

COMPSCI 732 S1 C

Software Tools and Techniques

COMPSCI 732

1

Course Content

The course covers advanced topics in tools that assist in the production of software, with a strong emphasis on practical aspects.

Part 1 (Gill Dobbie) investigates tools for the storage and management of XML documents

Part 2 (John Hosking) looks into integrated software development environments

Part 3 (Robert Amor) covers tools available for mapping information within integrated environments

COMPSCI 732

2

Course Assessment

- 3 assignments at 16.6% each (50%)
- Exam with equal parts from each lecturer (50%)

COMPSCI 732

3

Student Financials – from FoS

- The Change of Course Period for First Semester ends on Friday 11 March 2005. After this time a late enrolment fee of \$50 per course may be applied to late enrolments.
- Tuition fees for First Trimester are now overdue, and we require the outstanding amount to be paid promptly. If your account remains unpaid we may be forced to remove your current enrolment (s) and you will lose access to Library and Cecil. A reinstatement fee of \$25 will be charged.
- If you have not received an invoice by now or your mailing address has changed, you need to update your mailing address through nDeva, and e-mail us at fees@auckland.ac.nz to send you an invoice. Alternatively, you can call into the Student Financials Office, Room 108 of the ClockTower, or call us on (09) 3737599 ext 84422.
- If you have applied for a student loan and have not received payment confirmation from StudyLink please contact them on 0800 88 99 00.
- You can pay your tuition fees by Online Payment now. Log in using your netID and password at <http://ndeva.auckland.ac.nz> select the 'Financial Services' link on the left-hand menu and then select 'Pay your Account'. This will take you to the Direct Payment Solutions website where you can complete the payment process.

COMPSCI 732

4

Plagiarism – what is it?

- Using the work of other scholars or students when preparing coursework or writing an examination and pretending it is your own by not acknowledging where it came from. This is called plagiarism. Course coordinators, lecturers or tutors are the appropriate people with whom you should discuss how to appropriately use and acknowledge the work of others.
- Cheating is viewed as a serious offence by The University of Auckland. Penalties are administered by the Discipline Committee of the Senate, and may include suspension or expulsion from the University.

COMPSCI 732

5

Plagiarism – what happens?

- The Computer Science Department uses many ways to check that the work students submit for marking is their own and was not produced by, or copied from, someone else. In particular, for most programming assignments, the department uses a program comparison program to automatically compare all submissions from students. Also Turnitin.com may be used on essays and reports. This detects similarity to online material and submitted works in its own database.
- All assignments where plagiarism is detected are checked for similarity by the course supervisor or another suitable person associated with the course.
- All assignments deemed to be too similar are automatically allocated a zero mark.

To ensure you are not identified as cheating you should follow these points:

- Always do individual assignments by yourself.
- Never loan your code to another person.
- Never put your code in a public place (e.g., your web site).
- Never leave your PC without locking the screen (e.g., to get food, to have a drink, or to go to the toilet). You are responsible for the security of your account.
- Never get code from a tutor (e.g., private tutors). Several tutors have been caught giving the same code to all their students.
- Always reference the source for text you copy as part of the answer to an assignment.

COMPSCI 732

6

Who is this lecturer?

Gill Dobbie

Rm 303 475

Ph 09 373 7599 ext. 83949

Email gill@cs.auckland.ac.nz

- BTech, MTech at Massey Uni.
- PhD at the University of Melbourne (UoM).

“The foundations of deductive object oriented databases”.

- Worked for a couple of years in industry.
- Lectured at Massey University, UoM, VUW.
- Collaborate with researchers at VUW and NUS.
- Current research in databases: building a database specialized to XML or semistructured data, algorithm correctness, access control, data modeling, data mining.

COMPSCI 732

7



Why is data important?

- Organizational data is generally not valued as highly as other organizational assets such as cash or monetary resources, real estate, inventory and reputation.
- The data, then, is not well understood, often less well managed, and routinely underutilized.
- The wealth of information to be gleaned from stored data represents untapped resources that can be made available for corporate management to grow its business.

COMPSCI 732

8

Relational Databases

- Simple
- Deal with structured data
- Model constraints in schema
- Indexes to improve query performance
- Standard query language
- Have mathematical foundation that is used for modeling algorithms and reasoning about properties and correctness

XML Documents

XML documents fall into two broad categories:
data-centric and *document-centric*.

```
<?xml version="1.0"?>
<!-- This file represents a fragment of a book store inventory database -->
<bookstore>
  <book genre="autobiography">
    <title>The Autobiography of Benjamin Franklin</title>
    <author>
      <first-name>Benjamin</first-name>
      <last-name>Franklin</last-name>
    </author>
    <price>(US)$8.99</price>
  </book>
  <book genre="novel">
    <title>The Confidence Man</title>
    <author> Herman Melville</author>
    <price>(NZ)$11.99</price>
  </book>
</bookstore>
```

Data Centric

Document Centric

```
<Product>
  <Name>Turkey Wrench</Name>
  <Developer>Full Fabrication Labs, Inc.</Developer>
  <Summary>Like a monkey wrench, but not as big.</Summary>
  <Description>
    <Para>The turkey wrench, which comes in <i>both right- and left-handed versions (skyhook optional)</i>, is made of the <b>finest stainless steel</b>. The Readi-grip rubberized handle quickly adapts to your hands, even in the greasiest situations. Adjustment is possible through a variety of custom dials.
  </Para>
  <Para>You can:</Para>
  <List>
    <Item><Link URL="Order.html">Order your own turkey wrench </Link> </Item>
    <Item><Link URL="Wrenches.htm">Read more about wrenches </Link> </Item>
    <Item><Link URL="Catalog.zip">Download the catalog</Link></Item>
  </List>
  <Para>The turkey wrench costs <b>just $19.99</b> </Para>
</Description>
</Product>
```

Tools for the storage and management of XML documents

- This is the area where I currently do most of my research.
- I come to this area from a database perspective, as do many other researchers in this area.
- There are a lot of issues:
 - Removing redundancy
 - Modeling constraints on the data
 - Mapping XML to RDBs
 - Storing XML in native XML databases
 - Query performance in native XML databases
 - Indexes in native XML databases
 - Etc.

Summary

- Data is an asset that is undervalued.
- Relational databases are a well developed technology for structured data.
- Relational databases are not suitable for unstructured or semistructured data.
- Native XML databases are suitable for semistructured data but they are less well developed.
- In Gill's section of the course (Part 1), we will investigate tools and techniques for the storage and management of XML documents.

Source and further reading

- Ramakrishnan, Raghu and Gehrke, Johannes. Database Management Systems. McGraw Hill.
- Elmasri, Ramez and Navathe, Shamkant. Fundamentals of Database Systems. Addison Wesley.
- Bourett, Ronald. XML and Databases.
Available at :
<http://www.rpbouret.com/xml/XMLAndDatabases.htm>
- Bourett, Ronald. XML Database Products.
Available at :
<http://www.rpbouret.com/xml/XMLAndDatabases.htm>
- Obasanjo, Dare. An exploration of XML in DBMS.
Available at :
<http://www.25hoursaday.com/storingandqueryingxml.html>