Text Mining for Software Engineering: How Analyst Feedback Impacts Final Results

Jane Huffman Hayes,
Alex Dekhtyar,
Senthil Karthikeyan Sundaram

*Funded by NASA

Department of Computer Science
University of Kentucky
Question of the Day

What can Data Mining Do for Software Engineering ???
Question of the Day

What can Data Mining Do for Software Engineering ???

Answer 1

Help study the process

- After-the-fact
- Exploratory
- Conclusions help future projects
Question of the Day

Answer 1
Help study the process

Answer 2
Help improve the process

What can Data Mining Do for Software Engineering ???

- After-the-fact
- Exploratory
- Conclusions help future projects
Our Approach

Use Data Mining during the process
Use Mining During the Process?

Ultimately, we are interested in the accuracy of the final result.

Automated “Mining” Tool

Objective Study (RE’04, PROMISE’05)

Subjective Study

Feedback Loop

Task

Analyst

Final Result
Preliminary Study

**Question:** What would the analyst do with machine-generated data?

**Task:** Requirements Tracing

**Metrics:** Precision, Recall

Final Result

Automated “Mining” Tool

Analyst
Question: What would the analyst do with machine-generated data?

<table>
<thead>
<tr>
<th>Pr</th>
<th>Rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>20%</td>
<td>90%</td>
</tr>
<tr>
<td>80%</td>
<td>30%</td>
</tr>
</tbody>
</table>
Preliminary Study

**Question:** What would the analyst do with machine-generated data?

<table>
<thead>
<tr>
<th>Pr</th>
<th>Rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>20%</td>
<td>90%</td>
</tr>
<tr>
<td>80%</td>
<td>30%</td>
</tr>
</tbody>
</table>

Candidate link lists

<table>
<thead>
<tr>
<th>Pr</th>
<th>Rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>45%</td>
<td>56%</td>
</tr>
<tr>
<td>58%</td>
<td>65%</td>
</tr>
<tr>
<td>23%</td>
<td>27%</td>
</tr>
</tbody>
</table>
Preliminary Study

Question: What would the analyst do with machine-generated data?

<table>
<thead>
<tr>
<th>Pr</th>
<th>Rec</th>
<th>∆Pr</th>
<th>∆Rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>40%</td>
<td>60%</td>
<td>+5%</td>
<td>-4%</td>
</tr>
<tr>
<td>58%</td>
<td>65%</td>
<td>+38%</td>
<td>-25%</td>
</tr>
<tr>
<td>23%</td>
<td>27%</td>
<td>-57%</td>
<td>-2%</td>
</tr>
</tbody>
</table>

Candidate link lists:

```
1. Trend???
2. From RE2003 reg
```
(Not Quite) Conclusions

Task

Feedback Loop

Analyst

Automated “Mining” Tool

Final Result
(Not Quite) Conclusions

- New Field of Study
- Larger Study Needed
Call for Help!

WANTED! VOLUNTEERS!
Thank You!