

## **Items approved by convenor action:**

### **1. Changes to Machine Learning Practical:**

Please update as follows.

- Availability: Not available to Visiting Students.
- Summary: Please ADD the following at the end of this section.

"Note: this course is not a stand-alone introduction to machine learning. Please see Other Requirements for details."

- Other requirements: Please ADD the following at the beginning of this section.

"For Informatics UG and PG students only (including those on joint degrees), or by special permission of the School.

It is recommended that students have taken a previous course in machine learning (or with significant machine learning content). Those who have not MUST register for one of the following co-requisites: Introductory Applied Machine Learning, Machine Learning and Pattern Recognition, or Accelerated Natural Language Processing."

### **2. Changes to Text Technologies for Data Science:**

TTDS will change Delivery time and Assessment, and there are some updates to the rest of the course descriptor reflecting this. Updates to TTDS course descriptor; replace the following sections with the text below -

#### **Summary:**

This course teaches the basic technologies required for text processing, focussing mainly on information retrieval and text classification. It gives a detailed overview of information retrieval and describes how search engines work. It also covers basic knowledge of the main steps for text classification.

This course is a highly practical course, where at least 50% of what is taught in the course will be implemented from scratch in course works and labs, and students are required to complete a final project in small groups. All lectures, labs, and two course works will take place in Semester 1. The final group project will be due early Semester 2 by week 3 or 4.

#### **Course description:**

##### **Syllabus:**

- \* Introduction to IR and text processing, system components
- \* Zipf, Heaps, and other text laws
- \* Pre-processing: tokenization, normalisation, stemming, stopping.
- \* Indexing: inverted index, boolean and proximity search
- \* Evaluation methods and measures (e.g., precision, recall, MAP, significance testing).
- \* Query expansion
- \* IR toolkits and applications
- \* Ranked retrieval and learning to rank
- \* Text classification: feature extraction, baselines, evaluation
- \* Web search

**Course Start (delivery):** Full Year

**Assessment:** Written exam 50%, coursework 50%, Practical Exam 0%

**Additional Information (Assessment):**

Written examination will evaluate students' understanding of the fundamentals of text technologies and IR.

Coursework will include two practical assignments to show the depth of understanding of the basics of IR and text classification; and a group project that would require applying some of the knowledge gained during course to implement a running application by a team of students.

Coursework will be designed as follows:

- 1) Two assignments for student to work individually (worth 20% in total).
- 2) One course final project assignment, to be completed in small groups (worth 30%). This project is required to be submitted near the beginning of the second semester.

**Learning Outcomes:**

Knowledge objectives:

- 1: Describe the main algorithms for processing, storing and retrieving text.
- 2: Show familiarity with theoretical aspects of IR, including the major retrieval models.
- 3: Discuss the range of issues involved in building a real search engine
- 4: Evaluate the effectiveness of a retrieval algorithm
- 5: Understand the basic steps for building and effectively evaluating text classifiers.

Practical objectives:

- 1: Build basic search engines from scratch.
- 2: Build feature extraction modules for text classification
- 3: Learn new tools for searching massive collections of text documents
- 4: Implement evaluation scripts for IR and text classification
- 5: Work effectively in a team to produce working systems

**Reading List:**

"Introduction to Information Retrieval", C.D. Manning, P. Raghavan and H. Schutze

"Search Engines: Information Retrieval in Practice", W. Bruce Croft, Donald Metzler, Trevor Strohman

"Machine Learning in Automated Text Categorization". F Sebastiani "The Zipf Mystery"

Additional research papers and videos to be recommended during lectures

**3. Changes to Neural Computation and Bioinformatics 1:**

Once new course proposals submitted to change courses from Level 11/Year 5 to Level 11/Year 4, new courses will require -

1. Updating relevant UG DPTs to add these courses to appropriate course choice lists.
2. Updating the info at <http://course.inf.ed.ac.uk/>.

**4. Changes to Bioinformatics 1 and Bioinformatics 2:**

AI tag will need to be added to Bio1 (new course) and Bio 2, and update DPTs and Course Index Sortable List - <http://course.inf.ed.ac.uk/> as appropriate.

## 5. Additional change to Bioinformatics 2:

Change the Pre-requisites box to say: "It is RECOMMENDED that students have passed: Bioinformatics 1 (New Course Code)".

## 6. Minor updates and modification to descriptor of Social and Technological networks.

The following minor modifications to the DRPS descriptor of STN are suggested based recent experience in the course. The reasons for these changes are:

1. Based on in-class feedback of students, the new description has greater emphasis on algorithms and machine learning. Students are more interested in the computational aspects (than the modeling aspects), and this also helps them in the coursework on network analysis. As most students learn some degree of machine learning, they would like to know its links with network science.
2. The course "description" column previously had relatively specific topics with emphasis on network models. The current version has greater emphasis on computation and optimisation in line with the point above, and allows more flexibility in incorporating current topics and research.
3. The previous summary descriptor had a stronger emphasis on online social networks and applications. This has been reduced and the importance of fundamentals has been clarified.

### Updated columns in DRPS descriptor

**Summary:** In this course, we will study core properties of networks arising in various social, scientific and technological contexts. We will see techniques for their analysis, and applications in social networks, world wide web, Internet, IoT etc. The course will cover fundamental theories and techniques from machine learning, algorithms and mathematics. We will see the relevance of these techniques in real networks, as well as use of network-based techniques in more general data analysis. The course will involve theoretical analysis in class, development of algorithms, and writing of programs to analyse network data.

### **Course description:**

The course will study computational, mathematical and data analysis aspects of networks. Typical topics will include properties of social networks, epidemics, spread of innovation, random graphs, metric properties, preferential attachments and power law networks. It will cover relation to data analysis and machine learning: including clustering and community detection, submodularity, optimisation, embedding (dimension reduction) and classification. Other current topics will be covered as appropriate.