

A TraceLab-based Solution for Creating, Conducting, and Sharing Feature Location Experiments

Bogdan Dit, Evan Moritz, Denys Poshyvanyk



SEMERU



**WILLIAM
& MARY**

20th IEEE International Conference on Program Comprehension
(ICPC'12) - Passau, Germany

Have you ever tried to
reproduce a case study and you
couldn't?

Case Study in SE

- In general, it is difficult to reproduce case studies
 - Some details may be missing
- Results of case studies produce contradictory results *

* T. Menzies and M. Shepperd, "Special Issue on Repeatable Results in Software Engineering Prediction", Empirical Software Engineering (ESE), vol. 17, no. 1-2, 2012, pp. 1-17 ³

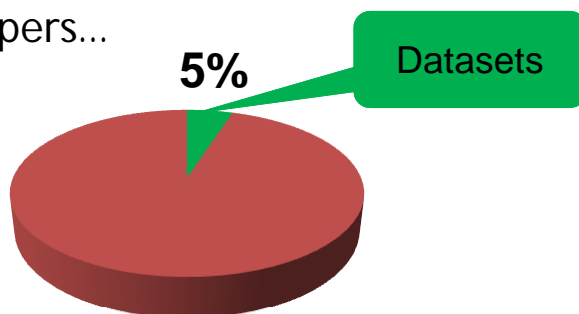
In Feature Location (FL)...

- Case studies are hard to reproduce**
 - Different datasets
 - Different implementations
 - Different metrics

** Dit, B., Revelle, M., Gethers, M., and Poshyvanyk, D., "Feature Location in Source Code: A Taxonomy and Survey", Journal of Software Maintenance and Evolution: Research and Practice (JSME), 2012, doi: 10.1002/smr.567 ⁴
<http://onlinelibrary.wiley.com/doi/10.1002/smr.567/abstract>

Different Datasets

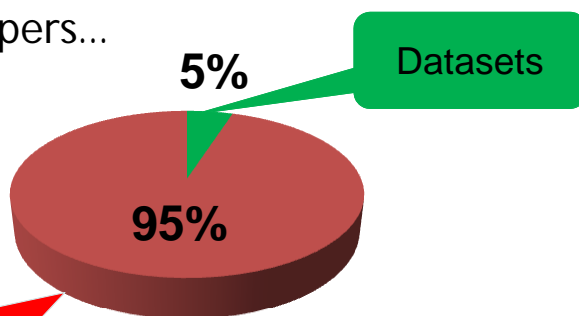
- Out of 60 papers...



5

Different Datasets

- Out of 60 papers...



Different software systems
Different versions of the systems
Different data points from the same system

6

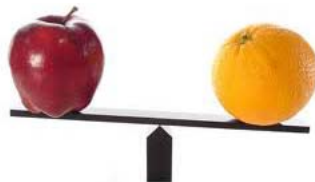
Different Implementation

- Often source code of FLTs is not available
- Configuration parameters are missing

7

Different Metrics

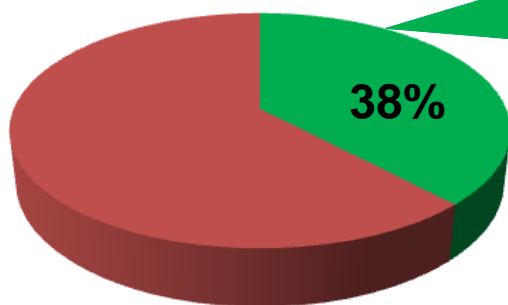
- Precision
- Recall
- F-measure
- Effectiveness measure



8

Other Observations

- Out of 60 papers...

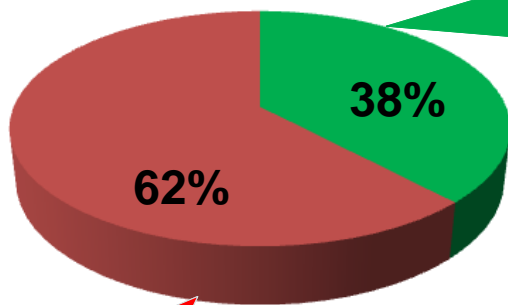


Compared their technique against a limited number of existing techniques

9

Other Observations

- Out of 60 papers...



Compared their technique against a limited number of existing techniques

No Comparison

10

Why is Reproducibility of Case Studies Important?

11

Why is Reproducibility of Case Studies Important?

- Confirm if field is making progress

12

Why is Reproducibility of Case Studies Important?

- Confirm if field is making progress
- (Many requests for tools and data at ICPC'12)

13

Why is Reproducibility of Case Studies Important?

- Results can be checked by reviewers
- Results can be verified by other researchers
- Clarify some aspects not explicitly stated in the paper
- Facilitate publishing new papers (others can use the data)

14

How to Publish a new FL Paper?



New FLT

15

How to Publish a new FL Paper?



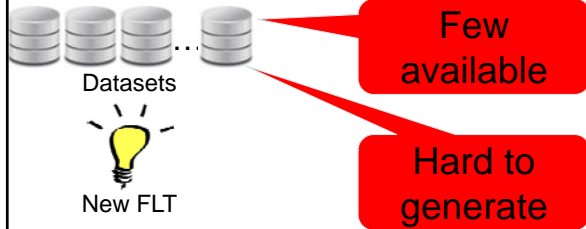
Datasets



New FLT

16

How to Publish a new FL Paper?



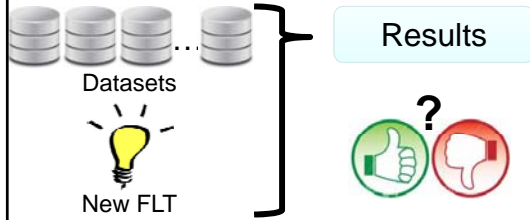
17

How to Publish a new FL Paper?



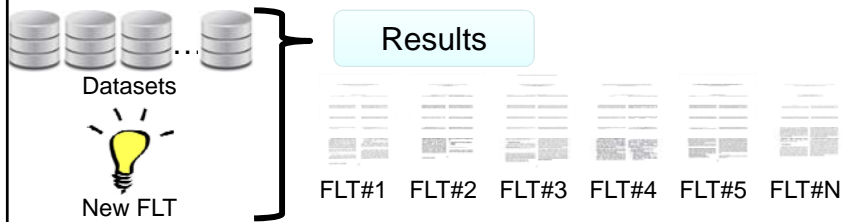
18

How to Publish a new FL Paper?

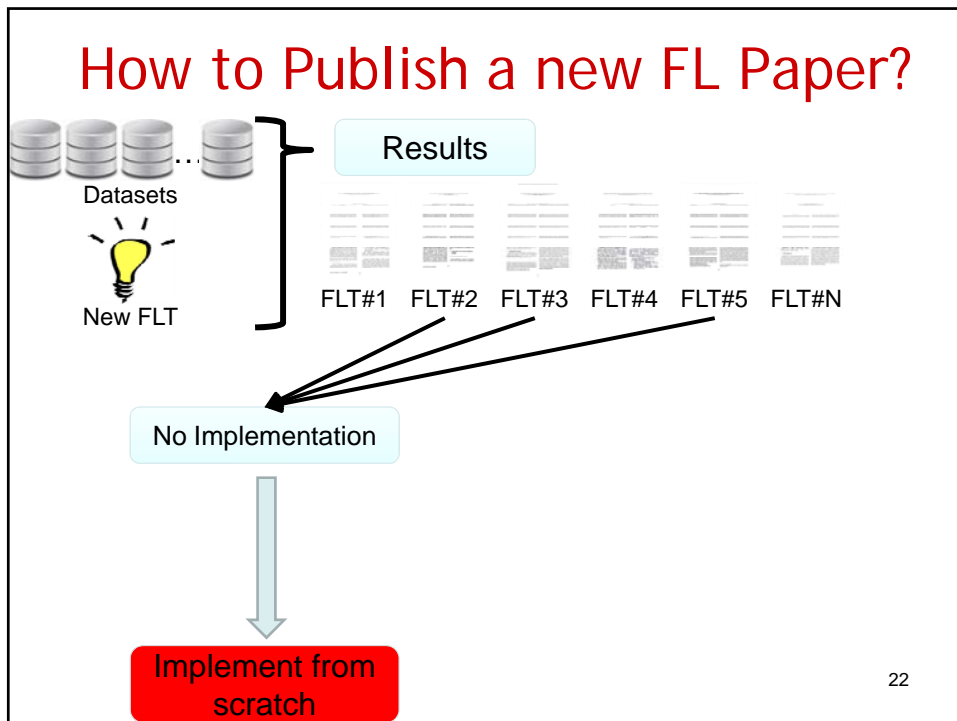
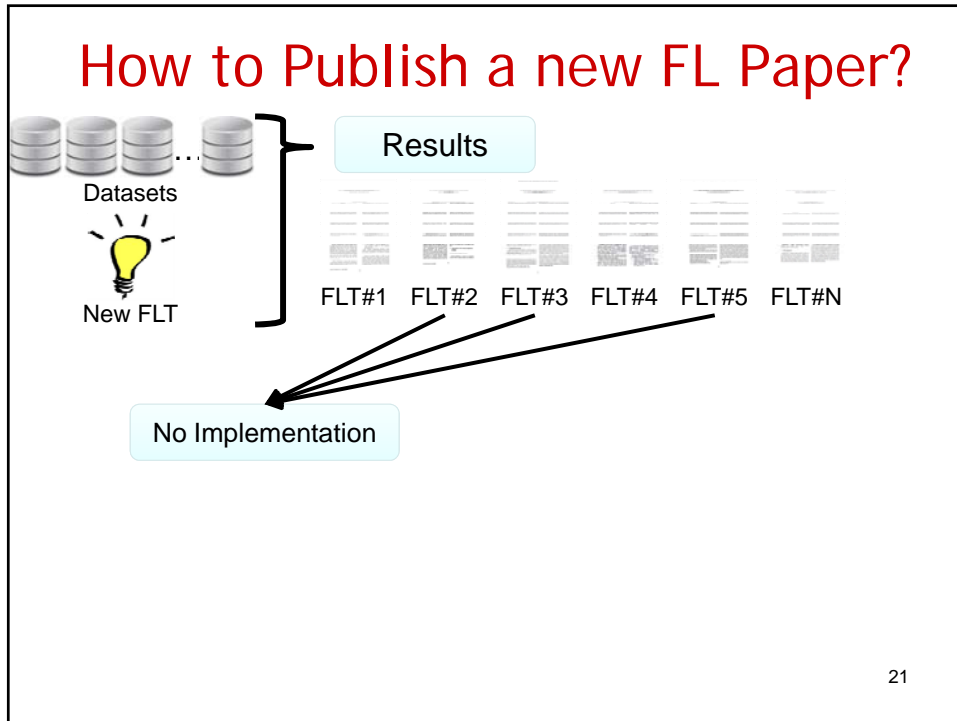


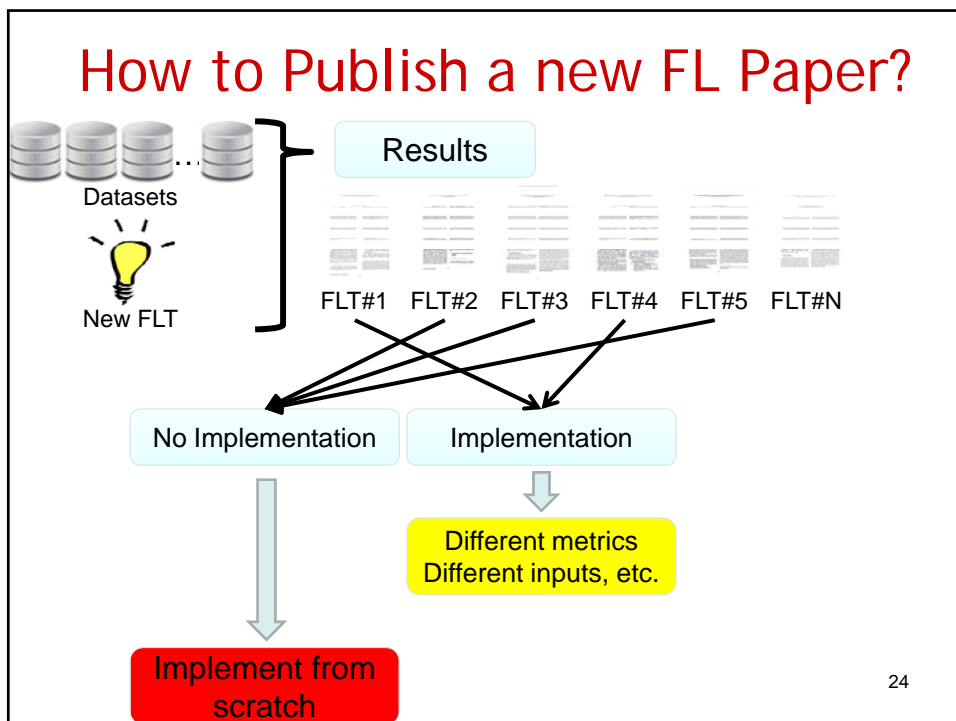
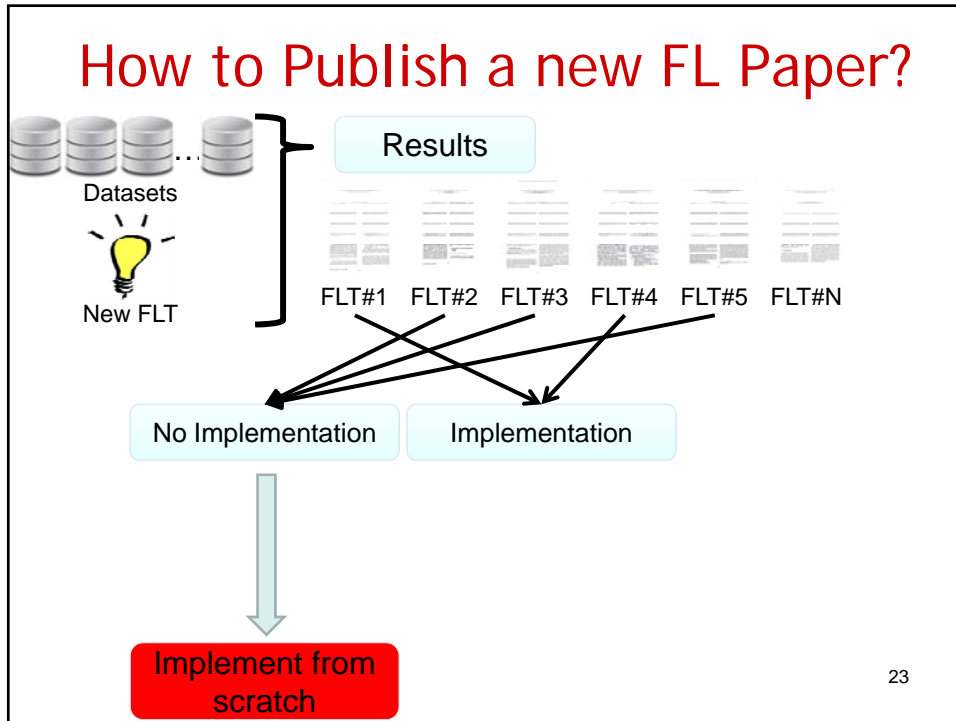
19

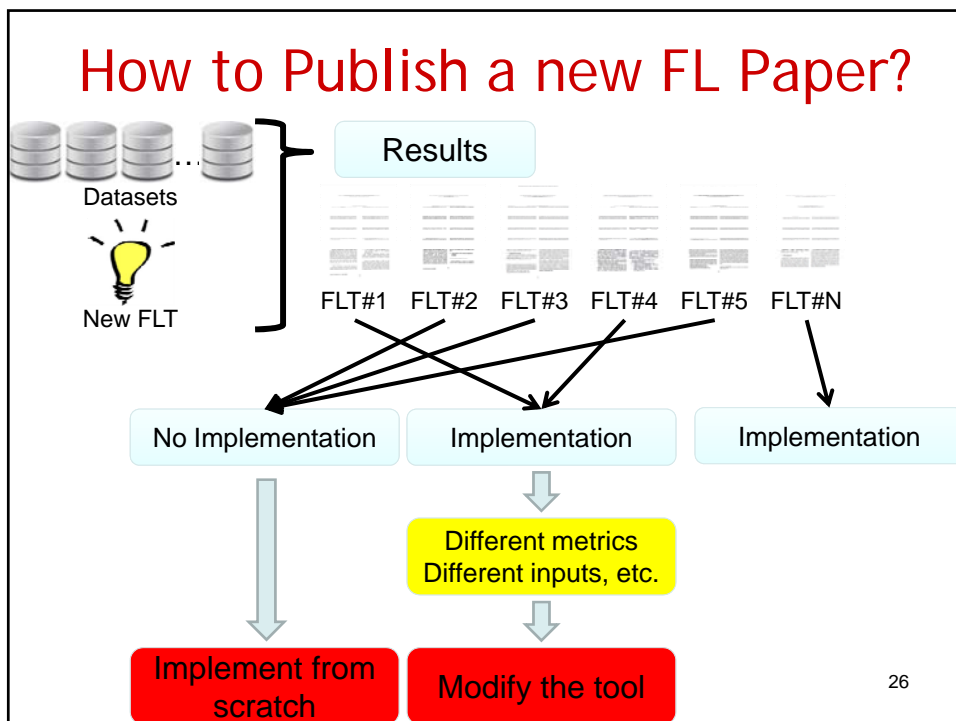
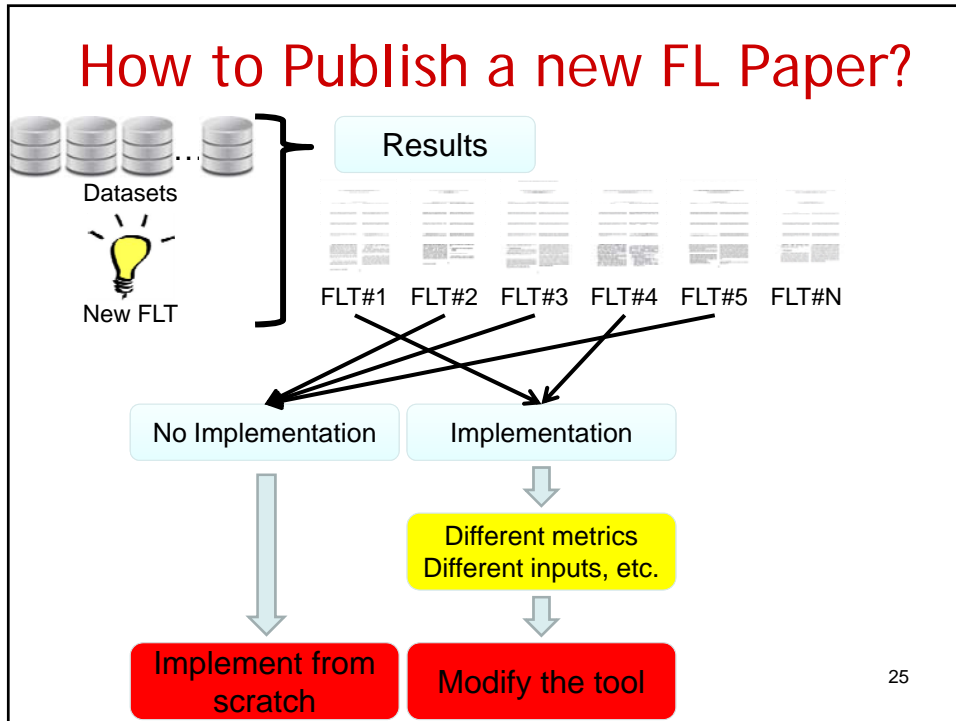
How to Publish a new FL Paper?

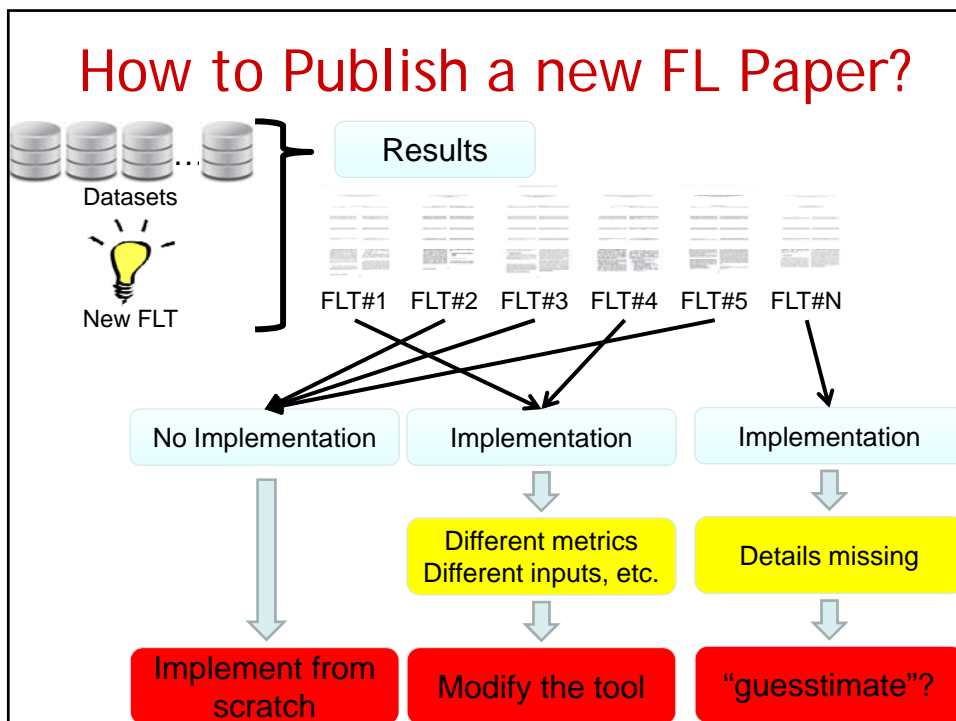
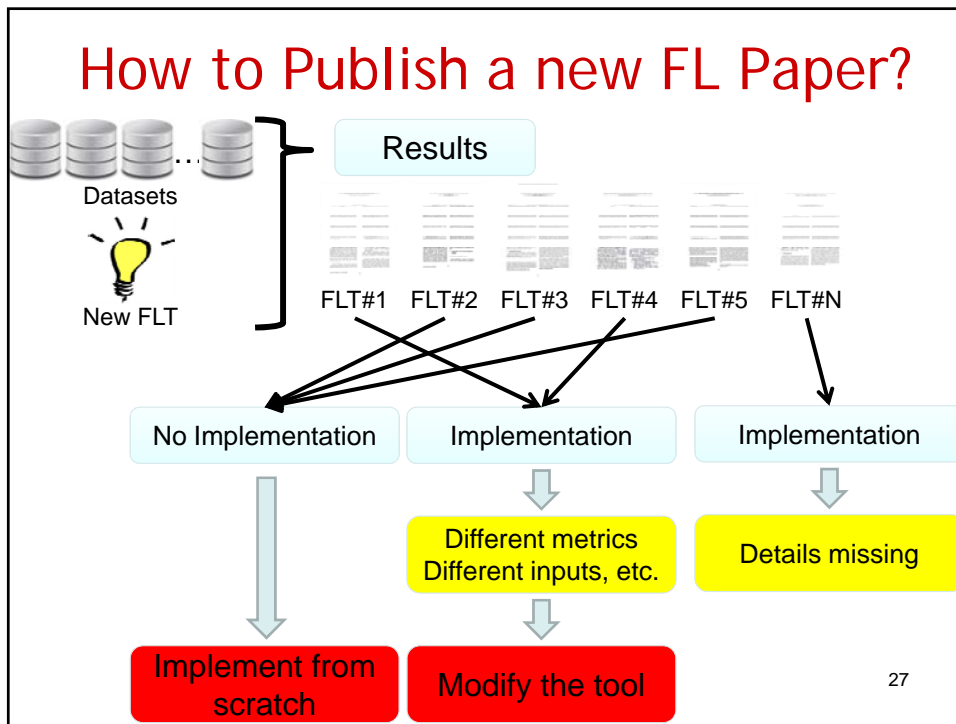


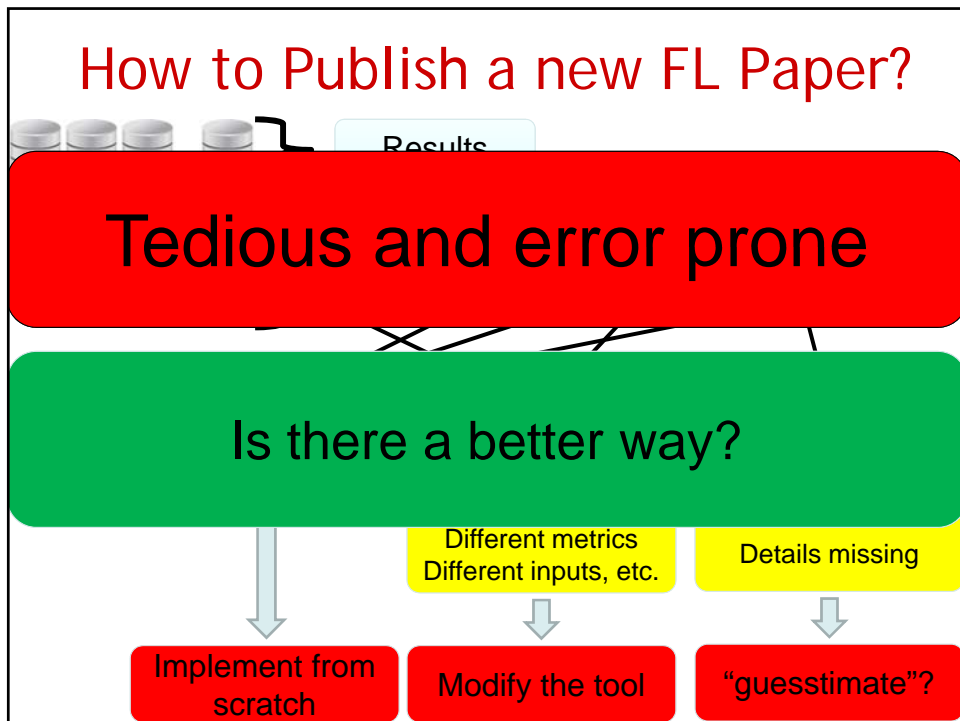
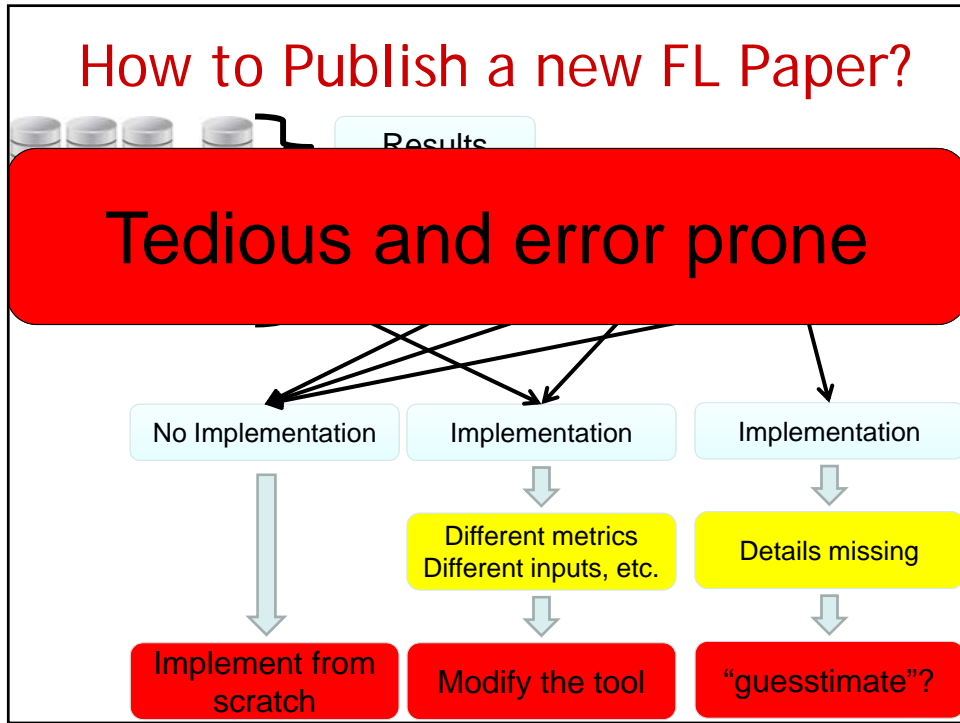
20











How to Publish a new FL Paper?

Results

Tedious and error prone

Is there a better way?

Different metrics

TraceLab

TraceLab

- Framework for:
 - Creating
 - Conducting
 - Sharingexperiments
- Designed to support traceability link recovery
- Developed by:



TraceLab

Name	Type	Value
QueryMap	System.Cc	System.Collectic
GoldSet	TraceLabS	TraceLabSDK.T
Stopwords	TraceLabS	[a, about, abov
Corpus	TraceLabS	TraceLabSDK.T
Document	System.Cc	System.Collectic
listOfArtifa	TraceLabS	TraceLabSDK.T
Queries	TraceLabS	TraceLabSDK.T
listOfArtifa	TraceLabS	TraceLabSDK.T
CorpusInd	TraceLabS	TraceLabSDK.T
VSMSimila	TraceLabS	TraceLabSDK.T

Status: Experiment done!

33

TraceLab

Experiment

Status: Experiment done!

34

Components (C#, Java) TraceLab Experiment

The screenshot shows the TraceLab interface with a workflow diagram on the right and a workspace view on the left. The workflow starts with 'Start', branching into 'Corpus Importer' and 'Gold Set Importer' (highlighted in green). 'Corpus Importer' leads to 'Corpus Remove Non-Literals', 'Corpus Split Identifiers', 'Corpus Stopwords', and 'Corpus Stemmer'. 'Gold Set Importer' leads to 'Stopwords Importer'. 'Queries Importer' leads to 'Queries Remove Non-Literals', 'Queries Split Identifiers', 'Queries Stopwords', and 'Queries Stemmer'. 'Stopwords Importer' and 'Queries Stopwords' both lead to 'Compute VSM Similarities'. 'Corpus Stemmer' leads to 'TFIDF Corpus'. 'Compute VSM Similarities' leads to 'Effectiveness All Ranks - VSM' and 'Effectiveness Best Ranks - VSM', both of which lead to 'End'. The workspace view on the left lists various components like 'QueryMap', 'GoldSet', 'Stopwords', 'Corpus', 'Document', 'ListOfArtifa', 'Queries', 'ListOfArtifa', 'CorpusInd', and 'VSMSimila'.

Status: Experiment done!

35

Components (C#, Java) TraceLab Experiment

This screenshot is identical to the one above, showing the TraceLab interface and workflow diagram. A callout box at the bottom center contains the text: 'Supports scripting languages: R, Python'. The workspace view and workflow diagram are the same as in the previous image.

Supports scripting languages:
R, Python

Status: Experiment done!

36

How we Envision Publishing FL Papers with TraceLab?

37

How we Envision Publishing FL Papers with TraceLab?



38

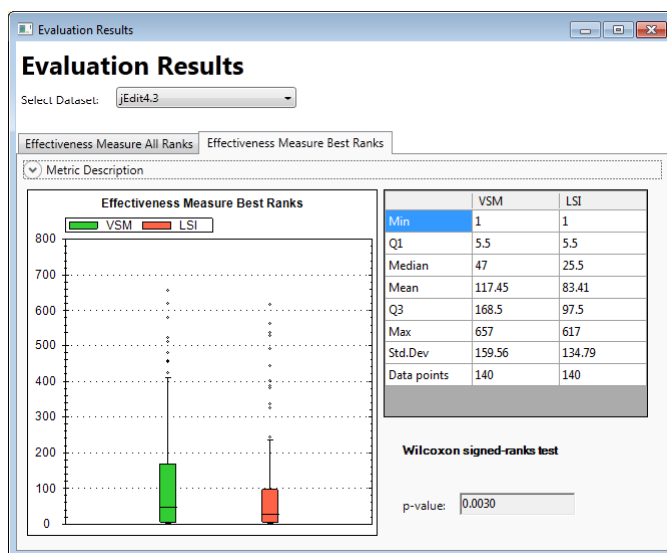
How we Envision Publishing FL Papers with TraceLab?

New FLT

Running the Experiment

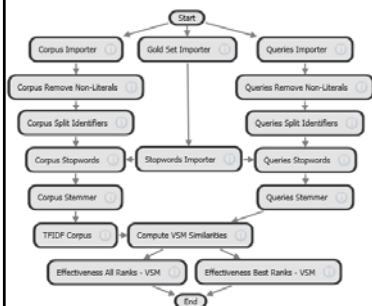
Status: Experiment done!

Running the Experiment



41

How to Create a Composite Component?



42

How to Create a Composite Component?

The flowchart on the left illustrates a complex workflow starting from a 'Start' node and ending at an 'End' node. It includes several intermediate steps such as 'Corpus Importer', 'Gold Set Importer', 'Queries Importer', 'Corpus Remove Non-Literals', 'Queries Remove Non-Literals', 'Corpus Split Identifiers', 'Queries Split Identifiers', 'Corpus Stopwords', 'Stopwords Importer', 'Queries Stopwords', 'Corpus Stemmer', 'Queries Stemmer', 'TFIDF Corpus', 'Compute VSM Similarities', 'Effectiveness All Ranks - VSM', and 'Effectiveness Best Ranks - VSM'.

The 'Export experiment as component' dialog on the right shows the configuration for this workflow. It includes a 'Define component info' section with fields for Component name, Author, Version, and Description. Below this are three tables: 'Import', 'Output', and 'Config Parameter'.

Import	Type	Include
GoldSet	TraceLabSDK.Types.TLSimilarityMatrix	<input checked="" type="checkbox"/>
DocumentMap	System.Collections.Generic.List`1[System.String, mscorlib, Ver	<input checked="" type="checkbox"/>
Corpus	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
Queries	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
QueryMap	System.Collections.Generic.Dictionary`2[System.Int32, mscorlib	<input checked="" type="checkbox"/>
Stopwords	TraceLabSDK.Types.TLStopwords	<input checked="" type="checkbox"/>

Output	Type	Include
Corpus	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
Queries	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
listOfArtifacts	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
VSMSimilarities	TraceLabSDK.Types.TLSimilarityMatrix	<input checked="" type="checkbox"/>
CorpusIndex	TraceLabSDK.Types.TLDictionaryIndex	<input checked="" type="checkbox"/>

Config Parameter	Alias	Type
Gold Set Importer Directory	Gold Set Importer Direct:	System.String
Corpus Importer Identifiers	Corpus Importer Identifie	TraceLabSDK.Component.Cc
Corpus Importer Documents	Corpus Importer Docume	TraceLabSDK.Component.Cc
Queries Importer QueriesDir	Queries Importer Queries	System.String
Queries Importer FeaturesIC	Queries Importer Feature	TraceLabSDK.Component.Cc
Queries Importer BugsIDFile	Queries Importer BugsID	TraceLabSDK.Component.Cc

How to Create a Composite Component?

This section is identical to the one above, showing the workflow flowchart and the 'Export experiment as component' dialog configuration.

Below the flowchart, a green square icon with the text 'VSM FLT' and an information symbol is shown, with an arrow pointing to it from the right.

How to Create a Composite Component?

Export experiment as component

Export as component

Define component info:

Component name: VSM FLT Experiment
 Author: Evan Moritz
 Version:
 Description: Performs the Vector Space Model Feature Location Technique.

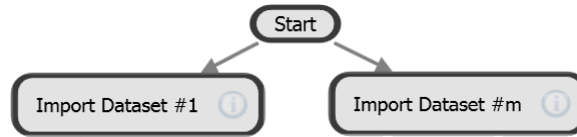
Import	Type	Include
GoldSet	TraceLabSDK.Types.TLSimilarityMatrix	<input checked="" type="checkbox"/>
DocumentMap	System.Collections.Generic.List`1[System.String, mscorlib, Ver	<input checked="" type="checkbox"/>
Corpus	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
Queries	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
QueryMap	System.Collections.Generic.Dictionary`2[System.Int32, mscorlib	<input checked="" type="checkbox"/>
Stopwords	TraceLabSDK.Types.TLStopwords	<input checked="" type="checkbox"/>

Output	Type	Include
Corpus	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
Queries	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
listOfArtifacts	TraceLabSDK.Types.TLArtifactsCollection	<input checked="" type="checkbox"/>
VSMSimilarities	TraceLabSDK.Types.TLSimilarityMatrix	<input checked="" type="checkbox"/>
CorpusIndex	TraceLabSDK.Types.TLDictionaryIndex	<input checked="" type="checkbox"/>

Config Parameter	Alias	Type
Gold Set Importer Directory	Gold Set Importer Directc	System.String
Corpus Importer Identifiers	Corpus Importer Identifie	TraceLabSDK.Component.Cc
Corpus Importer Documents	Corpus Importer Docume	TraceLabSDK.Component.Cc
Queries Importer QueriesDir	Queries Importer Queries	System.String
Queries Importer FeaturesIC	Queries Importer Feature	TraceLabSDK.Component.Cc
Queries Importer BugsIDFile	Queries Importer BugsID	TraceLabSDK.Component.Cc

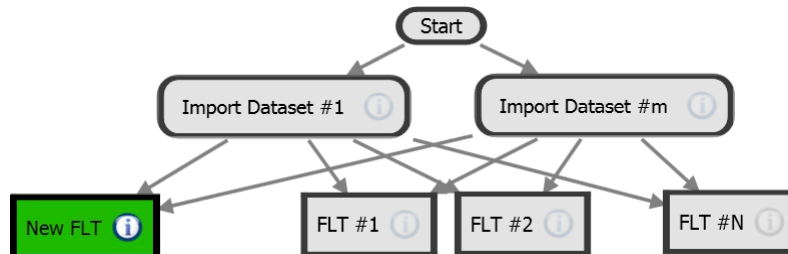
Comparing the Results

Comparing the Results



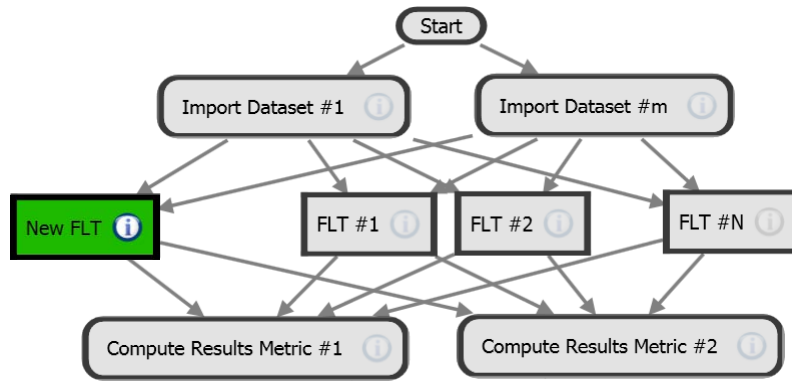
47

Comparing the Results



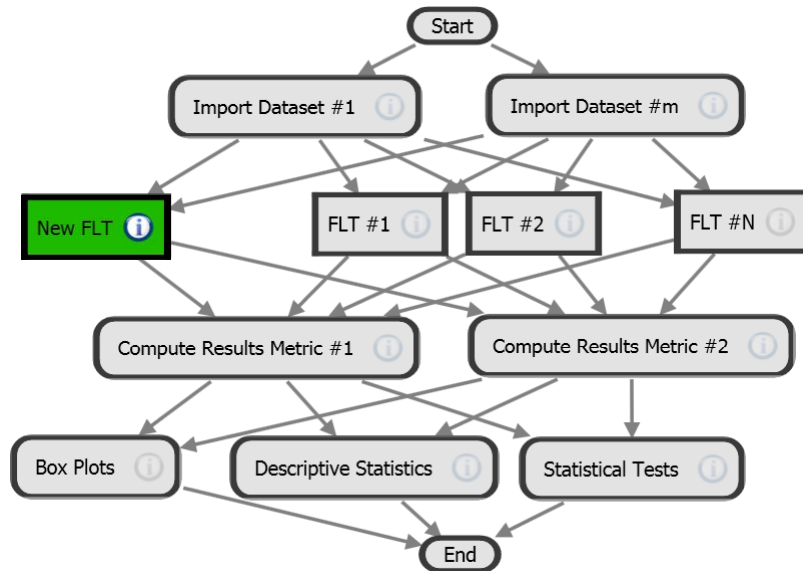
48

Comparing the Results



49

Comparing the Results



0

TraceLab Vision for FL

- Platform for creating, reproducing and sharing experiments
- In the future, submit a paper with a link to a TraceLab experiment:
 - Data
 - Code
 - Configuration parameters
 - Etc.

51

Status of TraceLab

- Beta version available by request
- TraceLab will be released in **July 2012**
- Features under development:
 - Packaging feature that bundles up:
 - Data
 - Code
 - Settings
 - Performance and GUI enhancements

52

Our Contribution

<http://www.cs.wm.edu/semeru/data/TraceLab-feature-location/>

- Datasets for 5 systems:
 - Features to methods mapping
 - Execution traces
 - Queries
 - Corpora
- Reusable components and sample experiments for FL
- Evaluation metrics

53

Conclusions and Future Work

- TraceLab & Feature Location = Reproducible experiments
- Research community welcomed to share:
 - More datasets
 - More components and experiments

54

Conclusions and Future Work

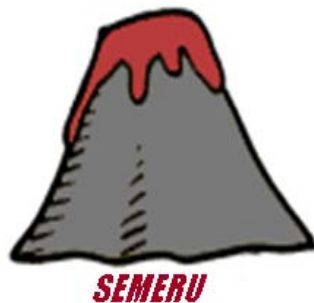
- Adapt TraceLab for other SE areas:
 - Impact analysis
 - Detection of duplicate bug reports
 - Developer recommendation
- At ICSE'12 TraceLab got a lot of interest from large software groups
- Hopefully adaption by ICPC community

55

Thank you! Questions?

SEMERU @ William and Mary

<http://www.cs.wm.edu/semeru/>



56

Generating the jEdit Datasets

SVN Commits between v4.2-v4.3

r13898 | shlomy | 2008-10-17 00:45:12 -0400 (Fri, 17 Oct 2008) | 2 lines
Changed paths:
M /jEdit/trunk/doc/CHANGES.txt
M /jEdit/trunk/org/gjt/sp/jedit/search/HyperSearchRequest.java

Fixed bug #2173112: Search within selection of regexp '\$' does not match the last line. This was caused by an exception - if the last selection line is empty, buffer.getSegment() was called with a negative length of -1. Also removed an unused variable.

57

Generating the jEdit Datasets

r13898 | shlomy | 2008-10-17 00:45:12 -0400 (Fri, 17 Oct 2008) | 2 lines
Changed paths:
M /jEdit/trunk/doc/CHANGES.txt
M /jEdit/trunk/org/gjt/sp/jedit/search/HyperSearchRequest.java

Fixed ~~bug~~ #2173112: Search within selection of regexp '\$' does not match the last line. This was caused by an exception - if the last selection line is empty, buffer.getSegment() was called with a negative length of -1. Also removed an unused variable.

SVN commit message

3pre18: regexp '\$' does not match last line in selection

Details:

- In a new buffer, make 10 empty lines. (NOTE: they don't have to be empty)
- Select any five lines (e.g. a text selection)
- Search for \$ (with hypersearch, selection, and regular expressions turned on)
- This only matches the first 4 lines. The fifth is not matched
- Now search for \$ and replace with z
- Since the fifth line was not matched, z is placed on only the first 4 lines.
- Please note:
 - * works fine. Just \$ demonstrates this behaviour
 - * the lines do not have to be empty. This is reproducible for non-empty lines as well
 - * and any mixture of empty/non-empty lines

Submitted: Carlos (solrac776) - 2008-10-16 23:00:21 UTC Assigned: Shlomy Reinstein
 Priority: 5 Category: search and replace
 Status: Closed Group: None
 Resolution: Fixed Visibility: Public

Title

+

Description

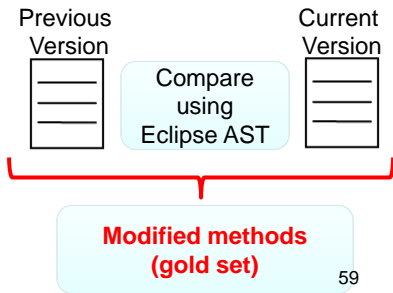
=

Query

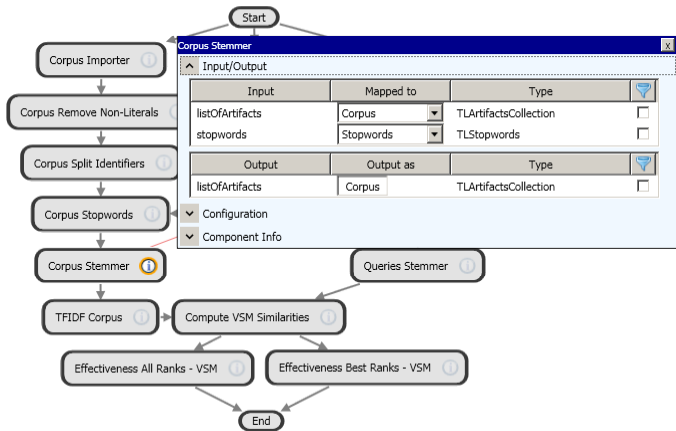
Generating the jEdit Datasets

```
r13898 | shlomy | 2008-10-17 00:45:12 -0400 (Fri, 17 Oct 2008) | 2 lines
Changed paths:
  H /jEdit/trunk/doc/CHANGES.txt
  H /jEdit/trunk/org/gjt/sp/jedit/search/HyperSearchRequest.java
Fixed bug #2173112: Search within selection of regexp '$' does not match
the last line. This was caused by an exception - if the last selection
line is empty, buffer.getSegment() was called with a negative length of
-1. Also removed an unused variable.
```

Changed files



Configuring Components



Comparing the Results (Alternative)

