Course Dynamics

Advanced Information Retrieval
Term: Jan-Feb 2020

Welcome to the Advanced Information Retrieval course. Information Retrieval has matured on several aspects in the last decade. Prominently, we have moved from keyword-based search to semantic search. Several concepts such as knowledgebases and entities have been widely accepted in practice. Ontologies, RDF Triplets and SPARQL have found its way into the industry. In this course, we survey these ideas as we delve into some recent trends of information retrieval.

This is a two-credit course. You **must have** completed Information Retrieval course with a passing grade to register for this course.

Key Learning Objectives

The key objectives of this course are as follows:

- To transition from keyword-based search to semantic search
- To take a deeper dive into query understanding, retrieval models and ranking
- To understand the recent trends in IR practice and research

Schedule

1. Review of Information Retrieval
    a. Science of Information [IT]
    b. Search Techniques [Chapter 4 of BBH]
    c. Advanced Search Techniques [Chapter 5 of BBH]
2. Query Understanding
    a. Query understanding challenges - Misspellings, Query Segmentation, Need for Query Expansion, Need for Query Reformulation [CPS]
    b. Query annotation with semantic information [SQA]
    c. Generic Intent Representation [AH]
    d. Semantic Query Understanding [SQU, TQI]
3. Retrieval Models
    a. Boolean, VSM, vs. Probabilistic [CPS]
    b. Probabilistic Relevance Feedback [CPS]
    c. Language Model [CPS]
4. Semantic Search
    a. Limitations of Keyword Search [SWP]
    b. Semantic Search [SWP]
    c. Embeddings - Word2Vec [MIK]
    d. Ontology [OWL]
    e. RDF Triplets, SPARQL [RDF, SPQL]

f. Constructing knowledgebases [KB]
g. Entities [RT]

Evaluation

- Assignments: 3 * 15% = 45 %
  - Lead a discussion in the class on the topic for 15 (to 30) minutes. Submit a one-page summary of the topic. Format for assignments is available at course git page [CGP]. Import it into overleaf.
- Quiz: 20% (2% * 10)
- Exam: 35%
- Bonus Project (Adjustable only against the exam score to a max of 25%. Deliverables - Upload your code to github. Write a one-page summary report describing your project.)

References

- [IT] Chapters 1, 2 (Sections 2.1 - 2.5) of A First Course in Information Theory
- [AH] Generalized Syntactic and Semantic Models of Query Reformulation
- [SQU] Semantic Query Understanding
- [SQA] Intent-Aware Semantic Query Annotation
- [TQI] Understanding Temporal Query Intent
- [CPS] An Introduction to Information Retrieval. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze.
- [MIK] Efficient Estimation of Word Representations in Vector Space
- [OWL] Sections 2, 3, 4 and 5 of OWL Primer at w3.rg. https://www.w3.org/2012/pdf/REC-owl2-primer-20121211.pdf
- [RDF] Chapter 1 of RDF Primer https://www.w3.org/TR/rdf11-concepts/
- [SPQL] Chapters 1 and 2 of SPARQL Primer https://www.w3.org/TR/rdf-sparql-query/.
- [KB] http://pages.cs.wisc.edu/~anhai/papers/kcs-sigmod13.pdf
- [RT] Chapter 4 of Rhode's thesis https://research.utwente.nl/files/6040940/thesis_H_Rode.pdf
- [SWP] A Semantic Web Primer, Grigoris Antoniou and Frank van Harmelen
- [SWR] The Semantic Web Revisited https://eprints.soton.ac.uk/262614/1/Semantic_Web_Revisted.pdf
- [CGP] https://github.com/vvtesh/vvtesh.github.io/blob/v1/teaching/advir2021/VLDBTemplate.zip