TRECVID-2005: Search Task

Alan Smeaton
Dublin City University
&
Tzveta Ianeva
NIST

Search Task Definition

- Given a test collection, a multimedia statement of information need (topic) and a common shot boundary reference, return a ranked list of at most 1,000 shots which best satisfy the need;
- Goal: promote progress in content-based retrieval from digital video via open, metrics-based evaluation;
- Many thanks to
  - Christian Petersohn (Fraunhofer Institute) for master shot reference
  - DCU team for formatting and selecting keyframes
  - Jonathan Lasko for the shot boundary truth data creation
  - CMU & Randy Paul for getting a government contractor to provide MT/ASR

Search Task Definition

- NIST created topics based on a number of basic search types: generic/specific and person/thing/event where there are multiple relevant shots coming from more than one video;
- Videos were viewed by NIST personnel (with sound turned off), notes taken on content, and candidates emerged and were chosen;
- 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; but the same searcher does not need to be used for all topics in a run.

Overarching Goals

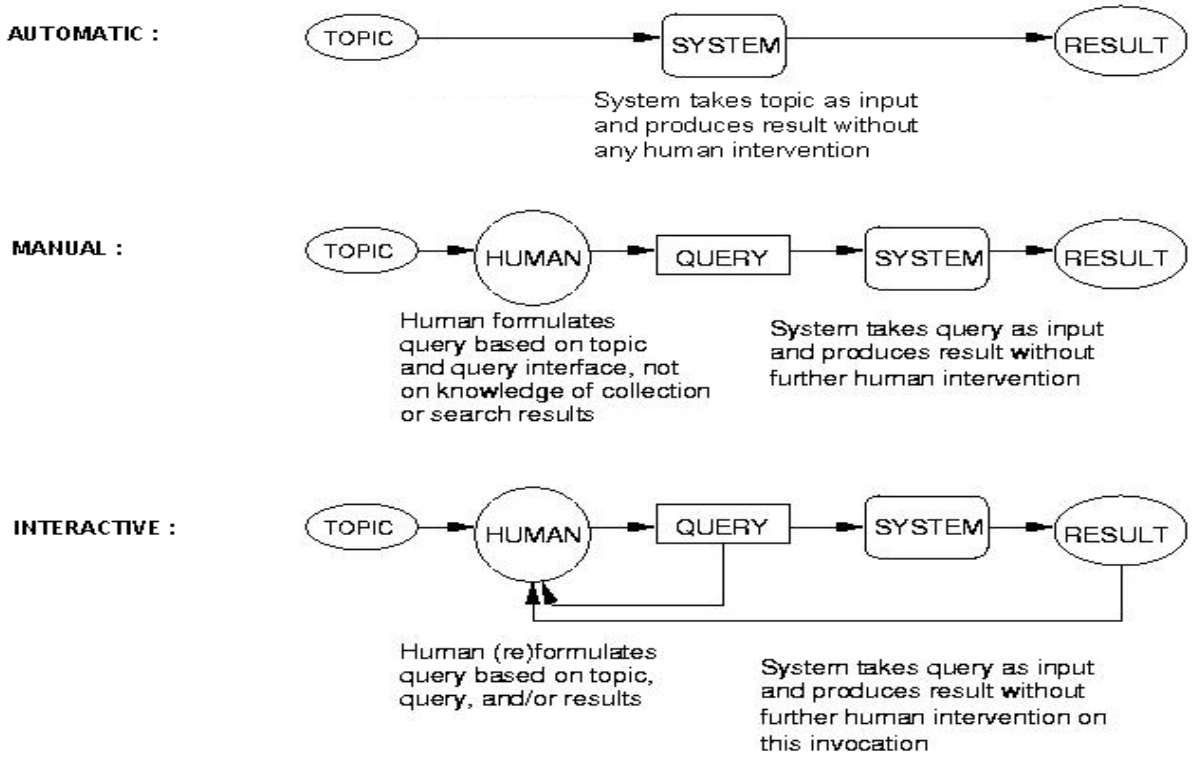
- Previous TRECVids show huge benefit from using text (ASR, closed captions, video OCR);
- TRECVid 2005 data is (deliberately) text-noisy with video from English language, Arabic & Chinese broadcasts;
- Text is derived from speech recognition and then machine translation, thus poorer quality than previously ?

- Net outcome is that task is harder, more emphasis on visual and less on text ?

2005: Search task participants (20, up from 16)

Bilkent University	Turkey
Carnegie Mellon University	USA
Columbia University	USA
Dublin City University	Ireland
Fudan University	China
FX Palo Alto Laboratory	USA
Helsinki University of Technology	Finland
IBM	USA
Imperial College London	UK
Language Computer Corporation (LCC)	USA
Lowlands Team (CWI, Twente, U. of Amsterdam)	Netherlands
Mediamill Team (Univ. of Amsterdam and TNO)	Netherlands
National University of Singapore (NUS)	Singapore
Queen Mary University of London	UK
SCHEMA-Univ. Bremen Team	EU
Tsinghua University	China
University of Central Florida / University of Modena	USA,Italy
University of Iowa	USA
University of North Carolina	USA
University of Oulu / MediaTeam	Finland

Search Types: Automatic, Manual and Interactive



Number of runs: 42 automatic (up from 23)
26 manual (down from 52)
44 interactive (down from 61)

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

149. Find shots of Condoleeza Rice [3, 6, 116]
150. Find shots of Iyad Allawi, the former prime minister of Iraq [3, 6, 13]
151. Find Find shots of Omar Karami, the former prime minister of Lebannon [2, 5, 301]
152. Find shots of Hu Jintao, president of the People's Republic of China [2, 9, 498]
153. Find shots of Tony Blair. [2, 4, 42]
154. Find shots of Mahmoud Abbas, also known as Abu Mazen, prime minister of the Palestinian Authority. [2, 9, 93]
155. Find shots of a graphic map of Iraq, location of Bagdhad marked – not a weather map [4, 10, 54]
156. Find shots of tennis players on the court – both players visible at the same time [2, 4, 55]
157. Find shots of people shaking hands [4, 10, 470]
158. Find shots of a helicopter in flight [2, 8, 63]
159. Find shots of George Bush entering or leaving a vehicle (e.g., car, van, airplane, helicopter, etc), he and vehicle both visible at the same time [2, 7, 29]
160. Find shots of something (e.g., vehicle, aircraft, building, etc.) on fire with flames and smoke visible [2, 9, 169]

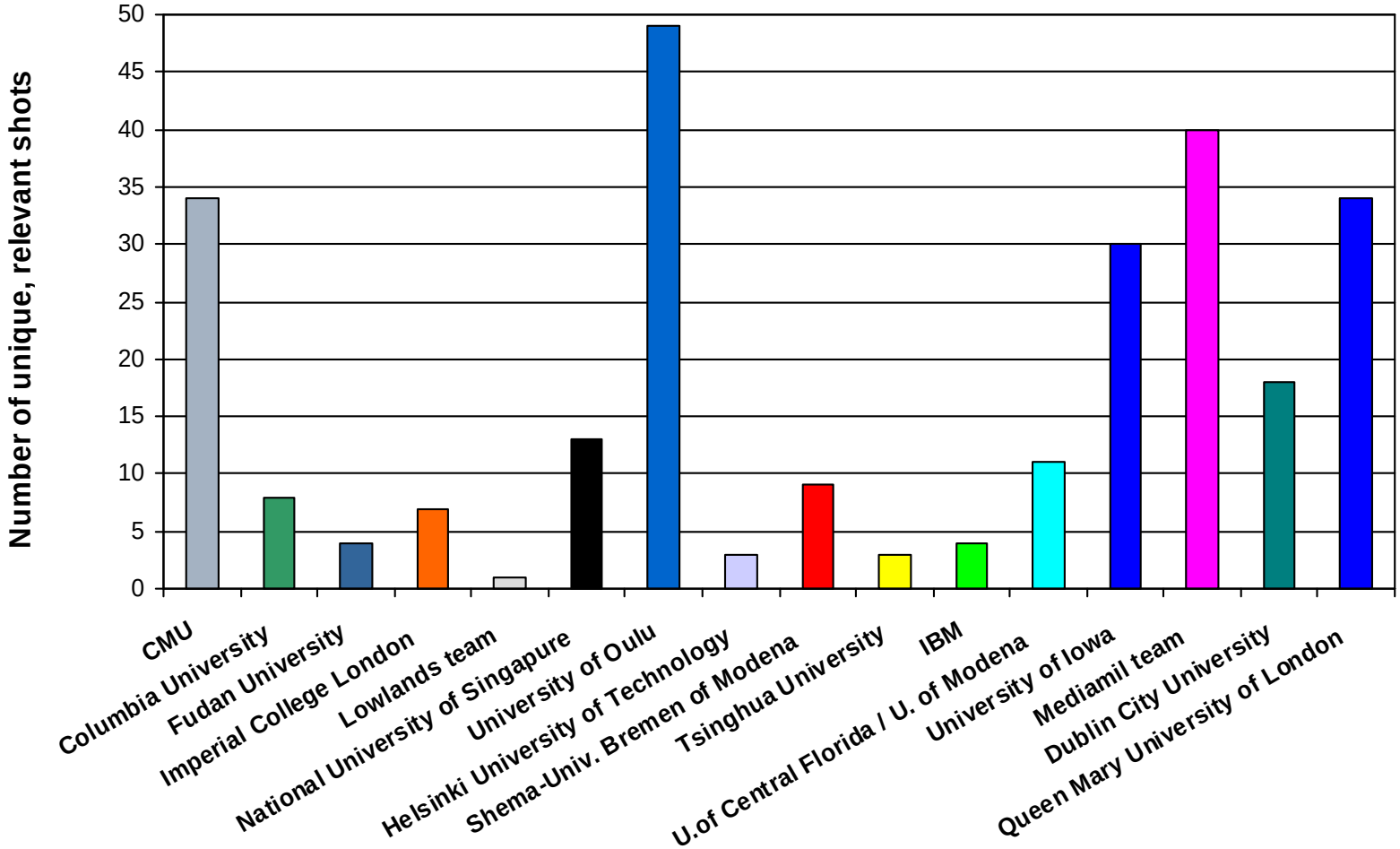
24 Topics [number of image, video examples and relevant found]

161. Find shots of people with banners or signs [2, 6, 1245]
162. Find shots of one or more people entering or leaving a building [5, 8, 385]
163. Find shots of a meeting with a large table and more than two people [2, 5, 1160]
164. Find shots of a ship or boat [3, 7, 214]
165. Find shots of basketball players on the court [2, 8, 254]
166. Find shots of one or more palm trees [2, 6, 253]
167. Find shots of an airplane taking off [2, 5, 19]
168. Find shots of a road with one or more cars [2, 5, 1087]
169. Find shots of one or more tanks or other military vehicles [3, 8, 493]
170. Find shots of tall building (with more than 5 floors above the ground) [3, 6, 543]
171. Find shots of a goal being made in a soccer match [1, 7, 49]
172. Find shots of an office setting, i.e., one or more desks/tables and one or more computers and one or more people [3, 8, 790]

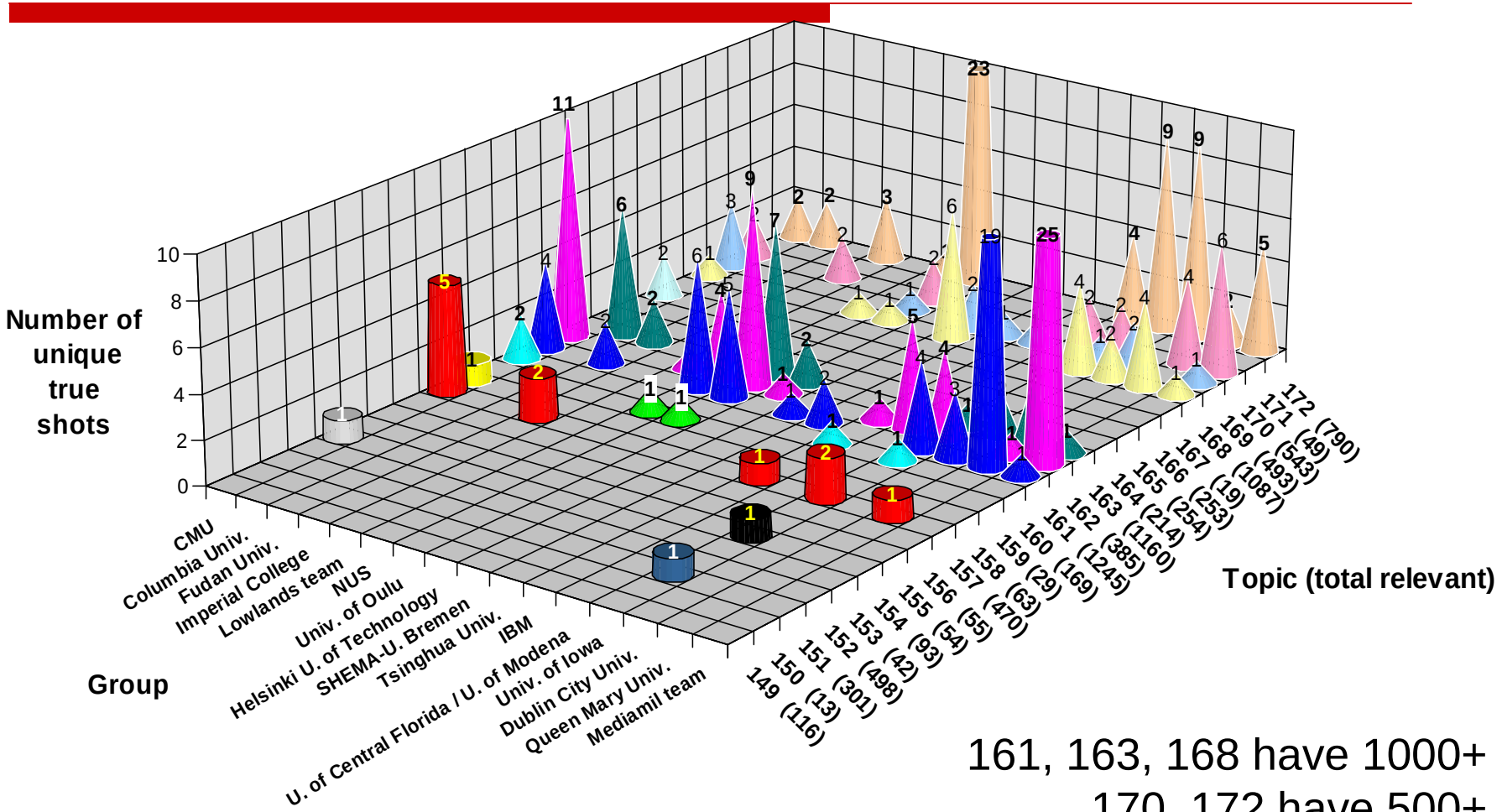
Some statistics

- **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**

2005: 16 sites contributed one or more unique, relevant shots (8 last year)



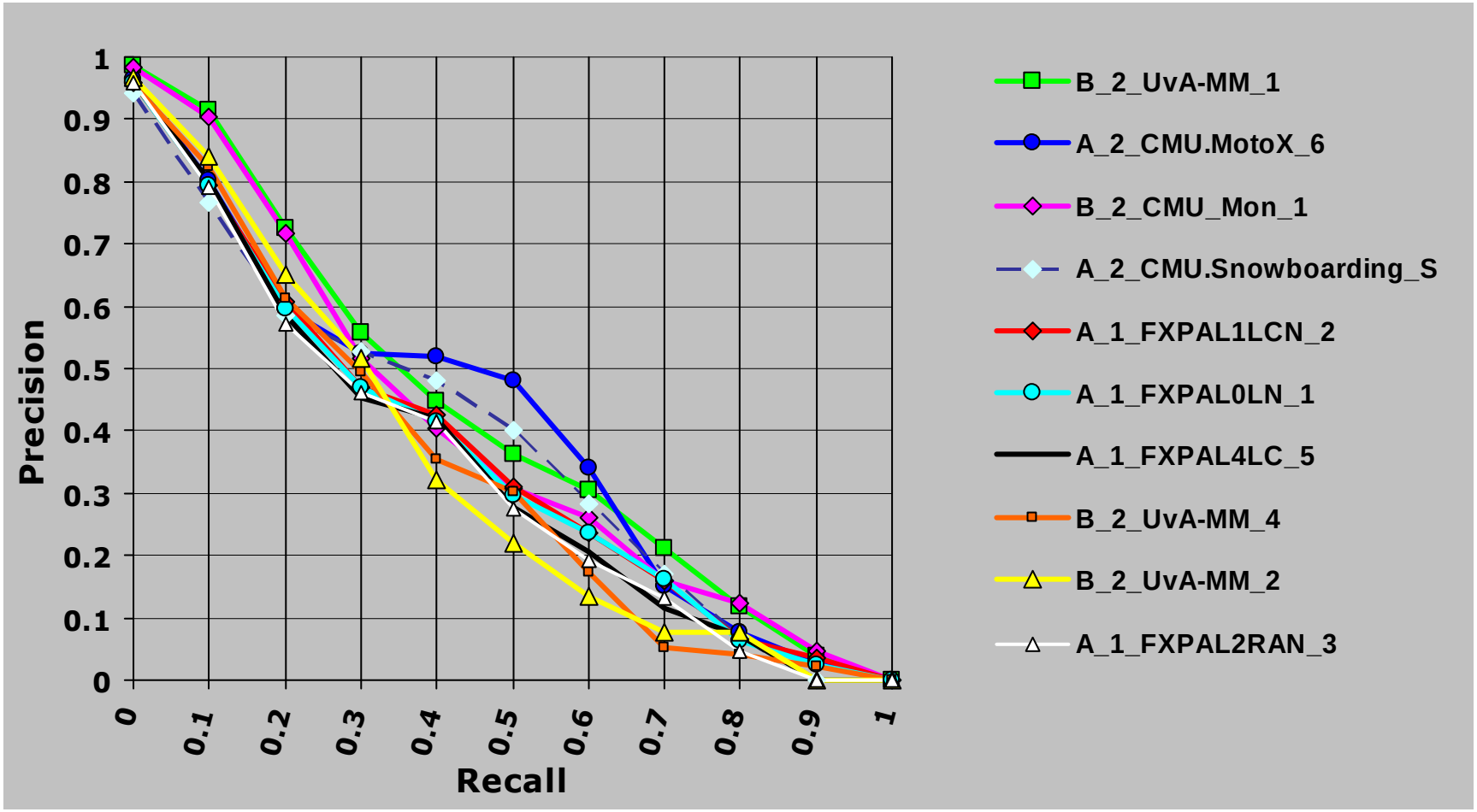
2005: Rel shots contrib. uniquely per topic by team



161, 163, 168 have 1000+
170, 172 have 500+

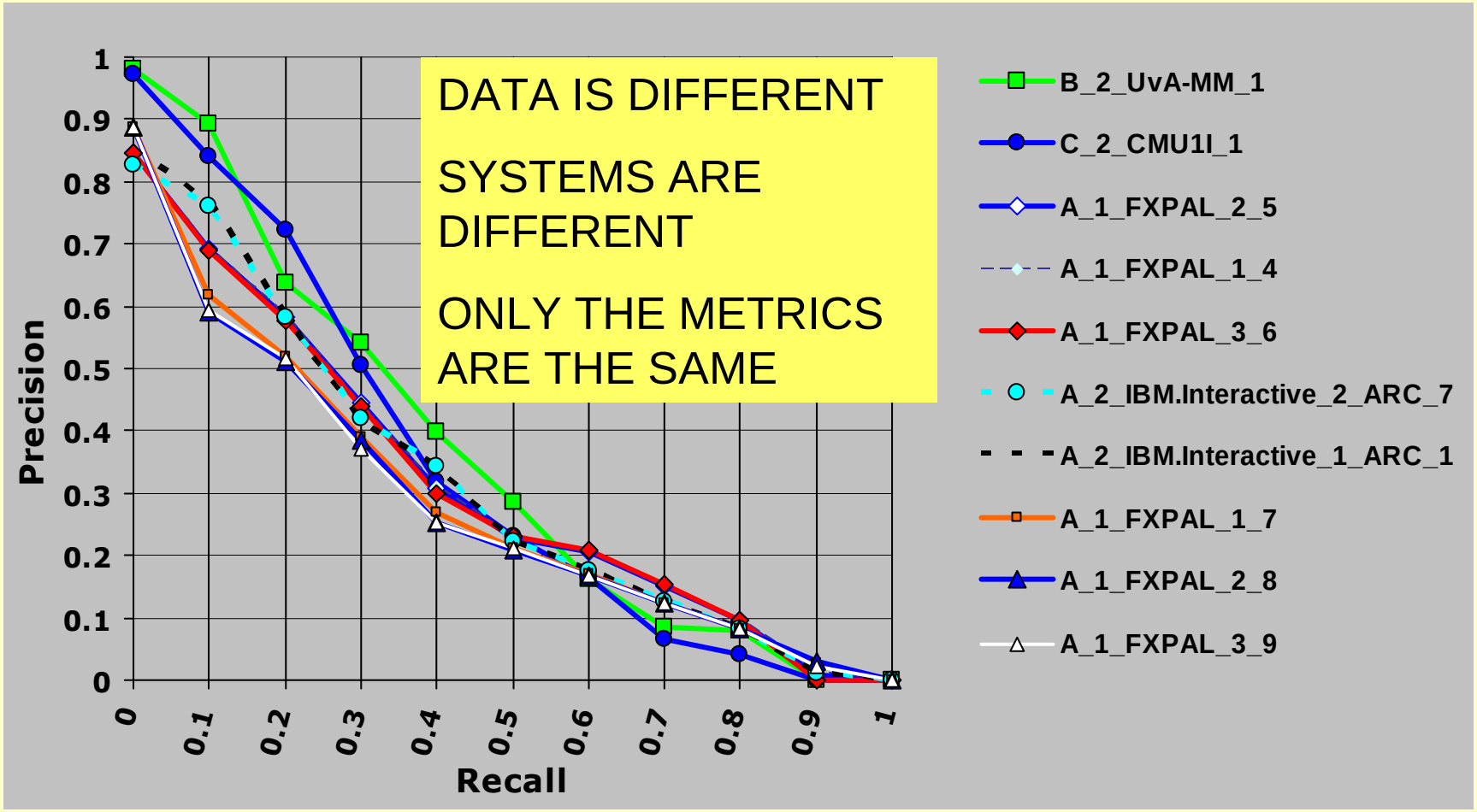
2005: Interactive runs - top 10 MAP (of 49)

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

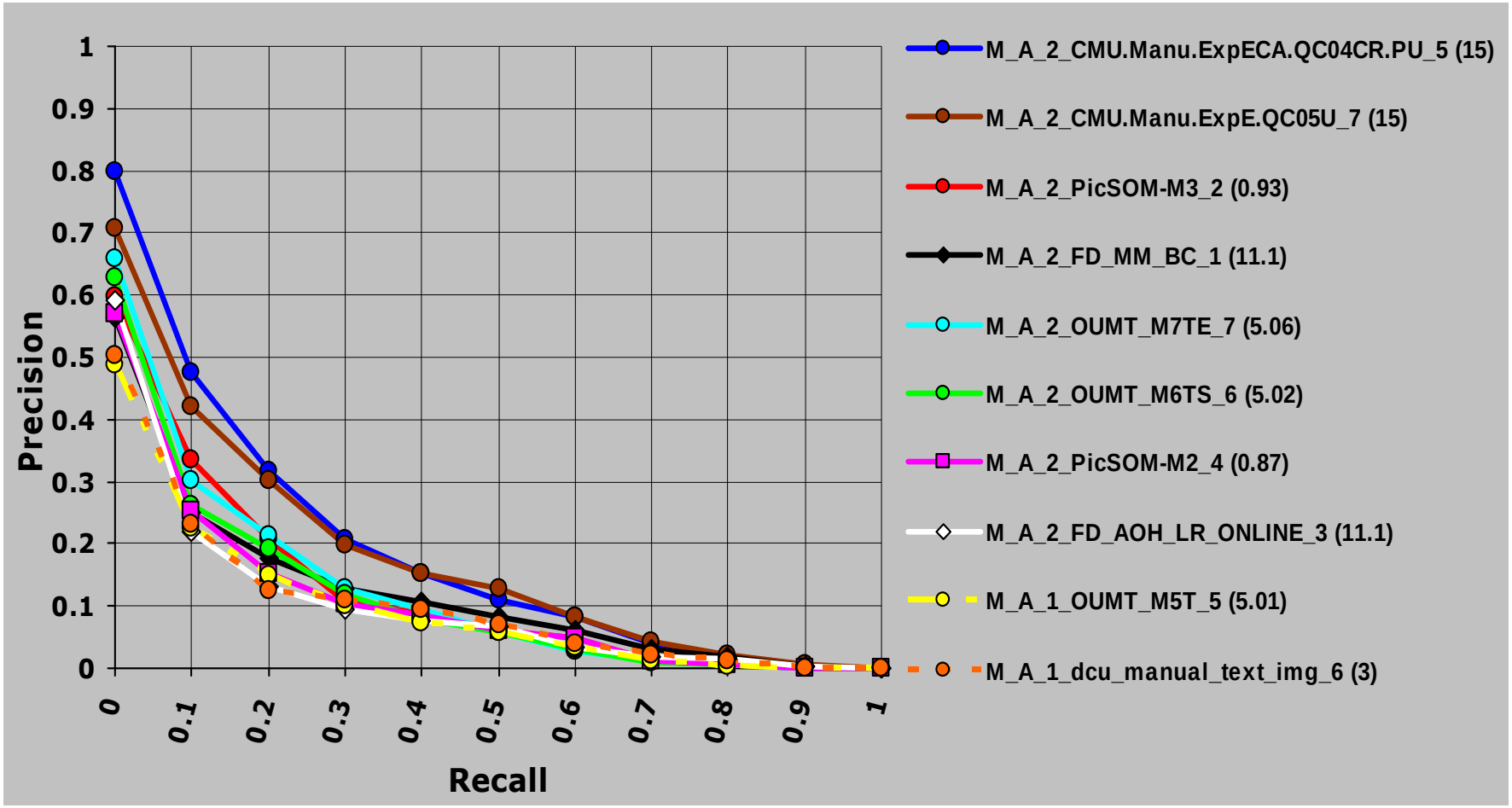


2004: Interactive runs - top 10 MAP (of 62)

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

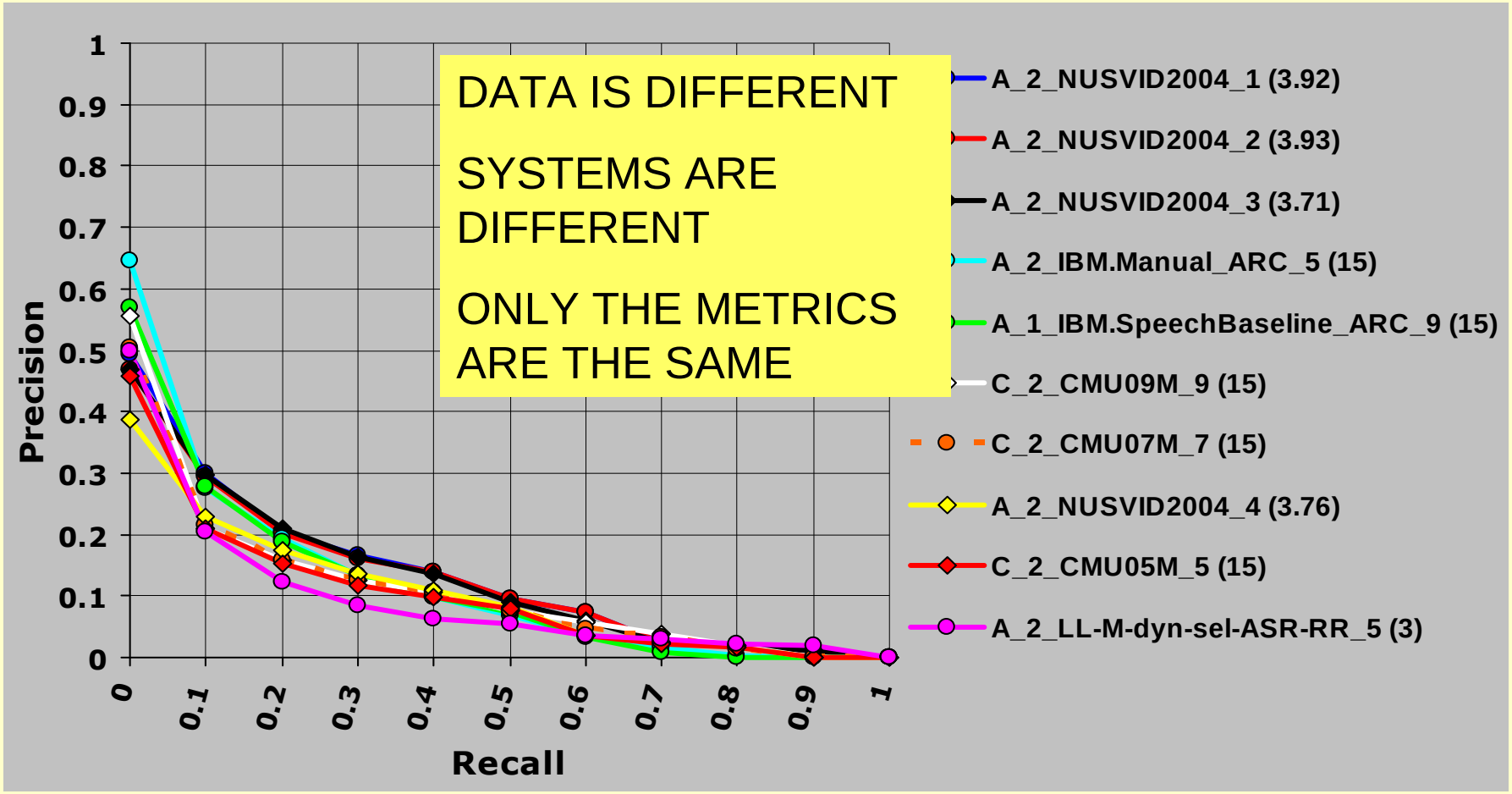


2005: Manual runs - top 10 MAP (of 26)
(mean human effort (mins) / topic)



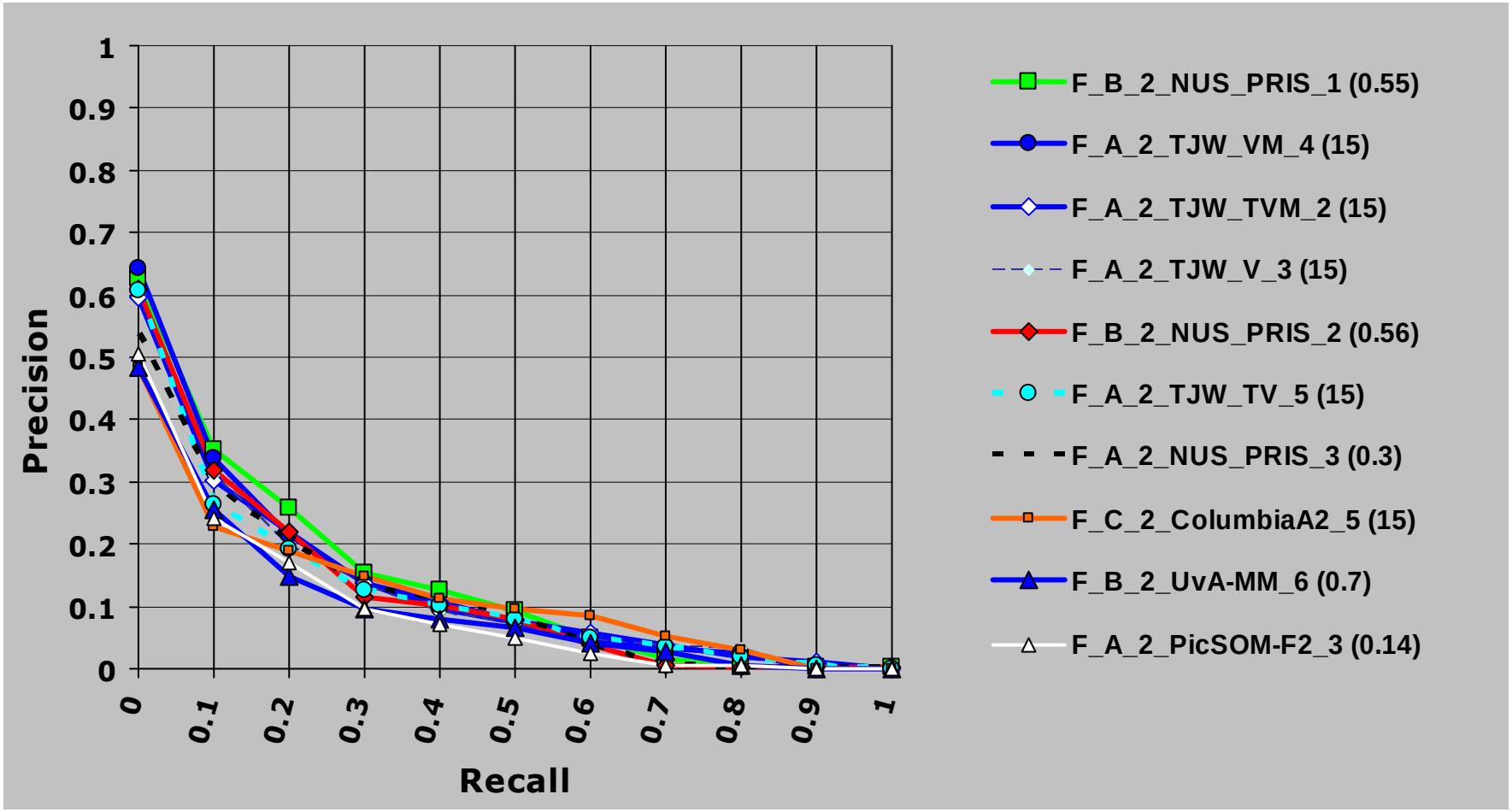
2004: Manual runs - top 10 MAP (of 52)

(mean human effort (mins) / topic)



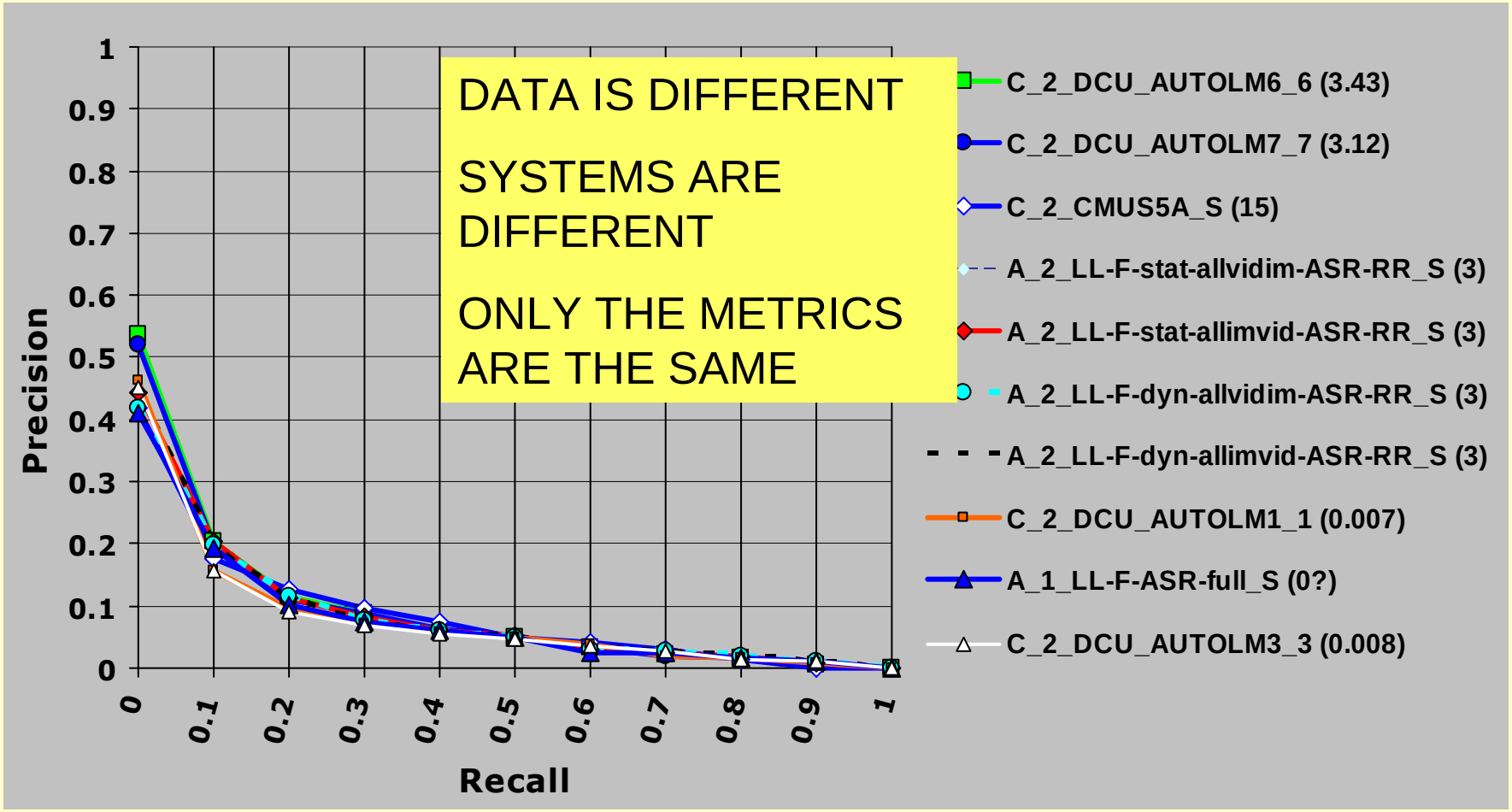
2005: Automatic runs (pilot) - top 10 MAP (of 23)

(mean elapsed time (mins) / topic)

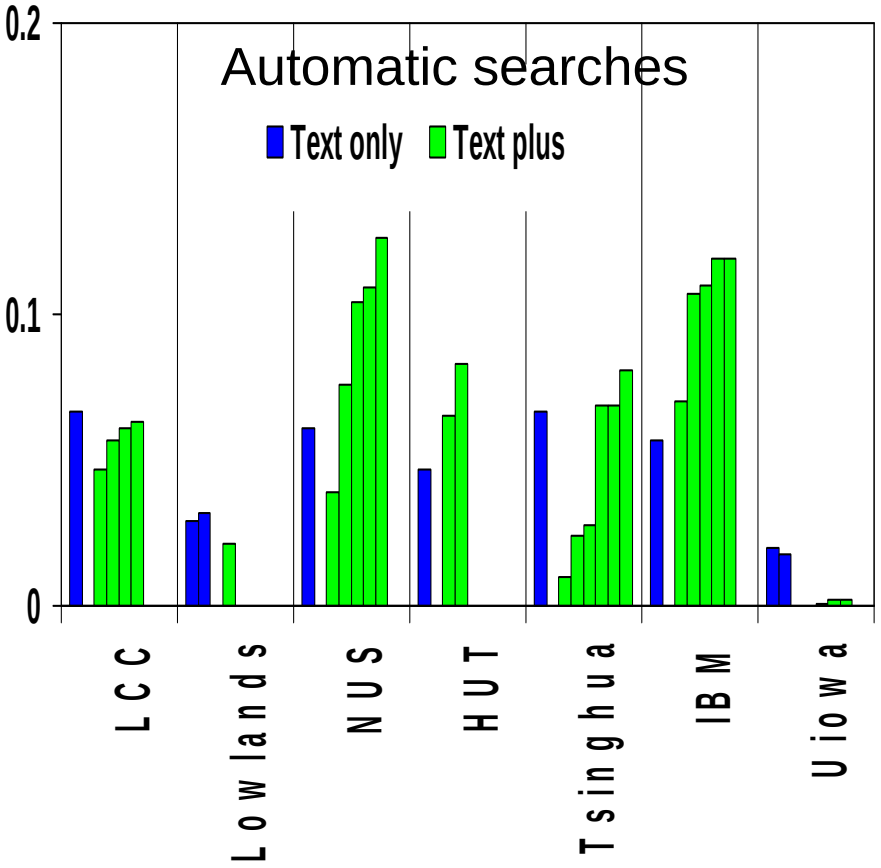


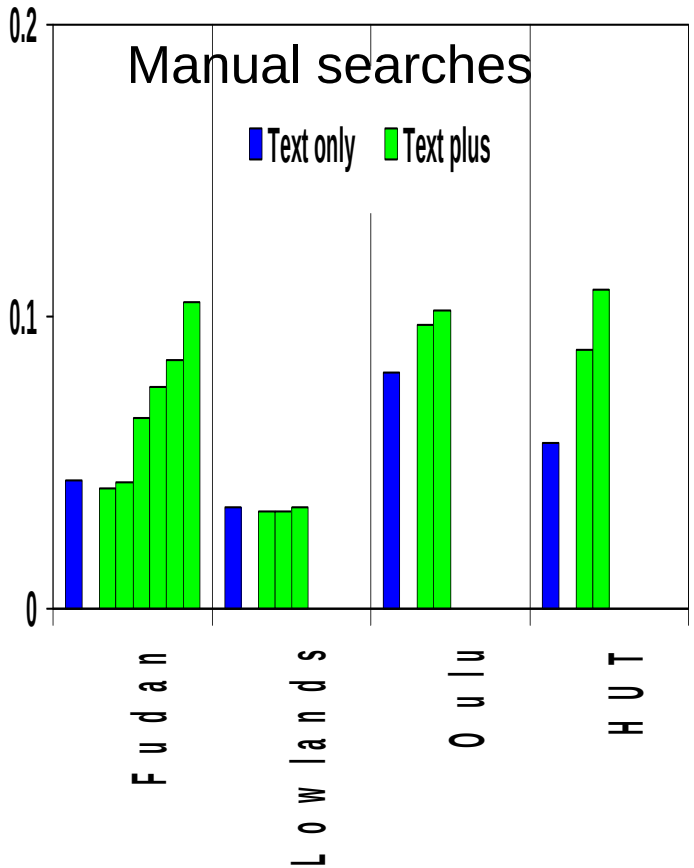
2004: Automatic runs (pilot) - top 10 MAP (of 23)

(mean elapsed time (mins) / topic)

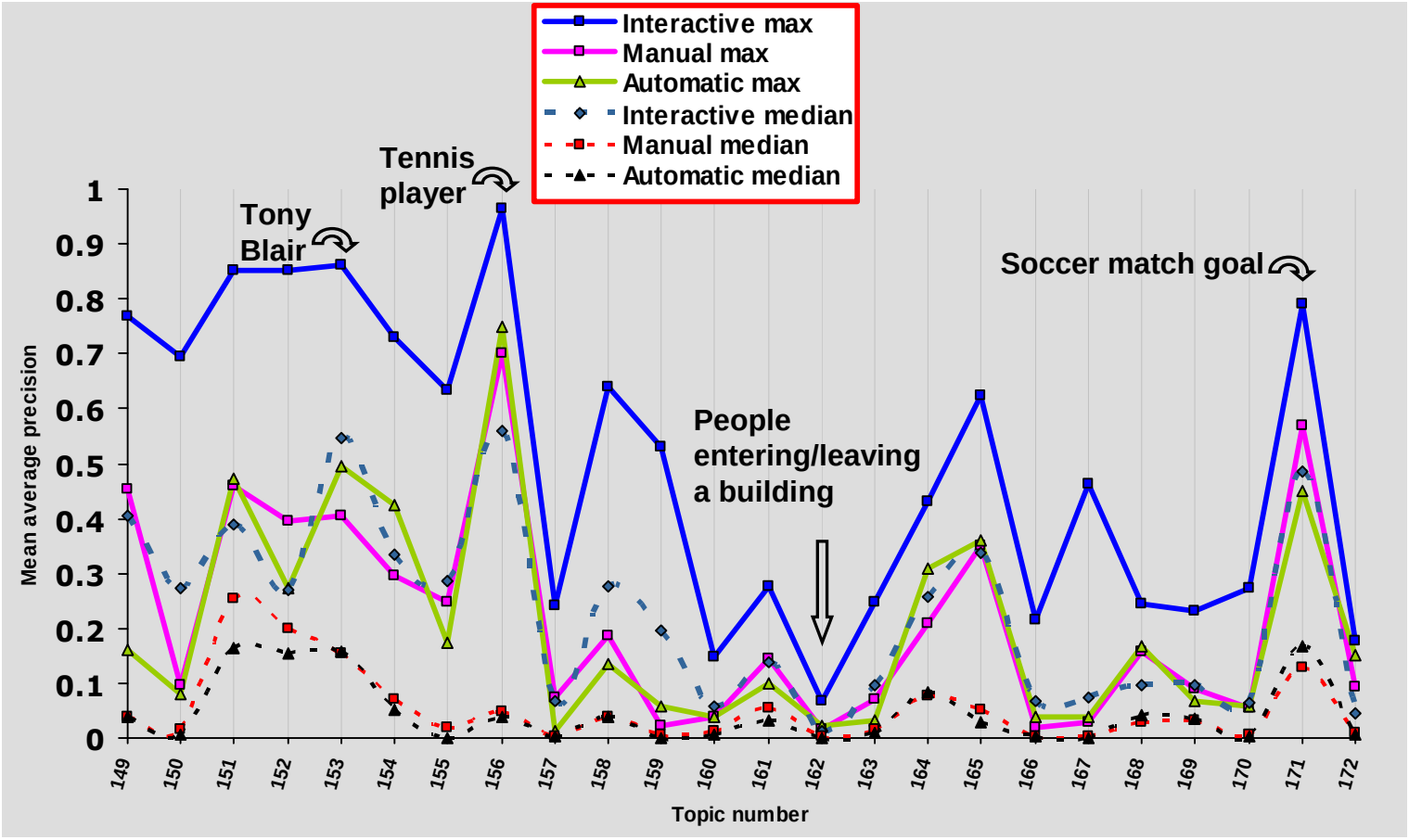


2005: Text-only versus Text-plus searches by group (using only common training data)

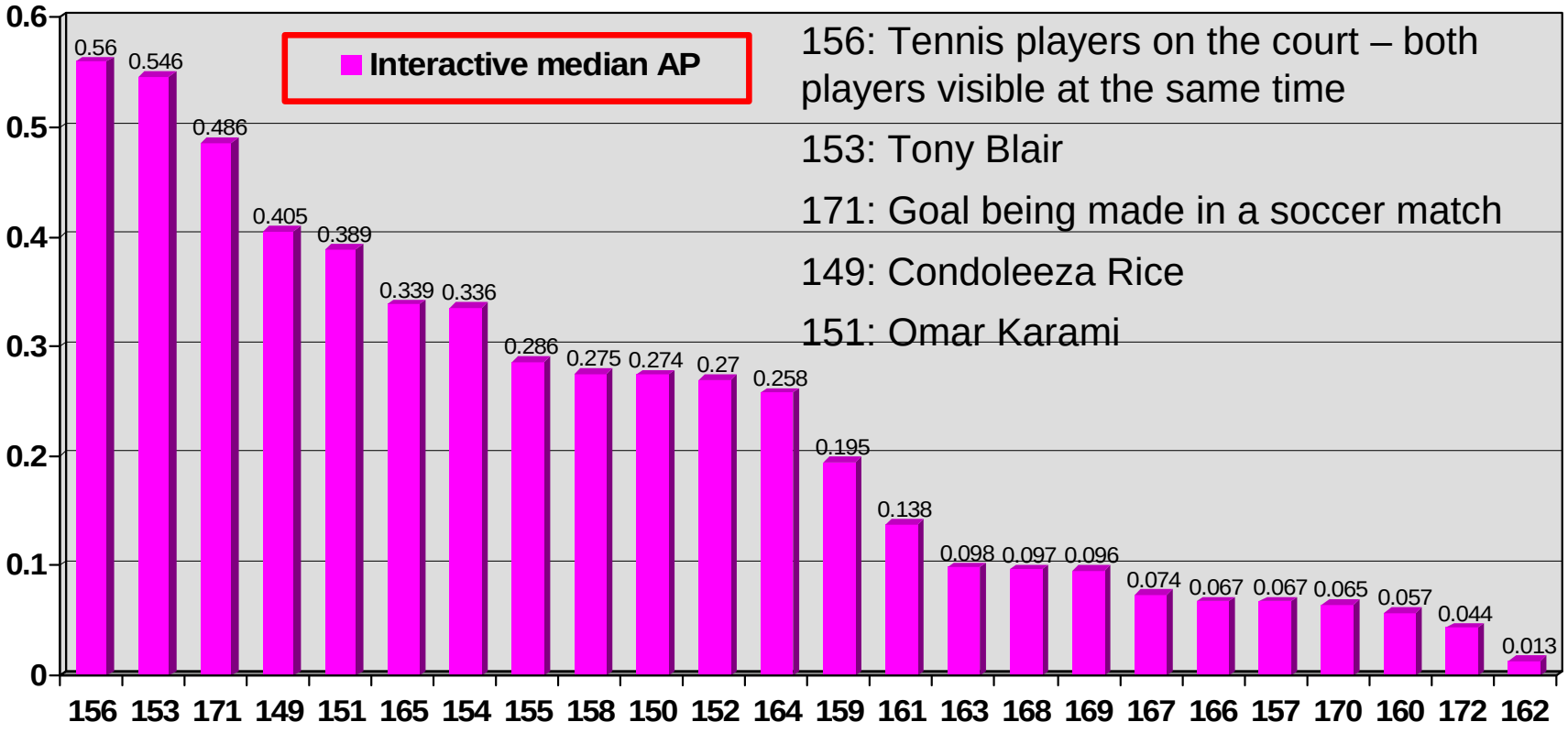




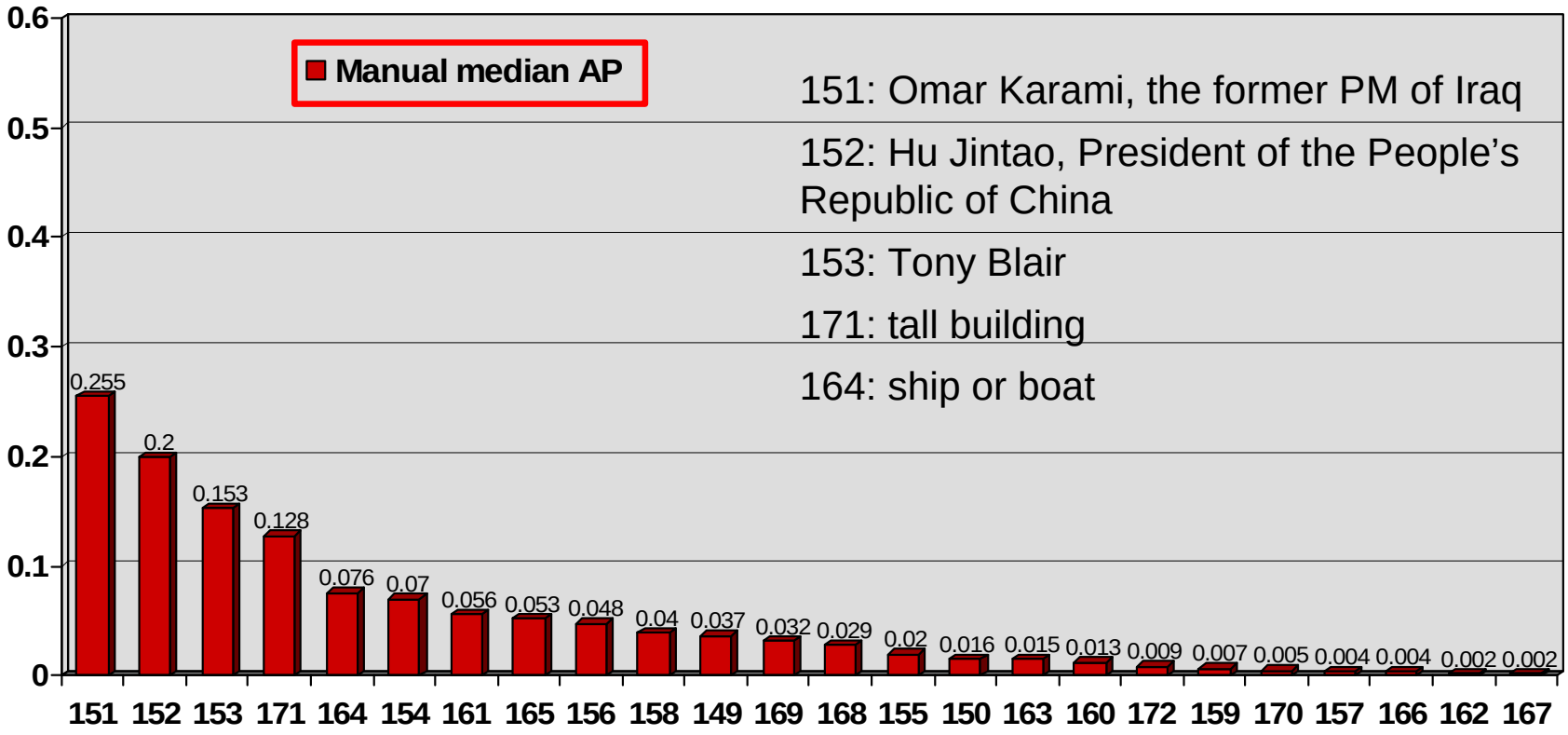
2005: Mean avg. precision by topic



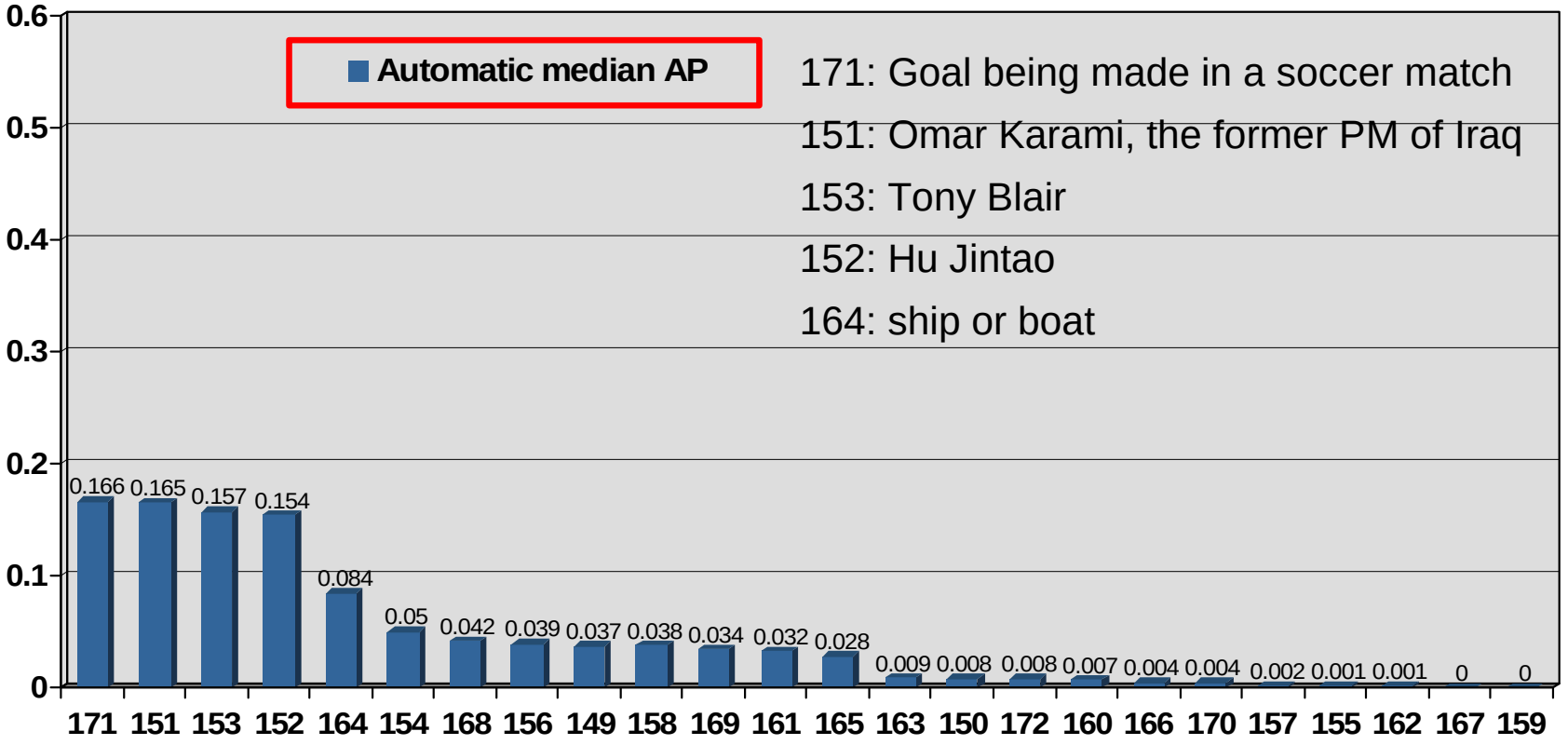
2005: Interactive runs' median average precision by topic



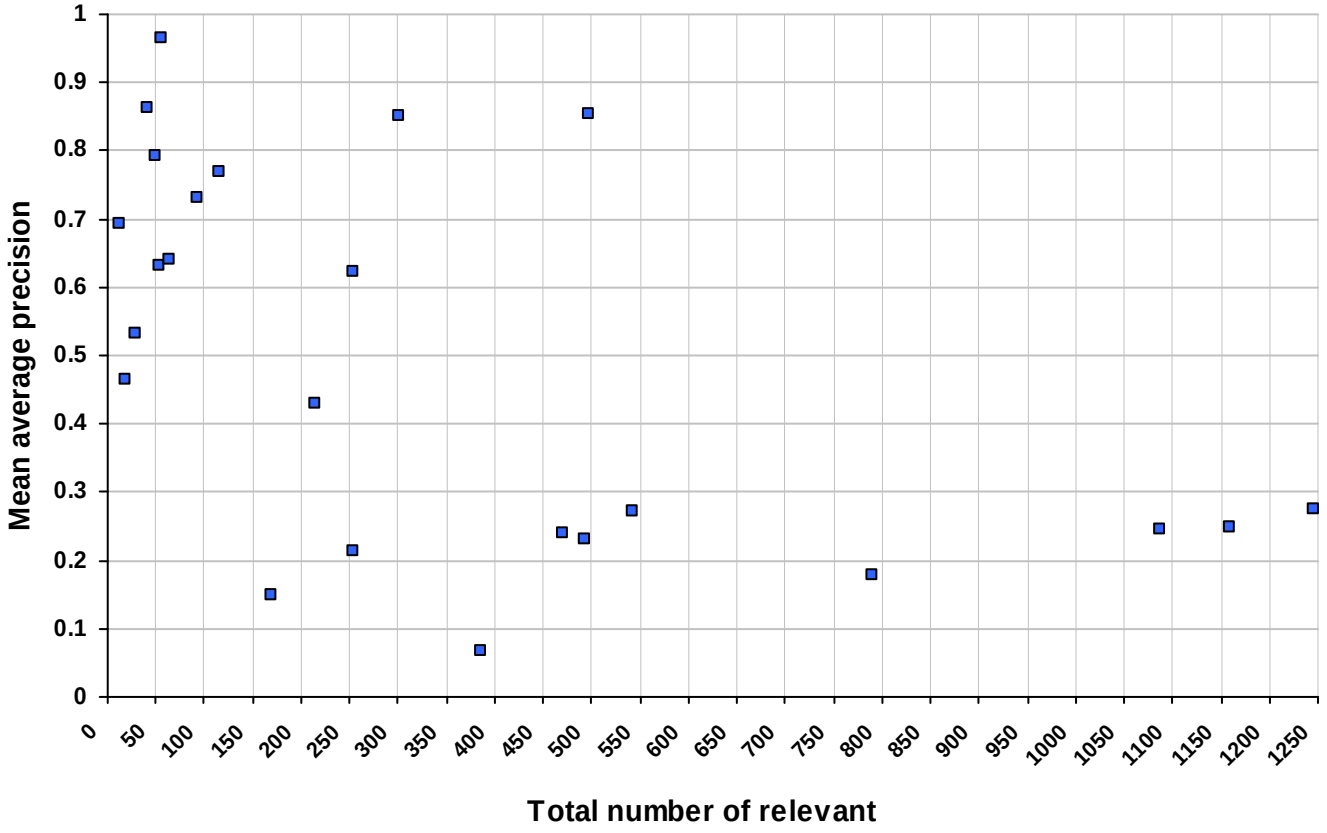
2005: Manual runs' median average precision by topic



2005: Automatic runs' median average precision by topic



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



Who did what ?

- Speaker slots to follow:
  - Carnegie Mellon University
  - IBM Research
  - MediaMill (University of Amsterdam and TNO)
  - National University of Singapore
  - University of Oulu/MediaTeam
- No papers from:
  - Bilkent University
  - QMUL
  - SCHEMA - University of Bremen
- Demos ?
- Posters ?

Interactive | Manual | Automatic

Columbia University

- Interactive search tool developed with
  - Text search, CBIR search, Story segmentation and story-level browsing, 39 visual concepts from LSCOM-Lite, near-duplicate detection, query-class dependent weights and cue-X re-ranking;
- Manual run with
  - Text, CBIR and visual concepts;
- Automatic runs with
  - Query-class dependent weights of some of the above;

Interactive | Manual | Automatic

Dublin City University

- Interactive search used a DiamondTouch collaborative tabletop interface from MERL, to text and image-based video searching;
- 2 versions
  - Increase user's awareness of other user thus forcing the collaboration;
  - More like "leave me alone" searching support for efficient solo searching;
- Aim was to explore user-user collaborative search;
- Findings are that group awareness benefits retrieval;
- Also did manual and automatic runs - exploring text-only vs. text+image searching;

Manual

Fudan University

- Submitted manual runs and explored multi-modal fusion;
- Found that relation expression fusion better than linear fusion using a variety of retrieval modalities;
  - Text;
  - 14 x visual concepts;
  - Pseudo relevance feedback;
  - Logistic regression
- Also explored training weights online vs. training weights offline

Interactive

FX Palo Alto Laboratory

- Participated in interactive search;
- Enhanced the 2004 system for efficient browsing and enhanced visualisation, by adding 29 concepts/semantic features;
- Story-level browsing, keyframe thumbnails, text dialogue overlays, story timelines;
- Query is text a/o image;
- Text-only search is as good as text+others (because the browser and visualisation is very strong ?);

Interactive | Manual | Automatic

Helsinki University of Technology

- Automatic, manual and interactive runs;
- Addressed text-only vs. text+mult-imodal querying;
- Multi-modal better than text-only !
- Interactive search used relevance feedback only with no “search” or shot browsing so very dynamic user control;

Interactive

Imperial College London

- Content-based search + NNk browsing in a 2D GUI map browser;
- Enhanced 2004 system with new kind of relevance feedback;
- Text-based search, content-based search with relevance feedback and temporal browsing integrated into a unified interface;
- Emphasis on supporting user task;

Automatic

Language Computer Corporation (LCC)

- Participated in automatic search;
- Used combinations of ASR text search (language modelling), image features, high-level features, alone and in combination;
- Image features used blobs;
- Text search alone was best-performing

Manual | Automatic

Lowlands (CWI, Twente, U. of Amsterdam)

- Manual and automatic search runs;
- Visual and text searching
- Weibull models and Gaussian mixture models for visual features, language modeling for text;
- No clear results differentiation;
- First steps towards developing parameterised search engines for each;

Automatic

Tsinghua University

- Three search modes - text, image match based on region matching, and concept matching in a concept;
- Concept/feature recognition approach based on their HLF submissions;
- Explore latent relationship (LSA) between (ASR) text and visual features and concepts;
- Tried each of these alone and in combinations using score fusion and query type-specific (2) weighting;
- Conclusion is that combinations work best;

Interactive

University of Central Florida

- UCF first time in search task;
- PEGASUS system, web-based, interactive, used ASR, OCR, keyframe global histograms and high level features;
- Submitted ASR-only & multi-modal runs;
- Multi-modal better than ASR-only;

Automatic

University of Iowa

- Automatic search runs;
- Text-only vs. text+image features;
  - Keyrame-keyframe pixel distances;
  - Text + colour information;
  - Text + texture information;
  - Text + edge information;
- Text-only was best - could have combined visual features ?

Interactive

University of North Carolina

- Investigate the effects of providing context and interactivity in a retrieval system, supporting the browsing of search result sets;
  - Basic Google-like video search
  - Enhanced with shot context browsing;
  - Further enhanced with interactive feedback, eg mouseover gives enlarged keyframes;
- For both performance and user perceptions, the Context+Interactive system was superior - higher recall, precision the same;

Observations

- 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;