TRECVID 2021
Ad-hoc Video Search (AVS) Task Overview

Georges Quénot
Laboratoire d'Informatique de Grenoble, France

George Awad
Retrieval Group, Information Access Division, Information Technology Laboratory, NIST;
Georgetown University

National Institute of Standards and Technology
U.S. Department of Commerce

Information Access Division
Information Technology Laboratory

Outline

Task Definition & Dataset
Topics (Queries)
Participating Teams
Evaluation & Results
General Observations

Task Definition

Goal: promote progress in content-based video retrieval based on end user **ad-hoc (generic) textual queries** that include searching for persons, objects, locations, actions and their combinations.

Task: Given a test collection, a query, and a master shot boundary reference, return a ranked list of at most 1000 shots (out of 1,082,657) which best satisfy the query.

Queries:

- Main : New queries each year
- Progress : A set of fixed queries for 3 years

Testing data: 7475 Vimeo Creative Commons Videos (V3C1), 1000 total hours with mean video durations of 8 min. Reflects a wide variety of content, style and source devices. Fixed testing data since 2019.

Development data: ≈2000 hours of previous IACC.1-3 (Internet Archive) data used between 2010-2018 with concept and ad-hoc query annotations.

Vimeo Creative Commons Collection

Partition	V3C1	V3C2	V3C3	Total
File Size	2.4TB	3.0TB	3.3TB	8.7TB
Number of Videos	7,475	9,760	11,215	28,450
Combined Video Duration	1000 hours, 23 minutes, 50 seconds	1300 hours, 52 minutes, 48 seconds	1500 hours, 8 minutes, 57 seconds	3801 hours, 25 minutes, 35 seconds
Mean Video Duration	8 minutes, 2 seconds	7 minutes, 59 seconds	8 minutes, 1 seconds	8 minutes, 1 seconds
Number of Segments	1,082,659	1,425,454	1,635,580	4,143,693

The Vimeo Creative Commons Collection (V3C)* consists of '**free**' video material sourced from the web video platform **vimeo.com**. *It is designed to contain a wide range of content which is representative of what is found on the platform in general*. All videos in the collection have been released by their creators under a **Creative Commons License** which allows for unrestricted redistribution.

* Rossetto, L., Schuldt, H., Awad, G., & Butt, A. (2019). V3C – a Research Video Collection. *Proceedings of the 25th International Conference on MultiMedia Modeling*.

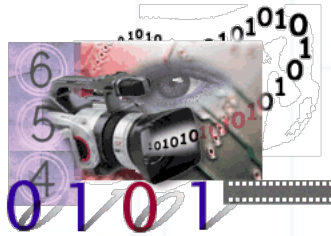
AVS 2021 (20 main) Queries by complexity

Query	Person	Action	Object	Location
Find shots of a hang glider floating in the sky on a sunny day		✓	✓	
Find shots of a woman wearing sleeveless top	✓			
Find shots of a person with a tattoo on their arm	✓			
Find shots of city street where ground is covered by snow			✓	✓
Find shots of an adult person wearing a backpack and walking on a sidewalk	✓	✓	✓	✓
Find shots of a man wearing a blue jacket	✓		✓	
Find shots of a person looking at themselves in a mirror	✓	✓	✓	
Find shots of a person wearing an apron indoors	✓	✓	✓	✓
Find shots of a woman holding a book	✓	✓	✓	
Find shots of a person painting on a canvas	✓	✓	✓	
Find shots of a man behind a pub bar or club bar	✓			✓
Find shots of a person wearing a cap backwards	✓		✓	
Find shots of a man pointing with his finger	✓	✓		
Find shots of a parachutist descending towards a field on the ground in the daytime	✓	✓		✓
Find shots of two or more ducks swimming in a pond		✓	✓	✓
Find shots of a white dog			✓	
Find shots of two boxers in a ring	✓			✓
Find shots of a man sitting on a barber chair in a shop	✓	✓	✓	✓
Find shots of a ladder with less than 6 steps			✓	
Find shots of a bow tie			✓	

2019-2021 (20 progress) Queries by complexity

Query	Person	Action	Object	Location
Find shots of a person holding an opened umbrella outdoors	✓	✓	✓	✓
Find shots of two people talking to each other inside a moving car	✓	✓	✓	✓
Find shots of people walking across (not down) a street in a city	✓	✓		✓
Find shots of a shark swimming under the water		✓	✓	✓
Find shots of a person reading a paper including newspaper	✓	✓	✓	
Find shots of fishermen fishing on a boat	✓	✓	✓	
Find shots of a person jumping with a motorcycle	✓	✓	✓	
Find shots of a person jumping with a bicycle	✓	✓	✓	
Find shots of one or more women models on a catwalk demonstrating clothes	✓	✓		
Find shots of people doing yoga	✓	✓		
Find shots of a person sleeping	✓	✓		
Find shots of people hiking	✓	✓		
Find shots of bride and groom kissing	✓	✓		
Find shots of a person skateboarding	✓	✓		
Find shots of people queuing	✓	✓		
Find shots of two people kissing who are not bride and groom	✓	✓		
Find shots of a man in a clothing store	✓			✓
Find shots of a person in a bedroom	✓			✓
Find shots of a person's shadow			✓	
Find shots showing electrical power lines			✓	

Task Parameters

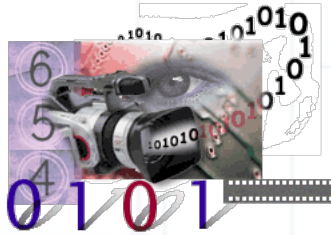


System Types	Description
Fully Automatic (F)	System uses official query directly
Manually-Assisted (M)	Query built manually
Relevance-Feedback (R)	Allow judging top-30 results up to 3 iterations

Training data categories	Description
A	Only IACC training data
D	Other training data sources
E	Only training data collected ***automatically*** using the query text
F	Only training data collected ***automatically*** using a query ***built manually*** from the official query text

->> Novelty (optional) run type to encourage retrieving non-common relevant shots easily found across systems.
->> Explainability of result items were allowed as extra optional information with the submitted shots

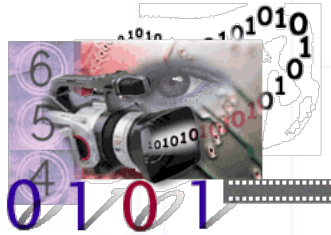
Teams – Main Task (39 runs)



DIGITAL VIDEO RETRIEVAL at NIST

Team (8 Finishers)	Organization	System Type		
		Manually assisted	Fully automatic	Novelty run
VIREO	Singapore Management University; City University of Hong Kong	4	4	1
Kindai_ogu_osaka	Kindai University; Osaka Gakuin University; Osaka University		4	1
GodSpeed	Kuaishou Tech		4	
ITI_CERTH	Information Technologies Institute, Centre for Research and Technology Hellas		4	
RUC_AIM3	Renmin University of China		4	
RUCMM	Renmin University of China		4	
WasedaMeiseiSoftbank	Waseda University; Meisei University; SoftBank Corporation	4	4	
DMT_CUC_01	Communication University of China	1		

Teams – Progress Task (112 runs)



Team 14 Finishers	Organization	System Type		
		Manually assisted	Fully automatic	Novelty run
VIdeoREtrievalGrOup	City University of Hong Kong	10	12	
FIU_UM	Florida International University; University of Miami		6	
Kindai_ogu	Kindai University; Osaka Gakuin University		9	1
SIRET (2019)*	Charles University	4		
ATL (2019)*	Alibaba group; ZheJiang University		4	
Inf (2019)*	Carnegie Mellon University; Monash University; Renmin University; Shandong University		4	
EURECOM (2019)*	EURECOM		3	
ITI_CERTH	Information Technologies Institute, Centre for Research and Technology Hellas		5	
RUC_AIM3	Renmin University of China		8	
RUCMM	Renmin University of China		12	
WasedaMeiseiSoftbank	Waseda University; Meisei University; SoftBank Corporation	12	9	
ZY_BJLAB	XinHuaZhiYun Technology CO,. Ltd.	4	4	
GodSpeed (2021)*	Kuaishou Tech		4	
DMT_CUC_01 (2021)*	Communication University of China	1		

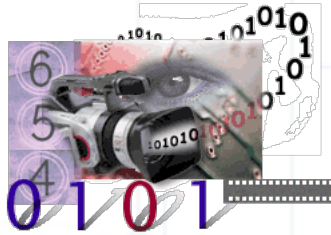
*Some teams only participated in 2019 or 2021

Evaluation Methodology

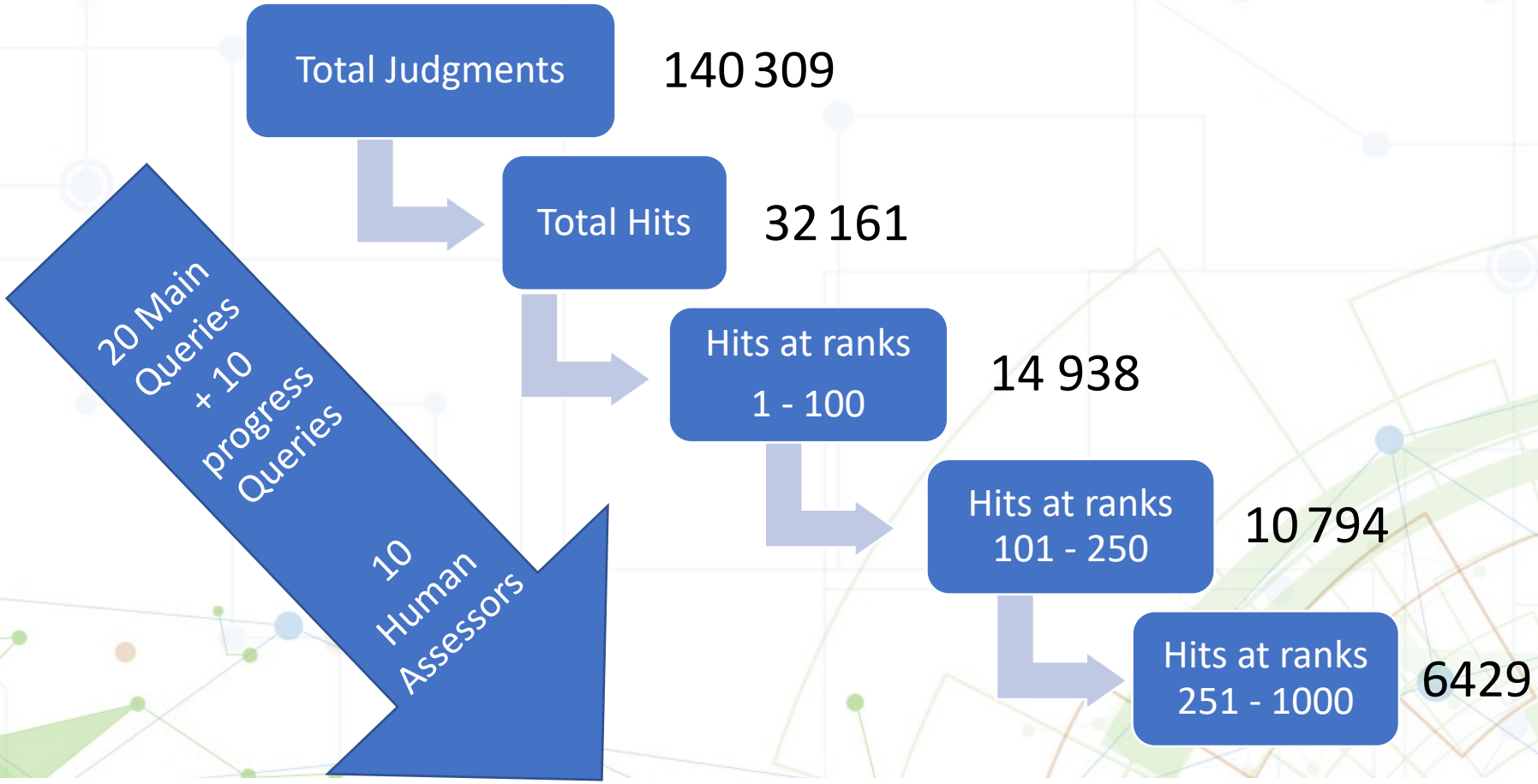
- NIST judged 100% of top (ranks 1 – 250) pooled results from all submissions and sampled 20% from the rest of pooled results (ranks 251 – 1000).
- Stats of sampled and judged clips (ranks 251 to 1000) across all runs and topics
  - At minimum, 16.9 % of any run and query results were sampled and judged
  - At maximum, 94.9 % of any run and query results were sampled and judged
  - On average, 73.9 % of any run and query results were sampled and judged
- One assessor per query, watched complete shot while listening to the audio.
- Each query assumed to be binary: absent or present for each master reference shot.
- Top submitted results were *double judged* if at least 10 runs submitted them, and assessor judged them as false positive.
- Extended inferred average precision (xinfAP) was calculated using the judged and unjudged pool by sample_eval[1] tool.
- Compared runs in terms of **mean** extended *inferred average precision* across the all evaluated queries.

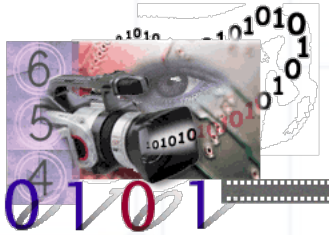
[1]https://www-nlpir.nist.gov/projects/trecvid/trecvid.tools/sample_eval/

Human Judgments



DIGITAL VIDEO RETRIEVAL at NIST





DIGITAL VIDEO RETRIEVAL at NIST

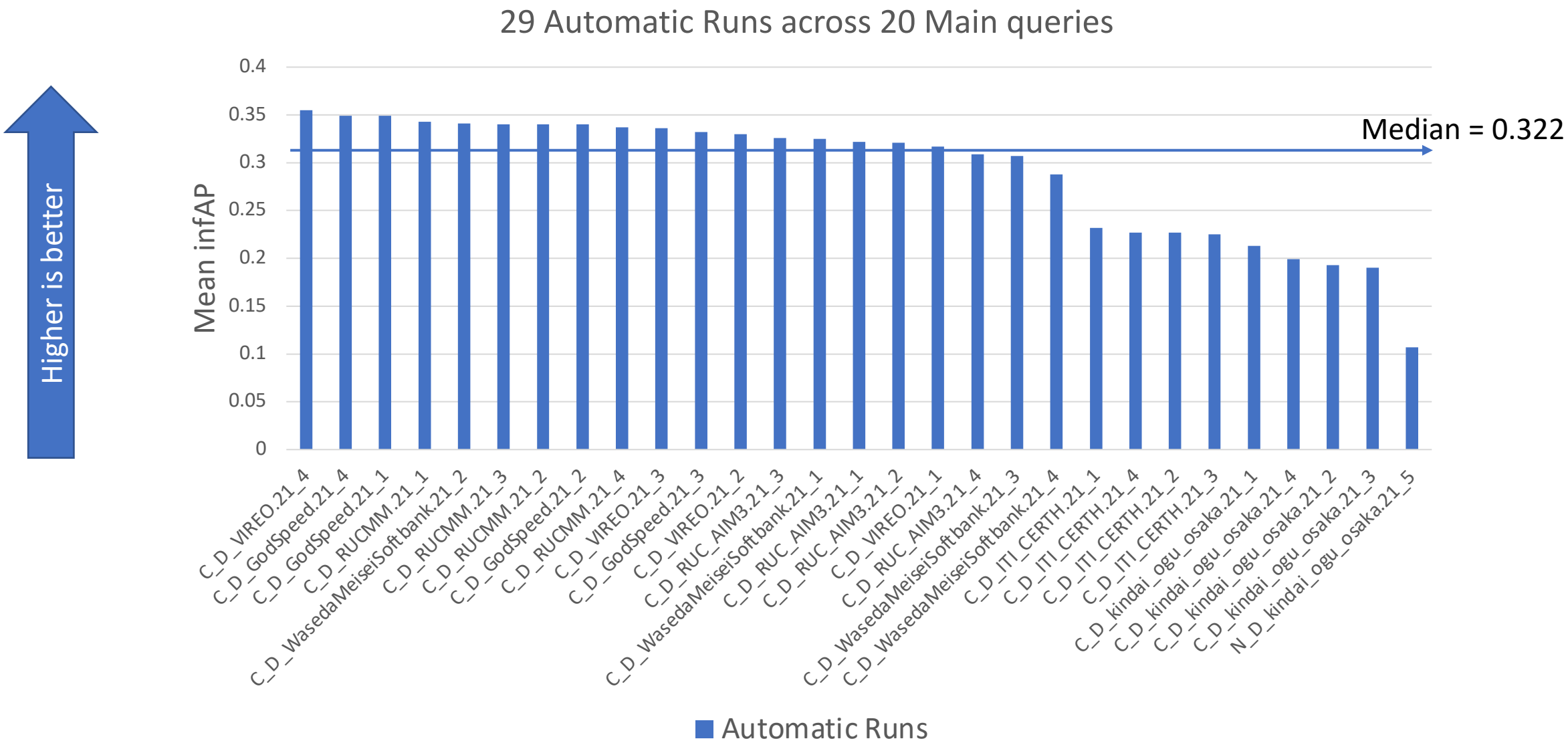
Main Task Results

Inferred average precision (InfAP)

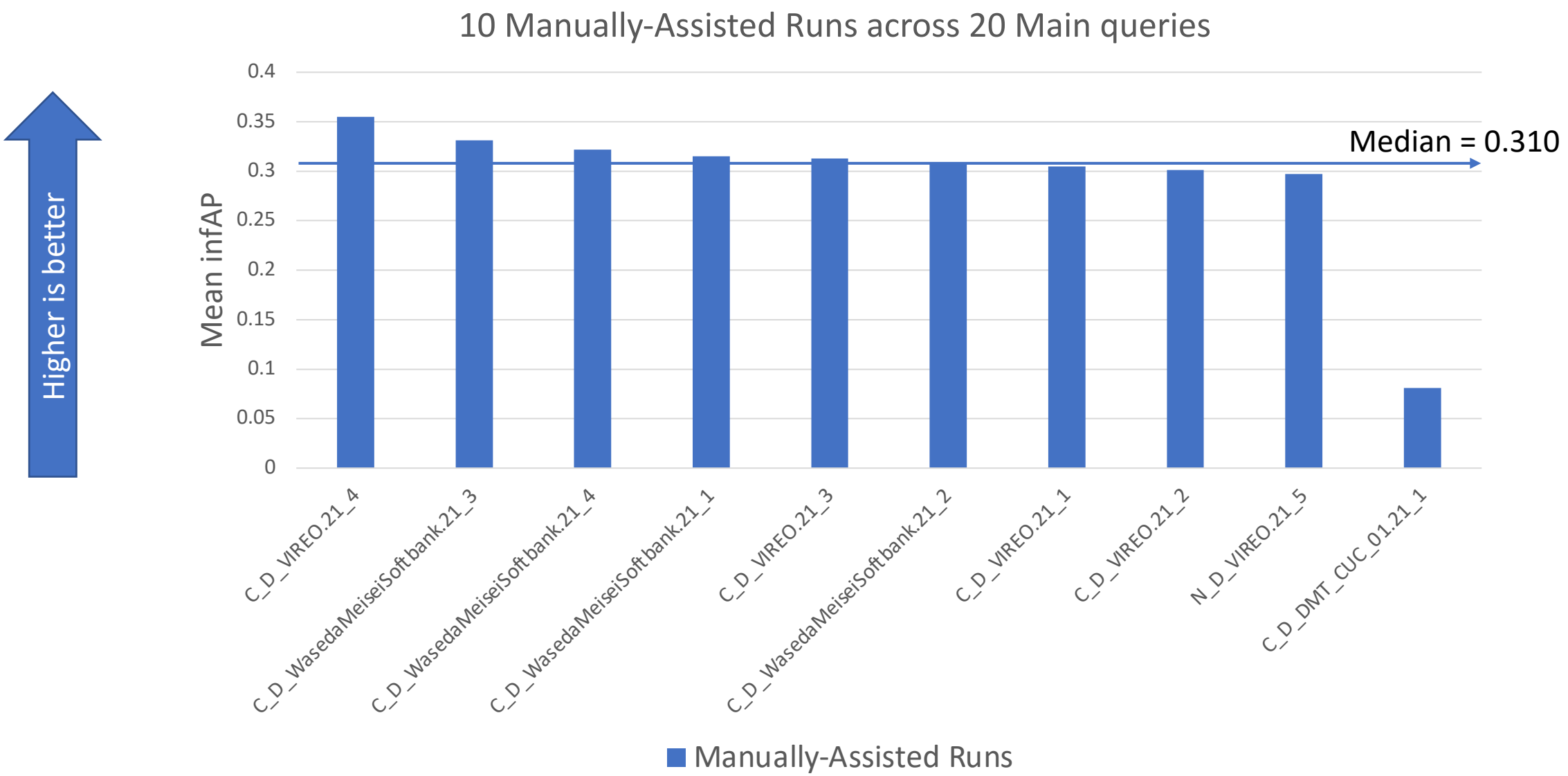
- Estimates average precision well using a small sample of judgments from the usual submission pools*
- Thus, more queries can be judged with same annotation effort.
- Experiments on previous TRECVID years confirmed the quality of the estimate in terms of actual scores and system ranking.
- Extended InfAP (xinfAP) allows the adjustment of sampling to match the relative importance of highest ranked items to average precision.

* J.A. Aslam, V. Pavlu and E. Yilmaz, Statistical Method for System Evaluation Using Incomplete Judgments Proceedings of the 29th ACM SIGIR Conference, Seattle, 2006.

Sorted Overall Scores



Sorted Overall Scores



Statistical Significance (top 10 runs)

Top 10 automatic runs - randomization test ($p < 0.05$)

Rank	Run	xInfAP (sorted scores)
1	C_D_VIREO.21_4	0.355
2	C_D_GodSpeed.21_4	0.349
3	C_D_GodSpeed.21_1	0.349
4	C_D_RUCMM.21_1	0.343
5	C_D_WasedaMeiseiSoftbank.21_2	0.341
6	C_D_RUCMM.21_3	0.340
7	C_D_RUCMM.21_2	0.340
8	C_D_GodSpeed.21_2	0.340
9	C_D_RUCMM.21_4	0.337
10	C_D_VIREO.21_3	0.336

C_D_VIREO.21_4 > C_D_VIREO.21_3

- VIREO run 4 is better than run 3.
- No significant difference between runs in rank 2 to 9.

Statistical Significance

Top 10 manually-assisted runs - randomization test ($p < 0.05$)

Run	xInfAP
C_D_VIREO.21_4	0.355
C_D_WasedaMeiseiSoftbank.21_3	0.331
C_D_WasedaMeiseiSoftbank.21_4	0.322
C_D_WasedaMeiseiSoftbank.21_1	0.315
C_D_VIREO.21_3	0.313
C_D_WasedaMeiseiSoftbank.21_2	0.308
C_D_VIREO.21_1	0.305
C_D_VIREO.21_2	0.301
N_D_VIREO.21_5	0.297
C_D_DMT_CUC_01.21_1	0.081

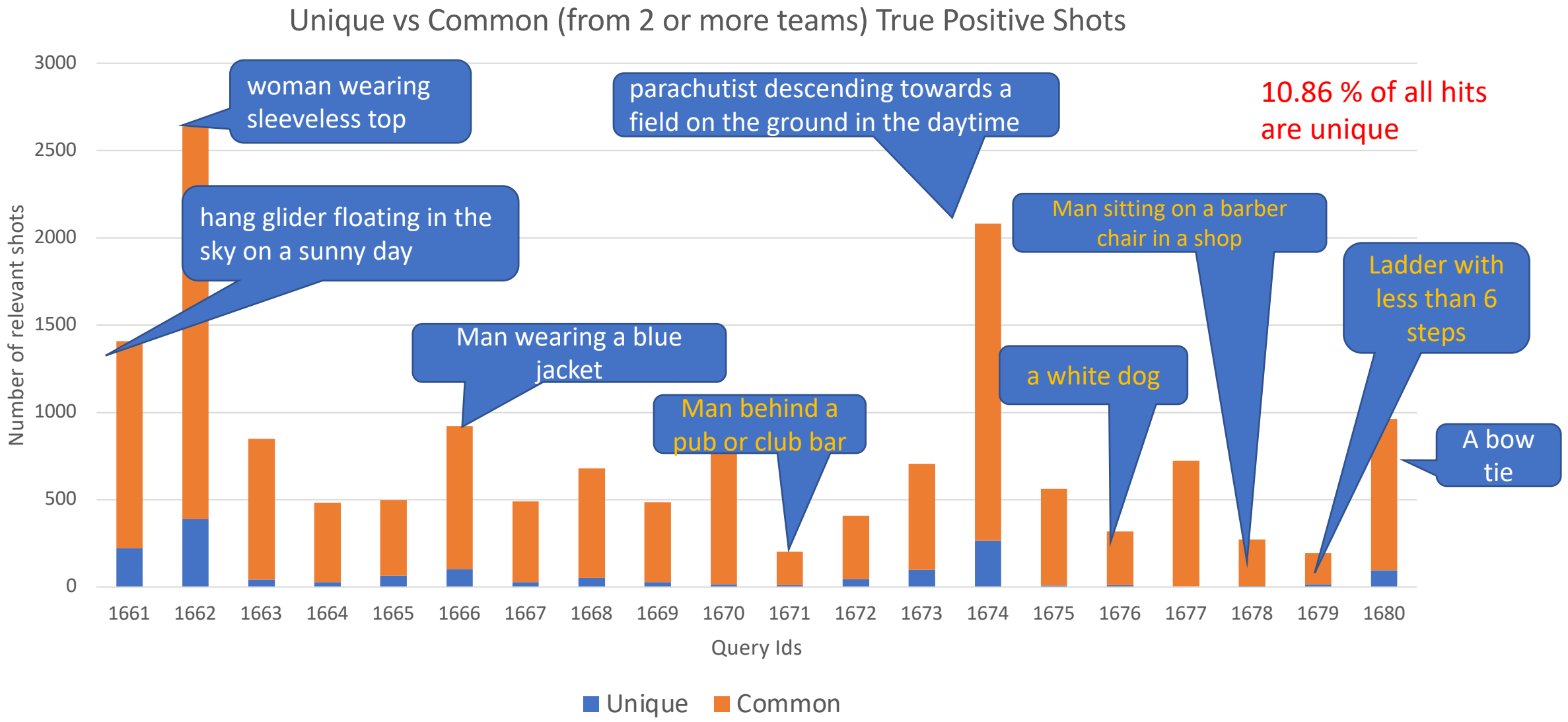
Hierarchy of significant differences between runs

- C_D_VIREO.21_4
  - C_D_VIREO.21_2
    - C_D_DMT_CUC_01.21_1
  - C_D_VIREO.21_3
    - C_D_VIREO.21_1
      - C_D_DMT_CUC_01.21_1
  - C_D_VIREO.21_5
    - C_D_DMT_CUC_01.21_1
- C_D_WasedaMeiseiSoftbank.21_3
  - C_D_WasedaMeiseiSoftbank.21_4
    - C_D_WasedaMeiseiSoftbank.21_2
      - C_D_DMT_CUC_01.21_1
  - C_D_WasedaMeiseiSoftbank.21_1
    - C_D_DMT_CUC_01.21_1

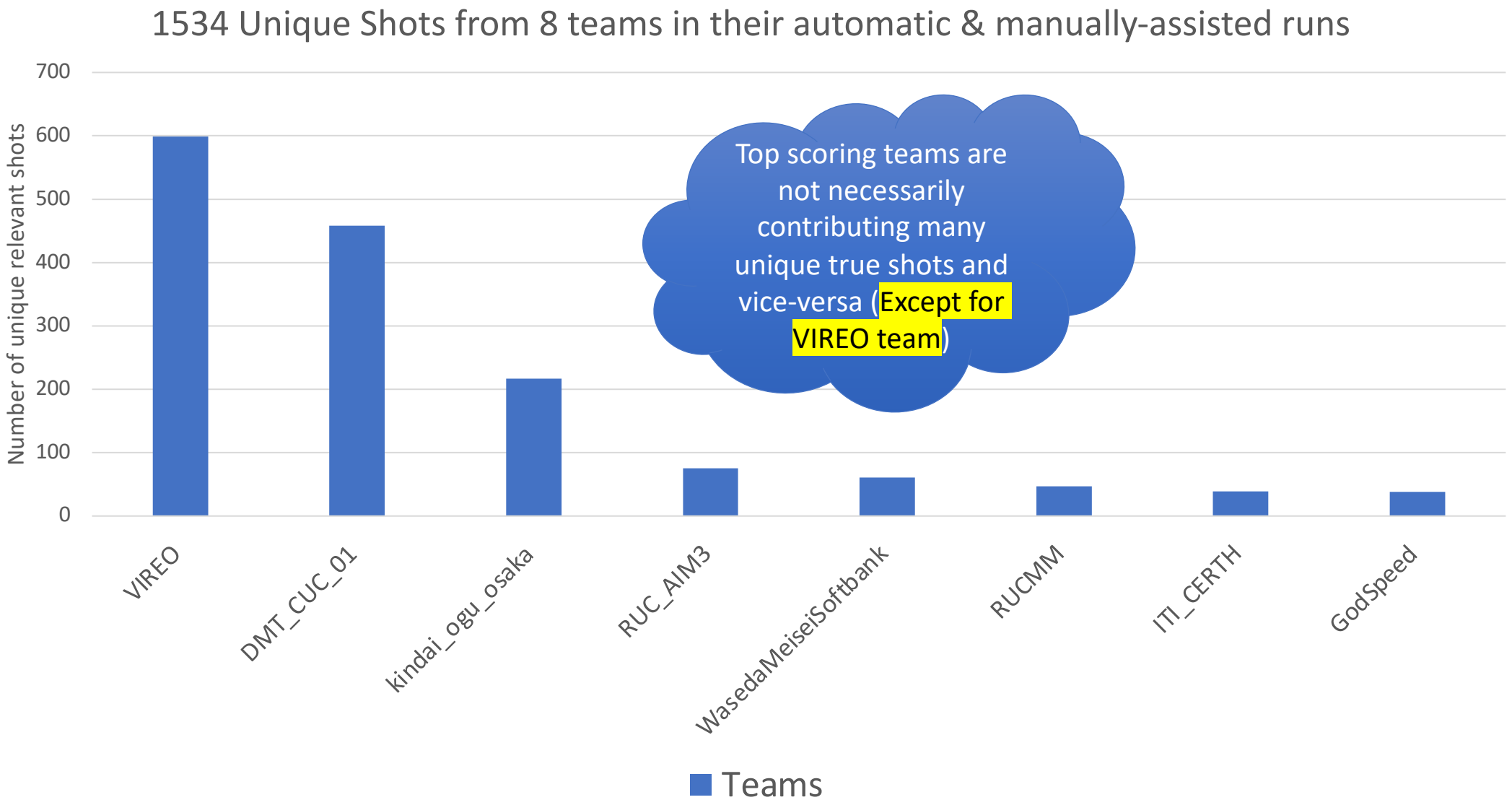
Indentation levels indicate significant difference. Outer indented run is better than inner indented run(s)

- No significant difference between VIREO run 4 and WasedaMeiseiSoftbank run 3.
- VIREO run 4 is better than all other VIREO runs.
- WasedaMeiseiSoftbank run 3 is better than all other Waseda runs.
- All runs are significantly better than DMT_CUC_01 run 1.

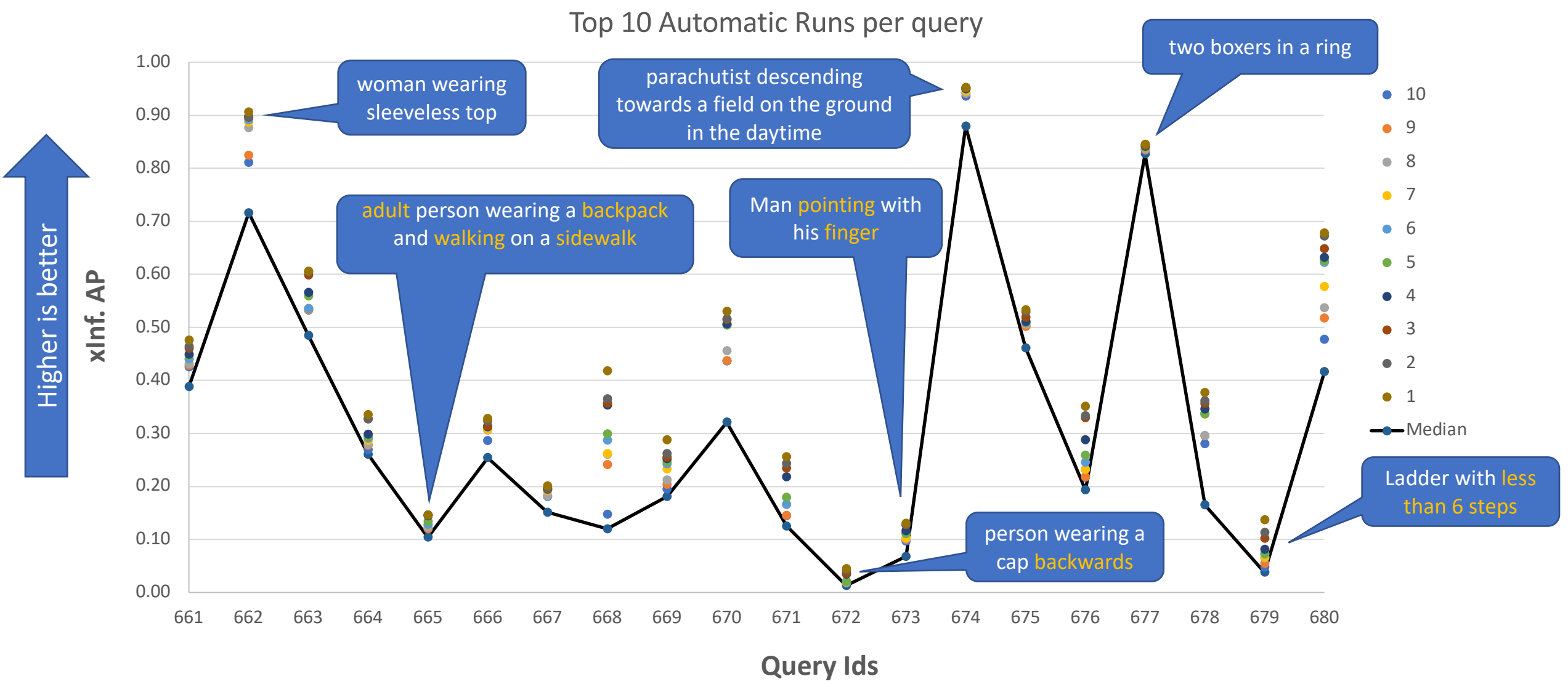
Hits Per Topic (Main Task)



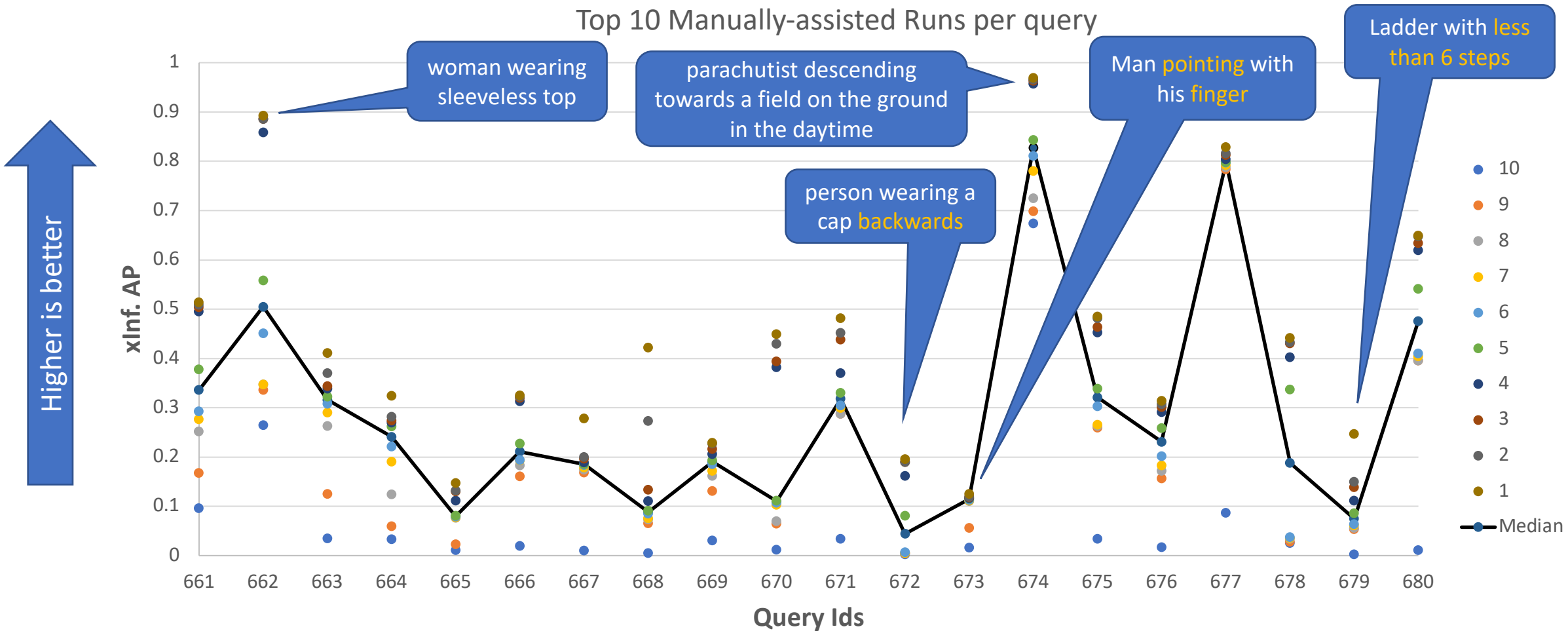
Sorted Unique Hits by Team



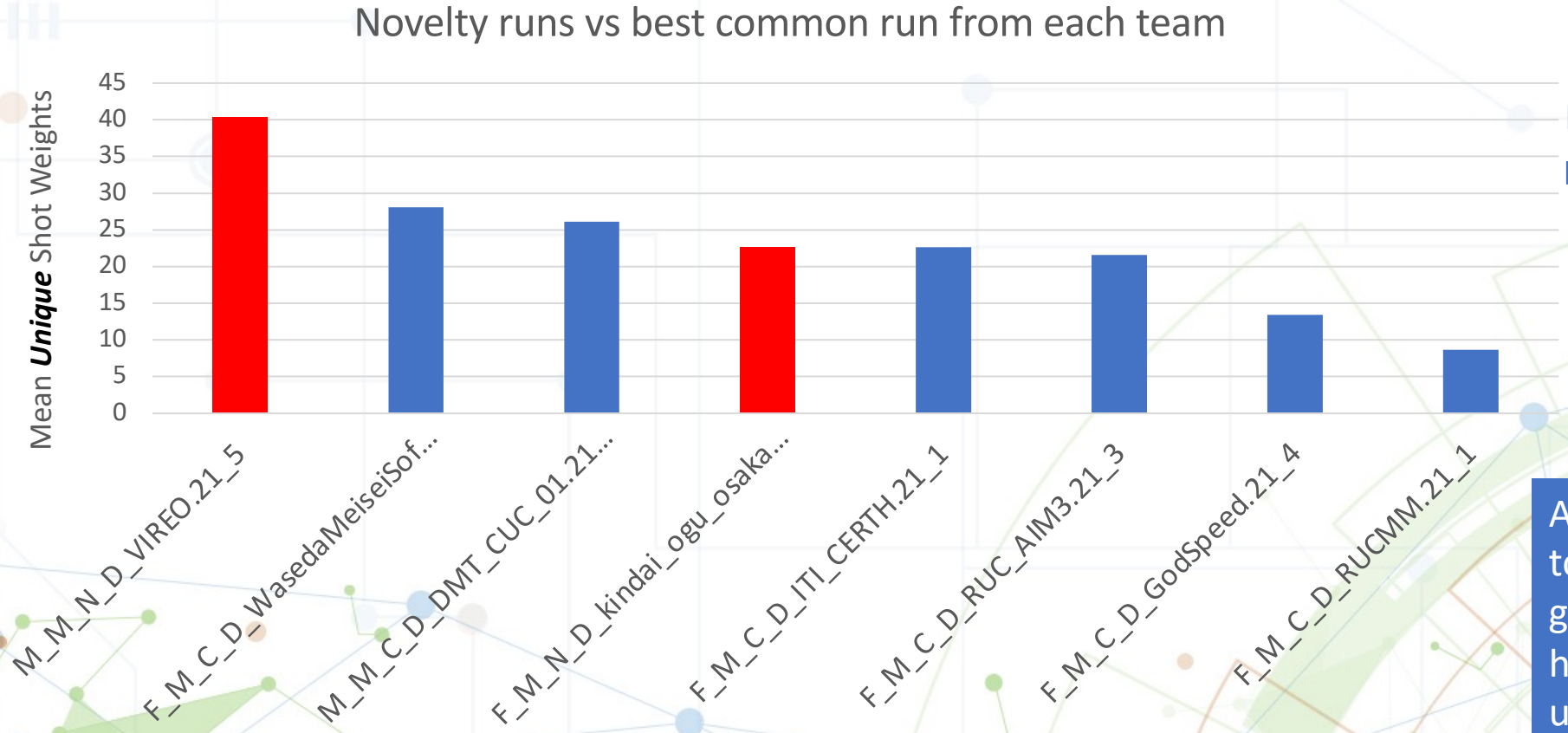
Top runs per query (Main Task)



Top runs per query (Main Task)

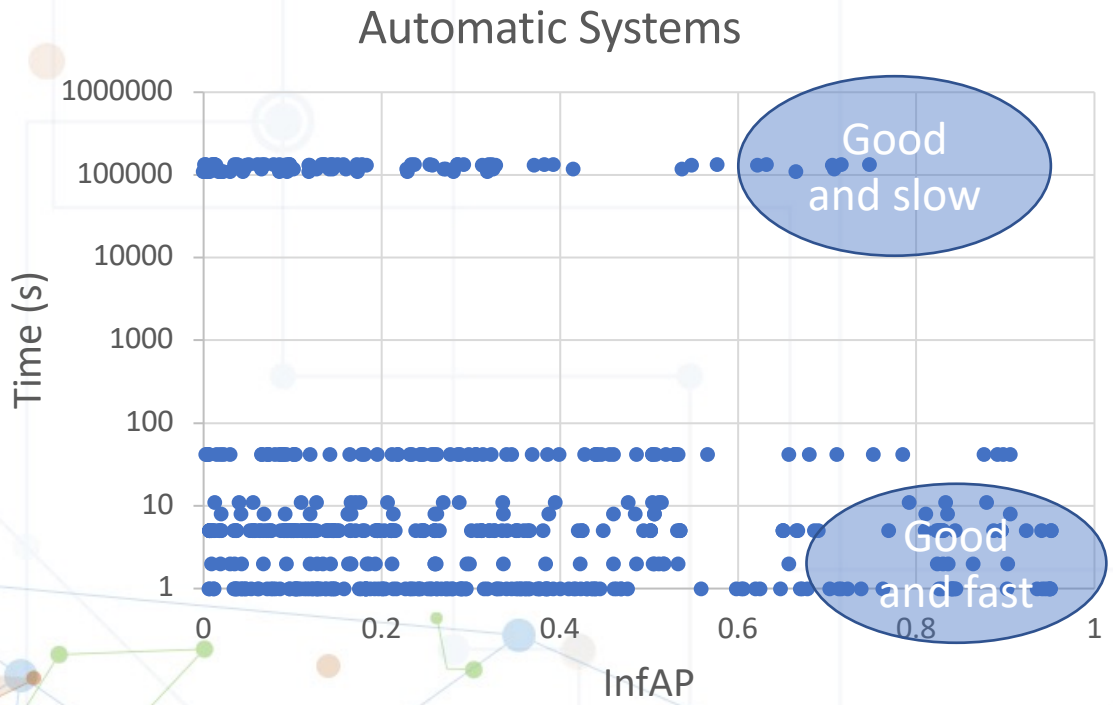


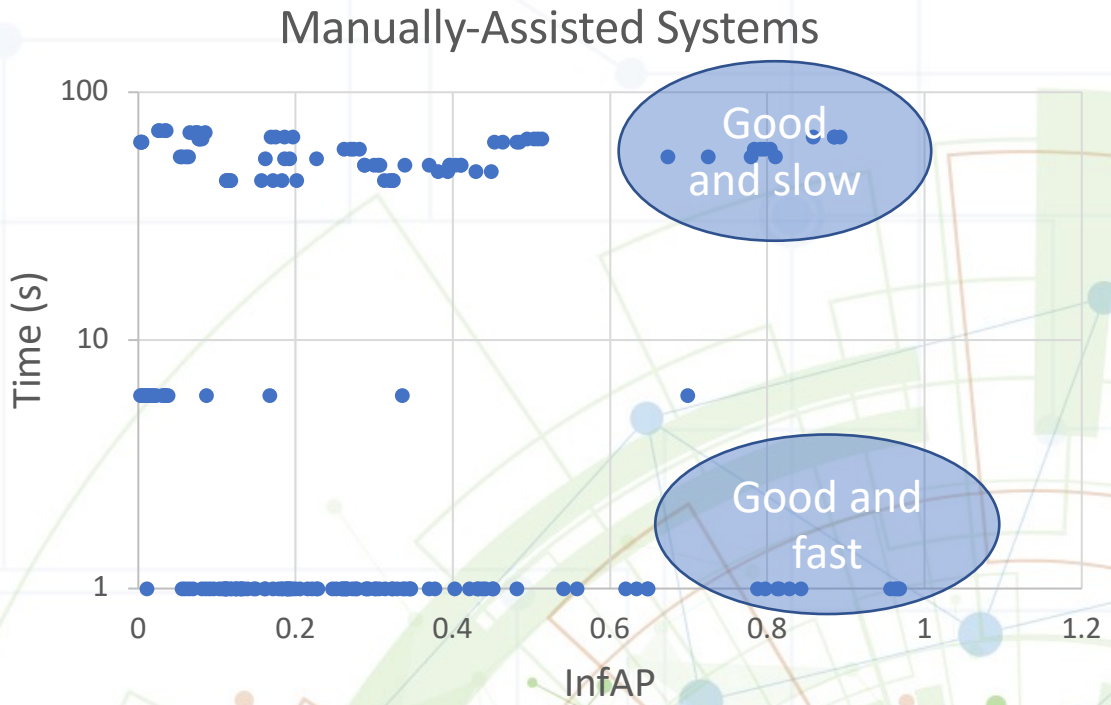
Novelty Scores



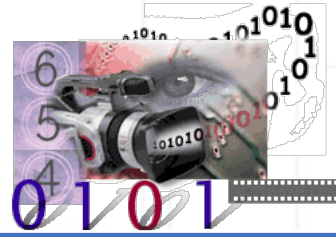
A weight is given to each topic and shot pairs in the ground truth such that highest weight is given to unique shots.

Efficiency


Automatic Systems
Time (s)
1000000
100000
10000
1000
100
10
1
0
0.2
0.4
0.6
0.8
1
InfAP
Good and slow
Good and fast


Manually-Assisted Systems
Time (s)
100
10
1
0
0.2
0.4
0.6
0.8
1
1.2
InfAP
Good and slow
Good and fast

Progress Task

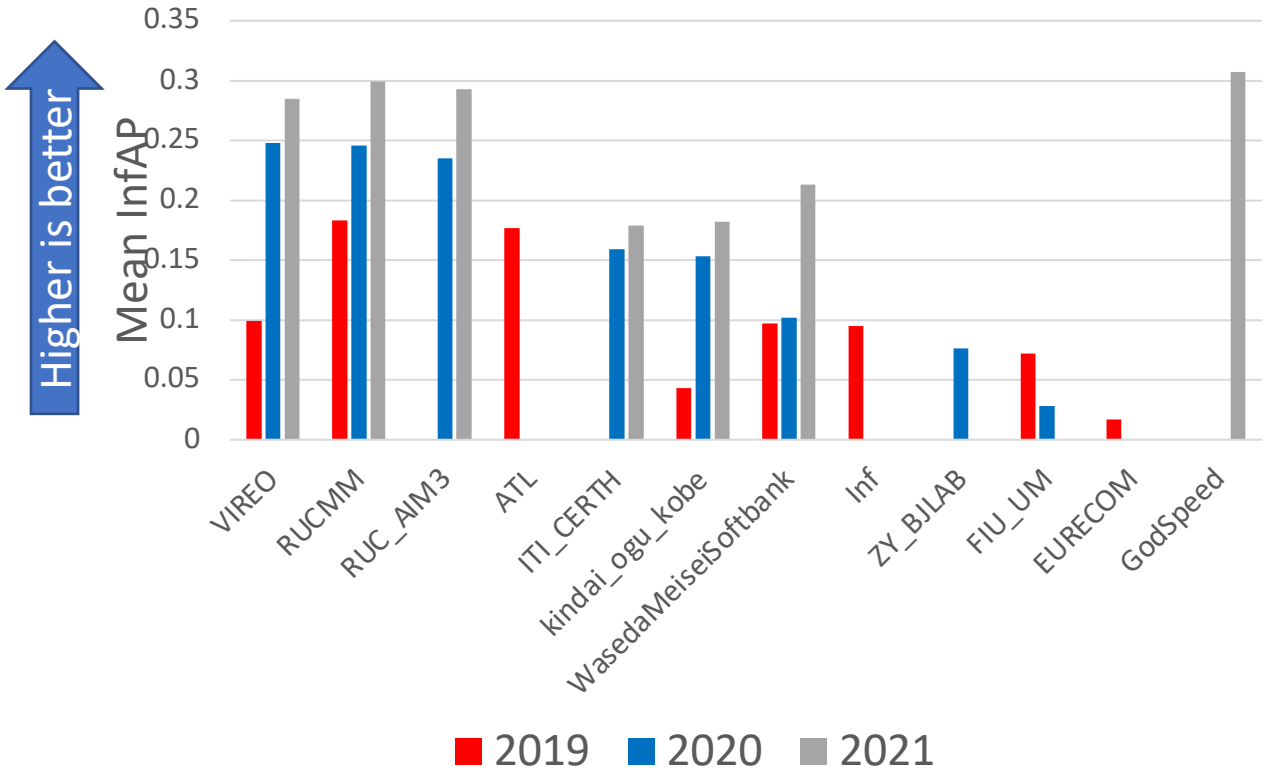


		Evaluation year		
		2019	2020	2021
Submission year	2019	***Systems:*** Submit 20 fixed progress queries		
	2020		***Systems:*** Submit 20 fixed progress queries ***NIST:*** Eval 10 queries (set A)	
	2021			***Systems:*** Submit 20 fixed progress queries ***NIST:*** Eval 10 queries (set B)

Goals : Evaluate 10 (set A) common queries submitted in 2 years (2019, 2020)
Evaluate 10 (set B) common queries submitted in 3 years (2019, 2020, 2021)

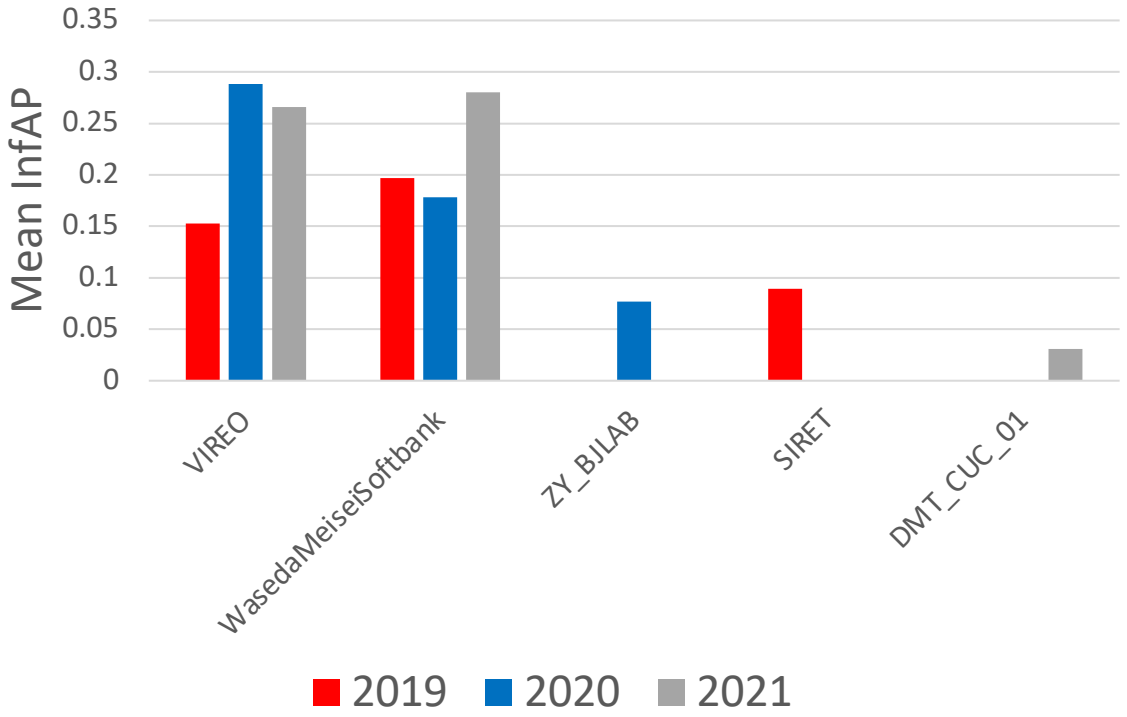
Progress set-A results (2019-2021)

Max performance per team (***automatic systems***) on 10 progress queries



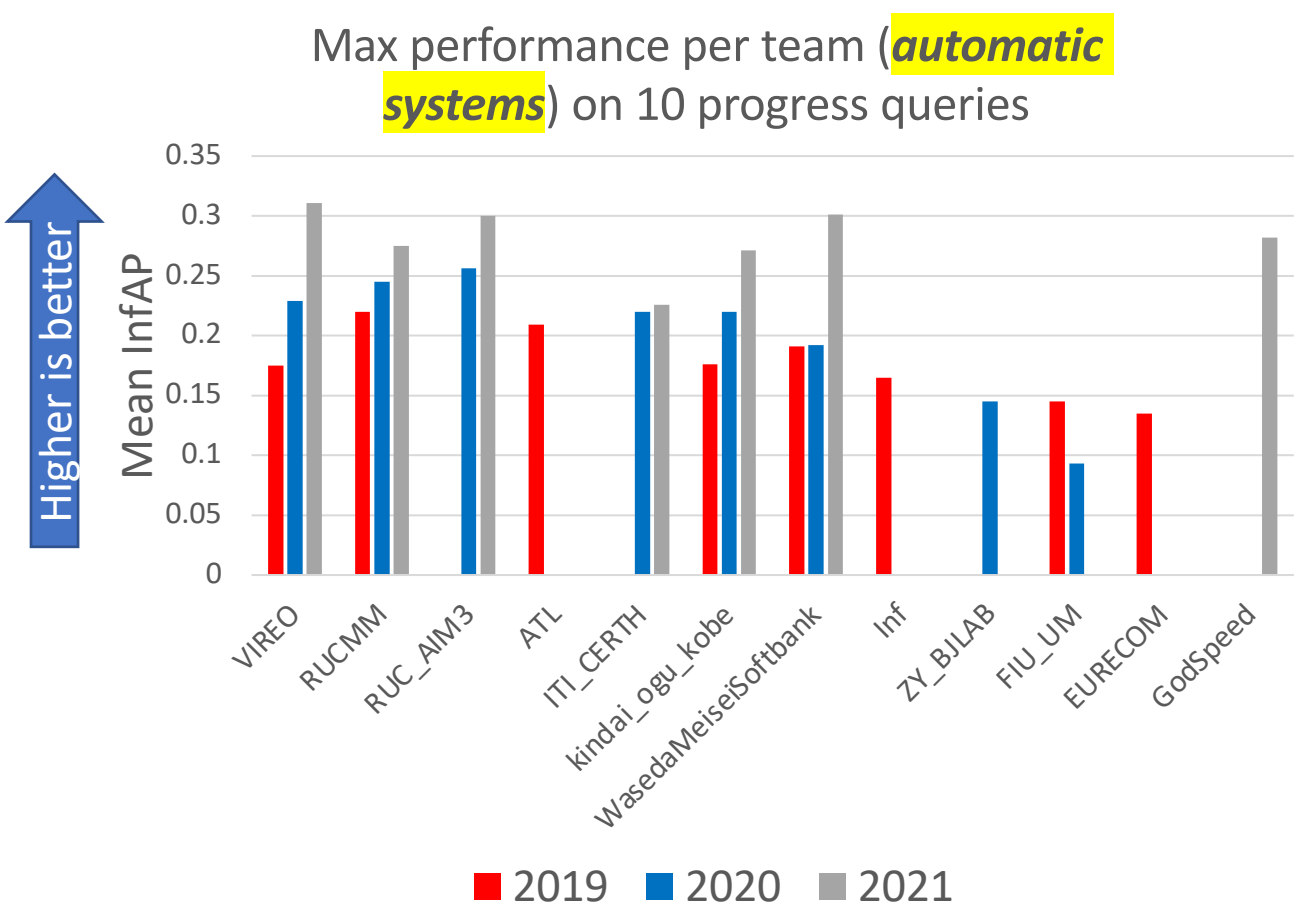
Majority of automatic systems performed better in later years on the fixed query set A

Max performance per team (***manually-assisted systems***) on 10 progress queries

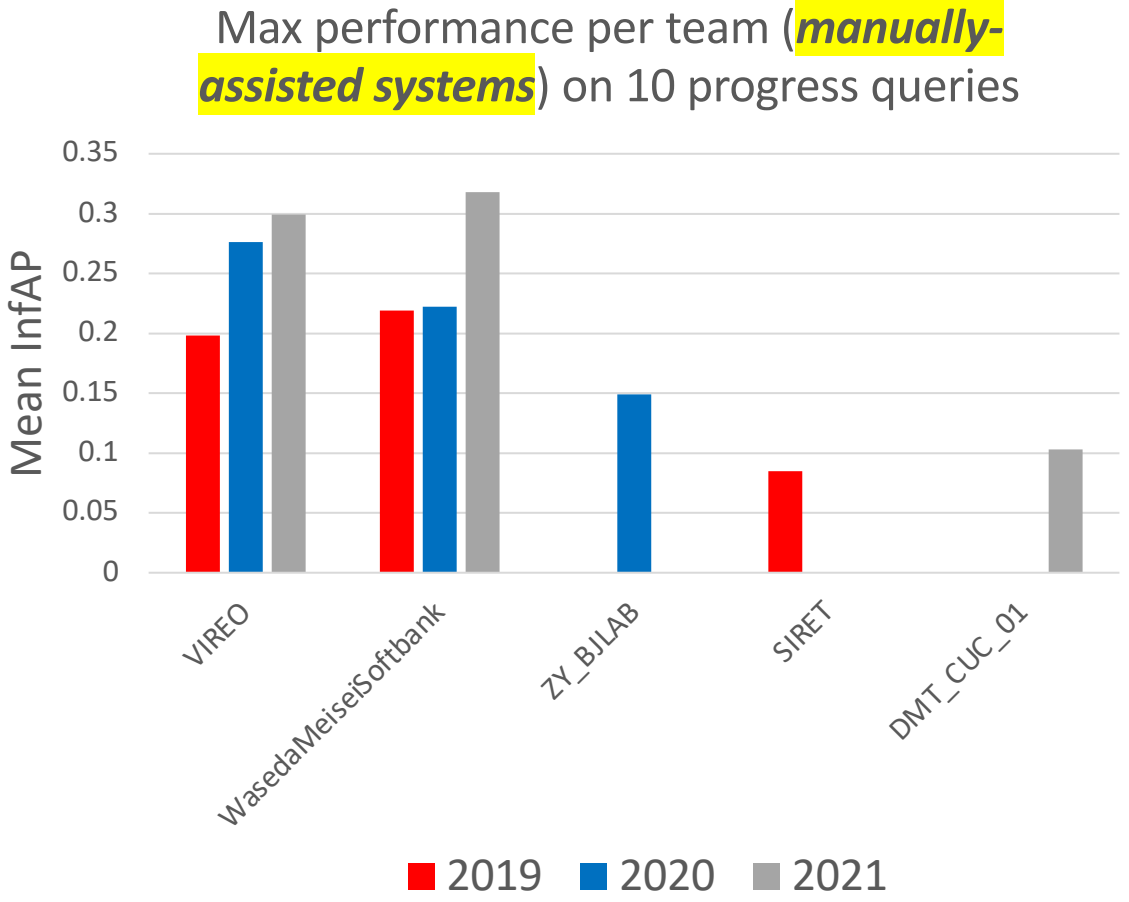


VIREO 2020 system is better, while Waseda's 2021 system is better. Not enough teams to make conclusions

Progress set-B results (2019-2021)



Majority of automatic systems performed better in later years on the fixed query set B

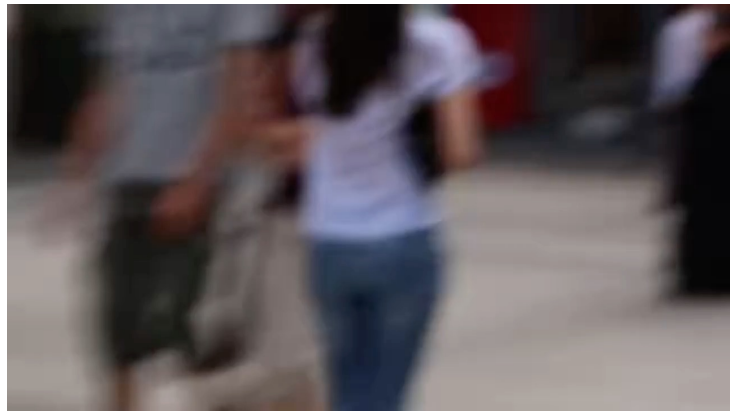


Two teams participated in 3 years with manually-assisted systems performed better in 2021. More participation is needed in manually-assisted runs.

Samples of frequent false positives

person looking at themselves in a mirror

person with a tattoo on their arm



adult person wearing a backpack and walking on a sidewalk

man pointing with his finger

man sitting on a barber chair in a shop

Samples of hard true positives

person with a tattoo on their arm

A woman holding a book

adult person wearing a backpack and walking on a sidewalk

person wearing an apron indoors

man sitting on a barber chair in a shop

a bow tie

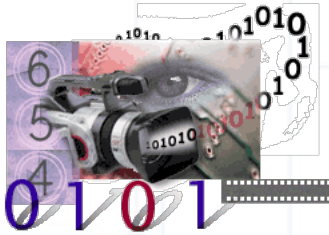
Easy vs Hard Queries

Top 5 easiest queries (ranked based on # of runs with infAP >= 0.5)

Query	Person	Action	Object	Location
Two boxers in a ring	✓			✓
Parachutist descending towards a field on the ground in the daytime	✓	✓		✓
Woman wearing sleeveless top	✓			
A bow tie			✓	
Person with a tattoo on their arm	✓			

Top 5 hardest queries (ranked based on # of runs with infAP < 0.5)

Query	Person	Action	Object	Location
Person wearing a cap backwards	✓	✓	✓	
Ladder with less than 6 steps			✓	
Man pointing with his finger	✓	✓		
Adult person wearing a backpack and walking on a sidewalk	✓	✓	✓	✓
Person looking at themselves in a mirror	✓	✓	✓	



DIGITAL VIDEO RETRIEVAL at NIST

General Observations

2021 Task Observations

- Submissions
  - 8 teams finished the main task and 14 teams finished the progress task.
  - 29 automatic systems and 10 manually-assisted systems submitted runs in the main task.
  - 112 total systems (31 manually-assisted and 81 automatic) were submitted between 2019-2021 in the progress subtask.
  - Run training types are dominated by “D” runs. No “E” or “F” runs.
  - No teams submitted “optional” explainability results with their runs!
  - Only 2 Novelty systems submitted. Better than common runs on novelty metric.
- Performance
  - Majority of 2021 systems performed higher than their 2019 & 2020 systems in the **progress subtask**
  - Few automatic systems are good and fast (< 10 sec). Additional processing time didn’t help most systems
  - High similarity between automatic and manually-assisted systems in terms of query performance relatively to each other.
  - Top scoring teams not necessary contributing a lot of unique true shots and vice-versa (Except for VIREO team)
  - About 11% of all hits are unique. All the rest are common hits across the runs.
  - Hard queries are the ones asked for unusual combinations of facets (compared to well-known concepts)

Interactive Video Retrieval

During the Video Browser Showdown (VBS)

At MMM 2022
28th International Conference on Multimedia Modeling,
April 2022, Qui Nhon, Vietnam

- 10 Ad-Hoc Video Search (AVS) topics : Each AVS topic has several/many target shots (from V3C1 + V3C2 datasets) that should be found.
- 10 Known-Item Search (KIS) tasks, which are selected completely random on site. Each KIS task has only one single 20 s long target segment.
- Registration for the task is now closed

Agenda

EST Time

7:30 – 7:50 AM	• VIREO
7:50 – 8:10 AM	• GodSpeed
8:10 – 8:30 AM	• RUCMM
8:30 - 8:50 AM	• Waseda_Meisei_SoftBank
8:50 - 9:10 AM	• Break
9:10 - 9:30 AM	• AVS Task Discussion