Are clickthrough data reliable as image annotations?

Theodora Tsikrika (CWI)
Christos Diou (AUTH, CERTH-ITI)
Arjen P. de Vries (CWI, TU Delft)
Anastasios Delopoulos (AUTH, CERTH-ITI)

1

Automatic concept-based image annotation

- ***Aim: unambiguously describe the visual content of images***

- Challenges when using **<u>supervised machine learning</u>** *techniques:*

  - *require labelled samples as training data*

  - *laborious and expensive task when performed manually*

  - *large number of semantic concepts*

  - *poor generalisation of concept classifiers in other domains*

  > How can we automatically supplement/replace the manually annotated training samples?
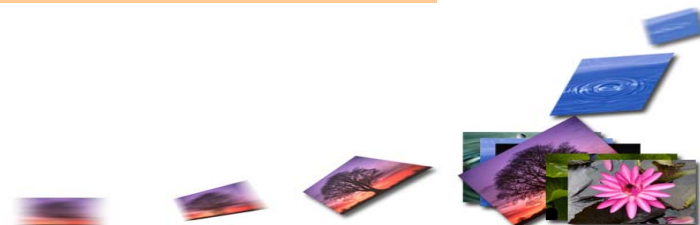
Approach

- **_Automatically generated annotated training samples_**
    - _user-defined tags (e.g., Flickr)_
    - _keywords extracted from Web pages where images are embedded_
    - _clickthrough data collected_

- **_<u>advantages</u>: large quantities, no user intervention, available to all content owners, collective annotations (assessments)_**
- **_<u>disadvantages</u>: sparse, noisy, user queries not based on strict visual criteria_**

Research questions

1) **How can we build classifiers for annotating images with concepts using clickthrough data?**

- *methods for searchlog-based positive sample selection*
- *random negative sample selection*

2) **What is the effectiveness of these concept classifiers?**

- *experiments using data provided by BELGA news agency*
  - *~100k photographic images (with their text metadata)*
  - *clickthrough data*

3) **What is the effect of the noise in the clickthrough data on the effectiveness of these concept classifiers?**

- *measure the reliability of clickthrough data as image annotations*
- *measure their effect on the effectiveness*

Concept definition

A concept is a clearly defined, non ambiguous entity represented by

– a short name

– keywords

– free-text short description

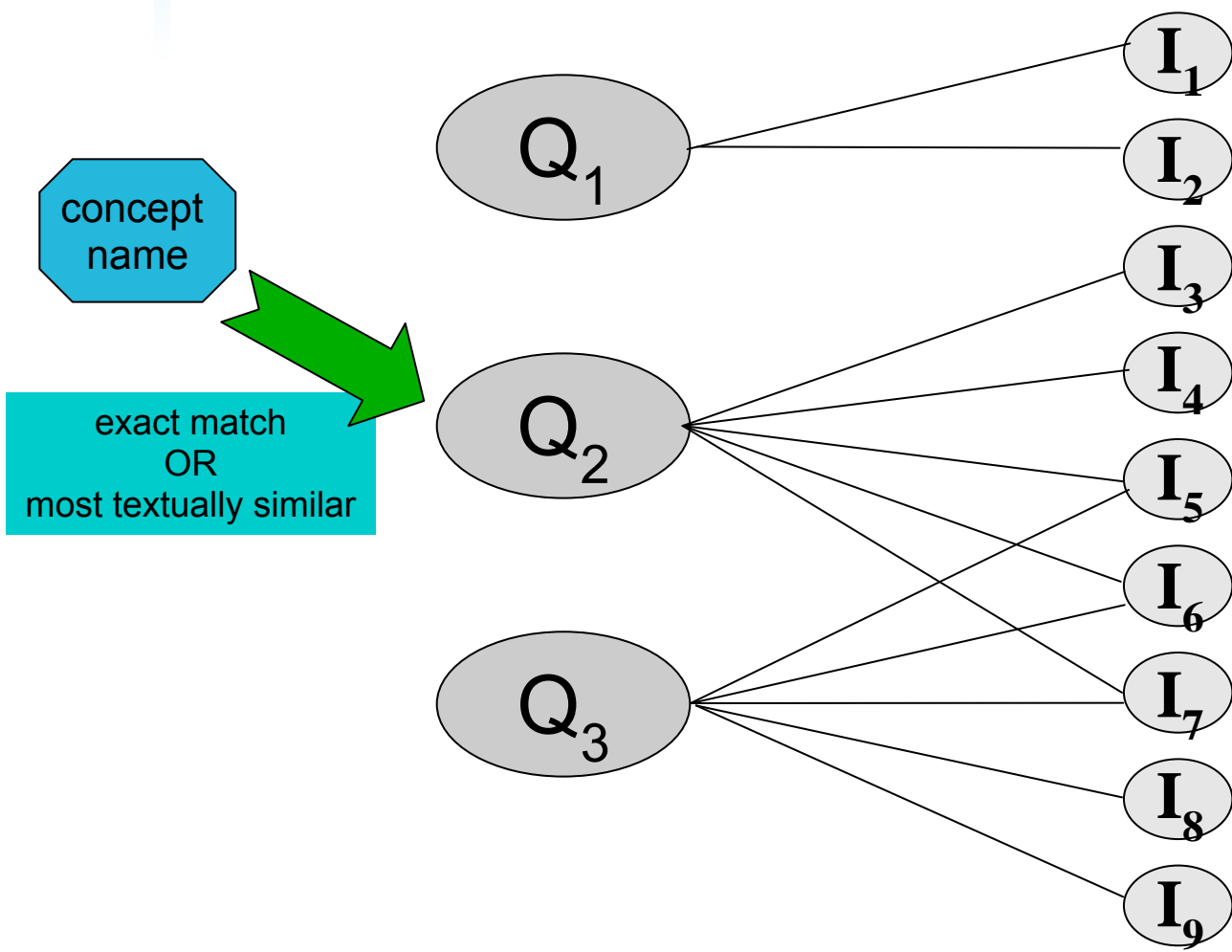| Concept		
Name	**Keywords**	**Description**
traffic	traffic, traffic jam, cars, road, highway	Image showing a high density of vehicles when on a road or highway.

Positive sample selection using search logs

- **Method exact**
  - select images clicked for queries exactly matching the concept name

- **Methods "textual similarity" (based on IR language models)**
  - annotate each image with all queries for which it has been clicked
    - apply stemming (yes/no)
  - select images retrieved
    - for query: concept name
    - using retrieval model: (i) language model (LM) (ii) smoothed LM (LMS)

- **Method clickgraph**
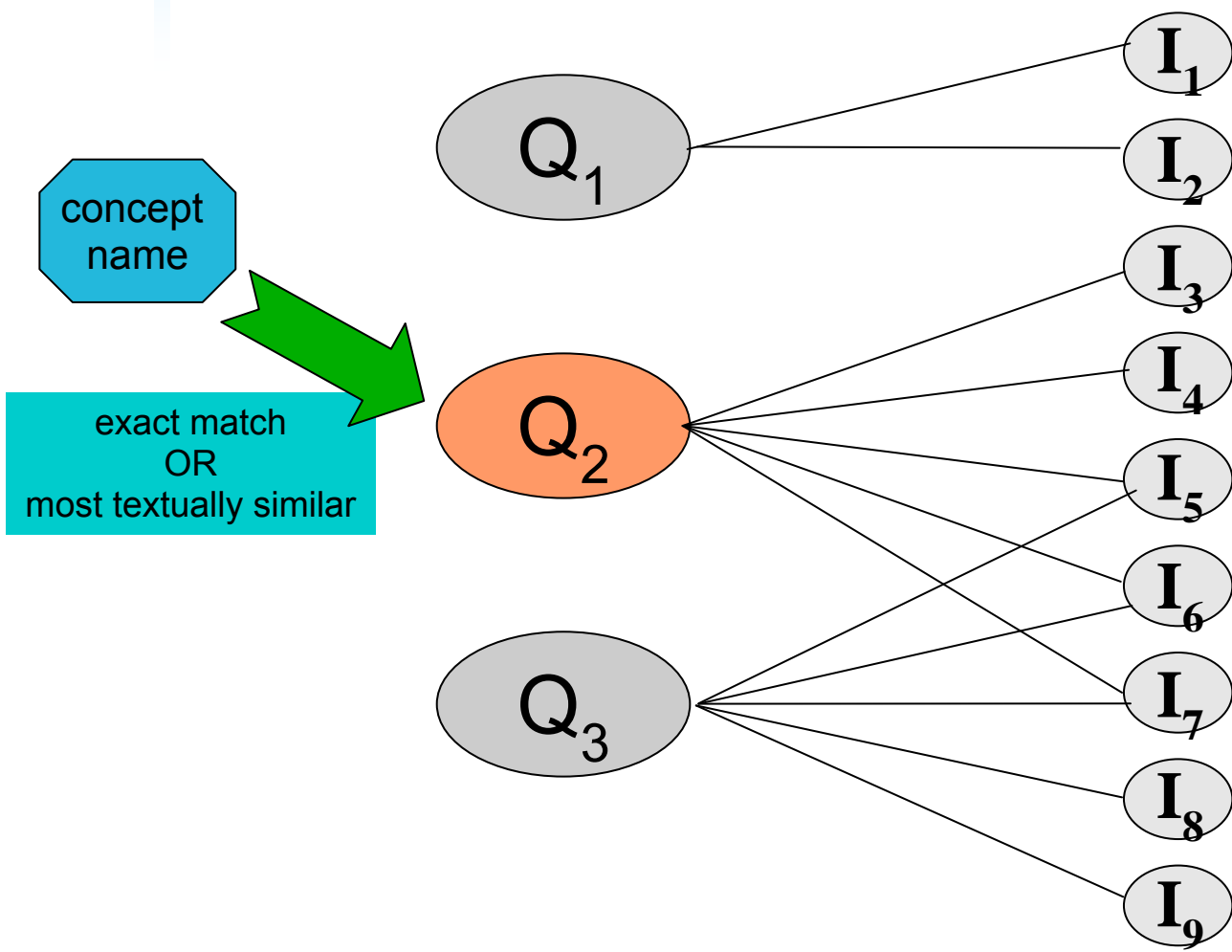 - images clicked for the same query are likely to be relevant to each other
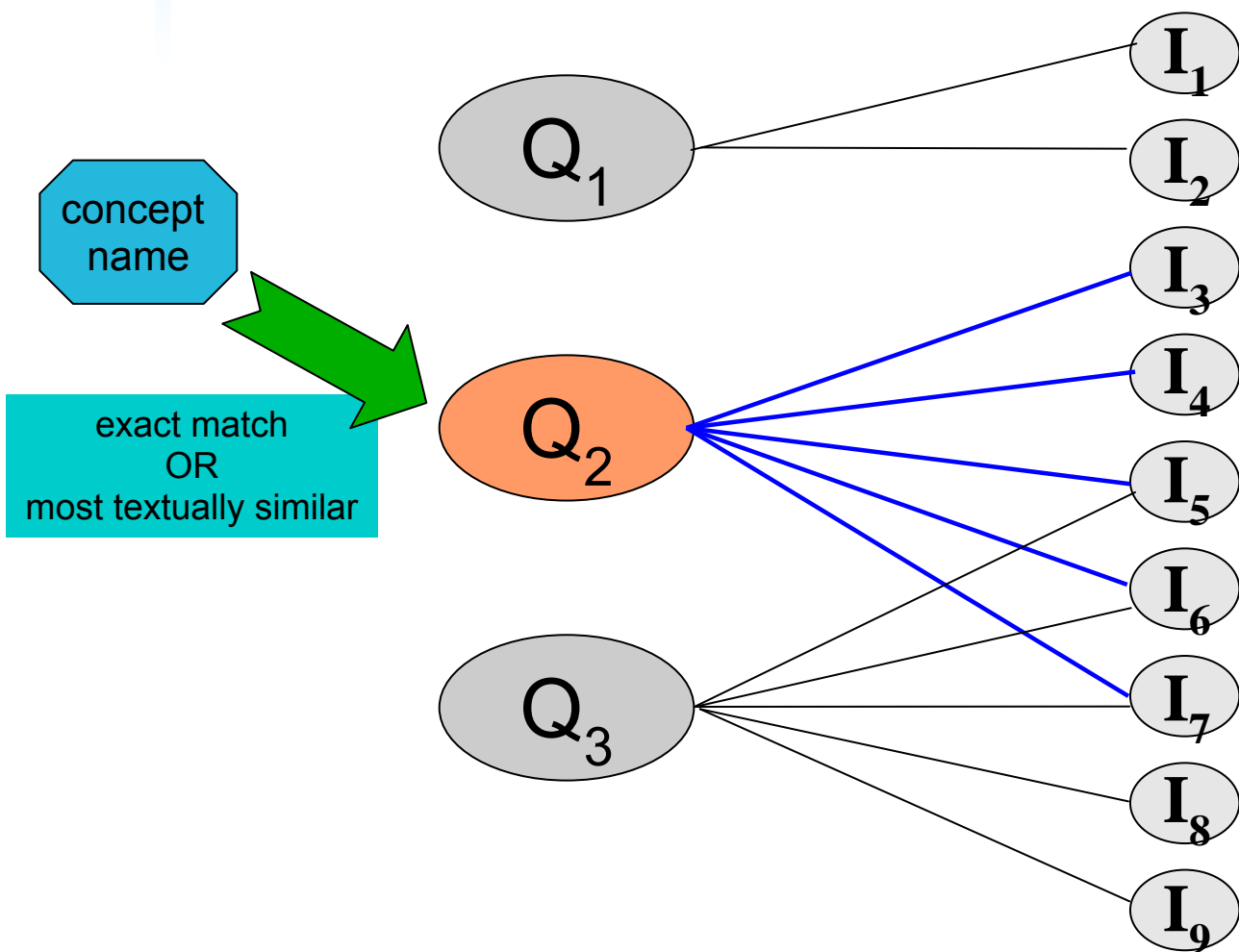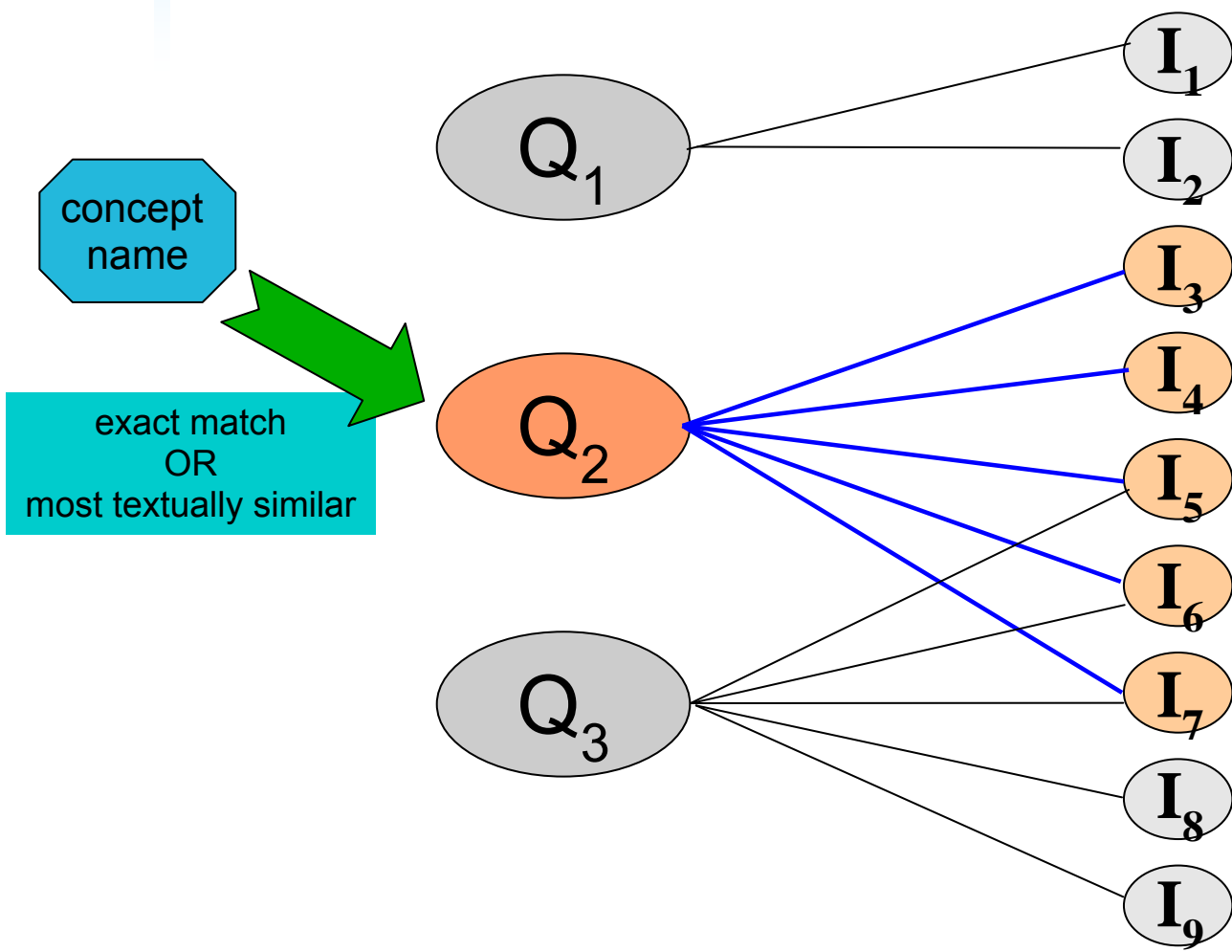
Method clickgraph

concept name

exact match
OR
most textually similar
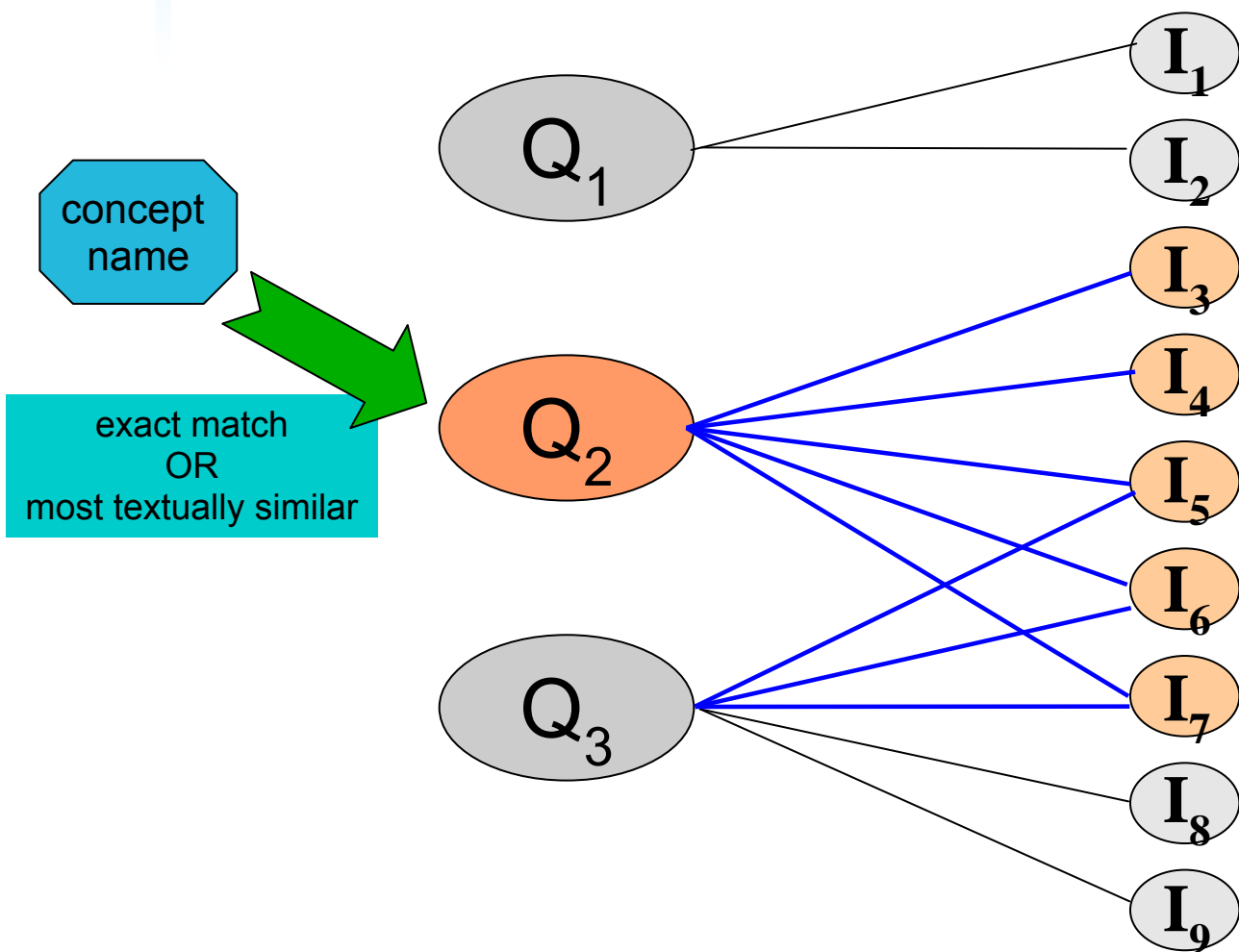
Q_1
Q_2
Q_3

I_1
I_2
I_3
I_4
I_5
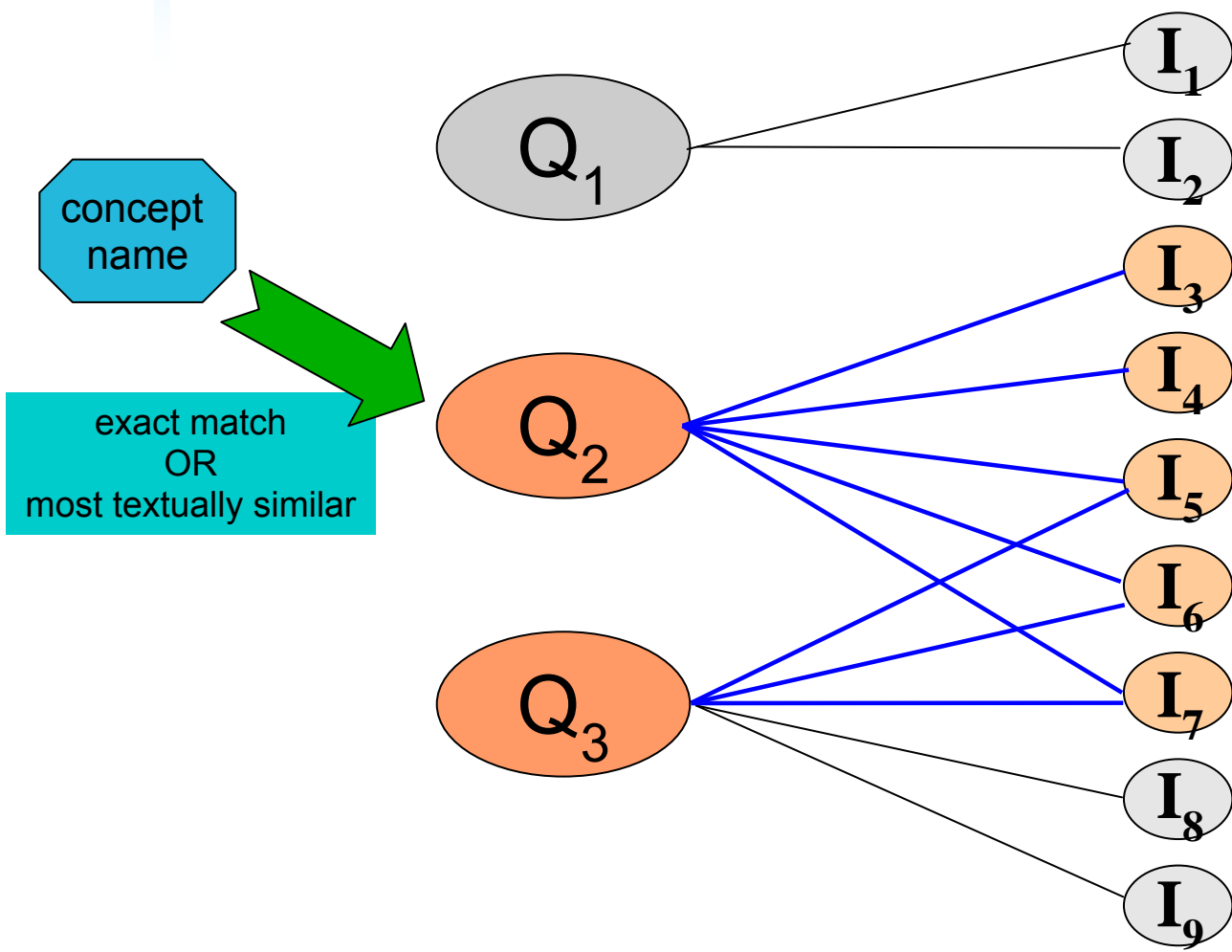I_6
I_7
I_8
I_9

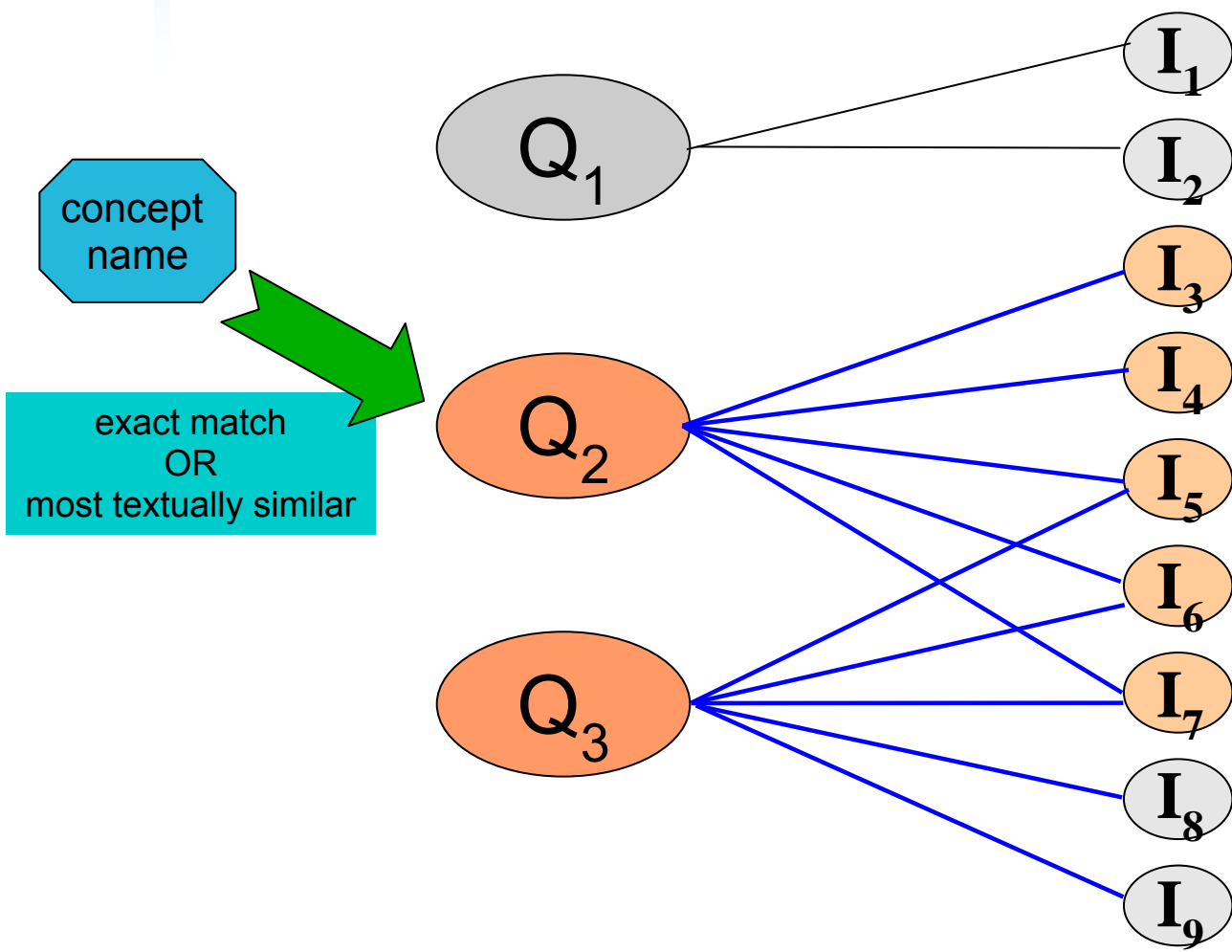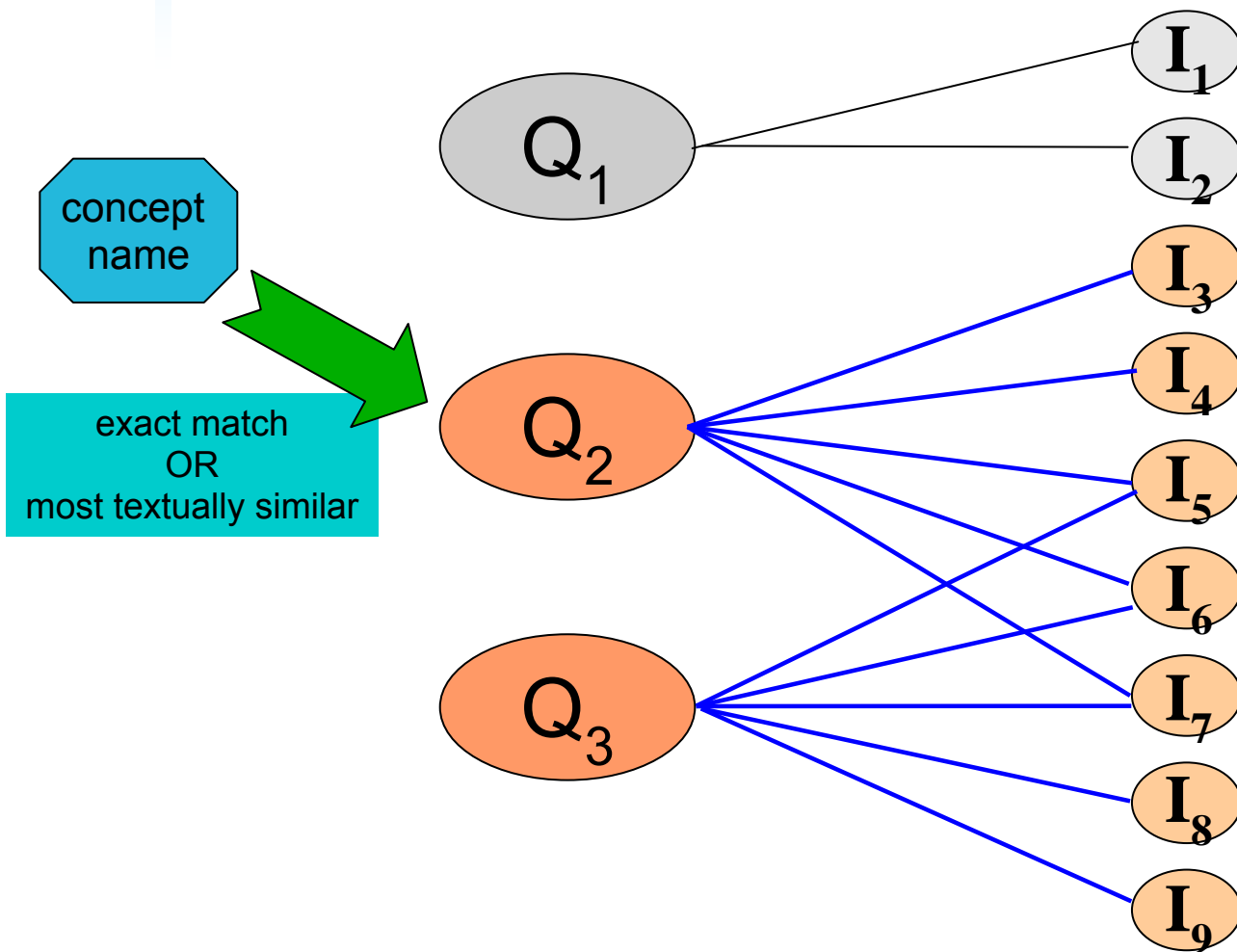# Method clickgraph

Method clickgraph

Method clickgraph

Method clickgraph

concept name

exact match
OR
most textually similar

Methods for positive sample selection using search logs

- ### *Method exact*

- ### Methods ``textual similarity''

  1. ### LM
  2. ### LM$_{stem}$
  3. ### LMS
  4. ### LMS$_{stem}$

- ### *Method clickgraph*
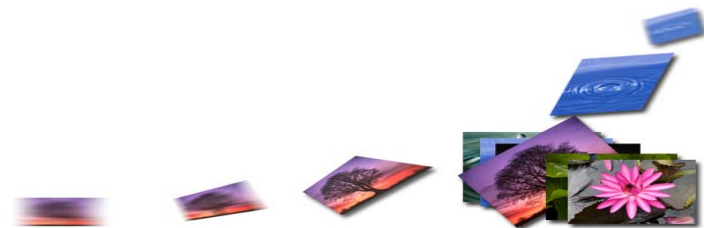
Research questions

1) How can we build classifiers for annotating images with concepts using clickthrough data?

- *methods for searchlog-based positive sample selection*
- *random negative sample selection*

2) What is the effectiveness of these concept classifiers?

T. Tsikrika, C. Diou, A. P. de Vries, and A. Delopoulos. Image annotation using clickthrough data. In CIVR 2009.

- contribution of searchlog-based training data is positive

3) What is the effect of the noise in the clickthrough data on the effectiveness of these concept classifiers?

- *measure the reliability of clickthrough data as image annotations*
- *measure their effect on the effectiveness*

Reliability of automatic clickthrough-based annotations

- How accurately do clickthrough-based image annotations correspond to explicit concept-based manual annotations

  - compare them to manual annotations

  - measure their agreement (i.e., find the true positives)

Experiments : datasets

- ***Image collection***
    - *photographic images provided by Belga news agency*

Title: GREECE BIKE DEMONSTRATION
Caption: Cyclists demonstrate in the centre of Athens on 20 October 2007. Hundreds of cyclists of all ages responded to a call made by cyclist associations on the Internet and turned up to take part in a protest bike ride against pollution and smog and in favour of a better traffic infrastructure, in spite of wet skies and rain.

Date:21/10/2007 18:18
Credit:EPA
Source:EPA
Dimensions:3000 x 2050
Photo:PANTELIS SAITAS
City:ATHENS
Country:Greece
ID:7705673

Experiments : datasets

- **Search logs provided by Belga news agency**
  - *101 days (June – October 2007)*
  - *professional users*
  - *17,861 unique ('lightly' normalised) queries*
    - *conversion to lower case*
    - *removal of punctuation, quotes, "and"*
    - *removal of names of major photo agencies*
  - *96,420 clicked images*
  - *power law distributions:*
    - *images-per-query*
   - *queries-per-image*

VITALAS concepts

- VITALAS concept lexicon consists of ~500 concepts
  - statistical analysis of image captions and image query logs
  - addition of concepts from MediaMill and LSCOM
  - feedback by Belga's professional archivists

- Selected 111 concepts for experiments
  - availability of search log-based positive samples
  - minimum: 10 clicked images

	number of clicked images per method					
	exact	LM	Lm_{stem}	LMS	LMS_{stem}	clickgraph
# concepts	17	63	71	88	99	49
mean	44.82	35.86	36.27	39.11	38.69	66.2
median	21	21	24	25.5	25	38
min	12	10	10	10	10	10
max	289	309	309	309	309	399

Experiments : clicks vs. manual annotations
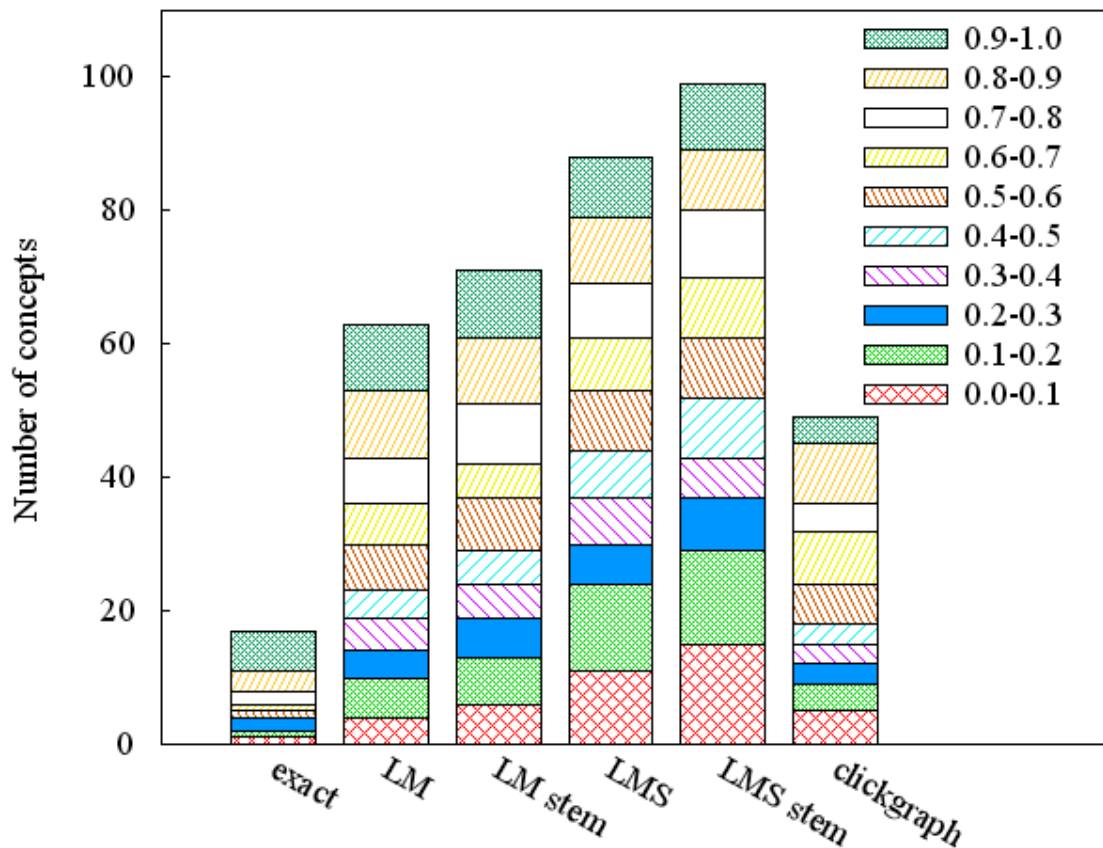
- Reliability of automatic clickthrough-based annotations
  - varies greatly across concepts
  - around 20% of the total number of concepts for each method reach agreement of at least 0.8.
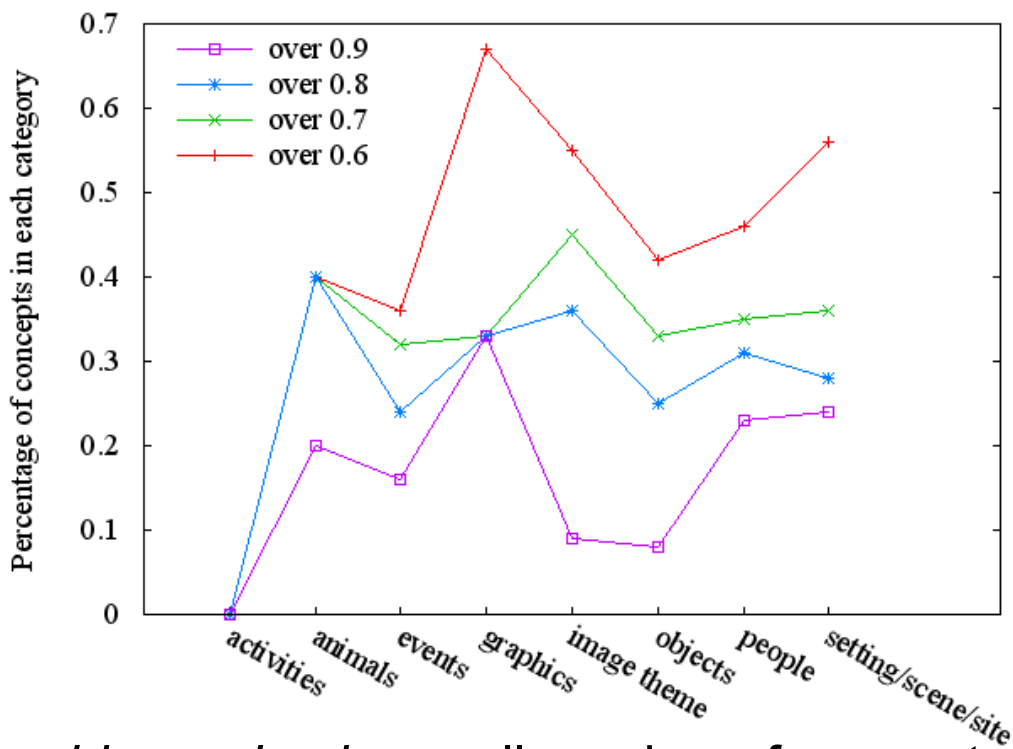
Which are the concepts with the most reliable annotations?

- Categorise concepts into LSCOM categories

Category	Category description (i.e., a concept is classified under this category if it describes ...)	Concepts	# of concepts	
image theme	a broad area of interest	classic, european-union, rally-motorsport, soccer, etc.	11	9.91 %
setting/scene/site	a specific place or land site or the environment in which something is set	airport, airport-terminal, amusement-park, art-gallery, atomium, bank, beach, bridge, church, court, damaged-building, disco, european-parliament, gas-station, highway, hospital-room, house, internet, mountain, parliament, road, school, stock-exchange, theatre-building, traffic	25	22.52 %
people	a human being, a group of human beings, or human body parts	abused-child, ac-milan-soccer, agricultural-people, anderlecht, arsenal-fc, artist, baby, belgium-royal, bush, child, club-brugge, factory-worker, fashion-model, federer, finger, girl, government, jacques-chirac, king, lawyer, parent, pope-benedict, queen, red-devils, sex, teenager	26	23.42 %
animals	an animal or a group of animals	animal, cat, dog, horse, lion	5	4.51 %
objects	a physical, tangible, and visible entity, excluding people, animals, & entities occupying land sites	apple, belgian-flag, boat, bus, car, computer, european-flag, fish, flag, street-sign, television, usa-flag	12	10.81 %
activities	a specific behaviour or action taking place	beach-leisure, car-racing, children-care, children-playing	4	3.60 %
events	something that happens at a given place and time, including natural events (e.g., fires and avalanches) and social events (e.g., shows, social functions, contests, and competitions)	airplane-crash, australian-open, award, cannes-festival, car-accident, champions-league, davis-cup, flood, demonstration, earthquake, election, explosion, fashion-show, festival, fire, flood, formula-one, gala, goal, hurricane-typhoon, memorial-services, olympic-games, parade, roland-garros, storm, tour-of-flanders	25	22.52 %
graphics	any form of artificially generated visual content	cartoon, illustration, logo	3	2.71 %
Total			111	100.00%

- *activities, graphics, animals*: small number of concepts

- around 50% of concepts in each category (apart from *activities*) reach an agreement of at least 0.6

- concepts with agreement over 0.9: *setting/scene/site*, *people*, *events*

- *objects* type concepts do not perform that well
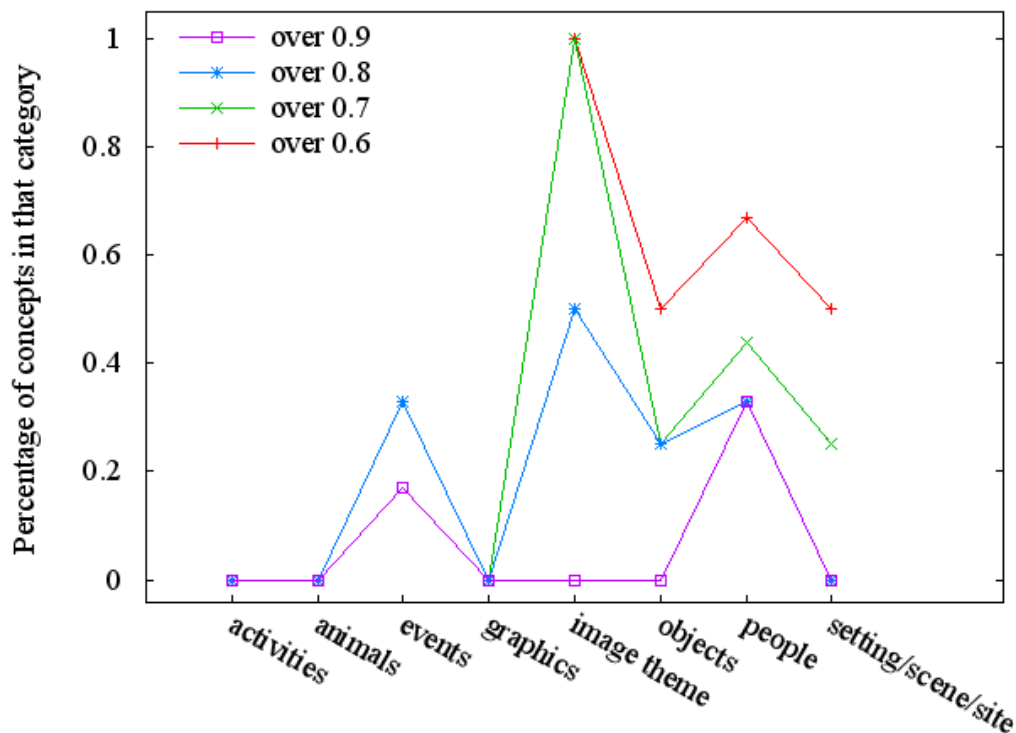   - queries/textual metadata do not usually refer to the visually obvious

Which are the concepts with the most reliable annotations?

- Examine only the concepts with the highest number of clicked images
  - top 20% of concepts with highest number of clicked images in a method
  - 25 of the 111 concepts

- concepts with agreement over 0.9: *people*, *events*

Conclusions

- Clickthrough-based image annotations are reliable for many concepts

- Concept categories are not good predictors of their reliability

- Despite being noisy, clickthrough data can be used as training samples for building effective concept classifiers

- Given their many advantages, should we aim to provide such clickthrough data in the context of an image annotation or image retrieval evaluation task? How?

THANK YOU!