

Master EIT ICT Labs in ICT Innovation

Universidad Politécnica de Madrid



- In-class assignment (groups of 2 students)
- Send presentation slides (up to 10 minutes) by 13th April at 17:00
- Implementation of a basic CBIR using Lucene
- This assignment's mark is the 25% of the course mark

Master EIT ICT Labs in ICT Innovation

Universidad Politécnica de Madrid



- Assignment mandatory:

a)

- OCR – obtain text from images
 - tesseract
 - Try to optimize extraction using layouts (structure information like author, abstract, etc)
- Dataset : <https://marg.nlm.nih.gov/files.html>

b)

- Lucene – index image descriptions obtained

Master EIT ICT Labs in ICT Innovation

Universidad Politécnica de Madrid



- Assignment
Optional

a)

- DICOM – obtain text from image headers
 - Dcm2che
 - imageJ
 - Try to structure image descriptiond from study
- Dataset : <http://www.osirix-viewer.com/datasets/>

Master EIT ICT Labs in ICT Innovation

Universidad Politécnica de Madrid



- Assignment

Optional

b)

- Metadata-Extractor – <https://drewnoakes.com/code/exif/>
- structure information for Lucene
- Dataset : use camera photos / flickr dataset

c)

- Basic color and shape descriptors – i.e. histogram ...
- structure information for Lucene
- Dataset : use camera photos / flickr dataset

Master EIT ICT Labs in ICT Innovation

Universidad Politécnica de Madrid



- Assignment

Optional

d)

- Integrate indexing of DICOM images and OCR using SOLR