TRECVID 2018 INSTANCE RETRIEVAL

INTRODUCTION AND TASK OVERVIEW

Wessel Kraaij
Leiden University; The Netherlands
Organisation for Applied Scientific Research TNO;

George Awad
Dakota Consulting ; National Institute of Standards and Technology

Keith Curtis
National Institute of Standards and Technology

Table of contents

- Task Definition
- Data
- Topics (Queries)
- Participating teams
- Evaluation & results
- General observation

Task

From 2013 – 2015

- The task asked systems to find a specific object, person or location in any context using a small set of image and video examples.

In 2016 - 2018

- A new query type was used: *find a specific person in a specific location.*

System task:

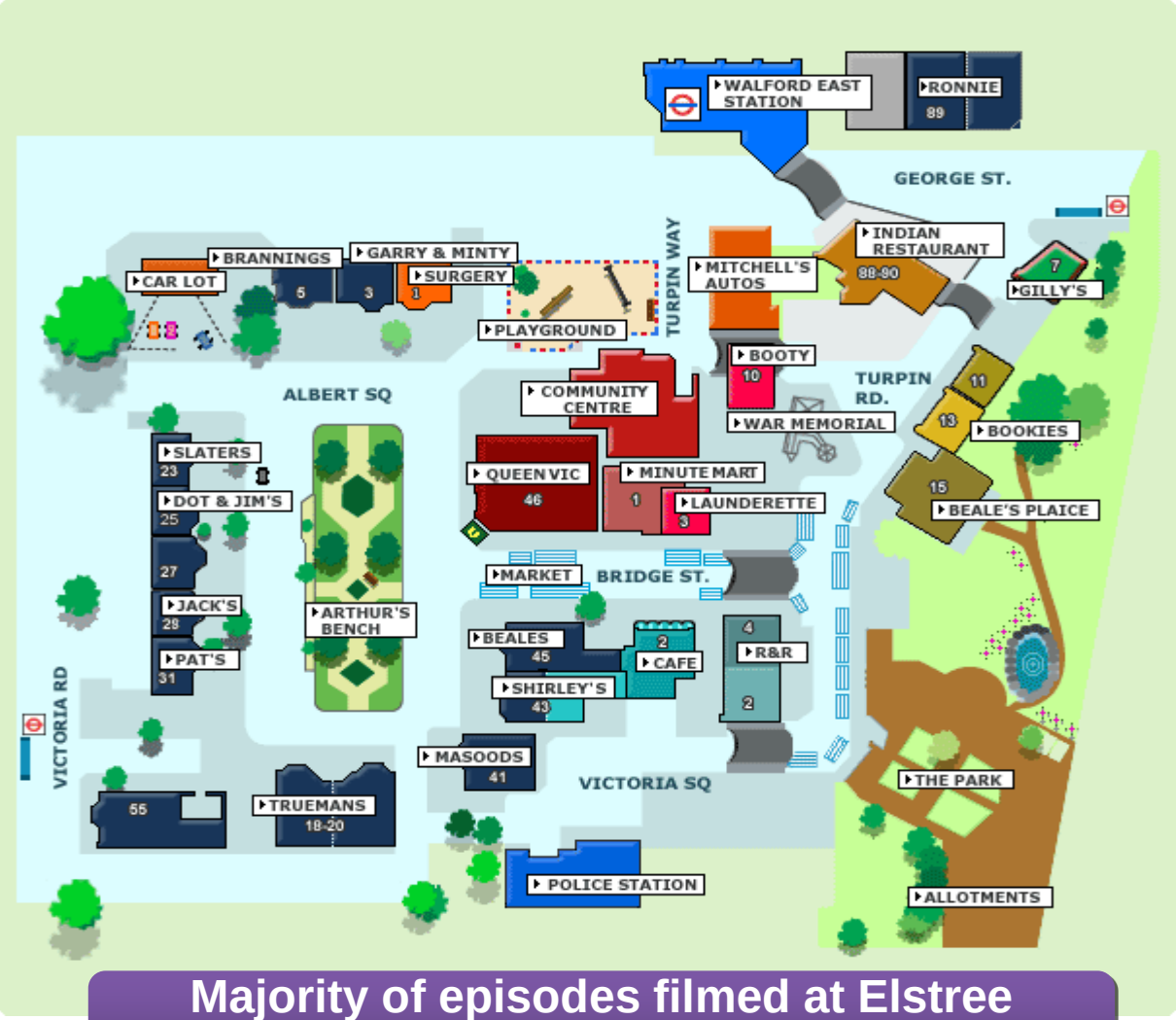
- Given a topic with:
  - 4 example images of the target person
  - 4 Region of Interest (ROI)-masked images of the target person
  - 4 shots from which the target person example images came
  - 6 to 12 image and video examples of a known location
- Return a list of up to 1000 shots ranked by likelihood that they contain the topic target person in the target location
- **Automatic** or **interactive** runs are accepted

Data …

- The British Broadcasting Corporation (BBC) and the Access to Audiovisual Archives (AXES) project made **464 h** of the BBC soap opera EastEnders available for research
  - 244 weekly "omnibus" files (MPEG-4) from 5 years of broadcasts
  - 471527 shots
  - Average shot length: 3.5 seconds
  - Transcripts from BBC
  - Per-file metadata

- Represents a "small world" with a slowly changing set of:
  - People (several dozen)
  - Locales: homes, workplaces, pubs, cafes, open-air market, clubs
  - Objects: clothes, cars, household goods, personal possessions, pets, etc
  - Views: various camera positions, times of year, times of day,
  - Use of fan community metadata allowed, if documented

EastEnders' world



Majority of episodes filmed at Elstree studios. Sometimes filmed on 'location'.

Topic creation procedure @ NIST

- Viewed several test videos to develop a list of recurring people, locations and their overlapping.
- Chose 10 master locations and identified 6 to 12 image and video examples to each depending on location type (private: kitchen, room, etc; public: pub, café, market, etc)
- Created ≈90 topics targeting recurring specific persons in specific locations.
- Chose representative sample of 30 topics. Each topic includes images for target persons from test videos, many from the sample video (ID 0) and a named location.
- Filtered example shots from the submissions if it satisfies the topic.

Global test condition: type of training data

Effect of examples – 2 conditions:

- A – one or more provided images – no video
- E - video examples (+ optional image examples)

Topics – segmented "person" example images

Chelsea

Darrin

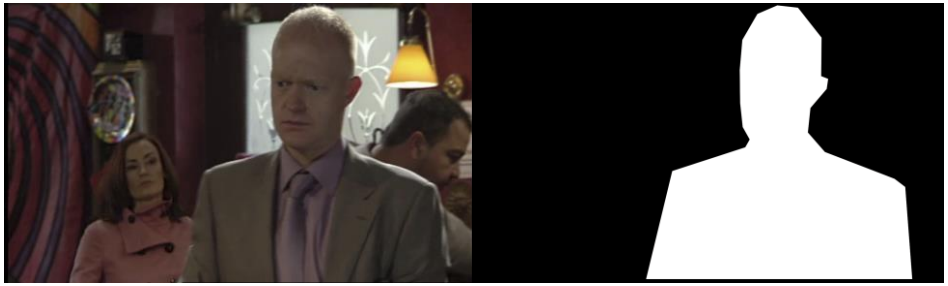
Garry

Heather

Topics – segmented example images

Jack

Jane

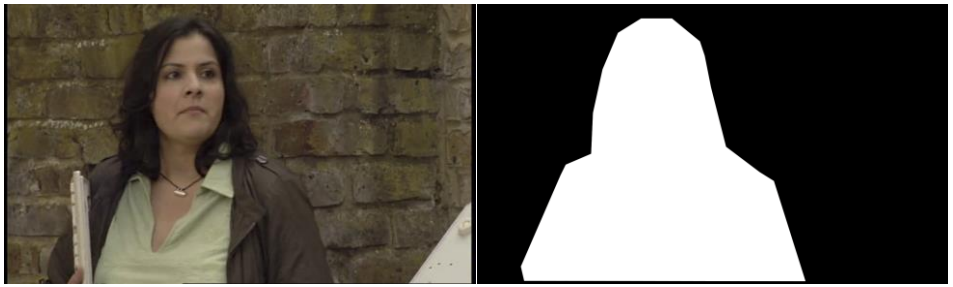


Max

Minty

Topics – segmented example images

Mo



Zainab

Topics – 10 Master locations


Foyer


Kitchen1


Kitchen2


LR1


LR2


Cafe1


Cafe2


Laundrette

Topics – 2018

	Jane	Chelsea	Minty	Garry	Mo	Darrin	Zainab	Heather	Jack	Max
Cafe2	x	x	x	x	x	x	x	x		x
Market	x	x	x				x	x	x	x
Pub	x	x	x	x	x	x			x	
Launderette				x	x	x	x	x	x	x

30 x topics : find {Chelsea, Darrin, Garry, Heather, Jack, Jane, Max, Minty, Mo, Zainab} in {Cafe2,Market,Pub,Launderette}

INS 2018: 8 Finishers (out of 17)

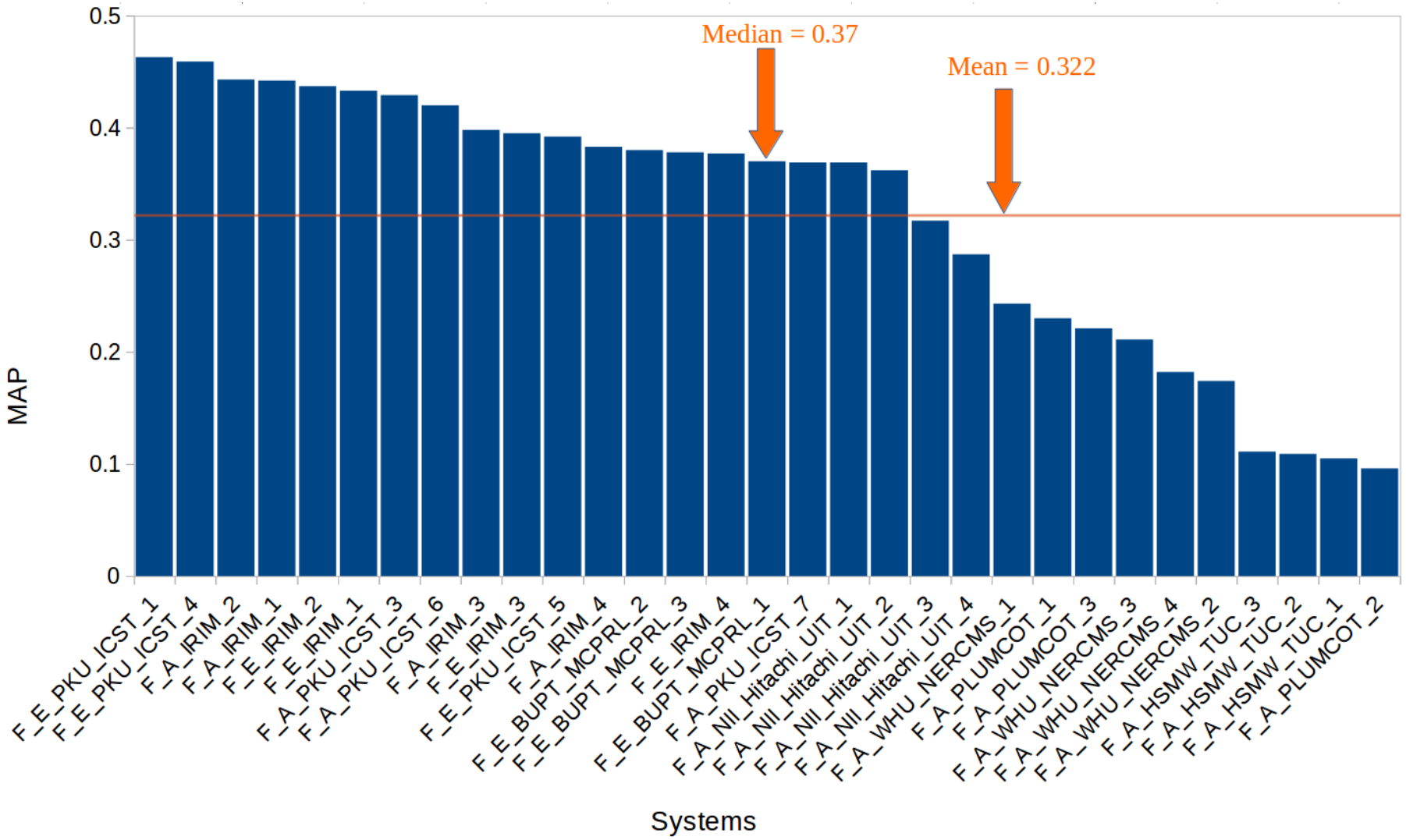
Team	Organization	Run Types Submitted F: automatic, I: Interactive
BUPT_MCPRL	Beijing University of Posts and Telecommunications	F_E (3), I_E (1)
HSMW_TUC	Chemnitz University of Technology, University of Applied Sciences Mittweida	F_A (3), I_A (1)
ITI_CERTH	Information Technologies Institute, Centre for Research and Technology Hellas	I_A (1)
IRIM	EURECOM; LABRI ; LIG ; LIMSI; LISTIC	F_A (4), F_E (4)
NII_Hitachi_UIT	National Institute of Informatics, Japan (NII); Hitachi, Ltd; University of Information Technology, VNU-HCM, Vietnam (HCM-UIT)	F_A (4) , I_A(1)
WHU_NERCMS	National Engineering Research Center for Multimedia Software, Wuhan University	F_A (4) , I_A (4)
PLUMCOT	LIMSI, Karlsruhe Institute of Technology	F_A (3)
PKU_ICST	Peking University	F_A (3), F_E (3), I_E (1)

Evaluation

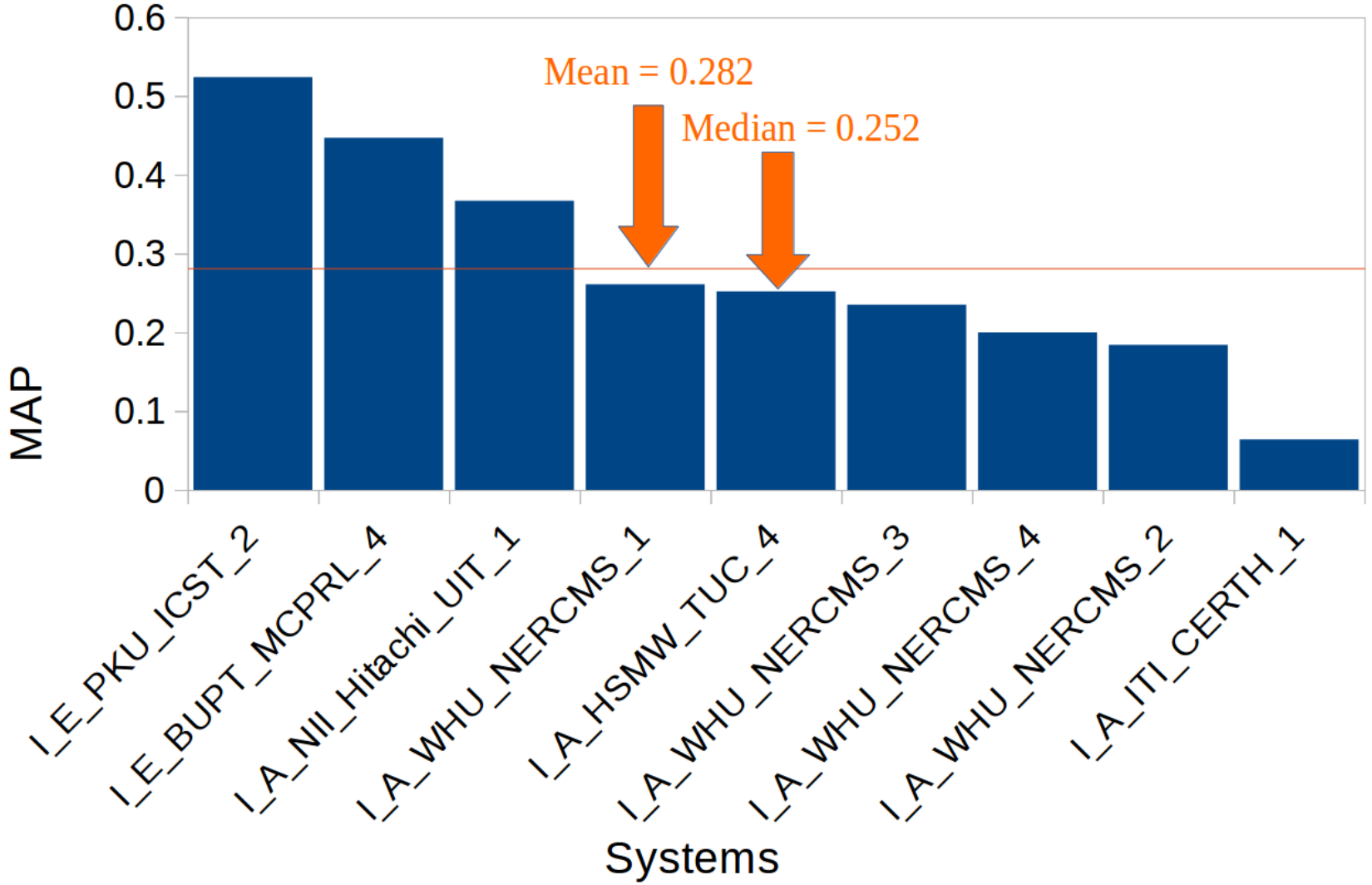
For each topic the submissions were pooled and judged down to max rank 520, resulting in 128 117 judged shots (≈ 480 person-h).

- 10 NIST assessors played the clips and determined if they contained the topic target or not.
- 11 717 clips (avg. 390 / topic) contained the topic target (9 %)
- True positives per topic: min 30 med 168 max 1340
- The task is treated as a form of ranking and thus the trec_eval_video tool was used to calculate average precision, recall, precision, etc.
- To measure efficiency, speed was also measured.
- In total, 31 automatic and 9 interactive runs were submitted.

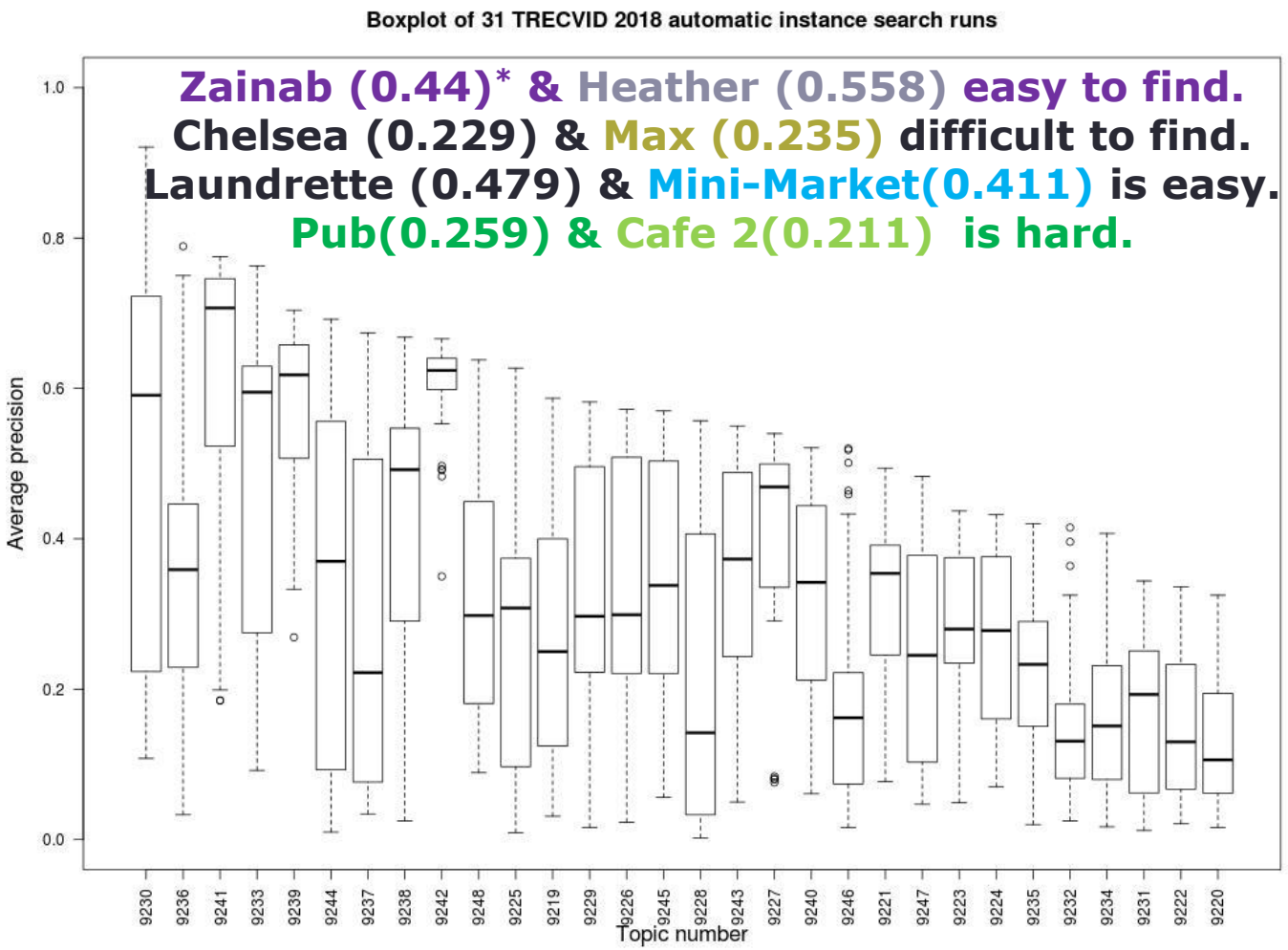
Results by team (Automatic)



Results by team (Interactive)



Results by topic - automatic



#	Query
9230	Find Garry in this Laundrette
9236	Find Darrin in this Laundrette
9241	Find Heather in this Laundrette
9233	Find Mo in this Laundrette
9239	Find Zainab in this Mini-Market
9244	Find Jack in this Laundrette
9237	Find Zainab in this Cafe 2
9238	Find Zainab in this Laundrette
9242	Find Heather in this Mini-Market
9248	Find Max in this Mini-Market
9225	Find Minty in this Cafe 2
9219	Find Jane in this Cafe 2
9229	Find Garry in this Pub
9226	Find Minty in this Pub
9245	Find Jack in this Mini-Market
9228	Find Garry in this Cafe 2
9243	Find Jack in this Pub
9227	Find Minty in this Mini-Market
9240	Find Heather in this Cafe 2
9246	Find Max in this Cafe 2
9221	Find Jane in this Mini-Market
9247	Find Max in this Laundrette
9223	Find Chelsea in this Pub
9224	Find Chelsea in this Mini-Market
9235	Find Darrin in this Pub
9232	Find Mo in this Pub
9234	Find Darrin in this Cafe 2
9231	Find Mo in this Cafe 2
9222	Find Chelsea in this Cafe 2
9220	Find Jane in this Pub

*Mean score of median MAP per character/location

Automatic Run results + Randomization testing

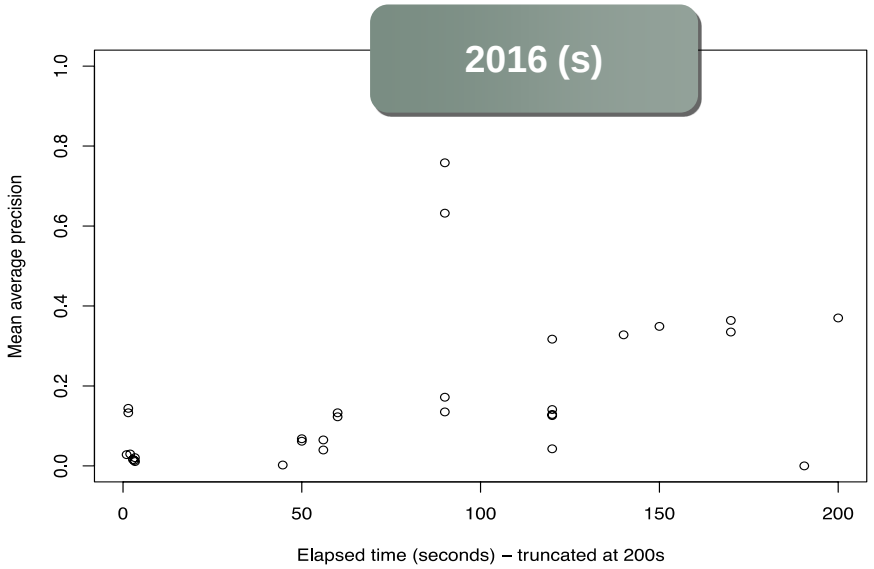
Top 10 runs across all teams (automatic)

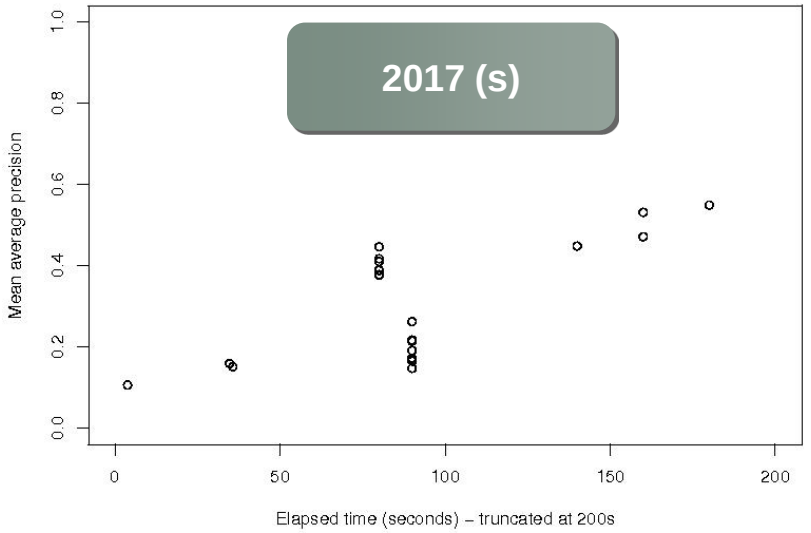
MAP		1	2	3	4	5	6	7	8	9	10
0.463	F_E_PKU_ICST_1	=	>				>	>	>	>	>
0.459	F_E_PKU_ICST_4		=					>	>	>	>
0.443	F_A_IRIM_2			=						>	>
0.442	F_A_IRIM_1				=					>	>
0.437	F_E_IRIM_2					=				>	>
0.433	F_E_IRIM_1						=			>	>
0.429	F_A_PKU_ICST_3							=	>		
0.420	F_A_PKU_ICST_6								=		
0.398	F_A_IRIM_3									=	
0.395	F_E_IRIM_3										=

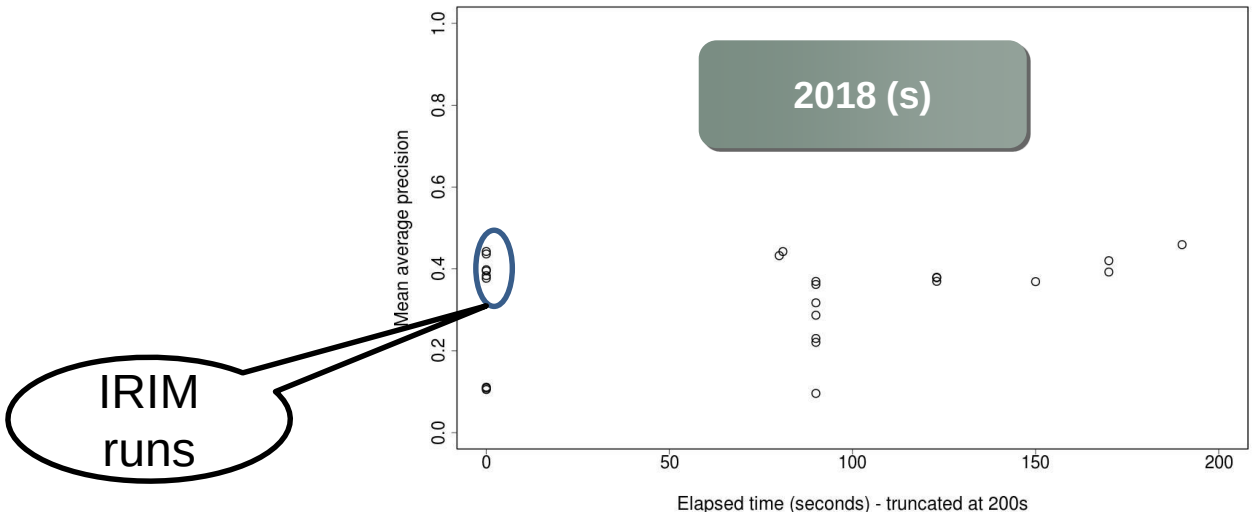
p = probability the row run scored better than the column run due to chance

> p < 0.05

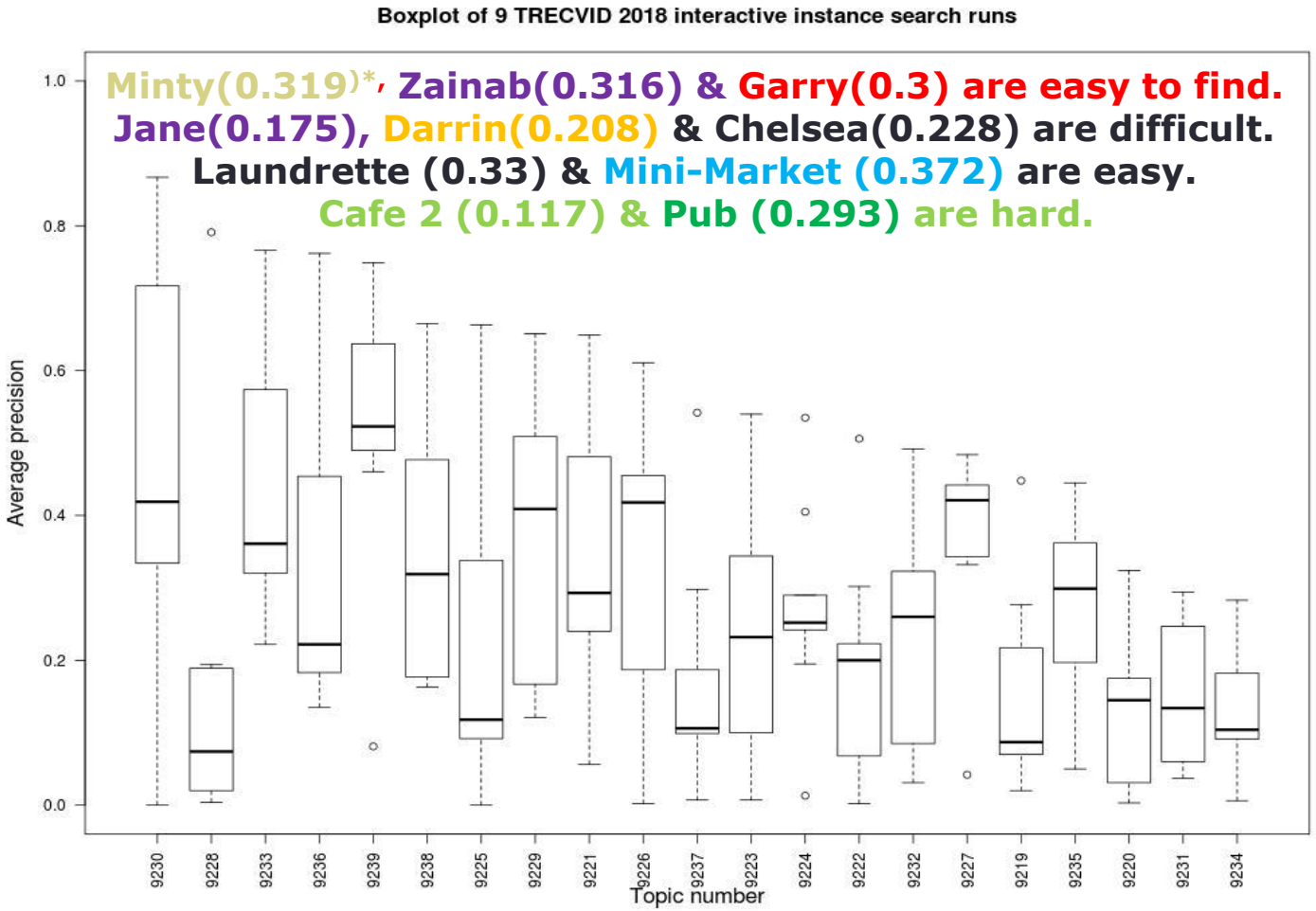
Mean Average Precision vs. per run clock processing time (automatic)







Results by topic - interactive



#	Query
9230	Find Garry in this Laundrette
9228	Find Garry in this Cafe 2
9233	Find Mo in this Laundrette
9236	Find Darrin in this Laundrette
9239	Find Zainab in this Mini-Market
9238	Find Zainab in this Laundrette
9225	Find Minty in this Cafe 2
9229	Find Garry in this Pub
9221	Find Jane in this Mini-Market
9226	Find Minty in this Pub
9237	Find Zainab in this Cafe 2
9223	Find Chelsea in this Pub
9224	Find Chelsea in this Mini-Market
9222	Find Chelsea in this Cafe 2
9232	Find Mo in this Pub
9227	Find Minty in this Mini-Market
9219	Find Jane in this Cafe 2
9235	Find Darrin in this Pub
9220	Find Jane in this Pub
9231	Find Mo in this Cafe 2
9234	Find Darrin in this Cafe 2

*Mean score of median MAP per character/location

Interactive Run Results, Randomization testing

ALL 9 runs by all teams (interactive)

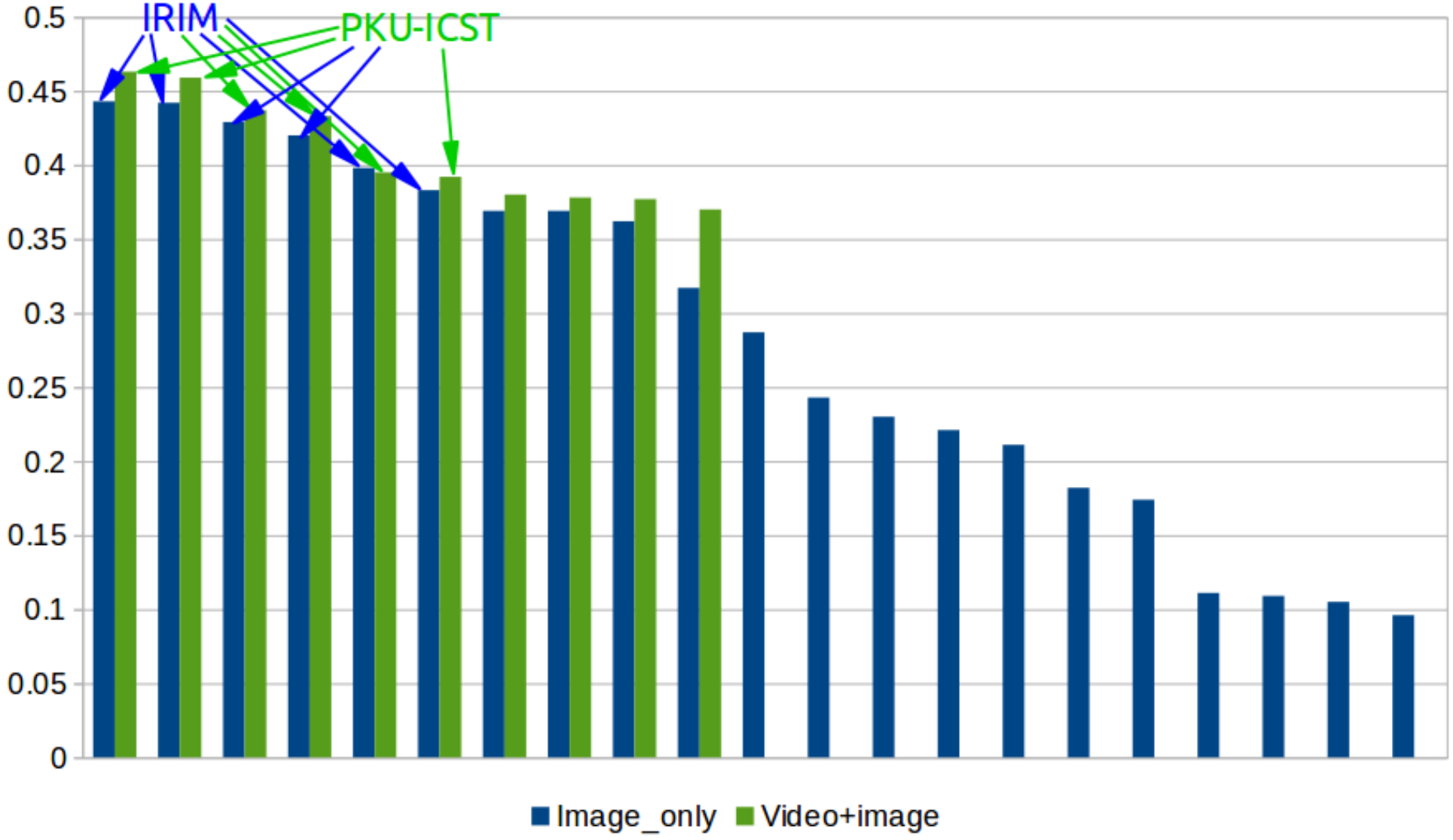
MAP

0.524	I_E_PKU_ICST_2	=	>	>	>	>	>	>	>	>
0.447	I_E_BUPT_MCPRL_4		=	>	>	>	>	>	>	>
0.367	I_A_NII_Hitachi_UIT_1			=	>	>	>	>	>	>
0.261	I_A_WHU_NERCMS_1				=		>	>	>	>
0.252	I_A_HSMW_TUC_4					=		>	>	>
0.235	I_A_WHU_NERCMS_3						=		>	>
0.200	I_A_WHU_NERCMS_4							=	>	>
0.184	I_A_WHU_NERCMS_2								=	>
0.064	I_A_ITI_CERTH_1									=
		1	2	3	4	5	6	7	8	9

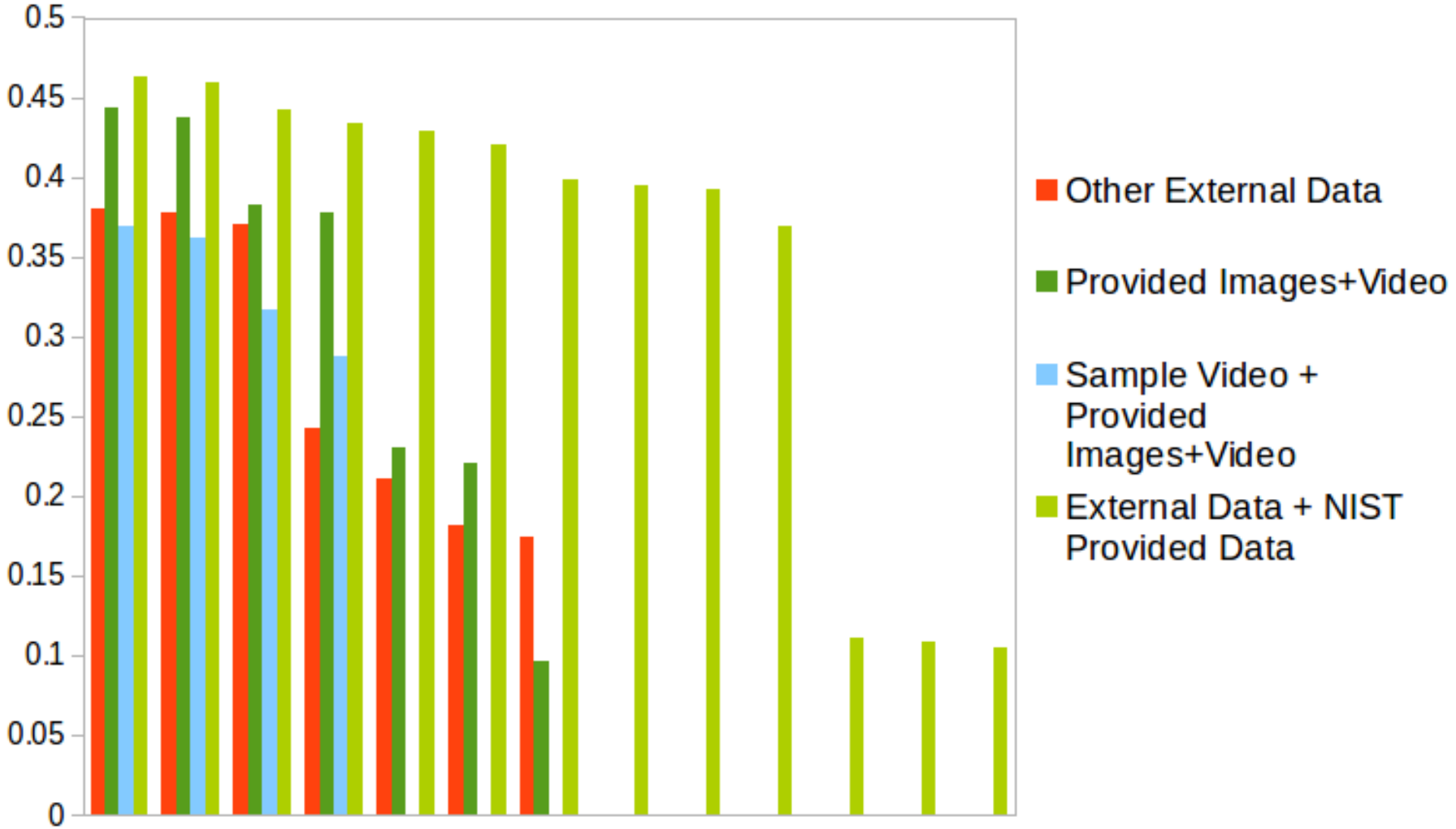
p = probability the row run scored better than the column run due to chance

> p < 0.05

Results by example set (A/E) - automatic



Results by Data Source



Some general observations about the task

- Slight decrease in number of participants but same number of finishers – higher % finished.
- Less teams are using E condition - training with video examples – (e.g tracking characters)
- Interactive search task:
  - Limited participation
- Third year: Slight decrease in best performances from 2nd year – Why? Queries more difficult?
- We encourage teams to test their 2016 or 2017 system on the 2018 topics or vice versa.

Some general observations about the task – Data Source

- Best results achieved using external data plus NIST provided data.
- Next best results are achieved using only the NIST provided images and video.
- Systems using only external data do not perform as well as systems which include the NIST provided data.

Observations over last three years of the task (Automatic):

- High Score down on last year but still up on 2016.
- Low Scores increasing year on year.
- Mean and Median scores increased significantly last year on 2016 but have since stabilized.
- Standard Deviation between MAP scores decreased on last year, now similar to 2016.
- Number of participants has been decreasing year on year.

Observations over last three years of the task - Locations

- Laundrette consistently shows to be among the easiest locations to recognize.
- Pub consistently shows to be among the most difficult locations to recognize.

	2016	2017	2018
Pub	0.021	N/A	0.259
Laundrette	0.172	0.338	0.479
Market	N/A	0.343	0.411

Average scores (automatic systems) across all topics for common locations per year

BUPT-MCPRL

- Location Retrieval: Two Independent Methods:
  - Hessian-Affine detector with RootSIFT, MSER detector with RootSIFT, and CNN features.
  - Fine-Tuned publicly available VGG-16 model, GoogleNet model, and ResNet-152 models.
- Person Retrieval: Face retrieval and transcript-based search:
  - Detect face on key frames captures from video by MTCNN, face representations extracted from bounding-box and cosine distance is employed to match faces.
  - Transcript search - locate character name in transcripts.
- Submitted 4 runs
  - Three automatic runs
  - One interactive run

ITI CERTH

- Focus on interactive task
- VERGE system includes several modes for navigation:
  - Visual similarity (DCNN)
  - Visual Concept Retrieval - 346 visual concepts
  - Face detection
  - Scene similarity
  - Multimodal Fusion
- Late fusion of DCNN face descriptors and scene descriptors
- Submitted 1 interactive run

TU Chemnitz

- Complete overhaul of INS system architecture using Docker
- Allows combination of various open face recognition and scene recognition pipelines
- All indexing is done offline, retrieval is very fast

Submitted 3 automatic and 1 interactive runs

IRIM (LaBRI, LIMSI, LIG)

- Combination of two person recognition methods
- One location recognition method
- Late fusion on person methods, additional late fusion to mix in the location scores
- 2018: focus on person recognition
  - Positive impact: data augmentation, faces reranking

Submitted 8 automatic runs (A and E)

Peking University (ICST)

- Location search: BOW plus CNN
- Person search:
  - Query preprocessing based on super resolution
  - Deep models for face recognition
  - Text based refinement
- Fusion based on combination of score and rank based fusion (boosting)
  - Filtering noisy shots (outliers)

Submitted 6 automatic runs (A and E), 1 interactive

Overview of submissions (1)

- 8 out of 8 teams described INS runs for the TV notebook
- 3 teams will present their INS experiments

3:40 - 4:10, (**HSMW_TUC – University of Applied Sciences Mittweida, Chemnitz University of Technology**)

4:10 - 4:40, (**NII_Hitachi_UIT – National Institute of Informatics, Japan Hitachi, Ltd., Japan University of Information Technology, VNU_HCMC, Vietnam**)

4:40 - 5:10, (**WHU_NERCMS – National Engineering Research Centre for Multimedia Software, Wuhan University**)

5:10 - 5:25, INS Discussion

INS 2019 plans

- Move on to a new query type
  - Action instances (drinking, walking, sleeping, talking, driving, etc)
  - Person + Action (Brad fighting)
  - Action + Location (e.g drinking in the cafe)
  - Mix of the above ?!
- Keep the newly added additional training data sources.
- Add manual run type ?

INS 2019 plans – Example Action Images

Jack Holding Money

Heather Talking on Phone

Ian Drinking in Cafe

Garry Holding Flowers

Minty and Mo Singing

Jane Holding Baby

Future INS plans – Common queries

- Plans to now include a set of common queries each year to measure yearly progress.
- 2019 teams will submit runs for 50 queries in total. 30 unique queries plus 20 common queries to be repeated each year.
- 2020 and 2021 teams to submit runs for 40 queries in total. 20 unique queries plus 20 common queries each year.
- The common queries each year provide a basis for the comparison of team performances year on year.

Future INS plans – Evaluations

	2019	2020	2021
2019	30 + 10A + 10B		
2020		20 + 10A + 10B	
2021			20 + 10A + 10B

- 2019 Assessors to evaluate just the 30 unique queries for that year.
- 2020 Assessors to evaluate the 20 unique queries for that year plus the first 10 common queries (10A), allows to measure progress on 2019.
- 2021 Assessors to evaluate the 20 unique queries plus the second 10 common queries (10B), allows to measure progress on 2019 and 2020.