KNOWN-ITEM SEARCH

Alan Smeaton
Dublin City University

Paul Over
NIST

NIST
National Institute of Standards and Technology

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. Its a natural, everyday scenario

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 (is 100 realistic ?),

-- 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.
- Task is replicable, has low judging overhead and is appealing

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

391 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

This is no YouTube in scale, but it is in nature. Its more like a digital library

Example topics

4

500 1-5 KEY VISUAL CUES: laptop, young man, swivel chair, lamp, dresser

QUERY: Find a video of a young man narrating a video showing a young man in jeans sitting in front of a laptop in a room with a desk, table lamp, and dresser and then moving to a bedroom with two females sleeping and being awoken in bed with the narrator mentioning "ambush cinematography" and asking what is on the tv.

501 1-5 KEY VISUAL CUES: Newsreel clips, Natilus Nuclear submarine, NY harbor, hunter killer helicopter, Pan AM passenger jet

QUERY: Find video featuring Newsreel clips of Nautilus Nuclear submarine entering NY harbor, a Hunter Killer helicopter and the first Pan Am commerical passenger jet

502 1-5 KEY VISUAL CUES: Staten Island ferry, Statue of Liberty, Ellis Island

QUERY: Find the video of people using ferry and touring Ellis Island

503 1-5 KEY VISUAL CUES: action-pack clip, man-flip, automatic weapon, light saber, car-spinning

QUERY: Find a video of an action-pack clip showing a man in a blue jacket doing a flip and hitting another man, a man firing an automatic weapon, a man with a light saber and a white car spinning around.

504 1-5 KEY VISUAL CUES: man, city backdrop,business suit

QUERY: Find the video with a man in a business suit broadcasting in front of a city backdrop, text on screen, relating to various news stories

TV2011 Finishers (9 from 40 applicants)

PicSOM	Aalto University
AXES-DCU *	Access to Audiovisual Archives
BUPT-MCPRL	Beijing University of Posts & Telecom.-MCPRL
ITI-CERTH *	Centre for Research and Technology Hellas
VIREO	City University of Hong Kong
DCU-iAD-CLARITY *	Dublin City University
KBVR	KB Video Retrieval
KSLab-NUT *	Nagaoka University of Technology in Japan
SCUC	Sichuan University of China (no paper !)

* - submitted interactive run(s)

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

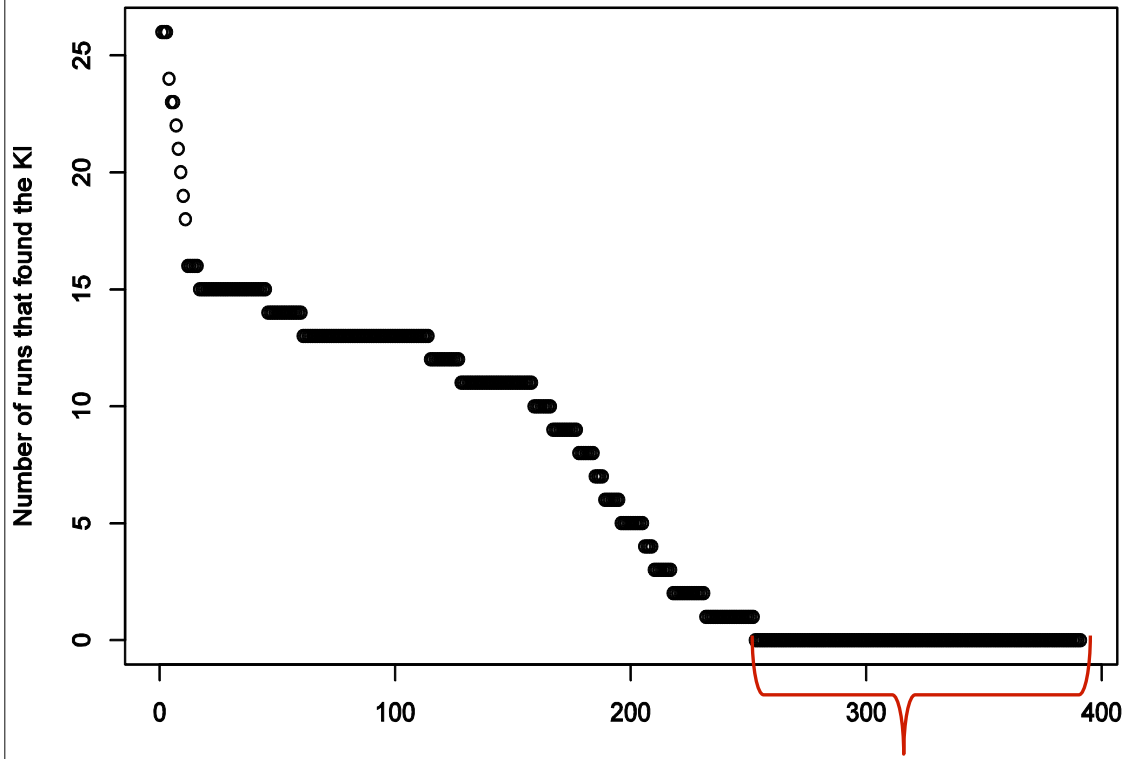
Evaluation

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
  - Calculated automatically using ground truth created with the topics
- mean elapsed time (mins.)
- user satisfaction (interactive) (1-7(best))

Results – topic variability

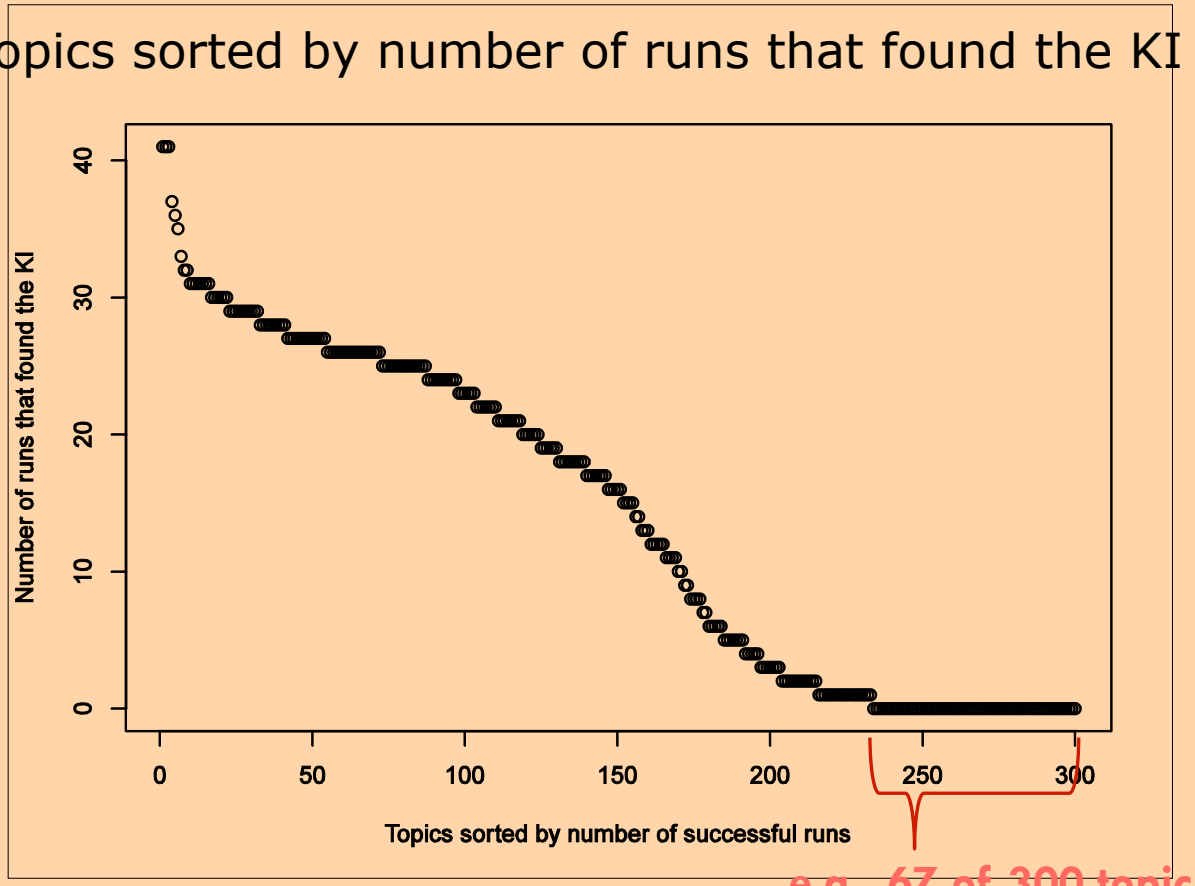
Topics sorted by number of runs that found the KI



e.g., 139 of 391 topics (35%) were never successfully answered

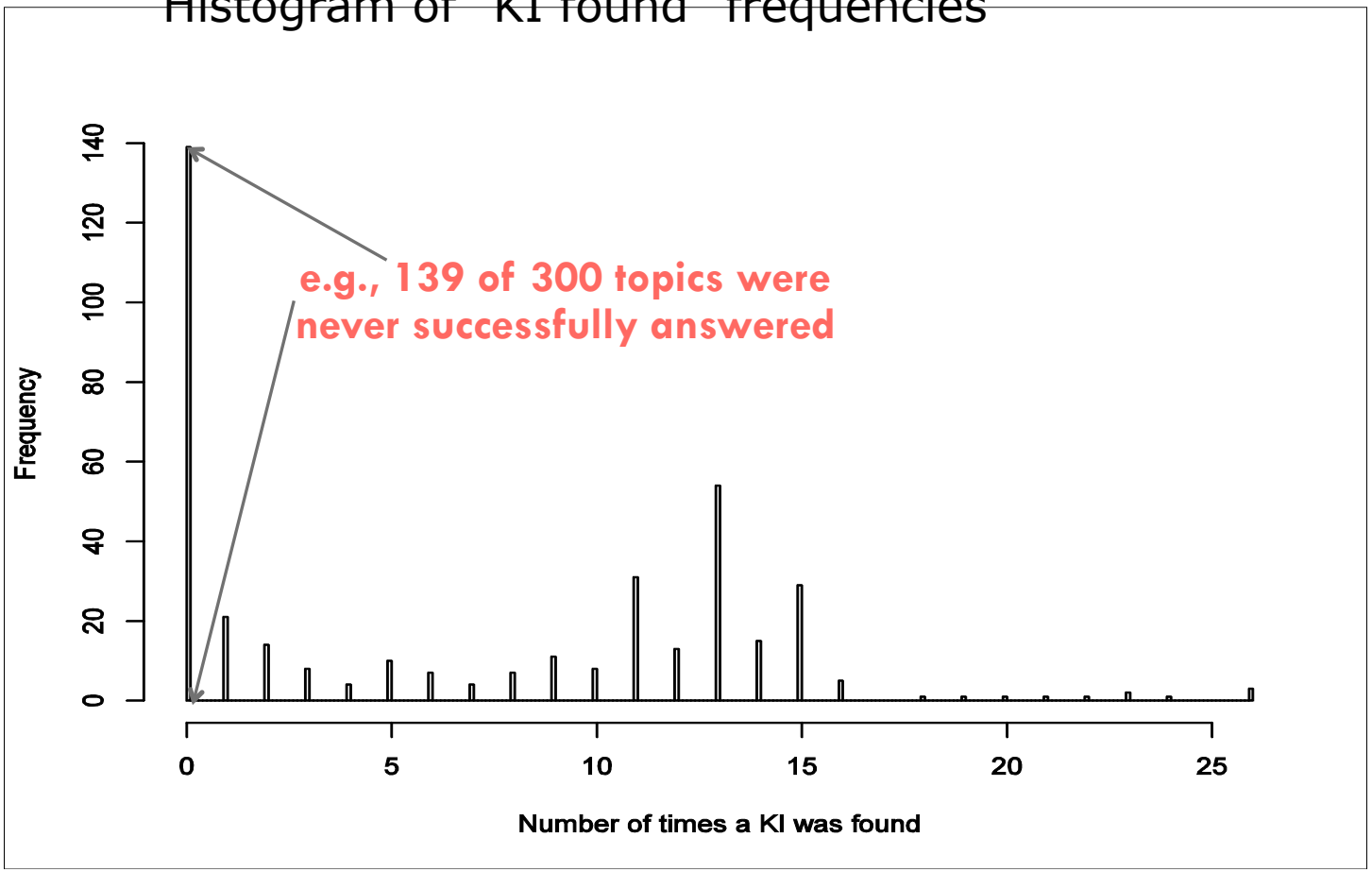
2010 Results – topic variability

Topics sorted by number of runs that found the KI

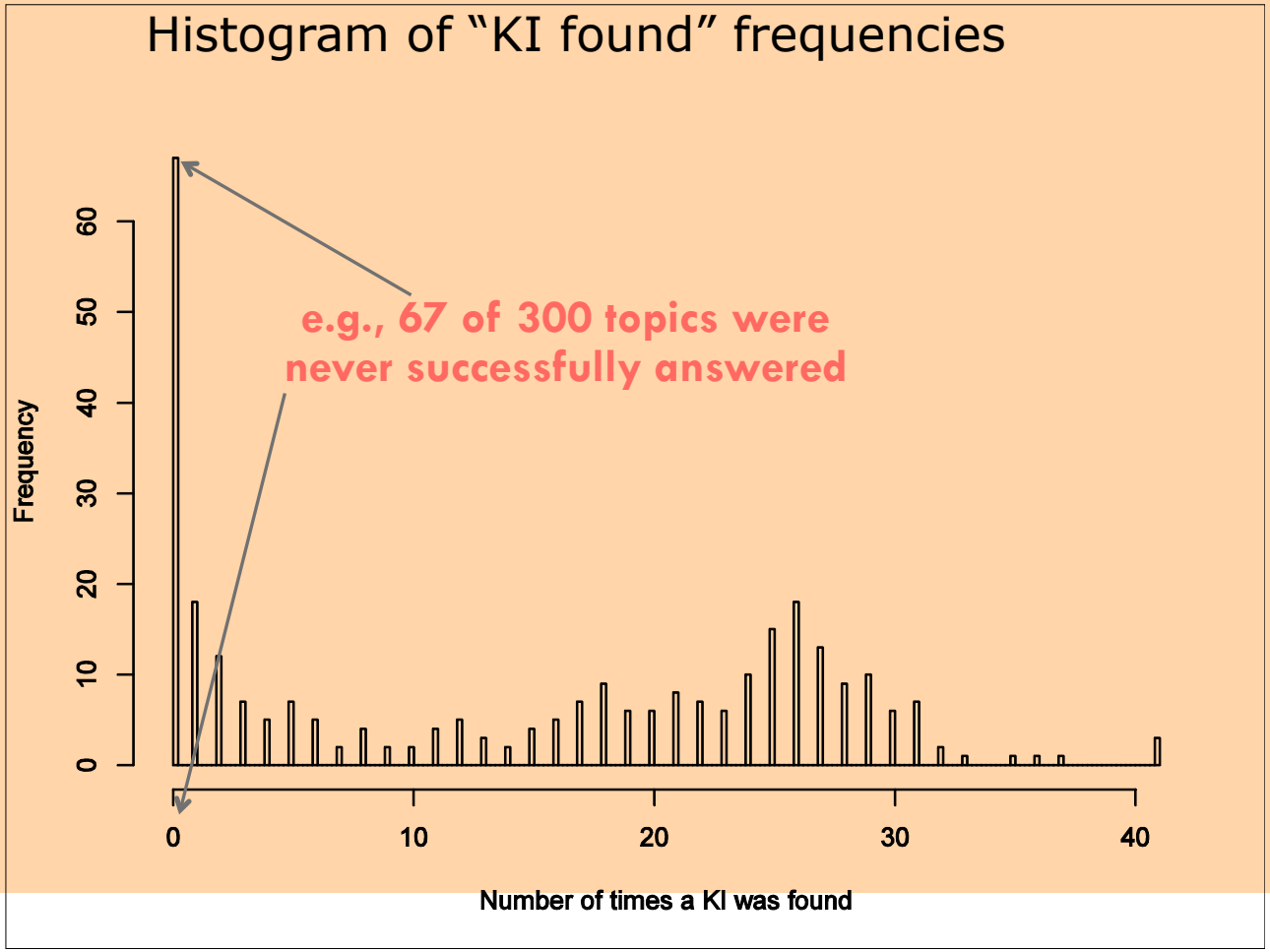


e.g., 67 of 300 topics (22%) were never successfully answered

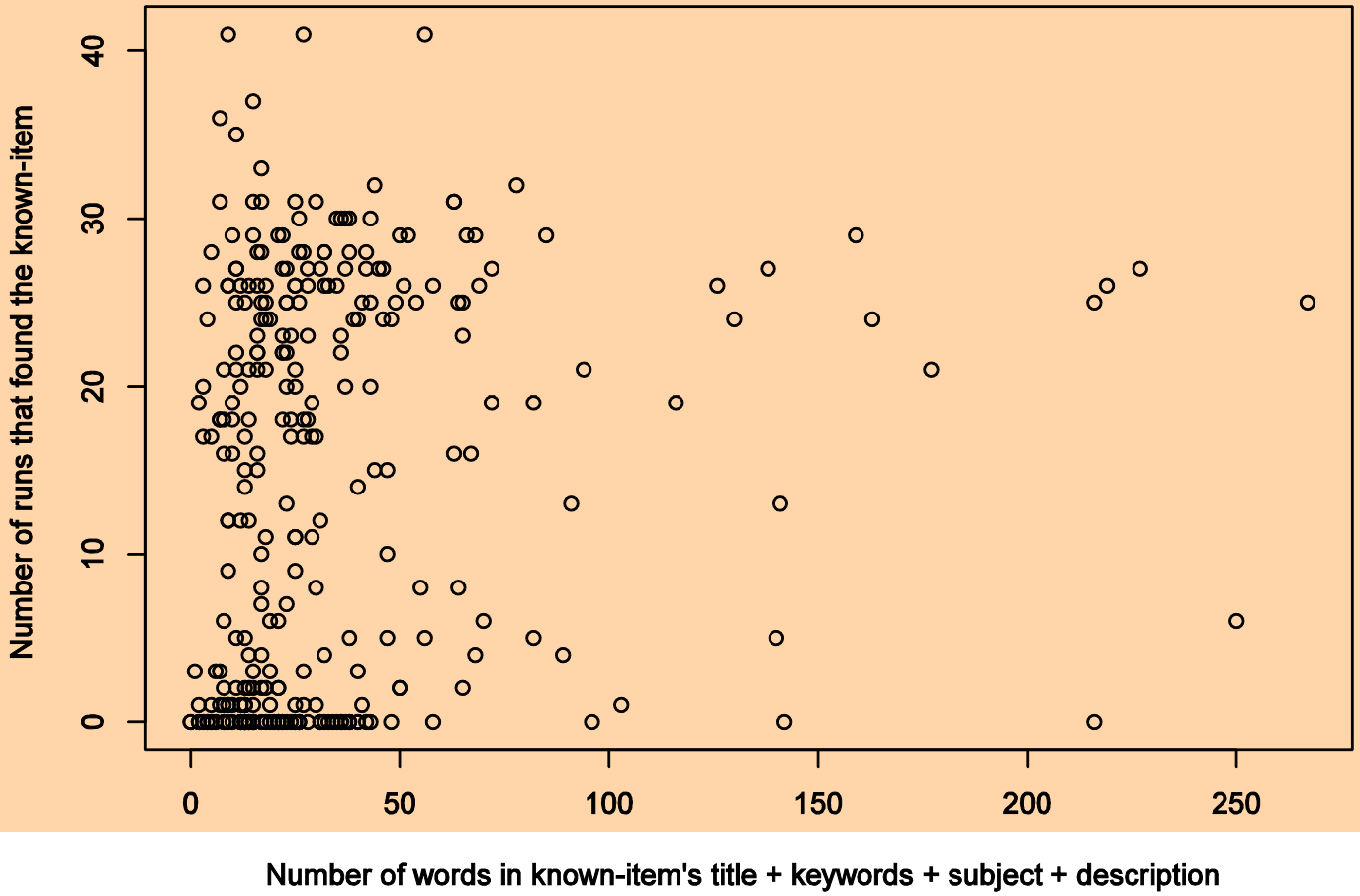
Results – topic variability



2010 Results – topic variability



2010 Results – does amount of metadata



Not found vs. found – what's the difference?

0860 CUES: woman, brown hair, brown couch, light blue shirt
0860 QUERY: Find the video with a woman who has shoulder length brown hair, a light blue shirt, sitting on a brown couch talking about how to talk to angels.

0861 CUES: President foreground curtin White Hpuse background
0861 QUERY: Find a video with President Bush in foreground and blue curtin with White House logo in background.

0862 CUES: Andalucia, goat herd, black dog, hills, flowers, woman, red jacket
0862 QUERY: Find the video of Andalucia with goat herd and black dog, flowering hillside, and woman in red jacket.

0863 CUES: man in blue shirt in a chair, hands moving wildly, web site
0863 QUERY: Find a video of a man in a blue shirt sitting in a chair yelling and complaining about Kim Kardashian and says "Kim Kardashian is a whore" and it shows the drinkingwithbob web address.

0864 CUES: man, greeting card display
0864 QUERY: Find a video with a man standing beside a greeting card display. He is facing the camera and talking.

0865 CUES: gorilla, wrecking ball
0865 QUERY: Find the video with various scenes that appear in sqare frames and circle frames, including a picture of a gorilla in a circle and a picture of a wrecking ball in a square.

0866 CUES: cat weapon
0866 QUERY: Find a video of a cat firing an automatic weapon.

0867 CUES: yellowstone park, gyser, music
0867 QUERY: Find the video showing yellowstone park and gyser going off as music is played in background

0868 CUES: drawing, man, large pink face, large mouth, teeth
0868 QUERY: Find a video with a drawing of the upper body of a man with a large pink face and a large mouth showing a lot of teeth. He is wearing a black shirt.

0869 CUES: man, t-shirt, German, stain, rubbing
0869 QUERY: Find a video of a man in t-shirt speaking in German trying to remove a stain by hard rubbing with a cloth.

Not found vs. found – what's the difference?

0870 CUES: SEE SAN DIEGO, WITH THE ULTIMATE PARTY, young people touring, dancing, drinking
0870 QUERY: Find a video ad of a bus tour of San Diego for happy hour titled "SEE SAN DIEGO WITH THE ULTIMATE PARTY" and shows young people touring, dancing, and drinking.

0871 CUES: four men, singers
0871 QUERY: Find a video of four men singing "Are You Going to Scarborough Fair" without accompaniment.

0872 CUES: Christof, Tony Blair, Harriet Tubman
0872 QUERY: Find a video showing a man in a baseball cap talking about a television newscast with his friend Christof,Tony Blair and honoring the Harriet Tubman Center.

0873 CUES: red and white plane, shore
0873 QUERY: Find the video of small red and white plane flying over shore.

0874 CUES: band, green light, guitar, white hat, audience
0874 QUERY: Find a video of a band playing with green light shining on them. One guitarist/singer wears a large white hat low over his eyes. A large audience is excited.

0875 CUES: t-shirt, sgirl, flag, photoshop
0875 QUERY: Find a video of demonstration of using photoshop to retouch picture of girl wearing yellow t-shirt standing in front of flag

0876 CUES: film, chapters, "My Video"
0876 QUERY: Find a video with no sound showing film clips identifying three chapters in a home film entitled "My Video".

0877 CUES: baby,chair, "Mary had a little lamb",man, "The Itsy Bitsy Spider."
0877 QUERY: Find a video of baby boy sitting in a chair while an unseen woman sings "Mary had a little Lamb" followed by a man holding the boy sings "The Itsy Bitsy Spider".

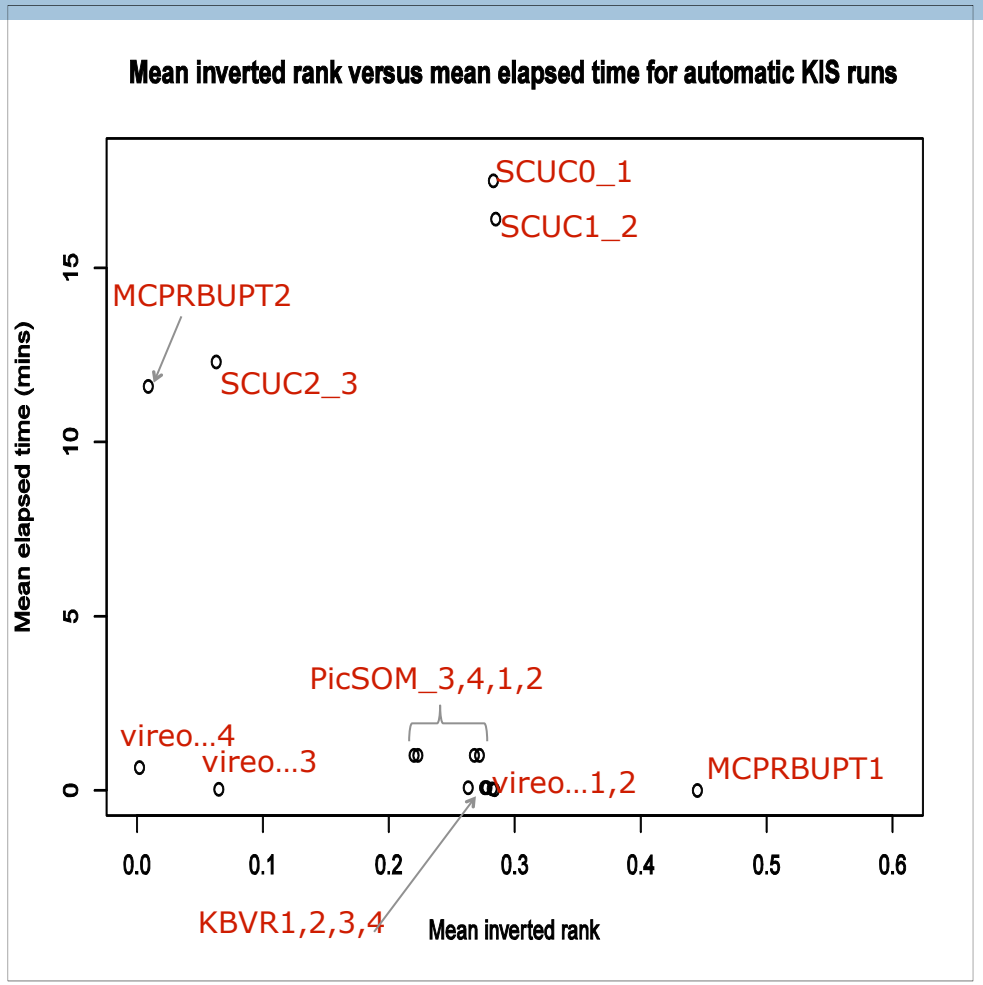
0878 CUES: people, movie, chandelier, projection screen
0878 QUERY: Find the video with people sitting at tables in a room with a chandelier watching a movie on a projection screen.

0879 CUES: man-gray hooded jacket, liquor store, man-white hat, T shirt-McDaddy, man-red Tshirt
0879 QUERY: Find a video of a man in gray jacket with hood over head, a liquor store, a man wearing a white hat, black T shirt with MacDaddy written on front buying a 12 pack of beer and man wearing a red T shirt and black hat

Results – automatic runs

15

	Mean Time	IR	Sat
F_A_YES_MCPRBUPT1_1	0.001	0.445	3.000
F_A_NO_SCUC1_2	16.400	0.285	1.000
F_A_YES_vireo_run...2	0.024	0.284	5.000
F_A_NO_SCUC0_1	17.500	0.283	1.000
F_A_YES_vireo_run...1	0.043	0.282	5.000
F_D_YES_KBVR_4	0.075	0.278	1.000
F_D_YES_KBVR_3	0.074	0.277	1.000
F_D_YES_KBVR_1	0.075	0.276	1.000
F_A_YES_PicSOM_2_2	1.000	0.272	7.000
F_A_YES_PicSOM_1_1	1.000	0.268	7.000
F_D_YES_KBVR_2	0.074	0.263	1.000
F_A_YES_PicSOM_4_4	1.000	0.223	7.000
F_A_YES_PicSOM_3_3	1.000	0.220	7.000
F_A_YES_vireo_run...3	0.028	0.065	5.000
F_A_NO_SCUC2_3	12.300	0.063	1.000
F_A_NO_MCPRBUPT2_2	11.601	0.009	3.000
F_A_YES_vireo_run...4	0.657	0.002	5.000

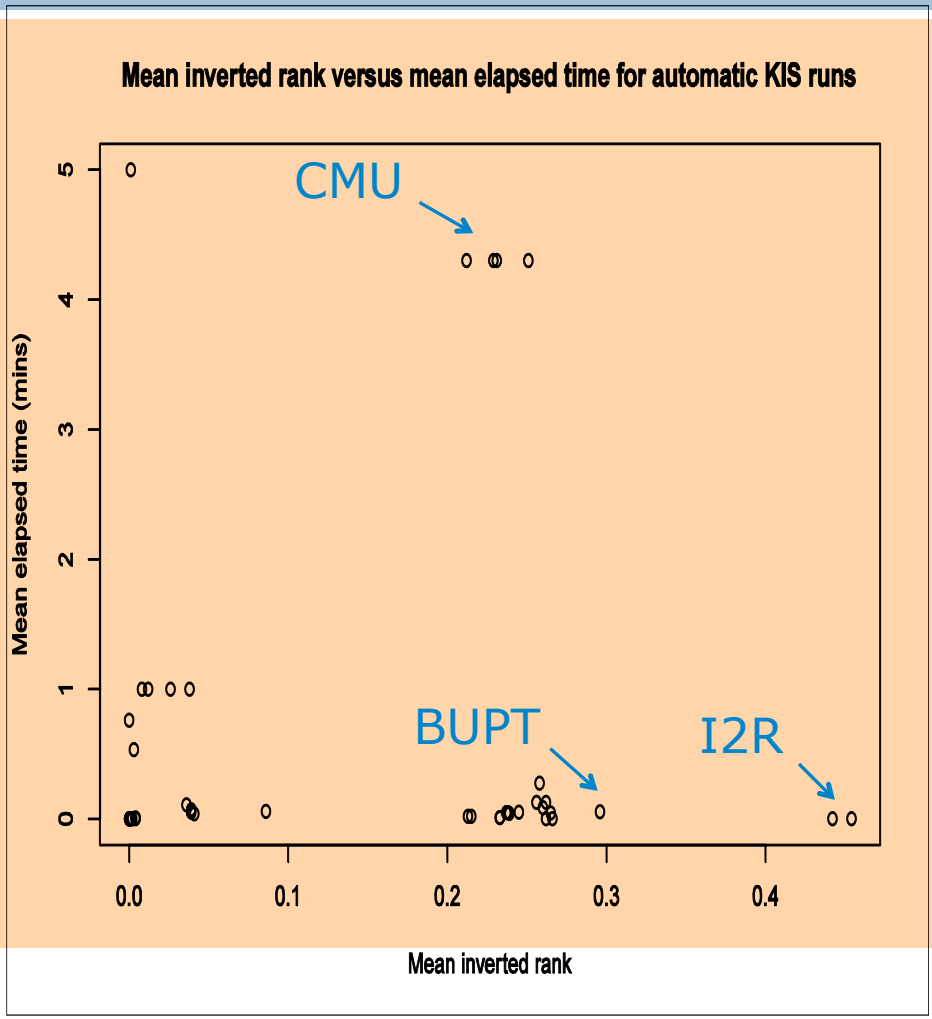


NIST
National Institute of Standards and Technology

TRECVID 2011

2010 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



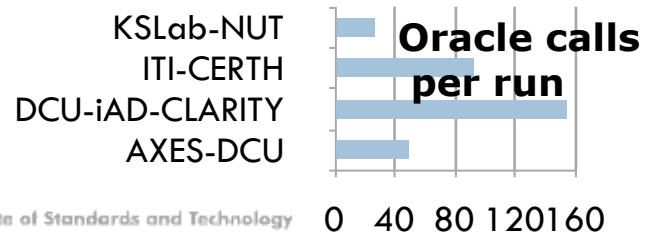
TRECVID 2011

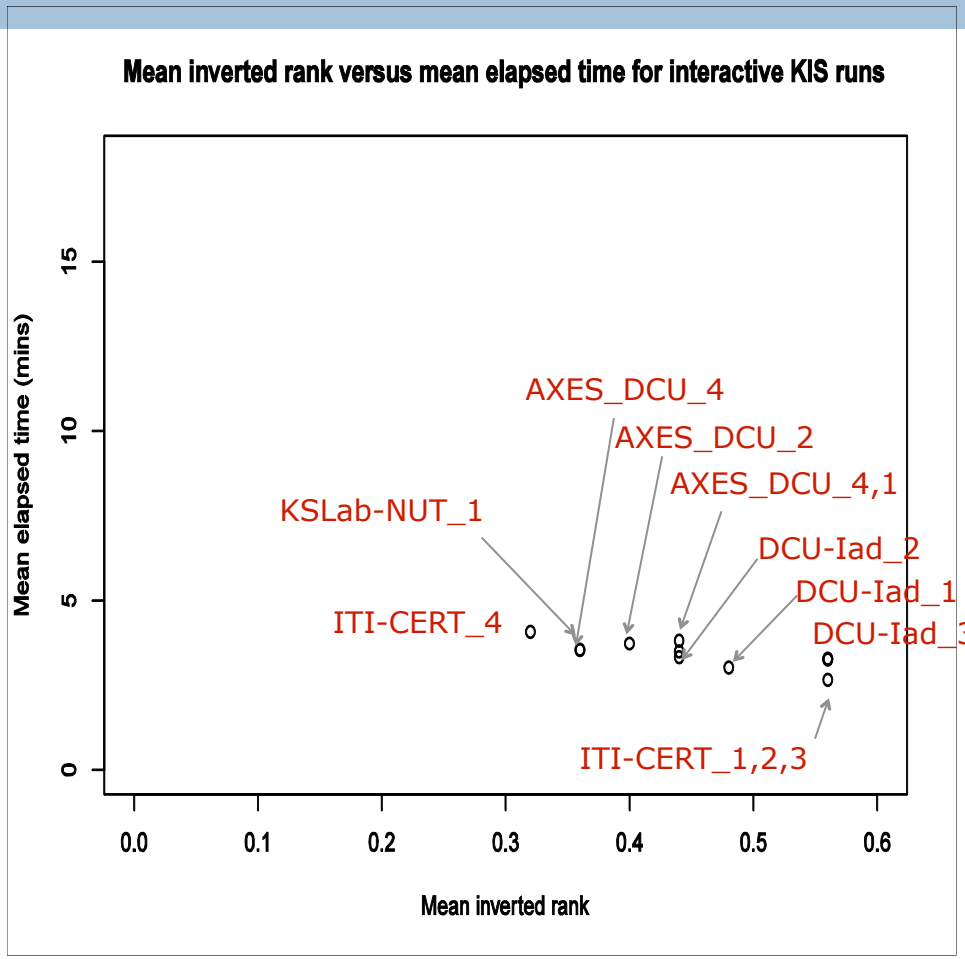
NIST National Institute of Standards and Technology

Results – interactive runs

17

	Mean Time	IR	Sat
I_A_YES_ITI-CERTH_3	3.274	0.560	5.000
I_A_YES_ITI-CERTH_2	3.284	0.560	6.000
I_A_YES_ITI-CERTH_1	3.257	0.560	6.000
I_A_YES_DCU-IAd...3	2.660	0.560	5.000
I_A_YES_DCU-IAd...1	3.022	0.480	5.000
I_A_YES_DCU-IAd...2	3.324	0.440	5.000
I_A_YES_AXES_DCU_4_4	3.811	0.440	7.000
I_A_YES_AXES_DCU_1_1	3.498	0.440	7.000
I_A_YES_AXES_DCU_2_2	3.729	0.400	7.000
I_B_YES_KSLab-NUT_1	3.544	0.360	4.000
I_A_YES_AXES_DCU_3_3	3.542	0.360	7.000
I_A_YES_ITI-CERTH_4	4.072	0.320	5.000





NIST National Institute of Standards and Technology

TRECVID 2011

2010 Results – interactive runs

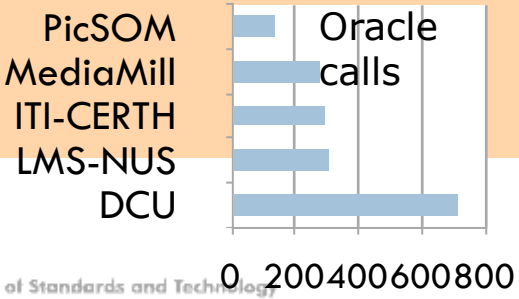
18

	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

Randomization tests (p<0.05)

I_A_YES_LMS-NUS_VisionGo_4
I_D_YES_LMS-NUS_VisionGo_1
I_A_YES_I2R_INTERACTIVE_KIS_2_1
> I_A_NO_ITI-CERTH_4
> I_A_YES_ITI-CERTH_1
> I_A_YES_ITI-CERTH_2
> I_A_YES_PicSOM_4_4
> I_A_YES_MM-Hannibal_1
> I_A_NO_ITI-CERTH_3
> I_A_YES_MM-Murdock_3
> I_A_YES_PicSOM_3_3
I_A_YES_DCU-CLARITY-iAD_novice1_1
> I_A_NO_ITI-CERTH_1
> I_A_NO_ITI-CERTH_3
> I_A_NO_ITI-CERTH_4
> I_A_YES_PicSOM_3_3

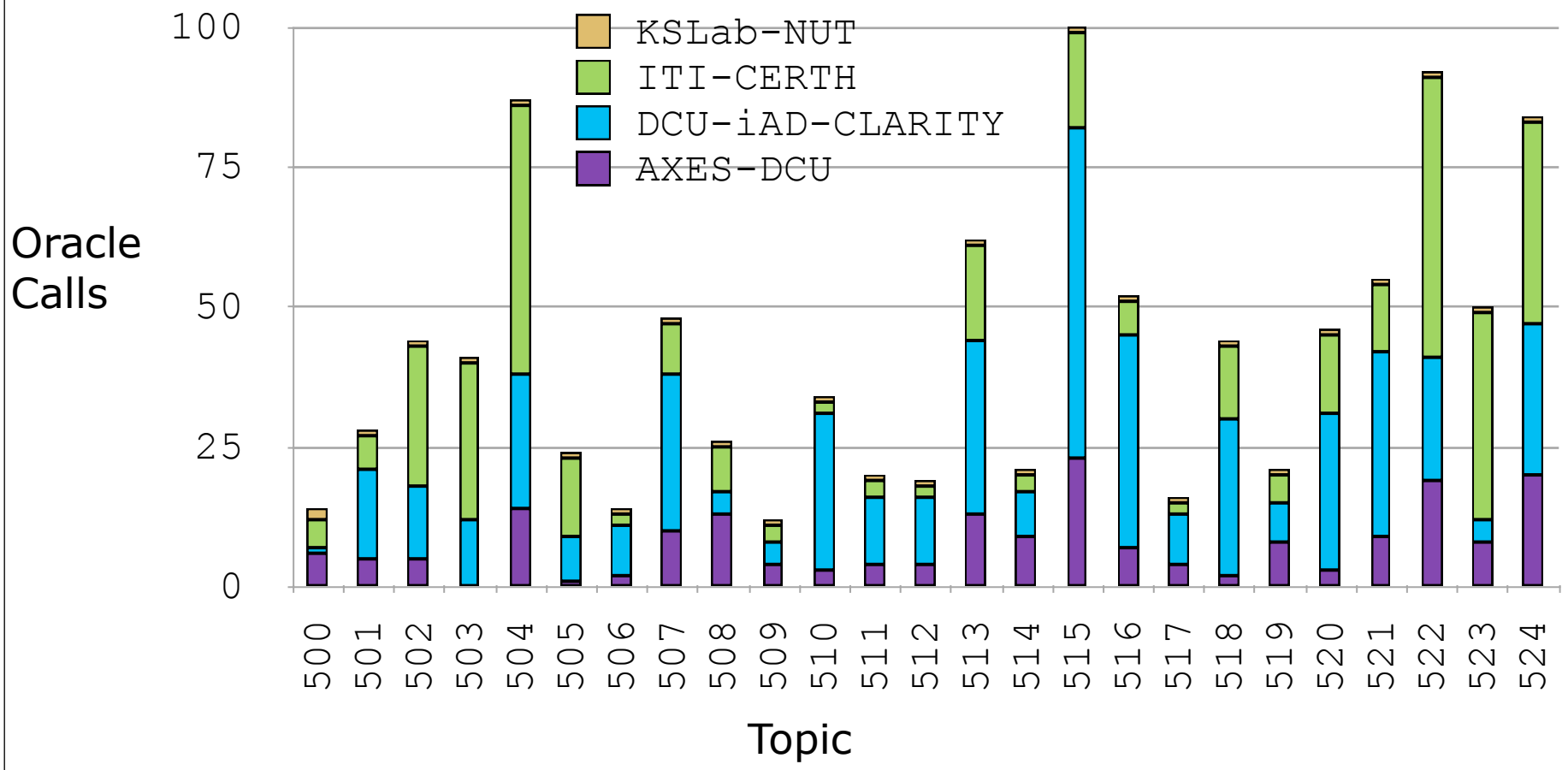
I_A_YES_DCU-CLARITY-iAD_run1_1
> I_A_NO_ITI-CERTH_3
> I_A_NO_ITI-CERTH_4
> I_A_YES_PicSOM_3_3
I_A_YES_I2R_INTERACTIVE_KIS_1_2
> I_A_NO_ITI-CERTH_3
> I_A_NO_ITI-CERTH_4
> I_A_YES_ITI-CERTH_1
> I_A_YES_ITI-CERTH_2
> I_A_YES_MM-Hannibal_1
> I_A_YES_MM-Murdock_3
> I_A_YES_PicSOM_3_3



NIST National Institute of Standards and Technology

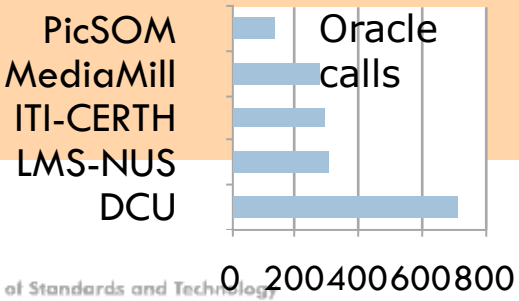
TRECVID 2011

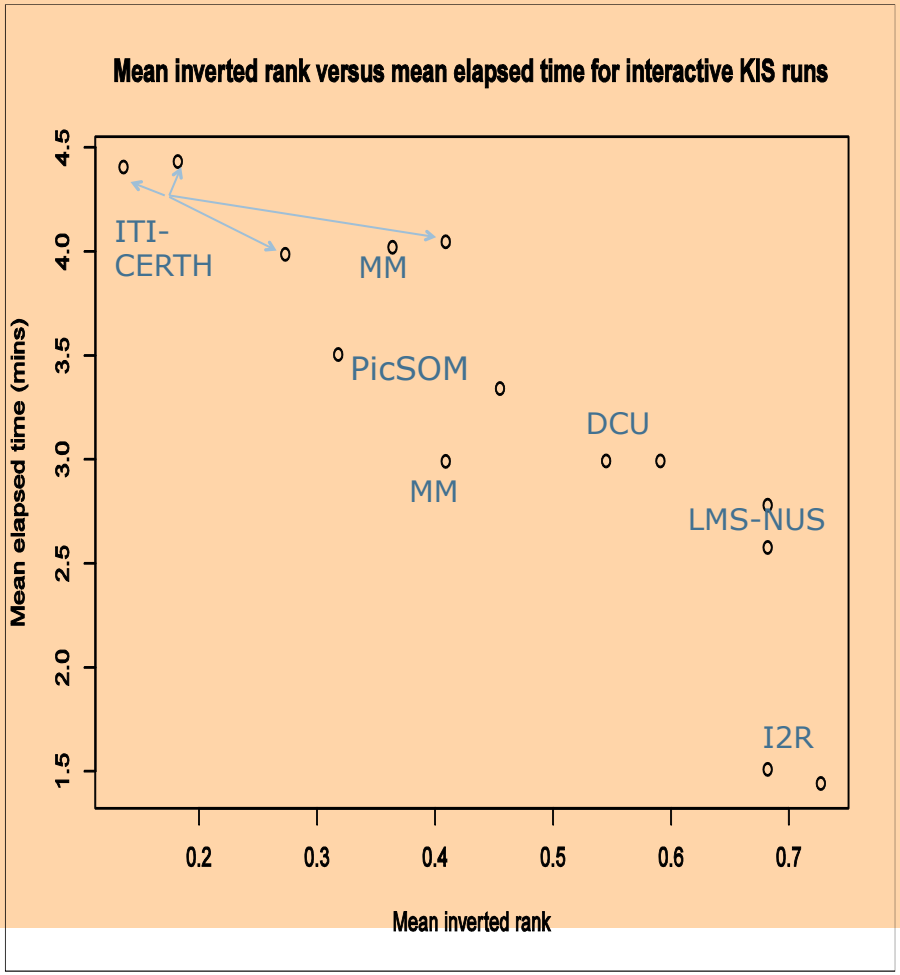
Results – oracle calls by topic and team



2010 Results – interactive runs

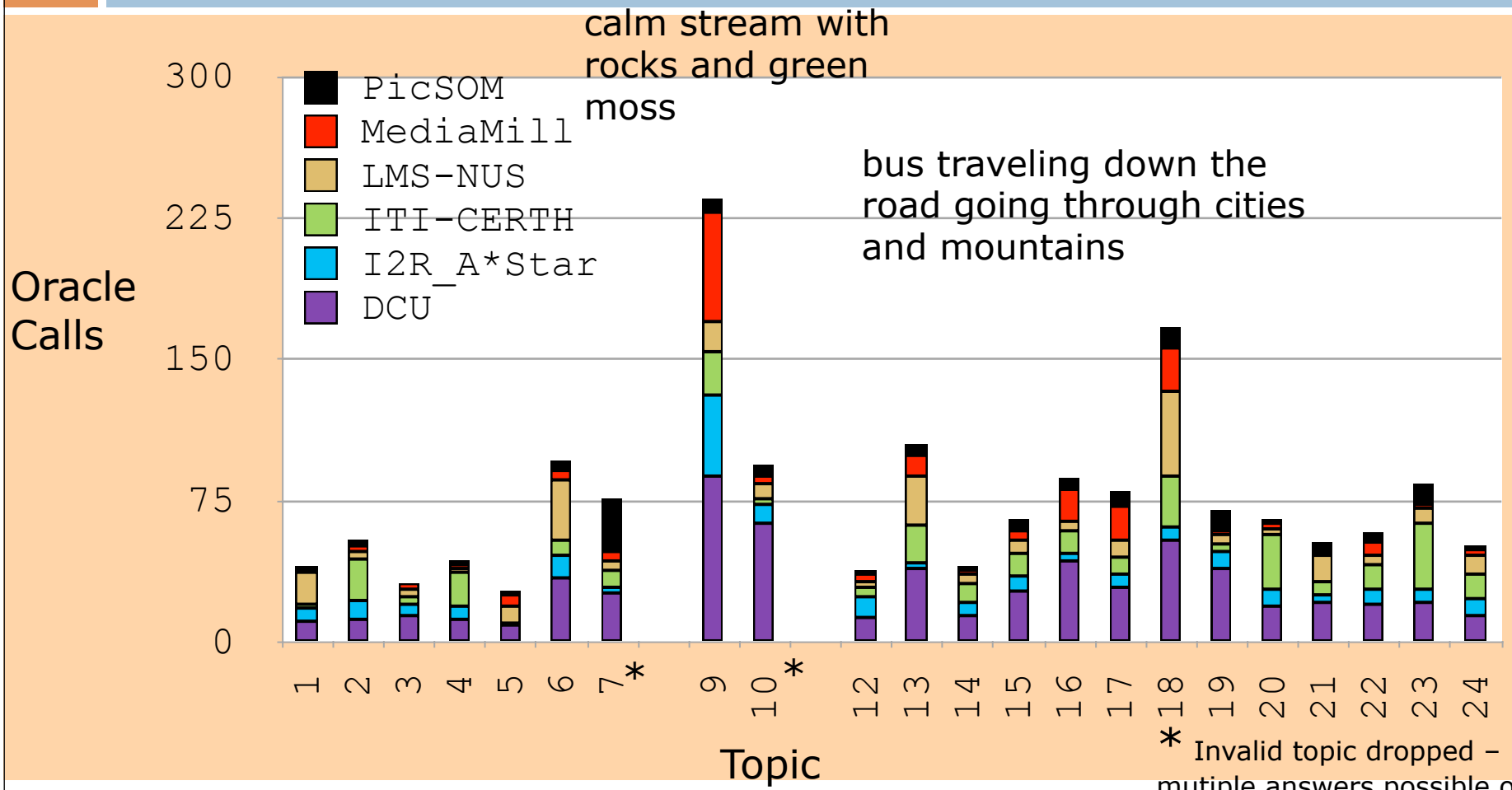
	Mean Time	IR	Sat
I_A_YES_I2R_INTERACTIVE_KIS_2_1	1.442	0.727	6.000
I_D_YES_LMS-NUS_VisionGo_1	2.577	0.682	6.000
I_A_YES_LMS-NUS_VisionGo_4	2.779	0.682	5.750
I_A_YES_I2R_INTERACTIVE_KIS_1_2	1.509	0.682	6.300
I_A_YES_DCU-CLARITY-iAD_novice1_1	2.992	0.591	5.000
I_A_YES_DCU-CLARITY-iAD_run1_1	2.992	0.545	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





TRECVID 2011

2010 Results – oracle calls by topic and team



* Invalid topic dropped – mutiple answers possible or answer not present in video

2010 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" ?

Personal overview of finishers

PicSOM	Aalto University
AXES-DCU *	Access to Audiovisual Archives
BUPT-MCPRL	Beijing University of Posts & Telecom.-MCPRL
ITI-CERTH *	Centre for Research and Technology Hellas
VIREO	City University of Hong Kong
DCU-iAD-CLARITY *	Dublin City University
KBVR	KB Video Retrieval
KSLab-NUT *	Nagaoka University of Technology in Japan
SCUC	Sichuan University of China (no paper !)

* - submitted interactive run(s)

PicSOM

- Based on combining simple text search with automatically matched semantic concepts using concept detectors from the semantic indexing task
- Tried to improve search by augmenting the metadata and ASR text with the output of optical character recognition
- Automatic runs used text search with a single video-level index containing all the ASR text plus the title, description and subjects from the meta data.
- Also included text detected by OCR, lemmatisation and used automatic selection of concepts based on matching keywords in the query text.
- Neither the concept detectors nor the lemmatisation managed to improve over baseline

AXES (speaker slot)

- Large FP7 team from DCU, U Twente, Erasmus University, NISV, Oxford University, IIIT and Fraunhofer - 18 authors
- Used text search on ASR, visual concepts and visual similarity.
- System was a year 1 build for a multi-year participation in interactive KIS and INS tasks this year

  1. Text search based on ASR and metadata used Lucene search engine
  2. Concepts based on a Pyramid Histogram of Visual Words and a very efficient SVM classifier
  3. Similarity search not based on the usual LLFs (colour, texture, etc) but on an object search, an elliptical region

- Three sources then fused together. Desktop user interface with 14 media professionals as users from NISV in Amsterdam

Beijing U Posts & Telecomms

- Regular participant, participated in 5 tasks
- Two methods proposed ... traditional text-based and a novel bio-inspired method.
- Text-based search consisted of text pre-processing, keyword extracting and processing, text-based retrieval, results fusion and re-ranking. Also used a manual ontology for query words, and used on top of Lucene
- KIS bio-inspired framework includes five parts:
  - a bottom-up attention model for determining salient regions,
  - a knowledge base containing various pre-trained object/concept (such as person, car) detectors,
  - a SOM (Self-Organizing Maps) network to map known-item keywords into seven image-related classes,
  - a SVM scene classifier for data filtering,
  - a fusion module to perform content-based retrieval, results fusion and ranking.

Text search was great, bio-inspired was not !

ITI-CERTH (speaker slot)

- Another long-term participant, using TRECVid annually in a series of build-on-build experiments
- Employed VERGE, an interactive retrieval application combining retrieval functionalities in various modalities and exploiting implicit user feedback
  - Implicit Feedback Capturing Module - time hovering over a shot, previewing
  - Visual Similarity Search Module - MPEG-7 based
  - Transcription Search Module
  - Metadata Processing and Retrieval Module
  - Video Indexing using Aspect Models and the Semantic Relatedness of Metadata - Bag-of-words representation of video
  - High Level Concept Retrieval and Fusion Module
  - High Level Concept and Text Fusion Module

VIREO (CUHK)

- Regular participant, several tasks, building on 2010 participation
- Set out to observe the effectiveness of different modalities (metadata, automatic speech recognition (ASR) and concepts) using same approach as in 2010
- Consistent with previous year's results, the evaluation once again shows that concept-based search is not helpful in known-item search
- Textual-based modalities continue to deliver reliable performance especially the metadata.
- Supplementing the metadata with the ASR feature is not longer able to boost the performance, unlike last year

iAD-DCU (speaker slot)

- Representing a collaboration with Norwegian Universities and groups, funded by Norwegian Research Council
- Building on participation in interactive KIS in 2010 which used novice users
- Implemented an iPad interface to a KIS video search tool to evaluate different display methodologies for KIS interaction.
- Keyframe clustering based on MPEG-7 features using k-means
- Employ concept detection for search and for choosing most representative keyframes.
- Compare baseline non-clustering to a clustering system on a topic by topic basis.
- 6 interactive users in Oslo and in Dublin

KBVR

- Baseline text-only runs plus pseudo-RF and semantic concept re-ranking
- Used Terrier system on ASR and metadata
- Semantic concept re-ranking assumes Known Item is retrieved but needs to be "bubbled up" the ranking
- Query, and initial 'documents' mapped into a semantic space defined using 130 LSCOM concepts with the description of each concept enhanced using a Wikipedia knowledge base

KSLab Nagaoka University of Tech.

31

- Also developed an iPad interface for interactive KIS, first participation
- Searched the metadata using Lucene, refining salient words integrated into retrieval
- Used video length as a cue for the user

Sichuan University of China

- No paper to judge

Questions for participants

Why the large(r) number of topics unanswered by all?

- 2010: 67 of 300 (22%)
- 2011: 139 of 391 (35%)

Did any groups run their 2011 system on 2010 test data or vice versa?

Any evidence use of metadata as crucial as in 2010?

- 2 automatic SCUC runs seem to be counter-examples