HESA

The Construction and Evaluation of Hierarchical Software Feature Repository

Yu Yue
INFLUX Group
National University of Defense Technology
Begin with an example
**Introduction**

1. **Massive number of OSS projects in open source communities**

2. **OSS projects are utilized for industry wildly**

<table>
<thead>
<tr>
<th>Community</th>
<th>Time</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freecode</td>
<td>1998</td>
<td>45,021+</td>
</tr>
<tr>
<td>SourceForge</td>
<td>1999</td>
<td>432,004+</td>
</tr>
<tr>
<td>Rubyforge</td>
<td>2003</td>
<td>9,349+</td>
</tr>
<tr>
<td>Ohloh</td>
<td>2004</td>
<td>500,000+</td>
</tr>
<tr>
<td>Google Code</td>
<td>2006</td>
<td>250,000+</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Actuate09] 4th Actuate Annual Open Source Survey
Stakeholders
What are you seeking for?
Mature Software

large-scale, uncategorized

There are 39.8% uncategorized software in SourceForge and 61.68% in Ohloh
Challenge

Reusable Module and Component

manually, difficultly
Challenge

Reusable Code Snippet

inaccuracy

Google Code Search
Krugle
Codase Source Code Search Engine
Software Feature

capture and identify commonalities and differences in a software domain
Features are any prominent and distinctive concepts or characteristics that are visible to various stakeholders.

a) Features are externally visible characteristics that can differentiate one product from the others.

b) Functions, objects, and aspects have been mainly used to specify the internal details of a system.
Social Feature
Recent releases

6.6.0 17 Sep 2012 20:57
Release Notes: This release adds an Outlook Add-in for Contacts and Calendar exchange syncing between Outlook and Atmail Server using ActiveSync.

6.5.0 17 Sep 2012 20:57
Release Notes: This release adds online file storage with sharing and a mobile-optimized UI.

6.3.5 27 Feb 2012 22:50
Release Notes: This is a minor release. It corrects uninitialized array usage in the dashboard controller during graph calculation. It corrects an unhandled exception in logsearch when...

(continued)

Norman Malware Cleaner description

Show Product Description

Here are some key features of "Norman Malware Cleaner":
- Easy to install and run
- Detect and Remove malware (viruses, Rootkit's, FakeAV, worms and more)
- Utilize advanced Anti-Roottkit technology
- Quarantine module to process the detected files
- Deep scan and cleaning including Norman patented Norman SandBox technology
- Supports Quick- and Deep Scan mode
- New command line function for better tailor scanning across several machines (businesses)
- Daily signature updates available

Description

Rigs of Rods is a 3D simulator game where you can drive, fly and sail various vehicles using an accurate and unique soft-body physics engine.

Rigs of Rods Web Site

Categories
Simulation, Simulations

License
GNU General Public License version 3.0 (GPLv3)

Features
- Rigid body physics simulation
- Can simulate cars, trucks, airplanes and boats and everything in between

What is GitX?

GitX is a git like clone written specifically for OS X Leopard and higher. This means that it has a native interface and tries to integrate with the operating system as good as possible. Examples of this are drag and drop support and Quicklook support.

Features

The project is currently still in its starting phases. As time goes on, hopefully more features will be added. Currently GitX supports the following:
- History browsing of your repository
- See a nicely formatted diff of any revision
- Browse based on author or revision subject
- Look at the complete tree of any revision
- Provide any file in the tree in a text view or with Quicklook
- Drag and drop files out of the tree view to copy them to your system
- Support for all parameters set in the .gitignore

License

GitX is licensed under the GNU version 2. For more information, see the attached COPYING file.

Download

GitX is currently hosted at GitHub. The project page can be found at http://github.com/pister/gitx. Recent binary releases can be found at http://github.com/pister/gitx/releases.
1. Synonymic Feature Element:

Kills the core of AdPower and not only symptoms;
Kills the core of Adroar and not only symptoms;
Kills the core of BANCOS.B and not only symptoms;

Ability to update that does not require downloading full package;
Incremental database updates and often to include information about latest threats;
2. Hybrid Semantic-level:

- Email Scanner enhanced email protection:
  - Email scanning for Microsoft Outlook, Outlook Express, Mozilla Thunderbird, Windows Live Mail, Windows Mail, and other POP3/IMAP mail clients, ensuring your email is free of viruses and other threats;

- Blocks spam mails, phishing attack mails, junk mails and porn mails before they reach your inbox;
Seeking Life’s Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here, two genome researchers with radically different approaches presented complementary views of the basic genes needed for life.

One research team, using computer analyses to compare known genomes, concluded that today’s organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn’t be enough.

Although the numbers don’t match precisely, those predictions are not all that far apart, especially in comparison to the 75,000 genes in the human genome, note Stan Anderson and James P. Lake of the University of Wisconsin. "One genome may not be enough," says Anderson, "but, coming up with a consensus number may be more than just a guess." As more and more genomes are completely mapped and sequenced, "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing all the genomes could "yield an estimate of the minimum modern and ancient genomes."
**ID1000**: Real-time windows registry monitor utility.
1. Email Scanner enhanced email protection;

2. Email scanning for Microsoft Outlook, Outlook Express, Mozilla Thunderbird, Windows Live Mail, Windows Mail, and other POP3/IMAP mail clients, ensuring your email is free of viruses and other threats;

3. Blocks spam mails, phishing attack mails, junk mails and porn mails before they reach your inbox;
Overview
### Evaluation

#### Table I
PREPROCESSED EXPERIMENT DATASETS

<table>
<thead>
<tr>
<th>Category</th>
<th>#Softpedia</th>
<th>#Sourceforge</th>
<th>#Total</th>
<th>#Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antivirus</td>
<td>2919</td>
<td>1105</td>
<td>4024</td>
<td>50</td>
</tr>
<tr>
<td>Audio-Player</td>
<td>3714</td>
<td>1283</td>
<td>4997</td>
<td>60</td>
</tr>
<tr>
<td>File-Manager</td>
<td>2270</td>
<td>970</td>
<td>3240</td>
<td>40</td>
</tr>
</tbody>
</table>

#### Table II
EVALUATION OF FEATURE-ONTOLGY QUALITY

<table>
<thead>
<tr>
<th>Category</th>
<th>Score-3</th>
<th>Score-2</th>
<th>Score-1</th>
<th>Likert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antivirus</td>
<td>33.3%</td>
<td>50.0%</td>
<td>16.7%</td>
<td>2.17</td>
</tr>
<tr>
<td>Audio-Player</td>
<td>39.1%</td>
<td>46.3%</td>
<td>14.6%</td>
<td>2.25</td>
</tr>
<tr>
<td>File-Manager</td>
<td>32.7%</td>
<td>52.4%</td>
<td>14.9%</td>
<td>2.18</td>
</tr>
<tr>
<td>Average</td>
<td>35.03%</td>
<td>49.57%</td>
<td>15.4%</td>
<td>2.20</td>
</tr>
</tbody>
</table>
Evaluation

(a) the average value

(b) the standard deviation

The clustering results for each category
Thank You!

Yu Yue
INFLUX Group
National University of Defense Technology