SINAI at ImageCLEFmed 2008

Arturo Montejo Ráez
amontejo@ujaen.es

University of Jaén
Spain

Introduction

- The 2008 data
- Query expansion
  - MeSH
  - UMLS
- Experiments
- Conclusions
- Further work

The 2008 data

- New collection: subset of Goldminer

- Organizers information: ImageCLEF2008.xml


```
<Record>
    <figureID>...
    <figureURL>...
    <caption>...
    <title>...
    <pmid>
    <articleURL>...
    <imageURL>...
    <imageLocalName>...
</Record>
```

Caption of the figure
(one or more images)

Title of article

URL of full text article

ID of list of results
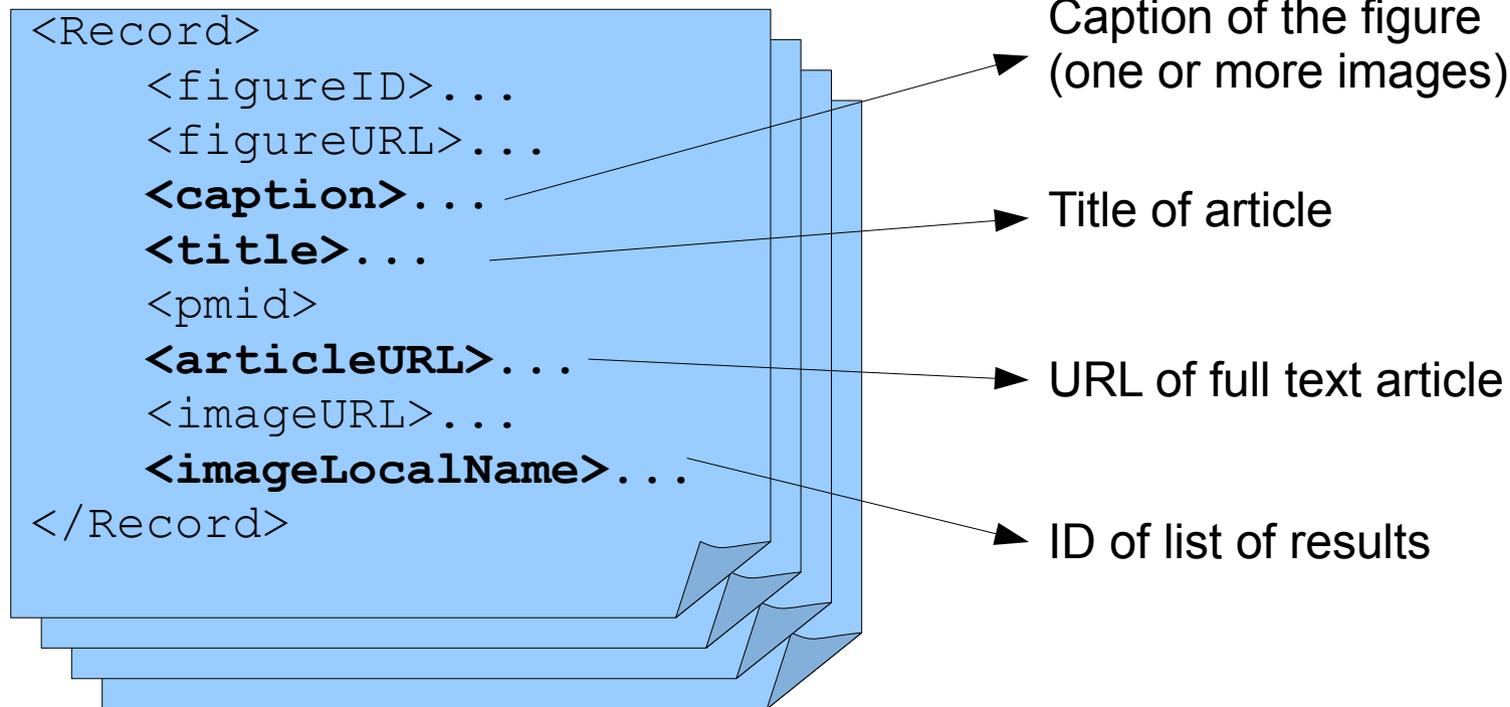
Extract textual information

```
<Record>
    <figureID>...
    <figureURL>...
    <caption>...
    <title>...
    <pmid>
    <articleURL>...
    <imageURL>...
    <imageLocalName>...
</Record>
```

Extract lists of URLs → Download articles → Filter HTML → Extract and add information

```
<Record>
    <FIGUREURL>...
    <IMAGEURL>...
    <TITLE>...
    <ARTICLEURL>...
    <IMAGELOCALNAME>...
    <CAPTION>...
    <PMID>...
    <FIGUREID>...
    <sinaiSection>...
    <sinaiArticle>...
    <sinaiSectionTitle>...
    <sinaiCaption>...
    <sinaiKeywords>...
    <sinaiID>...
</Record>
```
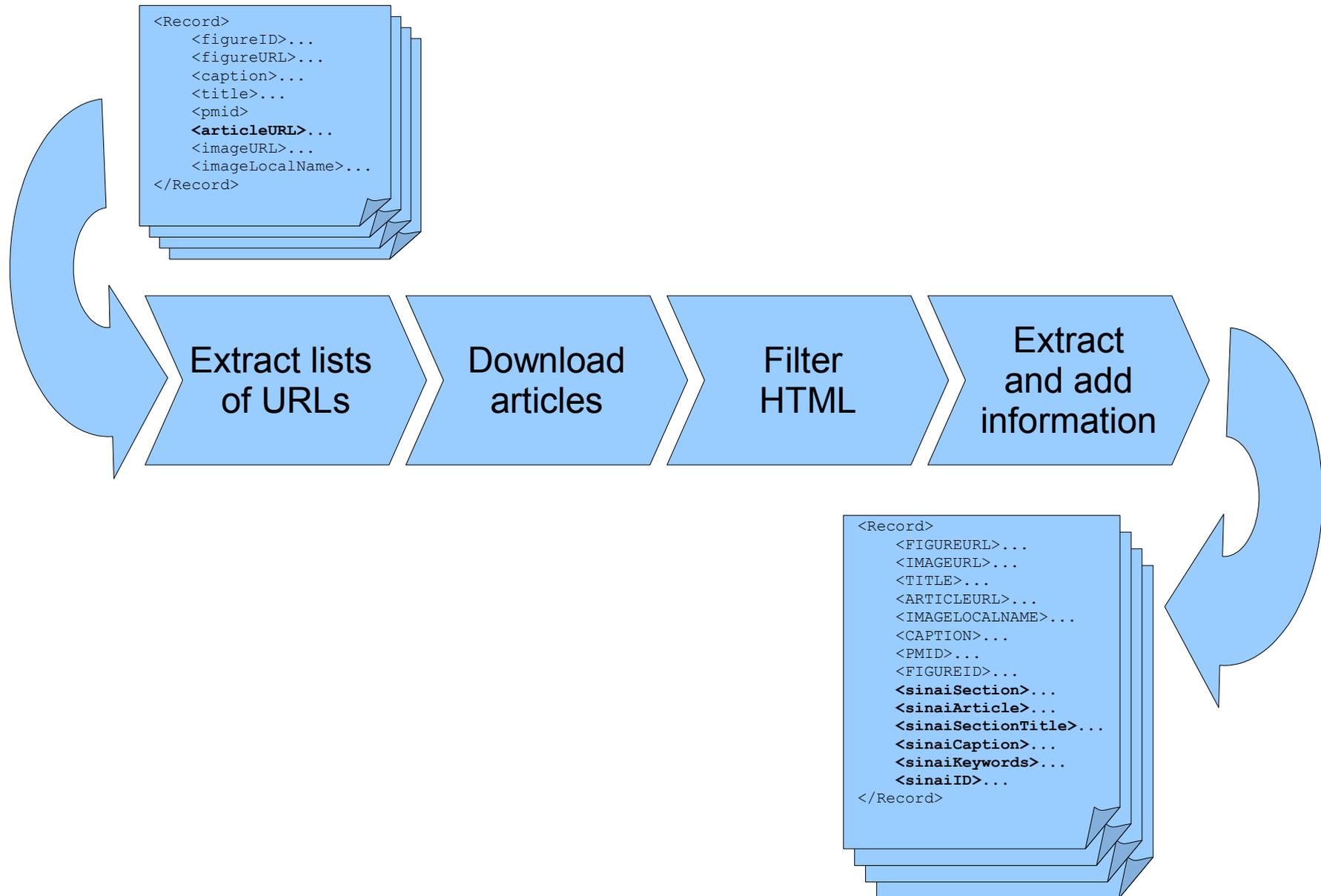
Generating new collections

```
<Record>
    <figureID>...
    <figureURL>...
    <caption>...
    <title>...
    <pmid>
    <articleURL>...
    <imageURL>...
    <imageLocalName>...
    <figureID>...
    <figureURL>...
    <caption>...
    <title>...
    <pmid>
</Record>
```

Caption + Title

```
<DOC>
    ...............
    ...............
    ...............
    ...............
    ...............
</DOC>
```

CT

Caption + Title
+ Article

```
<DOC>
    ...............
    ...............
    ...............
    ...............
    ...............
</DOC>
```

CTA

Caption + Title
+ Section

```
<DOC>
    ...............
    ...............
    ...............
    ...............
    ...............
</DOC>
```

CTS
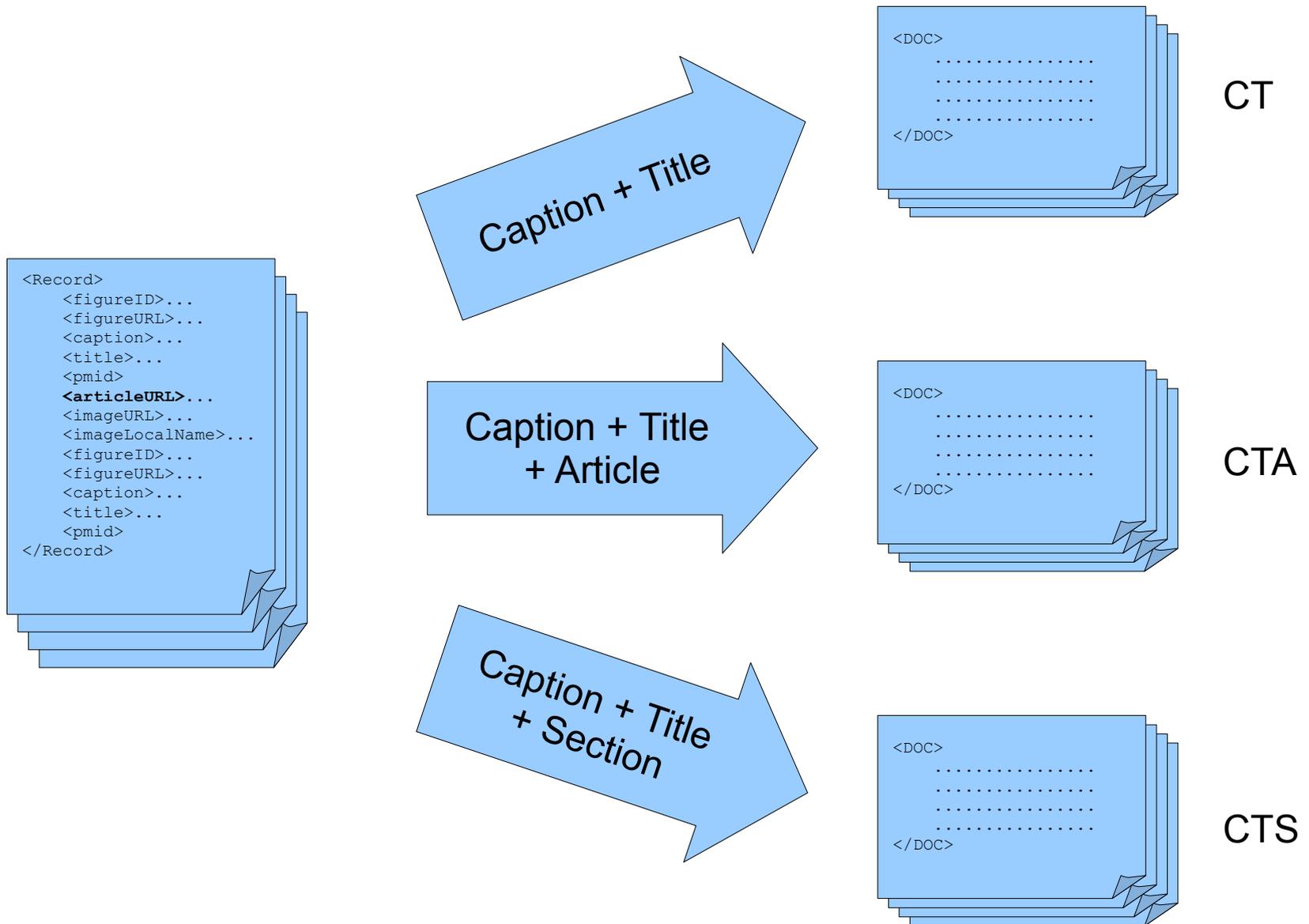
Query expansion

- MeSH:

  - Same as in ImageCLEFmed 2007

  - Good results in last year

  - 2006 version of the tree

- UMLS:

  - Using MetaMap

  - 2008 version

  - Several sources (included MeSH)

MeSH query expansion

- Restricted categories: [Chevallet, 2005]

  - **A**: Anatomy

  - **C**: Diseases

  - **E**: Analytical, Diagnostic and Therapeutic Techniques and Equipment

- MeSH records used:

  - MeSH Heading

  - Entry terms

MeSH query expansion

Query

a b c d ...

Expansion

Expanded Query

a, t1, t2
b
c d, t3, t4
...

MH = a
ENTRY = t1
ENTRY = t2
...

MH = t3
ENTRY = **c d**
ENTRY = t4
...

MeSH
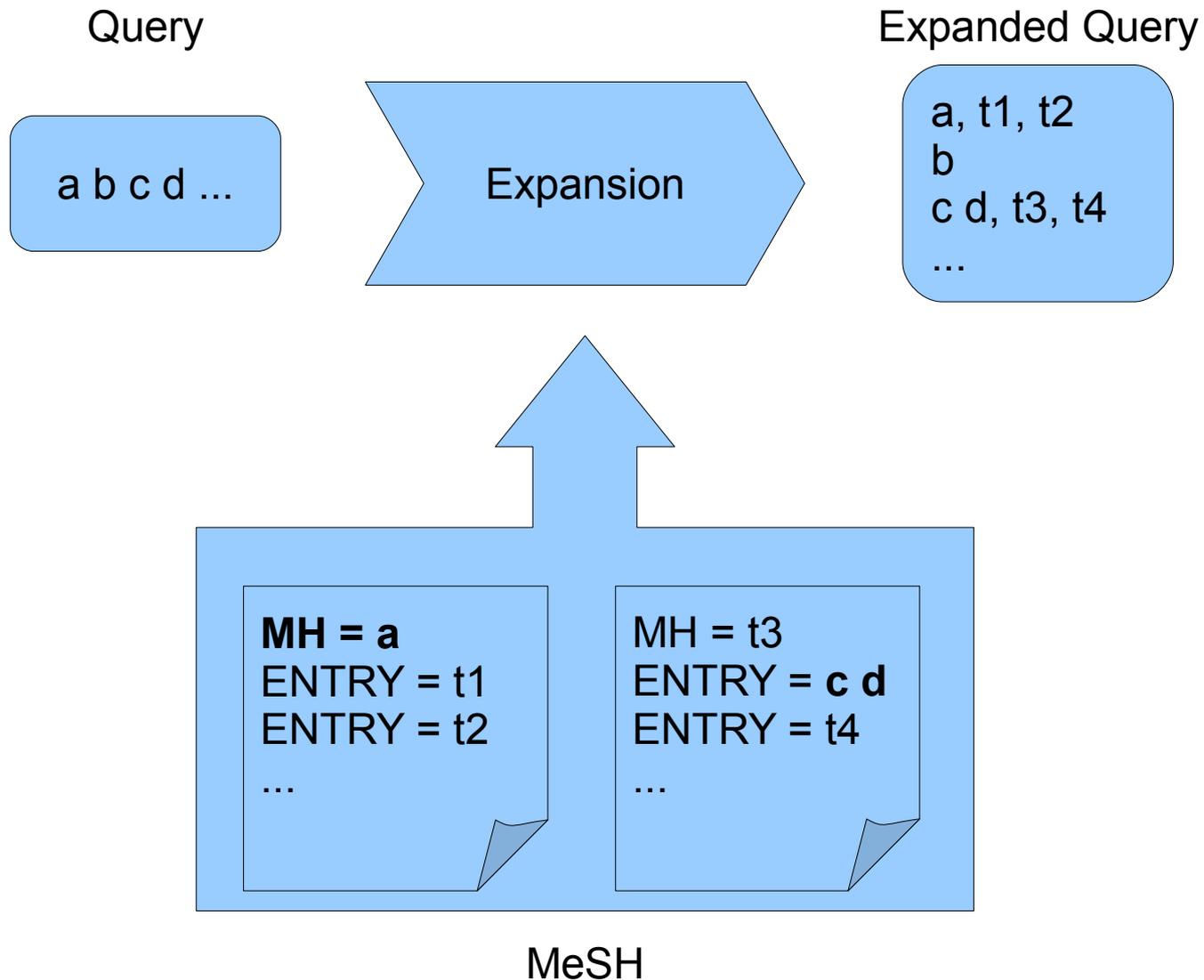
UMLS query expansion

- Restricted semantic types (based on 2007 data):

  - **bpoc**: Body Part, Organ, or Organ Component

  - **diap**: Diagnostic Procedure

  - **dsyn**: Disease or Syndrome

  - **neop**: Neoplastic Process

- Using MetaMap:

  - Use all Meta Candidates (not only Meta Mapping)

UMLS query expansion

Query

Show me
x-ray images
of a hip joint
with prosthesis

Phrases

Show

me x-ray images

of a hip joint

with prosthesis

Meta Candidates

<none>

Image
(Medical Imaging)
X-ray
(Diagnostic radiologic
examination)
X-ray
(Plain x-ray)

Hip (Bone structure of ischium)

<none>

Words to expand

Image
Medical
Imaging
X-ray
Diagnostic
radiologic
examination
Plain
Hip
Bone
structure
of
ischium

MetaMap

Experiments

- Textual and Mixed experiments.
  - Official textual experiments: 9
  - Official mixed experiments: 1
- Visual results obtained from organizers.
  - Obtained with FIRE software.

Textual experiments

- 3 collections:
    - CT (baseline)
    - CTA
    - CTS

- 3 topic sets:
    - No expansion (baseline)
    - MeSH expansion
    - UMLS expansion

MAP	CT	CTA	CTS
Baseline	0.2480	0.1982	0.1784
MeSH	**0.2792**	0.2057	0.1582
UMLS	0.2275	0.1781	0.1429

Mixed experiments

- Weighting score of visual and textual results.
    - scoreM = scoreV*alfa + scoreT*(1-alfa)
- Only one experiment (best weight of 2007)
    - Alfa = 0.2
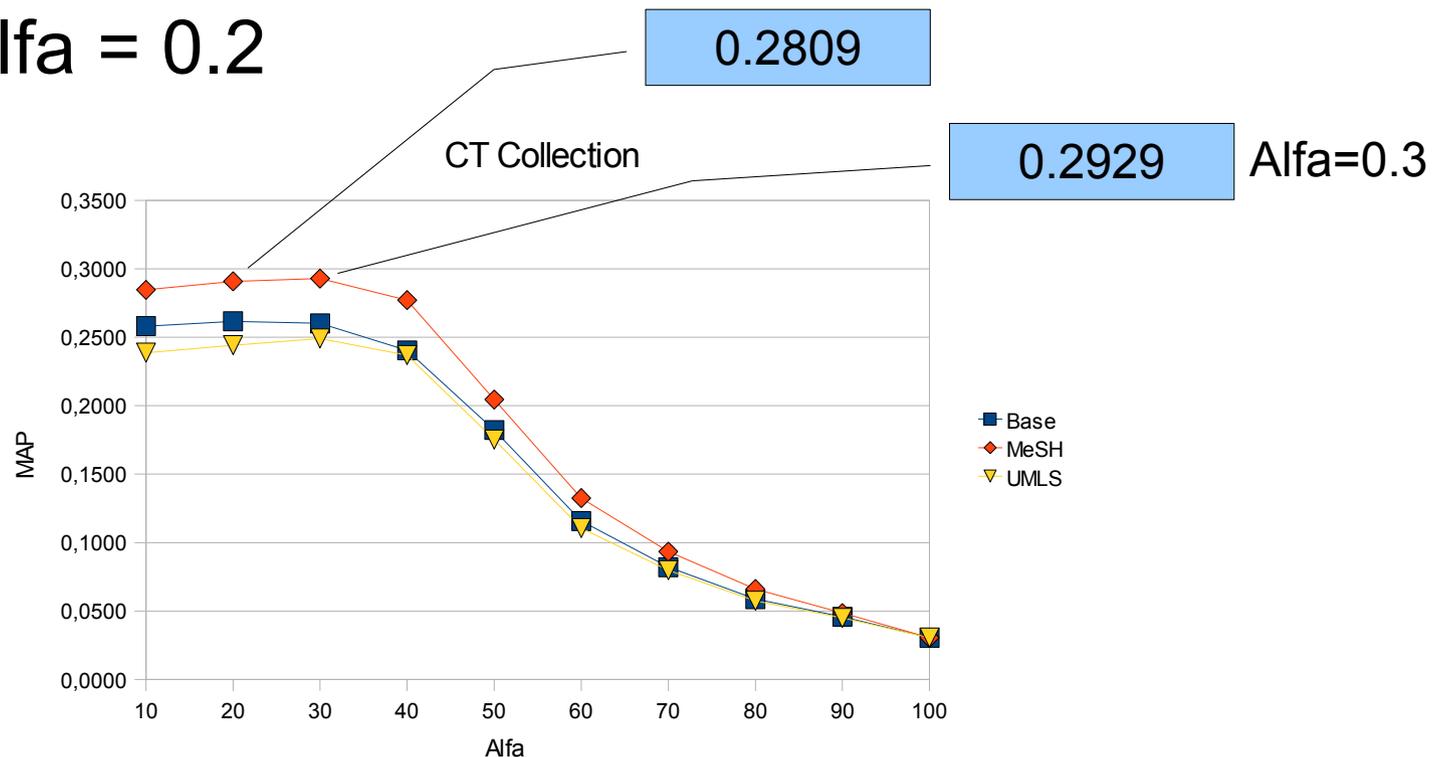
With different expansions

Alfa	Base	MeSH	UMLS
10	0,2581	0,2846	0,2387
20	0,2616	0,2908	0,2442
30	0,2602	0,2929	0,2491
40	0,2404	0,2771	0,2369
50	0,1822	0,2045	0,1753
60	0,1158	0,1325	0,1106
70	0,0821	0,0934	0,0798
80	0,0588	0,0661	0,0575
90	0,0459	0,0486	0,0450
100	0,0305	0,0305	0,0306
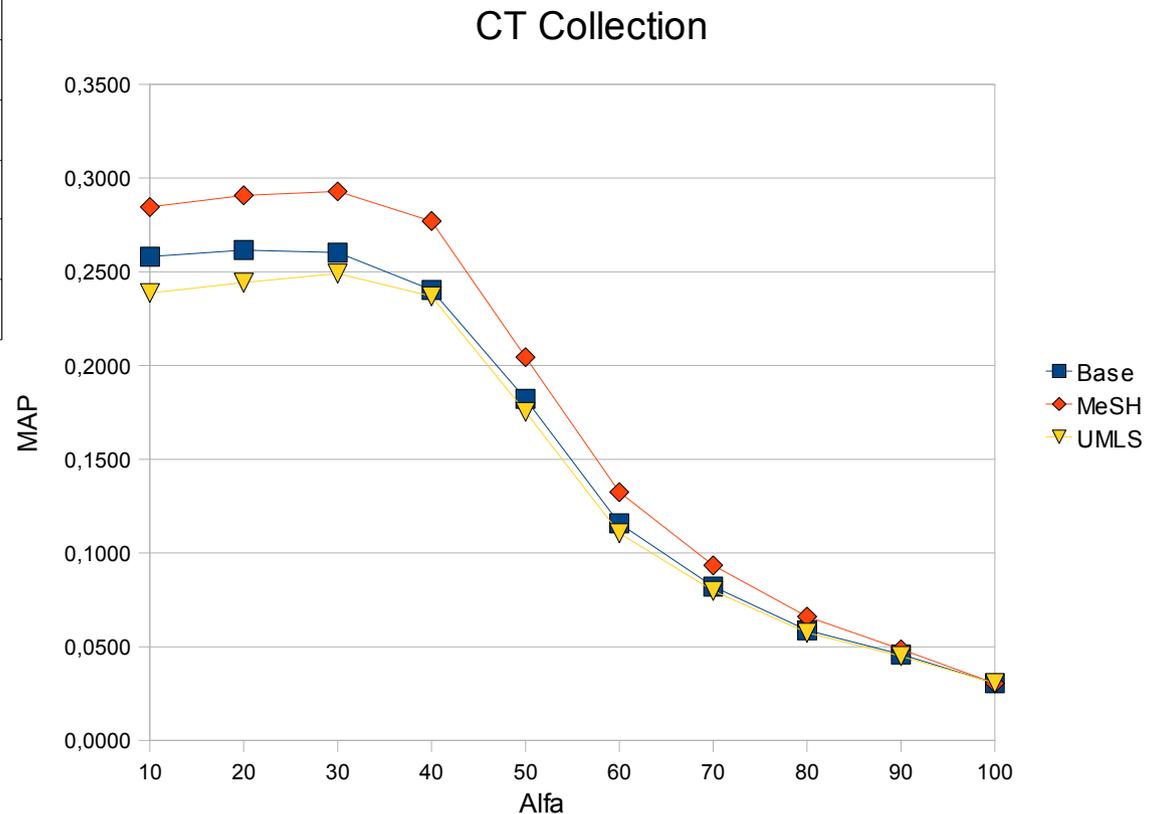
CT Collection

With different collections

Alfa	CT	CTA	CTS
10	0,2581	0,2001	0,1812
20	0,2616	0,1989	0,1819
30	0,2602	0,1983	0,1816
40	0,2404	0,2041	0,1761
50	0,1822	0,1757	0,1444
60	0,1158	0,1182	0,1027
70	0,0821	0,0851	0,0762
80	0,0588	0,0622	0,0572
90	0,0459	0,0462	0,0443
100	0,0305	0,0305	0,0306

Original topics
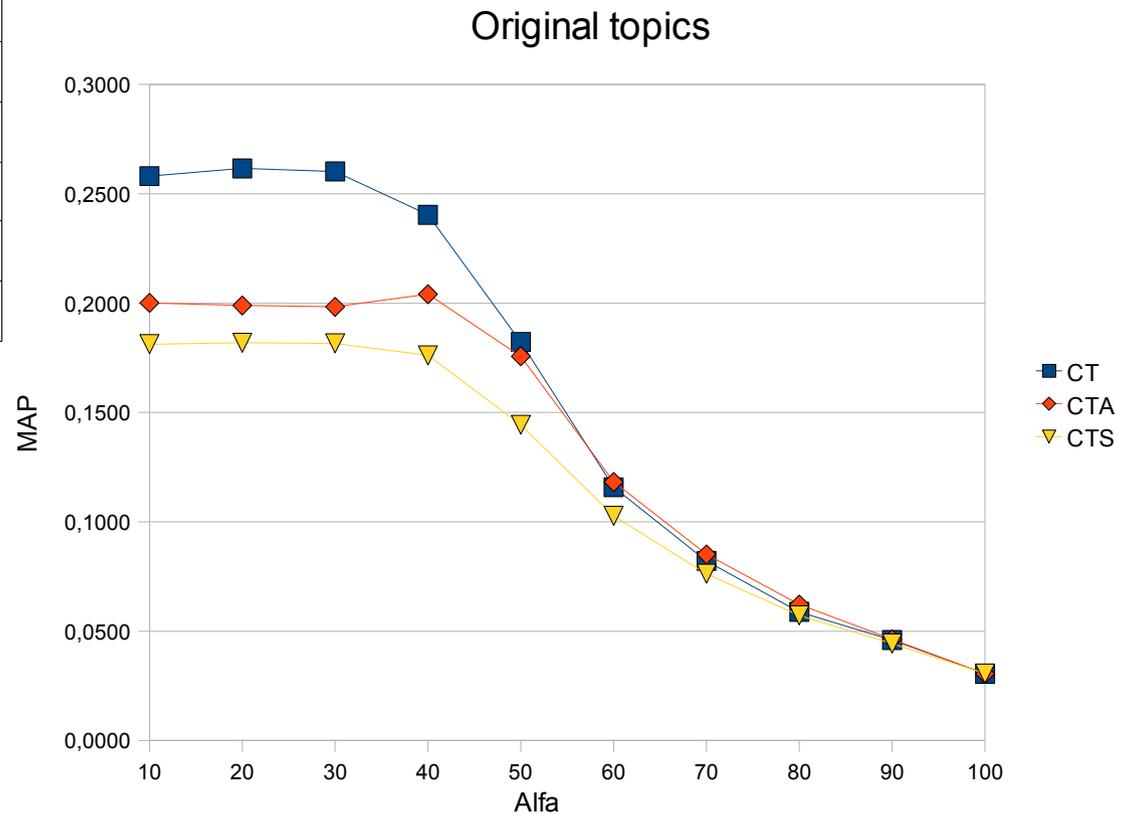
Conclusions

- Mix visual and textual results is better.

- Bad UMLS expansion algorithm.

- Select the textual information is very important.

  - Collection size: CT < CTS < CTA

  - Results MAP: CT > CTA > CTS

Further work

- Obtain a more precise textual information:

    – Find HTML references to the image.

    – Syntactic structures of type "Figure 1 ...".

- Expand the queries with UMLS Metathesaurus using other algorithm (candidates filtering?)

- Investigate when the expansion obtains an improvement. Only expand certain queries.

- Fusion algorithm: how to set alpha?