LONDON'S GLOBAL UNIVERSITY

UCL

Computer Science

Information Retrieval and Data Mining (COMP0084)

Description

Aims:

The module is aimed at an entry level study of information retrieval and data mining techniques.

It is about how to find relevant information and subsequently extract meaningful patterns out of it.

While the basic theories and mathematical models of information retrieval and data mining are covered, the course is primarily focused on practical algorithms of textual document indexing, relevance ranking, web usage mining, text analytics, as well as their performance evaluations.

Learning outcomes:

On successful completion of the module, a student will master both the theoretical and practical aspects of information retrieval and data mining, and will be able to understand:

1.the common algorithms and techniques for information retrieval (document indexing and retrieval, query processing, etc);

2.the quantitative evaluation methods for the IR systems and data mining techniques;

3.the popular probabilistic retrieval methods and ranking principles;

4.the techniques and algorithms existing in practical retrieval and data mining systems such as those in web search engines and recommender systems, including the recently popular topic of deep learning;

5.basic algorithms that can be used to make predictions out of data;

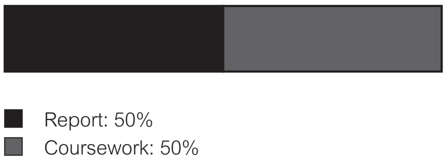
Content:

Overview of the fields:

Key information

Year	2018/19
Credit value	15 (150 study hours)
Delivery	PGT L7, Campus-based
Reading List	View on UCL website
Tutor	Prof Emine Yilmaz
Term	Term 2
Timetable	View on UCL website

Assessment



Find out more

For more information about the department, programmes, relevant open days and to browse other modules, visit ucl.ac.uk

Disclaimer: All information correct as of December 2018. Please note that aspects of the module may be subject to change. UCL will make best efforts to inform applicants of major changes.

Study some basic concepts of information retrieval and data mining, such as the concept of relevance, association rules, and knowledge discovery.

Understand the conceptual models of an information retrieval and knowledge discovery system;

Indexing and Text Processing:

Introduce various indexing techniques for textual information items, such as inverted indices, tokenization, stemming and stop words.

Techniques used for text compression, such as the Lempel-ziv algorithm and Huffman Coding will be covered;

Retrieval Methods:

Study popular retrieval models:

1 Boolean, 2.

Vector space, 3.

Binary independence, 4.

Language modelling.

Probability ranking principle.

Other commonly-used techniques such as relevance feedback, pseudo relevance feedback, and query expansion will also be covered;

Measurements:

Online and offline Evaluation techniques to evaluate retrieval quality.

Commonly used evaluation metrics such as average precision, NDCG, etc.

"Cranfield Paradigm" and TREC conferences, as well as some recently popular techniques such as interleaving will be discussed;

Data Mining:

Study basic techniques, algorithms, and systems of data mining and analytics, including frequent pattern and correlation and association analysis, basic machine learning algorithms such as linear regression and logistic regression.

Discussion on basic personalisation and usage mining techniques;

Emerging Areas:

Study new emerging areas such as learning to rank, deep learning, word embeddings and topic modelling;

Prerequisites:

In order to be eligible to select this module, students must have:

an understanding of probability and statistics;

and proficiency in java programming (as demonstrated by a least one programming project in the past);

LONDON'S GLOBAL UNIVERSITY

UCL

Computer Science

Information Retrieval and Data Mining (COMP0084)

Description

Aims:

The module is aimed at an entry level study of information retrieval and data mining techniques.

It is about how to find relevant information and subsequently extract meaningful patterns out of it.

While the basic theories and mathematical models of information retrieval and data mining are covered, the course is primarily focused on practical algorithms of textual document indexing, relevance ranking, web usage mining, text analytics, as well as their performance evaluations.

Learning outcomes:

On successful completion of the module, a student will master both the theoretical and practical aspects of information retrieval and data mining, and will be able to understand:

1.the common algorithms and techniques for information retrieval (document indexing and retrieval, query processing, etc);

2.the quantitative evaluation methods for the IR systems and data mining techniques;

3.the popular probabilistic retrieval methods and ranking principles;

4.the techniques and algorithms existing in practical retrieval and data mining systems such as those in web search engines and recommender systems, including the recently popular topic of deep learning;

5.basic algorithms that can be used to make predictions out of data;

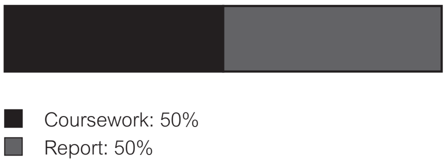
Content:

Overview of the fields:

Key information

Year	2018/19
Credit value	15 (150 study hours)
Delivery	UGM L7, Campus-based
Reading List	View on UCL website
Tutor	Prof Emine Yilmaz
Term	Term 2
Timetable	View on UCL website

Assessment



Find out more

For more information about the department, programmes, relevant open days and to browse other modules, visit ucl.ac.uk

Disclaimer: All information correct as of December 2018. Please note that aspects of the module may be subject to change. UCL will make best efforts to inform applicants of major changes.

Study some basic concepts of information retrieval and data mining, such as the concept of relevance, association rules, and knowledge discovery.

Understand the conceptual models of an information retrieval and knowledge discovery system;

Indexing and Text Processing:

Introduce various indexing techniques for textual information items, such as inverted indices, tokenization, stemming and stop words.

Techniques used for text compression, such as the Lempel-ziv algorithm and Huffman Coding will be covered;

Retrieval Methods:

Study popular retrieval models:

1 Boolean, 2.

Vector space, 3.

Binary independence, 4.

Language modelling.

Probability ranking principle.

Other commonly-used techniques such as relevance feedback, pseudo relevance feedback, and query expansion will also be covered;

Measurements:

Online and offline Evaluation techniques to evaluate retrieval quality.

Commonly used evaluation metrics such as average precision, NDCG, etc.

"Cranfield Paradigm" and TREC conferences, as well as some recently popular techniques such as interleaving will be discussed;

Data Mining:

Study basic techniques, algorithms, and systems of data mining and analytics, including frequent pattern and correlation and association analysis, basic machine learning algorithms such as linear regression and logistic regression.

Discussion on basic personalisation and usage mining techniques;

Emerging Areas:

Study new emerging areas such as learning to rank, deep learning, word embeddings and topic modelling;

Prerequisites:

In order to be eligible to select this module, students must have:

an understanding of probability and statistics;

and proficiency in java programming (as demonstrated by a least one programming project in the past);