

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

Answer 1

Help study
the process

- ➔ After-the-fact
- ➔ Exploratory
- ➔ Conclusions help
future projects

What can Data Mining



Do for Software
Engineering ???

Question of the Day

Answer 1

Help study
the process

- ➔ After-the-fact
- ➔ Exploratory
- ➔ Conclusions help
future projects

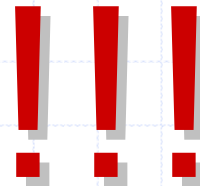
What can Data Mining



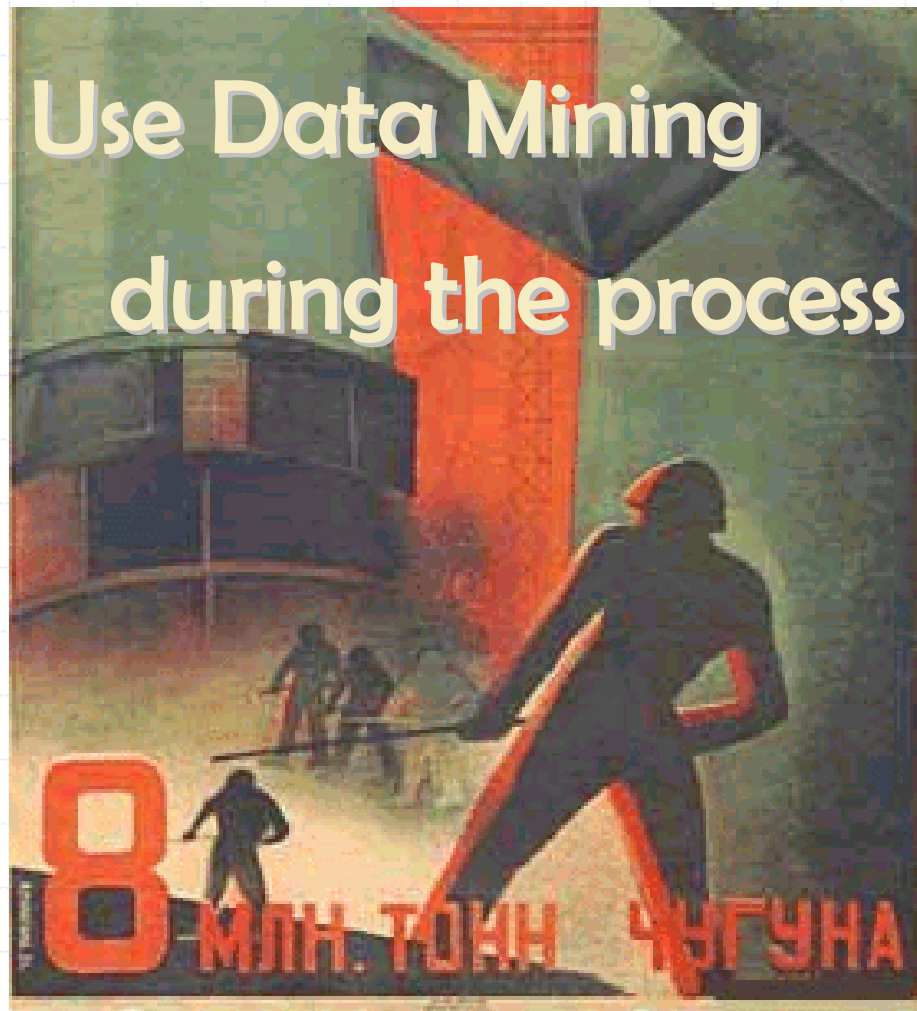
Do for Software
Engineering ???

Answer 2

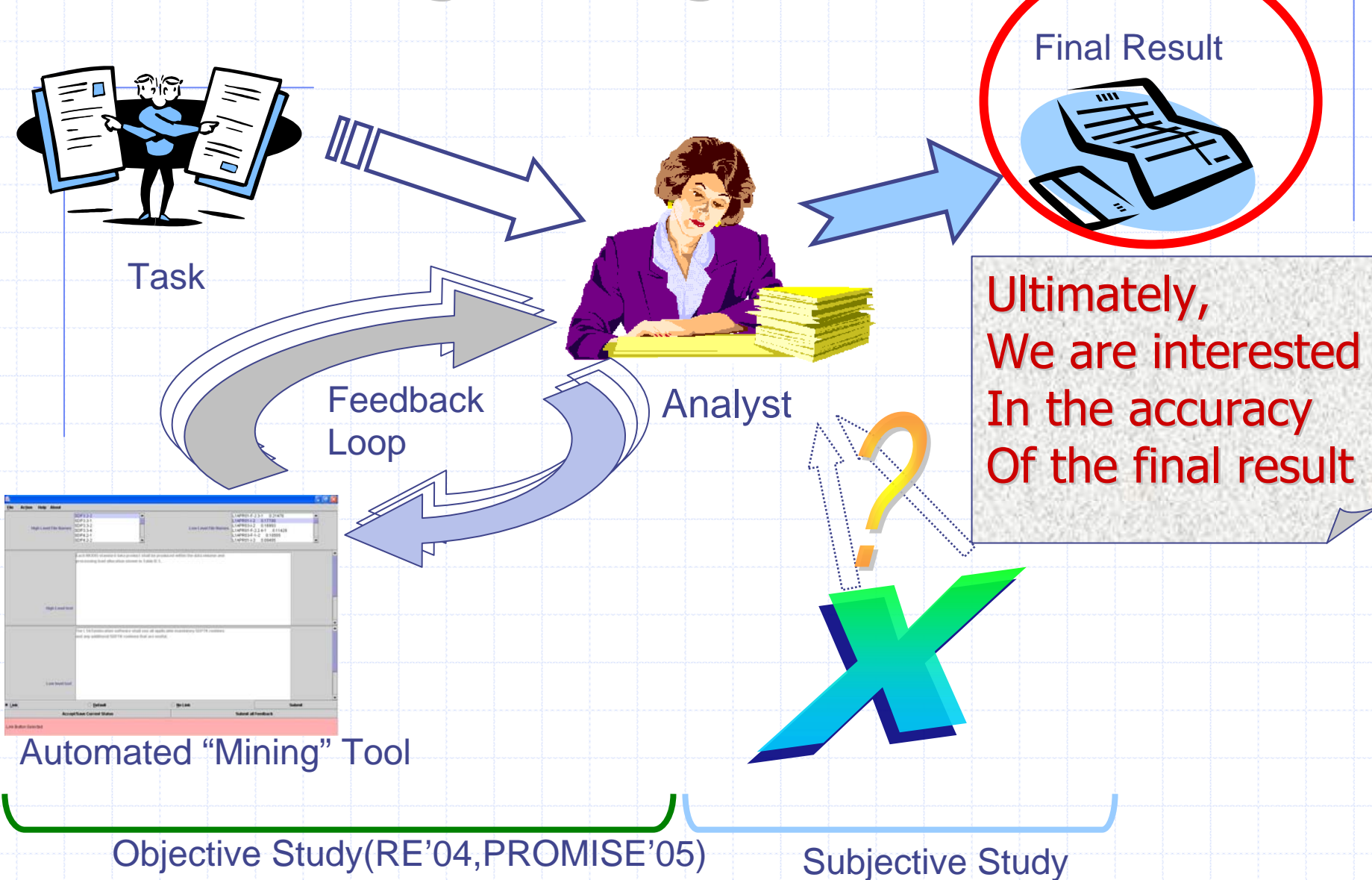
Help improve
the process



Our Approach



Use Mining During the Process?



Preliminary Study

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

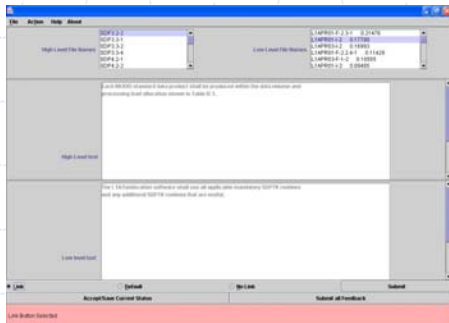


Analyst

Final Result



Task : Requirements Tracing
Metrics: Precision
Recall



Automated "Mining" Tool

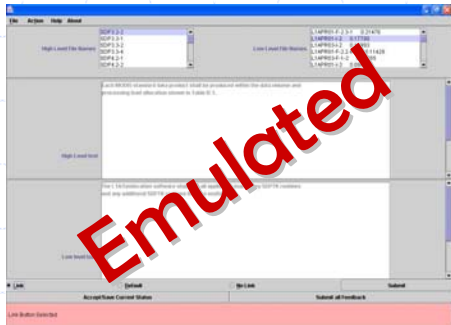
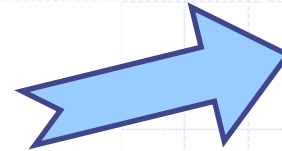
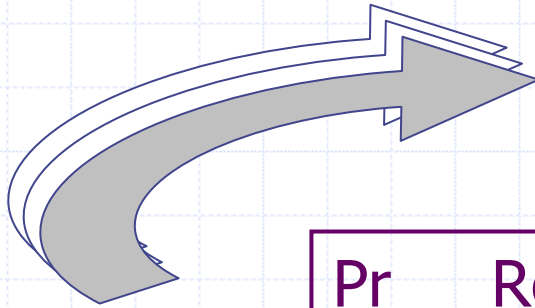
Preliminary Study

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

Final Result

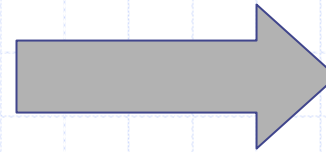


Analyst



Pr	Rec
40%	60%
20%	90%
80%	30%

Candidate link lists



Preliminary Study

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



Analyst

Pr	Rec
45%	56%
58%	65%
23%	27%

Pr	Rec
40%	60%
20%	90%
80%	30%

Candidate link lists

Preliminary Study

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

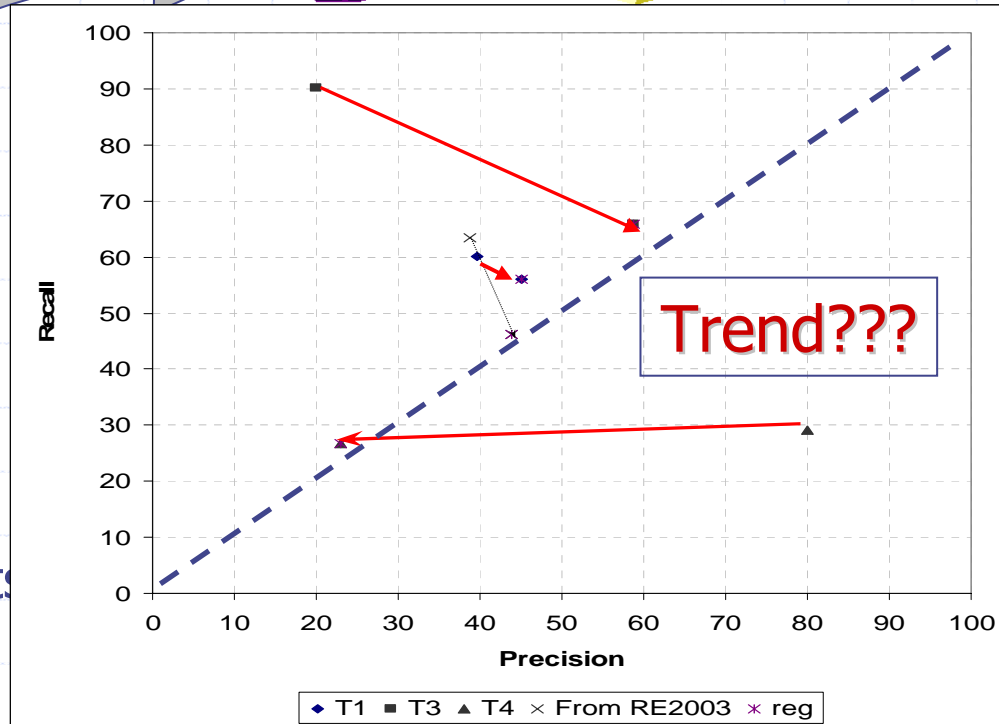


Pr	Rec
45%	56%
58%	65%
23%	27%

Δ Pr	Δ Rec
+5%	-4%
+38%	-25%
-57%	-2%

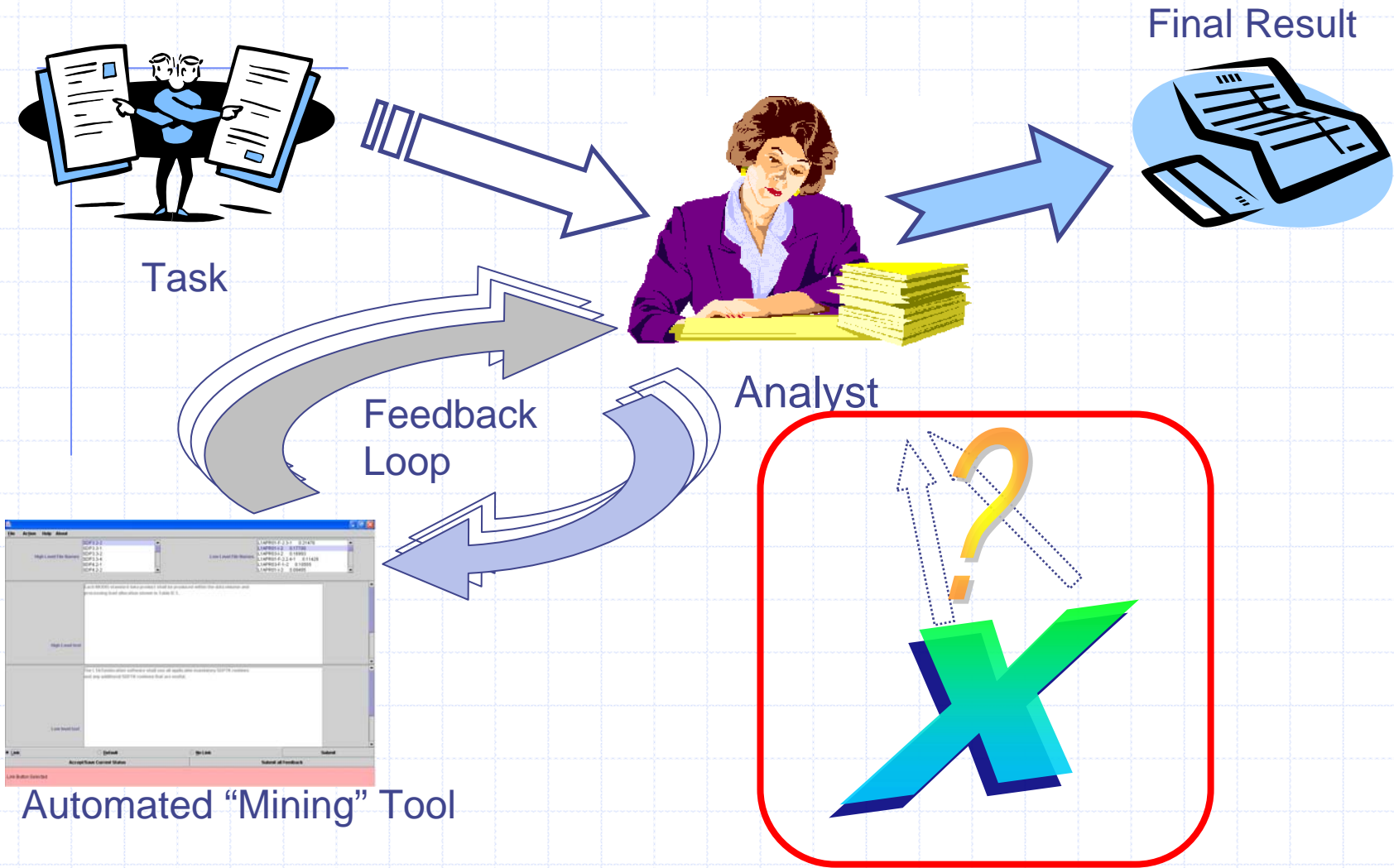
Pr	Rec
40%	60%
20%	90%
80%	30%

Candidate link lists

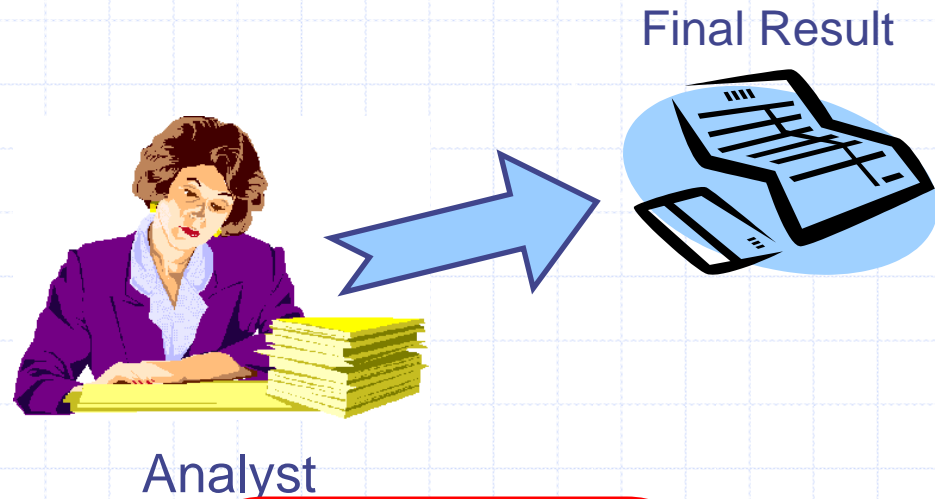


◆ T1 ■ T3 ▲ T4 × From RE2003 * reg

(Not Quite) Conclusions



(Not Quite) Conclusions



- New Field of Study
- Larger Study Needed



Call for Help!





Thank You!