Toward Mining “Concept Keywords” from Identifiers in Large Software Projects

Masaru Ohba
and
Katsuhiko Gondow
Tokyo Institute of Technology
What are “concept keywords”? 

- Most programmers try to name identifiers meaningfully.
- Concept keywords are defined terms that describe key concepts to aid in understanding. 
  - e.g. `readDirent()` : `dirent` is a concept keyword.

<table>
<thead>
<tr>
<th>Concept keywords</th>
<th><code>dirent</code>, <code>root</code>, <code>PTE</code>, <code>tss</code>, <code>path</code>, <code>signal</code>, <code>yield</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grouping words</td>
<td><code>kbd</code>, <code>vga</code>, <code>FAT12</code>, <code>sys</code>, <code>H</code>, <code>t</code></td>
</tr>
<tr>
<td>Attributes, less important concepts</td>
<td><code>busy</code>, <code>byte</code>, <code>offset</code>, <code>name</code>, <code>memory</code>, <code>end</code>, <code>int8</code>, <code>again</code></td>
</tr>
<tr>
<td>Generic verbs</td>
<td><code>read</code>, <code>set</code>, <code>is</code>, <code>move</code>, <code>wait</code>, <code>print</code>, <code>dump</code>, <code>make</code>, <code>init</code></td>
</tr>
</tbody>
</table>

Human-selected concept keywords and other category words in udos.
Suggestion

• We should use more “concept keywords” in program understanding tools.
  – concept keywords are concise and descriptive

• Our solution:
  – provides a way to mine concept keywords.
    • ckTF/IDF methods / Identifier Exploratory Framework
  – could be used to build tools that support and utilize extracted concept keywords (future work).
Future work

- Applying concept keywords to a Bug Tracking System (BTS) to see the relationship between bug report and corresponding problem source code.

Bug-report no.1
Overview:
It could not read directories.

Bug-report no.3
Overview:
I could not catch system calls.

Concept keyword can bridge the gap between bug-reports and source code.