Yet Another Visualization of Repository Information

Xinrong Xie, Denys Poshvyvanyk, Andrian Marcus

SEVERE Group @ WAYNE STATE UNIVERSITY
Visualization of CVS Repository Information

Xinrong Xie, Denys Poshyvanyk, Andrian Marcus

SEVERE Group @ Wayne State University
A little bit of history

1992 – SeeSoft
Source Viewer 3D - sv3D

IWPC 2003
ICSE 2003
sv3D principles

- Simple and abstract visual metaphors
- Ability to show a lot of data in one view
- Advanced user interaction
- Decouple from analysis
- Allow the user to define the mappings and the semantic - configurability - remember Jens Knodel on Tuesday
Mining software repositories

- Popular topic
- Important issues can be addressed with data from software repositories
- Lots of data being generated
- Need to be presented to the user. Users need to communicate the data to each other
  - Visualization - one solution
Motivation

- Comparison with Seesoft
- Ettore Merlo - IWPC 2003
- Harald Gall - IWPSE 2003
  - “... we plan to add more sophisticated visualization capability to enable viewing the identified relationships with 3-dimensional graphs (e.g. with [17] or [22]).”
CVS data + sv3D = cv3D

- CVS data extraction akin to Zimmerman’s approach
  - No novelty or improvement here
- Visualization decoupled from data extraction
  - Other repository data can be used
- Definition of multiple views - configurability
Questions vs. tasks

1. Who has been working on the artifacts?
2. Which authors work on the same files?
3. What kind of modifications did the author make?
4. When was a modification made?
5. Why was a modification made?
6. What files do frequently change together (co-change)?
7. Which parts of the code are stable?
8. Which parts of the code change frequently?
9. How many authors worked on a release of a system?
10. How many files or lines are added or deleted?
Views

- Granularity-based views
  - System, file (class), function (method), line of text
- The view granularity is given by the lowest level artifact visualized
- Each view helps answer specific questions
- Multiple ways to create the views
Perspectives

- CVS data is time dependent
- Version centric perspectives
  - One revision
  - A pair of revisions
  - Sequence of revisions
System level view

Q: 9-10

H – LOC added
D – LOC deleted x10
File/class level view

Q: 1, 2, 4-8

H: support for co-change

Confidence

- Blue: [0, 0.25)
- Green: [0.25, 0.5)
- Yellow: [0.5, 0.75)
- Orange: [0.75, 1]

WCORE2006
Function level view

Q: 7-8

H: #of LOC changed from v1.114 to v1.115

WCRE2006
Line level view

Q: 3-5, 7, 8
Evaluation

• Usage scenarios (example)

• Exploratory study?

• Case study?
<table>
<thead>
<tr>
<th>Views</th>
<th>Questions</th>
<th>Visual Metaphors</th>
<th>Tool</th>
<th>Mappings and data represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>10</td>
<td>Bar, Data Sheet</td>
<td>ADV</td>
<td># of lines changed</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Chart</td>
<td>EVC</td>
<td># of programs added/deleted</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Matrix</td>
<td>JRX</td>
<td>Area of bubble – number of changes</td>
</tr>
<tr>
<td></td>
<td>9, 10</td>
<td>Bar</td>
<td>CV3</td>
<td></td>
</tr>
<tr>
<td>File/Class</td>
<td>1,2,7,8</td>
<td>Fractal</td>
<td>FRF</td>
<td>Size - # submissions per author</td>
</tr>
<tr>
<td></td>
<td>1,3,4,5,</td>
<td>Graph, Text</td>
<td>XIC</td>
<td>Color – author name; LOC in text; Position of node – ordered by</td>
</tr>
<tr>
<td></td>
<td>7,8,10</td>
<td>Nested Graph</td>
<td></td>
<td>last commit date; CVS comments; Size of node - # of changes</td>
</tr>
<tr>
<td></td>
<td>3,7,8</td>
<td>Bar</td>
<td>JRX</td>
<td>Percentage of bar - # of type of modification</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Matrix</td>
<td>JRX</td>
<td>Show the type of change and time in Matrix</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Table</td>
<td>ROS</td>
<td>Show the sorted co-change files in table</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Graph</td>
<td>CCV</td>
<td>Color – subsystem name, File – filled circles</td>
</tr>
<tr>
<td></td>
<td>7,8</td>
<td>Matrix</td>
<td>EVM</td>
<td># attributes – width of node; # methods – height of node</td>
</tr>
<tr>
<td></td>
<td>4,7,8</td>
<td>DTF</td>
<td>DTF</td>
<td>color – # of revisions/bugs for period of time</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Chart</td>
<td>CHA</td>
<td>Percentage of bar - # of type of modification</td>
</tr>
<tr>
<td></td>
<td>1,2,4-8,10</td>
<td>Bar</td>
<td>CV3</td>
<td></td>
</tr>
<tr>
<td>Function/Method</td>
<td>3,7,8</td>
<td>Bar</td>
<td>JRX</td>
<td>Percentage of bar - # of type of modification</td>
</tr>
<tr>
<td></td>
<td>7,8</td>
<td>Bar</td>
<td>CV3</td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td>3,4,5,7,8</td>
<td>Text, Tooltip,</td>
<td>XIC</td>
<td>Show lines of source code in text; Color – time; Shows CVS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nested Graph</td>
<td></td>
<td>comments and documentation; Size of node - # of changes</td>
</tr>
<tr>
<td></td>
<td>3,7,8</td>
<td>PixelMap</td>
<td>SEE</td>
<td>Color – author last submitted; Color – age of code</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>PixelMap</td>
<td>CSS</td>
<td>Color – change status</td>
</tr>
<tr>
<td></td>
<td>3,4,5,7,8</td>
<td>Bar</td>
<td>CV3</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions and future work

• Multiple views and combination with perspectives helps in answering more questions

• User or usability studies?
  - sv3D already was subject of usability studies for other data/task (IWPC 2005)
  - Does it worth repeating the studies?
Acknowledgements

• Contributors to sv3d:
  - Louis Feng
  - Jonathan Maletic
  - Andrey Sergeyev