LABERINTO at ImageCLEF 2011
Medical Image Retrieval Task

Mariano Crespo Azcárate     Jacinto Mata Vázquez     Manuel J. Maña López
• Introduction
  - LABERINTO in ImageCLEF
  - MeSH Ontology

• Query Expansion Strategies
  - Using MeSH to Expand Queries
  - Techniques based on MeSH Tree Structure
  - Techniques based on Entry Terms

• Experiments and Results

• Conclusions and Future Works
1. INTRODUCTION

LABERINTO

- LABERINTO → 1st Participation in ImageCLEF.
- Medical Retrieval Task → Ad-hoc Image-Based Retrieval.
- 10 Runs sent.
- Retrieval type: Textual.

OBJECTIVE → To improve retrieval efficiency using MeSH to expand queries.
• MeSH (Medical Subject Headings) is a controlled vocabulary, produced and maintained by the U. S. National Library of Medicine.

• There are currently over 26,000 descriptors or Main Headings and almost 180,000 alternative expressions (ENTRY TERMS).
• MeSH offers many possibilities for expanding the query terms.
Using MeSH to Expand Queries in Medical Image Retrieval

2. Query Expansion Strategies

Query: “Mitral Valve”
- By Terms → There aren’t any descriptor
- Both Terms Together → MeSH descriptor.

Solution: Pre-process each query into N-grams
• QUERY: “breast cancer mammogram”

• N-GRAMS
  1) breast
  2) breast cancer
  3) breast cancer mammogram
  4) cancer
  5) cancer mammogram
  6) mammogram
2. QUERY EXPANSION STRATEGIES

TECHNIQUE BASED ON MeSH TREE STRUCTURE

• This strategy is based on the tree structure where MeSH organises its descriptors.

  Process each N-Gram of the query
  • If N-Gram is no descriptor → No expansion.
  • If N-Gram is a descriptor:
    Option 1. If descriptor is a parent node → Expansion with child descriptors.
    Option 2. If descriptor has no children → No expansion.

  Nervous System [A08]
  Central Nervous System [A08.186]
  ▶ Brain [A08.186.211]
    Blood-Brain Barrier [A08.186.211.035]
    Brain Stem [A08.186.211.132] +
    Cerebral Ventricles [A08.186.211.276] +
    Limbic System [A08.186.211.464] +
    Mesencephalon [A08.186.211.653] +
    Prosencephalon [A08.186.211.730] +
    Rhombencephalon [A08.186.211.865] +
    Meninges [A08.186.566] +
    Spinal Cord [A08.186.854] +
2. QUERY EXPANSION STRATEGIES

TECHNIQUES BASED ON ENTRY TERMS

1. If N-Gram is a descriptor
   → Expansion using all Entry Terms of descriptor.

2. If N-Gram is not a descriptor
   → Entry Term
      Option 1. Entry Term → Expansion using all Entry Terms of the descriptor.
      Option 2. No Entry Term → No Expansion.

PREFERRED CONCEPT

ENTRY TERMS

(1): Leukemia, Myeloid
(2): Myeloid Leukemia
(3): Leukemias, Myeloid
(4): Myeloid Leukemias

SELECTED ENTRY TERMS

(2): Myeloid Leukemia
• Three different indexes created:

- Captions (C) → Contains text of captions of each image.

- Image Reference (IR) → Contains sections of paper referred to each image indexed.

- Full Text (FT) → Contains full text of each paper.
• Three different runs for each indexing sent:

- Baseline (B) → Original Queries.

- Concept Tree (CT) → Queries expanded with technique based on MeSH Tree Structure.

- Entry Terms Preferred Concept (ETPC) → Queries expanded with techniques based on Entry Terms.

Moreover:

- Entry Terms (ET) → Queries expanded with techniques based on Entry Terms.
### 3. EXPERIMENTS AND RESULTS

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Run</th>
<th>MAP</th>
<th>P10</th>
<th>P20</th>
<th>Rprec</th>
<th>Bpref</th>
<th>Num_Rel_Ret</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Laberinto_CTC</td>
<td>0.2172</td>
<td>0.3467</td>
<td>0.3017</td>
<td>0.2369</td>
<td>0.2402</td>
<td>1471</td>
</tr>
<tr>
<td>4</td>
<td>Laberinto_BC</td>
<td>0.2133</td>
<td>0.3400</td>
<td>0.3067</td>
<td>0.2363</td>
<td>0.2384</td>
<td>1469</td>
</tr>
<tr>
<td>16</td>
<td>Laberinto_ETPCC</td>
<td>0.1939</td>
<td>0.2933</td>
<td>0.2617</td>
<td>0.2089</td>
<td>0.2198</td>
<td>1526</td>
</tr>
<tr>
<td>44</td>
<td>Laberinto_BIR</td>
<td>0.1496</td>
<td>0.3400</td>
<td>0.3000</td>
<td>0.1908</td>
<td>0.1992</td>
<td>1292</td>
</tr>
<tr>
<td>48</td>
<td>Laberinto_CTIR</td>
<td>0.1466</td>
<td>0.3433</td>
<td>0.2950</td>
<td>0.1868</td>
<td>0.1953</td>
<td>1293</td>
</tr>
<tr>
<td>50</td>
<td>Laberinto_ETPCIR</td>
<td>0.1411</td>
<td>0.3000</td>
<td>0.2850</td>
<td>0.1766</td>
<td>0.1887</td>
<td>1325</td>
</tr>
<tr>
<td>57</td>
<td>Laberinto_BFT</td>
<td>0.1146</td>
<td>0.2533</td>
<td>0.2267</td>
<td>0.1621</td>
<td>0.1786</td>
<td>1355</td>
</tr>
<tr>
<td>58</td>
<td>Laberinto_CTFT</td>
<td>0.1101</td>
<td>0.2500</td>
<td>0.2333</td>
<td>0.1512</td>
<td>0.1691</td>
<td>1348</td>
</tr>
<tr>
<td>59</td>
<td>Laberinto_ETFT</td>
<td>0.1050</td>
<td>0.2567</td>
<td>0.2250</td>
<td>0.1302</td>
<td>0.1640</td>
<td>1292</td>
</tr>
<tr>
<td>60</td>
<td>Laberinto_ETPCFT</td>
<td>0.1014</td>
<td>0.2400</td>
<td>0.2200</td>
<td>0.1253</td>
<td>0.1571</td>
<td>1310</td>
</tr>
</tbody>
</table>
CONCLUSIONS

• Technique based on MeSH Tree Structure → Obtain good results.

• This work verifies the difficulty of finding an appropriate strategy for query expansion.

FUTURE WORKS

• Further research on other query expansion strategies using other ontologies, such as UMLS.

• To build indexes using only medical concepts extracted from image captions.

• To experiment expanding both the queries and the indexed text.
THANK YOU FOR YOUR ATTENTION

ANY QUESTIONS