

Tutorial 8: Practice Exam Questions

Informatics 1 Data & Analysis

Week 10, Semester 2, 2017/18

Read this first: it is not the same as the other tutorials

Following the strike by university staff I have changed around some of the tutorial work. The material usually in Tutorial 8 has now moved a week later and will appear in Tutorial 9, by which time the topics it addresses will have been covered in lectures. In its place this week I am providing some exam question practice.

This worksheet includes two questions, both taken from the 2016/17 resit examination.

- Before your tutorial, attempt both questions. Write out your answers on paper and with all your working: just as you would expect to do in the exam itself.
- Bring your written work to the tutorial.
- At the tutorial you will use the marking guidelines from the original exam to assess these solutions. You will be working with others and with the help of your tutor.

You don't need to do this work under exam conditions. As usual with Inf1-DA tutorials, this is *formative* assessment entirely for your feedback and learning. No marks here will directly contribute to your final grade. Because of this you can freely ask for help on the questions, discuss on *Piazza*, and talk about your work with other students. Please do. The aim here is for you to get practice answering questions, find out what's involved, and prepare for doing this well in the exam itself.

If you have problems with some parts of the questions then write down what it is that you find challenging and ask your tutor about this in the meeting.

If you are uncertain about the meaning of some question, or think that there are several possible answers, then write all this down and raise it for discussion in the tutorial.

As usual, it's important both for your learning and other students in the group that you come to tutorials properly prepared. Students who have not attempted the exam questions will be sent away from the tutorial to do them elsewhere and return later.

Tutorials are a key part of course participation: if you are ill or otherwise unable to attend one week then email your tutor, and if possible attend another tutorial group in the same week.

Please send any corrections and suggestions to Ian.Stark@ed.ac.uk

1. [*This question is worth a total of 40 marks.*]

- (a) An online app store uses a relational database to keep track of everything in the store. One of the tables holds information about all apps in the store's online catalogue, and is defined as follows.

```
create table app (  
  title          varchar(100) not null, -- Each app has a different title  
  version        integer not null,      -- An app might have several versions  
  author         varchar(100) not null, -- One author can write multiple apps  
  cataloguecode  varchar(10) not null,  -- Unique identifier for items in the store  
  price          integer not null,      -- Price in pennies  
  primary key (cataloguecode)  
)
```

As you can see from the comments, the store might hold several versions of the same app, but they will always have different catalogue codes.

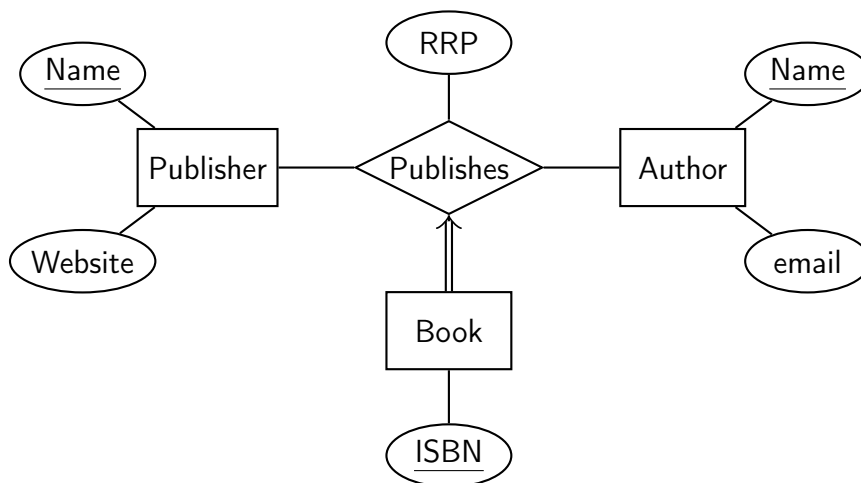
The following five terms have precise meanings when applied to a relational database. For each one, give a one-sentence description of its meaning and an example based on the **app** table declaration above.

- (i) Superkey
- (ii) Key
- (iii) Composite key
- (iv) Candidate key
- (v) Primary key

[10 marks]

QUESTION CONTINUES ON NEXT PAGE

The following entity-relationship diagram shows part of a conceptual design document for another database, this time for a bookshop.



The diagram is accompanied by the following information.

ISBN International Standard Book Number: A unique 13-digit identifier issued for every published book. Record as a 13-character string.

RRP Recommended Retail Price: Sale price suggested by the publisher. Record as an integer number of pence.

Names are required, email and web addresses are optional; and all of these are assumed to fit within 300 characters.

- (b) The double line and arrow joining the **Book** entity to the **Publishes** relationship indicates two distinct types of *constraint* in this link.
- What are the two constraints called?
 - Which graphical feature indicates which constraint?
 - What does each constraint mean?

[8 marks]

- (c) Design an SQL data declaration of tables suitable to implement this entity-relationship model. You should take into account all the constraints shown in the ER diagram and its accompanying information. Include **not null** and **on delete** declarations where appropriate.

[22 marks]

2. [This question is worth a total of 30 marks.]

The following small XML document is a marked-up version of a fragment from the play *Rosencrantz and Guildenstern Are Dead* by Tom Stoppard.

```
<speech speaker="Guildenstern">
  <line>
    <w>The</w>
    <w>colours</w>
    <w>red</w>
    <punct>,</punct>
    <w>blue</w>
    <w>and</w>
    <w>green</w>
    <w>are</w>
    <w>real</w>
  </line>
  <line>
    <w>The</w>
    <w>colour</w>
    <w>yellow</w>
    <w>is</w>
    <w>a</w>
    <w>mystical</w>
    <w>experience</w>
  </line>
</speech>
```

- (a) Draw the XPath data model for this document as a tree. [8 marks]
- (b) Write an XML DTD for a **Speech** document type to validate such speeches. Assume that every speech must have an identified speaker. [10 marks]
- (c) Suppose a large XML document contains many speeches, nested inside Plays, Acts, Scenes and so on. Write XPath expressions to extract the following.
- (i) All lines spoken by Rosencrantz.
 - (ii) All speakers using the word “yellow” in a speech.
- [6 marks]
- (d) Standard corpora for linguistic research like the *British National Corpus* bring together work from many sources. Building them requires balancing and sampling in order to ensure that they are representative.

Explain the meaning of *balancing*, *sampling* and *representative* here. [6 marks]