TRECVID-2004: Search Task Overview

Alan Smeaton
Dublin City University
&
Paul Over, Joaquim Arlandis
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;
- Many thanks to Georges Quenot (CLIPS) for providing common shot bounds and keyframes, and Jean Luc Gauvain (LIMSI) for providing ASR, and Alex Hauptmann (CMU) for donating feature detections;

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 run corresponds to 1 user search per topic;

2004: Seach task participants (16, up from 11)

Bilkent University	Turkey	--	--	--	SE
Carnegie Mellon University	US	--	--	FE	SE
Center for Research & Technology Hellas/ITI	Greece	--	--	--	SE
CLIPS-LSR-LIS	France	SB	SS	FE	SE
CWI/Twente	the Netherlands	--	--	--	SE
Dublin City University	Ireland	--	--	--	SE
FX Palo Alto Laboratory	US	SB	--	--	SE
IBM Research	US	SB	SS	FE	SE
Imperial College, London	UK	SB	SS	FE	SE
Indiana University	US	--	--	--	SE
Mediamill/University of Amsterdam	the Netherlands	--	--	FE	SE
National Taiwan University	Taiwan	--	--	--	SE
National University of Singapore	Singapore	--	--	FE	SE
Queen Mary, University of London	UK	SB	--	--	SE
University of North Carolina at Chapel Hill	US	--	--	--	SE
University of Oulu	Finland	--	--	--	SE

24 Topics [total relevant found]

125. Find shots of a street scene with multiple pedestrians in motion and multiple vehicles in motion somewhere in the shot. [154]
126. Find shots of one or more buildings with flood waters around it/them. [118]
127. Find shots of one or more people and one or more dogs walking together. [64]
128. Find shots of US Congressman Henry Hyde's face, whole or part, from any angle. [115]
129. Find shots zooming in on the US Capitol dome. [16]
130. Find shots of a hockey rink with at least one of the nets fully visible from some point of view. [162]
131. Find shots of fingers striking the keys on a keyboard which is at least partially visible. [86]
132. Find shots of people moving a stretcher. [41]
133. Find shots of Saddam Hussein. [46]
134. Find shots of Boris Yeltsin. [22]

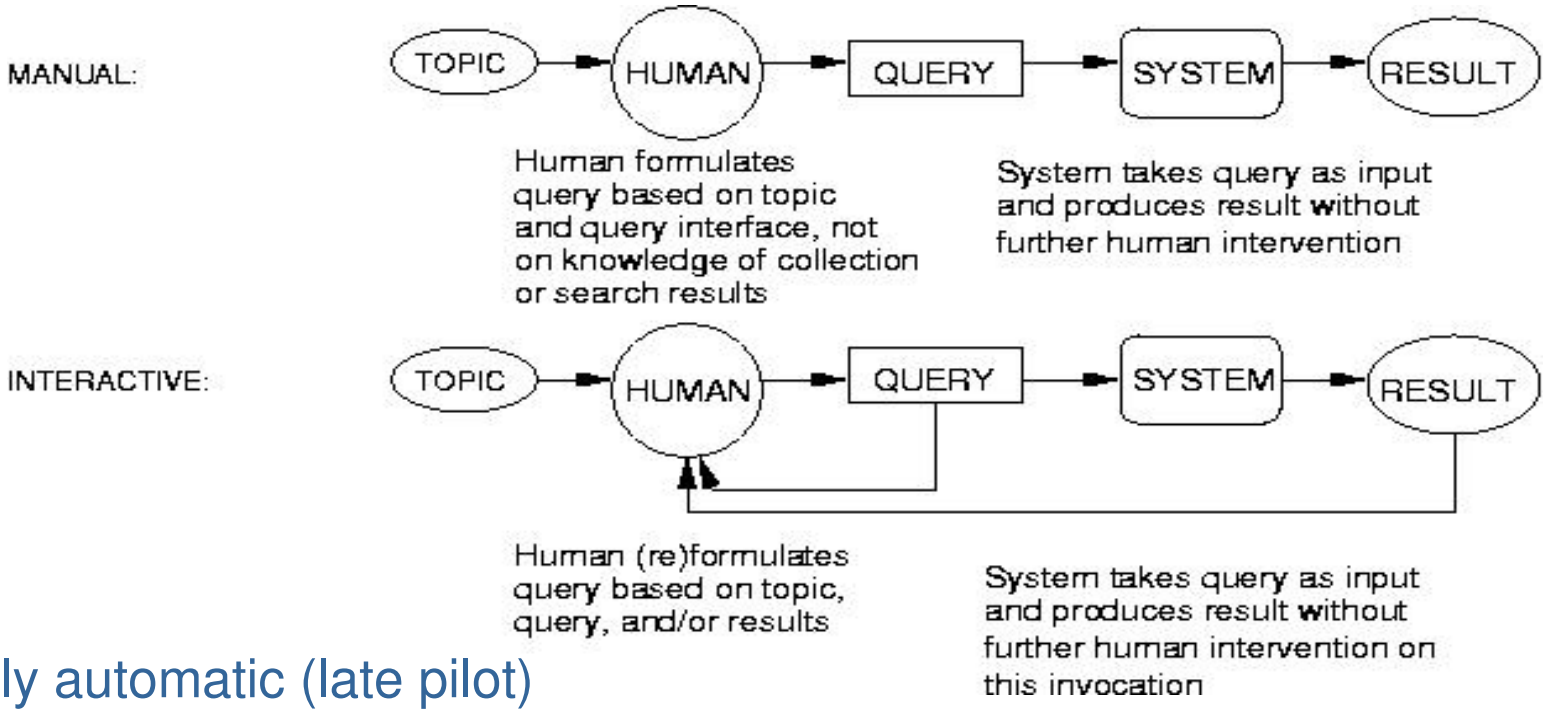
24 Topics [total relevant found]

135. Find shots of a person hitting a golf ball that then goes into the hole. [54]
136. Find shots of Benjamin Netanyahu. [19]
137. Find shots of one or people going up or down some visible steps or stairs. [106]
138. Find shots of a handheld weapon firing. [97]
139. Find shots of one or more bicycles rolling along. [55]
140. Find shots of one or more umbrellas. [69]
141. Find shots of Sam Donaldson's face - whole or part, from any angle, but including both eyes. No other people visible with him. [54]
142. Find more shots of a tennis player contacting the ball with his or her tennis racket. [41]
143. Find shots of one or more wheelchairs. They may be motorized or not. [39]
144. Find shots of Bill Clinton speaking with at least part of a US flag visible behind him. [96]

24 Topics [total relevant found]

145. Find shots of one or more horses in motion. [67]
146. Find shots of one or more skiers skiing a slalom course with at least one gate pole visible. [0]
147. Find shots of one or more buildings on fire, with flames and smoke visible. [85]
148. Find shots of one or more signs or banners carried by people at a march or protest. [194]

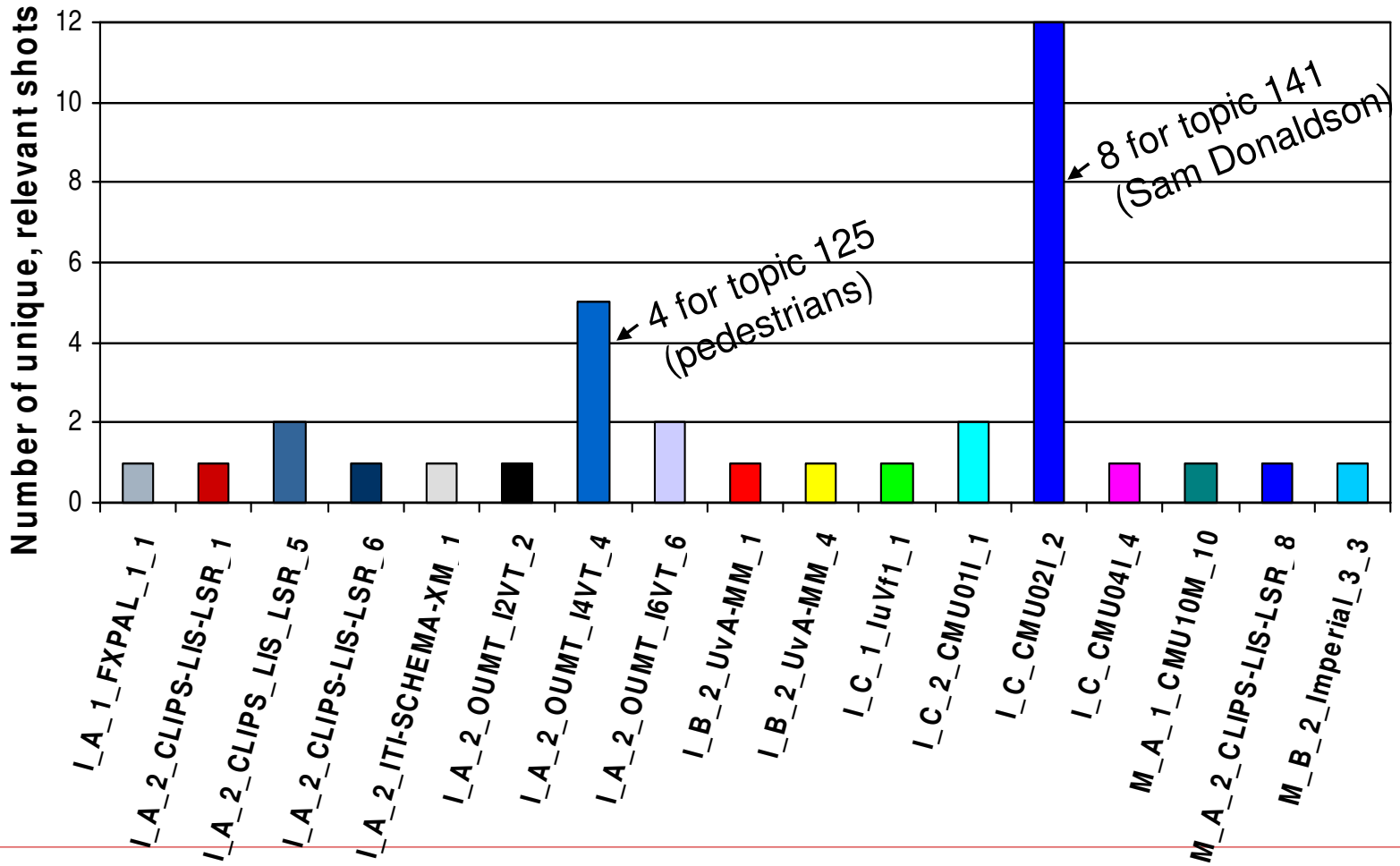
Search Types: Interactive and Manual



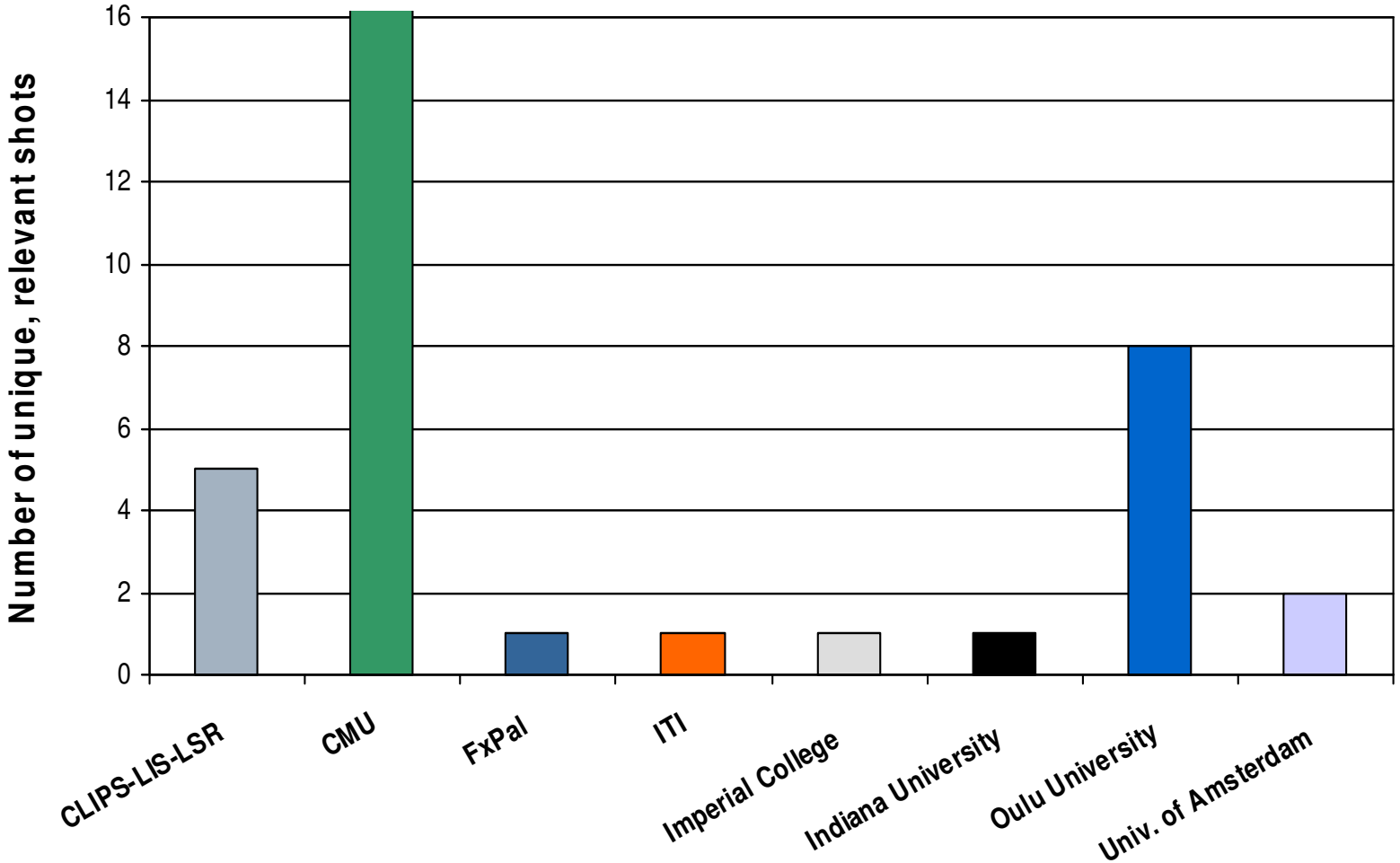
+ fully automatic (late pilot)

Number of runs: 62 interactive (up from 37)
52 manual (up from 38)
23 automatic

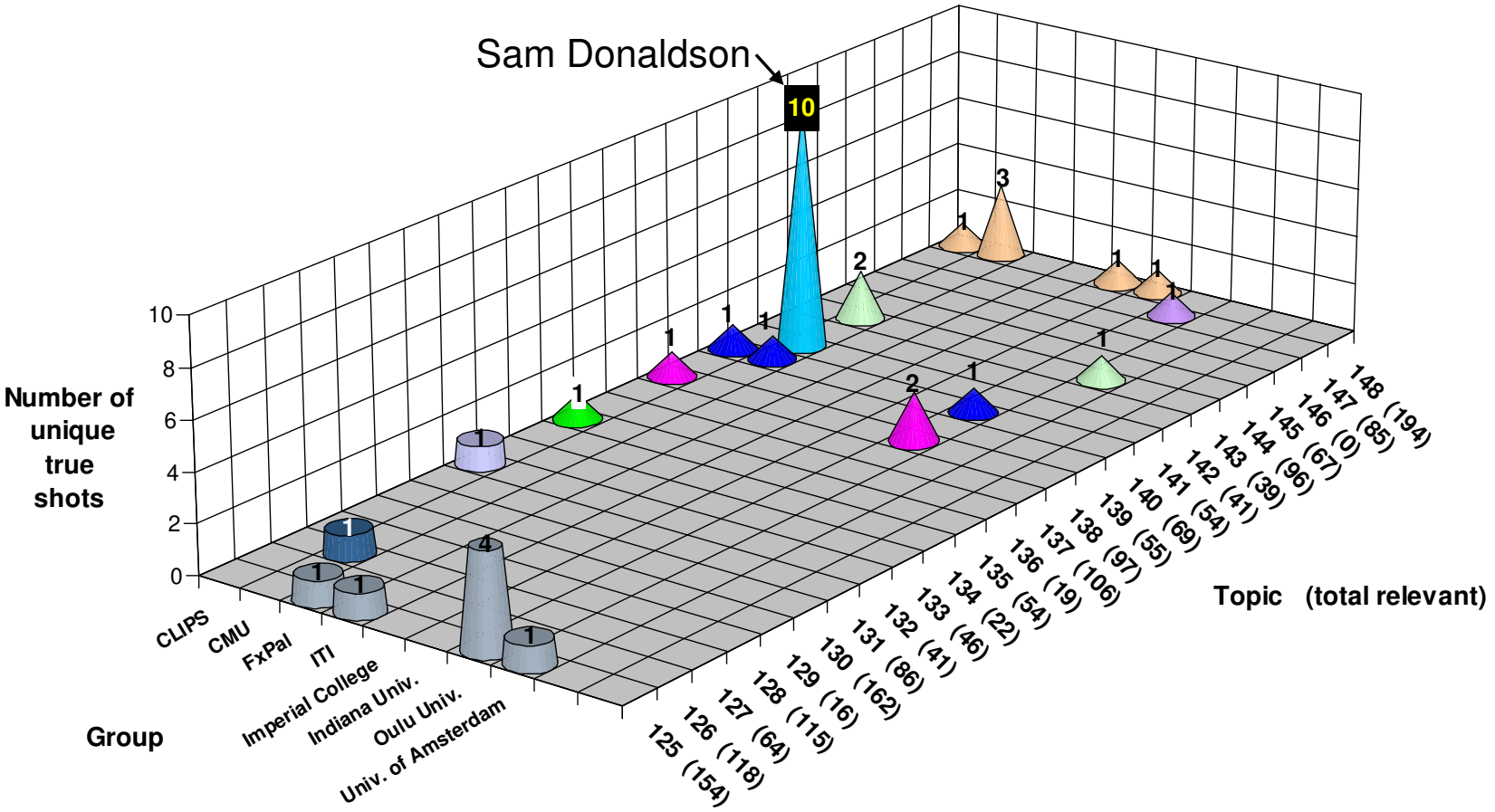
2004: 15 runs contributed one or more unique, relevant shots



2004: 8 sites contributed one or more unique, relevant shots

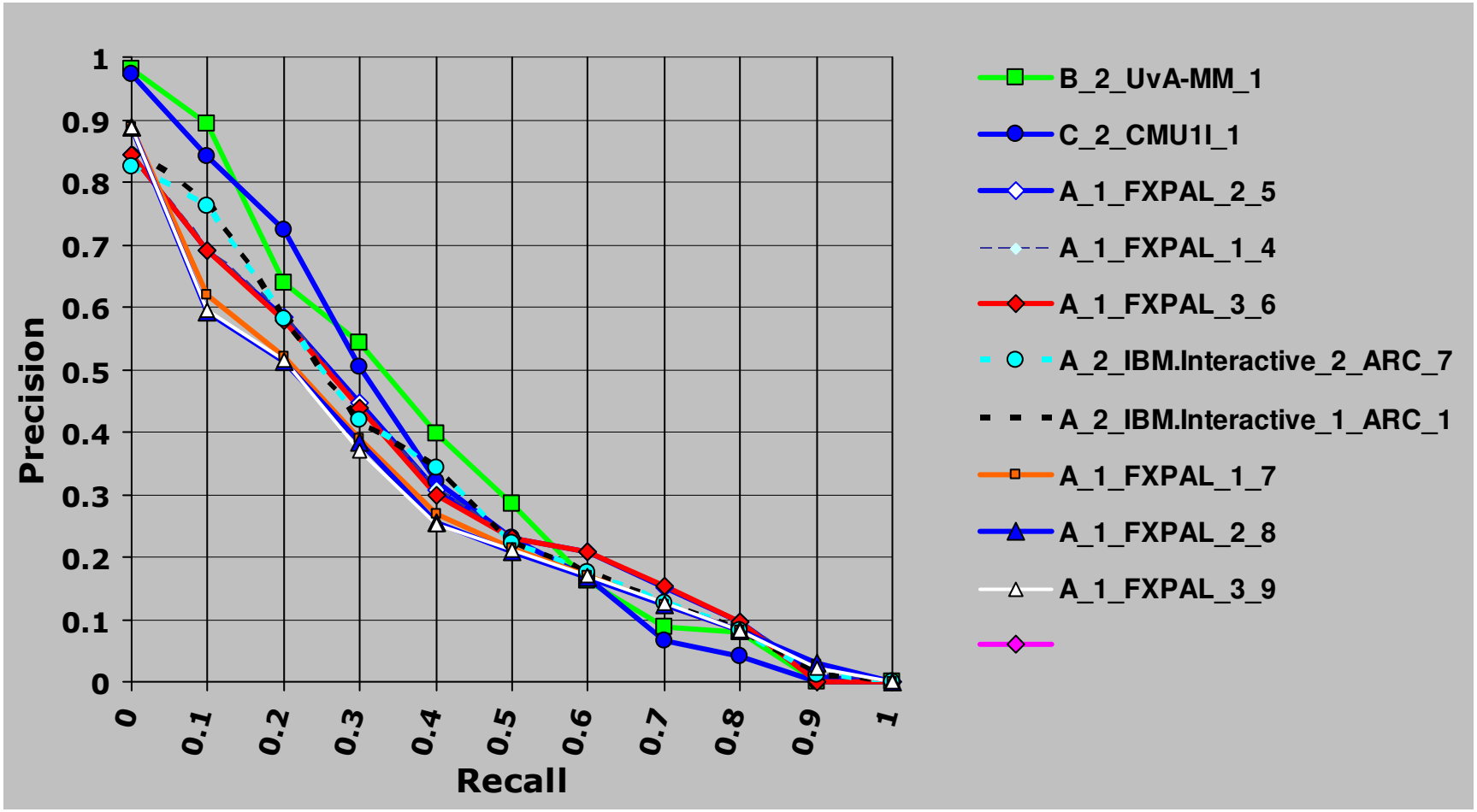


2004: Relevant shots contributed uniquely for a feature by group



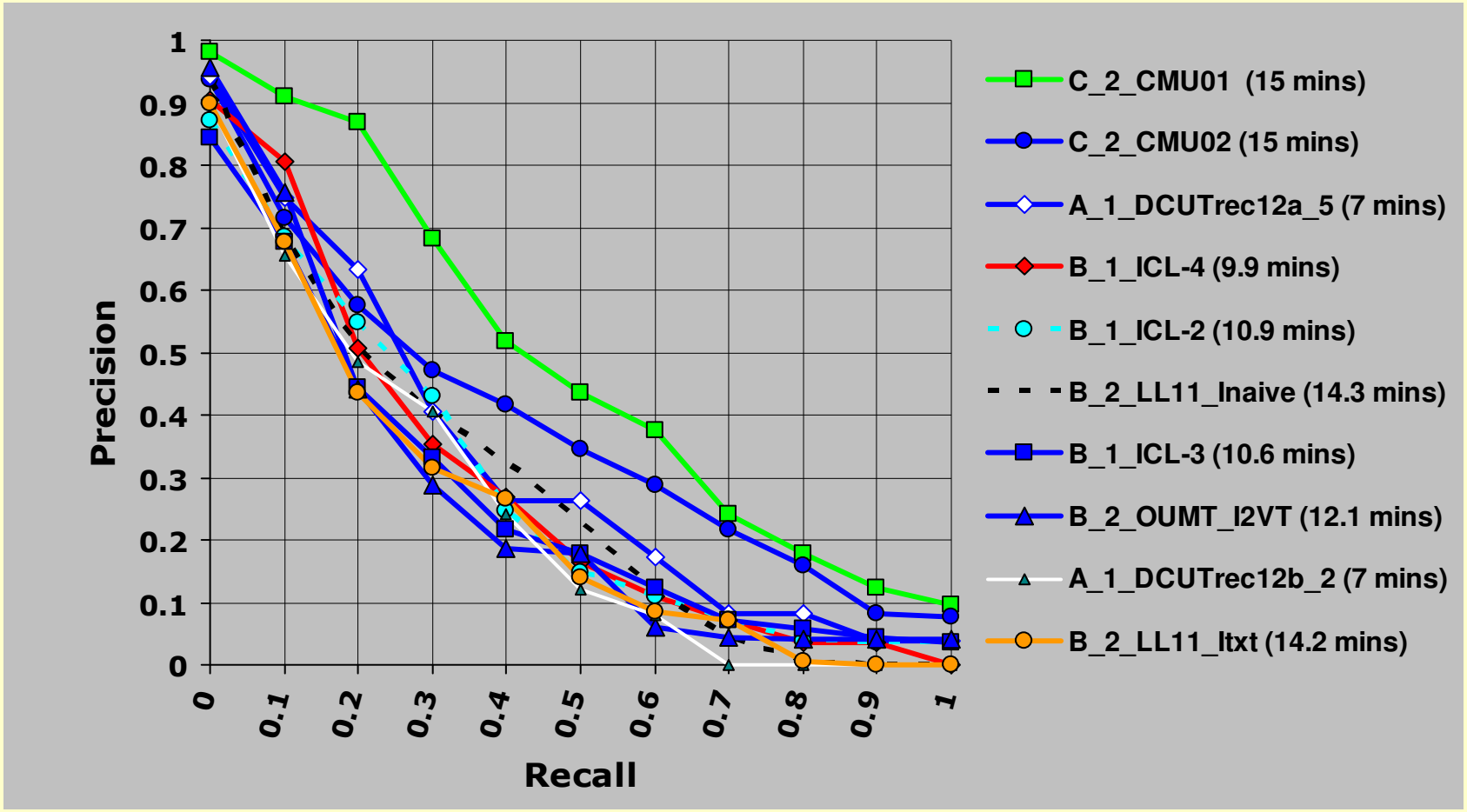
2004: Interactive runs - top 10 MAP (of 62)

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



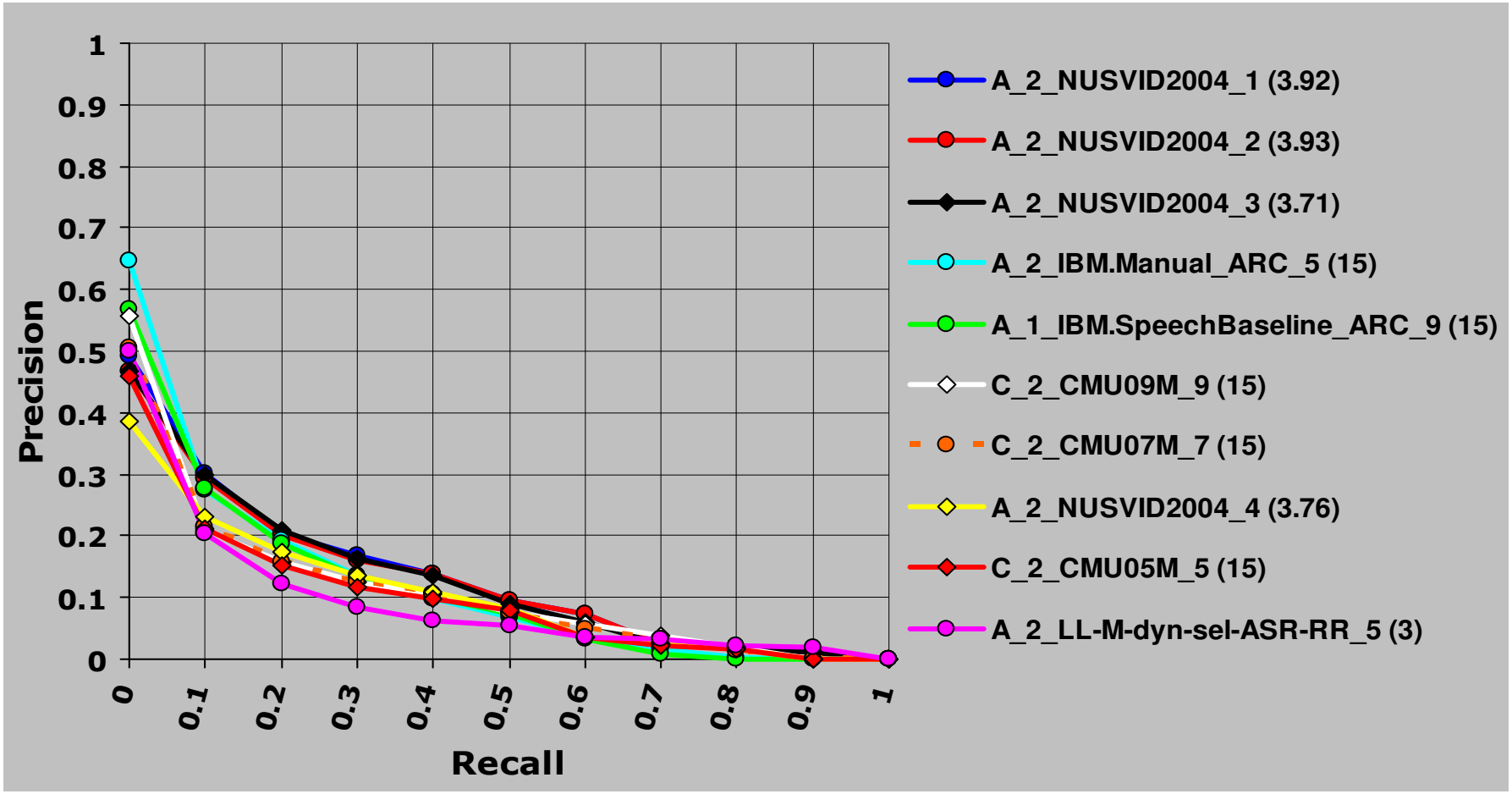
2003: Interactive runs - top 10 MAP (of 36)

(with mean elapsed time)



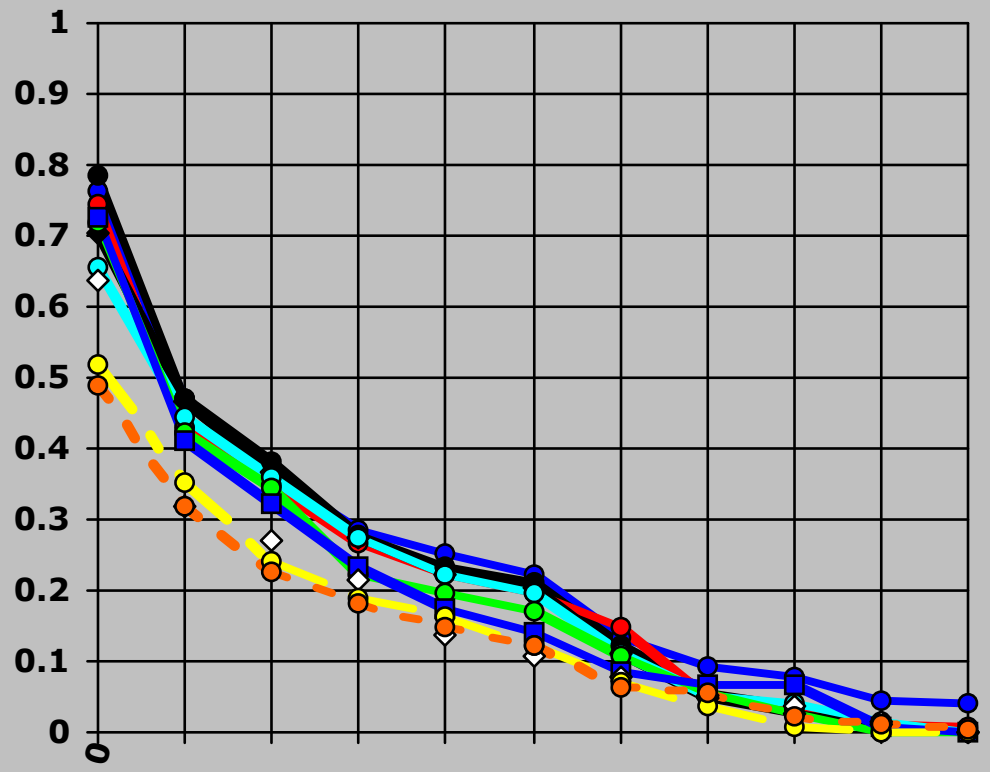
2004: Manual runs - top 10 MAP (of 52)

(mean human effort (mins) / topic)



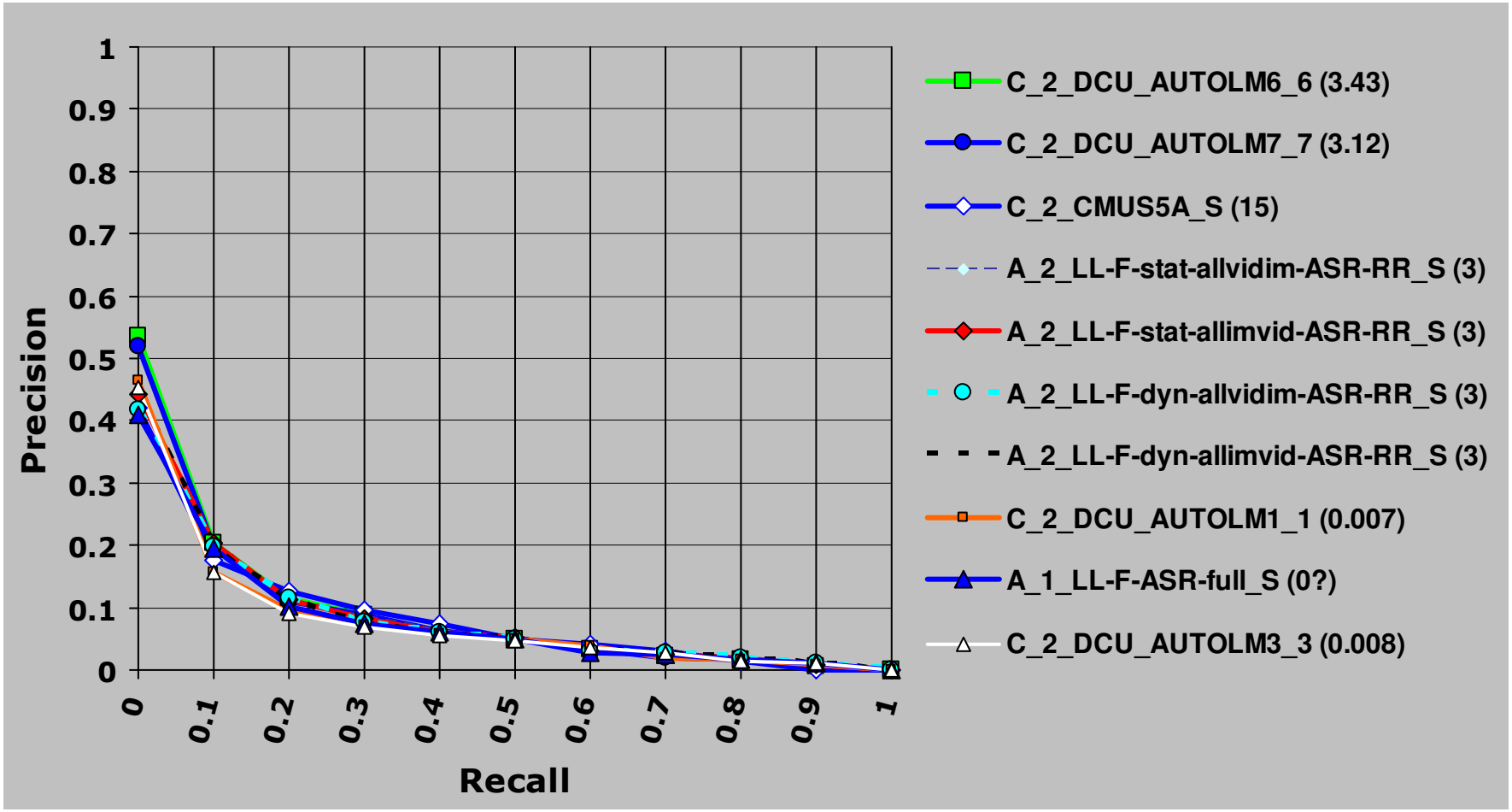
2003: Manual runs - top 10 MAP (of 38)

(with mean human effort / topic)

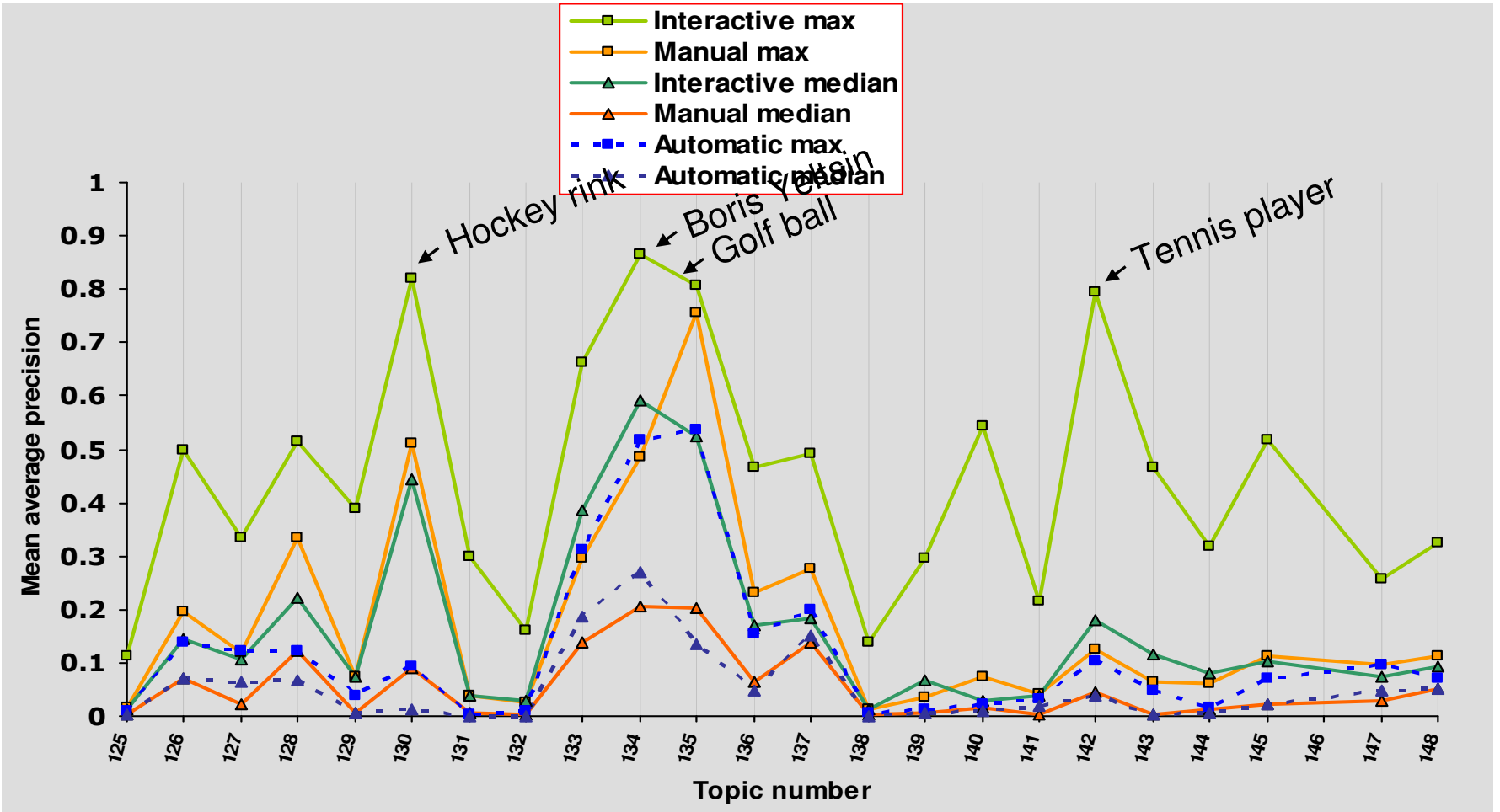


2004: Automatic runs (pilot) - top 10 MAP (of 23)

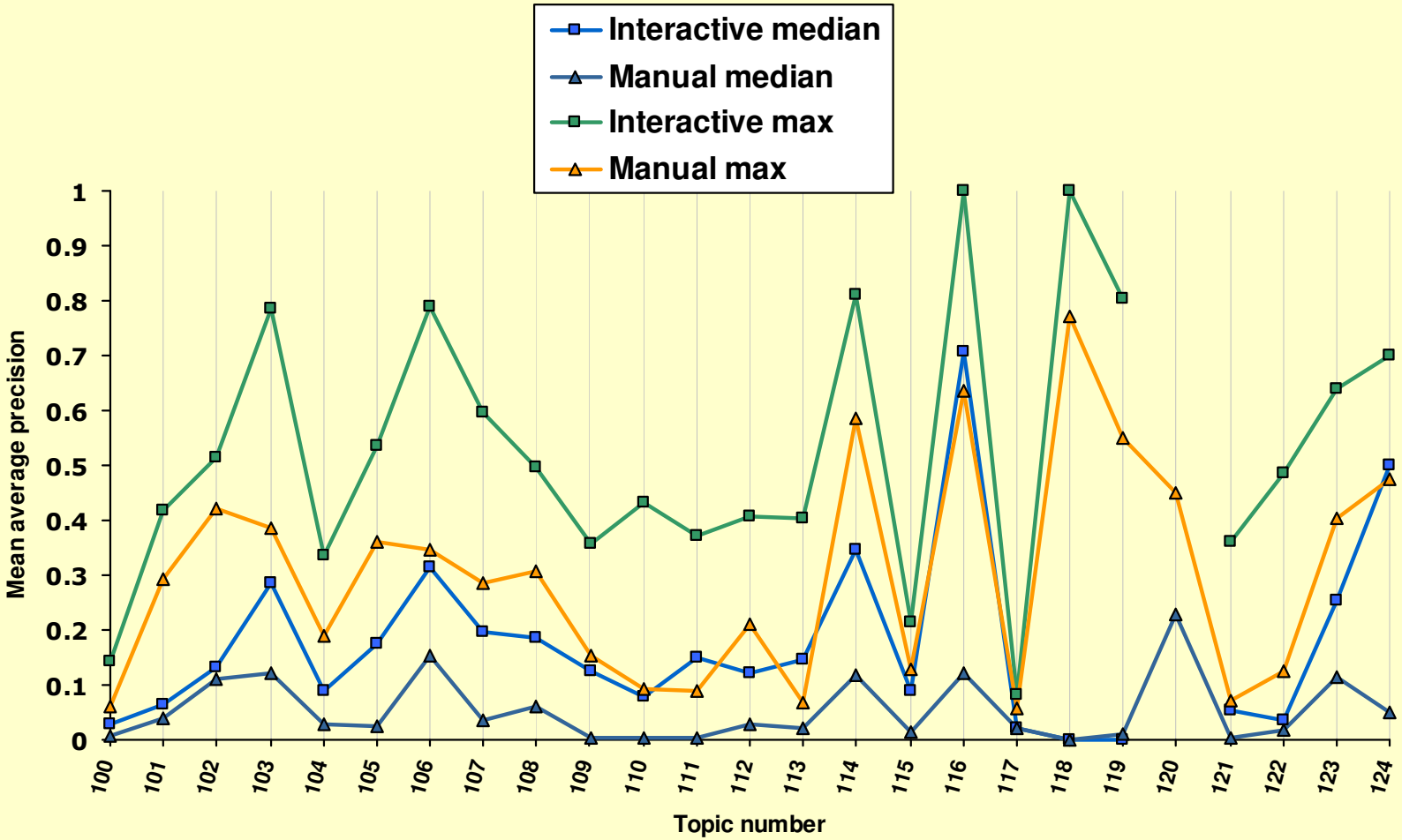
(mean elapsed time (mins) / topic)



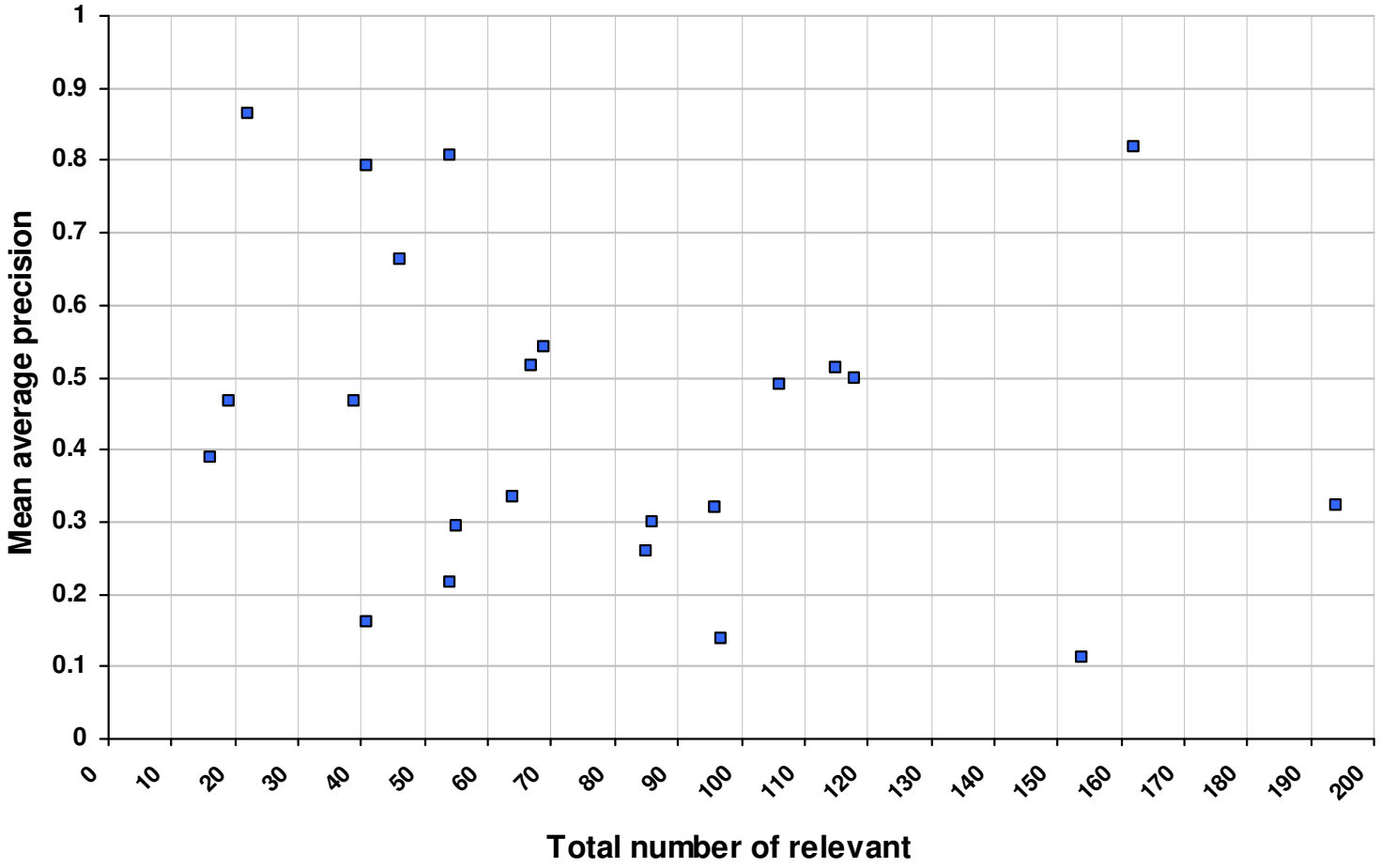
2004: Mean avg. precision by topic



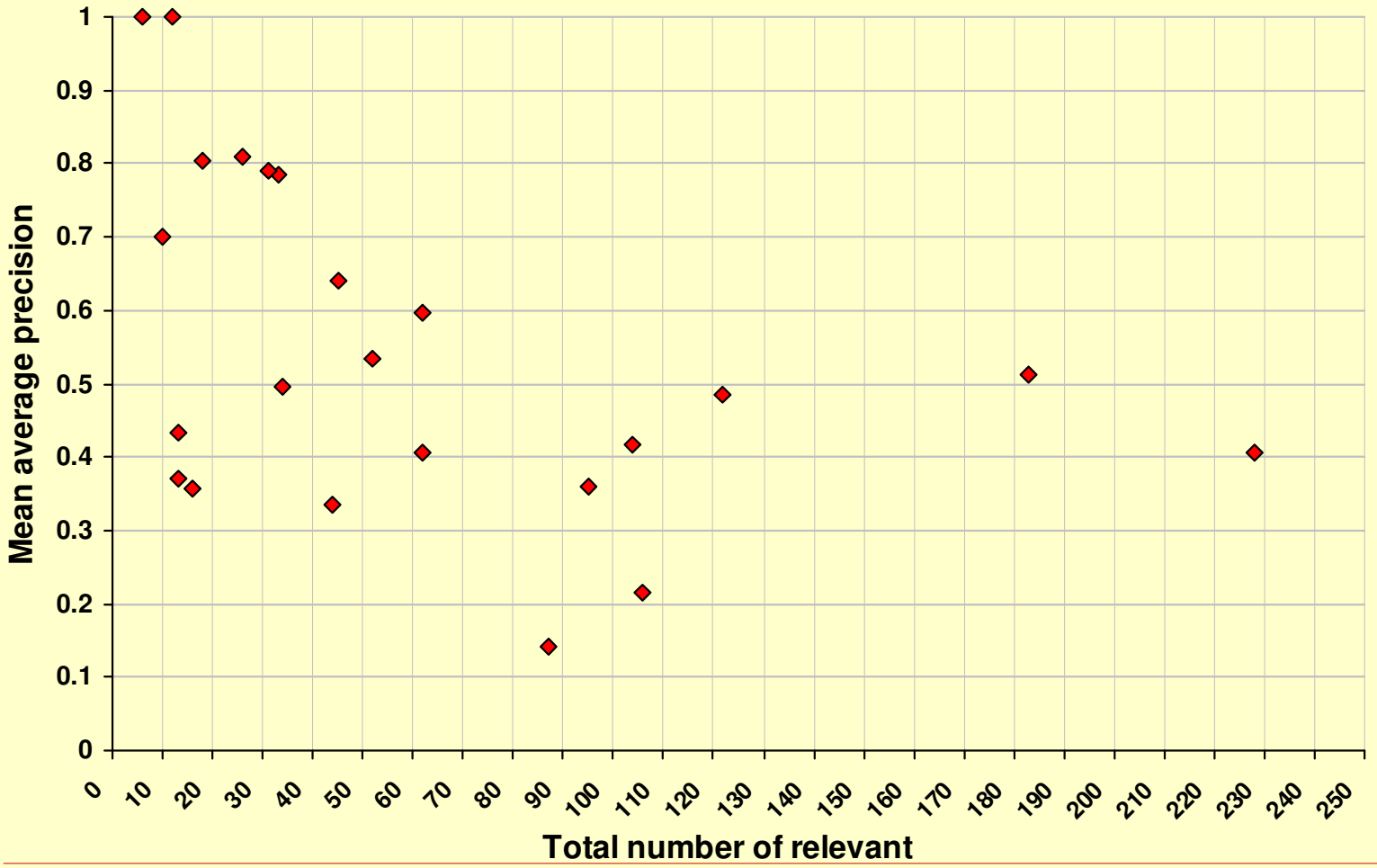
2003: Mean avg. precision by topic



2004: Mean average precision (interactive max) vs total number relevant

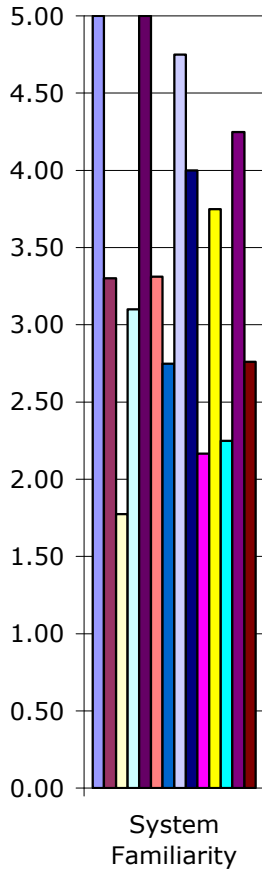


2003: Mean average precision (interactive max) vs total number relevant



Interactive Search: Who were the users ?

Who are the users

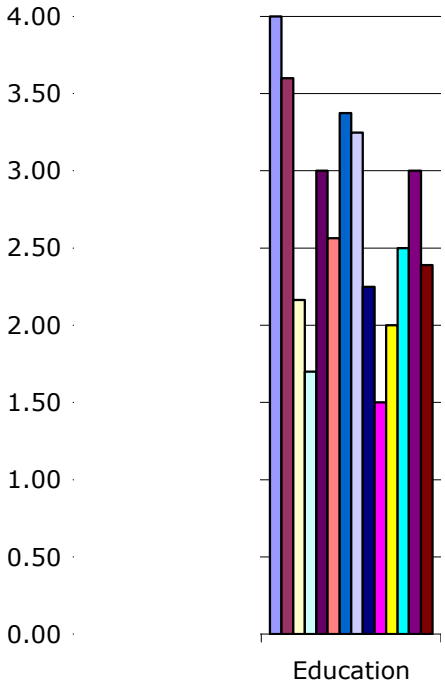


How familiar are you with the system(s) before your participation in this experiment ?

Questions

Interactive Search: Who were the users ?

Who are the users



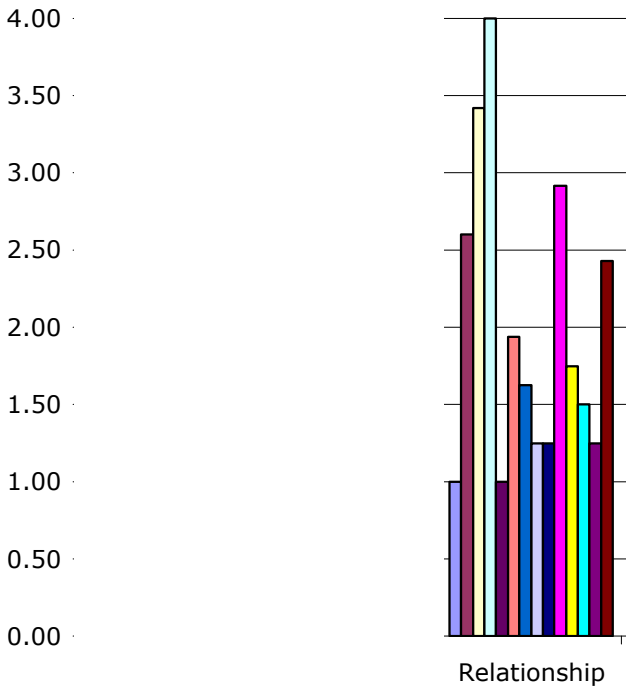
What is your educational level

1. Undergraduate
2. Graduate student
3. Researcher/advanced degree
4. Faculty/research staff

Questions

Interactive Search: Who were the users ?

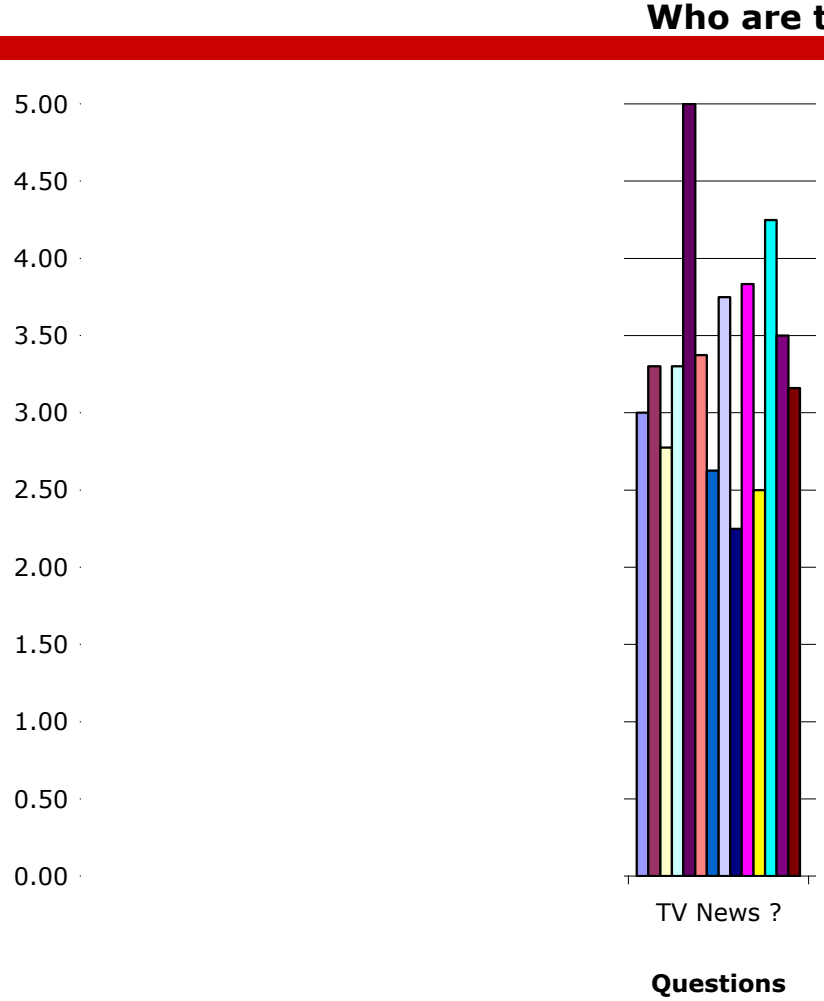
Who are the users



Relationship to the TRECVid group

1. Directly involved in development
2. Part of group
3. Not part of group but in department
4. No connection

Interactive Search: Who were the users ?



How often do watch TV news ?

1 - not at all

…

5 - very much

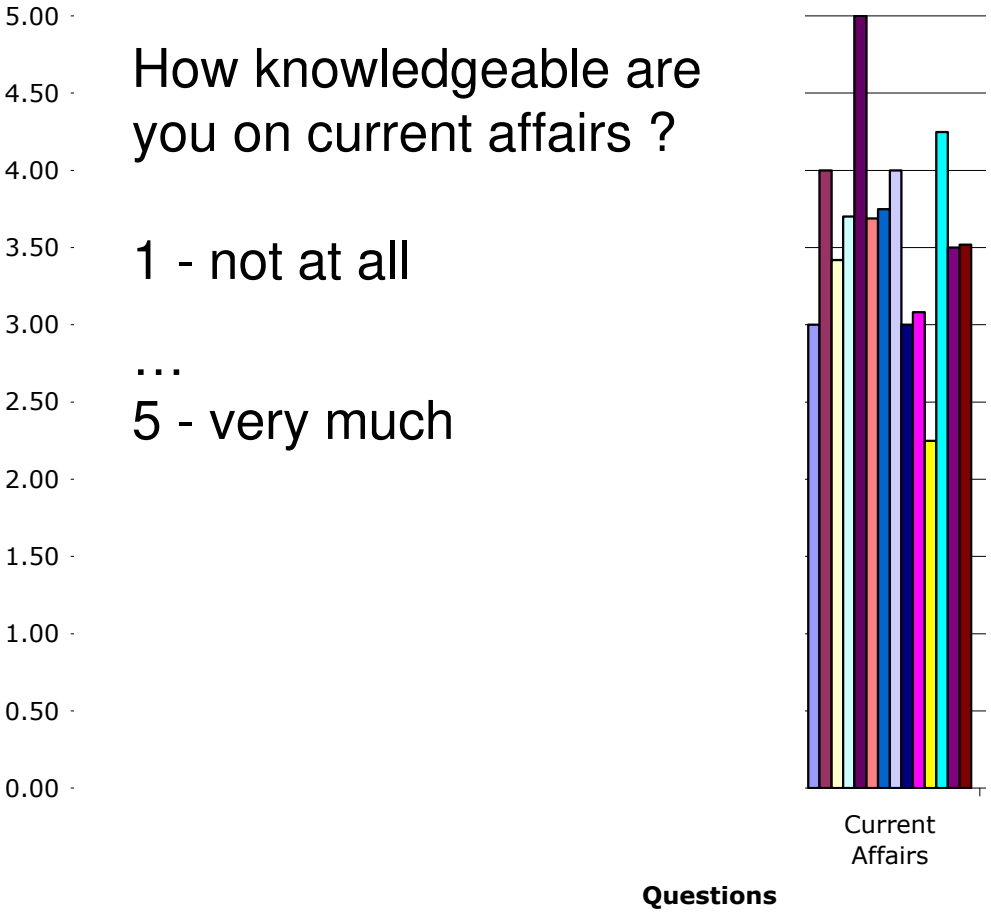
Interactive Search: Who were the users ?

How knowledgeable are you on current affairs ?

1 - not at all

…

5 - very much



Interactive Search: Who were the users ?

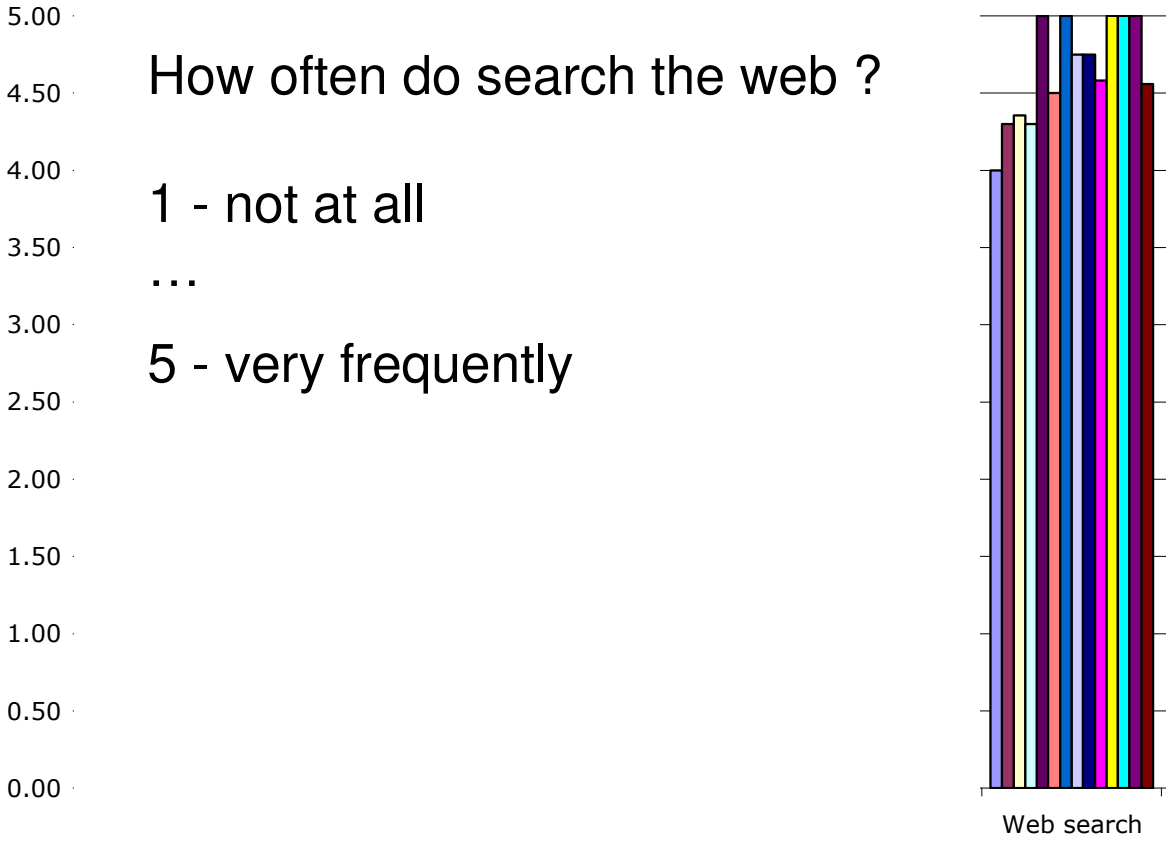
Who are the users

How often do search the web ?

1 - not at all

…

5 - very frequently



Interactive Search: Who were the users ?

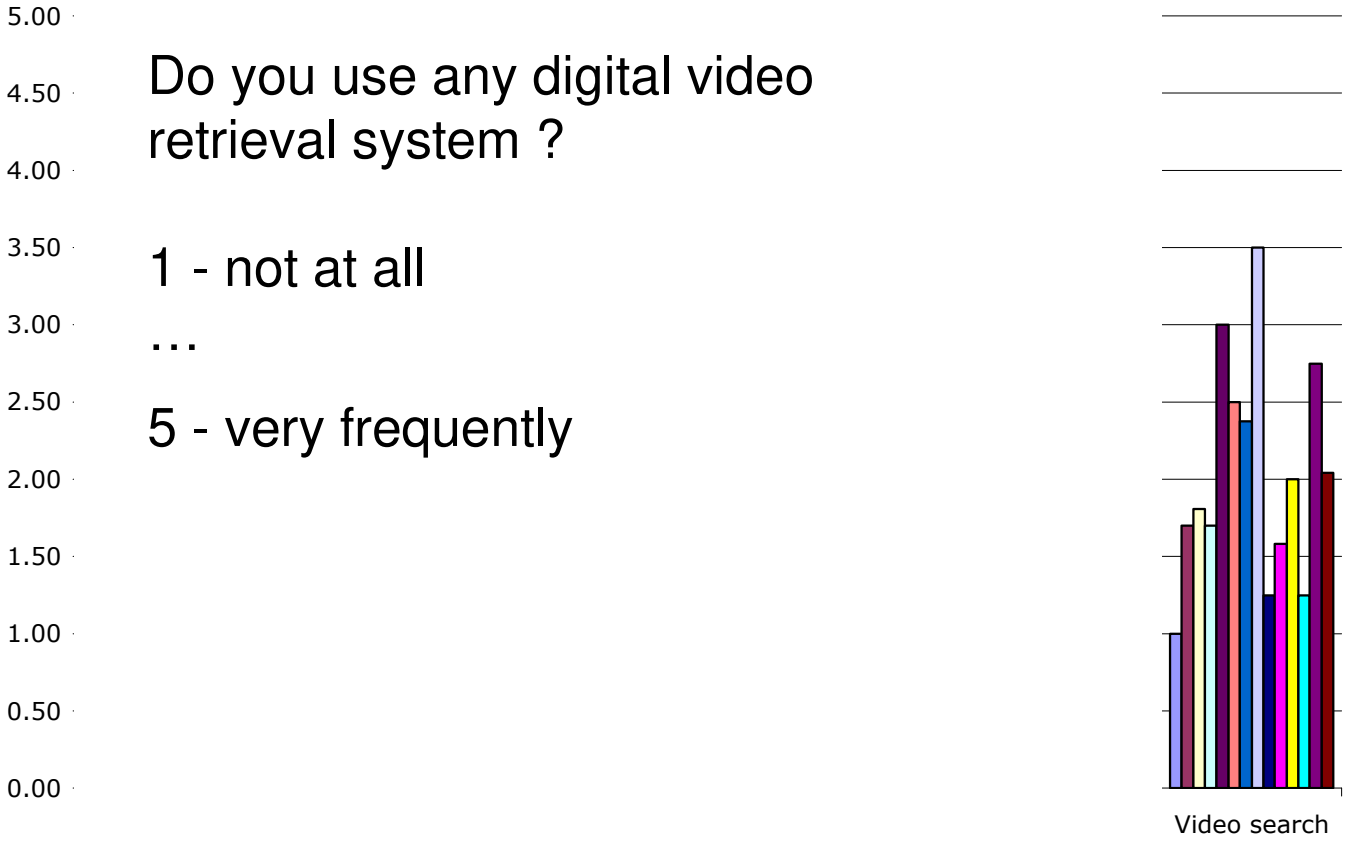
Who are the users

Do you use any digital video retrieval system ?

1 - not at all

…

5 - very frequently



Interactive Search: Who were the users ?

What did the users think ?

- We also have analysis of what users thought after each search (i.e. after each of the 1165 topic searches), whether it was difficult, whether example images were useful, etc.
- We can present this per topic / per site;

Bilkent University

- First time participant;
- Keyword based ASR and colour based keyframe image matching;
- Used CMU donated features;
- Also did face detection for later inclusion;

Carnegie Mellon University

- Speaker slot to follow

ITI/SCHEMA

- Collaboration among ITI Thessaloniki, Munich University of Technology, and Dublin City University;
- Speaker slot to follow

CLIPS-LSR-LIS

- User-controlled combination of 5 modalities:
  - Keyword search on ASR;
  - Visual similarity to example images based on colour histograms;
  - Use of 15 semantic categories (features);
  - Visual similarity to positive images (relevance feedback);
  - Temporal closeness to positive images;
- Manual and interactive runs with 4 users;

CWI/Twente

- Speaker slot to follow

Dublin City University

- Interactive search Text+Image vs. Text alone, finding Text+Image better;
- Manual search using face filtering - didn't help;
- Automatic runs based on language models over ASR and visual features (interpolated LMs better than other LMs)
- Multiple sources of text evidence (ASR, CC, OCR) and query expansion;
- More people on paper authorlist than IBM (16 vs. 15) !

FX Palo Alto Laboratory

- Speaker slot to follow

IBM Research

- Speaker slot to follow

Imperial College London

- Explored content based search + NN^k browsing vs. NN^k browsing alone;
- Integrated relevance feedback, temporal browsing, historical browsing, content based search into a tight user-oriented interface;
- Re-design on 2003 GUI image browser;
- Strong emphasis on supporting user task;

Indiana University

- Manual baseline run searching ASR output;
  - One user conducting all topics with five minute time limit
- One Interactive run searching ASR output;
  - One user conducting all topics with ten minute time limit and each topic beginning with query generated during manual run
- ASR search supported by tf•idf term weighting;
- Building upon previous years' ViewFinder image browsing interface;

MediaMill/University of Amsterdam

- Used a set of 32 semantic concepts - close link to their feature extraction task submissions;
- Global colour histograms for image querying also;
- Queries close to their 32 concepts worked well, e.g. *ice hockey, bicycle, Clinton*;
- Expert users, obtained best MAP and best performance in 7 topics;
- Success is based on the topics in their lexicon;

National Taiwan University

- Manual search runs only;
- Aligning ASR tokens and high level features (concepts) using WordNet for word-word distances based on distance to common ancestor in WN;
- Targeting the time delay in ASR vs image;

National University of Singapore

- Speaker slot to follow

Queen Mary University of London

- No details at time of writing;

University of North Carolina

- Used 3 different interactive search systems:
  1. Text only, using ASR and MySQL full text search;
  2. Features based using 10 IBM donated features;
  3. Semantic clusters of stories, used U.Iowa story segmentation and LSI on the ASR yielding semantic clusters;
- Extensive user questionnaire and user study;

University of Oulu

- Manual and interactive search experiments;
- Manual investigated combinations of different search engines on visual (low level image features), concept (feature) and text (ASR and CC);
- Interactive used ASR only vs. ASR+visual and found them equivalent;
- Improvements to the TV2003 browser with extra support for user tasks;

Observations

- At the end of the 2-year cycle …
- Last year I said "Lots of variation, interesting shot browsing interfaces, mixture of interactive & manual";
- Even moreso this year, plus we have automatic searching;
- Nearly 20 groups can now complete the search task and the demos are impressive;