Informatics 1: Data & Analysis

Lecture 1: Introduction

Ian Stark

School of Informatics
The University of Edinburgh

Tuesday 15 January 2013
Semester 2 Week 1

Welcome

Informatics 1: Data & Analysis

This course provides an introduction to representing and interpreting data from areas across Informatics; treating in particular structured, semi-structured, and unstructured data models.

Lecturer:	Ian Stark
Contact:	Ian.Stark@ed.ac.uk

Office Hours:	1030–1130 in AT4.09 on Thursdays
Web page:	http://www.inf.ed.ac.uk/courses/inf1/da

Degrees and Related Courses

Degrees

- Computer Science
- Software Engineering
- Artificial Intelligence
- Cognitive Science
- Cognitive Science (Humanities)
- Informatics

...perhaps with Mathematics, Electronics, Physics, or Management.

Other Courses

- Informatics 1: Functional Programming
- Informatics 1: Computation & Logic
- Informatics 1: Object-Oriented Programming
- Informatics 1: Cognitive Science
- Introduction to Linear Algebra
- Calculus and its Applications

Data

Data *(noun)*

Facts and statistics collected together for reference or analysis: *"there is very little data available".*

- The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.
- *Philosophy* Things known or assumed as facts, making the basis of reasoning or calculation.

Oxford Dictionaries Online

In Latin *"data"* is the plural, with singular *"datum"*, but it is now widely used as a singular mass noun: "data was collected".

Analysis

All definitions of "data" emphasise the requirement for further processing of data: invariably, the purpose of collecting data is to make some further use of it.

We shall be looking at several different kinds of data, but for all of them the topic of *data* goes hand in hand with that of the *analysis* necessary to process and interpret it.

Indeed, before even starting to collect data it's usually important to know what kind of analysis will be done with it, in order to gather, organise and manage the data appropriately.

Bits $\rightarrow$ Data $\rightarrow$ Information $\rightarrow$ Knowledge $\rightarrow$ Understanding $\rightarrow$ Wisdom

(We'll be spending most of our time towards the left)

Significance of Data and Analysis

- How much data is there on digital storage devices worldwide in total?
- How accurate/reliable is this data?
- How secure is this data?
- What do people do with this data?
- How much of it is personal data about you?

There are regular stories in the media that touch upon these questions — consequences of data inaccuracies; breaches of security with personal data; censorship of data by governments; etc.

Issues about data, how it is organised and analysed, are strongly relevant to our everyday lives.

This Course is About...

This course is not, however, about the political, legal, sociological and moral issues surrounding data. (Although these are both important and interesting.)

This course covers the methods and technologies used for large-scale collection, storage, retrieval, manipulation and analysis of data.

However, the technologies are a vehicle: this course is really about the *principles* which guide these technologies, which *challenges* they aim to address, and how they do so.

From Problems to Solutions

Example

Challenge — How can we use computers to help us extract information more efficiently from large quantities of data?

Principle — Use a custom language to describe the analysis required at a high level of abstraction, and have the computer identify the most efficient algorithm to carry it out.

Technology — SQL and query optimization engines.

Although we shall discuss specific technologies like SQL, and you should acquire skill in using them, the long-term goal is to understand the challenge and what it is that makes for a good solution.

Course Structure

Structured Data

Principles of relational databases. Entity-relationship diagrams. Relational model. Queries. Relational algebra and calculus. SQL.

Semistructured Data

Semistructured data models. XML. DTD's. XPath. Text corpora as semistructured data. Informatics applications of corpora.

Unstructured Data

Exploiting and analysing unstructured data. Information retrieval. Statistical analysis of data.

Intermission

Anonymous Survey

Q1: What is your nationality? Circle one answer below.

Scottish UK, not Scottish EU, not UK Not EU

Q2: About how many hours of physical exercise do you usually do in a week?

Q3: How many hours would you estimate you slept last night?

Lectures

Times and Places

Tuesday	11.10 – 12.00	Lecture Theatre 5, Appleton Tower
Friday	14.10 – 15.00	Lecture Theatre B, David Hume Tower

These are usually video recorded and available online later for reference.

I shall often in one lecture set reading or preparation for the next lecture.

You are strongly advised to attend all lectures, and to carry out the reading or preparation.

Textbooks

This is not a textbook course, and there is no single compulsory book.

For several parts of the course, however, I shall indicate one or more books which cover the current material — usually in much more depth and generality than required for this introductory course.

You can consult these books in the library, or borrow them, and you may find one or other helpful to you. Although the content is often similar, styles and tastes can differ significantly.

Occasionally I shall distribute photocopies of an individual textbook chapter when it is especially relevant to the course.

Exercises and Tutorials

Tutorials start in week 3 of lectures and and continue each week until the end of semester, except for Innovative Learning Week which falls between lecturing weeks 5 and 6.

You will be assigned a group and time before tutorials start. Your *course tutor* runs your tutorial group; this is not the same as your *personal tutor*. If you wish to move to a different tutorial group, please ask the ITO through their online contact form.

Exercise sheets for each tutorial are available the week before: you should work through these and bring your solutions to the tutorial for discussion.

All students registered for the course are required to attend the weekly tutorials. If you are ill or otherwise unable to attend one week then email your course tutor, and if possible attend another tutorial group in the same week.

Coursework Assignment

There will be a written assignment based on exam questions from previous years. Your course tutor will mark this and give you feedback, but it is not part of the overall course assessment.

- Available online by Thursday 7 March 2013
- Submission due by 4pm Thursday 21 March 2013
- Work returned at tutorial meetings in first week of April 2013

Final Assessment

Examination

Your overall mark for the course is based on a two-hour written exam at the end of the year. The paper follows a similar format each year and there are past papers are online.

The exam will be some time in April or May: the University Registry will publish a timetable later this semester.

Examinable Material

Any topic covered in lecture slides or tutorial exercises may tested in the exam unless indicated as *non-examinable*.

Note that many lectures will have non-examinable diversionary material at the half-way point, and some tutorial exercises will have additional "starred" exercises. This material will not be in the exam.

Online Resources

All online resources for the course are available from the Inf1-DA web page:

http://www.inf.ed.ac.uk/teaching/courses/inf1/da

Downloads	Lectures slides, videos and tutorial exercise sheets.
Blog	This forms a lecture log, with additional posts for announcements and supporting material.
Discussion group	**inf1** on http://www.forums.ed.ac.uk available to students, course tutors, and lecturing staff.
Mailing list	**inf1-da-students** for occasional important announcements.

Support

If you have questions about something in the lectures, difficulties with tutorial exercises, or want to find out more on the material, ask someone.

- Other students: in your tutorial group, in the lab, elsewhere.
- InfBASE: drop in Monday–Thursday 4–6pm in AT 5.02
- Your course tutor: in person or by email.
- The lecturer: in person after lectures, drop-in office hours, or by email.
- Online: in the discussion group, IRC #inf1, Facebook, etc.

For technical support when machines aren't working or you have problems with software on DICE, fill out the computing support form.

For administrative support in anything related to teaching, go to the Informatics Teaching Organisation (ITO) office on floor 4 of Appleton Tower, or fill out their contact form.

If you are having difficulties affecting all of your courses, or issues arising outside the University, contact your personal tutor.

Fin.

Reading for next lecture

Jeannette M. Wing.
Computational Thinking. *Communications of the ACM* 49(3):33–35.
DOI: 10.1145/1118178.1118215

Acknowledgements

This course includes material originally contributed by Frank Keller, Helen Pain, Alex Simpson and Stratis Viglas.