

Computer Science E-259

XML with Java

Lecture 1: Introduction

31 January 2007

David J. Malan

`malan@post.harvard.edu`

The Hype

In the Press

"XML, as a context-rich, data-neutral file format, is probably the most important new technology development of the last two years."

Michael Vizard, InfoWorld

"An idea almost as good as peanut butter and chocolate!"

XML/EDI Group

"'[A] spreading ooze of data' will expand as people exchange data more easily with XML."

Rita Knox, Gartner Group

"XML is like sex, even when it's bad it's still pretty good."

Tim Bray, tbray.org

The Hype

In the Classroom

*"XML is beautiful, but as with beautiful people,
it is neither easy to get along with nor quick."*

*"XML's strength is its wide adoption and excellent tools.
XML itself is not that exciting."*

Computer Science E-259

Motivations for the Course

- The hype surrounding XML is huge
- Strangely enough, the value can be huge as well
 - XML is the primary technology behind the “B2B revolution”
 - XML is the primary technology behind Web Services
 - XML is the primary technology behind Microsoft’s .NET
 - XML makes developing websites easier and more flexible than ever before

Computer Science E-259

Goals of the Course

- Cut through the hype and get to the value
- Focus on
 - practicality: what you need to know to do real work
 - applications: what are the tools and technologies necessary to put XML to use
 - possibilities: what are some of the most common ways XML is being used in applications
- Emphasize understanding from the bottom up

Computer Science E-259

Rewards of the Course

- A solid understanding of XML, the syntax, data structures, and algorithms that surround it
- Xperience
 - DTD, SVG, XForms, XInclude, XLink, XML Base, XML Encryption, XML Key Management, XML Namespaces, XML Schema, XML Signature, XPath, XPointer, XQuery, XSL-FO, and XSLT
 - SAX and DOM
 - JAXP and TrAX
 - JavaServer Pages and Java Servlet
 - HTTP, SOAP, web services, WSDL
 - FOP, Stylus Studio, Tomcat, Xalan, Xerces, XMLSpy...

Computer Science E-259

Agenda

- Computer Science E-259
- J2EE
- XML
 - What
 - Who
 - When
 - How
 - Why
- Computer Science E-259

J2EE

API

<http://java.sun.com/javaee/5/docs/api/>

J2EE

J2SE

<http://java.sun.com/j2se/1.5.0/docs/api/>

XML

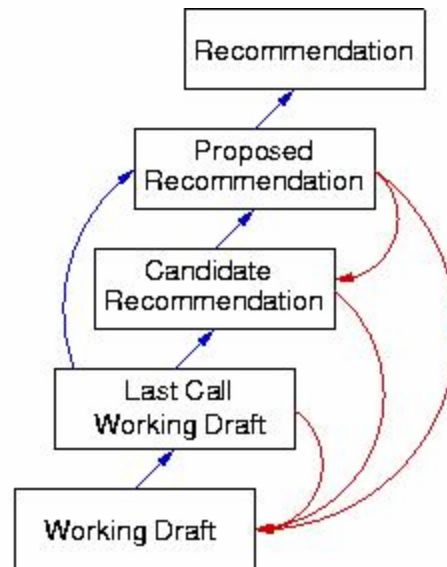
What

- XML is a language for creating other languages!
- XML lets you define schemas for tag-based languages (ergo, “markup language”)
- XML allows you to extend any existing language (schema) with your own tags (ergo, “eXtensible”)
- Examples of XML schemas
 - financial transactions (stock transactions)
 - business documents (purchase order, invoice)
 - remote procedure calls (SOAP)
 - configuration files (security, server properties)

XML

Who

World Wide Web Consortium (W3C)
<http://www.w3c.org/>



XML

When

- The World Wide Web Consortium (W3C) formed an XML Working Group in 1996 with these design goals
 1. XML shall be straightforwardly usable over the Internet
 2. XML shall support a wide variety of applications
 3. XML shall be compatible with SGML
 4. It shall be easy to write programs which process XML documents
 5. The number of optional features in XML is to be kept to the absolute minimum, ideally zero
 6. XML documents should be human-legible and reasonably clear
 7. The XML design should be prepared quickly
 8. The design of XML shall be formal and concise
 9. XML documents shall be easy to create
 10. Terseness in XML markup is of minimal importance

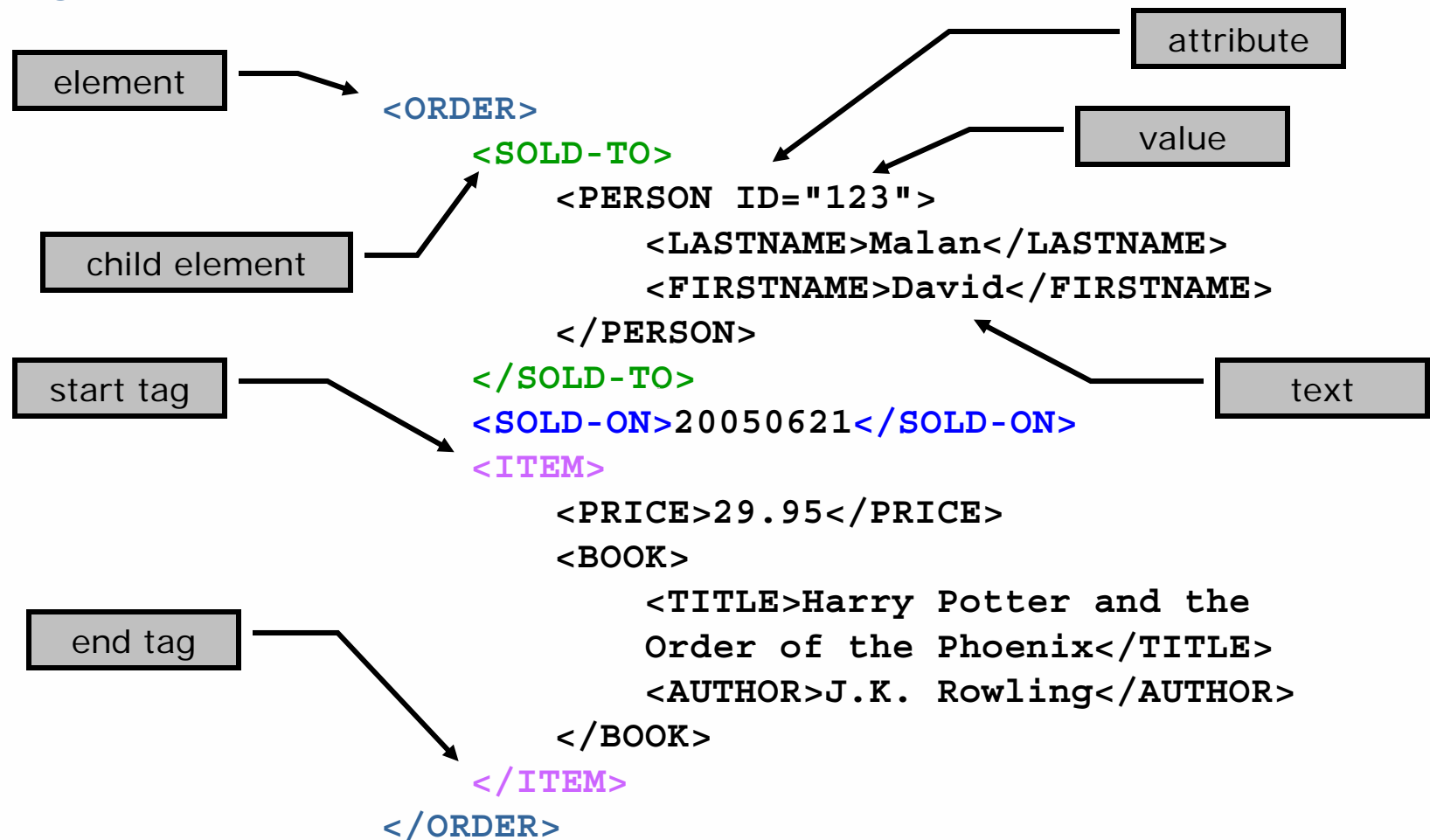
XML

When

- XML 1.0 became a standard (W3C recommendation) on 10 February 1998
 - effort grew out of experience with SGML
 - cooperation between various W3C member organizations (including Microsoft, Sun, IBM, HP, Adobe)
- XML 1.1 became a standard (W3C recommendation) on 4 February 2004
 - less rigid constraints on names
 - addresses issues involving character sets

XML

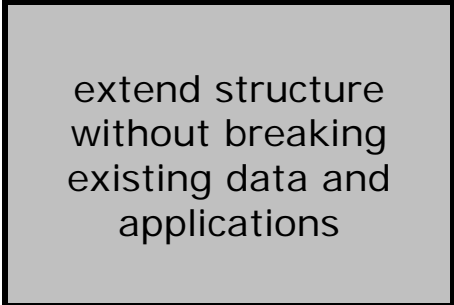
How



XML

How

```
<ORDER>
  <SOLD-TO>
    <PERSON ID="123">
      <LASTNAME>Malan</LASTNAME>
      <FIRSTNAME>David</FIRSTNAME>
      <INITIAL>J</INITIAL>
      <ADDRESS>
        <STREET>Oxford Street</STREET>
        <NUMBER>33</NUMBER>
        <CITY>Cambridge</CITY>
        <STATE>MA</STATE>
      </ADDRESS>
    </PERSON>
  </SOLD-TO>
  <SOLD-ON>20050621</SOLD-ON>
  <ITEM>
    ...
  </ITEM>
</ORDER>
```



extend structure
without breaking
existing data and
applications

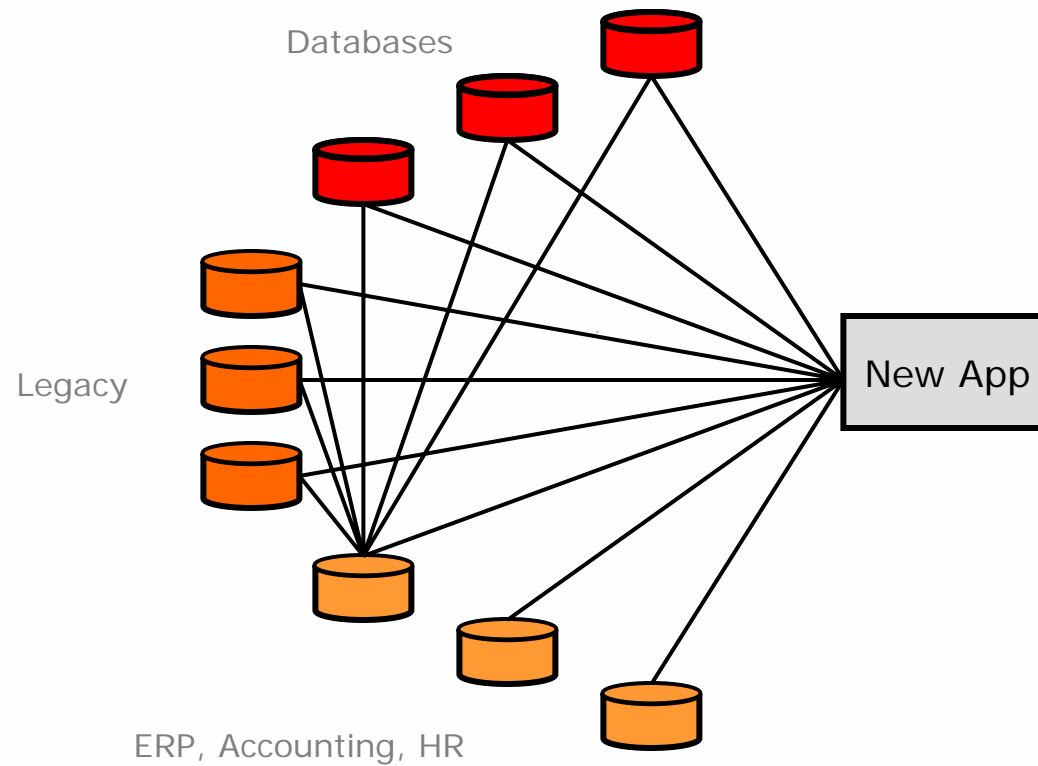
XML

Why: Content v. Presentation

<http://www.blockbuster.com/>

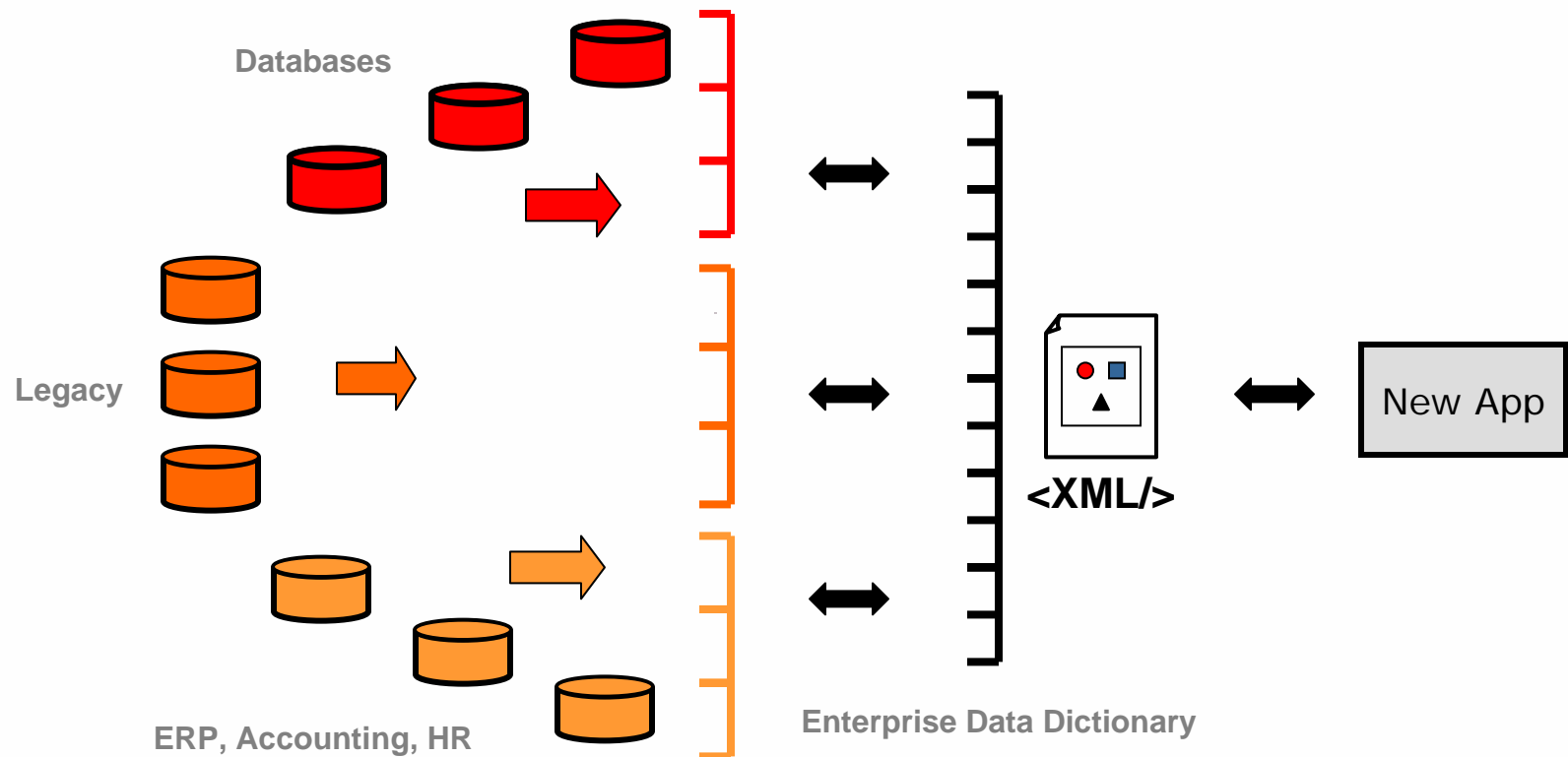
XML

Why: Application Integration



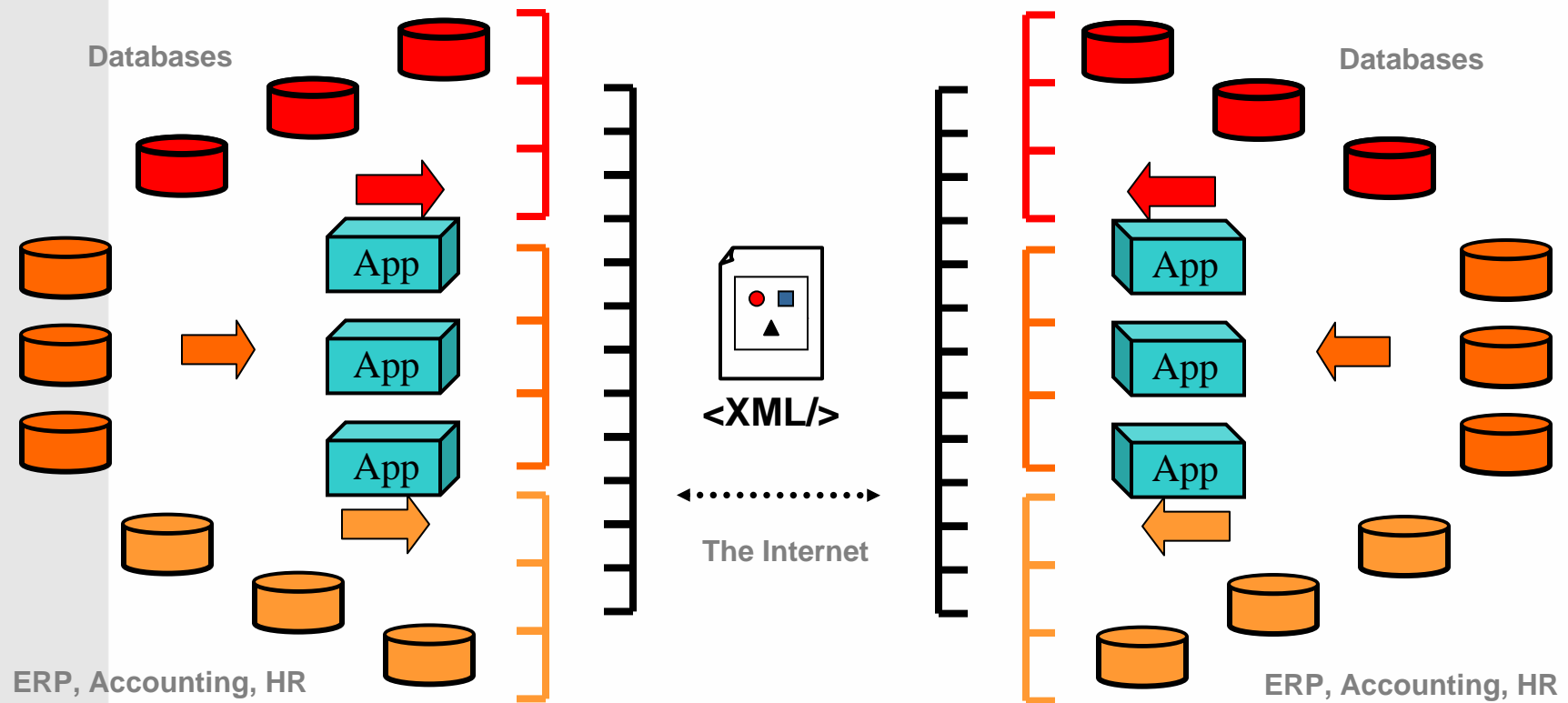
XML

Why: Application Integration



XML

Why: Application Integration



XML

Why: Platform-Independent Services

- In the past, functionality was exposed to remote clients via various remote object standards
 - COM/DCOM
 - CORBA
 - Java RMI
- Using these remote objects required a significant investment in a platform
- Web Services provide a new way to expose functionality
 - use XML and XML data types for transport
 - work with any platform
 - provide a bridge to existing business services

Computer Science E-259

Prerequisites

- Comfort with Java and HTML or XHTML is assumed.

Computer Science E-259

Expectations

- Attend or watch all lectures
- Complete four assigned projects
- Design and implement a final project

Computer Science E-259

Grades

- Project 1
- Project 2
- Project 3
- Project 4
- Final Project

Computer Science E-259

Lectures

- Lecture 1: Introduction
- Lecture 2: XML 1.1 and SAX 2.0.2
- Lecture 3: DOM Level 3
- Lecture 4: XPath 1.0 and XSLT 1.0
- Lecture 5: XPath 1.0 and XSLT 1.0, Continued
- Lecture 6: Namespaces in XML 1.1, SVG 1.1, and XSL (XSL-FO) 1.0
- Lecture 7: HTTP 1.1, JavaServer Pages 2.0, and Java Servlet 2.4
- Lecture 8: XQuery 1.0 and DTD
- Lecture 9: XML Schema (Second Edition)
- Lecture 10: XML Schema (Second Edition), Continued
- Lecture 11: Web Services, SOAP 1.2, and WSDL 1.1
- Lecture 12: Web Services, SOAP 1.2, and WSDL 1.1, Continued
- Lecture 13: To Be Announced
- Lecture 14: X{Forms,Link,Include} 1.0
and XML {Base,Encryption,Key Management,Signature}

Computer Science E-259

Projects

- Project 1: My First Sony® XML Parser
- Project 2: XTube
- Project 3: Wahoo!
- Project 4: Scamazon.com

Computer Science E-259

Final Project

- Pre-Proposal
- Proposal
- Status Report
- Design Document
- Implementation and Documentation

Computer Science E-259

Exams

- No Midterm
- No Final Exam

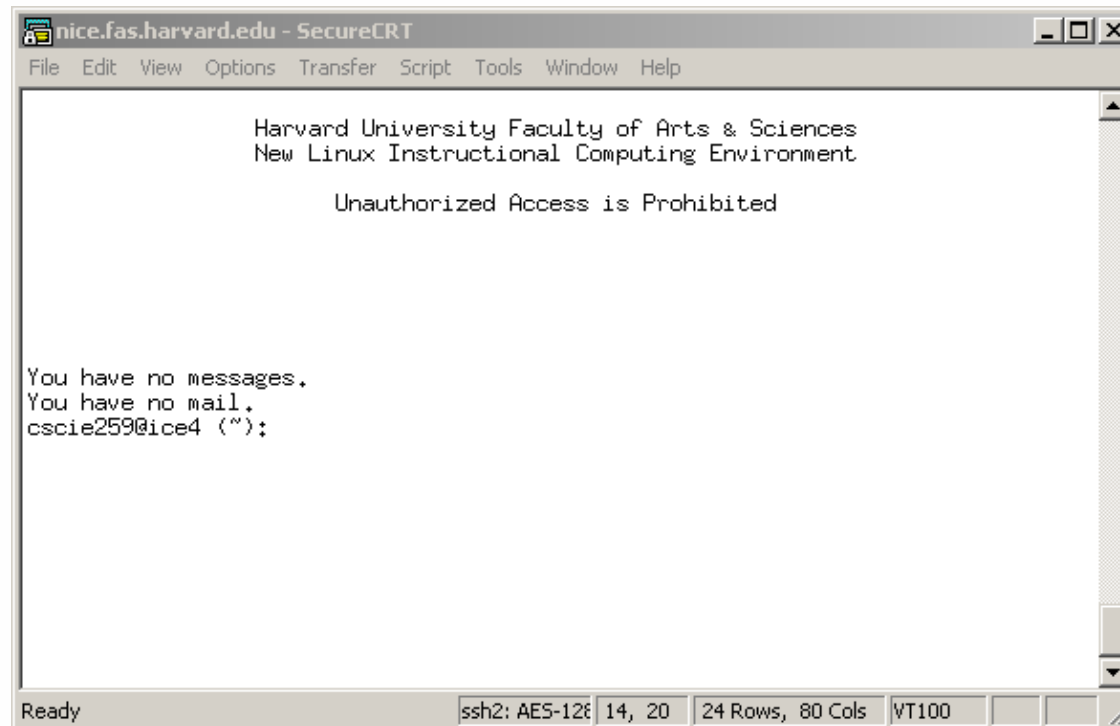
Computer Science E-259

Books

- *Core Servlets and JavaServer Pages, Vol. 1: Core Technologies*, Second Edition
- *Essential XML Quick Reference*
- *XML Pocket Consultant*
- *XSLT Programmer's Reference*, Second Edition

Computer Science E-259

nice.fas.harvard.edu



The image shows a SecureCRT terminal window titled "nice.fas.harvard.edu - SecureCRT". The window has a menu bar with "File", "Edit", "View", "Options", "Transfer", "Script", "Tools", "Window", and "Help". The terminal content displays the following text:

```
Harvard University Faculty of Arts & Sciences  
New Linux Instructional Computing Environment  
  
Unauthorized Access is Prohibited  
  
You have no messages.  
You have no mail.  
cscie259@ice4 (~):
```

The status bar at the bottom of the window shows "Ready", "ssh2: AES-128", "14, 20", "24 Rows, 80 Cols", and "VT100".

Computer Science E-259

Website

<http://www.fas.harvard.edu/~cscie259/>

Computer Science E-259

Staff

`cscie259@fas.harvard.edu`

Computer Science E-259

Listserv

`cscie259@lists.dce.harvard.edu`

Computer Science E-259

Academic Honesty

With the exception of the final project, on which some collaboration is allowed, all work that you do toward fulfillment of this course's expectations must be your own. Viewing or copying another individual's work (even if published in a world-accessible directory) or lifting material from a book, magazine, website, or other source—even in part—and presenting said matter as your own constitutes academic dishonesty, as does showing or giving your work, even in part, to another student.

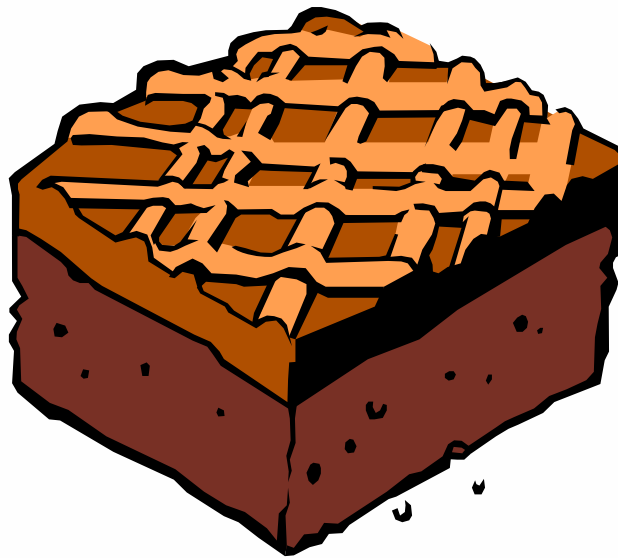
Similarly is dual submission academic dishonesty: you may not submit the same or similar work both to this class and to another. Moreover, submission of any work that you intend to use outside of the course (*e.g.*, at work) must be approved by the staff.

All forms of cheating will be dealt with harshly.

You are welcome to discuss the course's material with others in order to better understand it. You are also welcome to discuss projects' setups with others in order to resolve technical difficulties. But you may not discuss projects' content with other students beyond superficial details. If in doubt as to the appropriateness of some discussion, contact the staff.

Computer Science E-259

Brownie Points



Computer Science E-259

XML with Java

Lecture 1: Introduction

31 January 2007

David J. Malan

`malan@post.harvard.edu`