Characteristics of query topics for the WikipediMM task at ImageCLEF 2008

Jana Kludas, UNIGE, Switzerland
Theodora Tsikrika,CWI, Netherlands

ImageCLEF pre-workshop
16.9.2008, Aarhus Denmark

WikipediaMM Task

Description:

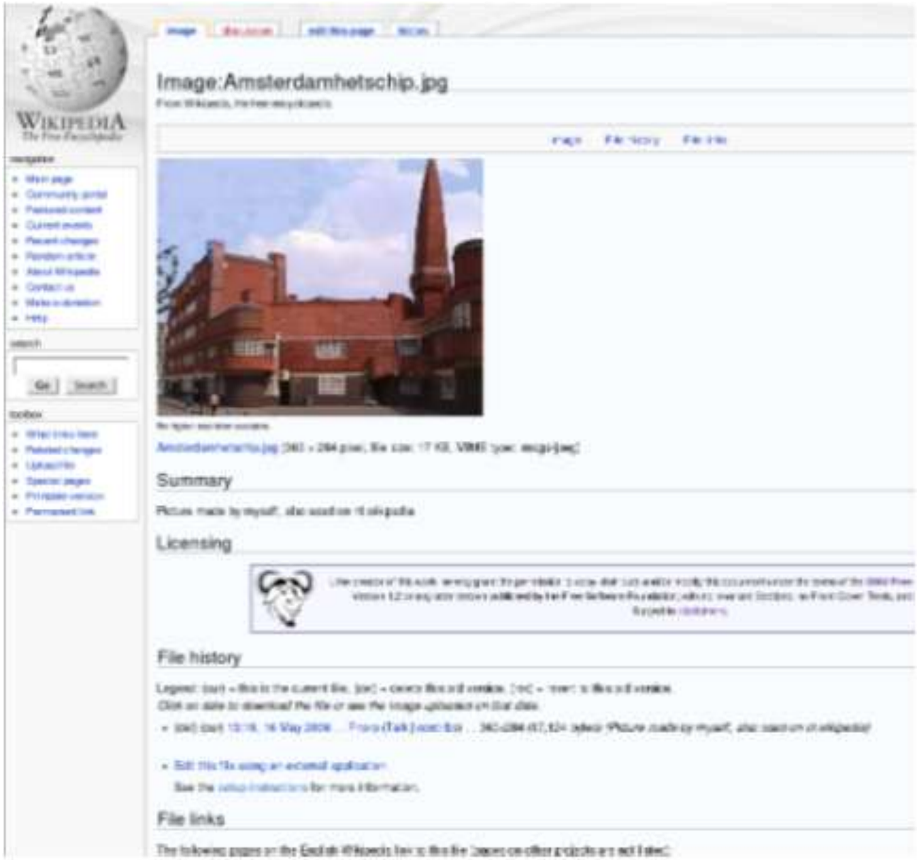
- ad-hoc image retrieval
- collection of Wikipedia images
  - large-scale
  - heterogeneous
  - user-generated annotations
  - availability of multi-lingual data
- diverse multimedia information needs

Aim:

- investigate mono-media and cross-media retrieval approaches
- focus on fusion/combination of evidence from different modalities
- attract researchers from both text and visual retrieval communities
- support participation through provision of appropriate resources

WikipediaMM collection

- 151,590 images
  - wide variety
  - global scope
  - JPEG, PNG formats
- Annotations
  - user-generated
  - highly heterogeneous
  - varying length
  - noisy
  - semi-structured
  - monolingual (English)



- Used in INEX MM 2006 - 2007

MM task topics @INEX 2006/2007

2006:

	MMfragments	MMimages	Total
Number of Topics	9	13	22
Avg. num. terms in <title>	2.7	2.4	2.6
Number of topics with src:	1	6	7
Number of topics with concept:	0	2	2

2007:

	MMfragments	MMimages	All
Number of topics	19	20	39
Average number of terms in <title>	3.21	2.35	2.77
Number of topics with <mmtitle>	6	10	16
Number of topics with src:	2	7	9
Number of topics with concept:	4	6	10
Number of topics with both src: and concept:	0	3	3

- NEXI format: XML based
- small number of topics
- not many multimedia hints
- text-based runs always best

Topic format

descriptions of multimedia information needs

- keywords
- optional: one or several visual and conceptual evidences

```
<topic>

<number> 62 </number>

<title> cities by night </title>
<concept> building </concept>
<image> http://www.bushland.de/hksky2.jpg </image>
```

```
<narrative> I am decorating my flat and as I like photos of cities at night,
I would like to find some that I could possibly print into posters. Photos
of cities (or the earth) from space are not relevant. I would like to find
photos of skylines or photos that contain parts of a city at night
(including streets and buildings).  </narrative>

</topic>
```

Topic development 2008

candidate topic pool:

(I) topics previously used in INEX 2006-2007 MM task

(II) topics submitted by this year's task participants

- initial topic statement
- exploration phase with assessment of top 25 results
- feedback search with assessment of top 100
- write <narrative>, optionally add <image> and <concept>
- finalize topic

goal: diverse set of topics with

- different characteristics (visual, semantic …)
- different amount of multimedia resources
- different domains: narrow/broad

Topic classification

visual:

- topics with visual, highly discriminating properties e.g. ‘blue flower’
- CBIR systems are likely to solve them

textual:

- topics containing proper nouns of persons, buildings, locations … e.g. ‘Da Vinci paintings’
- correctly annotated images easily found with text-only approaches

semantic:

- topics with a complex set of constraints, need world knowledge or contain ambiguous terms e.g. ‘labor demonstrations’, ‘plant’
- most likely no modality alone is effective

How to determine topics for each class?

Topics in 2008

intuitive classification of candidate topics according to class definition

- **5 visual**: blue flower, red ferrari, white cat …
- **35 textual**: oak tree, daily show, George W Bush, Golden gate bridge, can or bottle of beer …
- **35 semantic**: mountains under sky, winter landscape, people riding bicycles, famous buildings of Paris, plant …

topic statistics

	all	visual	textual	semantic
Number of topics	75	5	35	35
Average number of terms in title	2.64	2.2	2.3	2.9
Number of topics with image(s)	43	3	22	18
Number of topics with concept(s)	45	4	16	25
Number of topics with both image and concept	28	3	11	14
Number of topics with text only	15	1	8	6

Task submissions

- 12 groups submitted 77 runs
- many text-only runs, **but** same amount of fusion runs!
- type of methods:

textual	Txt	35
visual	Img	5
concept	Con	0
textual/visual	TxtImg	22
textual/concept	TxtCon	13
textual/visual/concept	TxtImgCon	2

- best run: still text-only

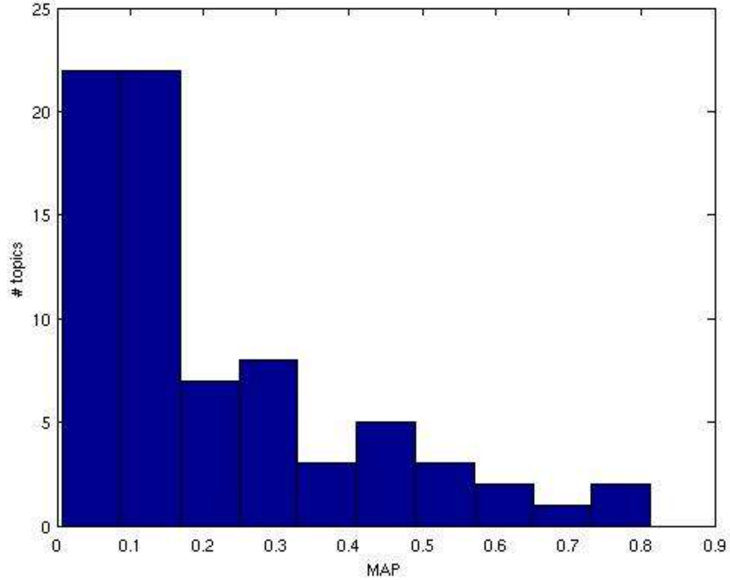
	group	runID	Modality	MAP	P@20
1	upeking	zzhou3	TXT	0.3444	0.3794
2	cea	ceaTxtCon	TXTCON	0.2735	0.3225
3	ualicante	IRnNoCamel	TXT	0.2700	0.3075

Evaluation on task results

- use only top 75% of the runs to eliminate noise and buggy runs

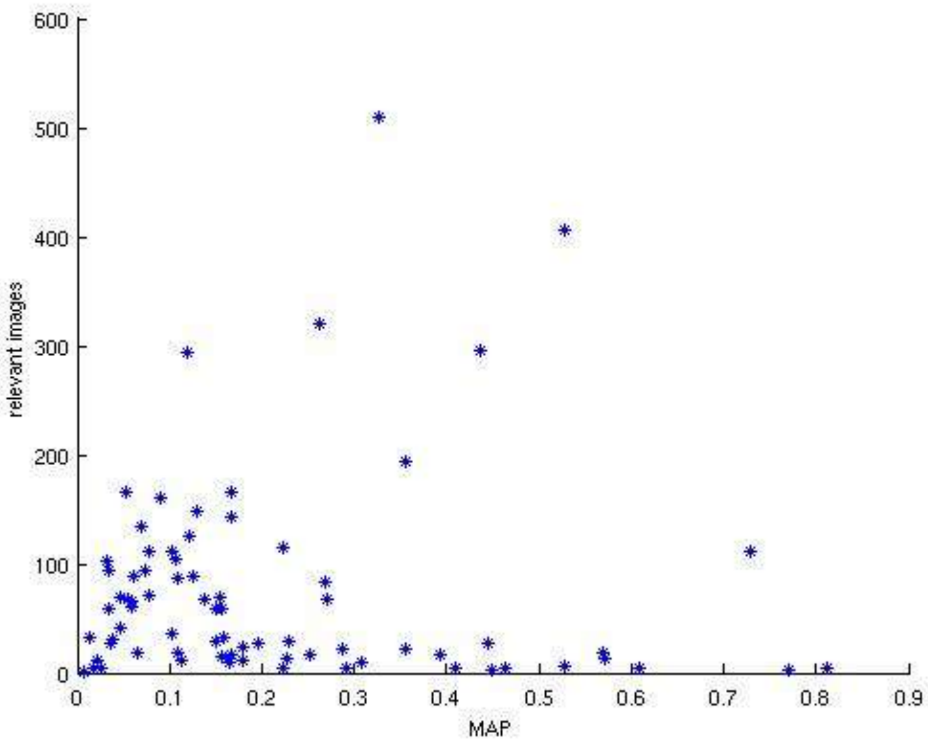
topic difficulty

easy: MAP>0.4
medium: 0.2<MAP≤0.4
hard: 0.1<MAP≤0.2
very hard: MAP≤0.1



Difficulty vs topic classification

	Total
eas	13
me	15
har	24
ver	23
Tot	75



- visual a
- too ma

(? due to many narrow topics: 29 with less than 25 relevant docs)

Topic classification vs best method

best method: maximum average MAP over runs that use the same resources

	TxtImg	Txt	TxtCon	TxtConImg	
visual	1	1	2	1	5
textual	5	3	16	11	35
semantic	6	4	13	12	35
	12	8	31	24	

- initial classification not accurate: 1/5, 19/35, 25/35

- most of the topics best solved with fusion methods: 67/75 - need for efficient fusion methods

Topic resources vs best method

	TxtImg	Txt	TxtCon	TxtConImg	
Images	4	6	15	18	43
Concepts	6	6	19	14	45
Images/Concepts	2	4	10	12	28
Text-only	4	0	7	4	15

- topic images and concepts can turn out as not useful
- characteristic of topic not depending on resources
- topic resources and resources of best runs not related – due to query expansion/feedback methods??

Conclusions

(I) hard to determine a topic’s class visual/textual/semantic, depends on:

- content of the topic
- relevant results in the collection (quality of annotation, images …)
- (a bit on) topic resources (relevance of example image)

(II) most topics best solved with fusion approach, also textual ones

Next year’s topics

- same amount of visual, textual and semantic topics based on this year’s experience

- avoid too hard topics (MAP~0)

- groups of topics
  - ‘mountains’, ‘mountains under sky’, ‘mountains under sky with snow’
  - ‘bridges’, ‘bridges at night’, ‘bridges at daylight’
  - see how constraints influence the retrieval result, which types of approaches perform best