moz://a

Teaching machines to handle bugs and test Firefox more efficiently

Using ML to automate bug management, test selection, and more.

Marco Castelluccio <mcastelluccio@mozilla.com>

FOSDEM 2023

My dogs

Luna

Nika

The Firefox Scale

- Hundreds of bug reports and feature requests per day
  - **1'813'700 bug reports** at this time!

- We release **every 4 weeks** with thousands of changes

- **13 major releases** during 2022
  - 45 minor releases

Whiteboard: [1.7 million bugs, 🎉 🎉] → [1.7 million boogs!, 🎉 🎉]

... ▼
Comment 2 • 4 years ago

with that bug id this bug looks so... intense.

... ▼
Comment 3 • 4 years ago

Yeah, this report should be used for a party (not for work :-)

@Alex: How much did you pay for the #?

The Firefox Scale

- One of the biggest and most complex software
    - With legacy and tech debt (Netscape was open sourced **25 years ago**)

- **800'000 commits** made by **9000 contributors** representing **25'000'000 lines of code**

- **37500 commits last year**

- **1000 unique contributors last year**

The Firefox Scale

	Language	Code Lines	Comment Lines	Comment Ratio	Blank Lines	Total Lines	Total Percentage	
	C++	6,613,111	1,488,254	18.4%	1,290,259	9,391,624		26.9%
	JavaScript	6,250,465	1,822,090	22.6%	1,214,091	9,286,646		26.6%
	HTML	3,554,582	103,565	2.8%	395,422	4,053,569		11.6%
	C	3,365,002	922,272	21.5%	492,233	4,779,507		13.7%
	Rust	2,673,304	498,267	15.7%	261,727	3,433,298		9.8%
	Python	1,106,815	311,689	22.0%	281,827	1,700,331		4.9%
	XML	603,157	9,797	1.6%	20,757	633,711		1.8%
	Assembly	375,139	28,882	7.1%	37,832	441,853		1.3%
	CSS	224,642	13,541	5.7%	31,965	270,148		0.8%
	Java	174,767	66,039	27.4%	27,325	268,131		0.8%
	Autoconf	100,500	1,584	1.6%	15,971	118,055		0.3%
	shell script	83,535	17,118	17.0%	13,283	113,936		0.3%
	Objective-C	54,334	9,289	14.6%	12,032	75,655		0.2%
	Make	43,659	13,187	23.2%	11,634	68,480		0.2%
	OpenGL Shading	35,435	10,713	23.2%	10,874	57,022		0.2%
	TypeScript	27,528	8,303	23.2%	3,082	38,913		0.1%
	Kotlin	23,250	1,121	4.6%	3,217	27,588		0.1%
	CMake	18,504	3,827	17.1%	2,720	25,051		0.1%

TODO
openhub

It's not a bug, it's a feature!

It's not a bug, it's a feature!
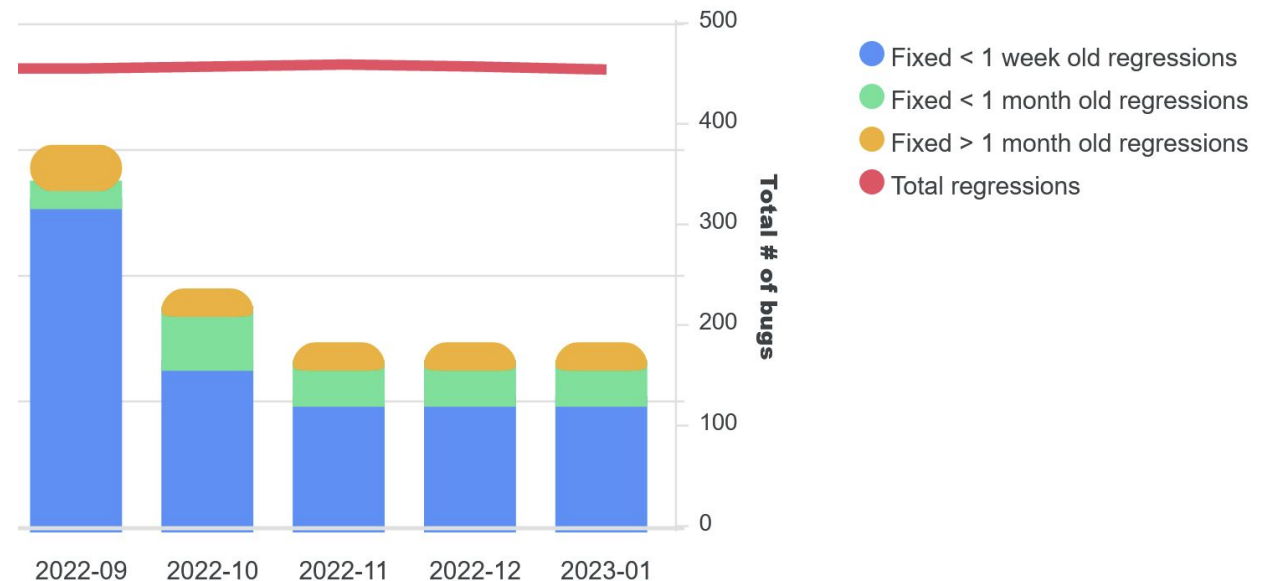
Where it all started

It's not a bug, it's a feature

The Why
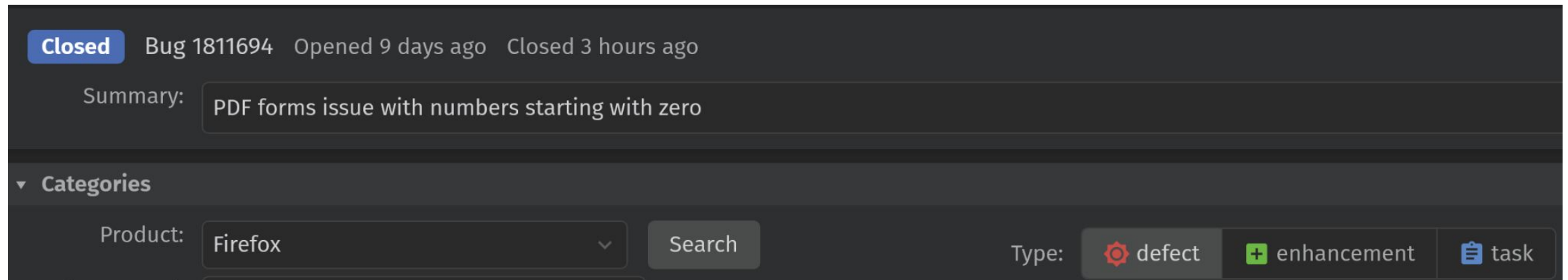
- **No way to differentiate** between defects and feature requests

- **Quality metrics**, e.g. how many defects are found in a given release

- **Harder to improve workflows**

Legend:
- Fixed < 1 week old regressions
- Fixed < 1 month old regressions
- Fixed > 1 month old regressions
- Total regressions

It's not a bug, it's a feature

The Result

- Introduction of a new *type* field

It's not a bug, it's a feature

The How

- ~2500 manually labelled bugs + ~9000 labelled with heuristics

~93% accuracy
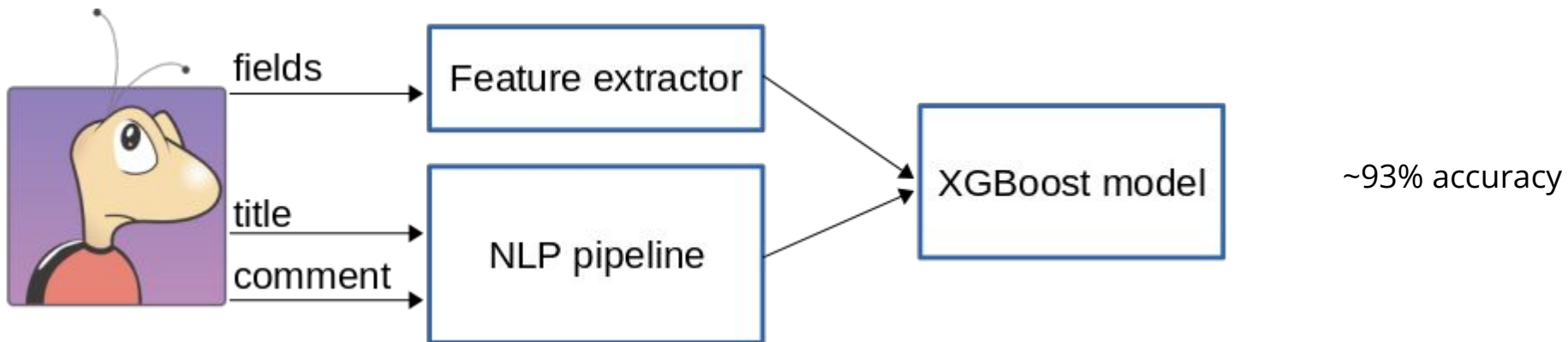
It's not a bug, it's a feature

The Result

bug   bug

https://github.com/mozilla/bugbug
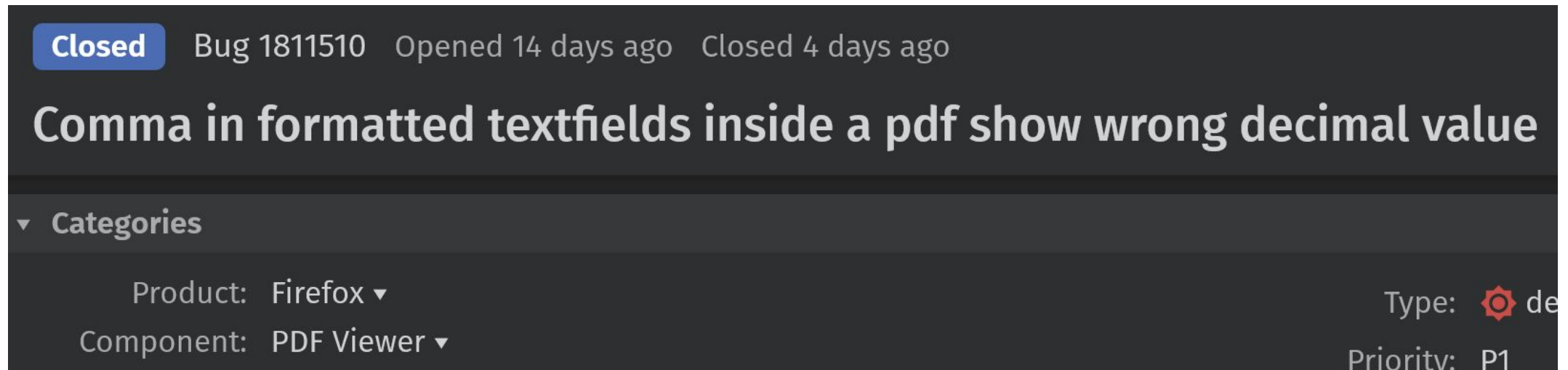
It's not a bug, it's a feature!

Assigning components faster

Assigning components faster

From 7 days to 7 seconds

Assigning components faster

The Why

- Use **Product and Component** to group bugs
- Manually done by volunteers and developers

Assigning components faster

The How

- Large curated dataset: **two decades worth of bugs**!

- **Roll-back of the bug history** to get back to the initial state

- Reduction of dataset to meaningful components
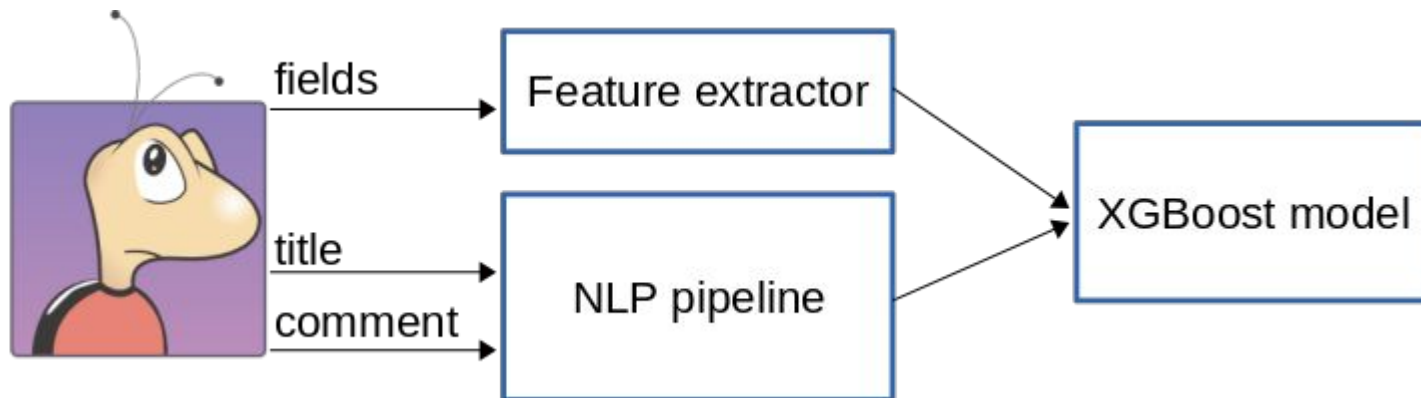
Assigning components faster

The How

- Use **confidence level** to only assign component when we are sure enough
  - Control precision and recall by **tweaking the threshold**

Hi there,
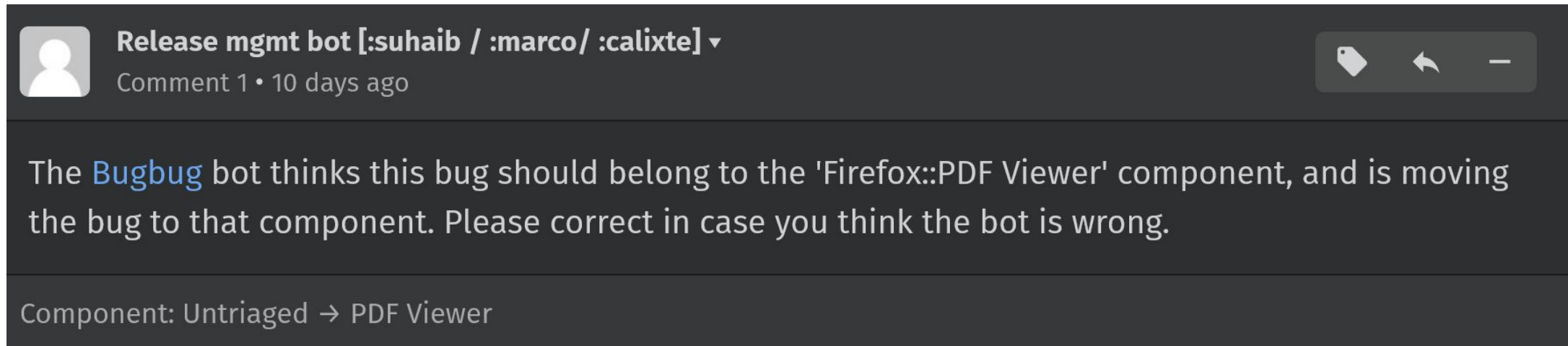
Classified components for the following bugs:

Bug	Summary	Component	Confidence (%)
1591214	Firefox hangs after waking from sleep	Core::Graphics: WebRender	81
1612261	PDF.js does not render borders consistently	Firefox::PDF Viewer	61

Assigning components faster

The Result

- Average time to assign down from ~1 week to a few seconds
- ~**20000 bugs autotriaged**
- Extended to help triage web compatibility issues from <u>webcompat.com</u> (filed on GitHub)

Release mgmt bot [:suhaib / :marco/ :calixte] ▾
Comment 1 • 10 days ago

The Bugbug bot thinks this bug should belong to the 'Firefox::PDF Viewer' component, and is moving the bug to that component. Please correct in case you think the bot is wrong.

Component: Untriaged → PDF Viewer

Building up the architecture

About time

It's not a bug, it's a feature!

Assigning components faster

Building up the architecture

Building up the architecture

The Pipeline
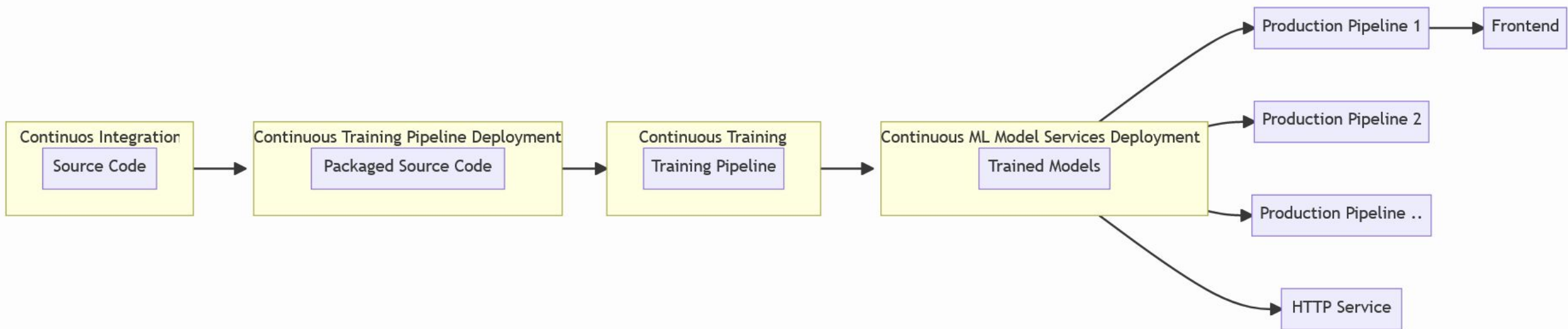
Rooting out spam

Cause nobody likes spam

It's not a bug, it's a feature!

Assigning components faster

Building up the architecture

Rooting out spam

It's not a bug, it's a feature!

Where it all started

It's not a bug, it's a feature!

Assigning components faster

Building up the architecture

Rooting out spam

Can we test more efficiently?

Assigning components faster

From 7 days to 7 seconds

It's not a bug, it's a feature!

Assigning components faster

Building up the architecture

Rooting out spam

Can we test more efficiently?

Predicting defects

Rooting out spam

The Why

- Spammers trying to increase links to their websites

- University students learning how to file bugs :(

- **Built by a contributor, Ayush Shridhar**

1700648 - Firefox Keeps on Crashing Netflix

Alex Henry ▾ [Reporter] [New to Bugzilla]
Description • 1 month ago

✕ spam

User Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/89.0.4389.90 Safari/537.36

Steps to reproduce:

Open Netflix on Mozilla Firefox.

Actual results:

When I open Netflix on my Mozilla Firefox. It keeps on crashing. But, whenever I open another browser it goes smoothly.

Expected results:

Netflix should work smoothly.

Alex Henry ▾ [Reporter] [New to Bugzilla]
Comment 1 • 1 month ago

✕ spam

If Netflix didn't work open: https://thexploretech.com/free-sports-streaming-sites/
Try other sites.

1675284 - Lockwise does not report website breaches or alerts

sriram ▾ [Reporter] [New to Bugzilla]
Description • 6 months ago

User Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:82.0) Gecko/20100101 Firefox/82.0

Steps to reproduce:

Enable "Show alerts about passwords for breached websites" in Options / Privacy & Security.
Go to "about:logins"

Actual results:

Breach alerts and the Sort by: "Alerts" is not displayed.

Expected results:

Breach alerts and sorting by "alerts" should be enabled.

1703459 - Lockwise does not report website breaches or alerts

John Joe ▾ [Reporter] [New to Bugzilla]
Description • 28 days ago

✕ spam

User Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/89.0.4389.114 Safari/537.36

Steps to reproduce:

User Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:82.0) Gecko/20100101 Firefox/82.0

Steps to reproduce: https://careofficecleaning.com/

Enable "Show alerts about passwords for breached websites" in Options / Privacy & Security.
Go to "about:logins"

Actual results:

Breach alerts and the Sort by: "Alerts" is not displayed.

Expected results:

Breach alerts and sorting by "alerts" should be enabled.

Part V

Can we test more efficiently?

Yes, we can

Can we test more efficiently?

Mozilla's CI

- **try** - On-demand CI

- **autoland** - Runs a subset of tests

- **mozilla-central** - Firefox Nightly is built from here, runs the full set of tests, merges from autoland + urgent fixes

Can we test more efficiently?

Mozilla's CI

- Around **100'000 unique test files**

- Over **150 unique configurations**, from combinations of:
  - Each major OS
  - High-level Firefox configurations (e.g. old style engine / new style engine)
  - Debug / Opt
  - ASAN / TSAN / Coverage / ...

- More than **300 pushes per day**, with the average full push taking **1500 hours**
  - 300 machine years per month
  - ~1 million machines per month

- Running all means running **2.3 billion test files per day**!

Can we test more efficiently?

Subset of test tasks

Can we test more efficiently?

The Difficulties

https://twitter.com/mozsheriffmemes

- Hard to pinpoint the cause of a given test failure
  - Lots of tests and **lots of intermittent** (flaky) tests
  - Not all tests are run on all pushes
  - We have **sheriffs** watching the CI, but human errors exist

Can we test more efficiently?

Patch Characteristics

- To link tests to patches, we need to analyze them

- mercurial for authorship info

- <u>rust-parsepatch</u> to parse patches efficiently

- <u>rust-code-analysis</u> to analyze source code (research partnership with Politecnico di Torino)

Can we test more efficiently?

The ML model

- Avoid huge multi-label model (where each test is a label): **input** is **tuple TEST / PATCH**, **label** is **"fail" / "not fail"**

- Features from both test, patch and test / patch, e.g.:
  - Past failures when the same files were touched
  - Distance from source files to test files
  - How often source files were modified together with test files

- Use *frequent itemset mining* to **remove test redundancies**

Part VI

Predicting defects

Can we?

Predicting defects

The Why

- **Reduce regressions**, by detecting riskier patches to scrutinize more

- **Reduce time spent** by reviewers **on less risky patches**

- Trigger **risk-control** operations for risky patches (e.g. code coverage)

- NOTE: **AI in support** of developers, will never replace a real reviewer

Predicting defects

Finding patches causing regressions

- **5'348 links** between bug-introducing and bug-fixing commit-sets (totaling **24'089 commits**)

- Research study on dataset and SZZ in collaboration with University of Zurich

Predicting defects

microannotate

```
161
162        skb->dev = dev;
163        skb_reset_mac_header(skb);
164
165        eth = (struct ethhdr *)skb->data;
166        skb_pull_inline(skb, ETH_HLEN);
167
168        if (unlikely(is_multicast_ether_addr_64bits(eth->h_dest))) {
169                if (ether_addr_equal_64bits(eth->h_dest, dev->broadcast))
170                        skb->pkt_type = PACKET_BROADCAST;
171                else
172                        skb->pkt_type = PACKET_MULTICAST;
173        }
174        else if (unlikely(!ether_addr_equal_64bits(eth->h_dest,
175                                                   dev->dev_addr)))
176                skb->pkt_type = PACKET_OTHERHOST;
177
```

Commit:
d54385ce68cd18ab002b

Alexander Duyck

Apr 30 2015

etherdev: Process
is_multicast_ether_addr
at same size as other
operations

Predicting defects

Result

Diff Risk Analysis - Risky with 79% confidence

increasing risk →   ← decreasing risk

I. **# of backouts in these directories** is large (2375), as in 56% of patches not introducing regressions. Show feature plot

II. **# of non-code lines added** (7) Show feature plot

III. **# of lines added in tests** is small (0), as in 57% of patches not introducing regressions. Show feature plot

IV. **Author seniority** is too small (154), as in 56% of patches introducing regressions. Hide feature plot

V. **Non-code files size** is too large (7592), as in 55% of patches introducing regressions. Show feature plot

See the most important features considered by the model

Predicting defects

Function-level Result

```
5333        aParams.offset += aParams.defaultLineThickness;          5333        aParams.offset += aParams.defaultLineThickness;
5334        aParams.offset -= aOffset.AsLength().ToAppUnits() / aAppUnitsPerDevPixel   5334        aParams.offset -= aOffset.AsLength().ToAppUnits() / aAppUnitsPerDevPixel;
     ;
5335    } else if (aRightUnderline) {                               5335    } else if (aRightUnderline) {

         ▲ Show 20 Lines • Show All 67 Lines • ▼ Show 20 Line(s)    5402        Float gfxWidth = (verticalRun ? aVisualOverflowRect->height
```

Risk Analysis Bot

The function 'nsTextFrame::UnionAdditionalOverflow' is risky (56% confidence).

Recent past bugs in this function:

- Bug 1426092 - wide text-strokes may be cut off at the bottom and right edges
- Bug 1529992 - Text shadow is improperly clipped

```
5403                          : aVisualOverflowRect->width) /       5403                          : aVisualOverflowRect->width) /
5404                    appUnitsPerDevUnit;                          5404                    appUnitsPerDevUnit;
5405    params.lineSize = Size(gfxWidth, underlineSize / appUnitsPerDevUnit);   5405    params.lineSize = Size(gfxWidth, underlineSize / appUnitsPerDevUnit);
5406    params.ascent = gfxFloat(mAscent) / appUnitsPerDevUnit;      5406    params.ascent = gfxFloat(mAscent) / appUnitsPerDevUnit;
5407    params.style = decorationStyle;                              5407    params.style = decorationStyle;
5408    params.vertical = verticalRun;                               5408    params.vertical = verticalRun;
5409    params.sidewaysLeft = mTextRun->IsSidewaysLeft();            5409    params.sidewaysLeft = mTextRun->IsSidewaysLeft();
5410    params.offset = underlineOffset / appUnitsPerDevUnit;        5410    params.offset = underlineOffset / appUnitsPerDevUnit;
5411    params.decoration = StyleTextDecorationLine_UNDERLINE;       5411    params.decoration = StyleTextDecorationLine::UNDERLINE;
```

Part VII

Privacy-friendly translations

Traduzioni con rispetto per la privacy*

* this was automatically translated using Firefox Translate.

Privacy-friendly translations

The Why

- Translation models improved a lot in recent times

- Current cloud-based services do not offer the **privacy guarantees** that we like to offer in Firefox

- **European Union 🇪🇺 funded project** to investigate client-side private translation capabilities

Privacy-friendly translations

Current Status

- Available as an **extension**
  - Firefox Translations – Get this Extension for 🦊 Firefox

- **Languages supported**
  - Spanish, Estonian, English, German, Czech, Bulgarian, Portuguese, Italian, French, Polish

- **In development**
  - Russian, Persian (Farsi), Icelandic, Norwegian Nynorsk, Norwegian Bokmål, Ukrainian, Dutch

Privacy-friendly translations

How Does It Work

- Use **machine learning on the client side** to do a local translation

- All the data remains on the system

- Firefox will download the model from our servers - just for the languages that you need

Privacy-friendly translations

Open Datasets

Search & download resources: [es (Spanish) ▾] [en (English) ▾] [all ▾] ☐ show all versions

Language resources: click on [tmx | moses | xces | lang-id] to download the data! (raw = untokenized, ud = parsed with universal dependencies, alg = word alignments and phrase tables)

| corpus	doc's	sent's	en tokens	es tokens	XCES/XML	raw	TMX	Moses	mono	raw	ud	alg	dic	freq		other files
CCMatrix v1	1	409.1M	7.7G	8.4G	xces en es	en es	tmx	moses	en es	en es				en es	sample	
WikiMatrix v1	1	6.5M	1.0G	381.7M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
UNPC v1.0	127036	20.0M	523.8M	607.4M	xces en es	en es	tmx	moses	en es	en es		alg		en es	sample	
ParaCrawl v8	161	16.0M	326.3M	348.0M	xces en es	en es	tmx	moses	en es	en es				en es		
MultiUN v1	73047	9.3M	254.4M	300.4M	xces en es	en es	tmx	moses	en es	en es		alg		en es query	sample	
EUbookshop v2	7637	5.3M	220.3M	217.6M	xces en es	en es	tmx	moses	en es	en es			dic	en es query	sample	moses/strict
wikimedia v20210402	1	1.3M	349.2M	75.8M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
CCAligned v1	1967	15.2M	161.0M	172.6M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
TildeMODEL v2018	6	3.8M	134.1M	102.3M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
EuroPat v3	1	44.1M	88.1M	88.1M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
DGT v2019	38658	3.6M	73.9M	84.7M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
Wikipedia v1.0	5	1.8M	44.9M	41.2M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es query	sample	
JRC-Acquis v3.0	11990	0.8M	34.0M	36.2M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
Europarl v8	656	1.3M	32.8M	34.2M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
XLEnt v1.1	1	8.7M	28.7M	27.6M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
QED v2.0a	13171	1.4M	21.2M	18.8M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
SciELO v1	2	0.4M	11.9M	13.2M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
ELITR-ECA v1	1124	0.4M	11.0M	12.4M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
TED2020 v1	3961	0.4M	8.3M	7.9M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
GlobalVoices v2018q4	32844	0.4M	6.7M	7.2M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
News-Commentary v16	8130	0.2M	5.4M	6.3M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
GNOME v1	2292	0.9M	5.6M	5.9M	xces en es	en es	tmx	moses	en es	en es		alg smt		en es	sample	
EMEA v3	1938	0.4M	5.3M	6.0M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es query	sample	
ELRC_2682 v1	1	0.2M	4.2M	4.8M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
UN v20090831	1	74.1k	3.7M	4.0M	xces en es	en es	tmx	moses	en es	en es		alg smt		en es query	sample	
ECB v1	1	0.1M	3.1M	3.5M	xces en es	en es	tmx	moses	en es	en es		alg		en es query	sample	
TED2013 v1.1	1	0.2M	3.1M	2.9M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es query	sample	
OpenSubtitles v2018	538	0.5M	3.1M	2.8M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es query	sample	xces/alt
Tanzil v1	15	0.1M	2.8M	2.4M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es query	sample	
KDE4 v2	2055	0.2M	2.3M	2.2M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es query	sample	
Books v1	18	92.8k	2.1M	2.0M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es query	sample	
bible-uedin v1	2	61.5k	1.5M	1.4M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
Tatoeba v2022-03-03	1	0.2M	1.4M	1.3M	xces en es	en es	tmx	moses	en es	en es				en es	sample	
Mozilla-I10n v1	1	93.0k	0.7M	0.6M	xces en es	en es			en es	en es				en es	sample	
OpenOffice v3	1738	38.6k	0.4M	0.5M	xces es	es			es	es				es		
WMT-News v2019	7	16.6k	0.4M	0.4M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
PHP v1	3217	38.3k	0.5M	0.2M	xces en es	en es	tmx	moses	en es	en es		alg smt		en es query	sample	
infopankki v1	245	12.9k	0.2M	0.2M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
ELRA-W0147 v1	1	32.2k	0.2M	0.2M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
EUconst v1	47	10.2k	0.2M	87.3k	xces en es	en es	tmx	moses	en es	en es		alg		en es query	sample	
ELRC_2922 v1	1	5.1k	0.1M	0.1M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
ELRC_3382 v1	1	3.7k	84.9k	0.1M	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
tico-19 v2020-10-28	1	3.1k	70.4k	85.5k	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
ELRC_2923 v1	1	0.5k	17.9k	20.5k	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
ELRA-W0305 v1	1	0.8k	10.4k	11.4k	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
ELRC_416 v1	1	0.4k	3.2k	3.3k	xces en es	en es	tmx	moses	en es	en es		alg smt	dic	en es	sample	
RF v1	1	25	0.2k	0.2k	xces en es	en es	tmx	moses	en es	en es		alg smt		en es query	sample	
Ubuntu v14.10					xces en es	en es	tmx	moses	en es	en es			dic	en es	sample	
total	332526	553.1M	11.1G	11.0G	553.1M		547.6M	622.0M							sample	

Privacy-friendly translations

Data Cleaning and Augmentation

- Basic **data cleaning** + ML-based detection of bad sentence pairs (<u>bicleaner</u>)

- **Back-translation** for data augmentation from monolingual data
  - Translate target to source
  - Add pair "noisy translated source" -> "good target"

🇮🇹 Mi piace mangiare

🇫🇷 J'aime manger

🇮🇹 Amo mangiare
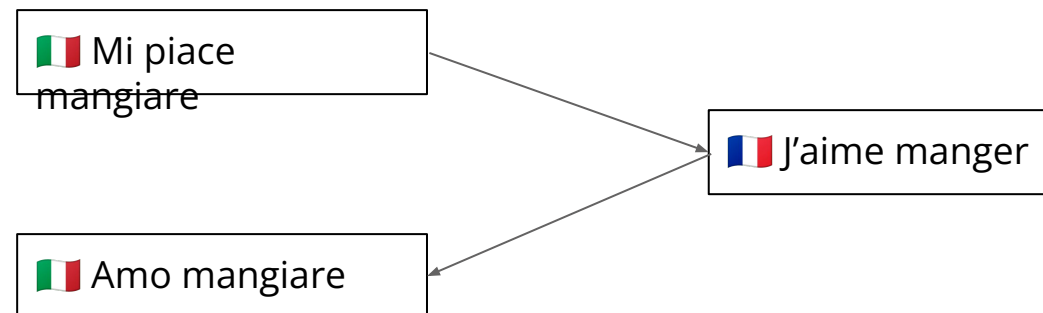
Privacy-friendly translations

Model Training and Compression

- **Training** of a large model

- Compression using **knowledge distillation** into a smaller one

- **Quantization** (float32 -> int8) for further compression and perf improvement
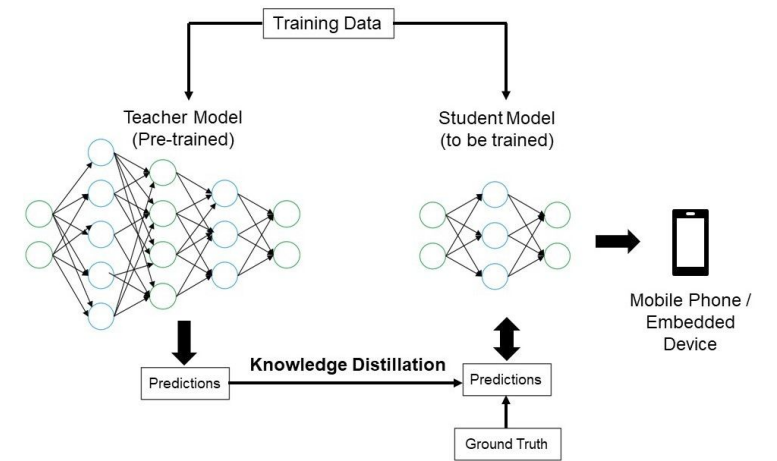
Model	Size	Total number of parameters	Dataset decoding time on 1 CPU core	Quality, BLEU
Teacher	798Mb	192.75M	631s	52.5
Student quantized	17Mb	15.7M	17.9s	50.7

Privacy-friendly translations

Firefox Integration via WASM

- NMT engine compiled to **WebAssembly**

- Using **SIMD instructions**
  - 10x perf improvement

- Engine and models downloaded and updated **on demand**

Privacy-friendly translations

Demo

Privacy-friendly translations

How Can You Help

- **Know any dataset** that we could use?

- Interested in contributing in building a great new feature in Firefox or **adding support for your language**?

- **Come and talk to us at our booth!**

My dogs (now grown)

moz://a

Thank You

Summary

1. It's not a bug, it's a feature!

2. Assigning components faster

3. Building up the architecture

4. Rooting out spam

5. Can we test more efficiently?

6. Predicting defects

7. Privacy-friendly translations