Bug linking techniques in software repositories

Advanced topics in Software Engineering

Giovanni Magoga

October 24, 2020

University of Gottingen

TABLE OF CONTENTS

1. Introduction

2. Problems

3. Linking techniques

4. Comparison

Introduction

PREREQUISITES

- What is a version control system?
- What is a bug tracking system?
- How do they interact?

VERSION CONTROL

A category of software tools that help a software team manage changes to source code over time by tracking modifications on a special kind of database.

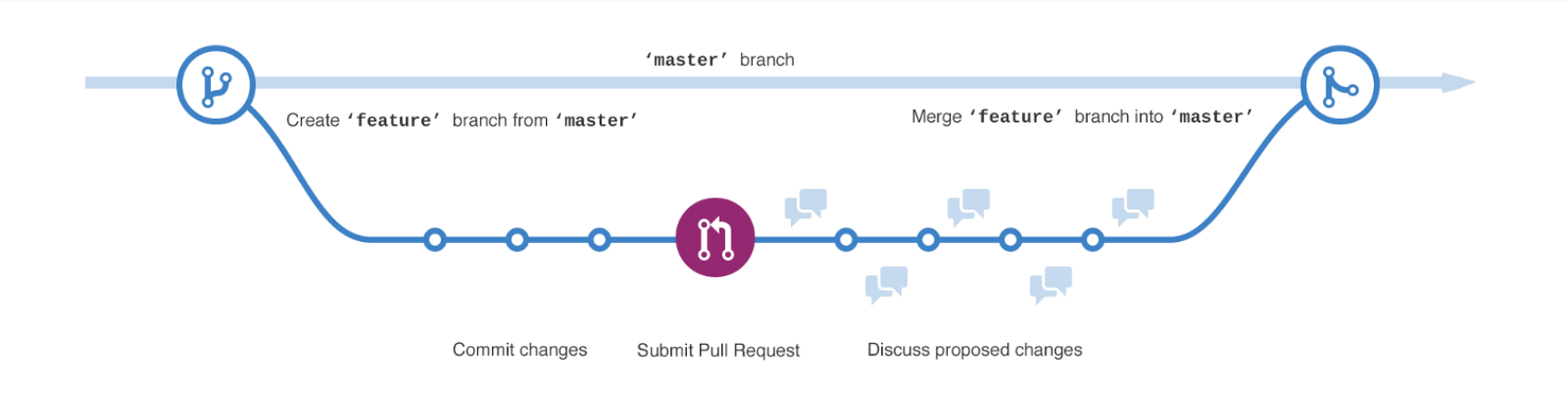


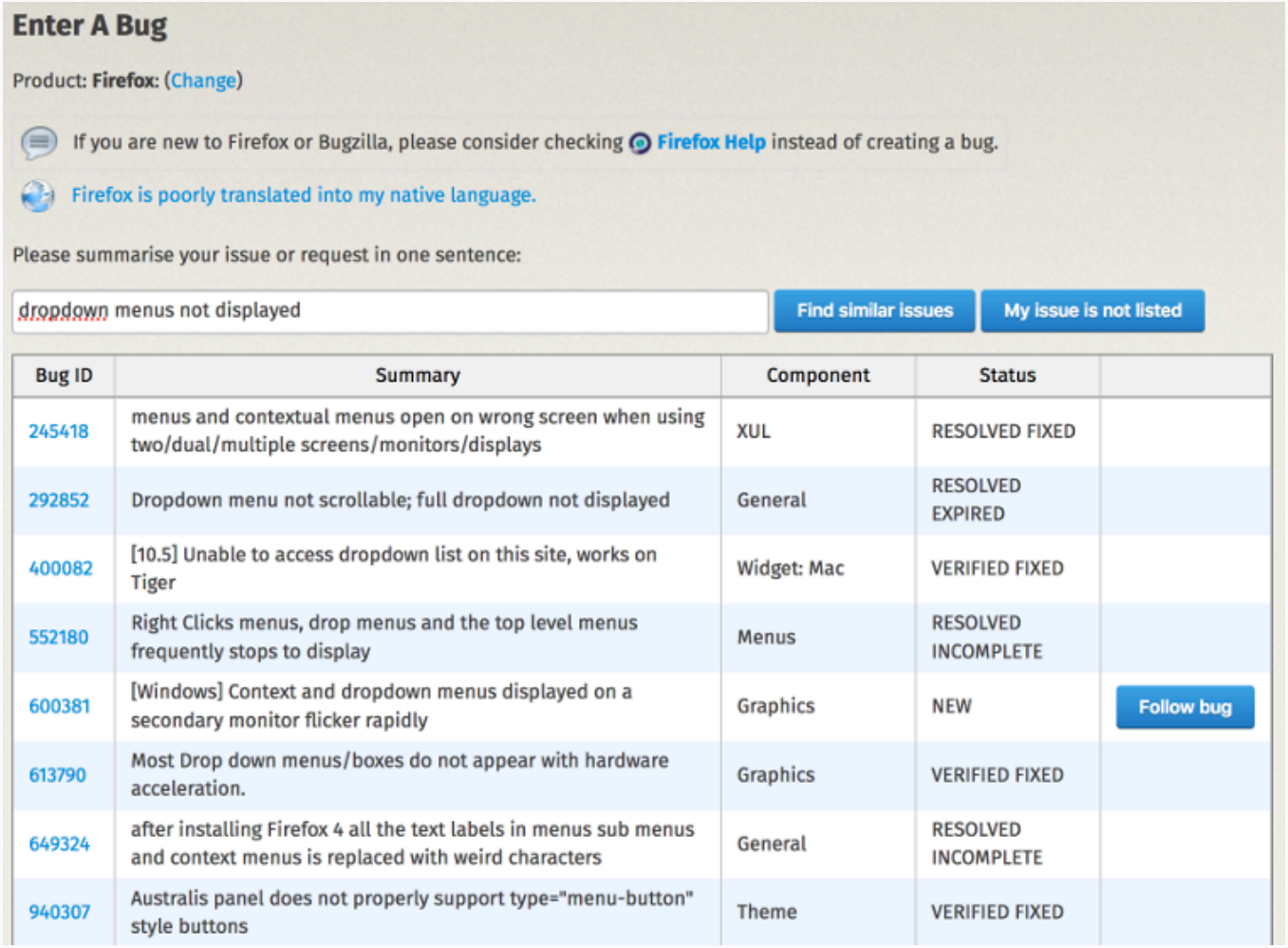
Figure 1: VCS Workflow example [1]

VERSION CONTROL SYSTEMS

- Decentralized (despite GitHub, Bitbucket)
- Created by Linus Torvalds in 2005
- Very popular in OSS communities

- Centralized
- Created by Apache in 2000
- Mainly used inside companies

BUG TRACKING



Enter A Bug

Product: **Firefox:** (Change)

If you are new to Firefox or Bugzilla, please consider checking Firefox Help instead of creating a bug.

Firefox is poorly translated into my native language.

Please summarise your issue or request in one sentence:

dropdown menus not displayed | Find similar issues | My issue is not listed

Bug ID	Summary	Component	Status	
245418	menus and contextual menus open on wrong screen when using two/dual/multiple screens/monitors/displays	XUL	RESOLVED FIXED	
292852	Dropdown menu not scrollable; full dropdown not displayed	General	RESOLVED EXPIRED	
400082	[10.5] Unable to access dropdown list on this site, works on Tiger	Widget: Mac	VERIFIED FIXED	
552180	Right Clicks menus, drop menus and the top level menus frequently stops to display	Menus	RESOLVED INCOMPLETE	
600381	[Windows] Context and dropdown menus displayed on a secondary monitor flicker rapidly	Graphics	NEW	Follow bug
613790	Most Drop down menus/boxes do not appear with hardware acceleration.	Graphics	VERIFIED FIXED	
649324	after installing Firefox 4 all the text labels in menus sub menus and context menus is replaced with weird characters	General	RESOLVED INCOMPLETE	
940307	Australis panel does not properly support type="menu-button" style buttons	Theme	VERIFIED FIXED	

Figure 2: Bugzilla

Figure 3: Issue tracker

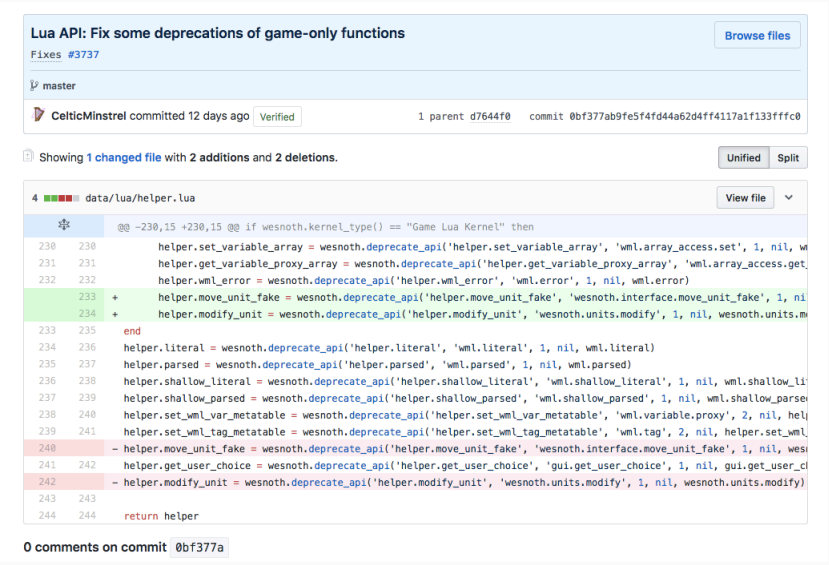


Figure 4: Bug fixing commit

Problems

Figure 5: Obligatory xkcd comic [5]

PROBLEMS

This introduces a number of problems:

- Makes release notes uninformative
- Slows down the reviewing process
- Makes software hard to maintain
- Loss of valuable production insights

SOLUTIONS

Just write better commit descriptions

- Hard to enforce in decentralized environments
- Inconsistencies between different developers
- Development timeline discrepancies

Or we can resort to bug-linking methods

Linking techniques

Research on linking methods

- When Do Changes Induce Fixes?, 2005 - J. Sliwerski et al. [9]
- ReLink, 2011 - R. Wu, H. Zhang, S. Kim, and S.-C. Cheung [11]
- MLink, 2012 - A. T. Nguyen, T. T. Nguyen, H. A. Nguyen [7]
- Bflinks, 2014 - L. Prechelt and A. Pepper [8]
- RCLink, 2015 - B. Le, M. Linares, D. Lo, and D. Poshyvanyk [6]
- FRLink, 2016 - Y. Sun, Q. Wang, and Y. Yang [10]

WHEN DO CHANGES INDUCE FIXES?

Syntactic analysis:

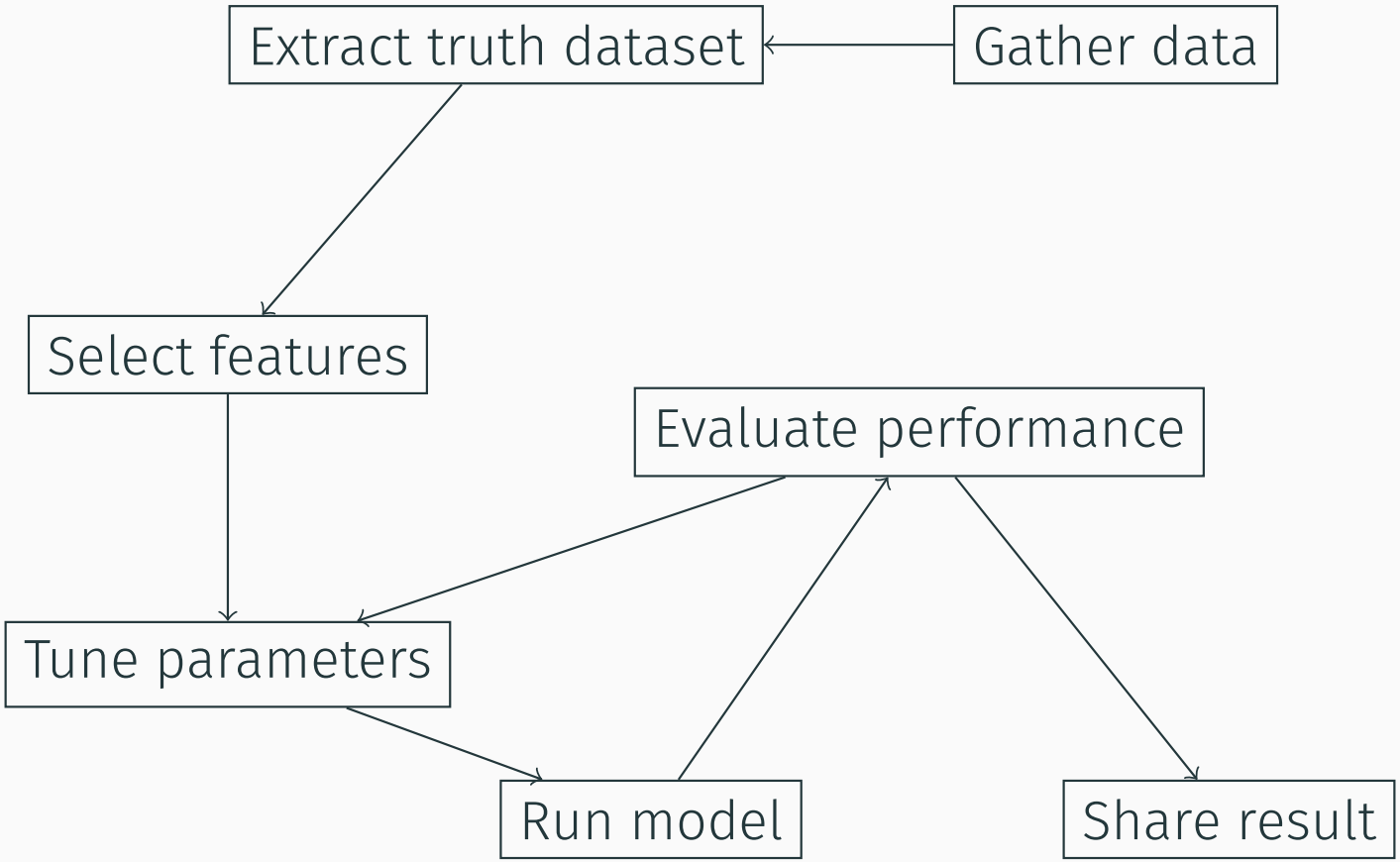
- strings containing `bug`, `fix`, `pr`, etc.
- plain numbers: `345933`
- alphanumeric strings: `a4d015b2`

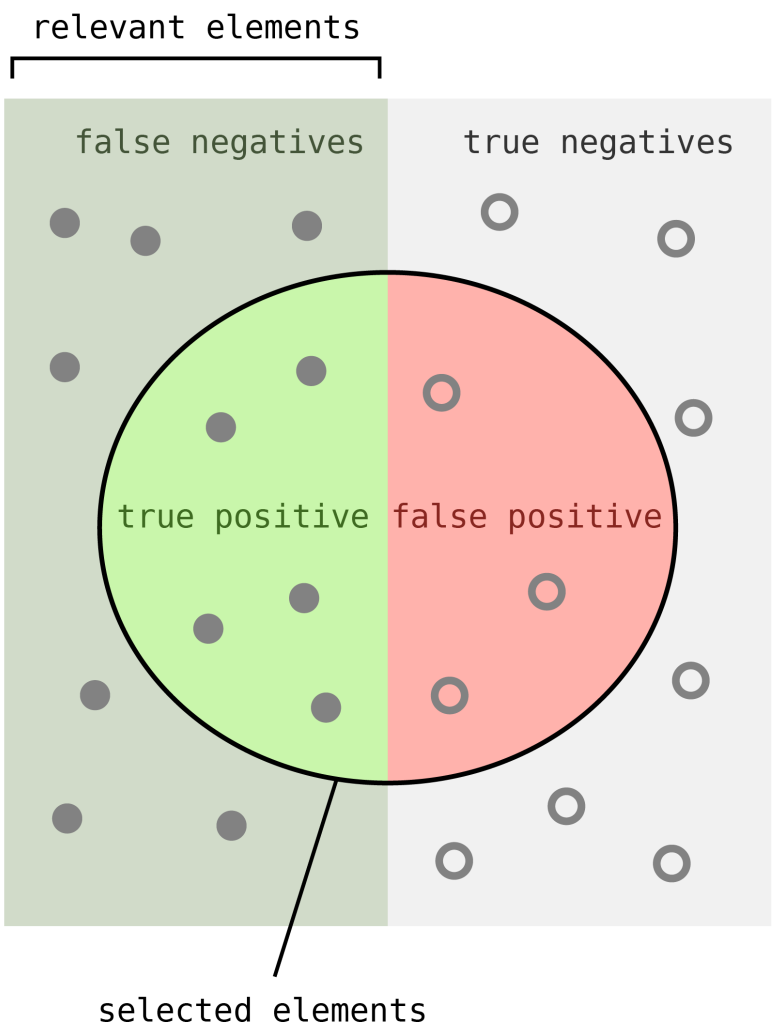
Semantic analysis:

- bug is flagged as `FIXED` in the bug tracker
- the author of the fixing commit has been assigned to the bug
- the bug report contains a description of the commit

Outcome: Don't program on fridays

WORKFLOW


Extract truth dataset
Gather data
Select features
Evaluate performance
Tune parameters
Run model
Share result



How many selected items are relevant?

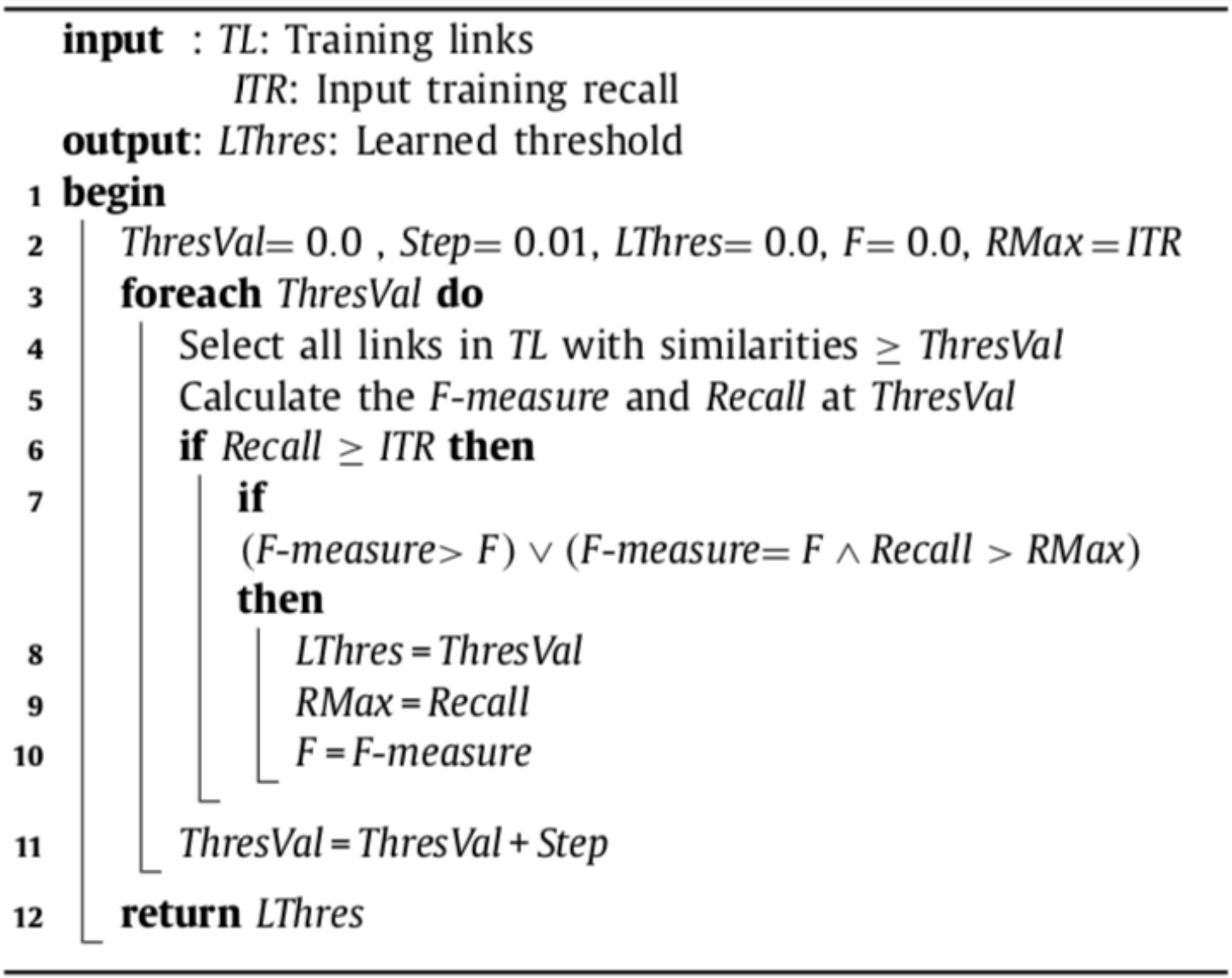
$$\text{Precision} = \frac{\text{true positive}}{\text{selected elements}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positive}}{\text{relevant elements}}$$

$$F_1 = \left(\frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1}$$

Figure 6: F-score measure [4]



input : *TL*: Training links
ITR: Input training recall
output: *LThres*: Learned threshold

```
1  begin
2      ThresVal= 0.0 , Step= 0.01, LThres= 0.0, F= 0.0, RMax=ITR
3      foreach ThresVal do
4          Select all links in TL with similarities ≥ ThresVal
5          Calculate the F-measure and Recall at ThresVal
6          if Recall ≥ ITR then
7              if
               (F-measure> F) ∨ (F-measure= F ∧ Recall > RMax)
               then
8                  LThres = ThresVal
9                  RMax = Recall
10                 F = F-measure
11         ThresVal = ThresVal + Step
12     return LThres
```

Figure 7: Basic threshold learning algorithm [10]

ReLink is the first formalized feature-based approach to bug linking, demonstrating major improvements over regex-based methods.

Datasets

Apache, ZXing, OpenIntents

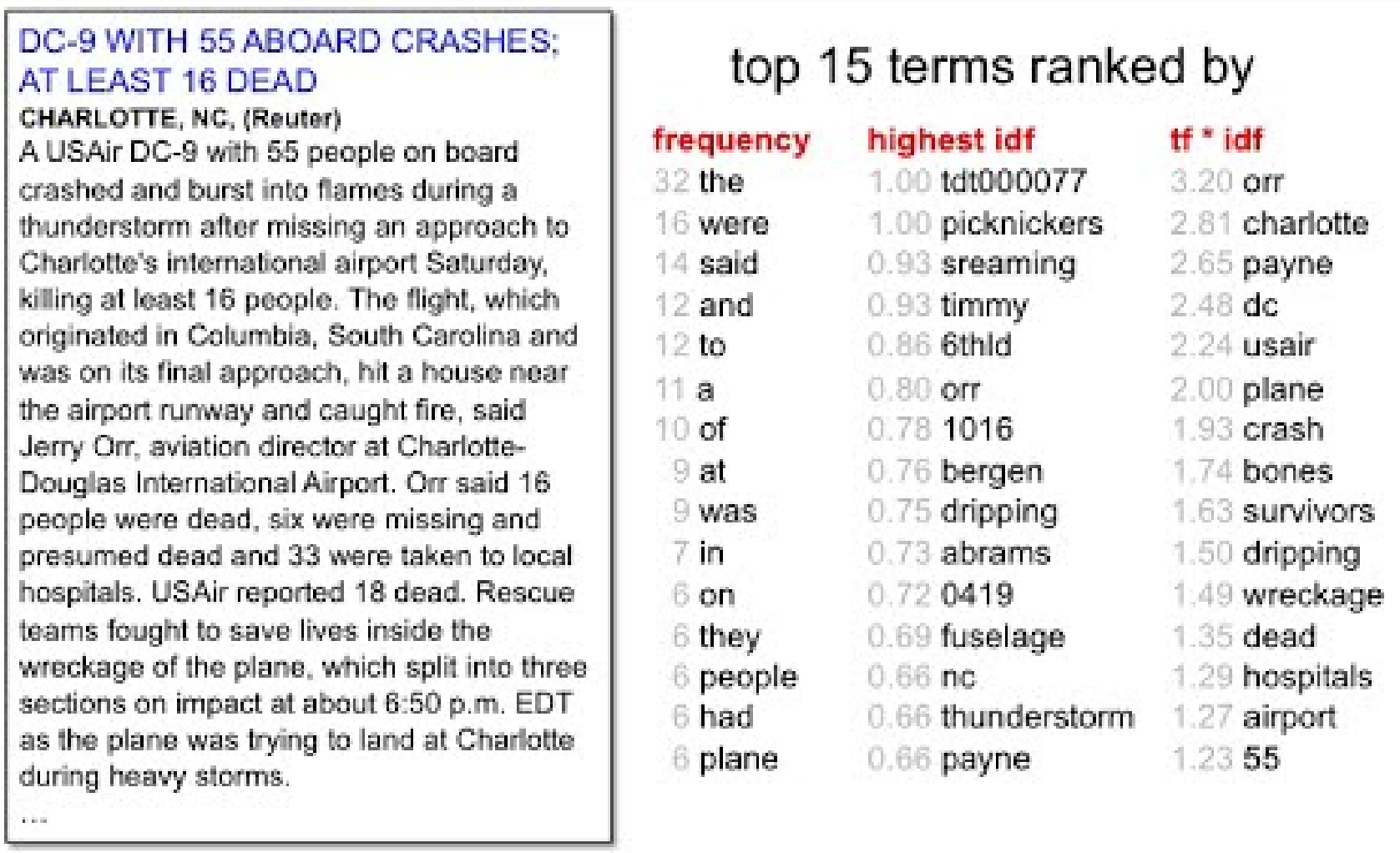
Average scores

Precision	–
Recall	–
F-score	–

Features

- Referential causality
- Bug owner and fixer
- Bug-commit descriptions similarity

TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY



DC-9 WITH 55 ABOARD CRASHES; AT LEAST 16 DEAD

CHARLOTTE, NC, (Reuter)

A USAir DC-9 with 55 people on board crashed and burst into flames during a thunderstorm after missing an approach to Charlotte's international airport Saturday, killing at least 16 people. The flight, which originated in Columbia, South Carolina and was on its final approach, hit a house near the airport runway and caught fire, said Jerry Orr, aviation director at Charlotte-Douglas International Airport. Orr said 16 people were dead, six were missing and presumed dead and 33 were taken to local hospitals. USAir reported 18 dead. Rescue teams fought to save lives inside the wreckage of the plane, which split into three sections on impact at about 6:50 p.m. EDT as the plane was trying to land at Charlotte during heavy storms.

...

top 15 terms ranked by

frequency	highest idf	tf * idf
32 the	1.00 tdt000077	3.20 orr
16 were	1.00 picknickers	2.81 charlotte
14 said	0.93 sreaming	2.65 payne
12 and	0.93 timmy	2.48 dc
12 to	0.86 6thld	2.24 usair
11 a	0.80 orr	2.00 plane
10 of	0.78 1016	1.93 crash
9 at	0.76 bergen	1.74 bones
9 was	0.75 dripping	1.63 survivors
7 in	0.73 abrams	1.50 dripping
6 on	0.72 0419	1.49 wreckage
6 they	0.69 fuselage	1.35 dead
6 people	0.66 nc	1.29 hospitals
6 had	0.66 thunderstorm	1.27 airport
6 plane	0.66 payne	1.23 55

Figure 8: The weight of a term in a document increases with its frequency in the specific document and decreases with its frequency in the others [3]

COSINE SIMILARITY

How do you measure distance between reports and fixes?

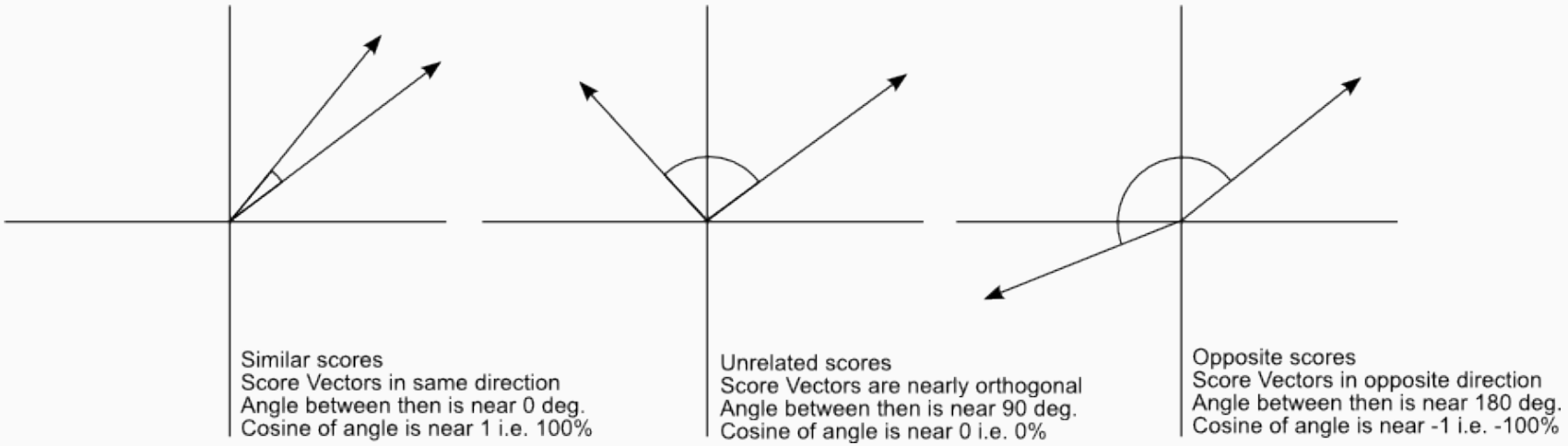


Figure 9: Document similarity as TF-IDF vectors distance

MLink improves over Relink by also taking into account entities with relationships crossing the source-description barrier.

Datasets

Apache, ZXing, OpenIntents

Average scores

Precision	0.56
Recall	0.18
F-score	0.27

Features

- Code metrics
- Entity names
- Term-entity association

Bflinks improves on the previous approaches with more specific time and frequency-based assumptions. Unidirectional links are discarded when taking into account dataset composition.

Datasets

Infopark

Average scores

Precision	0.93
Recall	0.65
F-score	0.76

Features

- ID candidates frequency
- Time intervals
- Bidirectional linking

RCLinker is a more ”machine learning” oriented approach, employing random forests as classifier. Commit descriptions are algoritmically enriched, providing the model with further contextual metadata.

Datasets

CLI, CC, CSV, IO, Lang, Math

Average scores

Precision	0.74
Recall	0.85
F-score	0.79

Features

- 20+ features based on multiple text, timing and priority metrics

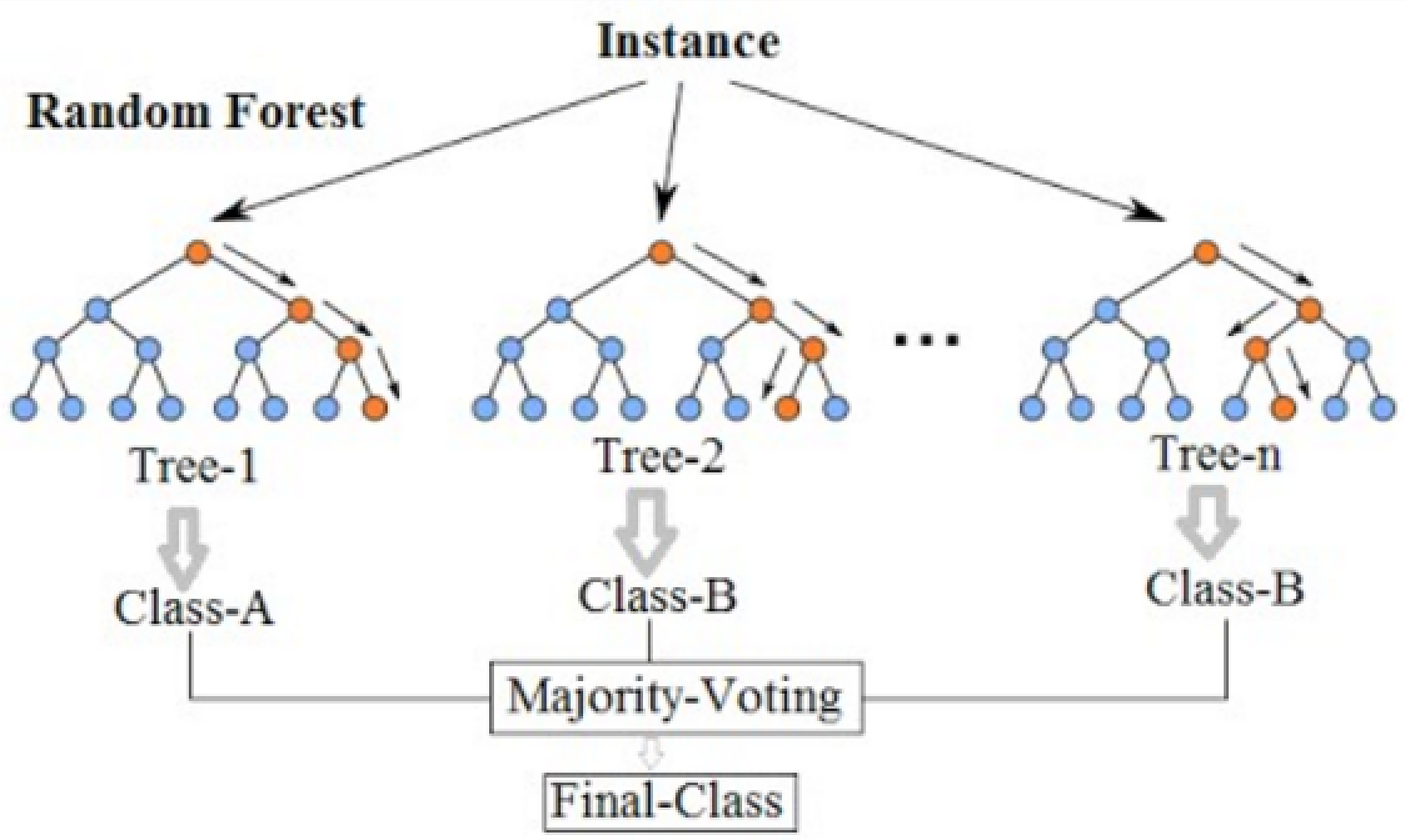


Figure 10: Multiple randomized trees prevent overfitting [2]

FRLink improves dataset composition by extracting additional data from non-source code and by filtering out textually irrelevant candidates by the source files pool.

Datasets

CLI, CC, CSV, IO, Lang, Math

Average scores

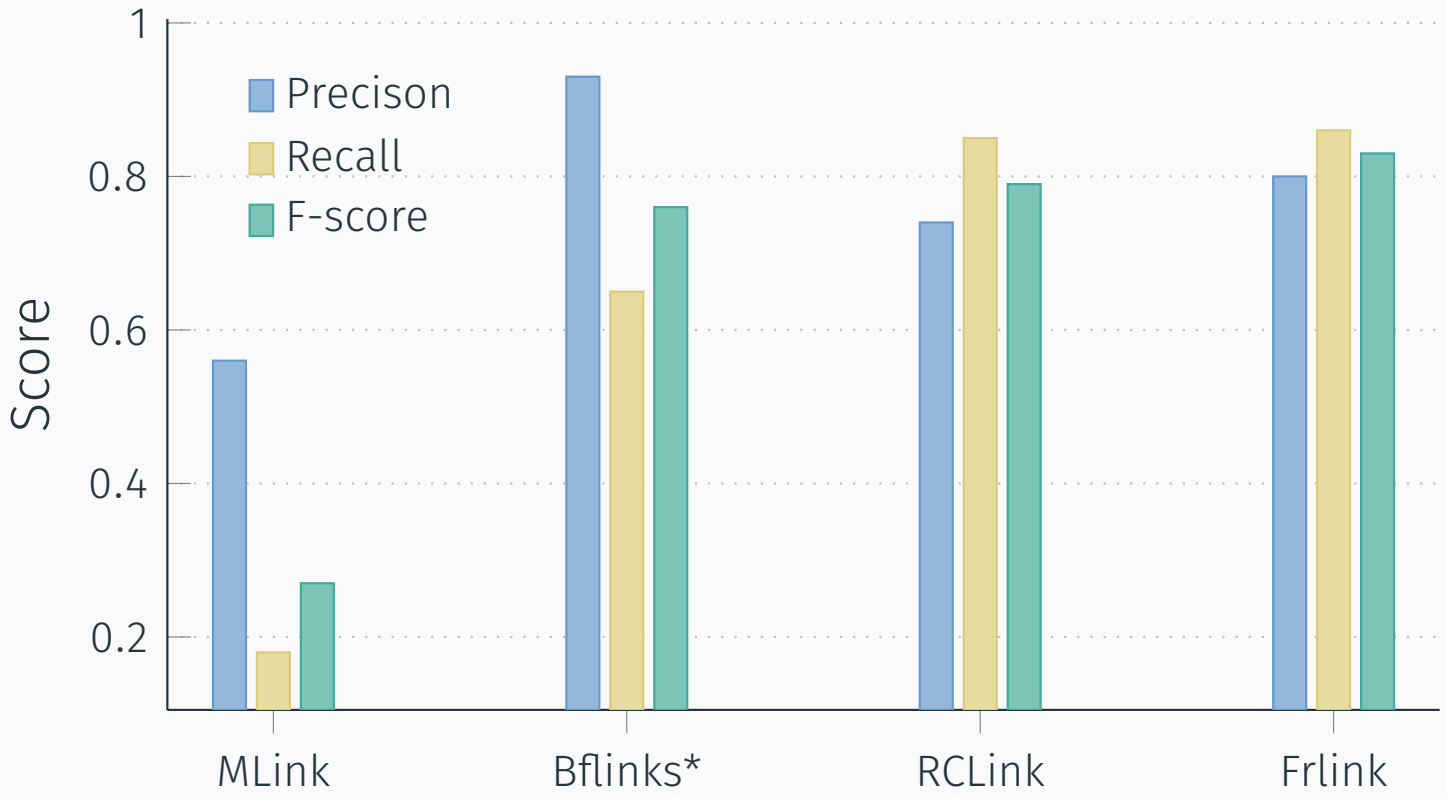
Precision	0.8
Recall	0.86
F-score	0.83

Features

- Comparable to the previous ones

Comparison

PERFORMANCE



* Infopark dataset

Thank you!

CREDITS

Get the source of this theme and the demo presentation from

`github.com/matze/mtheme`

The theme *itself* is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Questions?

REFERENCES I

- Git workflow - https://hackernoon.com/15-tips-to-enhance-your-github-flow-6af7ceb0d8a3.
- Random forests - https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d.
- Tf-idf - victor lavrenko © 2014.
- Walber - https://commons.wikimedia.org/wiki/precisionrecall.
- Xkcd - https://xkcd.com/1296/.

REFERENCES II

T.-D. B. Le, M. Linares-Vásquez, D. Lo, and D. Poshyvanyk.
Rclinker: Automated linking of issue reports and commits leveraging rich contextual information.
In *Proceedings of the 2015 IEEE 23rd International Conference on Program Comprehension*, ICPC '15, pages 36–47, Piscataway, NJ, USA, 2015. IEEE Press.

A. T. Nguyen, T. T. Nguyen, H. A. Nguyen, and T. N. Nguyen.
Multi-layered approach for recovering links between bug reports and fixes.
In *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering*, FSE '12, pages 63:1–63:11, New York, NY, USA, 2012. ACM.

REFERENCES III

L. Prechelt and A. Pepper.
Bflinks: Reliable bugfix links via bidirectional references and tuned heuristics.
In *International scholarly research notices*, 2014.

J. Śliwerski, T. Zimmermann, and A. Zeller.
When do changes induce fixes?
In *Proceedings of the 2005 International Workshop on Mining Software Repositories*, MSR '05, pages 1–5, New York, NY, USA, 2005. ACM.

Y. Sun, Q. Wang, and Y. Yang.
Frlink: Improving the recovery of missing issue-commit links by revisiting file relevance.
Information and Software Technology, 84, 12 2016.

R. Wu, H. Zhang, S. Kim, and S.-C. Cheung.
Relink: Recovering links between bugs and changes.
In *Proceedings of the 19th ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering*, ESEC/FSE '11, pages 15–25, New York, NY, USA, 2011. ACM.