G52APR CourseWork 2011

Football results system
(100% assessment, in 3 sections)

Mission

• To collect past football (soccer) results from the web.
• To save these in a database
• To query the database for information such as results, statistics etc.
• Purpose:
  – practice programming a larger system;
  – practice using Java libraries;
  – practice programming to interfaces.
• Section details to follow
Section 1 – Data Collection

Key:

- **given**
- **To do**
- **done**

Section 2 – Data Storage
Section 3 – Data Querying

Section 1 – Resources 1

- Main Site
  - http://www.footballresults.org/
- Bonus site
  - http://www.soccerstats.com/
- System.setProperty("proxyHost", "wwwcache-20.cs.nott.ac.uk");
- System.setProperty("proxyPort", "3128");
Section 1 – Resources 2

- For parsing HTML see eg
  org.htmlparser.visitors.NodeVisitor;
- Or
  org.htmlparser.parserapplications.StringExtractor;

Section 1 - hints

- You can parse the HTML how you like, but I found this easy to understand and debug...
  - extract data (ie items between HTML tags) into an array;
  - pattern match the array elements into types;
  - look for sequences of types;
  - extract the information from specific locations in the array relative to the sequences and save.
- Feel free to use other class libraries/methods for screen scraping!
Section 2 – Data Storage

• Fairly straight forward
• Need to design table(s)
• Need to implement SQL queries via JDBC

Section 3 – Data Querying

• Concurrent
• Singleton
• Cache to know which leagues/years have been retrieved
• Ability to fetch new leagues/years “on-the-fly”