



论 文 汇 报

# 2021 本周论文汇报

汇报人：王昭丹



## 5 Papers

Title	Publication source	Year
Network-Clustered Multi-Modal Bug Localization	TSE	2019
Pathidea: Improving Information Retrieval-Based Bug Localization by Re-Constructing Execution Paths Using Logs	TSE	2021
Tracking Buggy Files: New Efficient Adaptive Bug Localization Algorithm	TSE	2021

# Network-Clustered Multi-Modal Bug Localization — TSE 2019

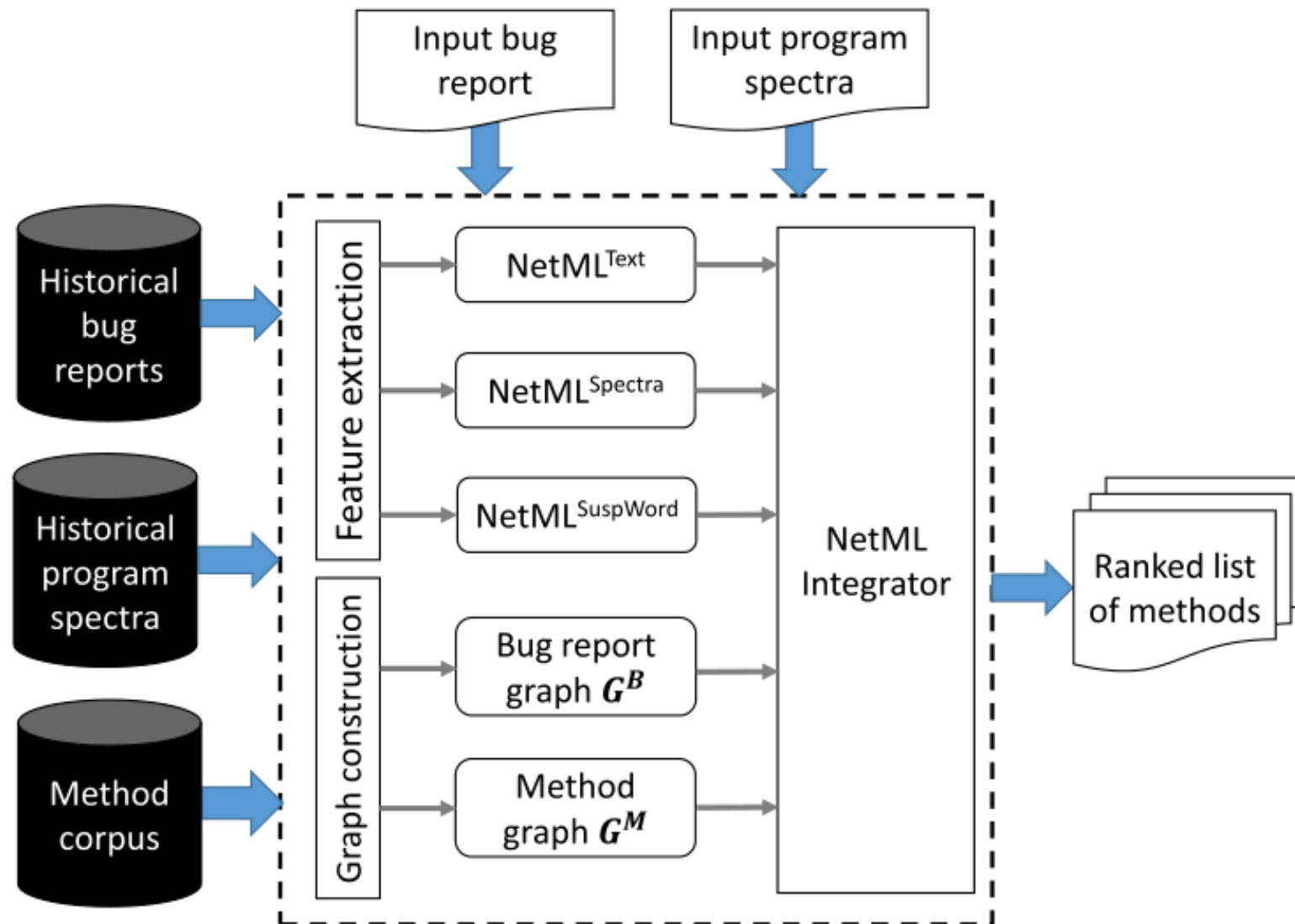


Fig. 2. The proposed NetML framework.

# Pathidea: Improving Information Retrieval-Based Bug Localization by Re-Constructing Execution Paths Using Logs

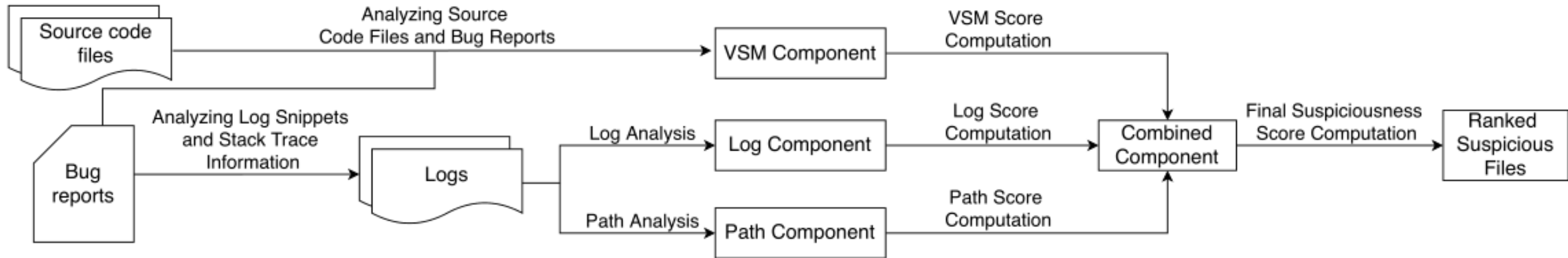


Fig. 3: An overview of Pathidea.

## Logs

```
2008-01-07 21:02:13 INFO org.apache.hadoop.mapred.ReduceTask: task_r_1 done copying task_m_0
2008-01-07 21:02:13 INFO org.apache.hadoop.mapred.ReduceTask: task_r_1 Copying task_m_1
...
2008-01-07 21:02:13 WARN org.apache.hadoop.mapred.ReduceTask: java.lang.OutOfMemoryError: Java heap space
    at org.apache.hadoop.io.SequenceFile$Reader.init (SequenceFile.java:1345)
    at org.apache.hadoop.mapred.ReduceTask.run (ReduceTask.java:1311)
2008-01-07 21:02:31 ERROR org.apache.hadoop.mapred.ReduceTask: java.lang.NullPointerException: Map output copy failure
    at org.apache.hadoop.fs.InMemoryFileSystem.close (InMemoryFileSystem.java:378)
    at org.apache.hadoop.fs.FileSystem.getLength (FileSystem.java:449)
    at org.apache.hadoop.mapred.ReduceTask.run (ReduceTask.java:665)
```

# Tracking Buggy Files: New Efficient Adaptive Bug Localization Algorithm

TABLE 2

Features used in ranking model - proposed by Ye et al.+ [12]. **Notation:**  $sim$  is the cosine similarity. *Query dependent* features relay on both the source code  $s$  and on the bug report  $r$ . We use new notation for features  $\phi_2$ ,  $\phi_3$ ,  $\phi_5$ . We propose  $\phi_2^*$  instead of  $\phi_2$ , see Section 3.1.1 for rationale.

Feature	Description	Formula	Query dep?
$\phi_1$	Surface lexical similarity	$\phi_1(r, s) = \max(\{sim(r, s)\} \cup \{sim(r, m) \mid m \in s\})$	Yes
$\phi_2$	API-enriched lexical similarity	$\phi_2(r, s) = \max(\{sim(r, s.api)\} \cup \{sim(r, m.api) \mid m \in s\})$	Yes
$\phi_3$	Collaborative filtering score	$\phi_3(r, s) = sim(r, \text{concat}(\{r.summary \mid r \in br(r, s)\}))$	Yes
$\phi_4$	Class name similarity	$\phi_4(r, s) =  s.main\_class  \cdot \mathbb{1}[s.main\_class \in s.summary]$	Yes
$\phi_5$	Bug-fixing recency	$\phi_5(r, s) = ((r.date - last(r, s).date).months + 1)^{-1}$	Yes (Timestamp)
$\phi_6$	Bug-fixing frequency	$\phi_6(r, s) =  br(r, s) $	Yes (Timestamp)
$\phi_7$	Summary-class names similarity	$\phi_7(r, s) = sim(r.summary, s.class)$	Yes
$\phi_8$	Summary-method names similarity	$\phi_8(r, s) = sim(r.summary, s.method)$	Yes
$\phi_9$	Summary-variable names similarity	$\phi_9(r, s) = sim(r.summary, s.variable)$	Yes
$\phi_{10}$	Summary-comments similarity	$\phi_{10}(r, s) = sim(r.summary, s.comment)$	Yes
$\phi_{11}$	Description-class names similarity	$\phi_{11}(r, s) = sim(r.description, s.class)$	Yes
$\phi_{12}$	Description-method names similarity	$\phi_{12}(r, s) = sim(r.description, s.method)$	Yes
$\phi_{13}$	Description-variable names similarity	$\phi_{13}(r, s) = sim(r.description, s.variable)$	Yes
$\phi_{14}$	Description-comments similarity	$\phi_{14}(r, s) = sim(r.description, s.comment)$	Yes
$\phi_{15}$	In-links = # of file dependencies of $s$	$\phi_{15}(r, s) = s.inLinks$	No
$\phi_{16}$	Out-links = # of files that depend on $s$	$\phi_{16}(r, s) = s.outLinks$	No
$\phi_{17}$	PageRank score	$\phi_{17}(r, s) = PageRank(s)$	No
$\phi_{18}$	Authority score (HITS)	$\phi_{18}(r, s) = Authority(s)$	No
$\phi_{19}$	Hub score (HITS)	$\phi_{19}(r, s) = Hub(s)$	No
$\phi_2^*$	full API-enriched lexical similarity	$\phi_2^*(r, s) = \max(\{sim(r, s.api^*)\} \cup \{sim(r, m.api^*) \mid m \in s\})$	Yes



论 文 汇 报

# 感谢您的聆听

汇报人：王昭丹



