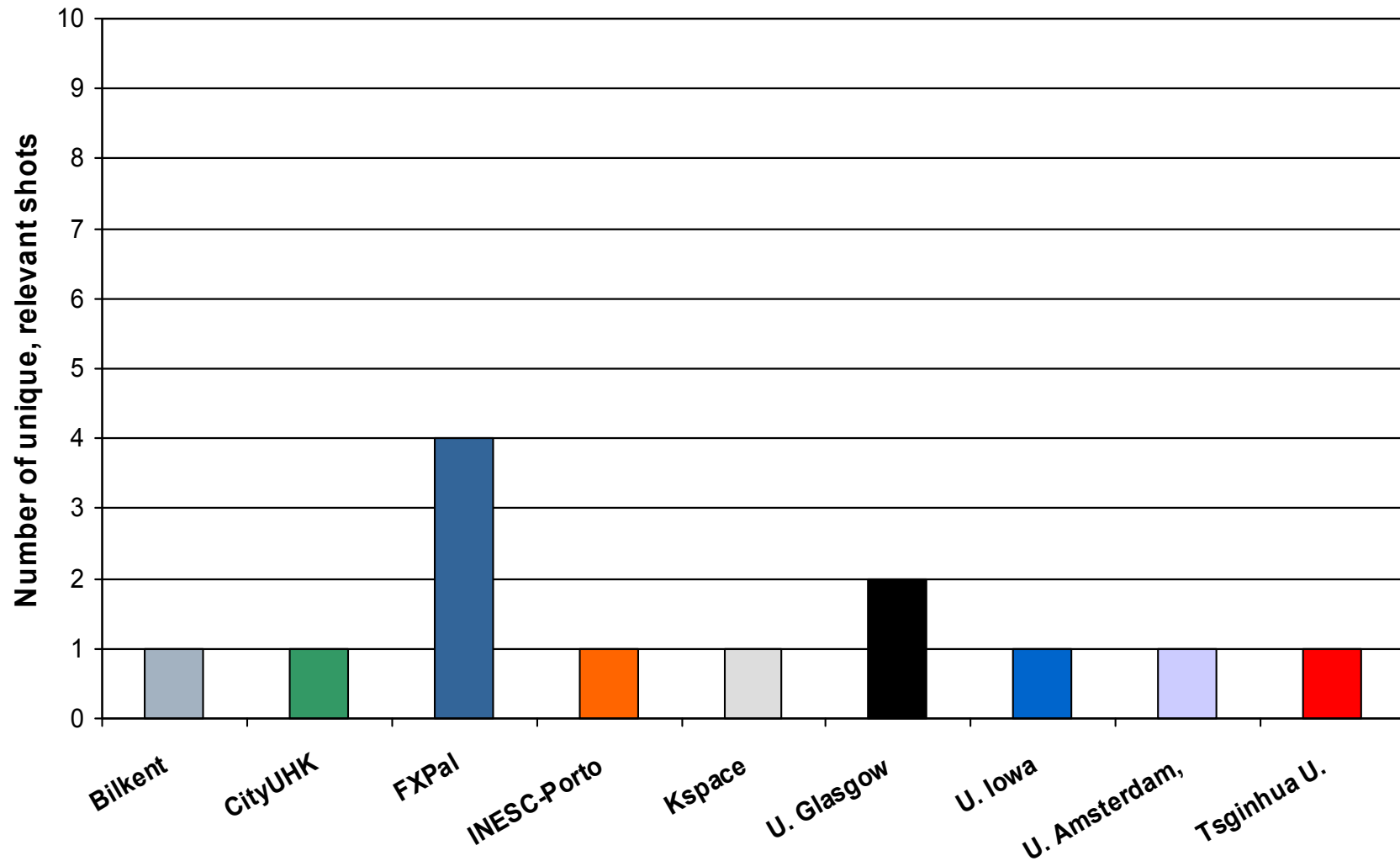


199 was intended as “person walking a bicycle or riding a bicycle” but was formulated as “Find shots of a person walking or riding a bicycle.” Doh!

Frequency of target topic-shots

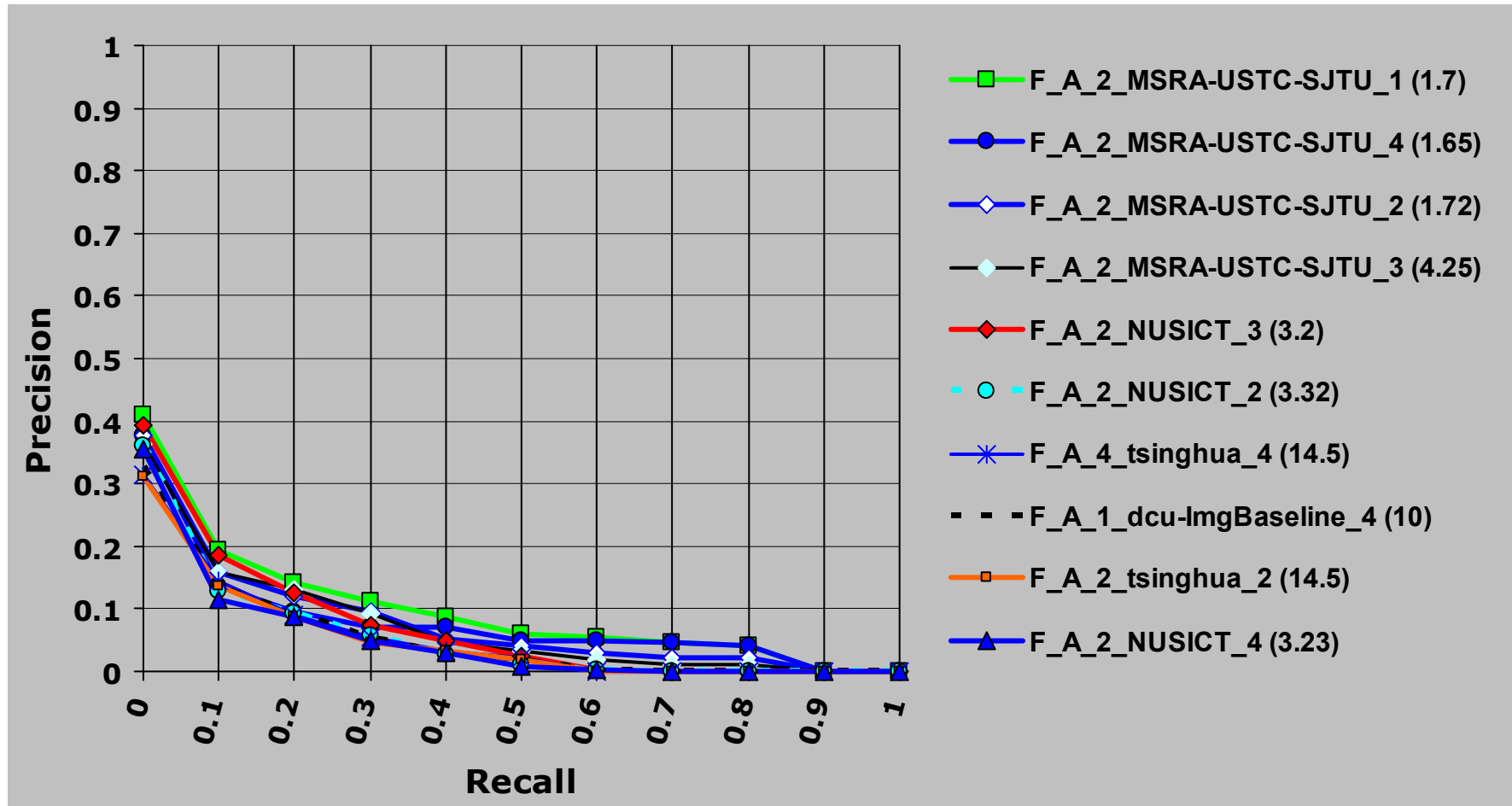
□	2007			
	■	Test shots * topics :	435,408	
	■	Relevant topic-shots:	4,704	1.1%
□	2006:			
	■	Test shots * topics:	1,907,616	
	■	Relevant topic-shots:	7,225	0.4%
□	2005			
	■	Test shots * topics:	1,098,360	
	■	Relevant topic-shots:	8,395	0.8%
□	2004			
	■	Test shots * topics:	800,808	
	■	Relevant topic-shots:	1,800	0.2%
□	2003			
	■	Test shots * topics:	775,632	
	■	Relevant topic-shots:	2,114	0.3%

Very few unique, relevant shots by group



Automatic runs - top 10 MAP (of 81)

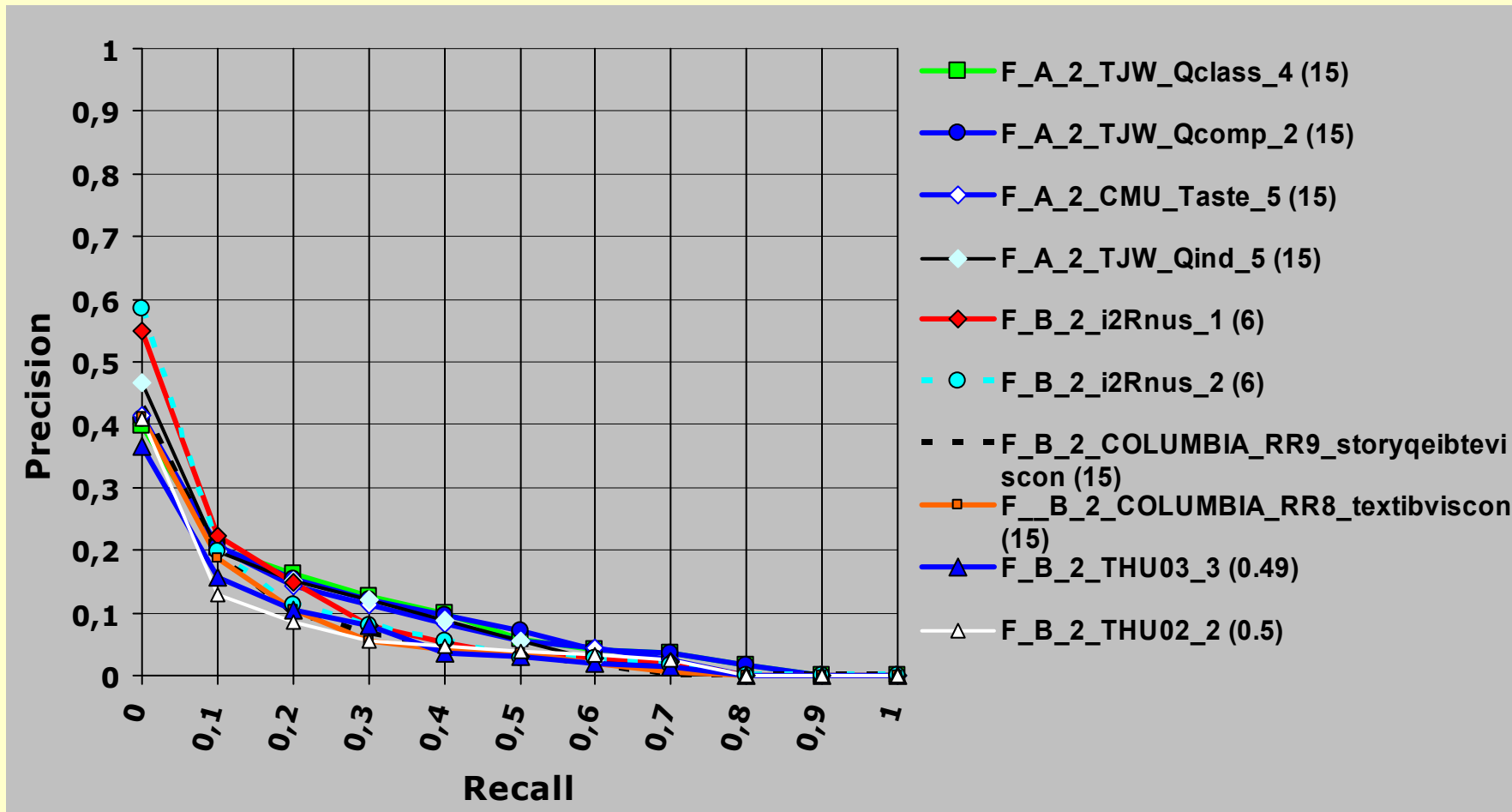
(mean elapsed time (mins) / topic)



Another view: in highest scoring run, on average a little more than 2 of the top 10 shots returned contained the desired video

2006: Automatic runs - top 10 MAP (of 76)

(mean elapsed time (mins) / topic)



Significant differences among top 8 automatic runs (using randomization test, 10^5 iterations, $p < 0.05$)

Run name	(MAP)	
* A_2_MSRA-..._1	(0.088)	A_2_MSRA-..._1
		➤ A_2_MSRA-..._2
		➤ A_2_MSRA-..._3
> A_2_MSRA-..._4	(0.066)	➤ A_2_MSRA-..._4
> A_2_MSRA-..._2	(0.065)	
> A_2_MSRA-..._3	(0.062)	A_2_NUSICT_3
= A_2_NUSICT_3	(0.061)	➤ A_2_NUSICT_2
		➤ A_2_NUSICT_4
= A_2_NUSICT_2	(0.044)	➤ A_2_tsinghua_2
		➤ A_2_tsinghua_4
= A_2_tsinghua_4	(0.044)	
= A_2_tsinghua_2	(0.043)	...

Significant differences among top 8 automatic runs (using randomization test, 10^5 iterations, $p < 0.05$)

without Topic 219

Run name	(MAP)	A_2_NUSICT_3
* A_2_NUSICT_3	(0.059)	➤ A_2_tsinghua_2 ➤ A_2_NUSICT_2
= A_2_MSRA-..._1	(0.054)	➤ A_2_NUSICT_4
> A_2_tsinghua_2	(0.043)	➤ A_2_MSRA..._2 ➤ A_2_MSRA..._5
> A_2_NUSICT_2	(0.043)	A_2_MSRA..._1
> A_2_NUSICT_4	(0.041)	➤ A_2_MSRA..._2 ➤ A_2_MSRA..._3
= A_2_MSRA-..._3	(0.041)	
> A_2_MSRA-..._2	(0.040)	...
> A_2_MSRA-..._5	(0.039)	

Full randomization test results - top 10 automatic search runs

Probability that
the difference in
runs (A>B) is
due to chance

p<0.01

p<0.05

p<0.10

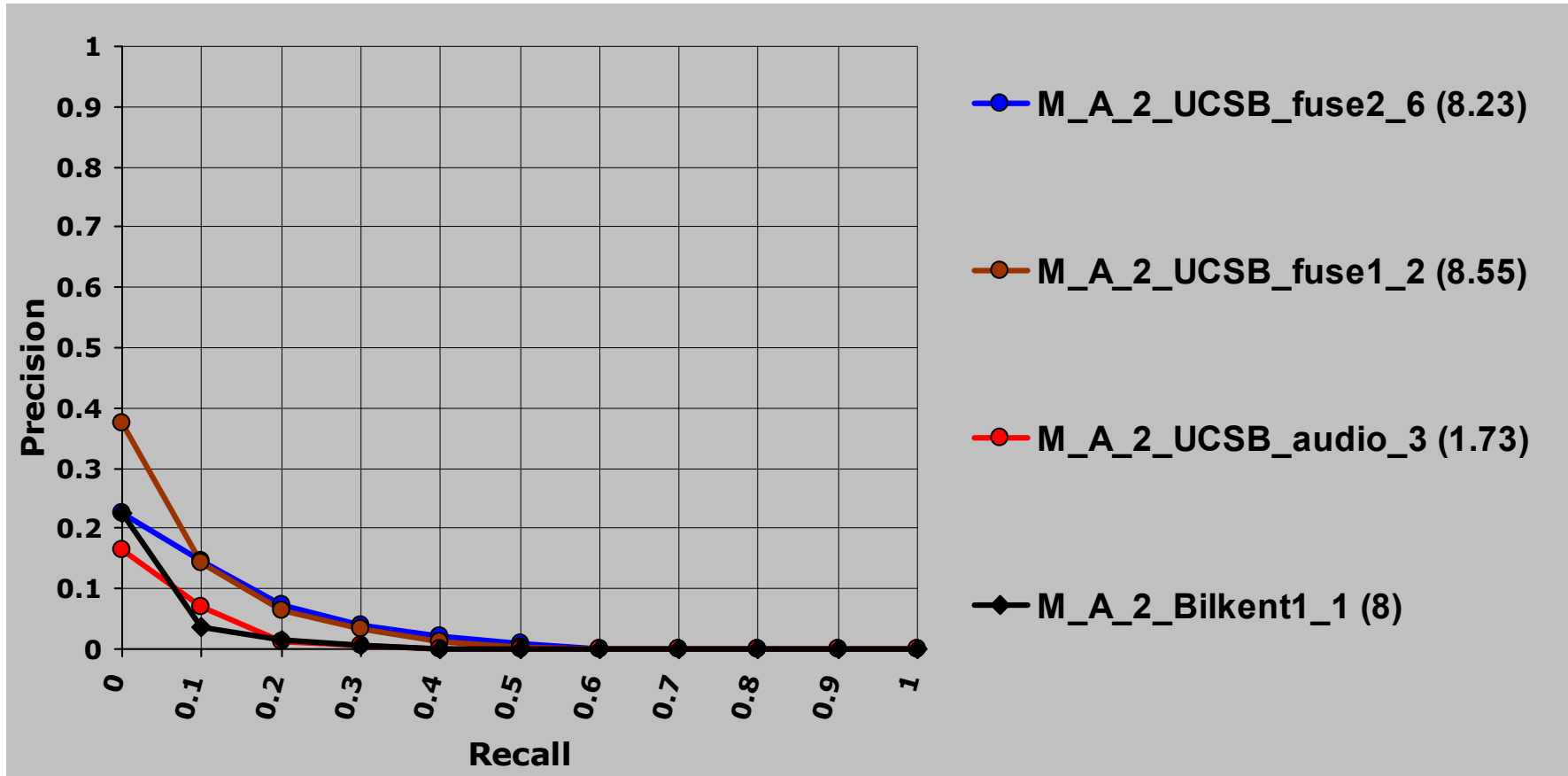
Run B

Run A

		F_A_1_dcu-ImgBaseline	F_A_2_NUSICT_2	F_A_2_NUSICT_3	F_A_2_NUSICT_4	F_A_2_tsinghua_2	F_A_2_tsinghua_4	F_A_2_MSRA-USTC-SJTU-SEARCH_1	F_A_2_MSRA-USTC-SJTU-SEARCH_2	F_A_2_MSRA-USTC-SJTU-SEARCH_3	F_A_2_MSRA-USTC-SJTU-SEARCH_4
F_A_1_dcu-ImgBaseline_4	*	0.502	*	0.492	0.495	*	*	*	*	*	*
F_A_2_NUSICT_2	0.499	*	*	0.006	0.433	*	*	*	*	*	*
F_A_2_NUSICT_3	0.171	0.001	*	0.000	0.007	0.024	*	*	*	*	*
F_A_2_NUSICT_4	*	*	*	*	*	*	*	*	*	*	*
F_A_2_tsinghua_2	*	*	*	0.384	*	*	*	*	*	*	*
F_A_2_tsinghua_4	0.491	0.421	*	0.252	0.094	*	*	*	*	*	*
F_A_2_MSRA-USTC-SJTU-SEARCH_1	0.005	0.103	0.319	0.076	0.099	0.123	*	0.016	0.016	0.013	*
F_A_2_MSRA-USTC-SJTU-SEARCH_2	0.032	0.312	0.478	0.245	0.286	0.328	*	*	0.345	*	*
F_A_2_MSRA-USTC-SJTU-SEARCH_3	0.023	0.294	0.488	0.235	0.276	0.323	*	*	*	*	*
F_A_2_MSRA-USTC-SJTU-SEARCH_4	0.202	0.387	0.471	0.356	0.372	0.400	*	0.447	0.423	*	*

Manual runs – All 4

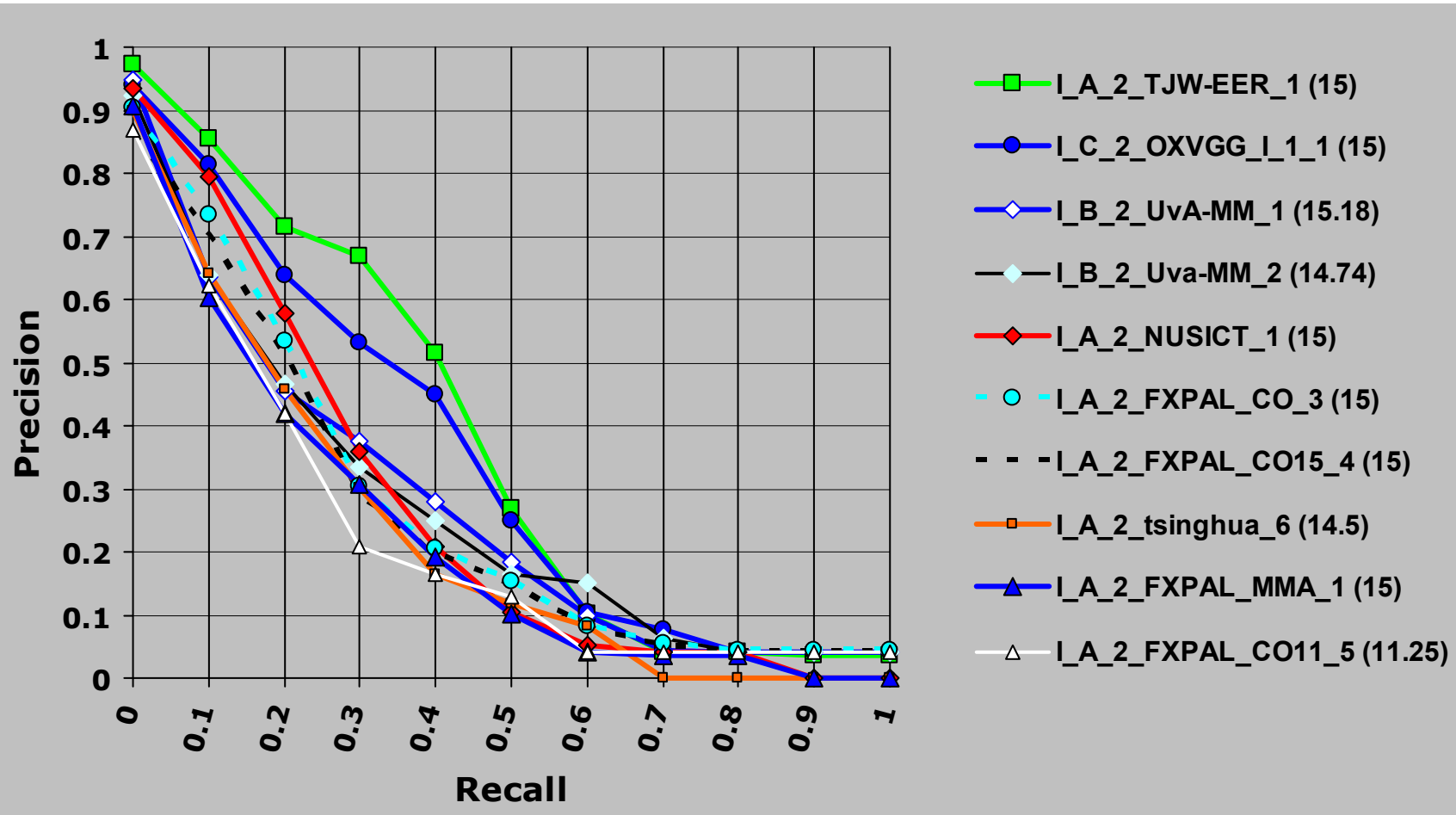
(mean human effort (mins) / topic)



Another view: in highest scoring run, on average not quite 1 of the top 10 shots returned contained the desired video

Interactive runs - top 10 MAP (of 33)

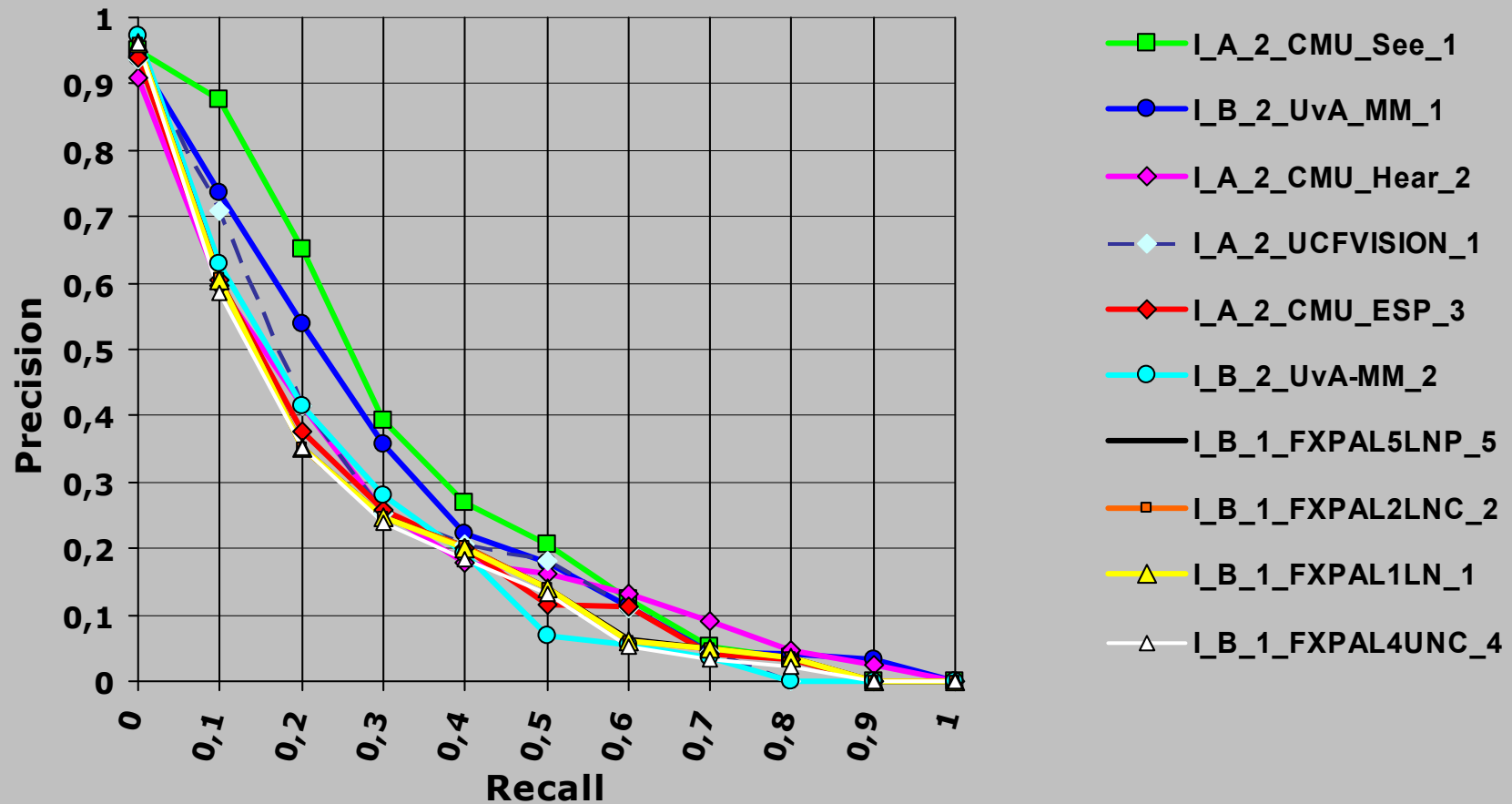
(mean elapsed time (mins) / topic)



Another view: in highest scoring run, on average 8 of the top 10 shots returned contained the desired video

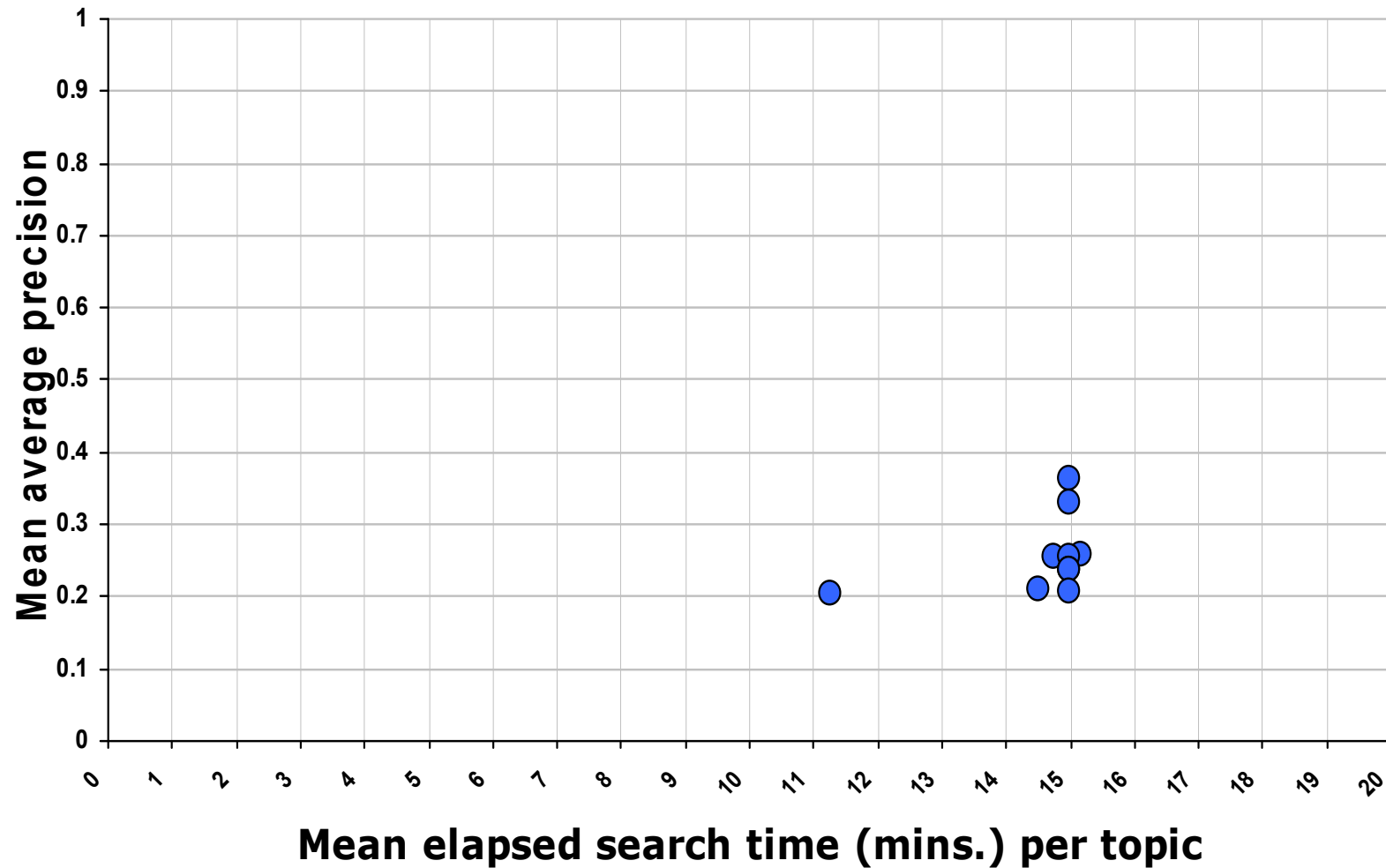
2006: Interactive runs - top 10 MAP (of 36)

(mean elapsed time for all == ~15 mins/topic)



Top 10 Interactive Runs

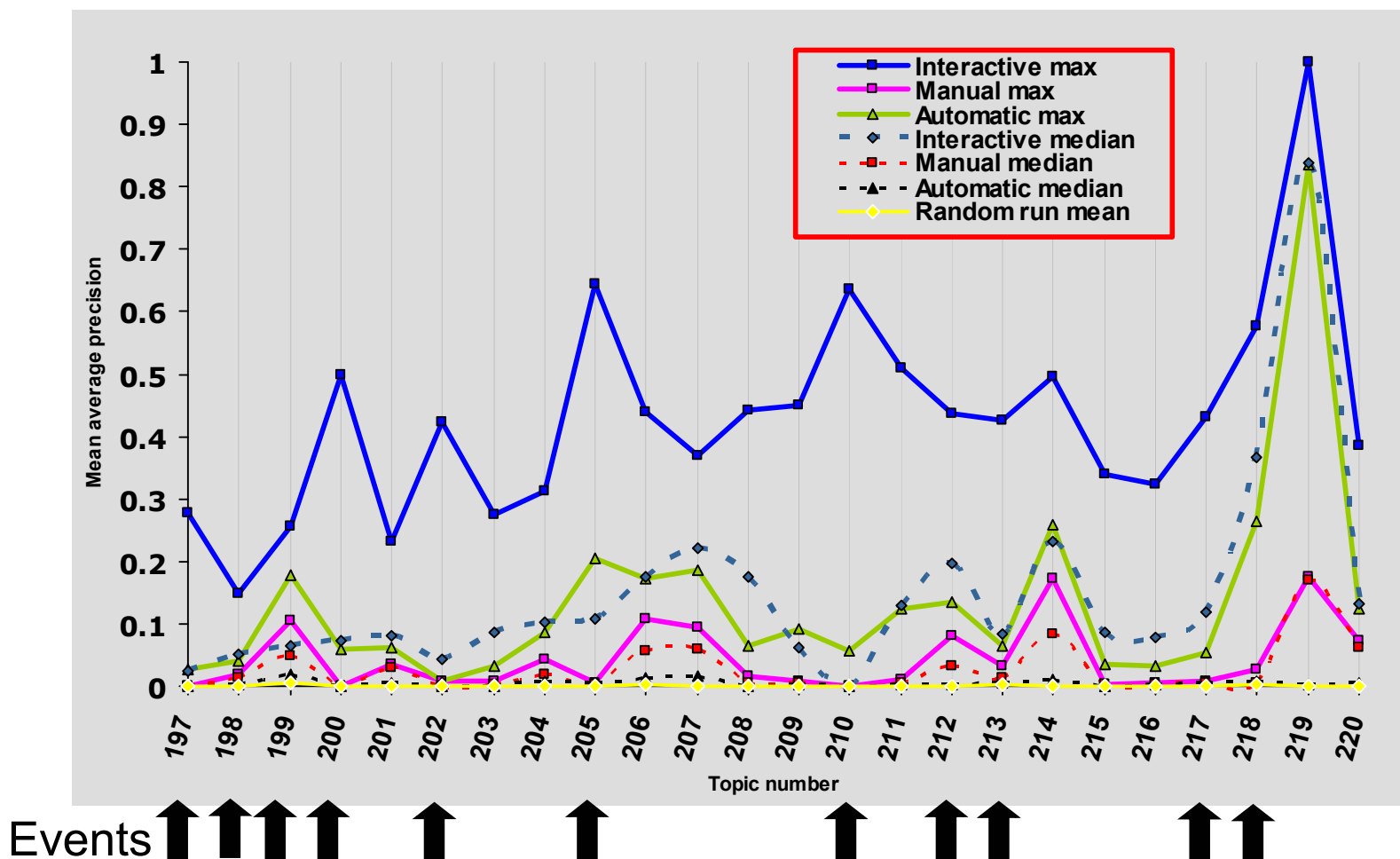
MAP vs mean elapsed search time



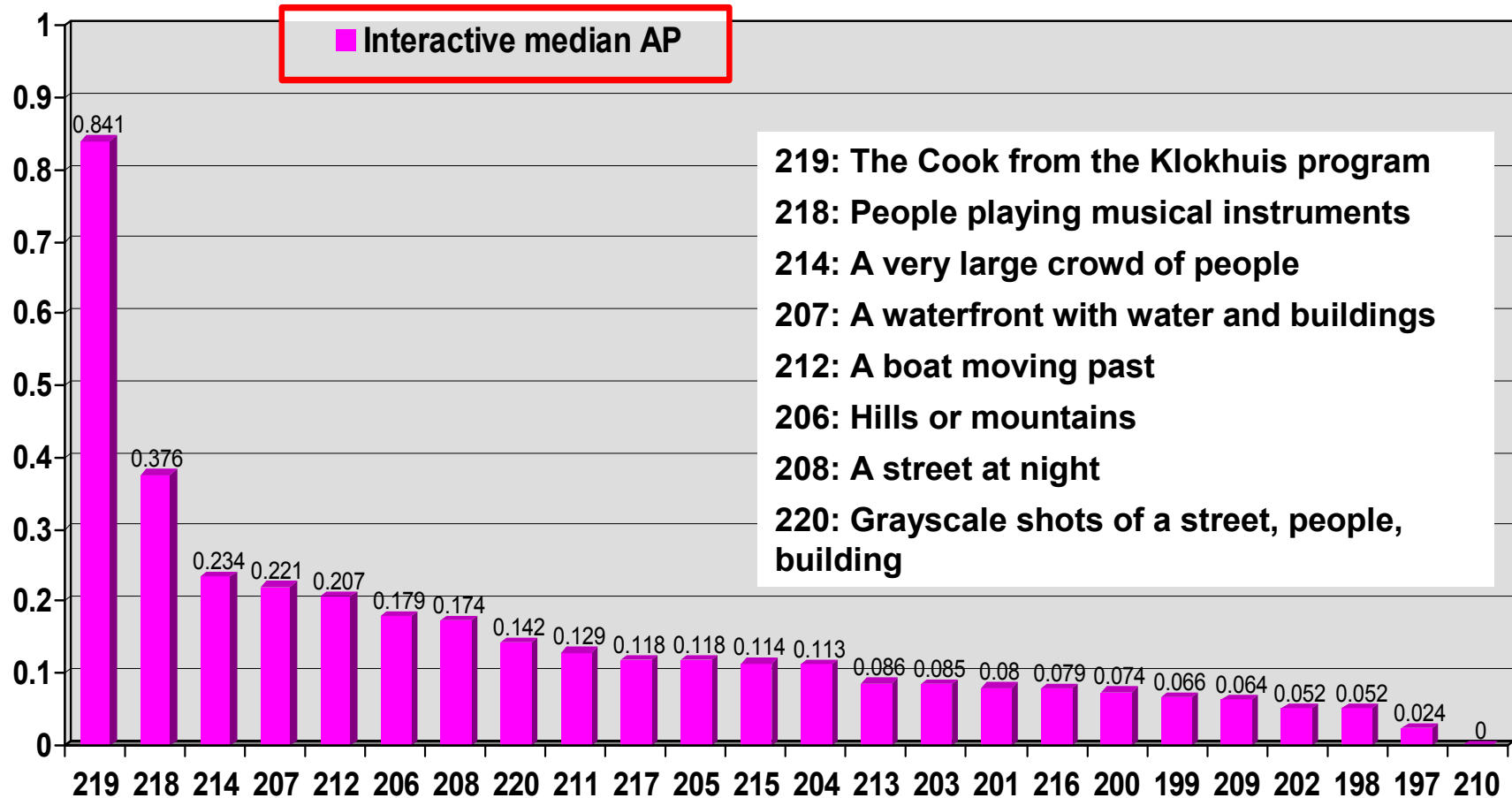
Significant differences among top 8 interactive runs (using randomization test, $p < 0.05$)

Run name	(MAP)	
• A_2_TJW_EER_1	(0.362)	A_2_TJW_EER_1
		➤ A_2_FXPAL_CO15_4
		➤ A_2_NUSICT_1
		➤ B_2_UvA-MM_1
= C_2_OXVGG_I_1	(0.329)	➤ B_2_UvA-MM_2
> B_2_UvA-MM_1	(0.259)	C_2_OXVGG_I_1
		➤ A_2_FXPAL_CO11_5
		➤ A_2_FXPAL_CO15_4
> B_2_UvA-MM_2	(0.256)	➤ A_2_FXPAL_CO_3
		➤ A_2_2_FXPAL_MMA_1
> A_2_NUSICT_1	(0.255)	➤ A_2_NUSICT_1
		➤ B_2_UvA-MM_1
		➤ B_2_UvA-MM_2
> A_2_FXPAL_CO15_4	(0.238)	➤ A_2_tsinghua_6

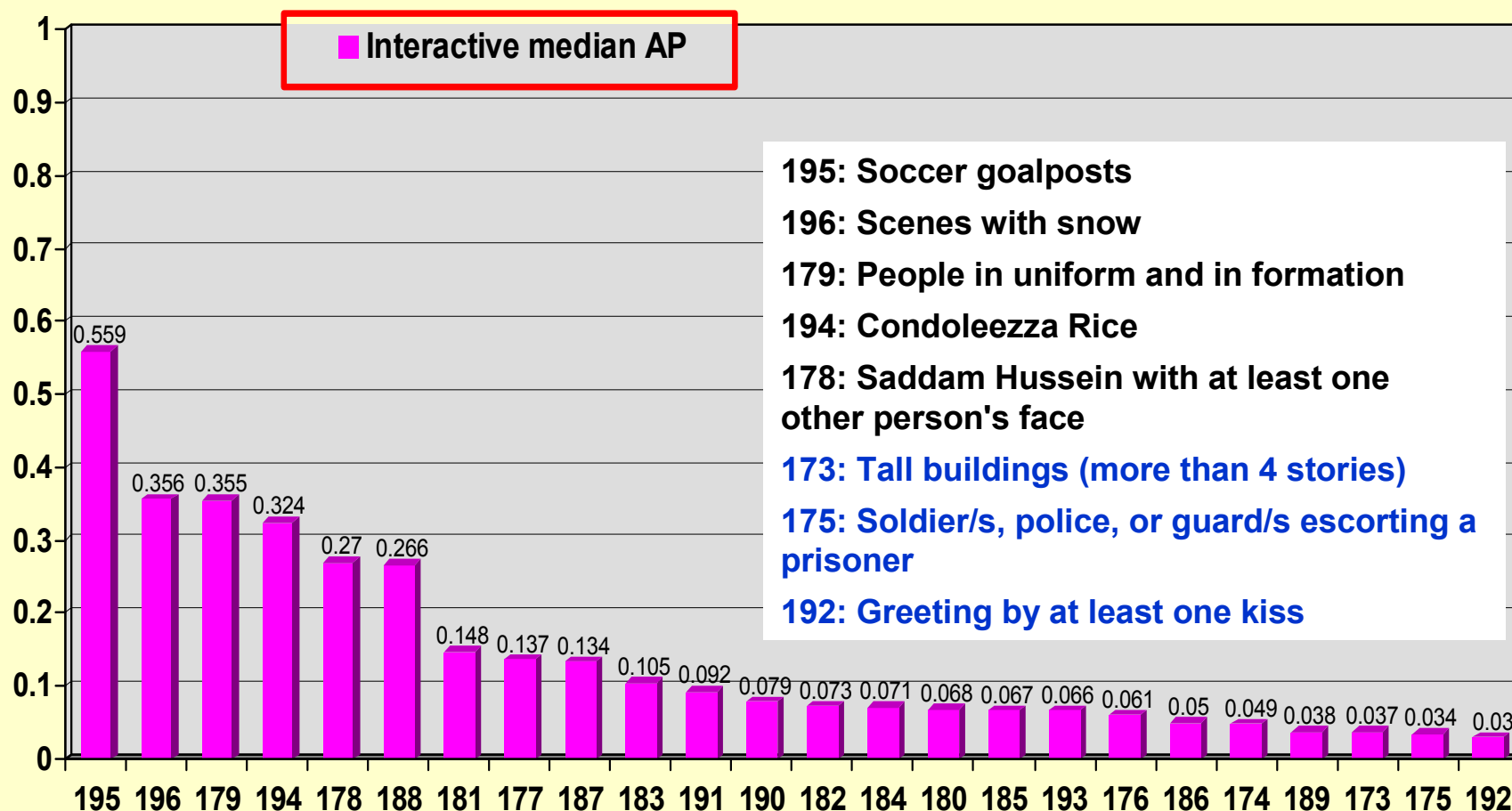
Average precision by topic



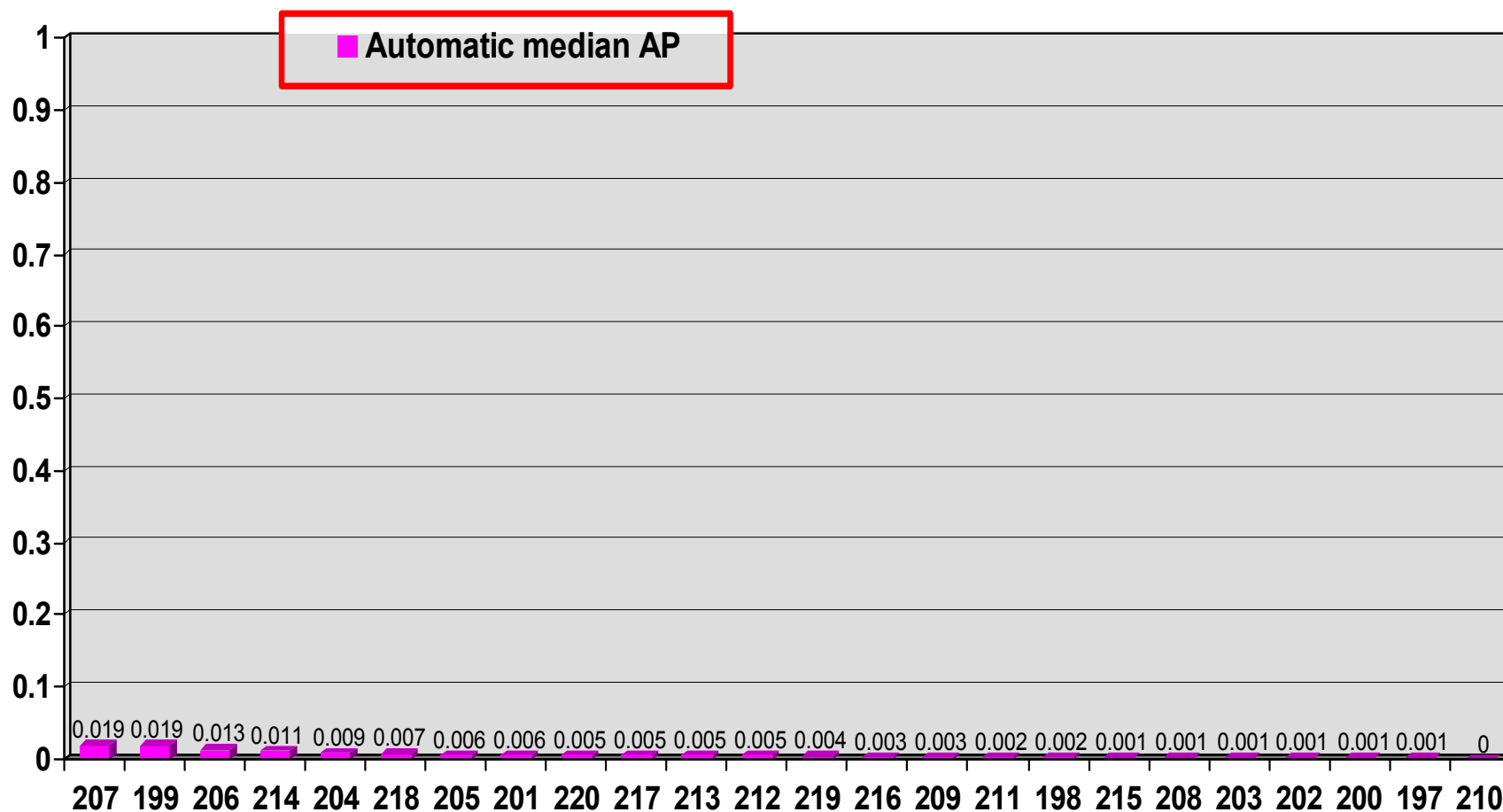
Interactive runs' median average precision by topic



2006: Interactive runs' median average precision by topic



Automatic runs' median average precision by topic



Who did what ?

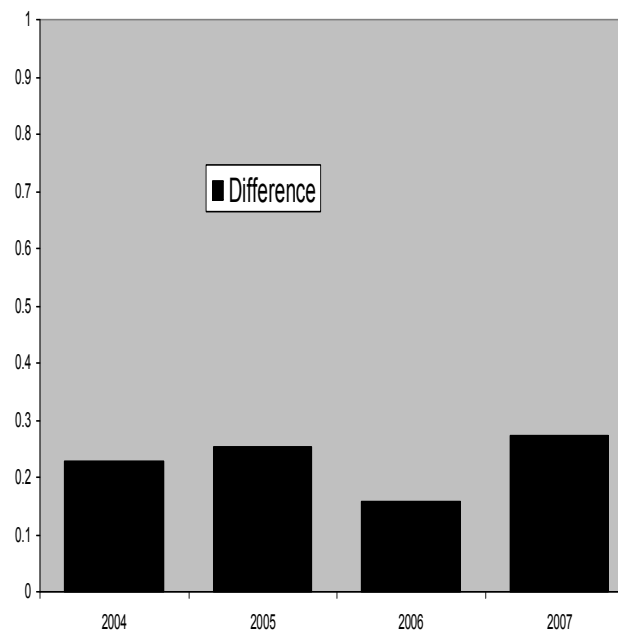
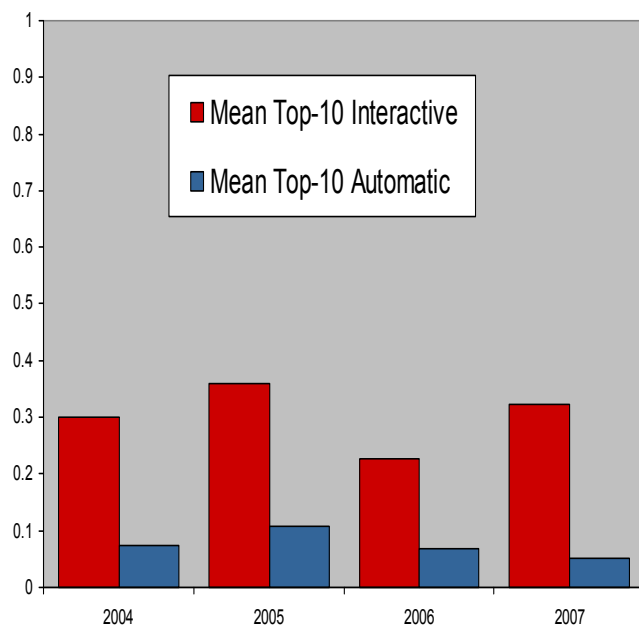
- ☐ Speaker slots to follow:
 - IBM
 - ☐ New interactive search system (tagging/browsing hybrid)
 - ☐ Expanded concept lexicon
 - FX Palo Alto Laboratories
 - ☐ Realtime search cooperation among 2 or more searchers
 - Tsinghua U. / Intel China
 - ☐ More work on concept-based search and example-to-concept mapping
 - University of Amsterdam
 - ☐ Crossbrowser meets Forkbrowser
 - National University of Singapore
 - ☐ Focus on low-level visual and motion features
 - ☐ User choice among 3 feedback strategies

Previous Observations

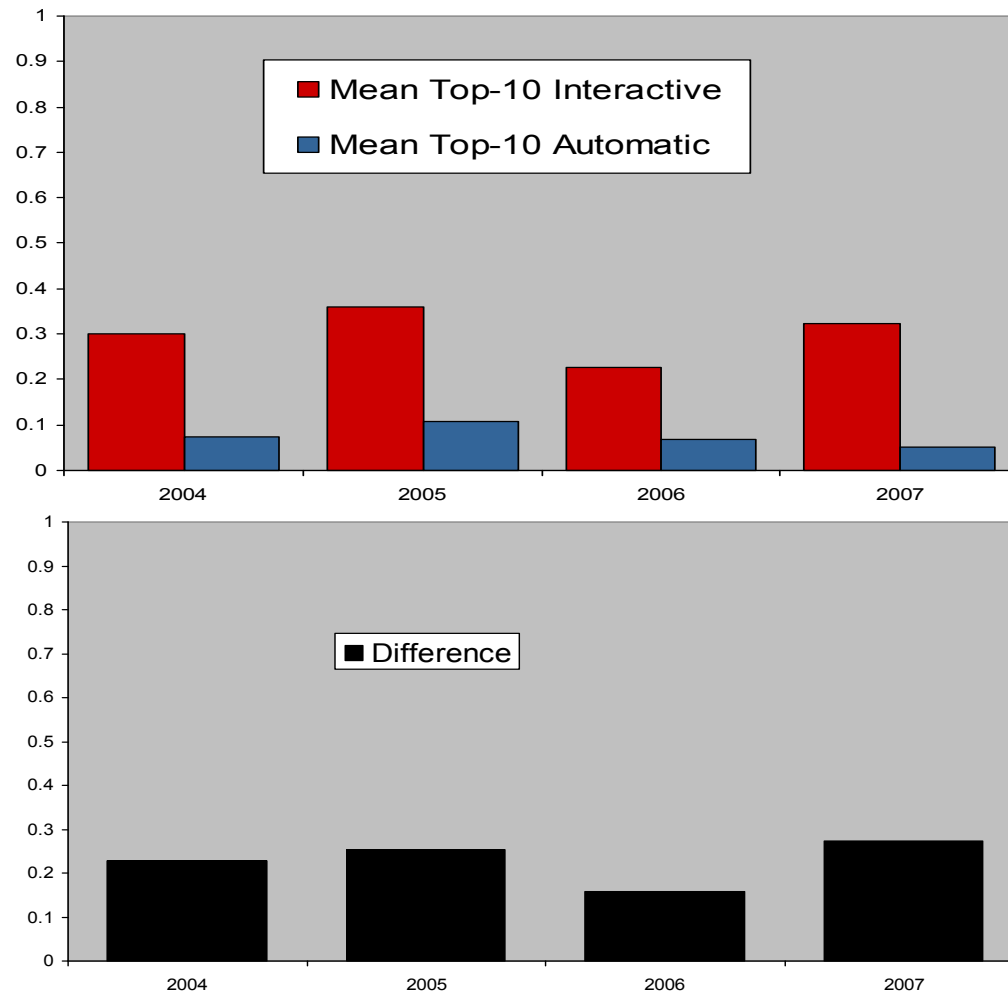
- Observations 2005 (new BN data, multi-lingual)
 - We're still getting "□ Lots of variation, interesting shot browsing interfaces, mixture of interactive & manual", and additionally automatic runs;
 - Top performances on all 3 search types are up, even with more difficult data,
 - Some leveraged the structured nature of B/News;
 - Many did automatic search & fewer did interactive
 - Most common issue was combination of text/image search/concepts
- Observations 2006 (bigger collection)
 - Top performances on all 3 search types are down (collection x2, half as many relevant shots, harder topics)
 - Increase in automatic search & fewer interactive search
 - Manual runs no longer outperform automatic – so few manual, and does it make sense to keep this processing type ?

Observations/Questions

- Still can't easily compare performance across very different data/topics (unless same system run on both to estimate data effect).
- However, difference between mean of top-10 interactive and the mean of the top-10 automatic runs has increased over 2006 but not over 2005



Top-10 interactive vs top-10 automatic runs (MAP)



Questions...

- ☐ Did systems adapt to new data/topic characteristics?
 - What old approaches stopped/continued working?
 - What new approaches were tried with(out) success?
- ☐ Did systems do anything special to support search for events?
- ☐ How did systems handle search for grayscale video?
- ☐ What is collaborative search all about?
- ☐ What experimental designs are being used to isolate the system effect from the search effect in interactive searches?

-
- ☐ Thanks to all who participated, contributed and organised
 - ☐ For citing TRECVID - see the website for preferred citation