

**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 as program understanding.
 - e.g. `read_dirent()` : `dirent` is a concept keyword.

Concept keywords	<code>dirent</code> , <code>root</code> , <code>PTE</code> , <code>tss</code> , <code>path</code> , <code>signal</code> , <code>yield</code>
Grouping words	<code>kbd</code> , <code>vga</code> , <code>FAT12</code> , <code>svs</code> , <code>H.t</code>
Attributes, less important concepts	<code>busy</code> , <code>byte</code> , <code>offset</code> , <code>name</code> , <code>memory</code> , <code>end</code> , <code>int8</code> , <code>again</code>
Generic verbs	<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>

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.

