

CT4100

Information Retrieval

# Lecturer contact details

Colm O' Riordan

[colm.oriordan@universityofgalway.ie](mailto:colm.oriordan@universityofgalway.ie)

Room 403

Discussion Board

# 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 the identification of relevant information based on users' information needs. e.g. web search engines, digital libraries, recommender systems

Raises many important and challenging questions.

(Manning, 2008)

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)

# Proposed Course Topics

Introduction to Information Retrieval and Filtering (basic architecture; components).	Collaborative Filtering/Recommender Systems
Introduction to Models	Web Search
Evaluation Approaches	Clustering
Relevance Feedback	Structured Retrieval
Weighting Schemes	Learning approaches
Further Techniques and Models	Query Difficulty
Indexing and String processing	Recent Trends and Applications

# 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
- Other measures - harmonic mean, E-measure
- User centered Measures
- Related issues

# Topic 4: Relevance Feedback

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

# Topic 5: Techniques and Models II

- Extended Boolean model
- Neural network approaches
- Probabilistic models
- Language Modelling
- Latent Semantic Indexing
- Fuzzy Set models

# Topic 6: Term weighting schemes

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

# Topic 7: Collaborative Filtering

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

# Topic 8: Web Search Engines

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

# Topic 10: Clustering in IR

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



# Topic 11: Structured Document Retrieval

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

# Topic 12: Learning in IR

- Learning in IR: introduction
- Evolutionary Computation approaches
- Neural Networks
- Learning to Rank

# Topic 13: Query Difficulty

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

# Topic 14: Trends/other topics

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