Introduction of Trustie Software Repository & Passion-Lab Data Center

Meng Li & Minghui Zhou
Peking University, Beijing, China

Software Institute, School of Electronics Engineering and Computer Science, Peking University
Key Laboratory of High Confidence Software Technologies, Ministry of Education
Agenda

• Introduction of TSR
• Trustworthiness Evaluation in TSR
• Passion-Lab Data Center
Introduction of TSR

• Background
  – **Software reuse** aims to improve software development via reusing software resources (SR).
  – *Typical SR*: components, Web services, tools, architectures, etc.
  – **Software resource repository** is the infrastructure that provides the SR management mechanism to support software reuse.
  – *Such as publishing, retrieving, classification, storage, feedback, evaluation, etc.*
Introduction of TSR

• About TSR
  – **Trustie Software Resource Repository (TSR)** is to provide a mechanism to describe, collect, evaluate, classify and manage SR’s trustworthiness, to support *trust software development*.
  – TSR is part of **Trustie Project** (Major 863 project, China)
  – Trustie Project: [http://www.trustie.net/](http://www.trustie.net/)
  – TSR: [http://tsr.trustie.net/](http://tsr.trustie.net/)
  – TSR is developed by *Peking University*
Introduction of TSR

• Functionalities of TSR
  – Resource management
  – *Publishing/editing/downloading/deleting resources*
  – *Retrieving resources*
  – *Recommendating resources*
  – User management
    – *Registration, sign in/out, etc.*
  – Feedback
    – *Rating or comment*
    – *Quality template-based feedbacks*
  – Tag management
  – Trustworthiness evaluation
  – Statistics
Introduction of TSR

- TSR Home Page
Introduction of TSR

• Architecture of TSR
Introduction of TSR

• Application of TSR
  – TSR has been deployed in several Software Incubators & Companies throughout China.
  – Installation for OW2 in Grenoble, France (in Nov. 2012).
Agenda

• Introduction of TSR
• Trustworthiness Evaluation in TSR
• Passion-Lab Data Center
Trustworthiness Evaluation in TSR

- Trustworthiness Evaluation in TSR
Trustworthiness Evaluation in TSR

- Trustworthiness Evaluation Methods
  - Feedback-Based Trustworthiness Evaluation (FBTE)
  - Internet-Based Trustworthiness Evaluation (IBTE)
Feedback-Based Trustworthiness Evaluation (FBTE)

• Rationales behind FBTE:
  – When I want to find and reuse a trustworthy SR, the most straightforward way is to choose SRs that are identified trustworthy by other users.
  – In other words, users’ feedbacks are an important kind of trustworthy evidences.
  – Moreover, if the user is trustworthy (e.g. expert), his feedbacks is more likely to be trustworthy.

• We collect users’ feedbacks in TSR, and provide FBTE to help users to find and reuse SRs.
Feedback-Based Trustworthiness Evaluation (FBTE)

- FBTE is based on the *feedback functionality* in TSR
  - Simple feedback: rating & comment
  - Template-based feedback: detailed feedback (stability, security, portability, etc.)
Feedback-Based Trustworthiness Evaluation (FBTE)

• FBTE Procedure:
  a) TSR admin assign trustworthiness evaluation experts;
  b) Experts evaluate the trustworthiness of a SR by providing template-based feedback;
  c) TSR admin confirm the feedbacks, and the latest overall rating is taken as SR’s trustworthiness level.

• FBTE is straightforward but widely used in software repositories/portals.
Internet-Based Trustworthiness Evaluation (IBTE)

• Background
  – SRs are diversifying
  – *From closed, static, code*
  – *To open, dynamic, service*
  – With the development of Web 2.0, there are more and more information on the Internet
  – *Feedbacks on different websites*
    • E.g. comments and ratings on Seekda, Service-finder

• We try to collect trustworthy evidences from the Internet and evaluate the trustworthiness of SRs in TSR with them.
Internet-Based Trustworthiness Evaluation (IBTE)

• Framework Overview

[Diagram showing the framework of IBTE with key components such as Trustworthiness Evaluation Model, SR List, Collecting Related Info from Internet, Trustworthy Evidence Extraction, and Evidence Storage.]
Internet-Based Trustworthiness Evaluation (IBTE)

• Taking Web Services for example.

Add Criterion

E.g. Is Availability > 80% ?

Is Response Time < 3000ms ?  Add

Customize Trustworthiness Evaluation Model

Evaluation Model

Level 1
Is availability > 50 %?
Add Criterion

Level 2
Is availability > 90 %?

Is response time < 5000ms?
Add Criterion

Level 3
Is availability > 90 %?

Is response time < 3000ms?  Add

Trustworthiness Information

Trustworthiness Evaluation

Evidences
Internet-Based Trustworthiness Evaluation (IBTE)

• Collecting trustworthy evidences from the Internet
  – Objective evidences (QoS)
    – *Implement a QoS monitor, invoke Web services on a regular basis.*
    – Over 2 million records
  – Subjective evidences (Reputation)
    – *Calculated with users’ feedbacks including ratings and comments*

<table>
<thead>
<tr>
<th>NO.</th>
<th>Comment</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>It works perfectly.</td>
<td>Positive (+1)</td>
</tr>
<tr>
<td>2</td>
<td>Cool service!</td>
<td>Positive (+1)</td>
</tr>
<tr>
<td>3</td>
<td>No data returned</td>
<td>Negative (-1)</td>
</tr>
<tr>
<td>4</td>
<td>Gave a wrong result (&quot;false&quot;) on a valid email address. Useless.</td>
<td>Negative (-1)</td>
</tr>
<tr>
<td>5</td>
<td>It’s not working.</td>
<td>Negative (-1)</td>
</tr>
</tbody>
</table>
# Internet-Based Trustworthiness Evaluation (IBTE)

- **Demo**

## Web Service Trust Classification Evidences

<table>
<thead>
<tr>
<th>Date</th>
<th>Evidence Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-01-02 21:16</td>
<td>是否有WSDL地址信息？</td>
<td>Yes</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>Web服务是否可用</td>
<td>Yes</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>Web服务可用性是否达到80%以上？</td>
<td>Yes</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>是否有在线测试工具</td>
<td>Yes</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>是否有功能描述信息</td>
<td>Yes</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>Web服务可用性是否达到90%以上？</td>
<td>No</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>是否有客户案例</td>
<td>Yes</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>是否有客户评估报告</td>
<td>Yes</td>
</tr>
<tr>
<td>12-01-02 21:16</td>
<td>是否有第三方提供的可靠评估</td>
<td>No</td>
</tr>
</tbody>
</table>

## Web Service QoS Info

- **Is Available?** Yes
- **Availability:** 0.98055737770418
- **Response Time:** 1219 ms
- **Calculated Date:** 12-01-02 21:16

## Web Service Feedbacks

<table>
<thead>
<tr>
<th>Date</th>
<th>Feedback</th>
<th>Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-02-23 22:32</td>
<td>It works fine. Resolves the ips correctly.</td>
<td>Positive</td>
</tr>
<tr>
<td>09-12-03 22:32</td>
<td>Service returning wrong city</td>
<td>Negative</td>
</tr>
<tr>
<td>09-12-03 22:32</td>
<td>great</td>
<td>Positive</td>
</tr>
<tr>
<td>09-12-03 22:32</td>
<td>Great</td>
<td>Positive</td>
</tr>
<tr>
<td>09-12-03 22:32</td>
<td>&quot;Test&quot; volume not really sufficient for development purposes. :-(</td>
<td>Negative</td>
</tr>
<tr>
<td>09-12-03 22:32</td>
<td>bad service</td>
<td>Negative</td>
</tr>
</tbody>
</table>
Agenda

- Introduction of TSR
- Trustworthiness Evaluation in TSR
- Passion-Lab Data Center
What are the best practices in OSS?

• Research question:
  – How people reuse code over time?
  – Which code is reused most often?
  – What attributes they have?

• Build infrastructure to review code reuse across OSS universe
Infrastructure of Big Data

Data Center

Passion-Lab
Software Engineering Institute
Peking University
Infrastructure

• We keep tracking various commercial and open source projects.

• This “universal” repository records data from:
  – Version control
  – Issue tracking
  – Email archives
  – ……. 

<table>
<thead>
<tr>
<th>Forge</th>
<th>Type</th>
<th>Files</th>
<th>File/Ver.</th>
<th>Unique File/Ver.</th>
<th>Branching</th>
<th>From</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large cmpny.</td>
<td>Var.</td>
<td>3,272K</td>
<td>12,585K</td>
<td>4,293K</td>
<td>2.9</td>
<td>1988</td>
</tr>
<tr>
<td>SourceForge</td>
<td>CVS</td>
<td>26,095K</td>
<td>81,239K</td>
<td>39,550K</td>
<td>2.1</td>
<td>1998</td>
</tr>
<tr>
<td>code.google</td>
<td>SVN</td>
<td>5,675K</td>
<td>14,368K</td>
<td>8,584K</td>
<td>1.7</td>
<td>1996</td>
</tr>
<tr>
<td>repo.or.cz</td>
<td>Git</td>
<td>2,519K</td>
<td>11,068K</td>
<td>5,115K</td>
<td>2.2</td>
<td>1986</td>
</tr>
<tr>
<td>Savannah</td>
<td>CVS</td>
<td>852K</td>
<td>3,623K</td>
<td>2,345K</td>
<td>1.5</td>
<td>1985</td>
</tr>
<tr>
<td>git.kernel.org</td>
<td>Git</td>
<td>12,974K</td>
<td>97,585K</td>
<td>856K</td>
<td>114</td>
<td>1988</td>
</tr>
<tr>
<td>OpenSolaris</td>
<td>Hg</td>
<td>77K</td>
<td>1,108K</td>
<td>91K</td>
<td>12.2</td>
<td>2003</td>
</tr>
<tr>
<td>FreeBSD</td>
<td>CVS</td>
<td>196K</td>
<td>360K</td>
<td>75K</td>
<td>4.8</td>
<td>1993</td>
</tr>
<tr>
<td>Kde</td>
<td>SVN</td>
<td>2,645K</td>
<td>10,162K</td>
<td>527K</td>
<td>19.3</td>
<td>1997</td>
</tr>
<tr>
<td>gnome.org</td>
<td>SVN</td>
<td>1,284K</td>
<td>3,981K</td>
<td>1,412K</td>
<td>2.8</td>
<td>1997</td>
</tr>
<tr>
<td>Gcc</td>
<td>SVN</td>
<td>3,758K</td>
<td>4,803K</td>
<td>395K</td>
<td>12.2</td>
<td>1989</td>
</tr>
<tr>
<td>Eclipse</td>
<td>CVS</td>
<td>729K</td>
<td>2,127K</td>
<td>575K</td>
<td>3.7</td>
<td>2001</td>
</tr>
<tr>
<td>OpenJDK</td>
<td>Hg</td>
<td>32K</td>
<td>747K</td>
<td>60K</td>
<td>12.4</td>
<td>2008</td>
</tr>
</tbody>
</table>
Hardware and data levels

- **Machines**
  - DELL R910(4U), 64GbRAM, 16-cores X7550
  - DELL MD3200, 12*2TB SAS
  - DELL R710 * 4, 64GbRAM

- **Data Levels**
  - Level0: raw data
  - Level1: filtered data
  - Level2-n: customized data
What we could do with this data?

• To enable better eco-system, user experience, and practices … …
  – Understand the past, predict the future

HTTP://passion-lab.org

A cloud for OSS best practices based on issue tracking data, version control data and …
Related Websites

- Trustie Software Repository
  http://tsr.trustie.net

- CoWS Web Services Search Engine
  http://www.cowebservices.com

- Trustworthiness Evaluation
  http://pingji.trustie.net

- API Example
  http://www.apiexample.com

- Passion-Lab
  http://passion-lab.org

- TSR @ OW2
  http://tsr.ow2.org/

All the systems are Open Source
Thank You!
Merci!
谢谢！