# Informatics 1: Data & Analysis

## Lecture 20: Course Review

Ian Stark

School of Informatics
The University of Edinburgh

Friday 27 March 2015
Semester 2 Week 10

## Plan

This is Teaching Week 10 of Semester 2, next week is Week 11, and the teaching block ends on Friday 3 April.

Your final Inf1-DA tutorial is next week, in which your tutor will return your coursework assignment with marks and feedback.

**But only if you go to the correct tutorial group!**

Find the course web page, Tutorial tab, "group membership" link.

The rest of this lecture will include:

- Exam arrangements and format;
- Summary of course topics;
- Review: XML and XPath.

This is the last lecture of Inf1-DA for this year.

# Exam: Time and Place                                               !

Informatics 1: Data & Analysis will be assessed by a single two-hour
written examination.

> Date:  Thursday 7 May 2015
> Time:  1430–1630
> Place: The Pleasance Sports Hall

This information for course code INFR08015 is current at 2015-03-27;
please check the link on the Inf1-DA web page nearer to the date to check
this and to confirm all of your exams.

## Exam: Format

As in previous years, the exam will have three compulsory questions.

- Read all questions before beginning the paper
- You don't need to do the questions in order
- Don't assume a question is only using one part of the course
- If you get stuck on one question: don't waste too much time on it; do go on to the next question; and don't give up!

Calculators are permitted, and will be provided at the exam hall. These are a standard scientific model: you can try one out at the ITO if you wish.

Do not bring your own calculator. Better, do not bring any electronic devices into the exam hall at all.

## Past Exam Papers

### Past Questions

All tutorial exercises have Examples and Solutions from past exam papers.

These are chosen to match each course topic. Solutions include comments about different possible answers, and notes on important things to include.

### Past Papers

The University Library keeps a full set of past papers online.

http://exampapers.ed.ac.uk

In each year there are two Inf1-DA exams, from the main and resit diets.

All questions since the 2009 exam are appropriate for the current syllabus.

## Using Past Exam Papers

Past papers are a good source of revision material, and I strongly recommend you attempt as many of these questions as you can.

However, please note the following.

- The key exam preparation task is to *master the course material:* understand it all and be able to apply it in practice.

- Memorizing the answers to previous questions is not a helpful way to tackle future ones.

- Attempting past exam questions yourself, to test your knowledge, and identify any gaps, is a much more effective way to learn.

If you are puzzled by a past question or the solution provided, ask on *Piazza* or email me directly.

# Exam Content

## Examinable Material

Unless otherwise specified, all of the following material is examinable:

- Topics covered in lectures

- Directed reading distributed in lectures

- Topics covered in the weekly exercise sheets

All non-examinable additional material in
lectures is marked by being on
purple-shaded slides with a "$+$" in the top
right-hand corner.

### Limits of Computation $+$

Matter organised to provide the greatest
possible computing power is fancifully
known as computronium.

In the 1960's Hans-Joachim Bremermann
was one of the first people to estimate
upper limits to computation.

His *Bremermann limit* is the computation
which could be performed using the earth,
over the period of its existence so far.

This is around $10^{93}$ bits of computation.

That's enough to solve the travelling
salesman problem for 300 cities.

But just the once.

## Topic Summary

The entity-relationship model, ER diagrams. The relational model, SQL DDL. Translating an ER model into a relational one. Relational algebra, tuple-relational calculus, SQL queries; translating between all three.

Semistructured data models and the XPath data tree. XML documents. Schema languages and DTDs. Relational data converted into XML. XPath as a query language.

Corpora: what they are and how they are made; examples. Annotations and tagging. Concordances, frequencies, $n$-grams, collocations. Methods for machine translation.

Information retrieval: what it is, evaluating and comparing performance of IR systems; the vector space model and cosine similarity measure.

Data scales, summary statistics, population vs. sample; hypothesis testing and significance; correlation coefficient, $\chi^2$ test.

# Some Specific Items

## Corpora

In general it is the principles of corpora that are examinable, rather then the precise details of individual corpora. Similarly, you should be familiar with the principles underlying POS-tagging and syntactic annotation, but you do not need to know detailed linguistics or specific tag sets.

You should however, be able to give examples of a corpus or a POS tag.

The CQP tool was used in a tutorial, so is examinable — although again for general principles and use, not every detail of syntax.

## Statistics

You are not expected to memorize critical value tables; however, you should be able to use one if provided.

You are expected to know the formulas for the various statistics used, and to be able to calculate with them.

# XML and XPath

XML trees  Nodes: root, internal, leaves. Parents, children, ancestors, descendants, siblings. XPath node types: root, element, text, attribute.

XML documents  Nested start and end tags; text content; tag attributes. Ordering of nodes matters; of attributes doesn't. Well-formed documents.

Validation  Specifying a hierarchy and attribute set. Validating document D against schema S. Schema languages, such as DTD.

XPath navigation  Path expressions identify a set of nodes. Navigation: axes, node tests, predicates.

# 2013/2014 Main Diet Question 2

```xml
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE libraries SYSTEM "libraries.dtd">
<libraries>
    <library village="Aberfoyle">
        <section name="Fiction">
            <book author="Scott" title="Rob Roy"></book>
        </section>
        <section name="Nonfiction">
            <book author="Ray" title="Learning XML"/>
        </section>
    </library>
    <library village="Kippen">
        <book author="Hill">Where's Spot?</book>
    </library>
</libraries>
```

# Draw the XPath data model tree for this XML document
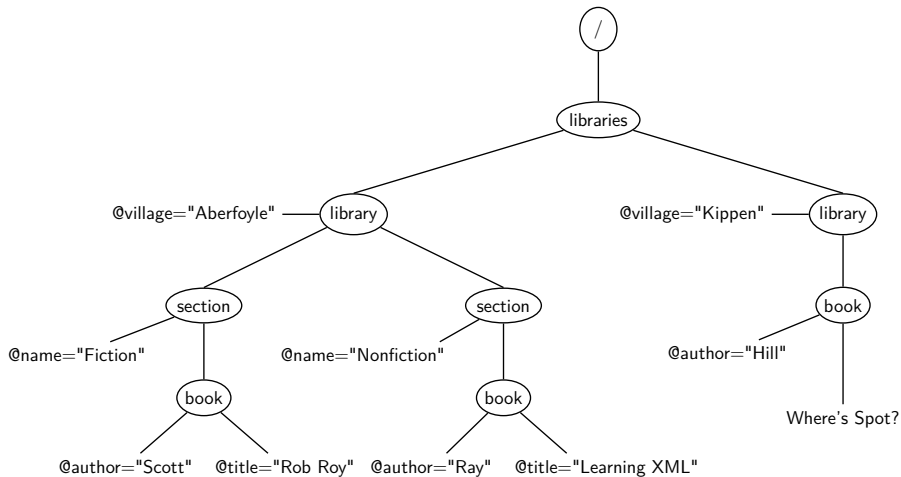
# Write a suitable DTD for libraries.dtd

&lt;!**ELEMENT** libraries (library)+&gt;
&lt;!**ELEMENT** library (section+|book+)&gt;
&lt;!**ELEMENT** section (book)*&gt;
&lt;!**ELEMENT** book (#**PCDATA**)&gt;

&lt;!**ATTLIST** library village **CDATA** #**REQUIRED**&gt;
&lt;!**ATTLIST** section name (Fiction|Nonfiction) #**REQUIRED**&gt;
&lt;!**ATTLIST** book author **CDATA** #**REQUIRED**&gt;
&lt;!**ATTLIST** book title **CDATA** #**IMPLIED**&gt;

# Some things not to put in libraries.dtd

**✗** <!**ELEMENT** library (section|book)∗>

> This states that library can be a mixture of both sections (which contain books) and individual books.

**✗** <!**ELEMENT** library (section+,book+)>

> This says that the XML record must contain some sections, followed by some books.
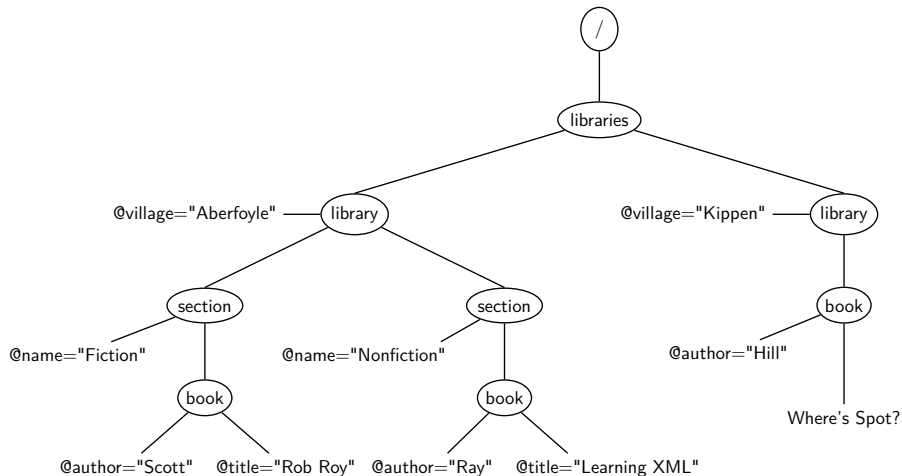
**✗** <!**ELEMENT** book (#**PCDATA**|**EMPTY**)>

> Using (#**PCDATA**) already includes the possibility of an element being empty. The **EMPTY** keyword is reserved for the case where an element *must* be empty.

Write XPath expressions to list the following from such a document:

(i) All of the villages included.

# Draw the XPath data model tree for this XML document

# XPath expressions (1/3)

Write XPath expressions to list the following from such a document:
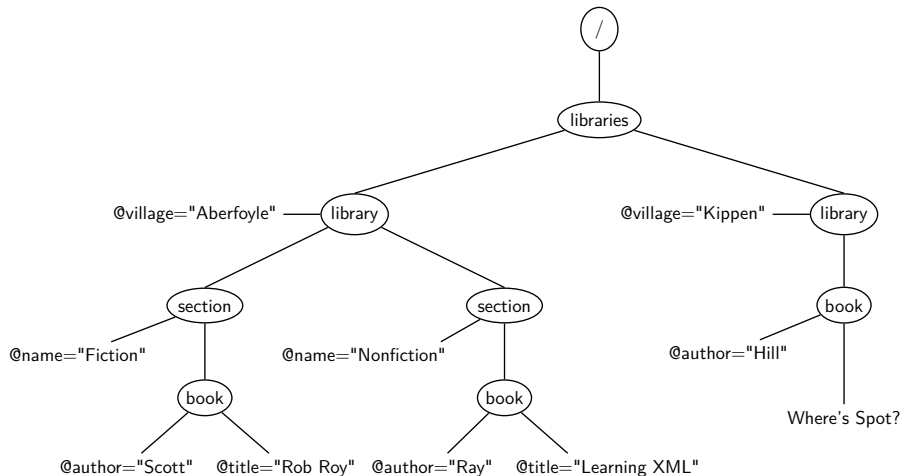
(i) All of the villages included.

- //@village

- / libraries / library /@village

# XPath expressions (2/3)

Write XPath expressions to list the following from such a document:

(ii) All authors of books listed as Nonfiction.

# Draw the XPath data model tree for this XML document

# XPath expressions (2/3)

Write XPath expressions to list the following from such a document:

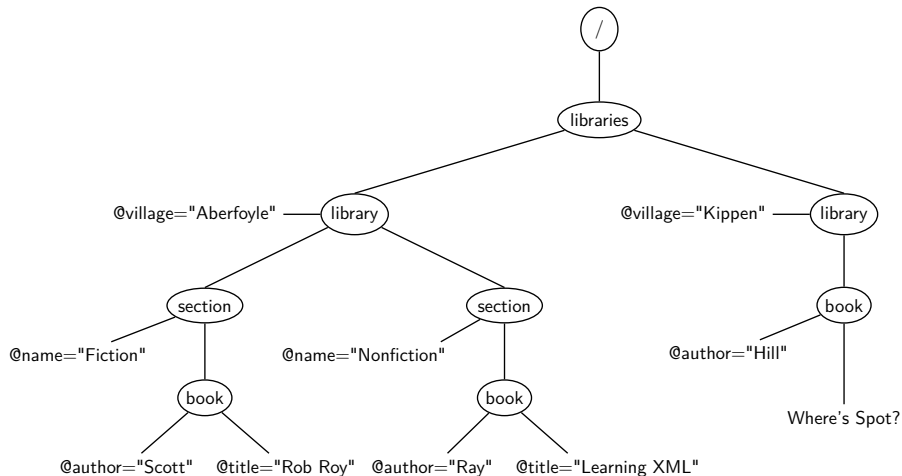(ii) All authors of books listed as Nonfiction.

- //section[@name="Nonfiction"]/book/@author

- //*[@name="Nonfiction"]//@author

- //book[../../section/@name="Nonfiction"]/@author

Write XPath expressions to list the following from such a document:

(iii) Names of villages whose school library holds a book by Dickens.

# Draw the XPath data model tree for this XML document

# XPath expressions (3/3)

Write XPath expressions to list the following from such a document:

(iii) Names of villages whose school library holds a book by Dickens.

- //library[.//@author="Dickens"]/@village

- //book[@author='Dickens']/**ancestor**::library/@village

## Other Revision Slides

The equivalent lecture in 2013/2014 contained different revision material.

- Summary statistics, estimates from samples, hypothesis testing.
- Tuple-relational calculus.

You can find last year's lecture slides through the Archive tab on the Inf1-DA web pages.

## You can do this

The Inf1-DA syllabus and exam questions are written to be achievable. Every year large numbers of students pass the exam writing straightforward correct answers about things they understand. You can do this too.

# Finis

## Anything Else?

If you have further questions about the course material, past lectures, exercises, the exam, or anything else, please:

- Post a question on *Piazza*; *or*
- Ask your course tutor, in person or by email; *or*
- Ask me, in person or by email.

## Course Survey

Please complete the online survey for this course. This is anonymous, and I read every submission. Ideally, do it immediately after this lecture.

> http://www.inf.ed.ac.uk/teaching/survey

Follow the link to the Semester 2 Course Surveys.