



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**



- 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.

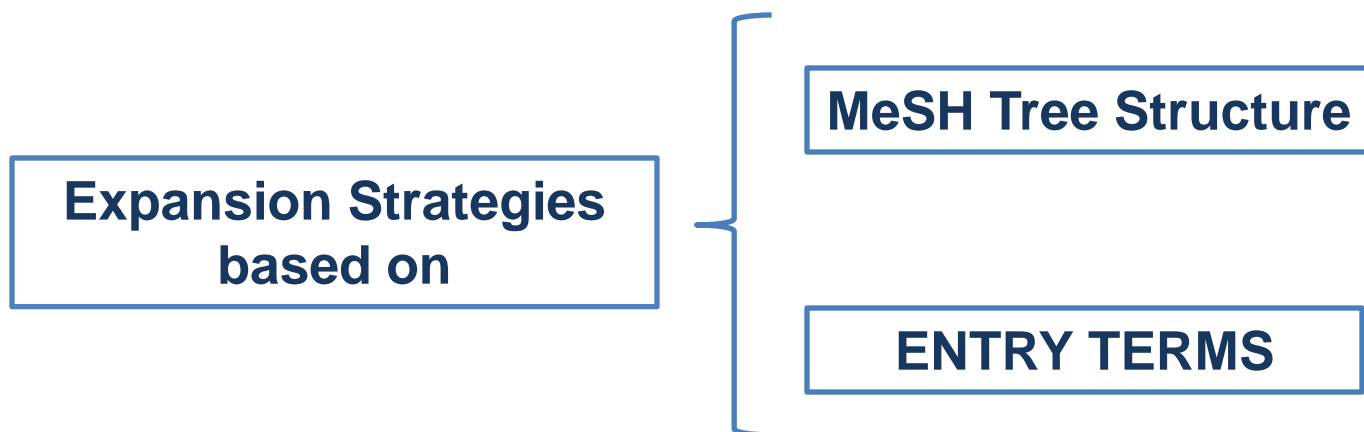


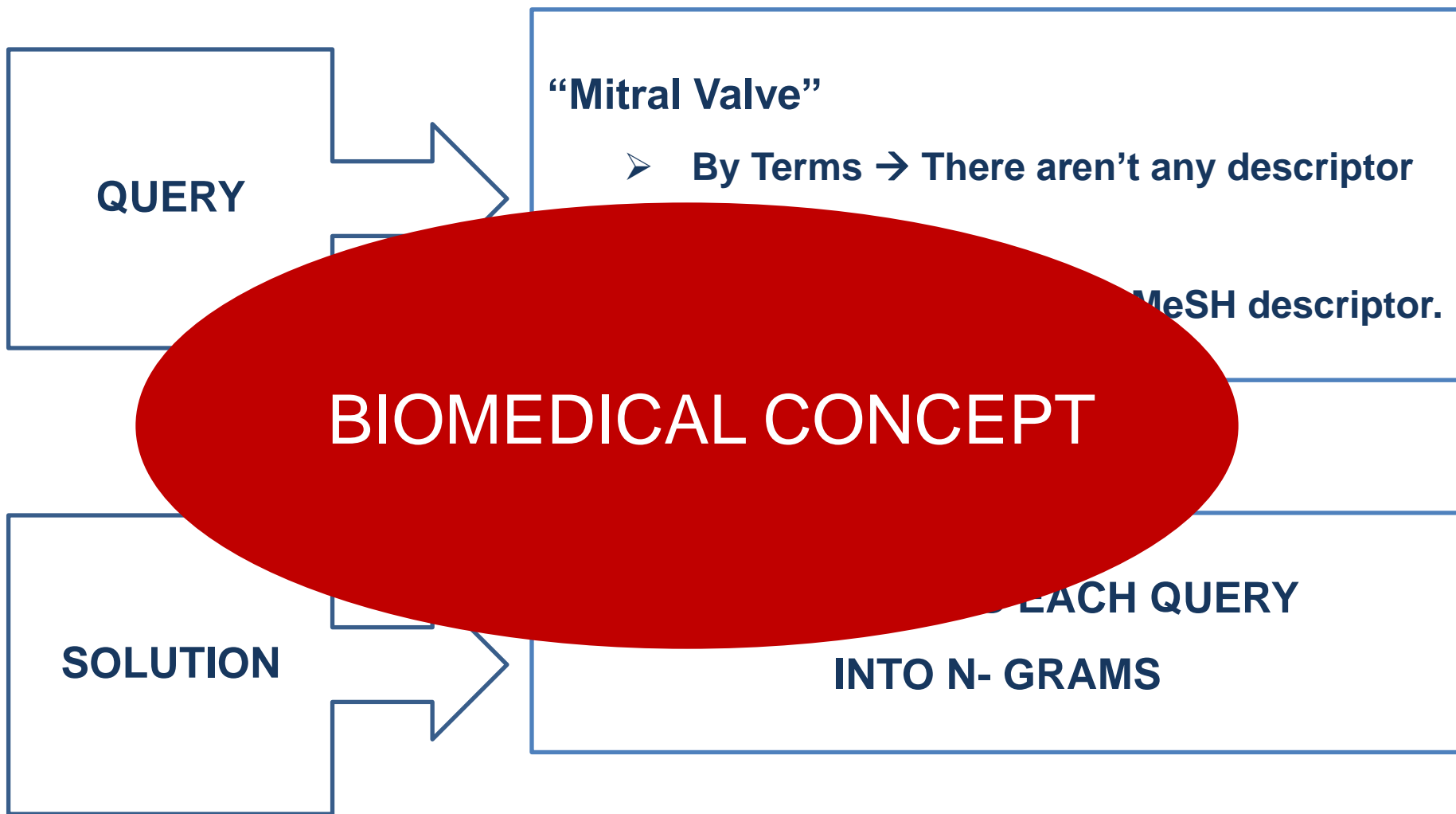
- MeSH (Medical Subject Headings) is a controlled vocabulary, produced and maintained by the U. S. National Library of Medicine.

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
 o [Bacterial Infections and Mycoses \[C01\]](#) +
 o [Virus Diseases \[C02\]](#) +
 o [Parasitic Diseases \[C03\]](#) +
 o [Neoplasms \[C04\]](#) +
 o [Cysts \[C04.182\]](#) +
 o [Hamartoma \[C04.445\]](#) +
 o [Neoplasms by Histologic Type \[C04.557\]](#) +

- 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.





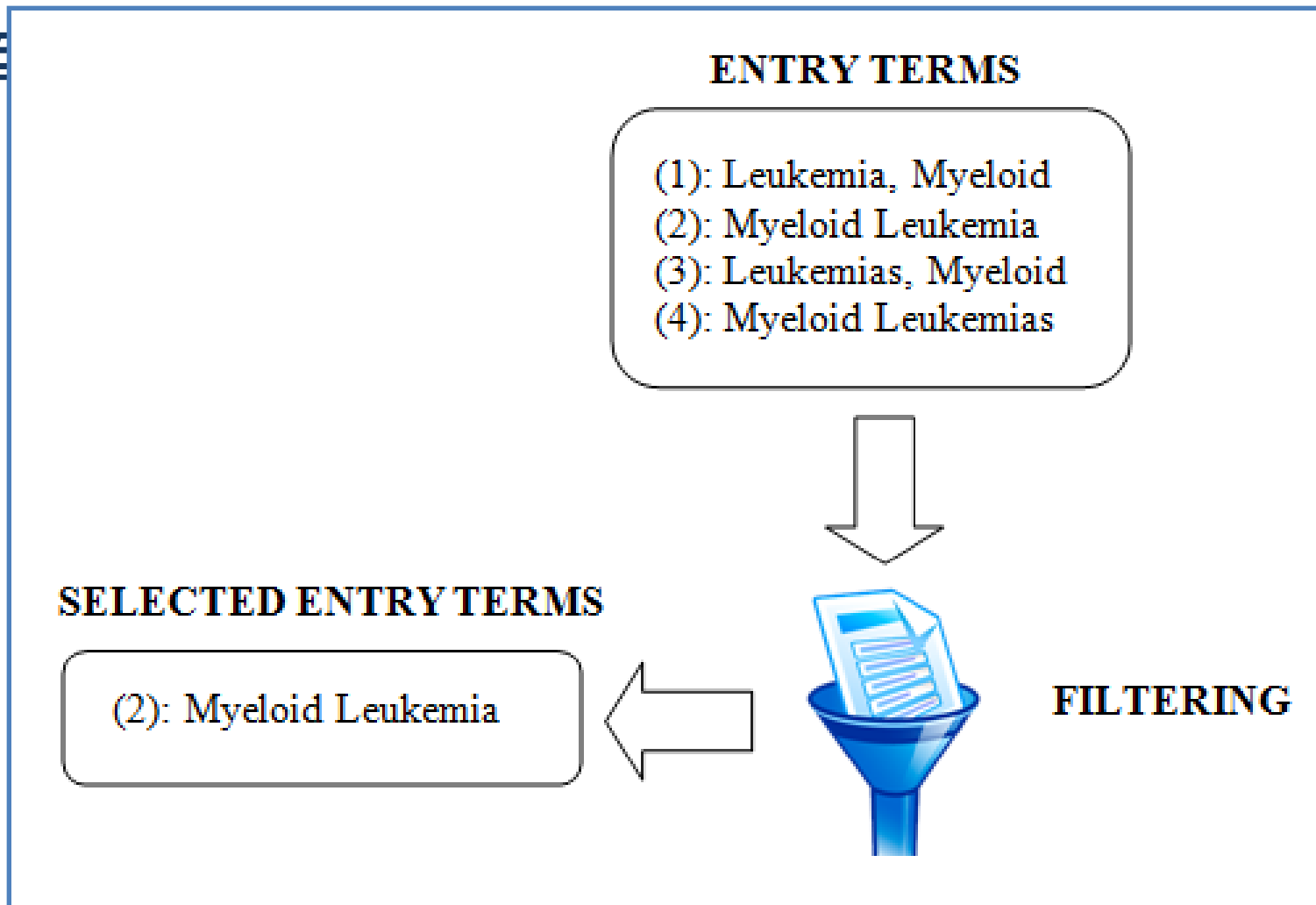
- QUERY: “breast cancer mammogram”
- N-GRAMS
 - 1) breast
 - 2) breast cancer
 - 3) breast cancer mammogram
 - 4) cancer
 - 5) cancer mammogram
 - 6) mammogram

TECHNIQUE BASED ON MeSH TREE STRUCTURE

- [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 \[A08.186.854\] +](#)


Option 2: If descriptor has no children -> No expansion

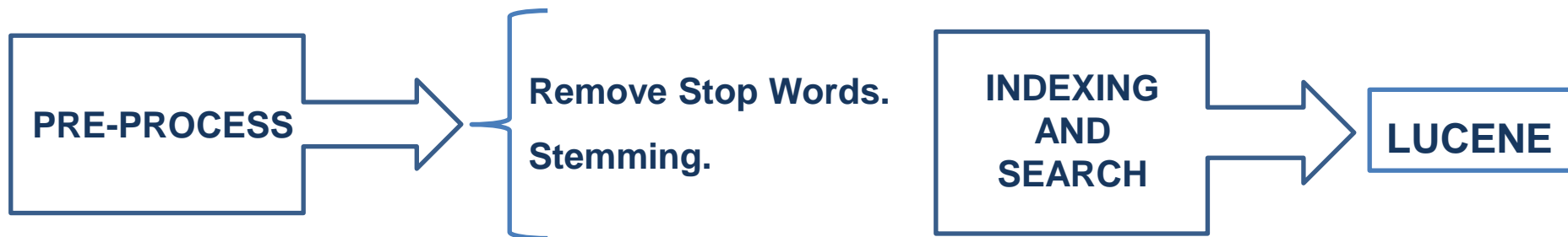
TE



of

the

- 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.**

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

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 ?



Universidad
de Huelva

