IT550 - Introduction to Information Retrieval

Parth Mehta

NLP Research Scientist, Parmonic

What this course will be…

- Up to date with the latest breakthroughs like Neural IR, LLMs and RAG
- Less theory and more hands on. Some lectures will be converted to labs.
- Programming heavy

What this course will not be…

- Easy
- Something that can be done in the week before exams. If you just want to skim through, this course is not for you
- Programming class. If you are not at least decent in coding, this course is not for you

Expected outcome

- Good grasp of traditional as well as neural IR
- Fluency in some of the most popular IR / NLP / ML concepts
- Exposure to at least one open source search engine (Terrier / Elastic / Vector DBs)
- “Some” exposure to what is directly useful in the industry

Class and Lab Format

- Topics to be discussed in class during a week will be shared in the preceding week (starting next week)
- Relevant datasets and lab manual of a lab session will be shared in the preceding week
- Participants are expected to come prepared with some background reading / code for classes and labs

Key Dates

- First Insem (29 August - 2 September)
- Project Topic Finalization - 5 min lightning talk (3-5 September)
- End Sem Exam (23 November - 02 December)
- Final Project Presentation - (04 - 06 December)*

* Tentative

Grading Scheme

- In-sem - 20%
- End-sem - 20%
- Lab - 20%
- Project - 40%
- Extraordinary effort bonus - 10%

Course outline

Classical IR (15 Lectures + 4 Labs) ~Insem 1

1. Basics of IR
2. Term weighting and Ranking
3. Evaluation
4. Query Operations
5. Probabilistic IR
6. Web Search

Course outline

Neural IR (11 Lectures + 4 Labs) ~Endsem

1. Distributed Word Representations (AKA embeddings)
2. Neural Information Retrieval
3. Search in Post LLM Era
4. Domain Specific Applications

Course Outline

Lab Only Topics (8 Lectures converted to 4 labs)

- Text Classification
- Text Clustering
- Distributed IR
- Elasticsearch and Kibana

Call for Voluntary TAs

- Help with planning and running the lab sessions, creating baseline systems
- Why?
  - Chance to learn something extra. More in-depth discussions with me then possible in lab/lectures
  - *No extra credits for volunteering*
  - *Possible extra credits IF the extra learning leads to better projects*

Volunteers will not be performing any type of grading

Useful Resources

Books

- Introduction to Information Retrieval - Christopher Manning, Prabhakar Raghavan and Hinrich Schütze
- An Introduction to Neural Information Retrieval - Bhaskar Mitra and Nick Crasswell

Mailing lists

- SIGIR, ML-NEWS, FIRE, CORPORA

Evaluation Forums

- TREC, CLEF, FIRE, NTCIR

Questions?

parth.mehta126@gmail.com