A Dataset of High Impact Bugs: Manually-Classified Issue Reports

Masao Ohira¹, Yutaro Kashiwa¹, Yosuke Yamatani¹, Hayato Yoshiyuki¹, Yoshiya Maeda¹, Nachai Limsettho², Keisuke Fujino², Hideaki Hata², Akinori Ihara² and Kenichi Matsumoto²

Background

A great number of bugs are found not only before but also after releasing products. A manager must prioritize reported bugs to be fixed. Many studies proposed methods to support bug-triaging. However, in these studies, each bug is equally treated without considering its impact on the bug management process and software products.

High impact bugs

Process bug

A process bug is a bug that affects a bug management process.

- **Blocking bug** [Valdivia¹¹ etc.]
  A bug that blocks other bugs from being fixed

- **Surprising bug** [Shihab¹¹]
  A bug that appears in unexpected timing and location

- **Dormant bug** [Chen¹⁴]
  A bug that was introduced in one version of a system but was not reported until the next immediate version

Product bug

A product bug is a bug that affects the quality of software products.

- **Security bug** [Gegick¹⁰ etc.]
  A bug that has security risks

- **Performance bug** [Nistor¹³]
  A bug that makes system’s performance degrade

- **Breakage bug** [Shihab¹¹]
  A functional bug that was made by code modifications to add new features or to fix bugs

Dataset

Issue report data has been collected from the Apache Ambari, Camel, Derby, and Wicket projects which JIRA is used for managing reported issues.

<table>
<thead>
<tr>
<th>All the issues in Nov. 2014</th>
<th>Our Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambari</td>
<td>BUG</td>
</tr>
<tr>
<td>8,389</td>
<td>871</td>
</tr>
<tr>
<td>Camel</td>
<td>8,063</td>
</tr>
<tr>
<td>Derby</td>
<td>6,772</td>
</tr>
<tr>
<td>Wicket</td>
<td>5,769</td>
</tr>
</tbody>
</table>

Manual Classification: Results

Reference


Dataset: http://goo.gl/r53j7w

Contact: masao@sys.wakayama-u.ac.jp