Information Retrieval

UA.DETI.RI - 2024/25

José Luís Oliveira

The course

- Scientific area
  - Informatics
- Weekly schooling:
  - 3 hours of TP classes
- ECTS credits: 6
- Code: 42596
- Motivation

Objectives

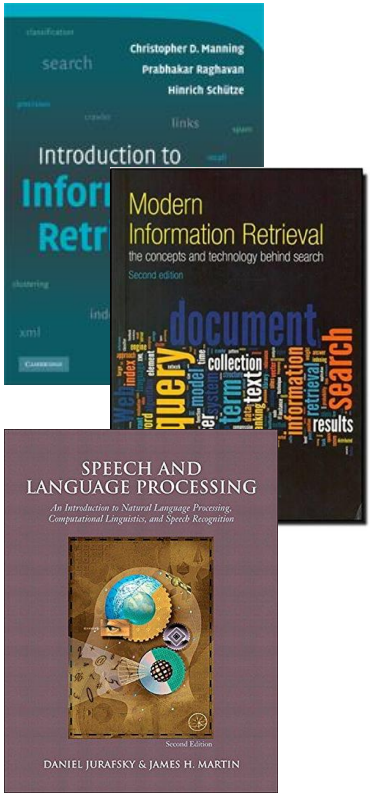
- Develop skills to:

- Understand fundamental concepts of processing, storing, indexing, and searching unstructured information.
- Identify and know how to use the best data structures and algorithms used in text-based information systems.
- Develop solutions based on rules, dictionaries and/or machine learning.
- Evaluate the performance of solutions.

Program

- General concepts and the boolean retrieval model
- Text processing and indexing
  - Documents, terms, stop-words, normalization
  - Stemming, lemmatization, part-of-speech
- Indexes, data structures and compression
- Vector space model
- Information retrieval evaluation
- Relevance feedback and term expansion
- Probabilistic and language-based models
- Distributed text representation
- Neural information retrieval
- Document classification
  - Naive Bayes, kNN, SVM
  - Neural networks and deep learning

Main bibliography



- Introduction to Information Retrieval, C.D. Manning, P. Raghaven, H. Schütze, 2008, Cambridge University Press.
  - https://nlp.stanford.edu/IR-book/information-retrieval-book.html
- Speech and Language Processing (3rd Edition): Daniel Jurafsky, James Martin, 2024,
  - https://web.stanford.edu/~jurafsky/slp3/
- Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition), R. Baeza-Yates, B. Ribeiro-Neto B., 2011, Addison Wesley Professional.
  - https://users.dcc.uchile.cl/~rbaeza/mir2ed/

Evaluation

- **Practical evaluation (AP) – 70%**
  - Assignment1 (40%)
    - 17-Oct Assignment 1 – Interim Presentation
    - 14-Nov Assignment 1 – Final Presentation
  - Assignment2 (30%)
    - 05-Dec Assignment 2 – Interim Presentation
    - 19-Dec Assignment 2 – Final Presentation
  - Submissions on the same dates on GitHub classroom
- **Monography (ATP) – 30%**
  - 28-Nov Topic selection
  - 8-Jan (26-Jan) PDF submission (submission on elearning.ua.pt)
  - 10-Jan (28-Jan) Presentation
- Minimum score for each component: 7 points.

Class notes

- After each TP session, students will submit a paragraph summarizing what they learned from the class along with questions.
  - Extra 5% of the final grade
- This card contains:
  - Student ID (NMEC, Name)
  - A paragraph summarizing what you learned
  - What questions are still on your mind

ECTS

- Education (T/TP/P): 0/2/2 - ECTS: 6

- The number of ECTS credits indicates the expected number of hours they must study for this subject.
  - 1 ECTS = 25-30 hours of study.
  - 6 ECTS = 150-180 hours of study.
- In a 15-week semester, they must study at least 10 hours per week.
- These hours include face-to-face classes, reading books, solving exercises, studying for tests and exams, etc.

Resources

- elearning.ua.pt
  - Slides
  - Assignments (1 and 2)
  - Submission instructions
  - Information and results
- GitHub classroom

- Public resources
  - Books, blogs
  - Hugging Face
  - Kaggle
  - ... And many others

Teachers

- José Luis Oliveira (jlo@ua.pt)
  - IEETA
- Tiago Almeida(tiagomeloalmeida@ua.pt)
  - IEETA

Bons estudos e bom semestre!