

CT4100 Information Retrieval

Lecturer contact details

- Colm O'Riordan
- colm.oriordan@universityofgalway.ie
- Discussion Board on canvas

CT4100 Motivations

- To study/analyse techniques to deal suitably with the large amounts (and types) of information.
- Emphasis on research and practice in Information Retrieval

Related Fields

- Artificial Intelligence
- Database and Information Systems
- Algorithms
- Human Computing Interaction

Recommended Texts

- Modern Information Retrieval Riberio-Neto & Baeza-Yates (several copies in library)
- Information Retrieval Grossman
- Introduction to Information Retrieval Christopher Manning et al (available online)
- Extra resources: Will also place research papers online as recommended extra reading

Introduction

Information retrieval deals with identifying relevant information based on users' information needs. e.g. web search engines, digital libraries, and recommender systems. Raises many important and challenging questions.

Information retrieval (IR) is finding material (usually documents) of an unstructured nature . . . that satisfies an information need from within large collections (usually stored on computers) (Manning, 2008)

Topics

Topic 1: Introductory Techniques and Models

- Boolean model
- Vector space model

Topic 2: System Architecture

- Pre-processing
- Indexing techniques
- Compression

Topic 3: Evaluation

- Precision/Recall
- Mean Average Precision
- Other measures - harmonic mean, E-measure
- User centered Measures

Topic 4: Relevance Feedback

- Local analysis; global analysis
- Vector space feedback
- Rocchio method etc.
- Clustering - association clusters; metric clusters etc.
- Open issues

Topic 5: Term weighting schemes

- Weighting Schemes
- Early tf-idf approaches
- BM25
- Pivoted normalisation
- Axiomatic approaches

Topic 6: Collaborative Filtering

- Recommender systems
- Neighbourhood based approaches
- Correlation algorithms- Pearson's, constrained Pearson's etc.
- Evaluation

Topic 7: Web Search Engines

- Structure of the web
- Link analysis techniques
- HITS
- PageRank
- Extensions

Topic 8: Clustering in IR

- Uses of clustering in IR
- K means algorithm – discussion
- Issues re: clustering

Topic 9: Structured Document Retrieval

- Document representation
- Problems in structured retrieval
- IR Approaches to dealing with structure
- Extensions to vector space

Topic 10: Learning in IR

- Evolutionary Computation approaches
- Neural Networks
- Learning to Rank

Topic 11: Query Difficulty

- Estimating query difficulty
- Pre-retrieval
- Post-retrieval

(Time permitting): Trends/other topics

- Sentiment analysis
- Expert finder
- Music retrieval
- Graph representations
- Image retrieval
- Conversational IR

Grading

Grading

- Exam: 70%
- Assignment 1: 15%
- Assignment 2: 15%

Exercise Sheets

- There will be exercise sheets posted for most lectures; these are not mandatory and are intended as a study aid.