Web-based technology for cheminformatics

John Woods

Director of Product Management, Cheminformatics, Oxford Molecular Group
Web technologies...

- HTML / HTTP / forms / Javascript / …
- CGI
- Plug-ins
- Browser-side Java (applets)
- Server-side Java (servlets)
- CORBA, RMI, …
Chemical Database

Web server

LAN, WAN or Internet

Browser
No “best” architecture...

- Depends on database server and web server architecture
- Depends on user needs
Oracle 7 or 8 with RS3 Discovery

Any web server (e.g. Apache)

Netscape or IE 4+
The specific problem: server architecture

- Database is RS3 Discovery / RS3 Discovery HTS
  - Built in Oracle (Oracle 7 and 8 SQL connectivity)
  - Chemical structures (queries and results) in MOL format
  - Links to arbitrary data in Oracle
    - Biological test results
    - Inventory
    - MSDS info
    - …etc
  - Server has built-in query management

- Any web server with servlet support
The specific problem: user needs

- Retrieve simple, fixed “datasheets”
  - All recent assay results
  - Compound safety data
  - …other data from Oracle
  - Configurable by administrator with HTML but no Java

- Query by:
  - structure / substructure
  - compound identifier

- Register single compounds from the browser
  - Business logic to control registration

- IE 4+ / Netscape 4+
Browser-side needs

- Structure rendering
- Query structure specification
- Other data (query constraints, registration info)
Plug in or Java structure renderer

HTML table

HTML forms

Find keyword:

Corros

Search
**Browser-side approach**

- Can use any structure renderer
  - ActiveX / plugins
    - Chemdraw, Isisdraw, Kekule etc
    - What users are used to
    - Needs client-side install
    - Needs different approach for IE and Netscape
  - Java
    - Look and feel are not yet as acceptable to users
    - Slower to load
    - Same approach works for both IE and Netscape
- No "best" answer yet, but easy to change
Browser-side approach

- Communicate with renderer via Molfiles
- Javascript for control
  - IE / Netscape compatibility issues
- HTML forms
  - Molfile in a hidden field
  - Other data / constraints handled directly
- Molfile size would be an issue for Internet
  - Not yet addressed - compression schemes possible
Server-side approach

- CGI lacks power and flexibility
  - Much too difficult to create and maintain pages

- Considered JSP 1.0
  - Java for server-side execution embedded in HTML
  - Too difficult to create / maintain (needs Java programming + HTML skills)
  - JSP 1.1 would be ideal but not ready

- Proprietary solutions e.g. Cold Fusion

- Best available standards-based solution - XML
  - Sun XML parser saved time
  - Harder for users to write than HTML
XML template syntax

- Tags for common functions
  - Connect to database
  - Embed structure display
  - Create and execute substructure query (managed by RS3)
  - Display query results from Oracle/RS3
  - Register structure to Oracle/RS3

- Can include conditional logic (e.g. “display all available assay results for this compound”, business rules for registration)

- Can link to PL/SQL scripts for complex functions
XML example

...<statement name="data" sql="select * from rs3_structure where structure_id = :str_id">
<bind name="str_id" value="$Request.structureId" dir="in" type="string"/>
</statement>

<resultset name="datarset" stmt="data">
<table rules="rows" frame="void" border="1" cellspacing="4" cellpadding="0">
<tr><td><table width="100%">
<tr><td class="header">Registration Date</td></tr>
<tr><td class="data">$datarset.registration_date.value</td></tr>
</table></td></tr>
<tr><td><table width="100%">
<tr><td class="header">Molecular Formula</td></tr>
<tr><td class="data">$datarset.formula.value</td></tr>
</table></td></tr>
<tr><td><table width="100%">
<tr><td class="header">Molecular Weight</td></tr>
<tr><td class="data">$datarset.weight.value</td></tr>
</table></td></tr>
...</
**XML example**

```xml
...  
<tr><td><table width="100%">
<tr><td class="header">Registration Date</td></tr>
<tr><td class="data">
    $dataset.registration_date.value
</td></tr>
</table></td></tr>
...  
```
RS3 Intranet Servlets

- Implemented in Java for any Web server
- Parse XML template files to give HTML
- Connectivity to Oracle
- Provide results from database as requested by tags
  - Communicate with Oracle/RS3 via JDBC
- Interface with RS3 query management
- Connection pooling, security and housekeeping
Web server

JVM

RS3 Intranet Servlets

XML-based templates

Browser

Oracle

RS3
1. Browser requests XML page (may include structure / form query)
2. Web server activates RS3 servlet

Oracle
RS3

Web server
JVM
RS3 Intranet Servlets

XML-based templates

Browser
3. Servlet reads and parses XML template.
4. XML template instructs servlet to query Oracle/RS3
5. Results from Oracle/RS3 merged into template to generate HTML (may include embedded structure display)
Dynamically-created HTML page

6. Browser displays page

Web server

JVM

RS3 Intranet Servlets

XML-based templates

Oracle

RS3

Browser
**Example - XML**

```
...<tr><td class="header">Registration Date</td></tr>
<tr><td class="data">
    $dataset.registration_date.value</td></tr>
</table></td></tr>
...```

```
Example - HTML

...<td> <table width="100%"> <tr> <td class="header">Registration Date</td> </tr> <tr><td class="data">1999-07-13 15:00:21.0</td> </tr> </table> ...
Example - XML

...<displaystruct
molfile="$structRset.object_contents.value"
xsize="400" ysize="400" /></td>

...
Example - HTML

...<object classid="clsid:0C176...
    height="400"
    width="400"
    id="struct92">
</object>
<script type="text/javascript">
document.struct92.hexStructure = "58310A202020...";
</script>
...
...
Demonstration
Register a structure

Search by substructure

Search by Corporate ID

Simple Substructure Query

No Structure

Execute Query
Query Results: 17 hit(s) - Page 1/2

OMG000002
OMG000003
OMG000004
Results to date

- Standards-based intranet interface
- Full chemical display and searching
  - choice of renderer / editor
  - registration also possible
- Easy links to other Oracle data
  - Searching
  - Retrieval
- Pages are configurable by editing XML templates - no programming
Future work

- Commercial product
- Will switch from XML to HTML with JSP 1.1 when standard is fixed and supported
  - will allow templates to be created with standard HTML editors
- Better structure component for browser - Java?
- Molfile compression for Internet use
Conclusions

- No best technology - you must define the problem
- Intranet and Internet impose different design constraints
- XML / JSP work well, relational database platform helps a lot
- Browser-side technology still mixed - need to support multiple combinations for now
- Intranet is a powerful way to rapidly build and deploy chemical database applications
Acknowledgements

- Glenn Atter
- Joe McDaniel
- Tim Regulski
- Steve Sowerby