KNOWN-ITEM SEARCH

Alan Smeaton
Dublin City University

Paul Over
NIST

Task

Use case: You've seen a specific given video and want to find it again but don't know how to go directly to it. You remember some things about it.

System task:

- Given a test collection of short videos and a topic with:
  - some words and/or phrases describing the target video
  - a list of words and/or phrases indicating people, places, or things visible in the target video
- Automatically return a list of up to 100 video IDs ranked according to the likelihood that the video is the target one,

  OR
- Interactively return a single video ID believed to be the target
  - Interactive runs could ask a web-based oracle if a video X is the target for topic Y. Simulates real user's ability to recognize the known-item. All oracle calls were logged.

Data

~ 200 hrs of Internet Archive available with a Creative Commons license

- ~8000 files
- Durations from 10s – 3.5 mins.
- Metadata available for most files (title, keywords, description, …)

122 sample topics created like the test topics – for development

300 test topics created by NIST assessors, who …

- Looked at a test video and tried to describe something unique about it
- Identified from the description some people, places, things, events visible in the video
- No video examples, no image examples, no audio; just a few words, phrases

Example topics

0001 KEY VISUAL CUES: man, clutter, headphone

QUERY: **Find the video of bald, shirtless man showing pictures of his home full of clutter and wearing headphone**

0002 KEY VISUAL CUES: Sega advertisement, tanks, walking weapons, Hounds

QUERY: **Find the video of an Sega video game advertisement that shows tanks and futuristic walking weapons called Hounds.**

0003 KEY VISUAL CUES: Two girls, pink T shirt, blue T shirt, swirling lights background

QUERY: **Find the video of one girl in a pink T shirt and another in a blue T shirt doing an Easter skit with swirling lights in the background.**

0004 KEY VISUAL CUES: George W. Bush, man, kitchen table, glasses, Canada

QUERY**: Find the video about the cost of drugs, featuring a man in glasses at a kitchen table, a video of Bush, and a sign saying Canada.**

0005 KEY VISUAL CUES: village, thatch huts, girls in white shirts, woman in red shorts, man with black hair

QUERY: **Find the video of a Asian family visiting a village of thatch roof huts showing two girls with white shirts and a woman in red shorts entering several huts with a man with black hair doing the commentary.**

TV2010 Finishers

---	***	KIS	***	---	SIN	Aalto University School of Science and Technology
CCD	INS	KIS	---	SED	SIN	Beijing University of Posts and Telecom.-MCPRL
---	***	KIS	MED	SED	SIN	Carnegie Mellon University - INF
***	***	KIS	---	---	***	Chinese Academy of Sciences - MCG
CCD	---	KIS	---	***	SIN	City University of Hong Kong
---	INS	KIS	---	---	---	Dublin City University
***	INS	KIS	---	---	***	Hungarian Academy of Sciences
---	INS	KIS	MED	---	SIN	Informatics and Telematics Inst.
---	---	KIS	---	---	---	Institute for Infocomm Research
---	INS	KIS	MED	***	SIN	KB Video Retrieval
---	***	KIS	***	***	---	National University of Singapore
---	---	KIS	---	---	SIN	NTT Communication Science Laboratories-UT
---	INS	KIS	***	***	SIN	University of Amsterdam
***	***	KIS	---	---	---	University of Klagenfurt
***	---	KIS	***	***	***	York University

Interactive runs

** : group applied but didn't submit

-- : group didn't apply for the task

TV2010 Run conditions

Training type (TT):

- A used only IACC training data
- B used only non-IACC training data
- C used both IACC and non-IACC TRECVID (S&V and/or Broadcast news) training data
- D used both IACC and non-IACC non-TRECVID training data

Condition (C):

- NO the run DID NOT use info (including the file name) from the IACC.1 *_meta.xml files
- YES the run DID use info (including the file name) from the IACC.1 *_meta.xml files

Evaluation

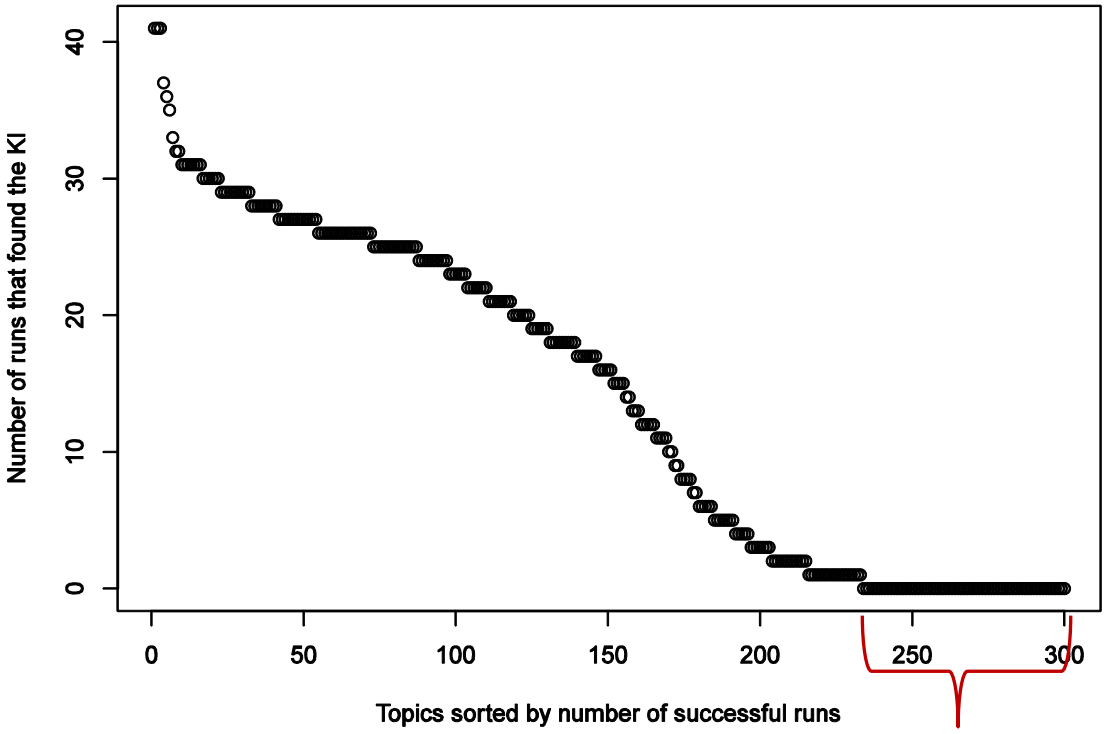
Three measures for each run (across all topics):

- mean inverted rank of KI found (0 if not found)
  - for interactive (1 result per topic) == fraction of topics for which KI found
- mean elapsed time (mins.)
- user satisfaction (interactive) (1-7(best))

Calculated automatically using the ground truth created with the topics

Results – topic variability

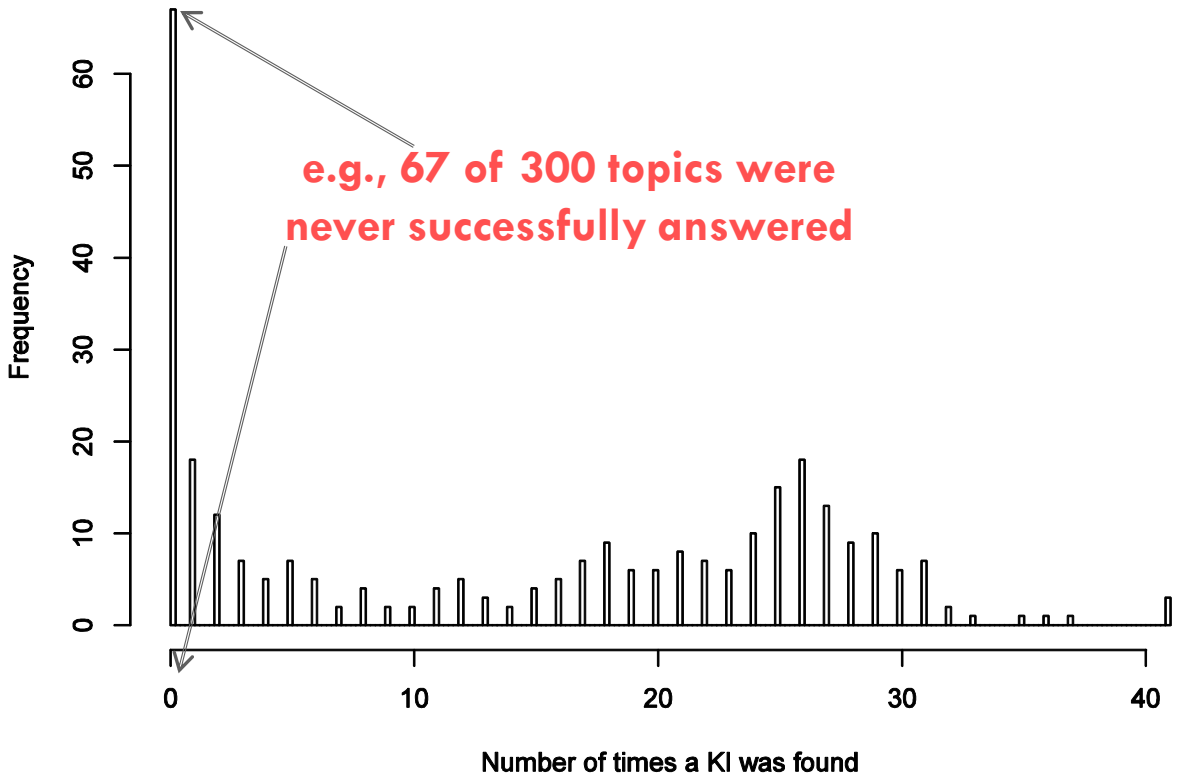
Topics sorted by number of runs that found the KI



e.g., 67 of 300 topics were never successfully answered

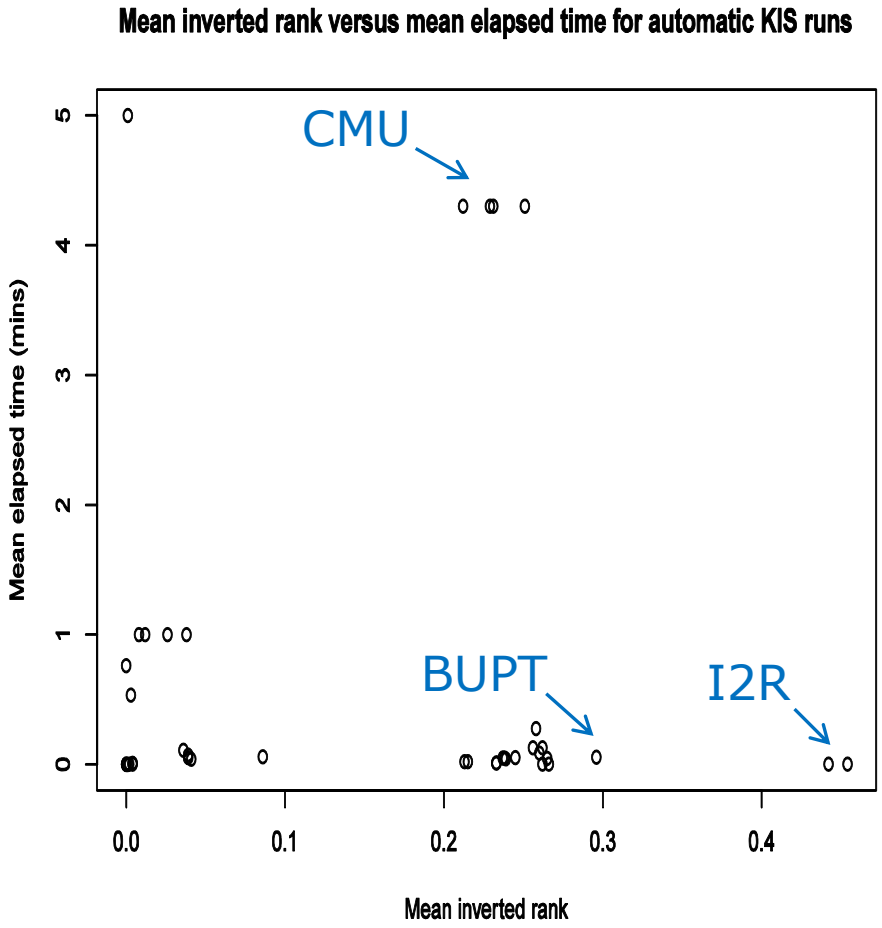
Results – topic variability

Histogram of "KI found" frequencies



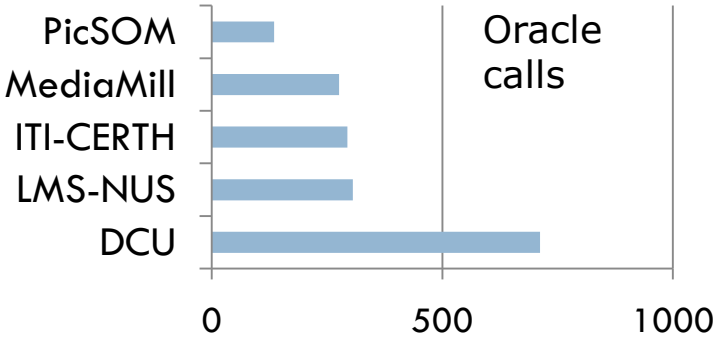
Results – automatic runs

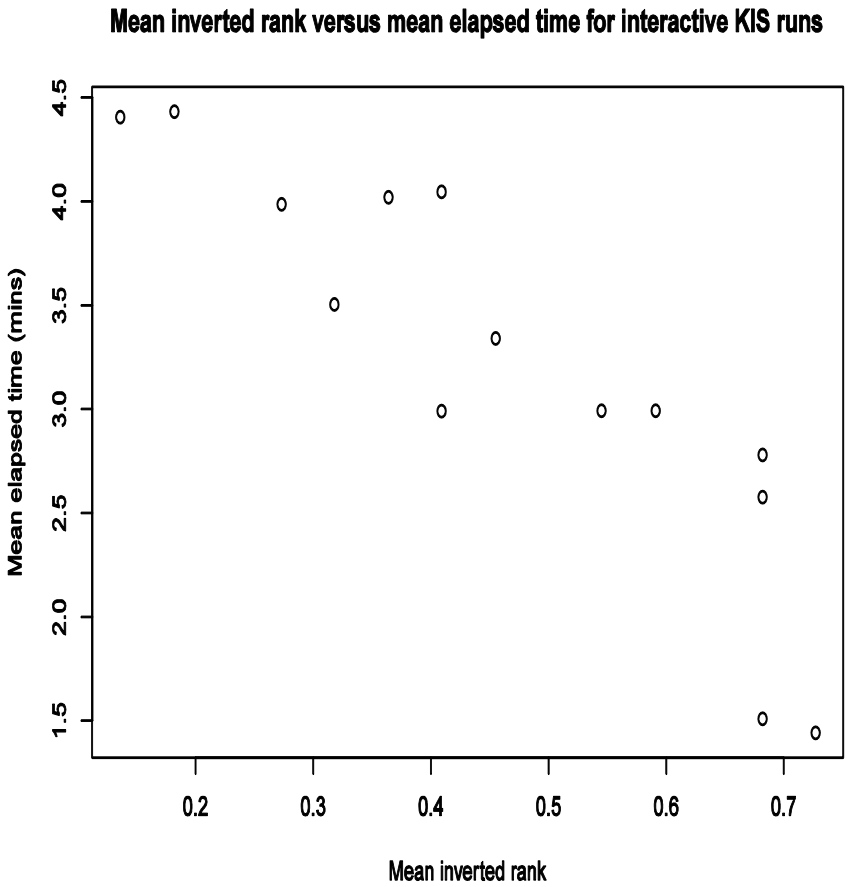
	Mean Time	IR	Sat
F_A_YES_I2R_AUTOMATIC_KIS_2_1	0.001	0.454	7.000
F_A_YES_I2R_AUTOMATIC_KIS_1_2	0.001	0.442	7.000
F_A_YES_MCPRBUPT1_1	0.057	0.296	3.000
F_A_YES_PicSOM_2_2	0.002	0.266	7.000
F_A_YES_ITEC-UNIKLU-1_1	0.045	0.265	5.000
F_A_YES_PicSOM_1_1	0.002	0.262	7.000
F_A_YES_ITEC-UNIKLU-4_4	0.129	0.262	5.000
F_A_YES_vireo_run1_metadata_asr_1	0.088	0.260	5.000
F_A_YES_ITEC-UNIKLU-2_2	0.276	0.258	5.000
F_A_YES_ITEC-UNIKLU-3_3	0.129	0.256	5.000
F_A_YES_CMU2_2	4.300	0.251	2.000
F_A_YES_vireo_run2_metadata_2	0.053	0.245	5.000
F_D_YES_MCG_ICT_CAS2_2	0.044	0.239	5.000
F_A_YES_MM-BA_2	0.050	0.238	5.000
F_D_YES_MCG_ICT_CAS1_1	0.049	0.237	5.000
F_A_YES_MM-Face_4	0.010	0.233	5.000
F_A_YES_MCG_ICT_CAS3_3	0.011	0.233	5.000
F_A_YES_CMU3_3	4.300	0.231	2.000
F_D_YES_CMU4_4	4.300	0.229	2.000
F_A_YES_LMS-NUS_VisionGo_3	0.021	0.215	6.000
F_D_YES_LMS-NUS_VisionGo_1	0.021	0.213	6.000
F_A_YES_CMU1_1	4.300	0.212	2.000



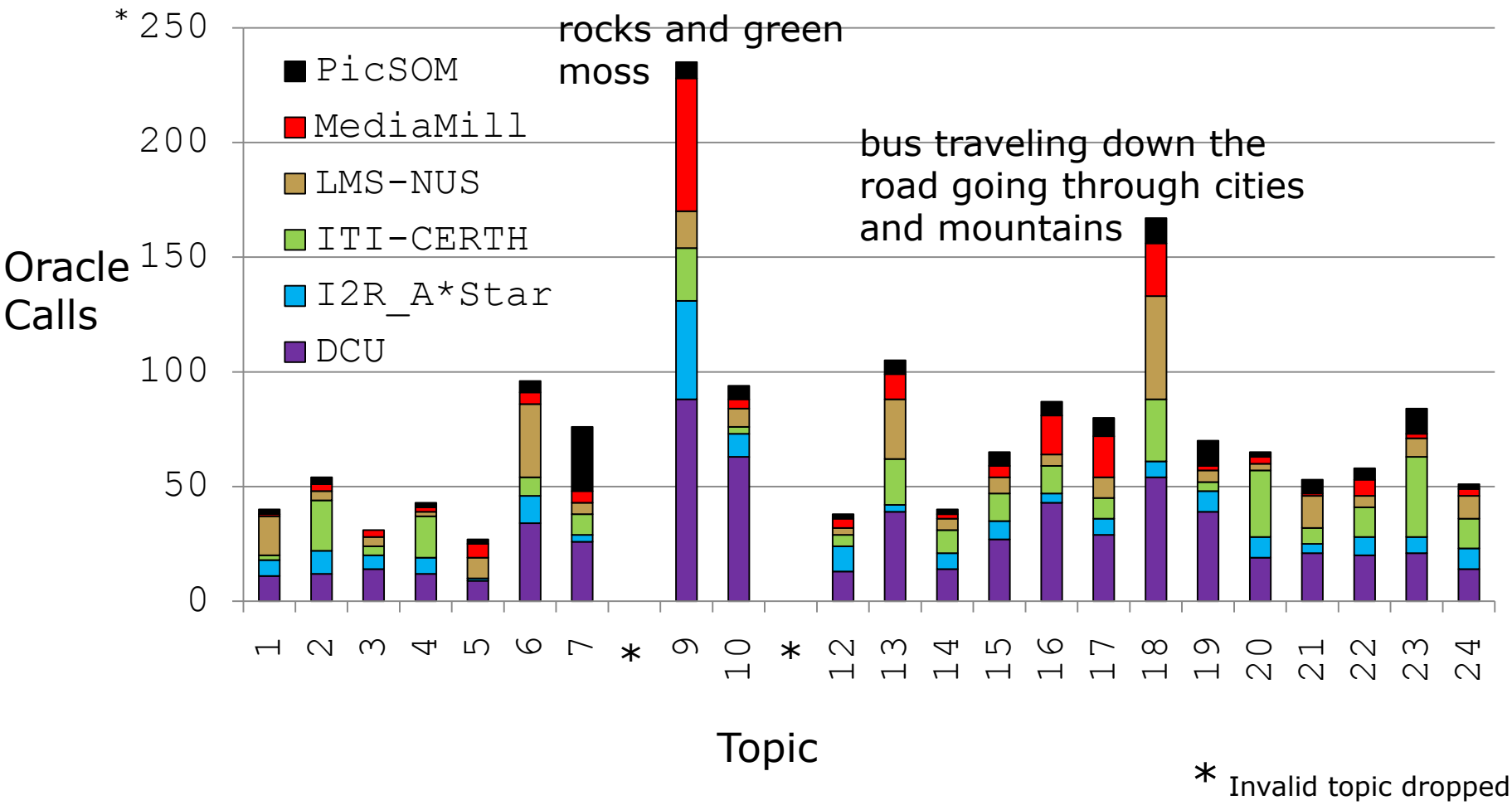
Results – interactive runs

	Mean Time	IR	Sat
I_A_YES_I2R_INTERACTIVE_KIS_2_1	1.442	0.727	6.000
I_D_YES_LMS-NUS_VisionGo_1	2.577	0.682	6.000
I_A_YES_LMS-NUS_VisionGo_4	2.779	0.682	5.750
I_A_YES_I2R_INTERACTIVE_KIS_1_2	1.509	0.682	6.300
I_A_YES_DCU-CLARITY-iAD_novice1_1	2.992	0.591	5.000
I_A_YES_DCU-CLARITY-iAD_run1_1	2.992	0.545	5.500
I_A_YES_PicSOM_4_4	3.340	0.455	5.000
I_A_YES_MM-Hannibal_1	2.991	0.409	3.000
I_A_YES_ITI-CERTH_2	4.045	0.409	6.000
I_A_YES_MM-Murdock_3	4.020	0.364	3.000
I_A_YES_PicSOM_3_3	3.503	0.318	6.000
I_A_YES_ITI-CERTH_1	3.986	0.273	5.000
I_A_NO_ITI-CERTH_4	4.432	0.182	4.000
I_A_NO_ITI-CERTH_3	4.405	0.136	4.000



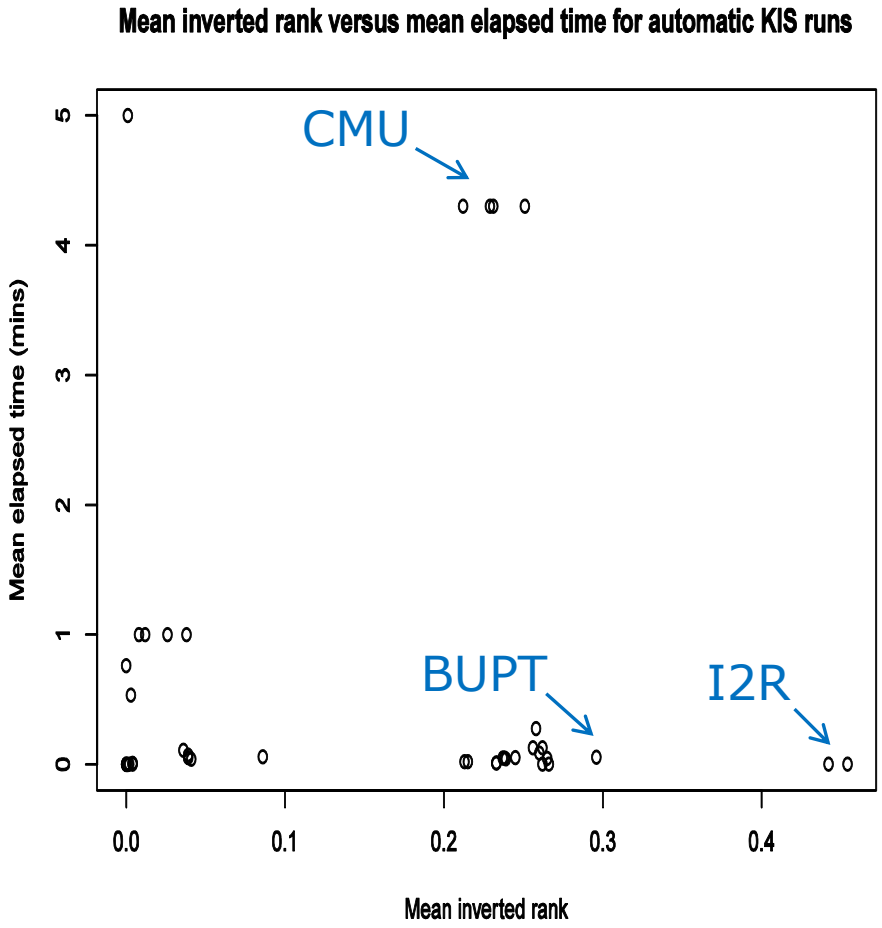


Results – oracle calls by topic and team



Results – automatic runs

	Mean Time	IR	Sat
F_A_YES_I2R_AUTOMATIC_KIS_2_1	0.001	0.454	7.000
F_A_YES_I2R_AUTOMATIC_KIS_1_2	0.001	0.442	7.000
F_A_YES_MCPRBUPT1_1	0.057	0.296	3.000
F_A_YES_PicSOM_2_2	0.002	0.266	7.000
F_A_YES_ITEC-UNIKLU-1_1	0.045	0.265	5.000
F_A_YES_PicSOM_1_1	0.002	0.262	7.000
F_A_YES_ITEC-UNIKLU-4_4	0.129	0.262	5.000
F_A_YES_vireo_run1_metadata_asr_1	0.088	0.260	5.000
F_A_YES_ITEC-UNIKLU-2_2	0.276	0.258	5.000
F_A_YES_ITEC-UNIKLU-3_3	0.129	0.256	5.000
F_A_YES_CMU2_2	4.300	0.251	2.000
F_A_YES_vireo_run2_metadata_2	0.053	0.245	5.000
F_D_YES_MCG_ICT_CAS2_2	0.044	0.239	5.000
F_A_YES_MM-BA_2	0.050	0.238	5.000
F_D_YES_MCG_ICT_CAS1_1	0.049	0.237	5.000
F_A_YES_MM-Face_4	0.010	0.233	5.000
F_A_YES_MCG_ICT_CAS3_3	0.011	0.233	5.000
F_A_YES_CMU3_3	4.300	0.231	2.000
F_D_YES_CMU4_4	4.300	0.229	2.000
F_A_YES_LMS-NUS_VisionGo_3	0.021	0.215	6.000
F_D_YES_LMS-NUS_VisionGo_1	0.021	0.213	6.000
F_A_YES_CMU1_1	4.300	0.212	2.000



Questions

How did use of IACC metadata affect system performance?

For example:

F_A_YES_MCPRBUPT1_1	0.296
F_A_NO_MCPRBUPT_2	0.004
F_A_NO_ MCPRBUPT_3	0.004
F_A_NO_ MCPRBUPT_4	0.002
F_D_YES_MCG_ICT_CAS2_2	0.239
F_D_YES_MCG_ICT_CAS1_1	0.237
F_A_YES_MCG_ICT_CAS3_3	0.233
F_D_NO_MCG_ICT_CAS4_4	0.001

How useful were the "1-5 KEY CUES" ?

Overview of submissions

15 teams completed the task, 6 interactive, 9 automatic

Here are the teasers

1. Aalto University School of Science and Technology (I)

- Picsom, formerly Helsinki University of Technology ?
- automatic and interactive runs submitted
- text search used Lucene on metadata and ASR, incl. WordNet synonyms, separate and combined indexes (best), concept matching (expanding definitions)
- concept detectors alone were inadequate, text much better, so integrated concepts and text via
  - (1) weighting detector scores and
  - (2) re-ranking based on concepts
- interactive search based on automatic then 1 of 2 search interfaces

2. Beijing University of Posts and Telecomms. - MCPRL

- concentrated on concept/feature based retrieval using 86 concepts with several suggested boosting approaches
- text alone was run against metadata and ASR
- other runs based on 86 of 130 concepts boosted by B&W detector, music/voice audio detector, motion detector
- also boosted by concept co-occurrence matrix
- text alone (i.e. no visual) performed best

3. Carnegie Mellon University-INF (S)

- used metadata, ASR (released and own), OCR combined via Lemur
- built and cued colour concept detectors (12 topics had colour)
- used LDA to describe joint description of text, and also SIFT, bag-of-words features
- included further topic examples taken from Google images
- performed query type classification (x5) and chose fusion based on this
- speaker slot to follow

4. Chinese Academy of Sciences-MCG

- submissions based on text search of metadata (no visual), but a visual baseline
- apply two text enrichment algorithms separately, based on Wikipedia and on Google
- results indicate content of web video too diverse to be usefully exploited
- Wikipedia based expansion using named entities etc., added value

5. City University of Hong Kong

- VIREO group - CUHK, and Sichuan University, China
- explored metadata, ASR and concept based search
- results are that text-only (metadata) is best, ASR has a complimentary role, concepts not effective
- reasons might be that query-to-concept mapping onto 130 concepts is too difficult and …
- performance of concept detectors is poor

6. Dublin City University (I,S)

- CLARITY: Centre for Sensor Web Technologies & and iAD: Information Access Disruptions (Bnorway)
- first year of multi-year plan, developed from scratch
- target non-expert users,
- iPad interface, multimodal retrieval
- experiment was novice (BI School of Management, Oslo) vs. expert (DCU) users
- have poster, demo, and speaker slot to follow

7. Hungarian Academy of Sciences

- and FBK Trento in JUMAS consortium, who gave ASR
- linear combination of text retrieval on metadata, ASR (from FBK), feature classifiers and ImageCLEF annotations
- feature detectors didn't use KIS training data but other sources (ImageCLEF, MIR Flickr)
- Interested in cross-domain application of feature detectors
- then run on KIS
- Surprisingly (!), metadata yielded best performance

8. Informatics and Telematics Institute, Thessaloniki (I,S)

- used VERGE, interactive retrieval application combining basic retrieval functionalities in various modalities
- included visual similarity, text, metadata, HLFs. concept fusion
- visual similarity used Color Layout, Color Structure, Scalable Color, Edge Histogram, and Homogeneous Texture
- text similarity based on ASR
- for high-level features, used 72 of the 130 concepts selected for the semantic indexing task
- concept fusion based on interaction from user
- results showed metadata was best, and content-based didn't add benefit
- speaker slot to follow

9. Institute for Infocomm Research I2R, Singapore (I,S)

- how to adapt traditional information retrieval, specifically video search methods to KIS in both automatic and interactive setting
- automatic query formulation is a focus, refine the query by formulating query phrases and weighing different query terms
- used multi-modal information sources, including text metadata, ASR, OCR, HLFs, audio classes and language type
- interactive targets user interface to facilitate browsing and fast rejection via a storyboard
- speaker slot to follow

10. KB Retrieval (David Etter)

- approach (in multiple tasks) is a knowledge-based one, using 400 classifiers based on LSCOM
- in KIS task, runs were based on different numbers (from the 400) of related concepts, so 3, 5, 10 and 15 related concepts

11. National University of Singapore (I,S)

- speaker slot to follow

12. NTT Communication Science Laboratories, Japan

- I only had the paper on CBCD !

13. University of Amsterdam (I)

- interactive and automatic runs submitted
- used combination of metadata, transcript (ASR) and concept detectors in official and post-submission runs
- query-independent fusion of results (not enough known about search types ?)
- interactive based on combining metadata, and several content-based searches with extensive visualisation on a large storyboard, a detail pane on the selected video, plus video categorization

x

14. University of Klagenfurt

- used text search on metadata as first step (pseudo relevance feedback ?)
- then used highest-ranked to create queries for content-based retrieval in 3 different feature spaces, then combined
- 3 subsequent searches are
  - -- colour and edge directivity descriptor
  - -- local feature histogram
  - -- global motion histogram
- fusion by a rank-based interlacing scheme
- content-based retrieval had little impact on text/metadata

15. York University

- first time, automatic runs, tough baptism
- focus on metadata search, query expansion
- used Lemur
- poor results, failure analysis to follow

Overall conclusions …

- This was a hard task !
- Metadata was great, OCR helped, concepts were not much assistance
- Reasons ?

Speakers

Institute for Infocomm Research, Singapore

Dublin City University

Carnegie Mellon University

National University of Singapore