

**UNIVERSITY OF DUBLIN
TRINITY COLLEGE**

**Faculty of Engineering, Mathematics and Science
School of Computer Science & Statistics**

**BA (Mod) Business & Computing
Integrated Computer Science Programme**

Trinity Term 2014

Information Management 1

Prof. Declan O'Sullivan

Monday 28th April, 2014

GoldHall

14:00 – 15:30

Instructions to Candidates:

Answer two questions
All questions carry equal marks.
Answer each question in a separate answer book.
Hand up complete question paper.

Materials permitted for this examination:

Calculator

**SOME QUESTIONS REFER TO THE FOLLOWING
FIGURE A:**

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<news>
  <news_item timestamp='1'>
    <title> Gorilla Corporation acquires YouNameItWeIntegrateIt.com </title>
    <content>
      <par> Today, Gorilla Corporation announced that it will purchase
YouNameItWeIntegrateIt.com. The shares of YouNameItWeIntegrateIt.com dropped
$3.00 as a result of this announcement. </par>
      <par> As a result of this acquisition, the CEO of YouNameItWeIntegrateIt.com Bill
Smarts resigned. He did not announce what he will do next.</par>
      <par>YouNameItWeIntegrateIt.com is a leading systems integrator that enables
<quote>brick and mortar</quote> companies to have a presence on the web. </par>
    </content>
    <date>1-20-2000</date>
    <author>Mark Davis</author>
    <author>John Smith</author>
    <news_agent>News Online</news_agent>
  </news_item>
  <news_item timestamp='2'>
    <title>Foobar Corporation releases its new line of Foo products today</title>
    <content>
      <par> Foobar Corporation releases the 20.9 version of its Foo products. </par>
      <par> The President of Foobar Corporation announced that they were proud to
release 20.9 version of Foo products
      </par>
      <figure>
        <title>Presidents of Foobar Corporation and TheAppCompany
        Inc. Shake Hands</title> <image source='handshake.jpg' />
      </figure>
    </content>
    <date>1-20-2010</date>
    <news_agent>Foobar Corporation</news_agent>
  </news_item>
  <news_item >
    <title>Foobar Corporation is suing Gorilla Corporation for patent infringement </title>
    <content>
      <par> In surprising developments today, Foobar Corporation announced that it is
suing Gorilla Corporation for patent infringement. </par>
      <par>The tension between Foobar and Gorilla Corporations has been increasing
ever since the Gorilla Corporation acquired more than 40 engineers who have left
Foobar Corporation, TheAppCompany Inc. and YouNameItWeIntegrateIt.com over the
past 3 months. </par>
    </content>
    <date>2-15-2014</date></news_item></news>
```

Question 1

Using UML and XML, design an information system to support aspects of the operation of a radio broadcasting company (e.g. 2FM). Include cardinality and other constraint information.

- a) Model at least 6 UML classes (each with several attributes) with named associations between the classes representing your information.
- b) Model 2 UML use cases (diagrams and textual descriptions) that will be supported by these classes.
- c) Provide a UML sequence diagram for each use case that indicates the sequence of operations involved that will implement the UML use cases designed in part b) above.
- d) Provide a detailed commentary on the design decisions you took during the modelling task.

(25 Marks)

Question 2

- a) What are XPath and XQuery used for, and when is XQuery more suitable to use than XPath? Describe (using examples) 3 queries you can express in XQuery that are impossible to express in XPath.

(4 Marks)

- b) Use DTD notation to fully describe the XML document shown in Figure A. Provide explanations for your design decisions.

(8 Marks)

- c) Define and explain XQuery statements for each of the following queries posed over the document in Figure A. Show expected results and explain your design decisions.

- i) Return just the values of timestamp attributes.
- ii) Return each value of a date element; within a new element called "US_date".
- iii) Return in one single element called <titles>, the name of each title mentioned in the document, separated by a semi-colon ";", without the individual <title> tags.
- iv) Return the second author of the news_item with timestamp of "1".

(13 Marks)

Question 3

- a) Explain the difference between a Primary Index and a Secondary Index.
(4 Marks)
- b) Describe the algorithm for inserting a key into a B-Tree. Use diagrams with examples to illustrate.
(8 Marks)
- c) Given the B-Tree in Figure B (with $m=5$), show the growth of the B-Tree as you add the keys in the following order:

18, 20, 23, 5, 24, 27, 16, 32, 37, 44, 55, 25, 57, 69, 72

- Show the tree at each stage and provide an explanation for any splits
- What is the average number of node splits per new key inserted?
- What is the average search time?

(13 Marks)

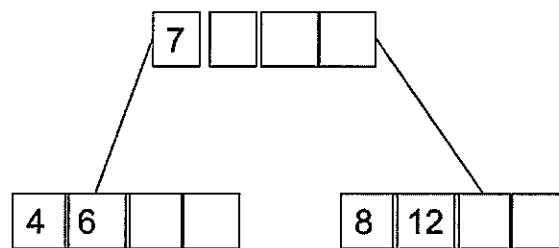


FIGURE B

Question 4

The success of Information Retrieval depends primarily on the quality of indexing of text in content or metadata. Discuss.

Cover at least the following points in your answer:

- the role of indexing of text content in information retrieval;
- what is meant by index effectiveness;
- explain in detail the different steps/approaches used in indexing a document (e.g. stemming);
- explain the usefulness of "Zipf's law" and term weighting.

(25 Marks)

© UNIVERSITY OF DUBLIN 2014