

An Approach to Detecting Duplicate Bug Reports using Natural Language and Execution Information

Xiaoyin Wang¹, Lu Zhang^{1*}, Tao Xie^{2*}, John Anvik³ and Jiasu Sun¹

¹Key laboratory of High Confidence Software Technologies, Ministry of Education, Institute of Software, EECS, Peking University, Beijing, 100871, P. R. China, {wangxy06, zhanglu, sjs}@sei.pku.edu.cn

²Department of Computer Science, North Carolina State University, Raleigh, NC 27695, USA
xie@csc.ncsu.edu

³Department of Computer Science, University of Victoria, PO Box 3055, STN CSC, Victoria, BC, Canada, V8W 3P6
janvik@cs.uvic.ca

*

Bug-260331 Bug-239223
Bug-260331
Bug-239223

Bug-260331: After closing Firefox, the process is still running. Cannot reopen Firefox after that, unless the previous process is killed manually

Bug-239223: (Ghostproc) – [Meta] firefox.exe doesn't always exit after closing all windows; session-specific data retained

-
-
-
-

Bug-244372 Bug-219232

Bug-244372: "Document contains no data" message on continuation page of NY Times article

Bug-219232: random "The Document contains no data." Alerts

Bug-244372

Bug-219232
Bug-219232



$$w_i = \frac{tf_i}{\sum_{i=1}^n tf_i} \times \frac{idf_i}{\sum_{i=1}^n idf_i} \quad 1 \leq i \leq n$$

$$w_i = \frac{tf_i}{\sum_{i=1}^n tf_i} \times \frac{idf_i}{\sum_{i=1}^n idf_i}$$

tf_i term frequency idf_i
 inverse document frequency tf_i
 i

$$idf_i = \log \left(\frac{D_{sum}}{Dw_i} \right)$$

D_{sum} i Dw_i

■

■

■

4.3.1 Basic Heuristic

$$SIM_{combined} = f(SIM_{nlp}, SIM_{exe})$$
$$SIM_{combined} = \frac{SIM_{nlp} + SIM_{exe}}{2}$$

$$SIM_{combined} = \frac{SIM_{nlp} + SIM_{exe}}{2}$$

4.3.2 Classification-Based Heuristic

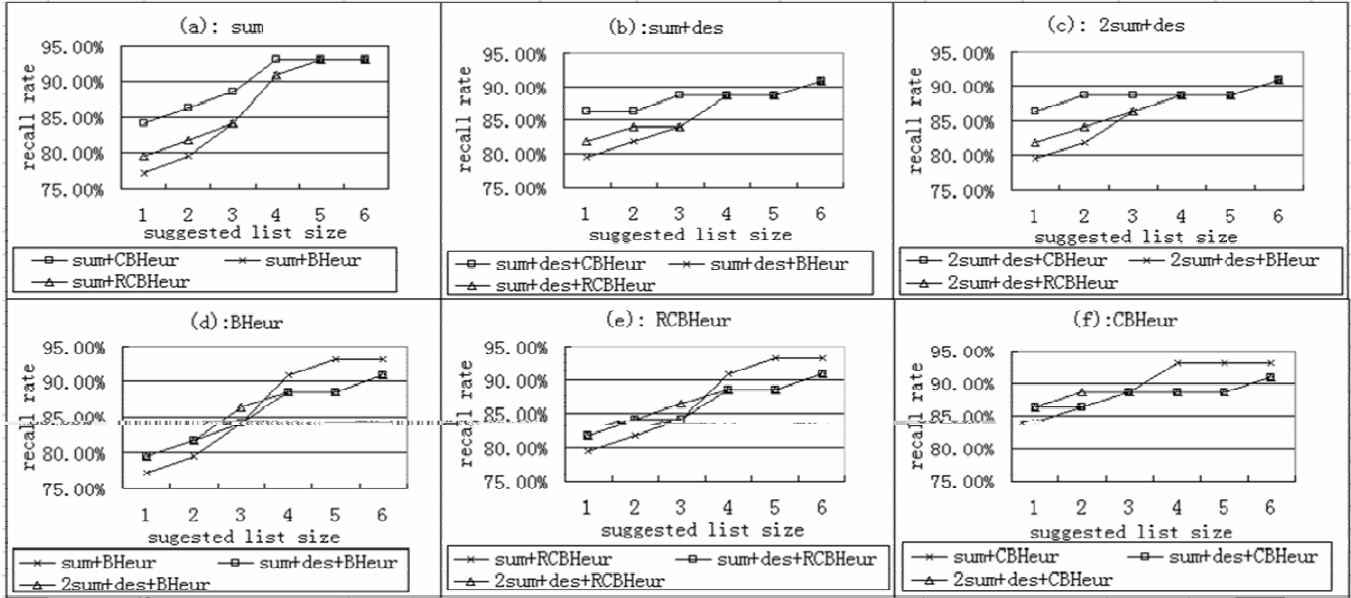
■

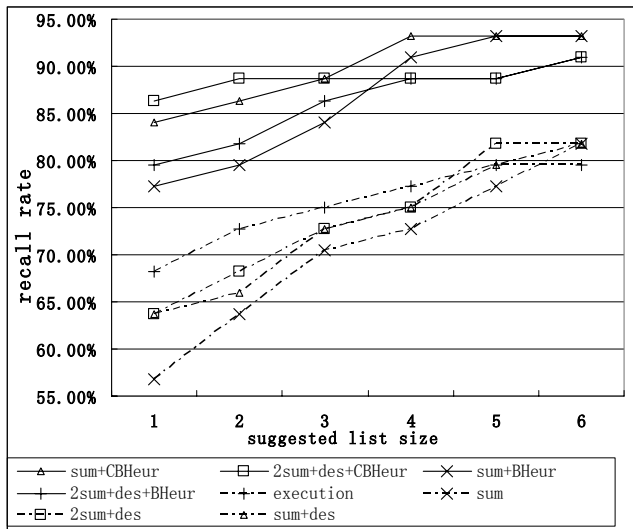
$$\text{recall rate} = \frac{N_{\text{recalled}}}{N_{\text{total}}}$$

•

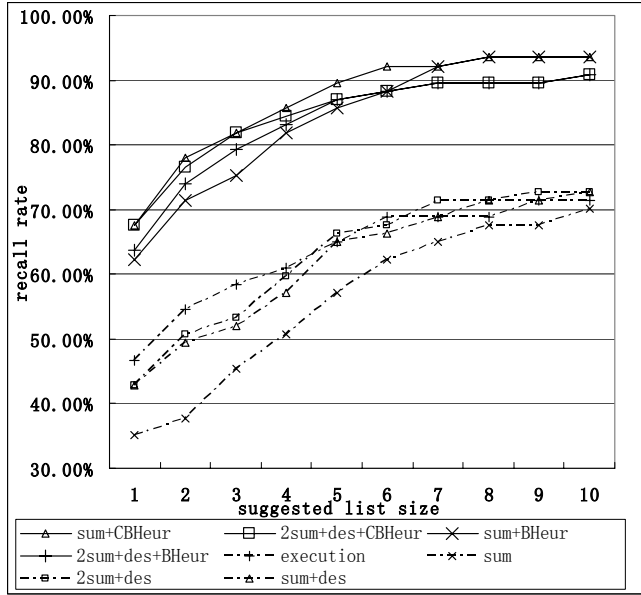
•

•





inverse duplicate problem



idf

idf

In Proc. ICSE.

*In Proc. of OOPSLA Workshop on Eclipse
Technology eXchange (ETX)*

In Proc. ICSE

In Proc. SEKE

IEEE TSE

In Proc. ISSRE

Assisted Detection of Duplicate Bug Reports

In Proc. ISSTA

*In
Proc. of IEEE Conf. on Visual Language and Human-Centric
Computing (VL/HCC)*

In Proc. PLDI

In Proc. ICSM,

ACM TOSEM

In Proc. ICSE

*Journal of the American So-
ciety for Information Science*

Proc. ICSE

In

In Proc. ICSE,

In Proc. MSR,