
TRECVID-2006: Search Task

Alan Smeaton
Dublin City University
&
Tzveta Ianeva
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;
- NIST created more topics asking for **general** (vs. **specific**)
- NIST created 10 of 24 topics to ask for video of events – encouraging exploration beyond *one-keyframe-per-shot*
- Videos were viewed by NIST personnel, notes taken on content, and candidates emerging were chosen;

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.

Search Task Definition

- Bing Xiang, John Makhoul, and Ralph Weischedel at BBN for providing MT/ASR
- Christian Petersohn (Fraunhofer Institute) for master shot reference
- DCU team for formatting and selecting keyframes
- MediaMill team for 101 features baseline results donation
- CMU and IBM for 449 LSCOM features annotations

Data characteristics

- TRECVID 2006 data is again (deliberately) text-noisy with video from English language, Arabic & Chinese broadcasts;
- **32.2%** of the test video comes from programs not represented in the development data
- Text is derived from speech recognition and then machine translation, thus poorer quality than with English-only sources but ASR/MT from “state-of-the-art” GALE system.

2006: Search task participants (26, up from 20)

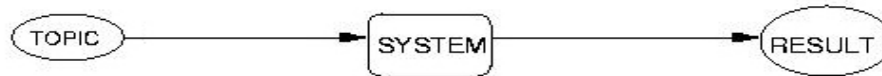
AT&T Labs - Research	USA
Beijing Jiaotong U.	China
Bilkent U.	Turkey
Carnegie Mellon U.	USA
Chinese U. of Hong Kong	China
City University of Hong Kong	China
CLIPS-IMAG	France
Columbia U.	USA
Dublin City U.	Ireland
Fudan U.	China
FX Palo Alto Laboratory Inc	USA
Helsinki U. of Technology	Finland
IBM T. J. Watson Research Center	USA
Imperial College London / Johns Hopkins U.	UK, USA

2006: Search task participants (continued)

NUS / I2R	Singapore
Mediamill / U. of Amsterdam	Netherlands
RMIT U. School of CS&IT	Australia
Tsinghua U.	China
U. of Central Florida	USA
U. of Glasgow / U. of Sheffield	UK
U. of Iowa	USA
U. of Oxford	UK
U. Rey Juan Carlos	Spain
Zhejiang U.	China
COST292 (www.cost292.org)	France, Netherlands, UK, Ireland, Greece, Turkey, Serbia and Montenegro, Slovakia
K-Space (kspace.qmul.net)	UK, Germany, Austria, Greece, Ireland, Netherlands, France, Switzerland, Czechia

Search Types: Automatic, Manual and Interactive

AUTOMATIC :



System takes topic as input and produces result without any human intervention

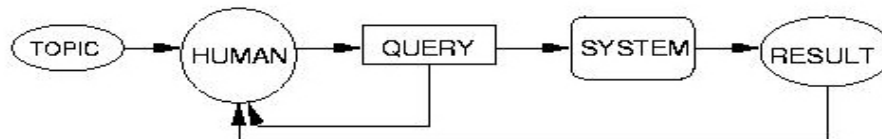
MANUAL :



Human formulates query based on topic and query interface, not on knowledge of collection or search results

System takes query as input and produces result without further human intervention

INTERACTIVE :

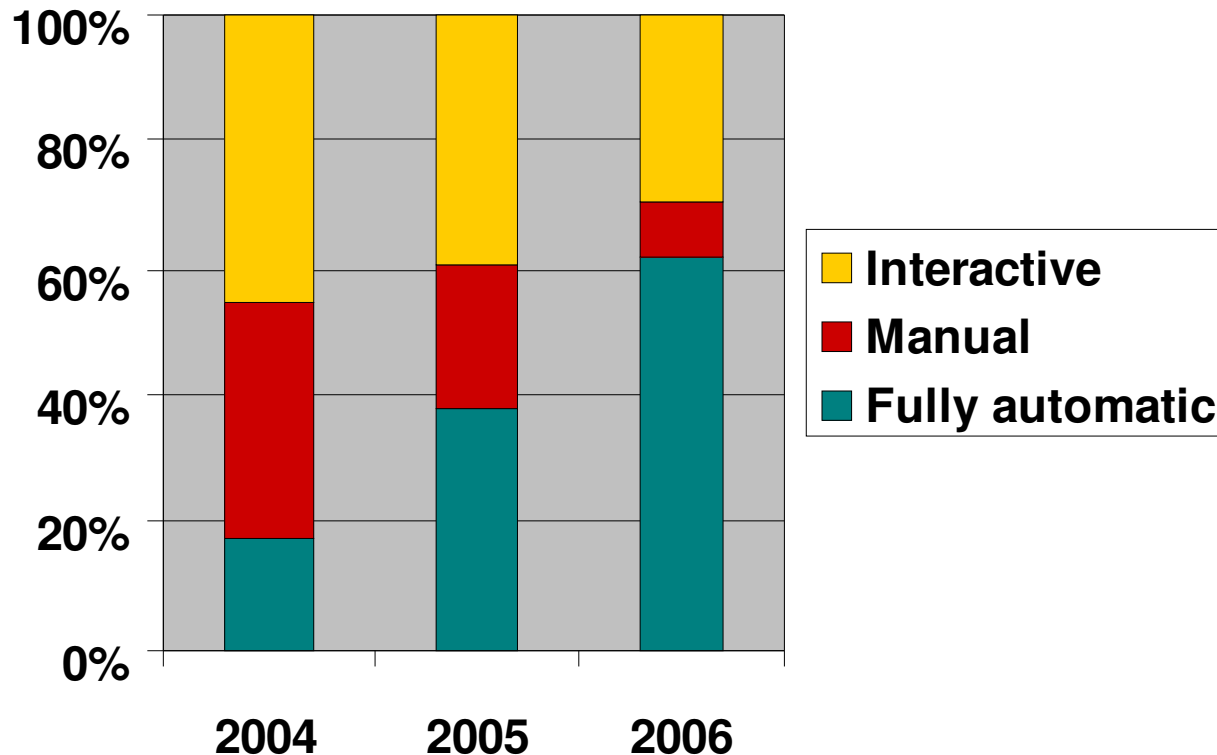


Human (re)formulates query based on topic, query, and/or results

System takes query as input and produces result without further human intervention on this invocation

Number of runs: **76 automatic**
11 manually assisted
36 interactive

Everybody likes to search automatically, dislikes manually



24 Topics [number of image, video examples and relevant found]

- 173. Find shots with a view of one or more tall buildings (more than 4 stories) and the top story visible [3, 4, 142]
- 174. Find shots with one or more people leaving or entering a vehicle [0, 10, 675]
- 175. Find shots with one or more soldiers, police, or guards escorting a prisoner [0, 4, 204]
- 176. Find shots of a daytime demonstration or protest with at least part of one building visible [4, 4, 111]
- 177. Find shots of US Vice President Dick Cheney [3, 3, 393]
- 178. Find shots of Saddam Hussein with at least one other person's face at least partially visible [8, 0, 99]
- 179. Find shots of multiple people in uniform and in formation [3, 5, 191]
- 180. Find shots of US President George W. Bush, Jr. walking [0, 5, 197]

24 Topics [number of image, video examples and relevant found]

- 181. Find shots of one or more soldiers or police with one or more weapons and military vehicles [2, 6, 128]
- 182. Find shots of water with one or more boats or ships [3, 5, 307]
- 183. Find shots with one or more emergency vehicles in motion (e.g., ambulance, police car, fire truck, etc.) [0, 4, 299]
- 184. Find shots of one or more people seated at a computer with display visible [3, 4, 440]
- 185. Find shots of one or more people reading a newspaper [3, 4, 201]
- 186. Find shots of a natural scene - with, for example, fields, trees, sky, lake, mountain, rocks, rivers, beach, ocean, grass, sunset, waterfall, animals, or people; but no buildings, no roads, no vehicles [2, 4, 523]
- 187. Find shots of one or more helicopters in flight [0, 6, 119]

24 Topics [number of image, video examples and relevant found]

- 188. Find shots of something burning with flames visible [3, 5, 375]
- 189. Find shots of a group including at least four people dressed in suits, seated, and with at least one flag [3, 5, 446]
- 190. Find shots of at least one person and at least 10 books [3, 5, 295]
- 191. Find shots containing at least one adult person and at least one child [3, 6, 775]
- 192. Find shots of a greeting by at least one kiss on the cheek [0, 5, 98]
- 193. Find shots of one or more smokestacks, chimneys, or cooling towers with smoke or vapor coming out [3, 2, 60]
- 194. Find shots of Condoleezza Rice [3, 7, 122]
- 195. Find shots of one or more soccer goalposts [3, 4, 333]
- 196. Find shots of scenes with snow [3, 6, 692]

Some statistics

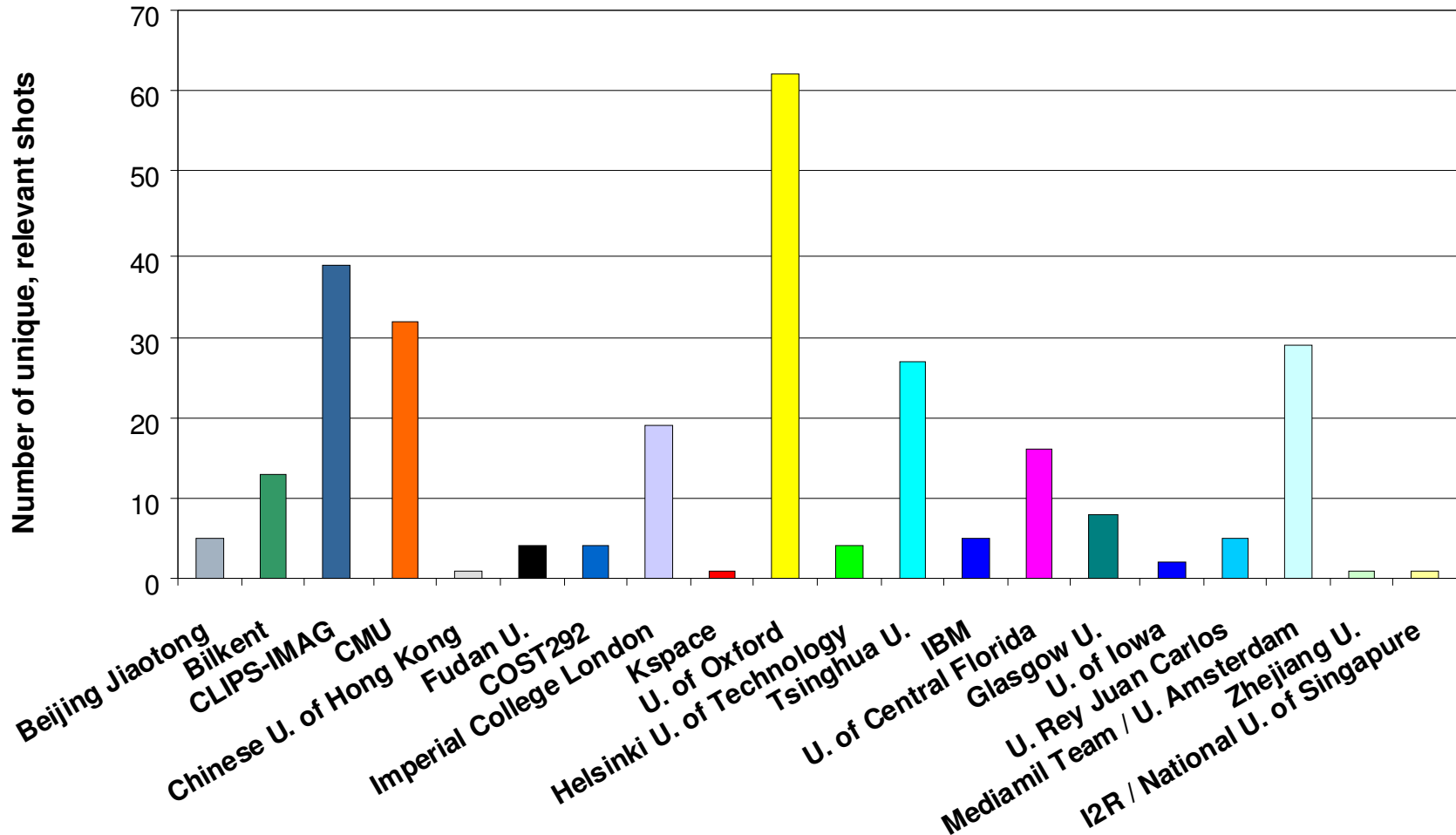
- **2006:**
 - Number of shots in test collection: 79.484
 - ~9.1% relevant shots found: 7.225

- **2005**
 - Number of shots in test collection: 45.765
 - ~18.3% relevant shots found: 8.395

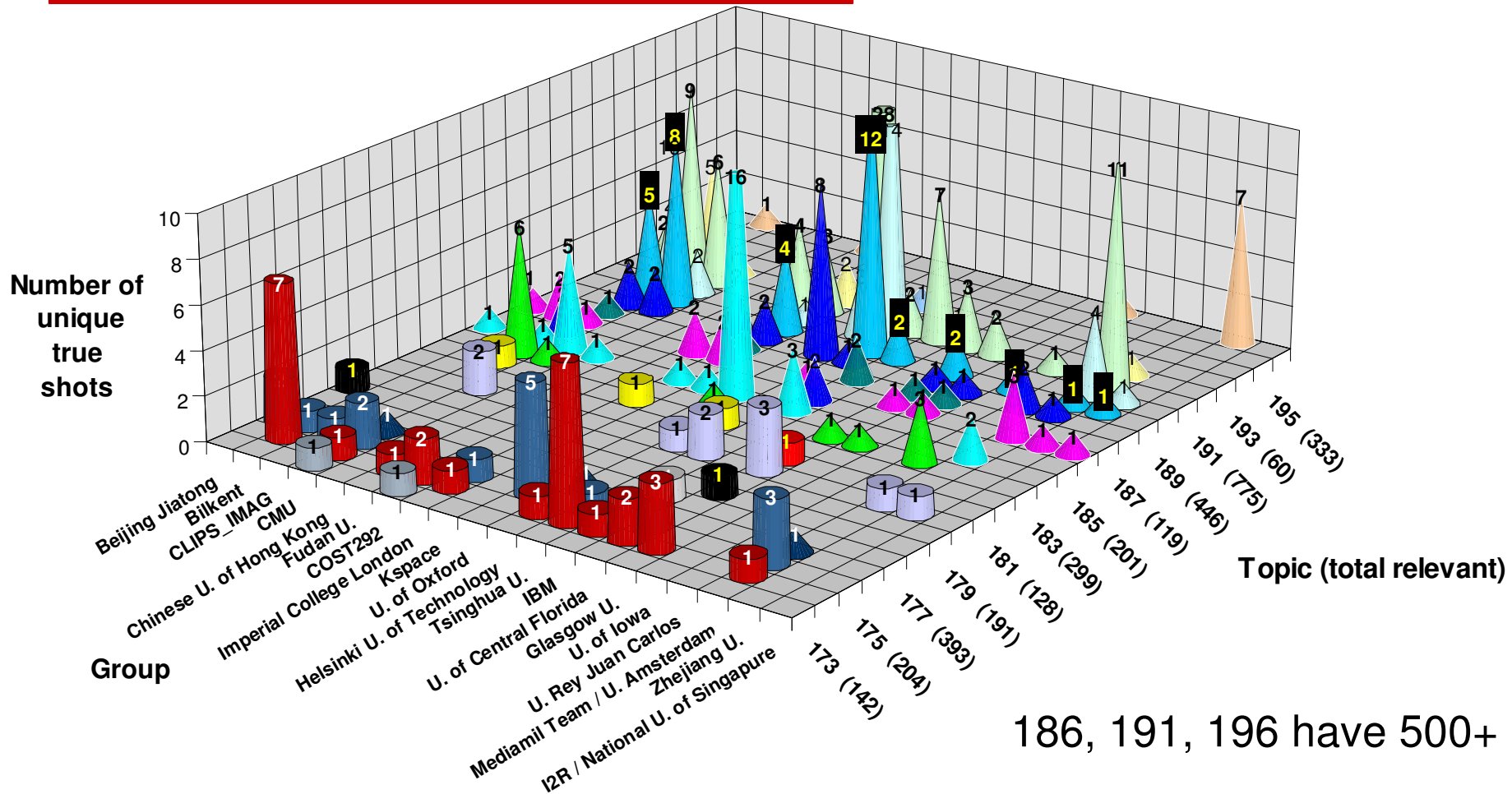
- **2004**
 - Number of shots in test collection: 33.367
 - ~5.4% relevant shots found: 1.800

- **2003**
 - Number of shots in test collection: 32.318
 - ~6.5% relevant shots found: 2.114

2006: 20 sites contributed one or more unique, relevant shots

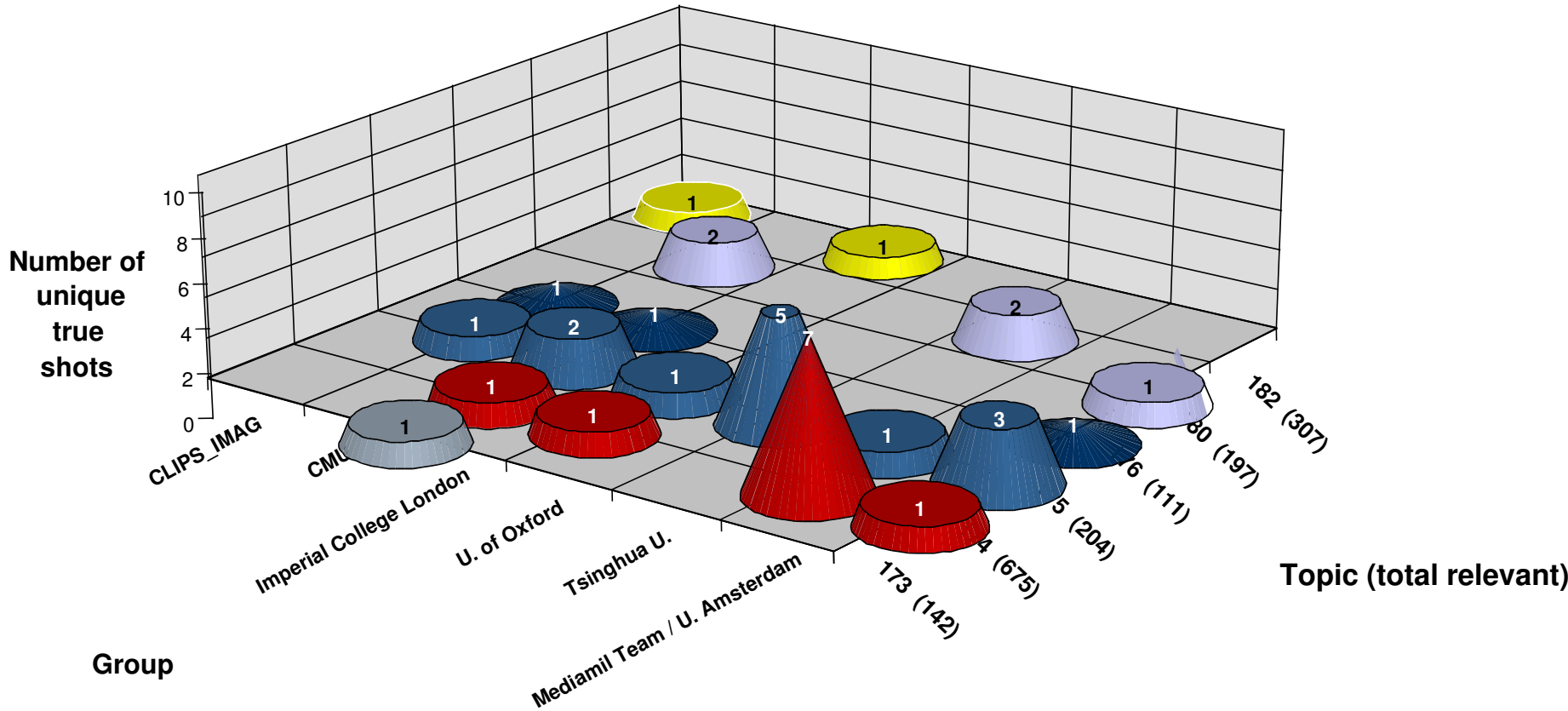


2006: Rel shots contrib. uniquely per topic by team



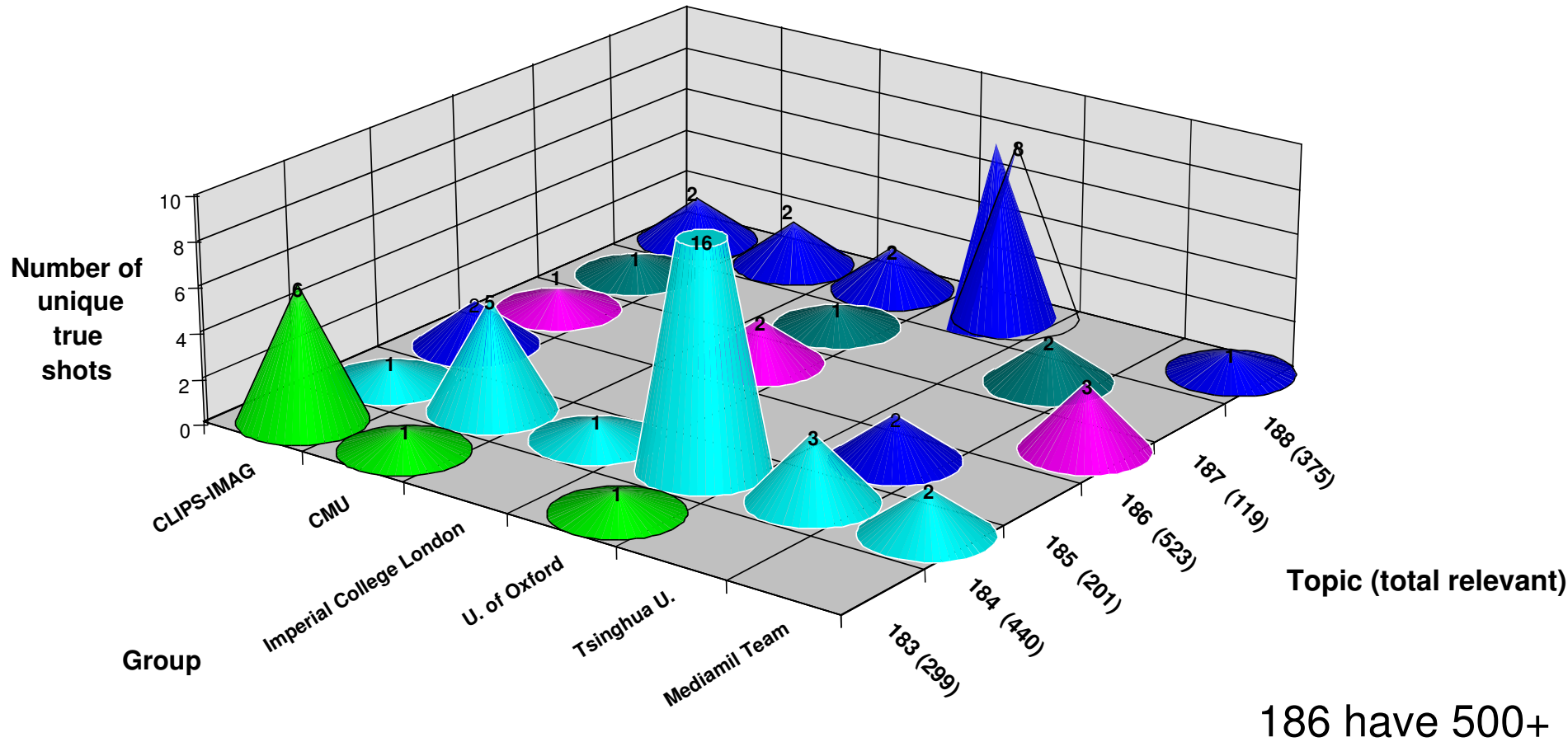
186, 191, 196 have 500+

2006: Most rel shots uniquely returned by topic & team



186, 191, 196 have 500+

2006: Most rel shots uniquely returned by topic & team

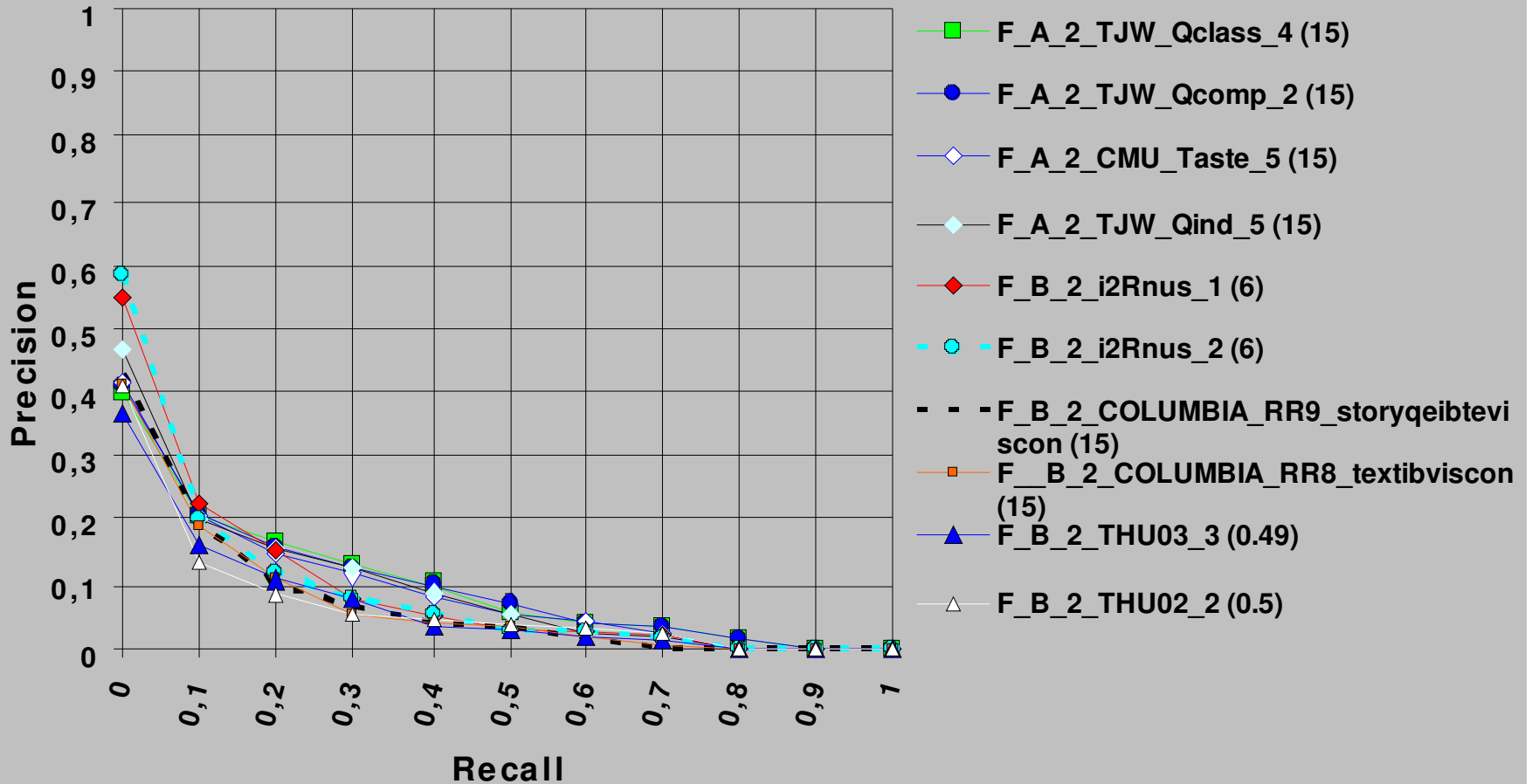


Unique relevant shots return by Oxford U. for Topic 191 (“adult and child”)



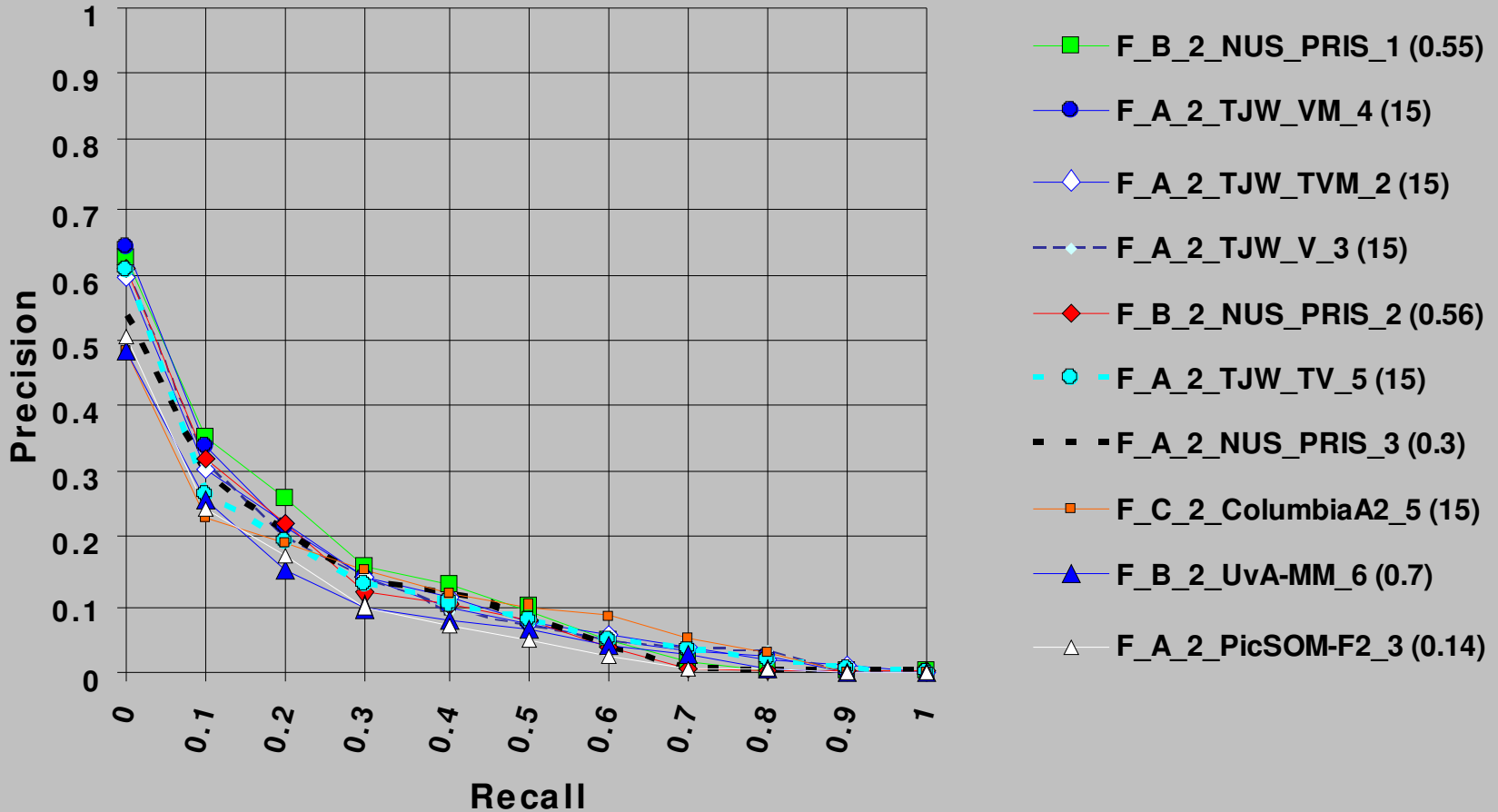
2006: Automatic runs - top 10 MAP (of 76)

(mean elapsed time (mins) / topic)



2005: Automatic runs - top 10 MAP (of 42)

(mean elapsed time (mins) / topic)

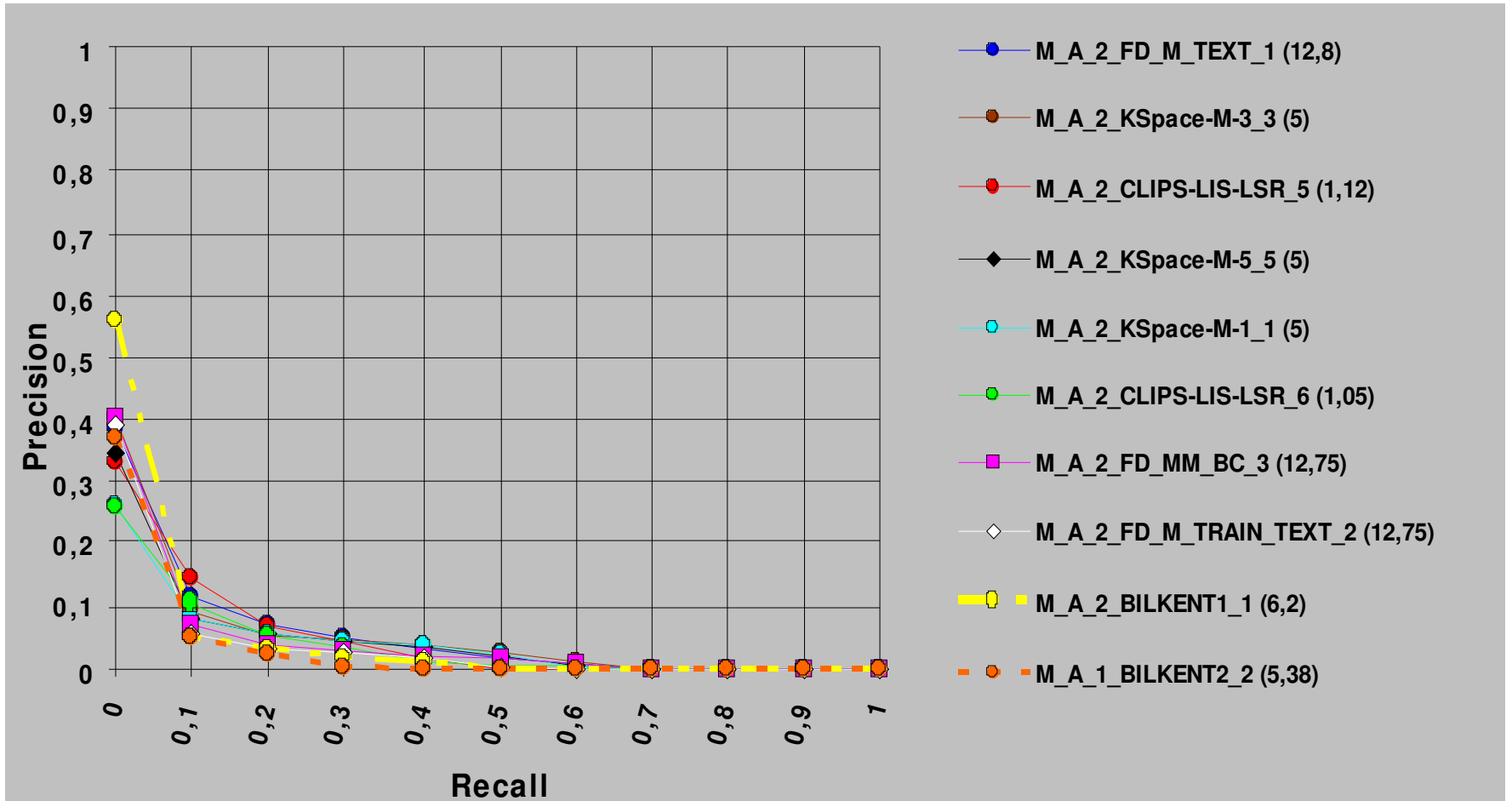


Significant differences among top 8 automatic runs (using randomization test, $p < 0.05$)

Run name	(MAP)	
* A_2_TJW_Qclass_4	(0.087)	A_2_TJW_Qclass_4
= A_2_TJW_Qcomp_2	(0.086)	↘ B_2_COLUMBIA_RR9_storyqeibteviscon_1
= A_2_CMU_Taste_5	(0.079)	↘ B_2_COLUMBIA_RR8_textibviscon
= A_2_TJW_Qind_5	(0.076)	↘ B_2_i2Rnus_2
= B_2_i2Rnus_1	(0.075)	A_2_TJW_Qcomp_2
> B_2_i2Rnus_2	(0.067)	↘ B_2_i2Rnus_2
> B_2_COLUMBIA_RR9...	(0.060)	↘ B_2_COLUMBIA_RR9_storyqeibteviscon_1
> B_2_COLUMBIA_RR8...	(0.056)	↘ B_2_COLUMBIA_RR8_textibviscon
		A_2_CMU_Taste_5
		↘ B_2_COLUMBIA_RR9_storyqeibteviscon_1
		↘ B_2_COLUMBIA_RR8_textibviscon
		B_2_i2Rnus_1
		↘ B_2_COLUMBIA_RR9_storyqeibteviscon_1
		↘ B_2_COLUMBIA_RR8_textibviscon

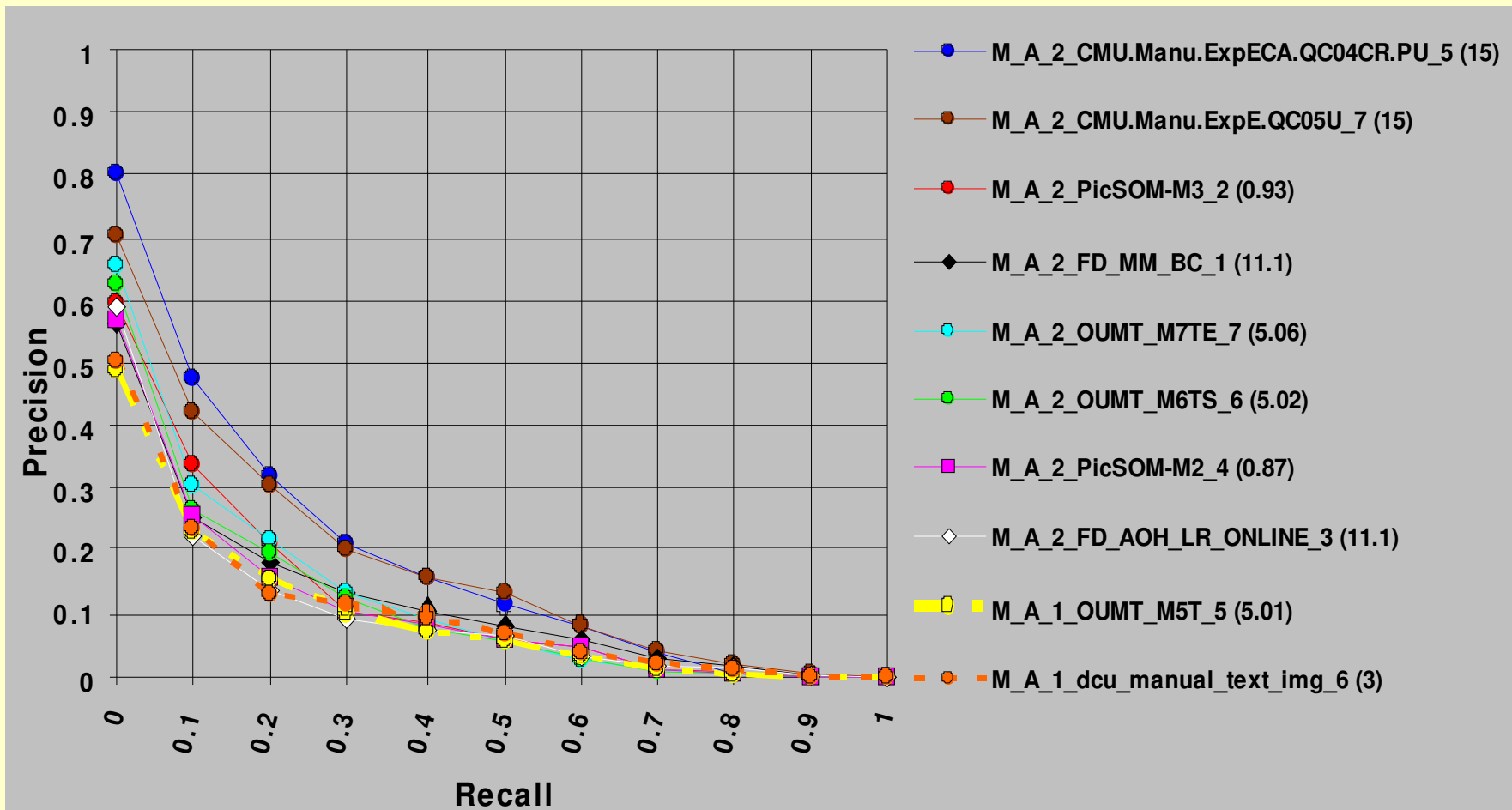
2006: Manual runs - top 10 MAP (of 11)

(mean human effort (mins) / topic)



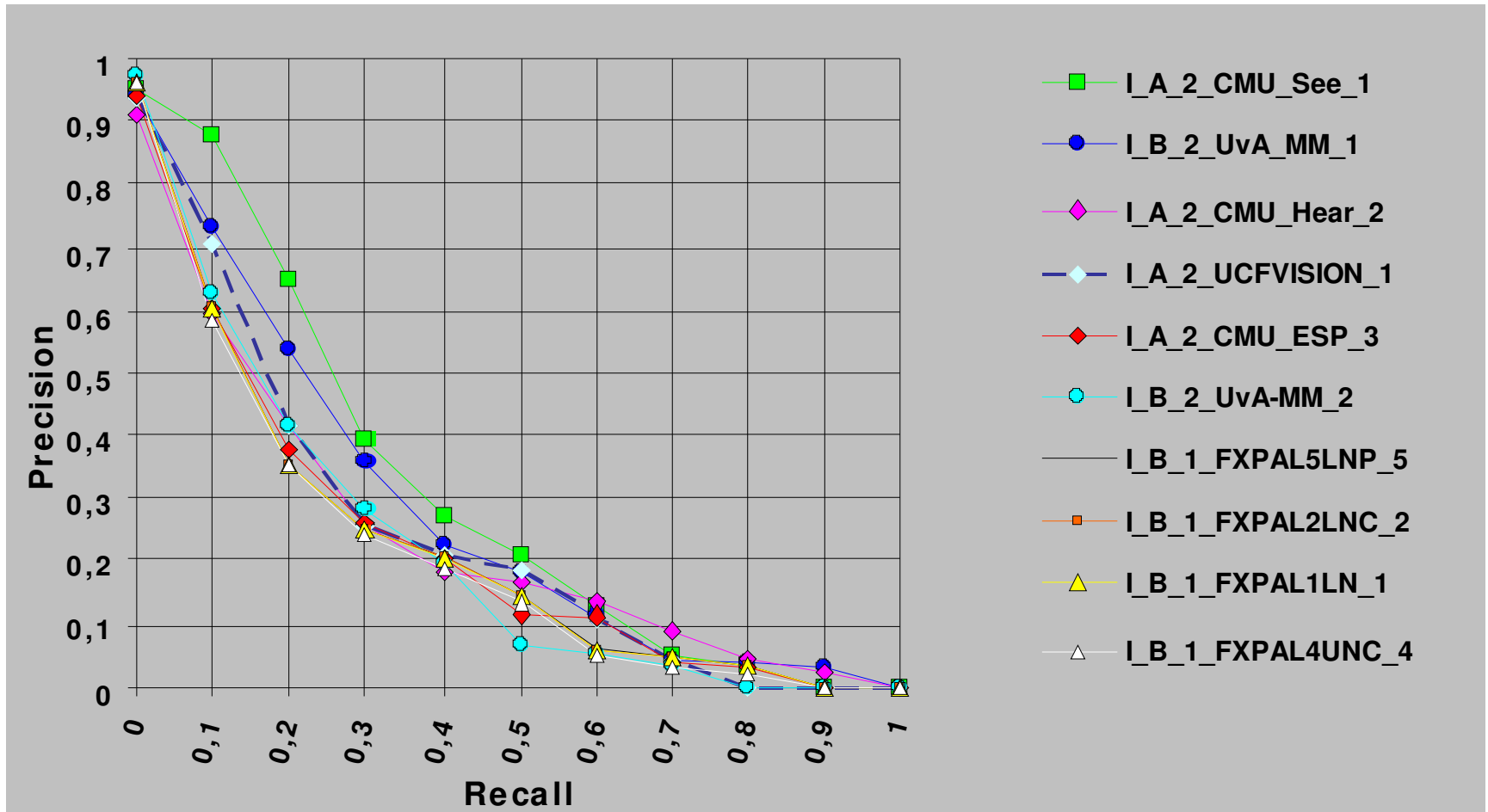
2005: Manual runs - top 10 MAP (of 26)

(mean human effort (mins) / topic)



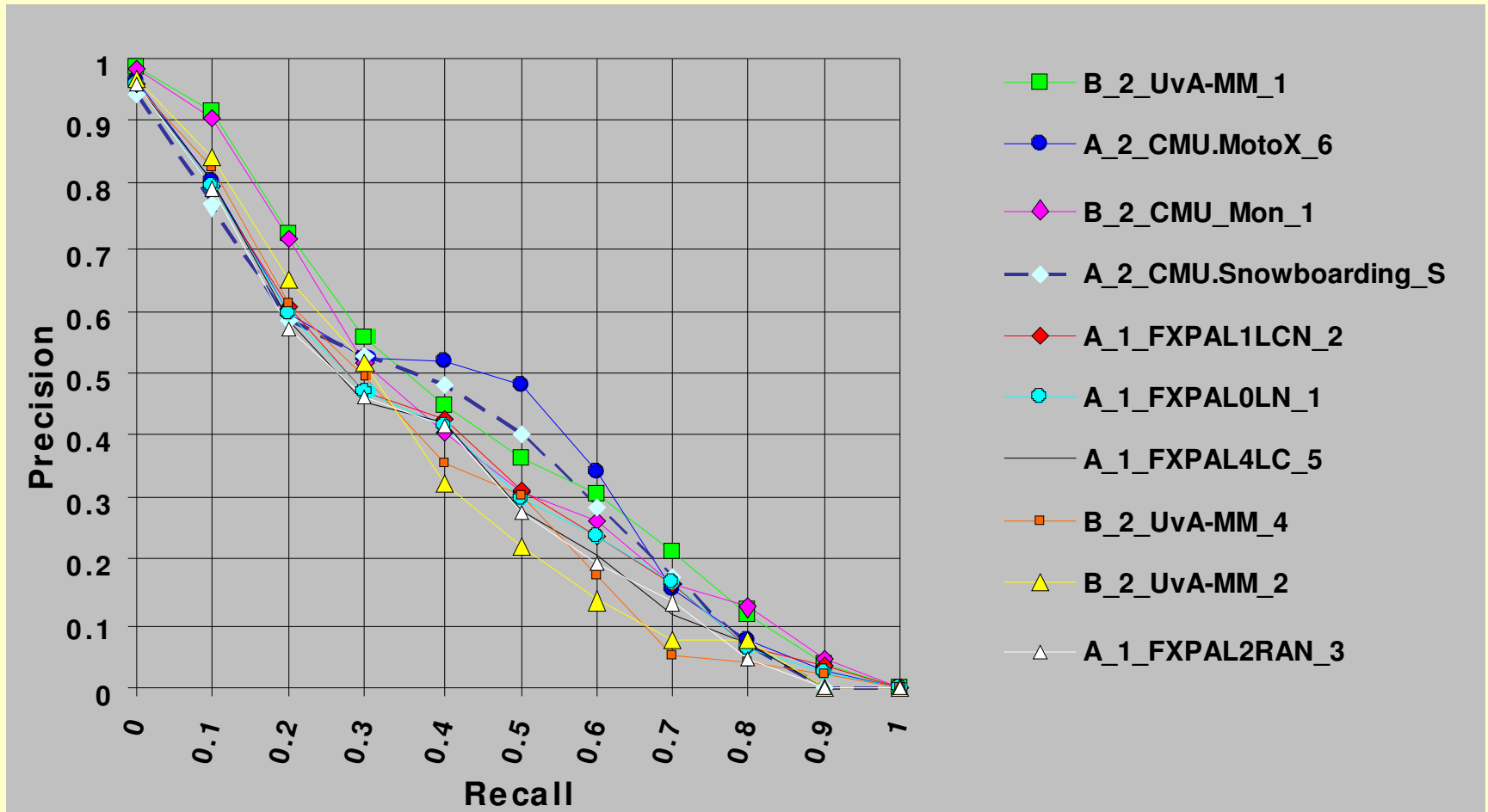
2006: Interactive runs - top 10 MAP (of 36)

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



2005: Interactive runs - top 10 MAP (of 44)

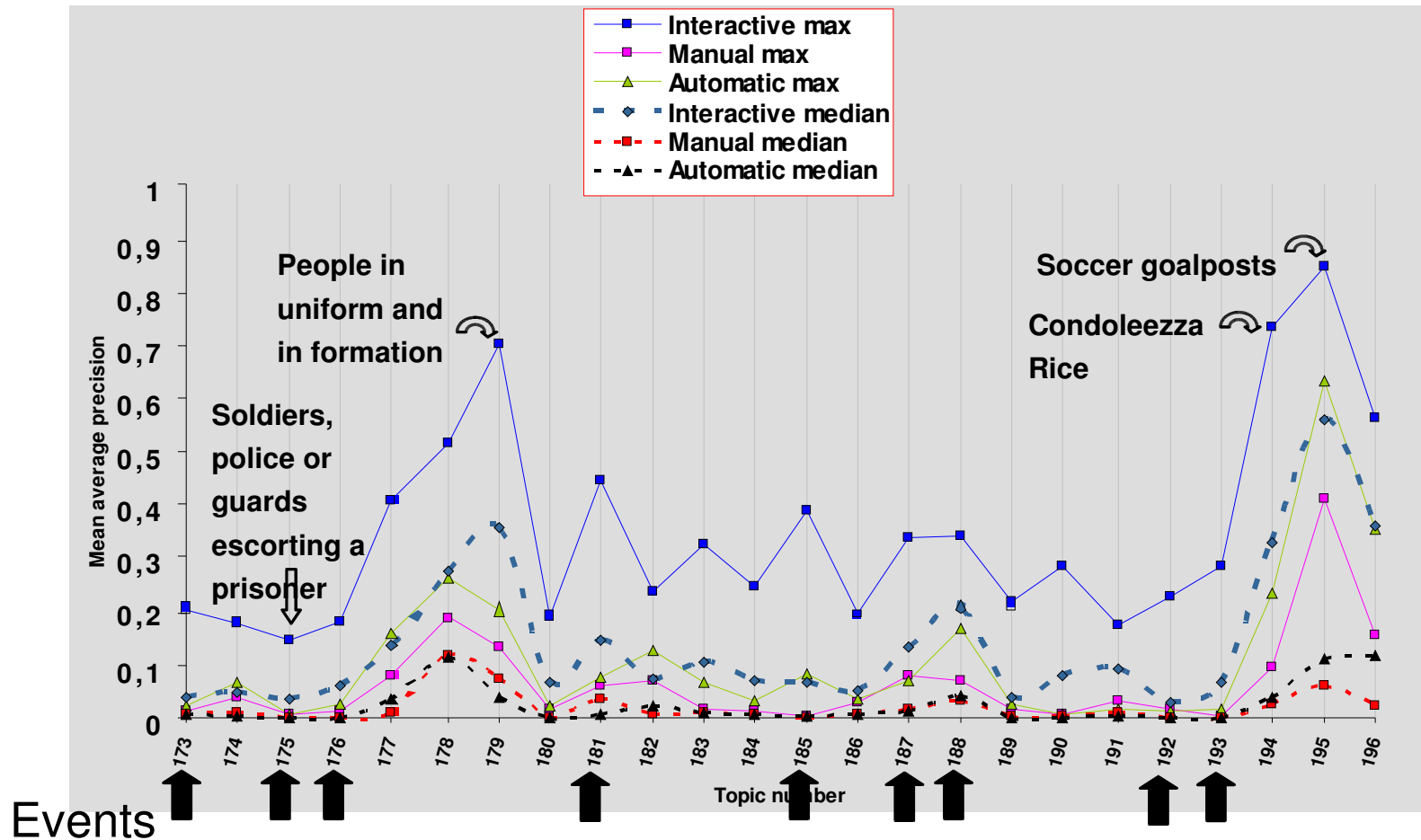
(mean elapsed time for all == ~15 mins/topic)



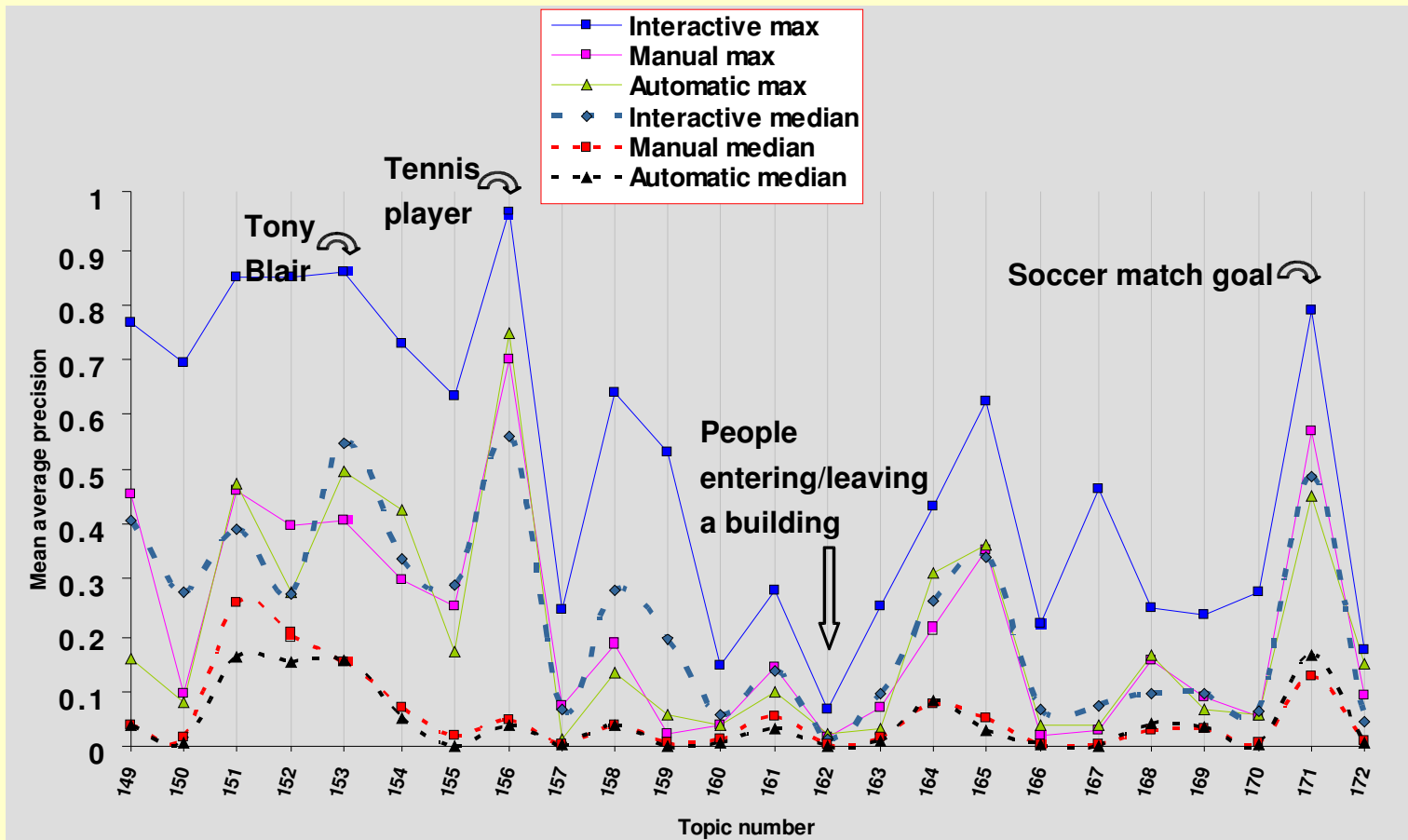
Significant differences among top 8 interactive runs (using randomization test, $p < 0.05$)

	Run name	(MAP)	
*	A_2_CMU_See_1	(0.303)	
>	B_2_UvA-MM_1	(0.267)	A_2_CMU_See_1
>	A_2_CMU_Hear_2	(0.226)	> B_2_UvA-MM_1
>	A_2_UCFVISION_1	(0.225)	> A_2_UCFVISION_1
>	A_2_CMU_ESP_3	(0.216)	> A_2_CMU_ESP_3
>	B_2_UvA-MM_2	(0.212)	> B_2_UvA-MM_2
>	B_1_FXPAL5LNP_5	(0.210)	> B_1_FXPAL5LNP
>	B_1_FXPAL4UNC_4	(0.210)	> B_1_FXPAL4UNC
			> A_2_CMU_Hear_2

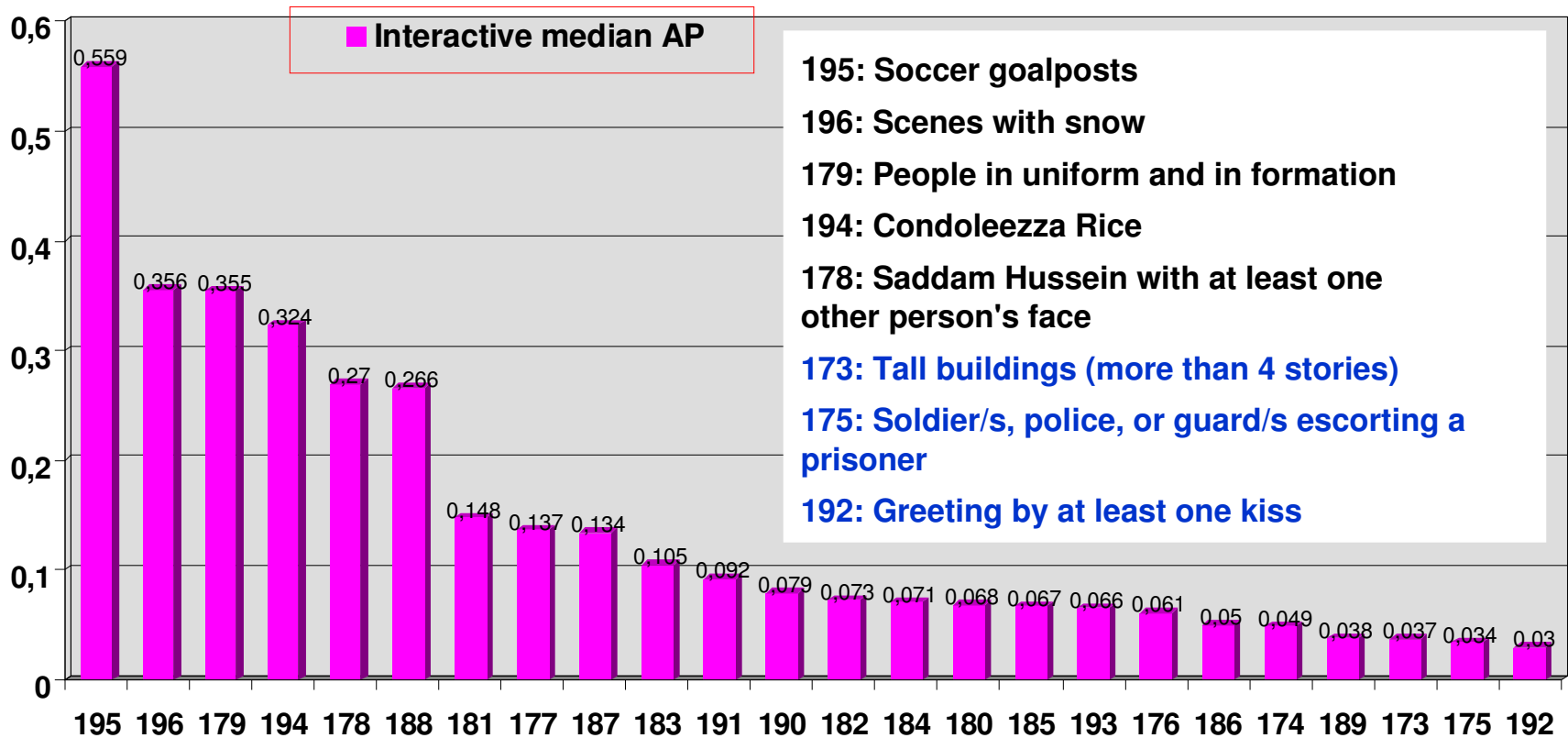
2006: Average precision by topic



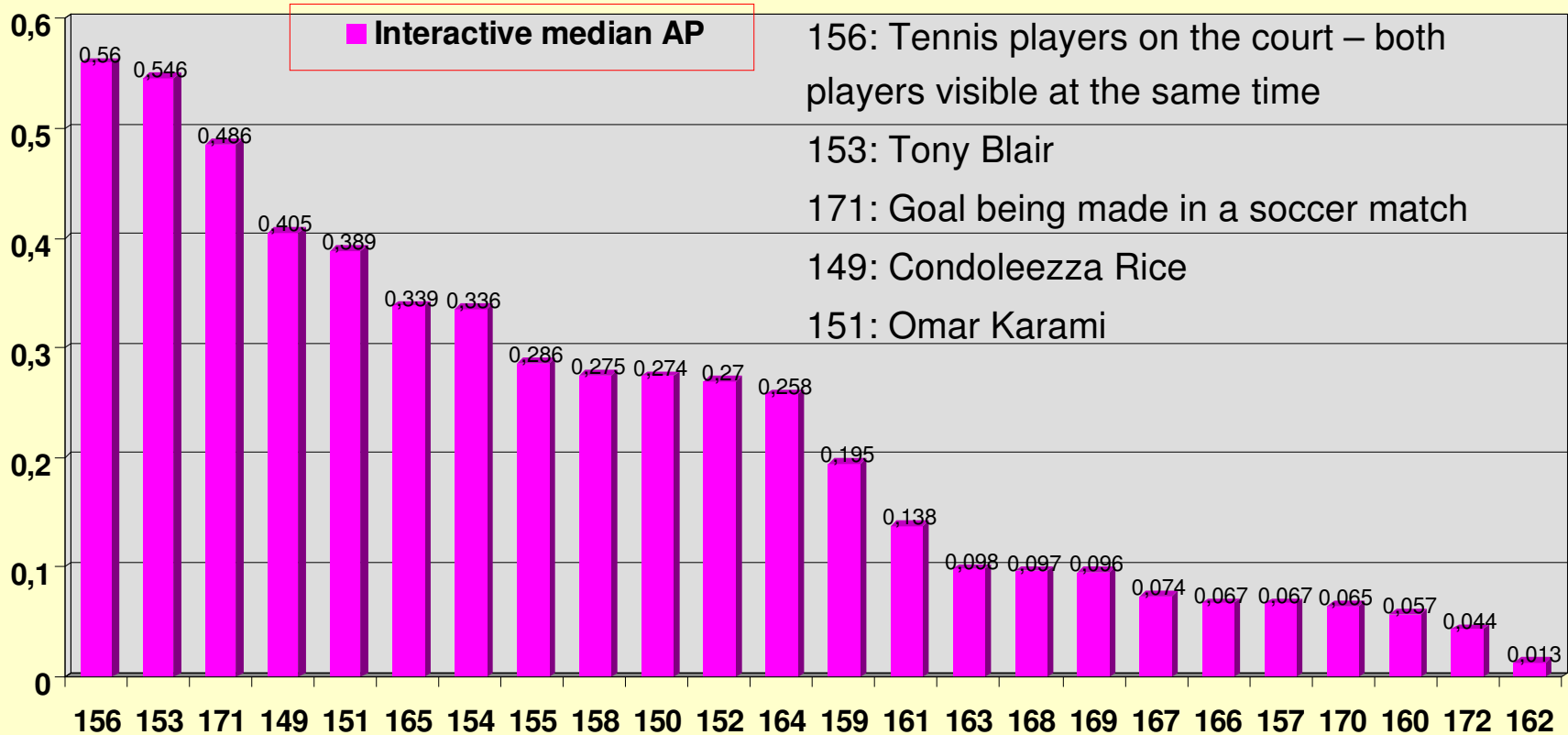
2005: Average precision by topic



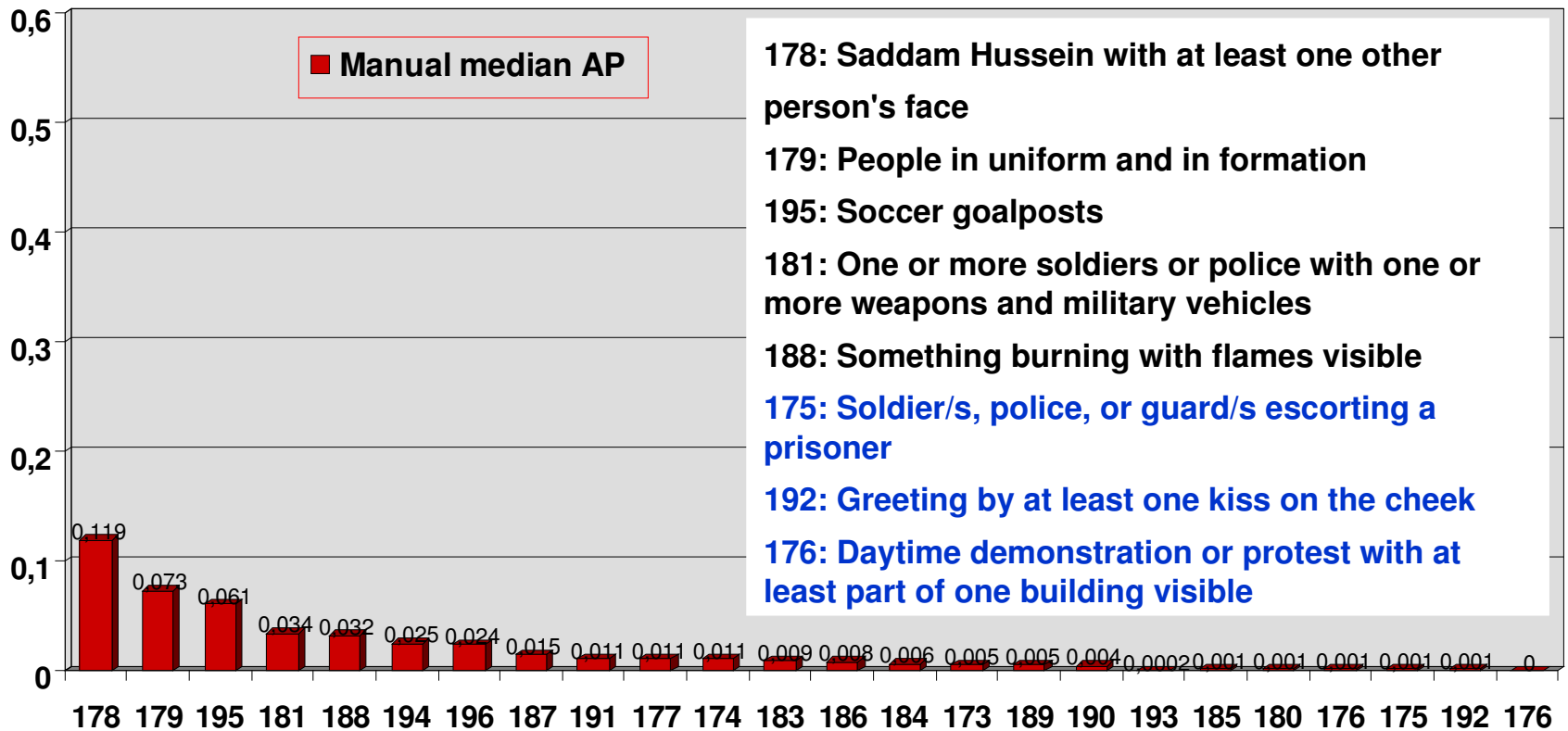
2006: Interactive runs' median average precision by topic



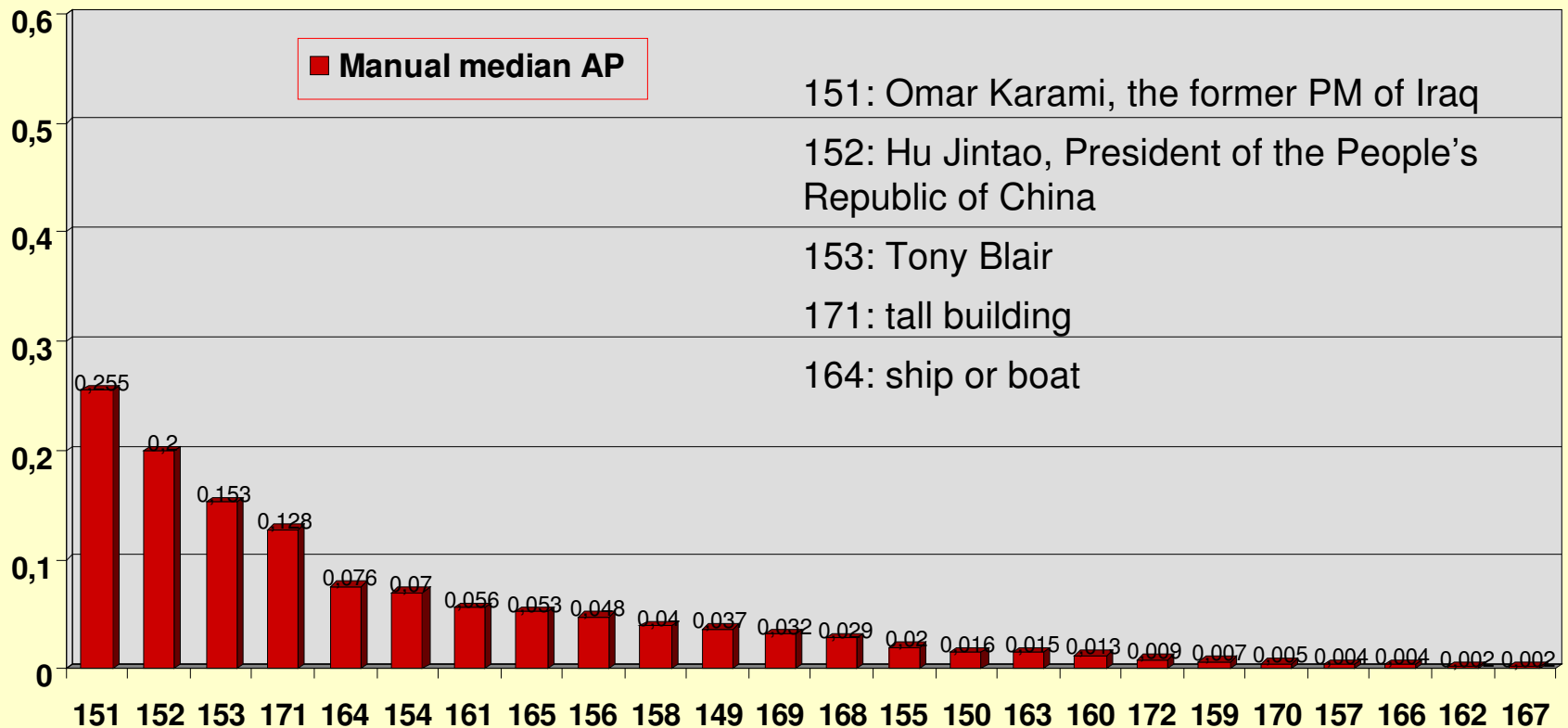
2005: Interactive runs' median average precision by topic



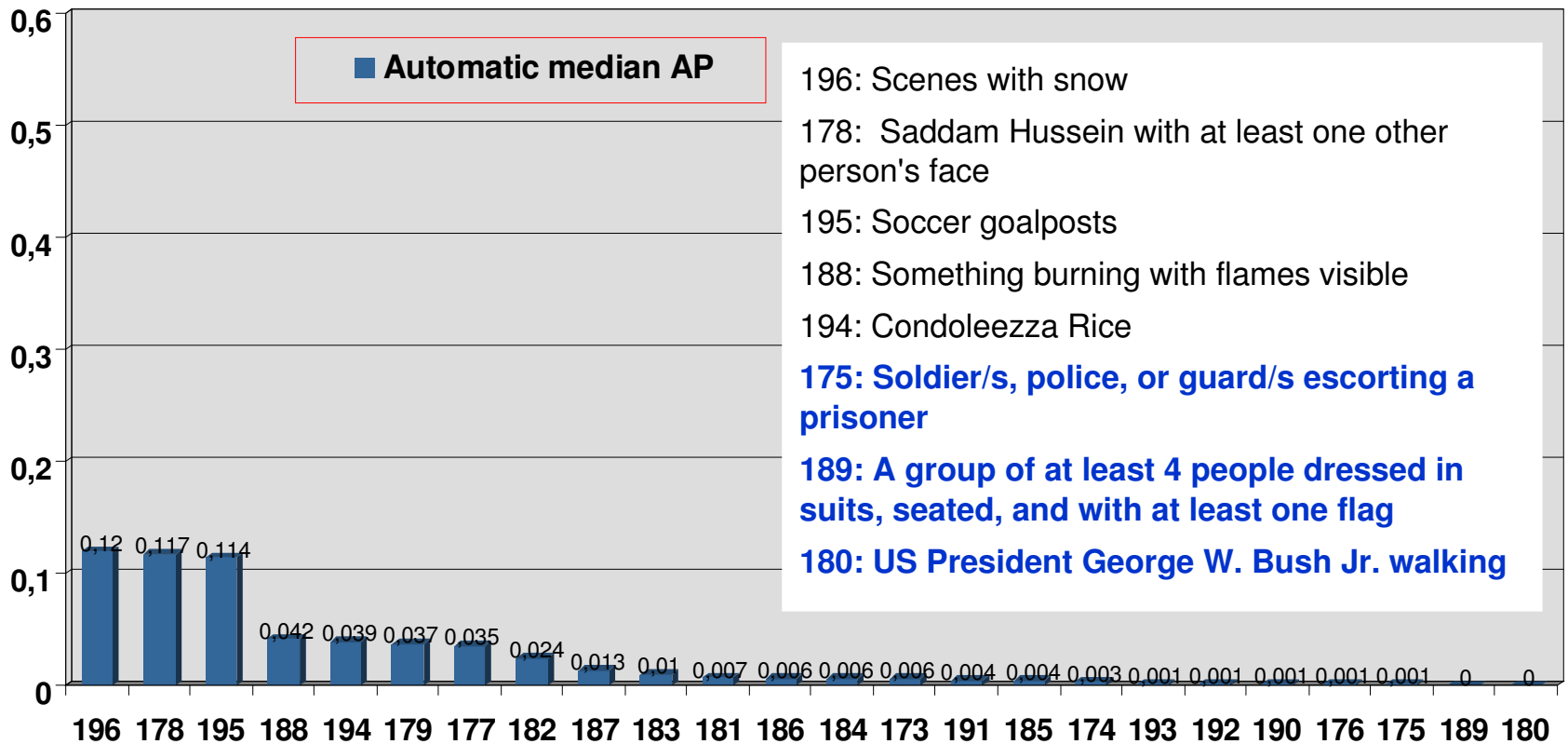
2006: Manual runs' median average precision by topic



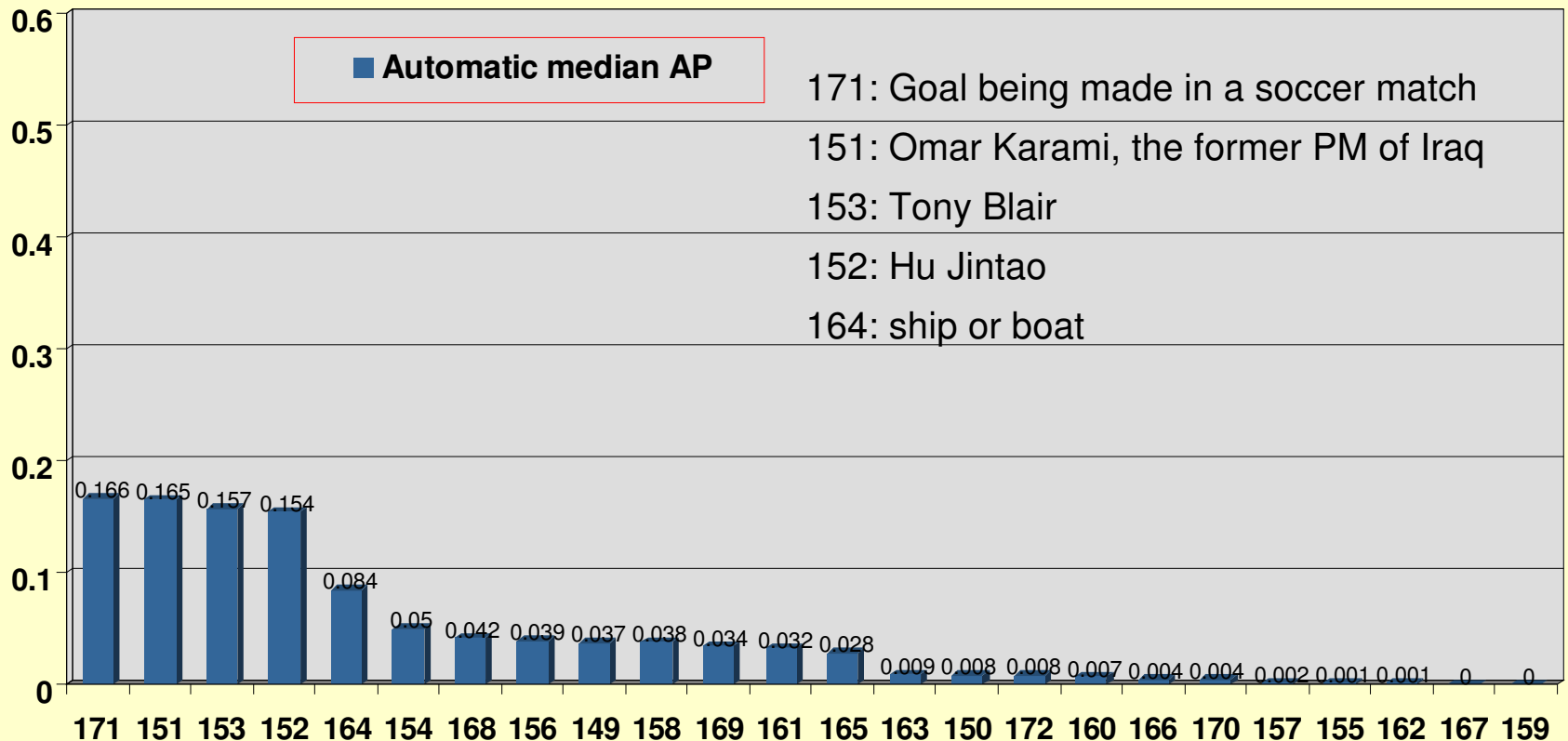
2005: Manual runs' median average precision by topic



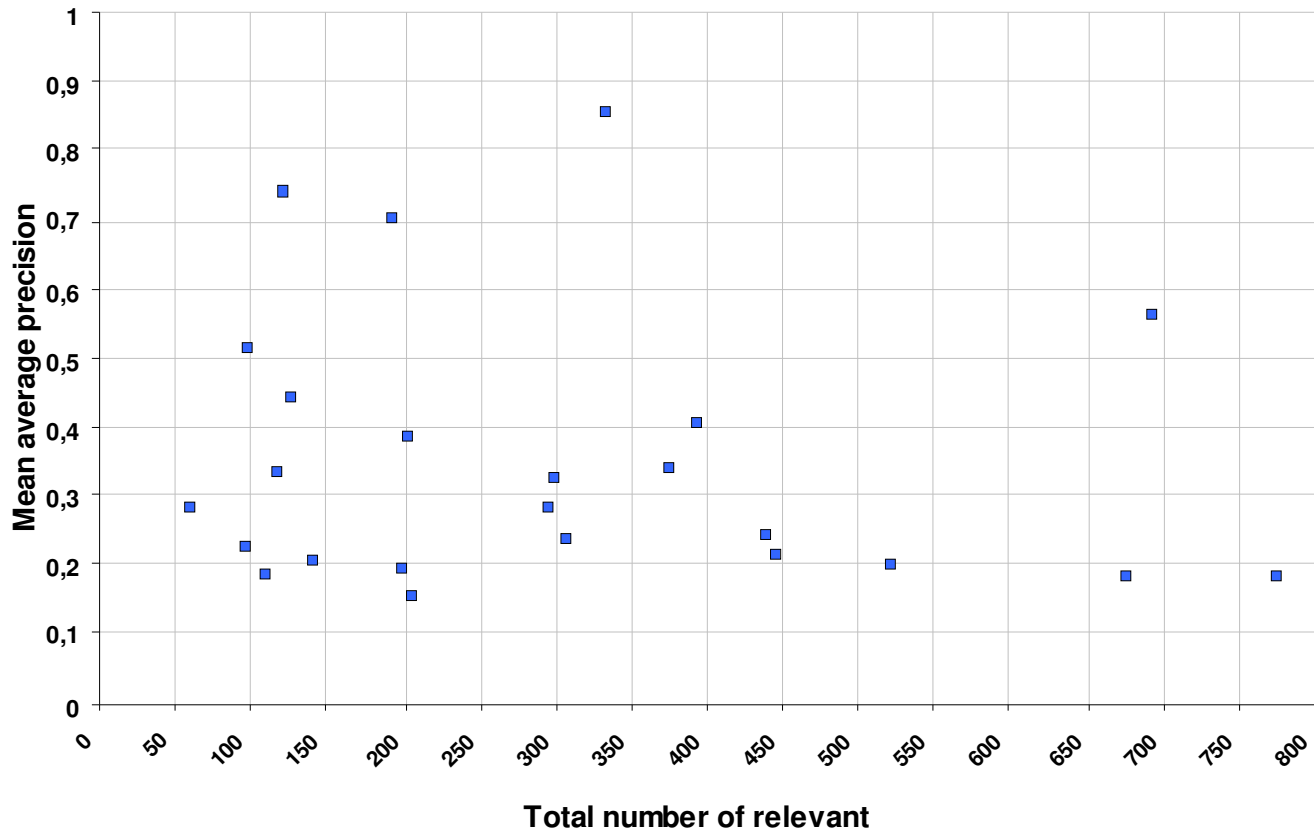
2006: Automatic runs' median average precision by topic



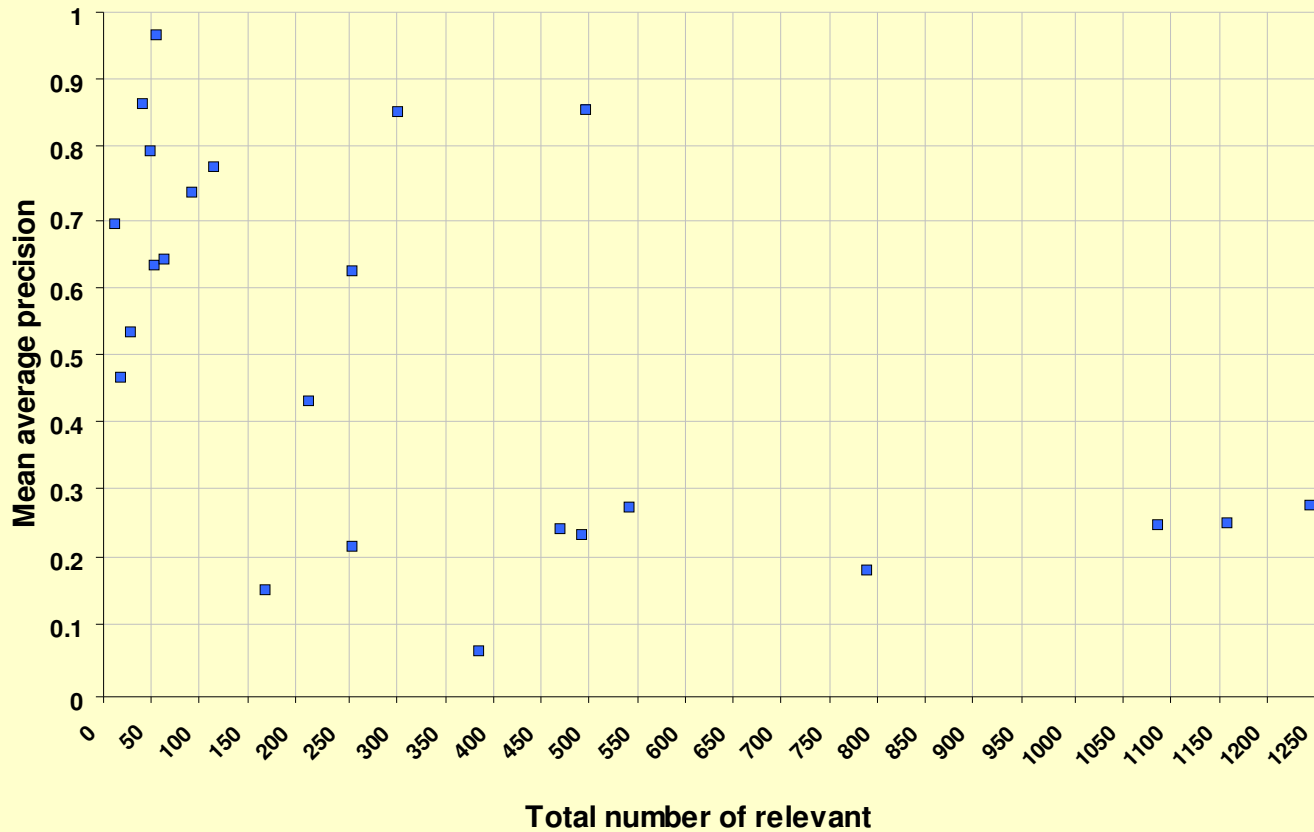
2005: Automatic runs' median average precision by topic



2006: Mean average precision (interactive max) vs total number relevant



2005: Mean average precision (interactive max) vs total number relevant



Who did what ?

- Speaker slots to follow:
 - Carnegie Mellon University
 - University of Amsterdam
 - Columbia University
 - IBM
- Demos ?
- Posters ?

Observations 2005 !

- 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, but data is different, systems are different ... anybody run 2004 system on 2005 data ?
- Some leveraged the structured nature of B/News;
- Many did automatic search & fewer did interactive search - because its easier (no users) ?
- Most common issue explored was the best combination of text vs. image search vs. concept/features;
- Search participants are the “regulars” plus new groups, some bigger, some smaller;

Observations 2006

- Top performances on all 3 search types are down
 - Test collection is twice as big
 - Half as many relevant shots
 - Harder topics ? Data ? 'Events' in topics ?
- Again, increase in automatic search & fewer did interactive search, almost nobody manual
 - It's easier (no users)?
 - Topic to query translation good enough?
 - ?
- Manual runs no longer outperform automatic – is this because so few manual, and does it make sense to keep this processing type ?