

# **VIREO-TNO @ TRECVID 2014**

## **Zero-Shot Event Detection and Recounting**

Speaker: Maaïke de Boer (TNO)

Yi-Jie Lu<sup>1</sup>, Hao Zhang<sup>1</sup>, Chong-Wah Ngo<sup>1</sup>

Maaïke de Boer<sup>2</sup>, John Schavemaker<sup>2</sup>, Klamer Schutte<sup>2</sup>, Wessel Kraaij<sup>2</sup>

<sup>1</sup>VIREO Group, City University of Hong Kong, Hong Kong

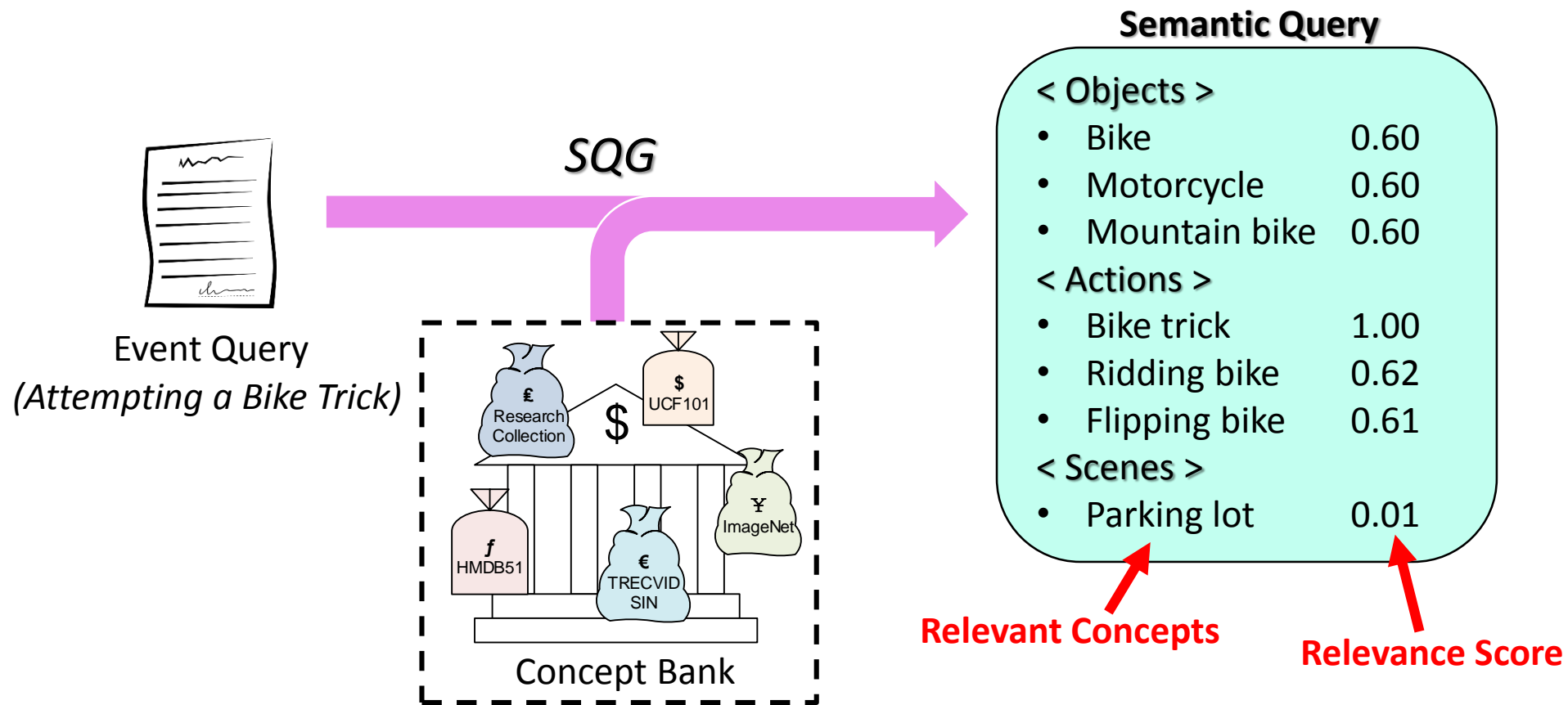
<sup>2</sup>Netherlands Organization for Applied Scientific Research (TNO), Netherlands

# Outline

- **0-Shot System**
  - System Overview
  - Findings
- **MER System**
  - System Workflow
  - Results

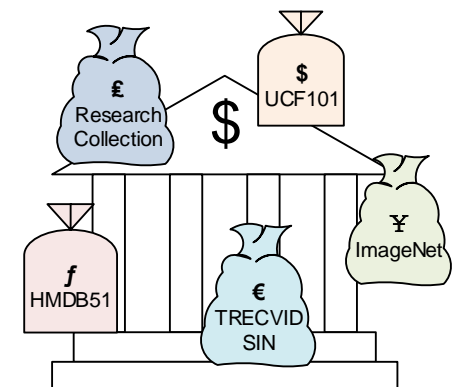
# Semantic Query Generation (SQG)

- Given an *event query*, SQG translates the query description into a representation of *semantic concepts*



## ● Concept Bank

- Research collection (497 concepts)
- ImageNet ILSVRC'12 (1000 concepts)
- SIN'14 (346 concepts)



## ● Event Search

- Ranking according to the SQ and concept responses

### Semantic Query $q$

#### < Objects >

- Bike 0.60
- Motorcycle 0.60
- Mountain bike 0.60

#### < Actions >

- Bike trick 1.00
- Ridding bike 0.62
- Flipping bike 0.61

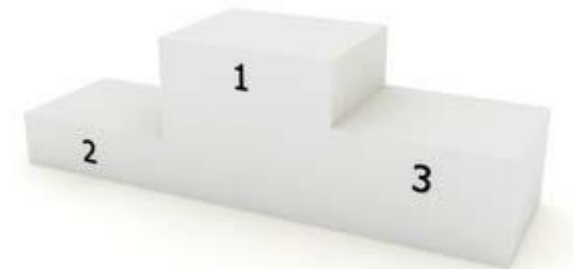
#### < Scenes >

- Parking lot 0.01

*Event Search*  $s_i = qc_i$



Concept Response  $c_i$



**Video Ranking**

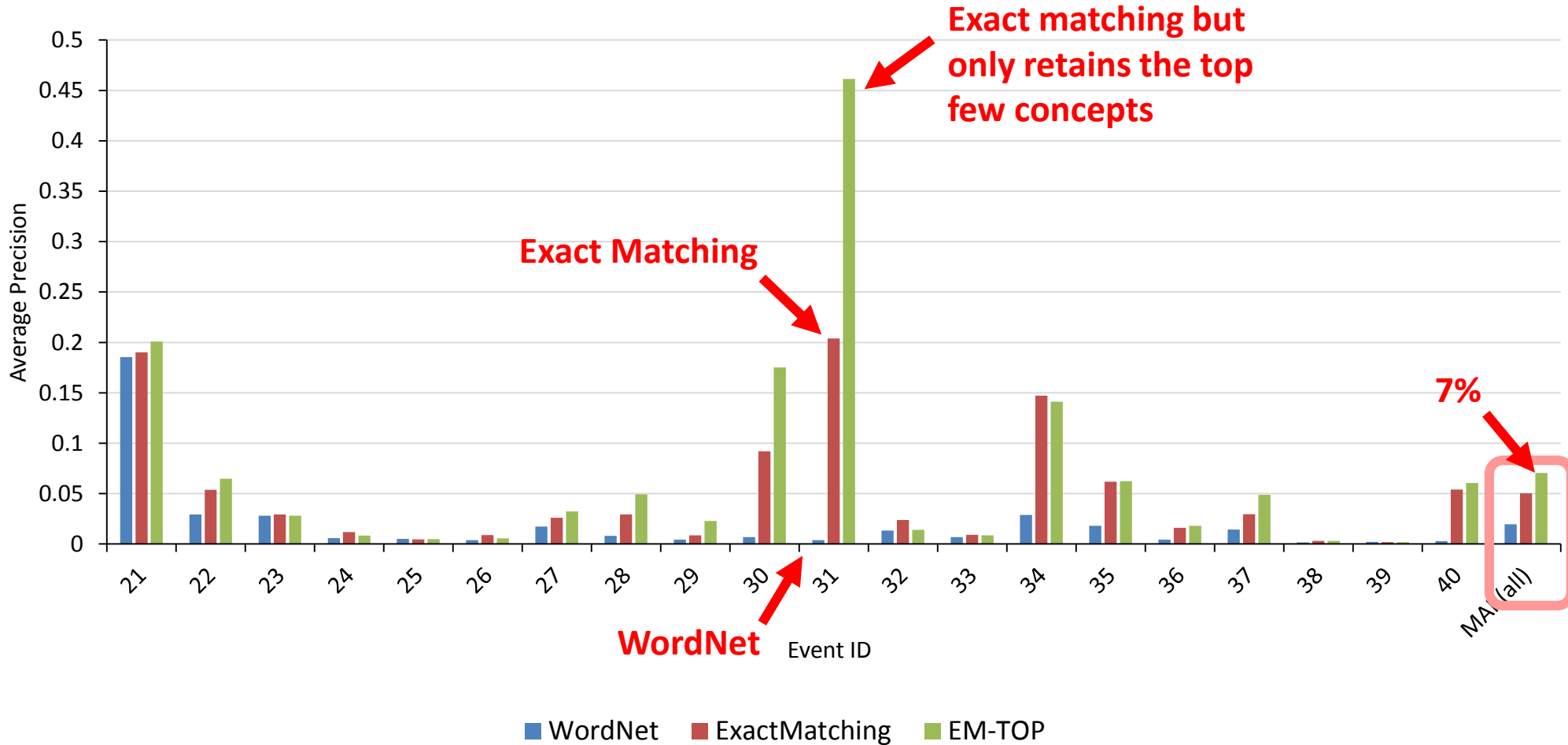
# Outline

- **0-Shot System**
  - System Overview
  - Findings
- **MER System**
  - System Workflow
  - Results

- **SQG Experiments**

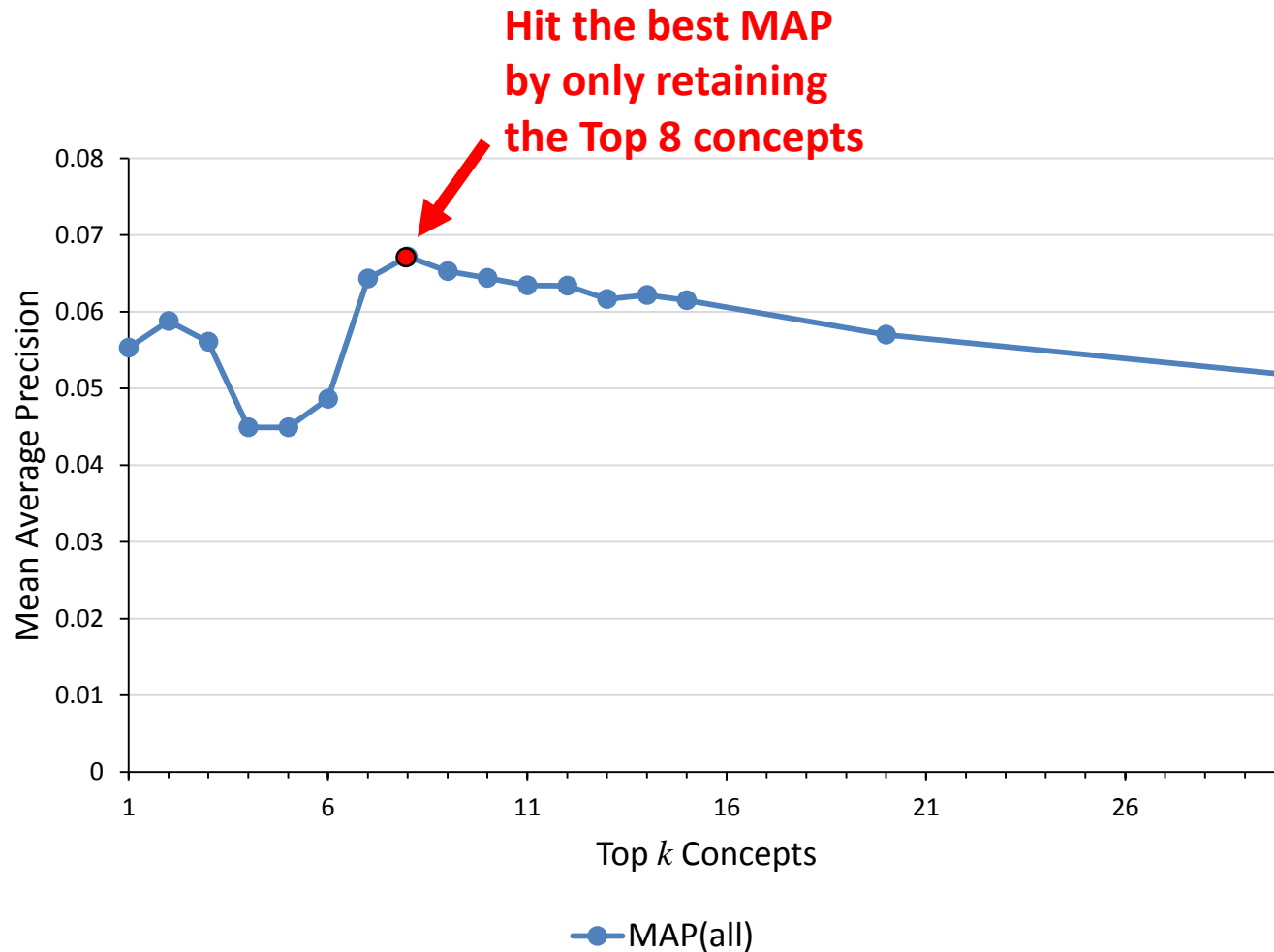
- Exact matching vs. *WordNet/ConceptNet* matching
- How many concepts are used to represent an event?
- To further improve the weighting:
  - *TF-IDF*
  - *Term specificity*

# Exact matching vs. WordNet matching

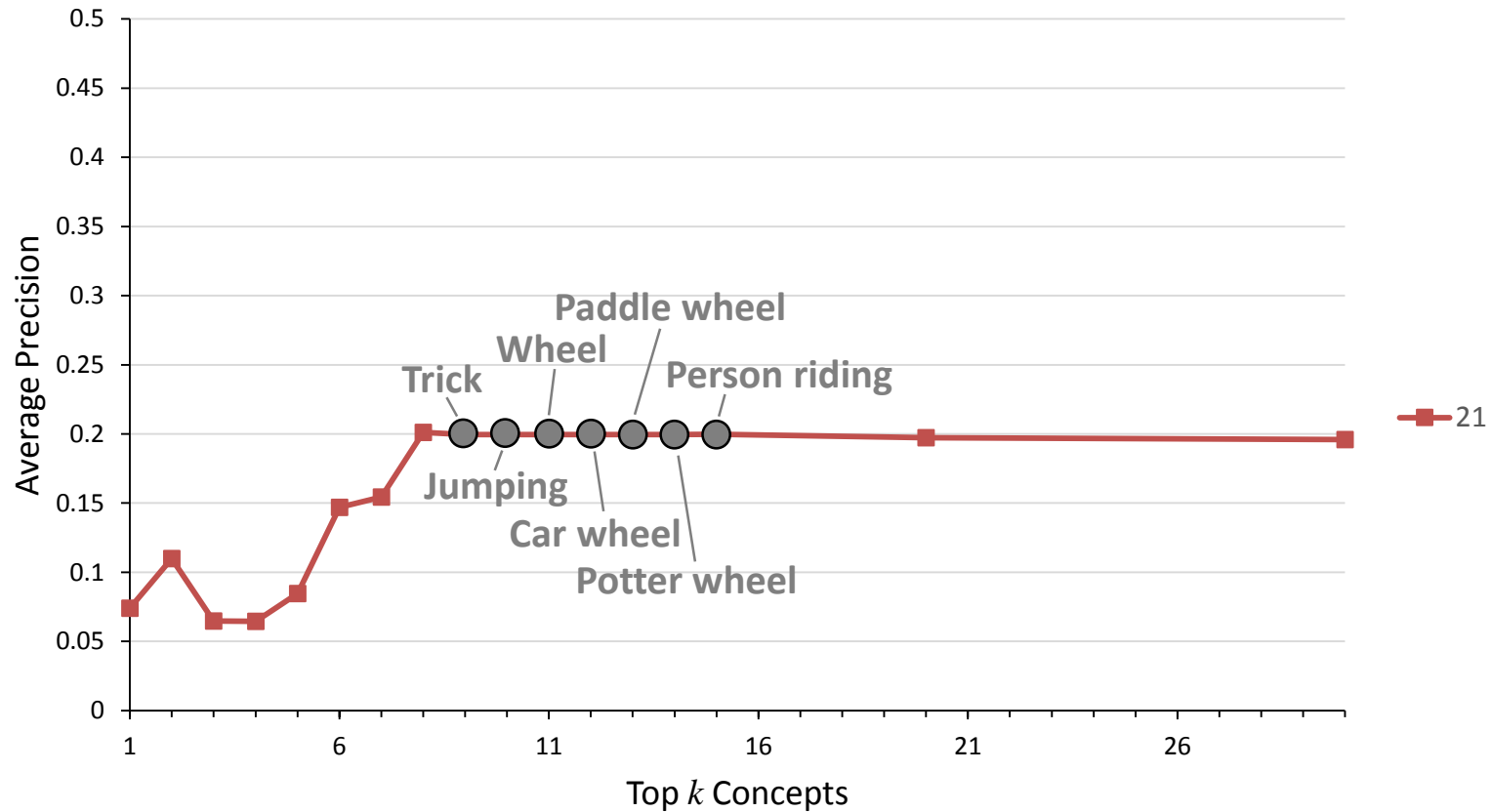




- Amount of concepts used to represent event

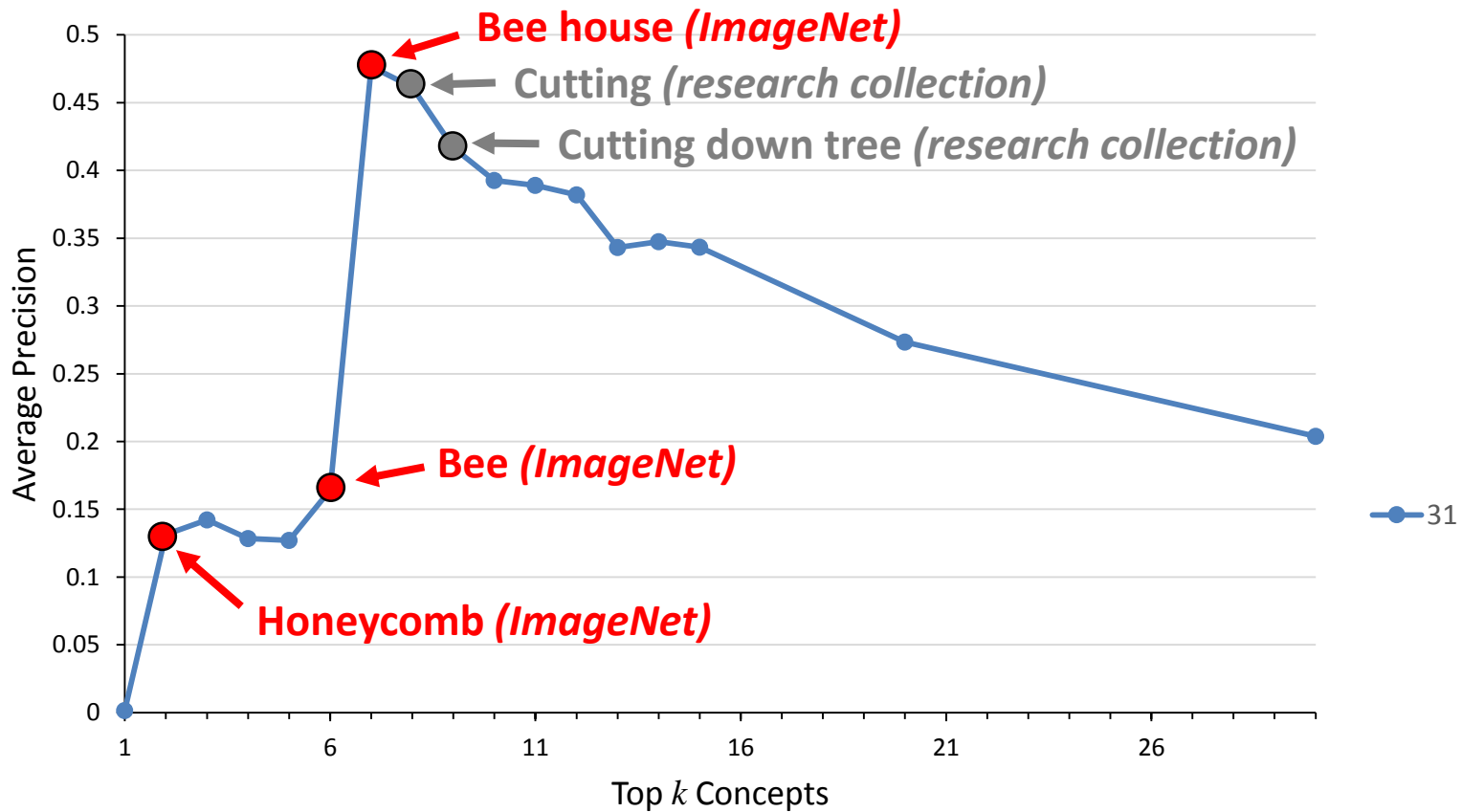


# Insights



Event 21: Attempting a bike trick

# Insights



Event 31: Beekeeping

# Insights

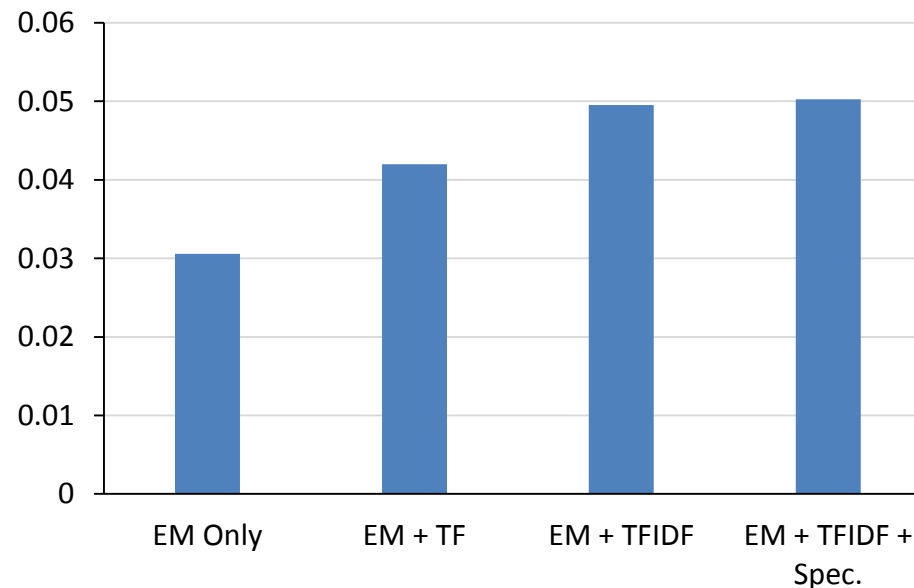


Event 23: Dog show



- Improvements by **TF-IDF** and **word specificity**

Method	MAP (on MED14-Test)
Exact Matching Only	0.0306
Exact Matching + TF	0.0420
Exact Matching + TFIDF	0.0495
Exact Matching + TFIDF + Word Specificity	0.0502

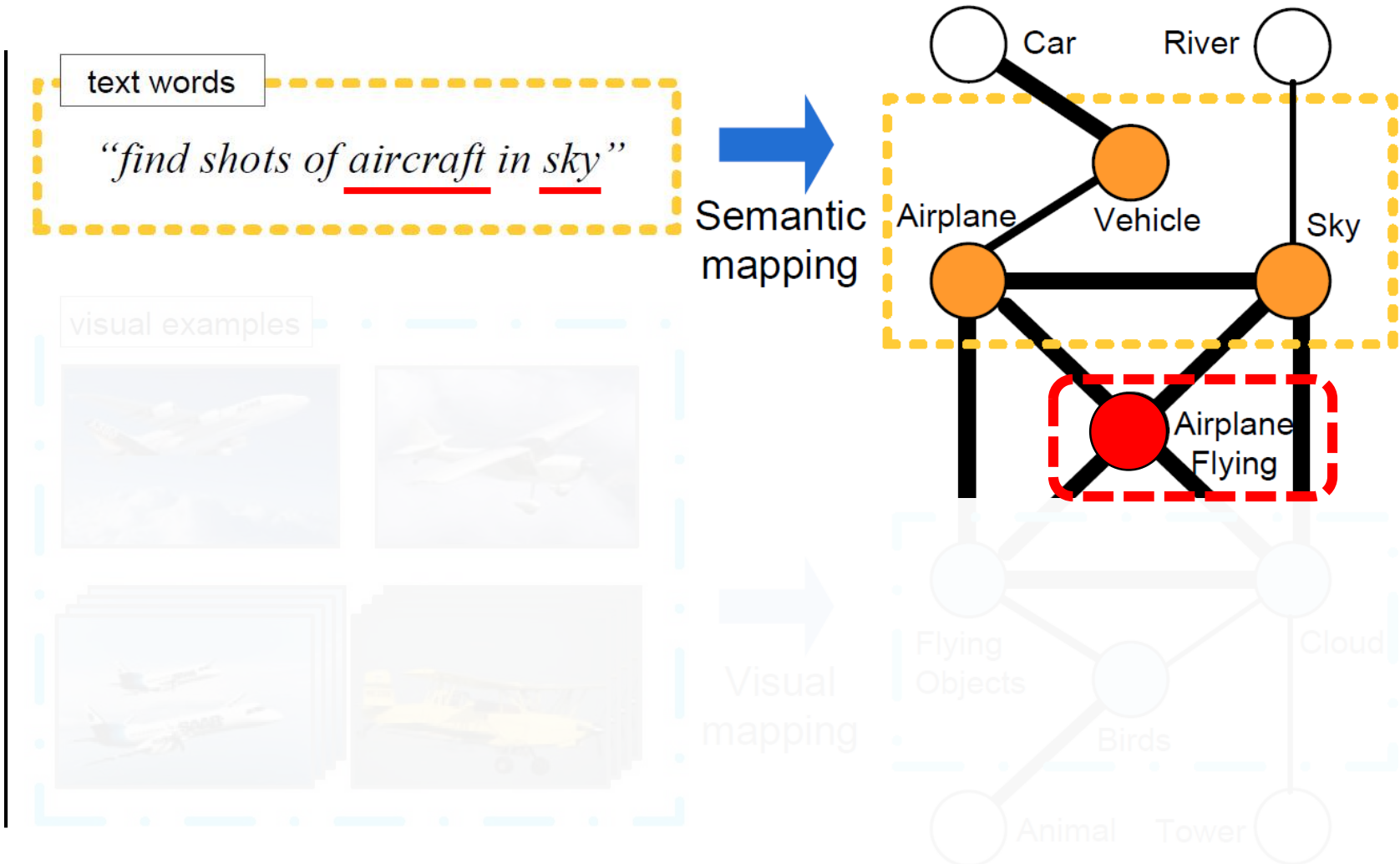


# Findings

1. **Exact matching** performs better than matching with WordNet and/or ConceptNet
2. Performance is even better by only retaining the **top few** exactly matched concepts
3. Adding both **TF-IDF** and **Word Specificity** increases performance

- **Why ontology-based mapping would not work?**

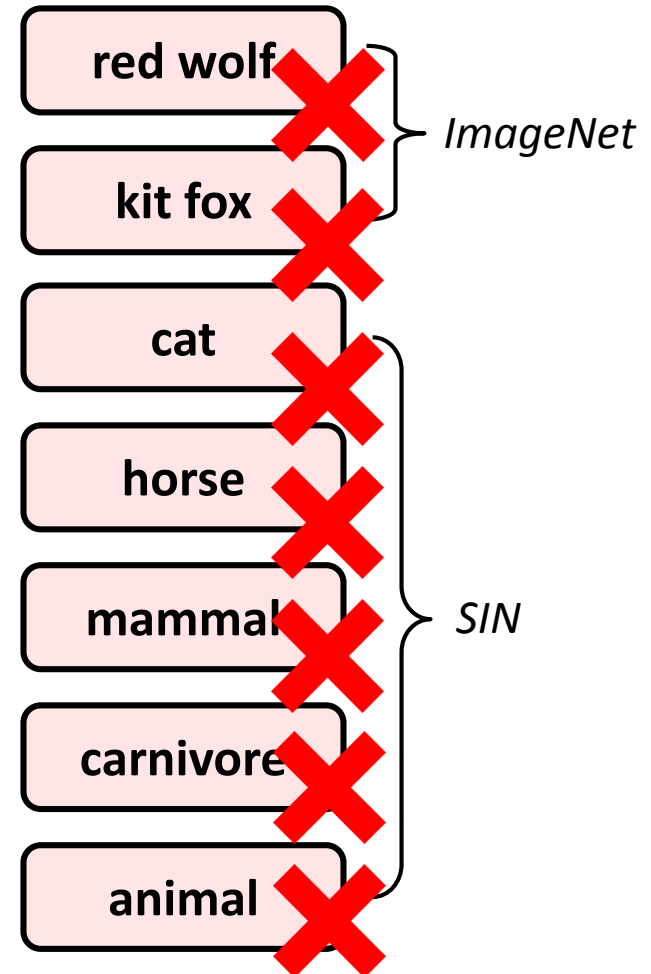
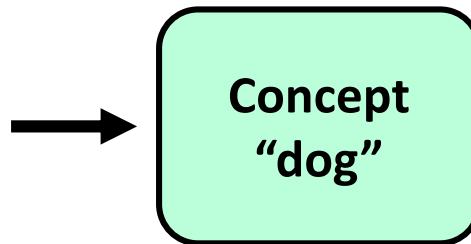
A sample query in TRECVID 2009



- Why ontology-based mapping would not work?

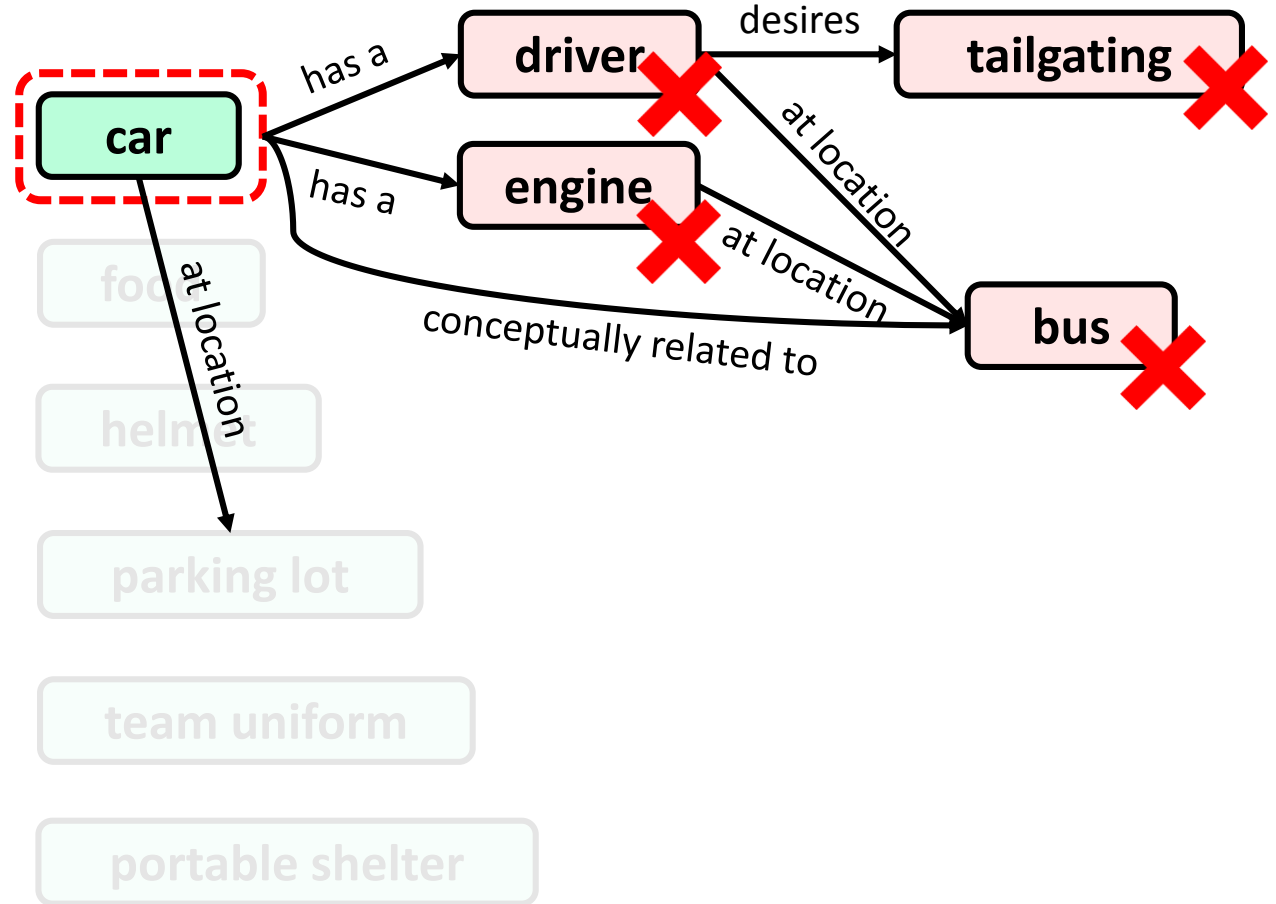


Dog Show





● **Why *ConceptNet* mapping would not work?**



**Tailgating**

# Findings

- **It is difficult to**
  - harness the ontology-based mapping while constraining the mapping by event context

- In the *Ad-Hoc* event “Extinguishing a Fire”

- Key concepts are missing:

- *Fire extinguisher*



- *Firefighter*



# Findings

- **It is reasonable to**
  - **Scale up** the number of concepts, thus increasing the chance of exact matching

# MED14-Eval-Full Results

- ***PS 000Ex***
  - Automatic semantic query generation and search
  - Fusion of *0-Shot* and *OCR* system
  - Achieves the MAP of **5.2**
- ***AH 000Ex***
  - System is the same as in *PS 000Ex*
  - Achieves the MAP of **2.6**
  - Performance drops due to the lack of key concepts

# Outline

- **0-Shot System**
  - System Overview
  - Findings
- **MER System**
  - System Workflow
  - Results

# MER System

- **In algorithm design, we aim to optimize**
  - *Concept-to-event relevancy*
  - *Evidence diversity*
  - *Viewing time of evidential shots*

# MER System

- In algorithm design, we aim to optimize
  - *Concept-to-event relevancy*
    - First, we require that candidate shots are *relevant to the event*;
    - Second, we do *concept-to-shot alignment*.
  - *Evidence diversity*
  - *Viewing time of evidential shots*



# MER System

- In algorithm design, we aim to optimize
  - **Concept-to-event relevancy**
    - First, we require that candidate shots are *relevant to the event*;
    - Second, we do *concept-to-shot alignment*.
  - **Evidence diversity**
    - In concept-to-shot alignment, we recount each shot with a *unique* concept different from other shots.
  - *Viewing time of evidential shots*

# MER System

- In algorithm design, we aim to optimize
  - **Concept-to-event relevancy**
    - First, we require that candidate shots are *relevant to the event*;
    - Second, we do *concept-to-shot alignment*.
  - **Evidence diversity**
    - In concept-to-shot alignment, we recount each shot with a *unique* concept different from other shots.
  - **Viewing time of evidential shots**
    - Select only the *three* most confident shots as key evidence
    - Basically, each shot is in about *5 seconds*

# Outline

- **0-Shot System**
  - System Overview
  - Findings
- **MER System**
  - System Workflow
  - Results

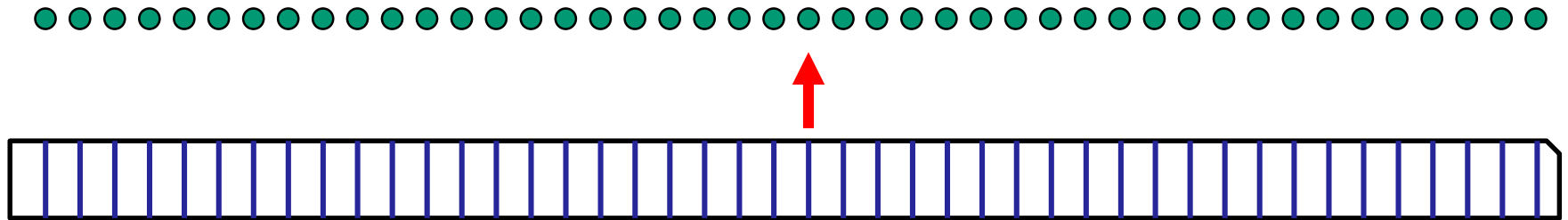
- **Key Evidence Localization**



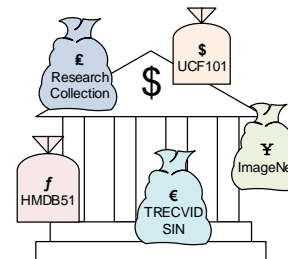
Extract keyframes uniformly

- **Key Evidence Localization**

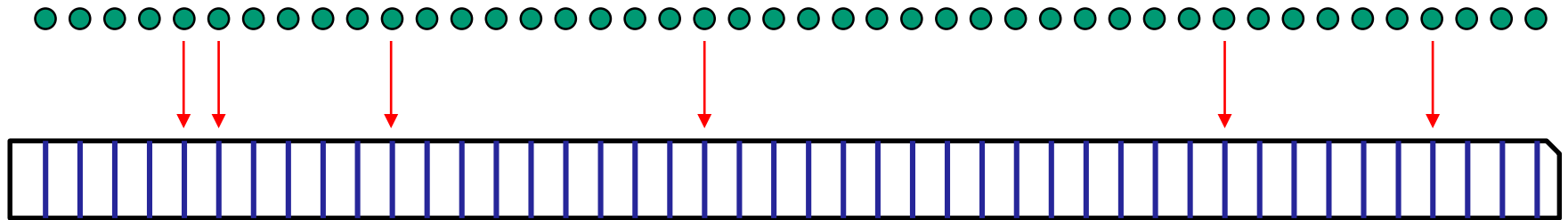
Concept Responses



Apply concept detectors



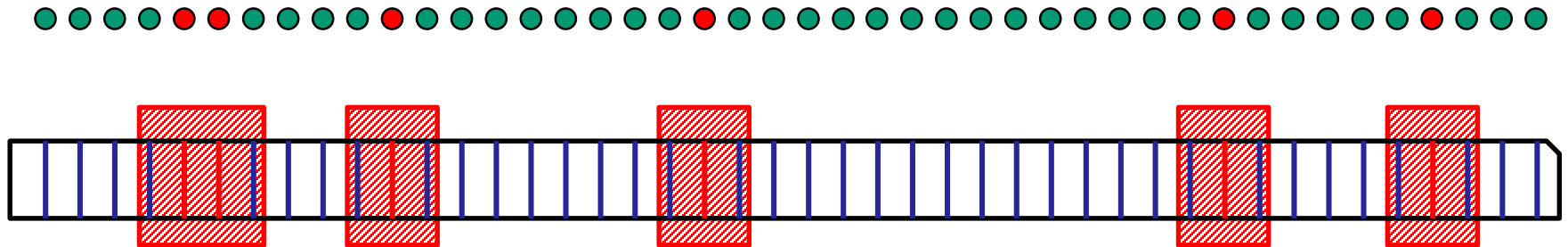
## ● Key Evidence Localization



Choose keyframes that are most relevant to this event

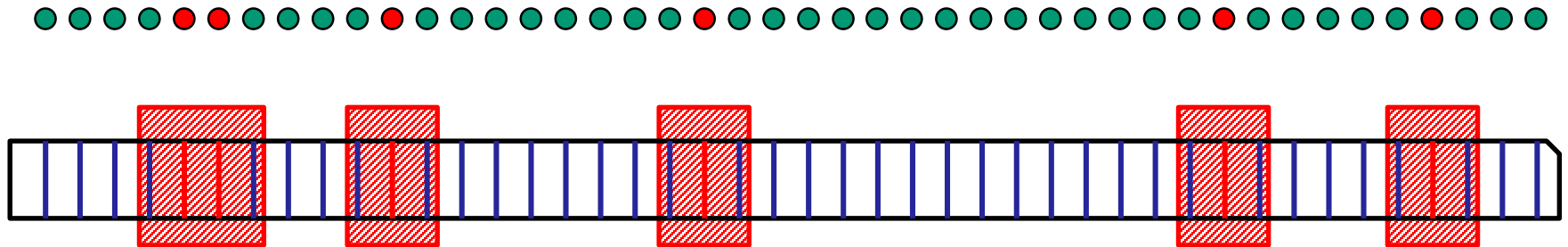
- *All concepts in semantic query are taken into account by calculating the weighted sum  $s_i = \mathbf{w}r_i$*

- **Key Evidence Localization**



Expand keyframes to shots

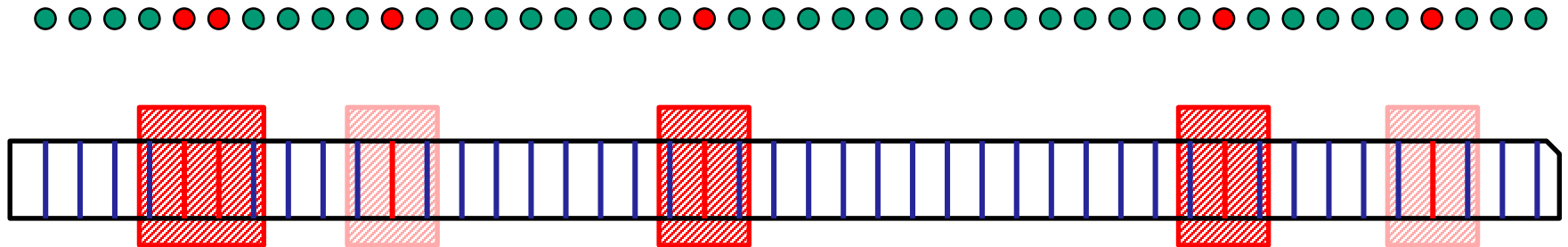
- **Key Evidence Localization**



The top 3 shots are selected as key evidences

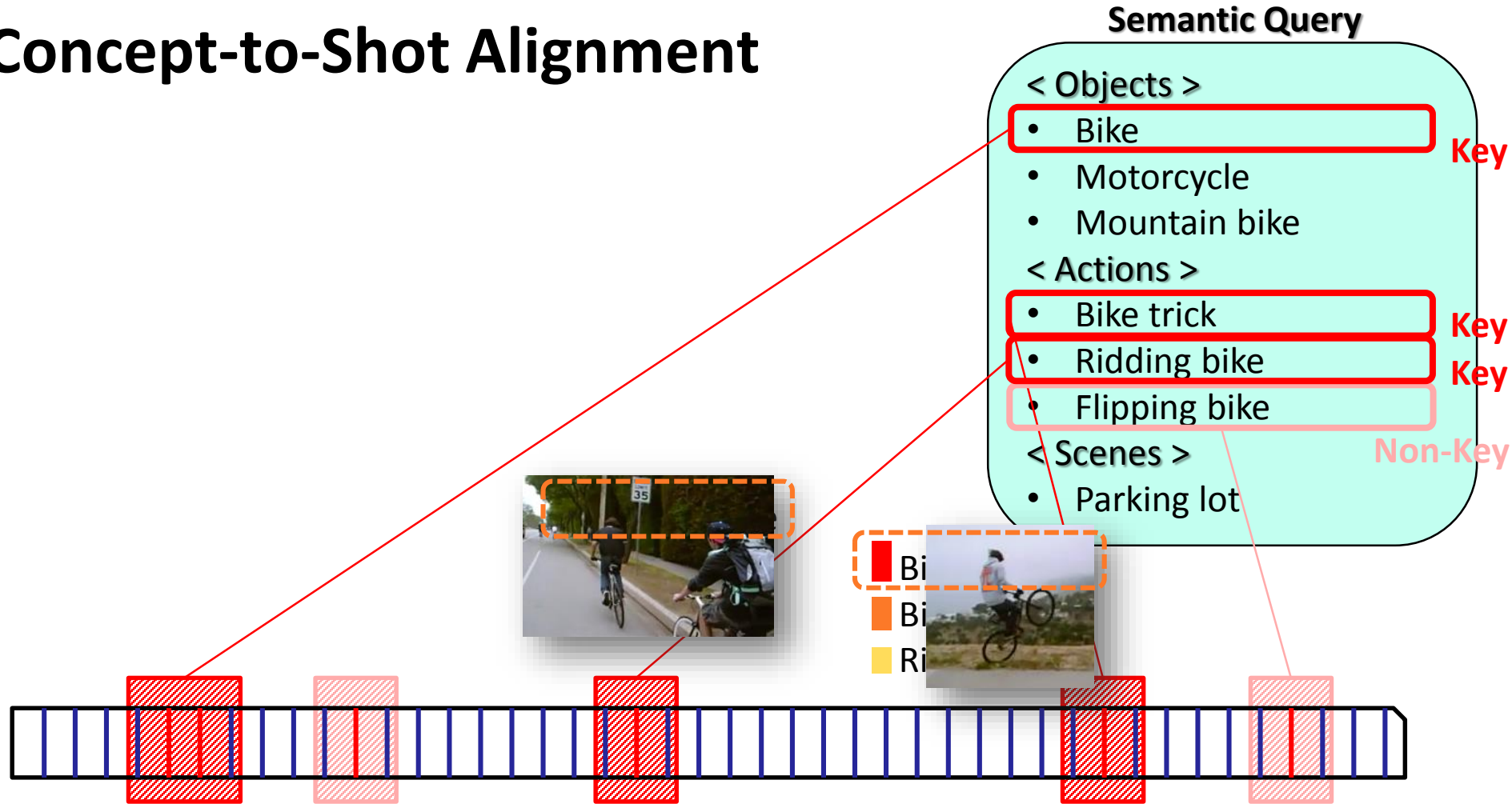


- **Key Evidence Localization**



The rests are non-key evidences

# ● Concept-to-Shot Alignment

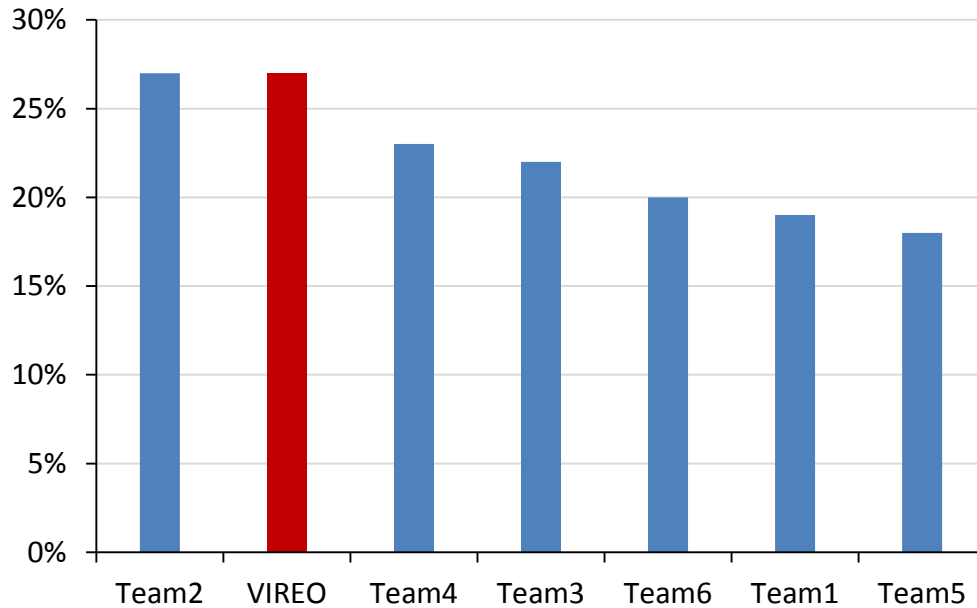


The top concept in the key evidence is selected as the representative concept

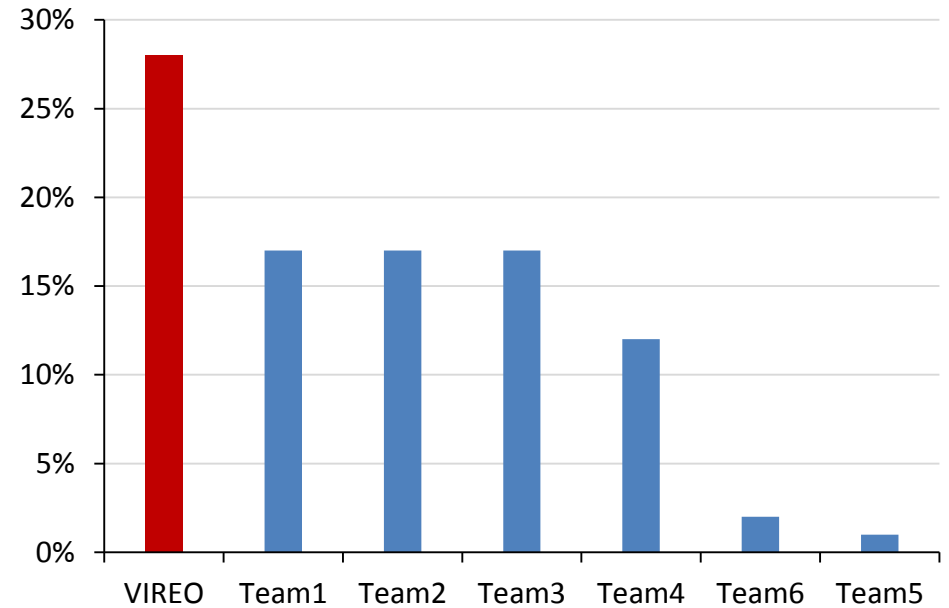
\* We choose unique concept for each shot

# MER14 Results

The percentage of strongly agree



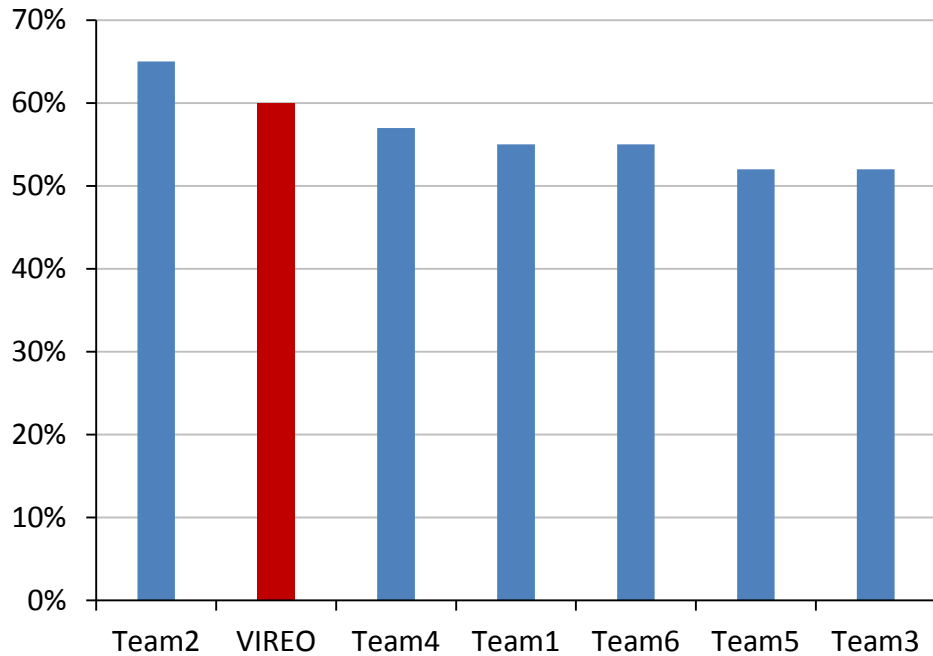
(a) Evidence quality



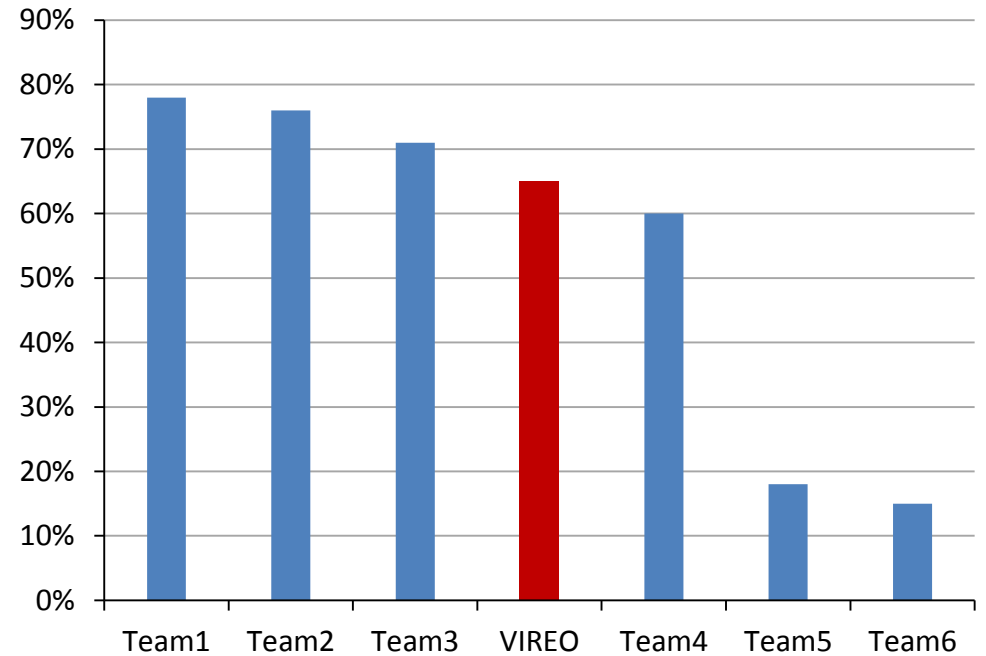
(b) Event query quality

# MER14 Results

The percentage of both agree and strongly agree



(a) Evidence quality



(b) Event query quality

# Summary

- **0-Shot System**

- The simple *exact matching* performs the best
- The *quality* of concepts selected to represent an event is more important than *quantity*
- It's an open problem of how to harness the *ontology-based mapping*

# Summary

- **MER System**

- In key evidence localization, we emphasize the *event relevancy* first, then the *hot concepts*
- We recommend *three* shots as key evidences and each in about *5 seconds*

Thanks!