

# ImageCLEF: Understanding and Improving Image Retrieval in Biomedicine

William Hersh, MD  
Professor and Chair  
Department of Medical Informatics & Clinical Epidemiology  
Oregon Health & Science University  
hersh@ohsu.edu  
www.billhersh.info

## References

- Hersh, W. (2003). *Information Retrieval: A Health and Biomedical Perspective (Second Edition)*. New York. Springer-Verlag.
- Voorhees, E. and Harman, D., eds. (2005). *TREC: Experiment and Evaluation in Information Retrieval*. Cambridge, MA. MIT Press.
- Hersh, W., Bhupatiraju, R., et al. (2006). Enhancing access to the bibliome: the TREC 2004 Genomics Track. *Journal of Biomedical Discovery and Collaboration*, 1: 3. <http://www.j-biomed-discovery.com/content/1/1/3>.
- Peters, C., Gey, F., et al., eds. (2006). *Accessing Multilingual Information Repositories - 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005*. Berlin, Germany. Springer-Verlag.
- Hersh, W., Müller, H., et al. (2006). Advancing biomedical image retrieval: development and analysis of a test collection. *Journal of the American Medical Informatics Association*, 13: 488-496.
- Müller, H., Michoux, N., et al. (2004). A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics*, 73: 1-23.
- Müller, H., Deselaers, T., et al. (2006). Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks. *Working Notes for the CLEF 2006 Workshop*, Alicante, Spain. [http://www.clef-campaign.org/2006/working\\_notes/workingnotes2006/deselaersOCLEF2006.pdf](http://www.clef-campaign.org/2006/working_notes/workingnotes2006/deselaersOCLEF2006.pdf).
- Lacoste, C., Chevallet, J., et al. (2006). IPAL Knowledge-based Medical Image Retrieval in ImageCLEFmed 2006. *Working Notes for the CLEF 2006 Workshop*, Alicante, Spain. [http://www.clef-campaign.org/2006/working\\_notes/workingnotes2006/lacosteCLEF2006.pdf](http://www.clef-campaign.org/2006/working_notes/workingnotes2006/lacosteCLEF2006.pdf).
- Hersh, W., Kalpathy-Cramer, J., et al. (2006). Medical Image Retrieval and Automated Annotation: OHSU at ImageCLEF 2006. *Working Notes for the CLEF 2006 Workshop*, Alicante, Spain. [http://www.clef-campaign.org/2006/working\\_notes/workingnotes2006/hershCLEF2006.pdf](http://www.clef-campaign.org/2006/working_notes/workingnotes2006/hershCLEF2006.pdf).

## ImageCLEF: Understanding and Improving Image Retrieval in Biomedicine

William Hersh, MD  
Professor and Chair  
Department of Medical Informatics & Clinical Epidemiology  
Oregon Health & Science University  
hersh@ohsu.edu  
www.billhersh.info



1

## Acknowledgements

- Funding
  - NSF Grant ITR-0325160
- Collaborators
  - Jeffery Jensen, Jayashree Kalpathy-Cramer, OHSU
  - Henning Müller, University of Geneva, Switzerland
  - Paul Clough, University of Sheffield, England
  - Cross-Language Evaluation Forum (Carol Peters, ISTI-CNR, Pisa, Italy)



2

## Overview of talk

- Introduction to image retrieval
- Primer on generic information retrieval
- Indexing and retrieval of images
- ImageCLEF medical image retrieval project
  - Test collection description
  - Results and analysis of experiments
- Future directions



3

## Image retrieval

- Biomedical professionals increasingly use images for research, clinical care, and education, yet we know very little about how they search for them
- Most image retrieval work has focused on either text annotation retrieval or image processing, but not combining both
- Goal of this work is to increase our understanding and ability to retrieve images



4

## Image retrieval issues and challenges

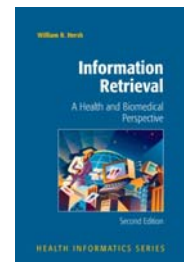
- Image retrieval is a “poor stepchild” to text retrieval, with less understanding of how people use systems and how well they work
- Images are not always “standalone,” e.g.,
  - May be part of a series of images
  - May be annotated with text
- Images are “large”
  - Relative to text
- Images may be compressed, which may result in loss of content (e.g., lossy compression)



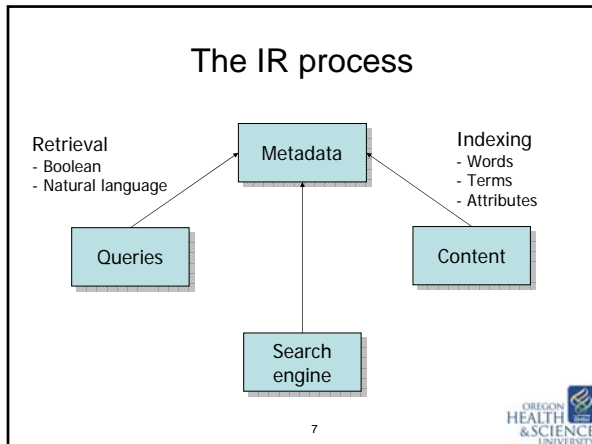
5

## Information retrieval (IR) (Hersh, 2003)

- [www.irbook.info](http://www.irbook.info)
- Systems that provide indexing and retrieval of usually knowledge-based information
- Historically centered on text in documents, but increasingly associated with many types of content



6



- ### Evaluation of IR systems
- System-oriented – how well system performs
    - Historically focused on relevance-based measures
      - Recall – # relevant retrieved / # relevant in collection
      - Precision – # relevant retrieved / # retrieved by search
    - When content output is ranked, can aggregate both in measure like mean average precision (MAP)
  - User-oriented – how well user performs with system
    - e.g., performing task, user satisfaction, etc.
- 8

- ### System-oriented IR evaluation
- Historically assessed with test collections, which consist of
    - Content – fixed yet realistic collections of documents, images, etc.
    - Topics – statements of information need that can be fashioned into queries entered into retrieval systems
    - Relevance judgments –by expert humans for which content items should be retrieved for which topics
  - Calculate summary statistics for all topics
    - Primary measure usually MAP
- 9

### Calculating MAP in a test collection

Average precision (AP) for a topic:

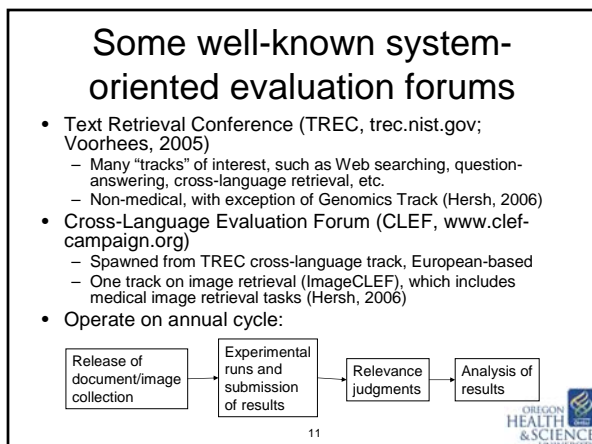
1: REL	1/1 = 1.0
2: NOT REL	
3: REL	2/3 = 0.67
4: NOT REL	
5: NOT REL	
6: REL	3/6 = 0.5
7: NOT REL	
N: REL	0
N+1: REL	0

Mean average precision (MAP) is mean of average precision for all topics in a test collection

Result is an aggregate measure but the number itself is only of comparative value

$$(1.0 + 0.67 + 0.5) / 5 = 0.43$$

10



- ### Image retrieval – indexing
- Two general approaches (Müller, 2004)
    - Textual or semantic – by annotation, e.g.,
      - Narrative description
      - Controlled terminology assignment
      - Other types of textual metadata, e.g., modality, location
    - Visual or content-based
      - Identification of features, e.g., colors, texture, shape, segmentation
      - Our ability to “understand” content of images less developed than for textual content
- 12

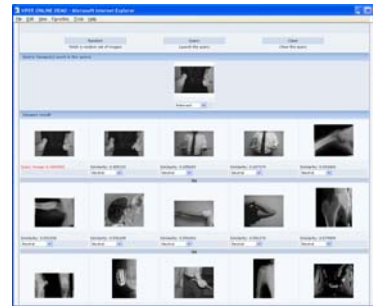
## Image retrieval – searching

- Based on type of indexing
  - Textual – typically uses features of text retrieval systems, e.g.,
    - Boolean queries
    - Natural language queries
    - Forms for metadata
  - Visual – usual goal is to identify images with comparable features, i.e., find me images similar to this one

13



## Example of visual image retrieval



14



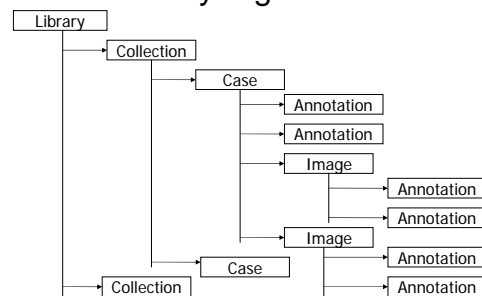
## ImageCLEF medical image retrieval

- Aims to simulate general searching over wide variety of medical images
- Uses standard IR approach with test collection consisting of
  - Content
  - Topics
  - Relevance judgments
- Has operated through two cycles of CLEF (2005-2006)
  - Both used same image collection
  - Developed new topics and performed relevance judgments for each cycle

15



## ImageCLEF medical collection library organization



16



## ImageCLEF medical test collection

Collection	Predominant images	Cases	Images	Annotations	Size
Casimage	Mixed	2076	8725	English – 177 French – 1899	1.3 GB
Mallinckrodt Institute of Radiology (MIR)	Nuclear medicine	407	1177	English – 407	63 MB
Pathology Education Instructional Resource (PEIR)	Pathology	32319	32319	English – 32319	2.5 GB
PathoPIC	Pathology	7805	7805	German – 7805 English – 7805	879 MB

17



## Example case from Casimage

Images



Case annotation

**ID:** 4272  
**Description:** A large hypochoic mass is seen in the spleen. CDFI reveals it to be hypovascular and distorts the intrasplenic blood vessels. This lesion is consistent with a metastatic lesion. Urinary obstruction is present on the right with pelvocaliceal and ureteral dilatation secondary to a soft tissue lesion at the junction of the ureter and bladder. This is another secondary lesion of the malignant melanoma. Surprisingly, these lesions are not hypervascular on doppler nor on CT. Metastasis are also visible in the liver.  
**Diagnosis:** Metastasis of spleen and ureter, malignant melanoma  
**Clinical Presentation:** Workup in a patient with malignant melanoma. Intravenous pyelography showed no excretion of contrast on the right.

18



## Annotations vary widely

- Casimage – case and radiology reports
- MIR – image reports
- PEIR – metadata based on Health Information Assets Library (HEAL)
- PathoPIC – image descriptions, longer in German and shorter in English

19



## Topics

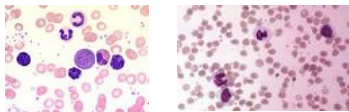
- Each topic has
  - Text in 3 languages
  - Sample image(s)
  - Category – judged amenable to visual, mixed, or textual retrieval methods
- 2005 – 25 topics
  - 11 visual, 11 mixed, 3 textual
- 2006 – 30 topics
  - 10 each of visual, mixed, and textual

20



## Example topic (2005, #20)

Show me microscopic pathologies of cases with chronic myelogenous leukemia.  
Zeige mir mikroskopische Pathologiebilder von chronischer Leukämie.  
Montre-moi des images de la leucémie chronique myélogène.



21



## Relevance judgments

- Done in usual IR manner with pooling of results from many searches on same topic
- Pool generation – top N results from each run
  - Where N = 40 (2005) or 30 (2006)
  - About 900 images per topic judged
- Judgment process
  - Judged by physicians in OHSU biomedical informatics program
  - Required about 3-4 hours per judge per topic
- Kappa measure of interjudge agreement = 0.6-0.7 (“good”)

22



## Categories of runs

- Query type – human preparation
  - Automatic – no human modification
  - Manual – human modification of query
  - Interactive – human modification of query after viewing output (not designated in 2005)
- System type – feature(s)
  - Textual – searching only via textual annotations
  - Visual – searching only by visual means
  - Mixed – textual and visual searching
  - (NOTE: Topic types have these category names too)

23



## ImageCLEF medical retrieval task results

- Primary measure – MAP
- 2005 results reported in Hersh (JAMIA, 2006)
- 2006 results reported in track overview on CLEF Web site (Müller, 2006) and in following slides
  - Runs submitted
  - Best results overall
  - Best results by query type
  - Comparison by topic type
  - Comparison by query type
  - Comparison of measures
  - Interesting finding from OHSU runs

24



## Runs submitted by category

System Type	Visual	Mixed	Textual	Total
Automatic	11	37	31	79
Manual	10	1	6	17
Interactive	1	2	1	4
Total	22	40	38	100

25



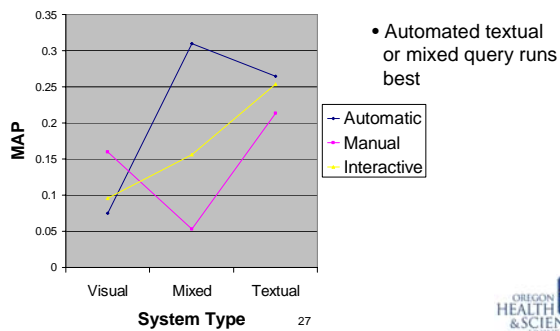
## Best results overall

- Institute for Infocomm Research (Singapore) and IPAL-CNRS (France) (Lacoste, 2006)
- Used combination of image and text processing
  - Latter focused on mapping terms to semantic categories, e.g., modality, anatomy, pathology, etc.
- MAP – 0.3095
- Precision at
  - 10 images – 0.6167 (6.2 images)
  - 30 images – 0.5822 (17.4 images)
  - 100 images – 0.3977 (40 images)

26



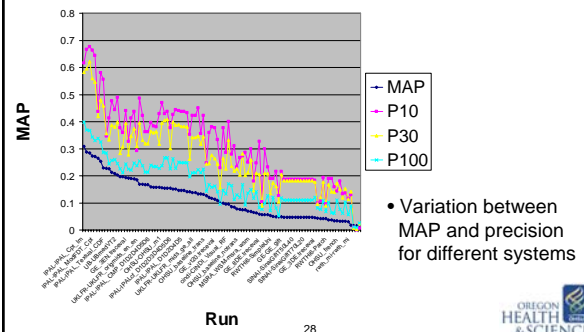
## Best performing runs by system and query type



27



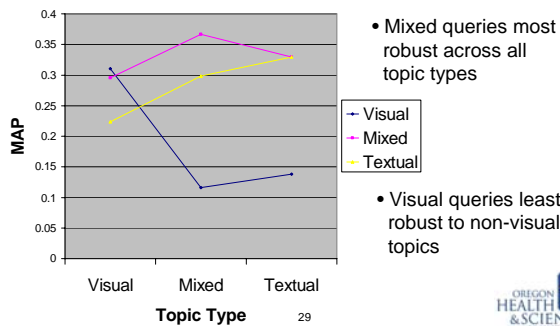
## Results for all runs



28



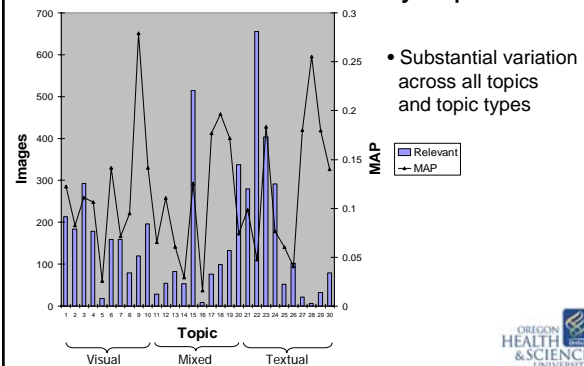
## Best performing runs by topic type for each system type



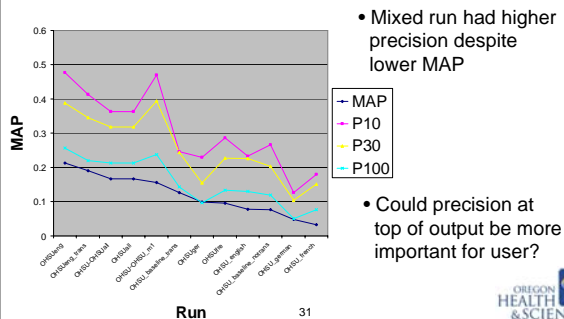
29



## Relevant and MAP by topic



## Interesting finding from OHSU runs



## Conclusions

- A variety of approaches are effective in image retrieval, similar to IR with other content
- Systems that use only visual retrieval are less robust than those that solely do textual retrieval
- MAP might not be the best performance measure for the image retrieval task
- Need broader understanding of system use followed by better test collections and experiments based on that understanding

## Limitations

- This test collection
  - Topics artificial – may not be realistic or representative
  - Annotation of images may not be representative or of best practice
- Test collections generally
  - Relevance is situational
  - No users involved in experiments

## Future directions

- ImageCLEF 2007
  - Continue work on annual cycle
  - Funded for another year from NSF grant
  - Aim to expand image collection, add new topics
- User experiments with OHSU image retrieval system
  - Aim to better understand real-world tasks and best evaluation measures for those tasks
- Continued analysis of 2005-2006 data
  - Improved text retrieval of annotations
  - Improved merging of image and text retrieval

## For more information

- Web site: <http://ir.ohsu.edu/image/>
- Another way you can contribute: Donate your images to IR science!
- Questions?