UNIVERSITY OF SASKATCHEWAN

It Will Never Work in Theory

April 2023

Technical Debt in R Packages

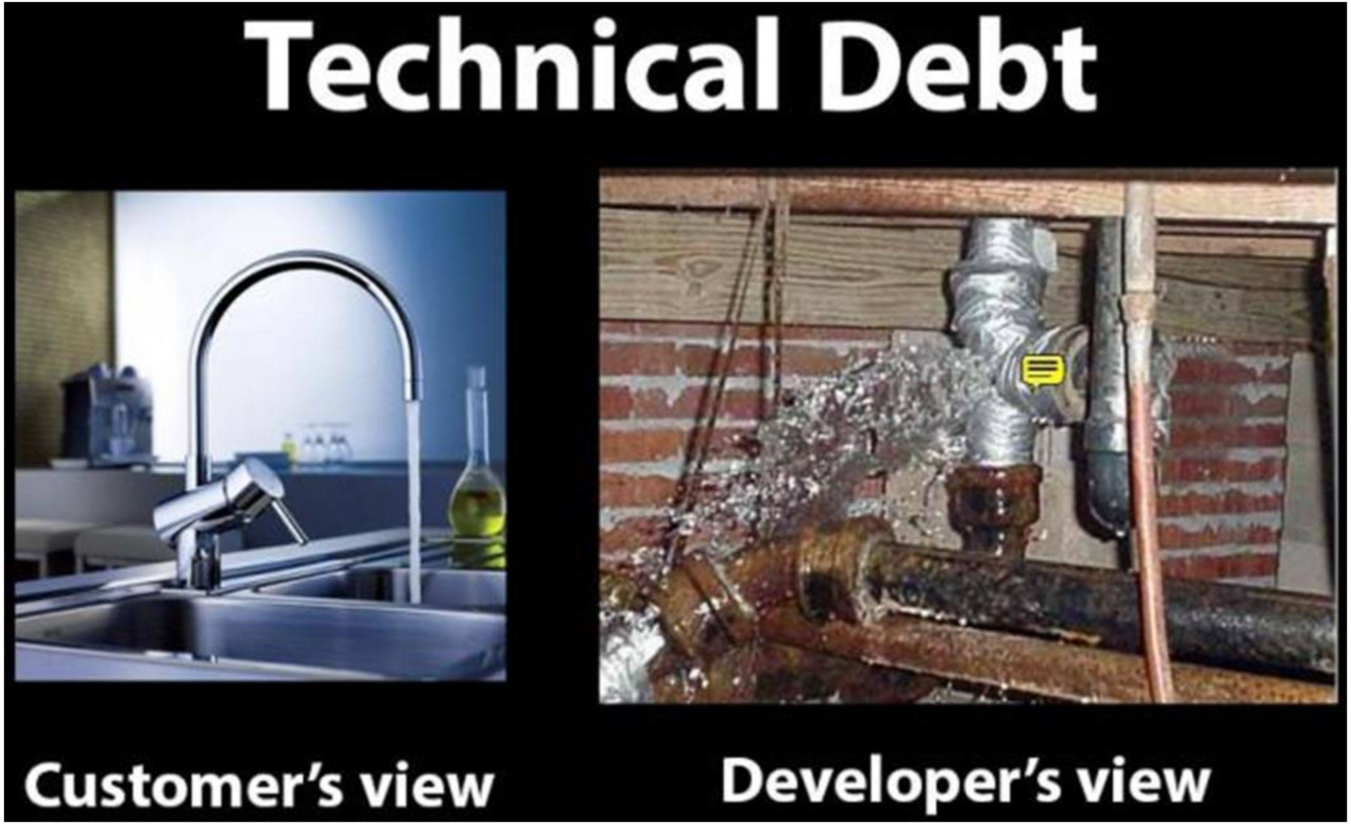
Zadia Codabux

University of Saskatchewan, Canada

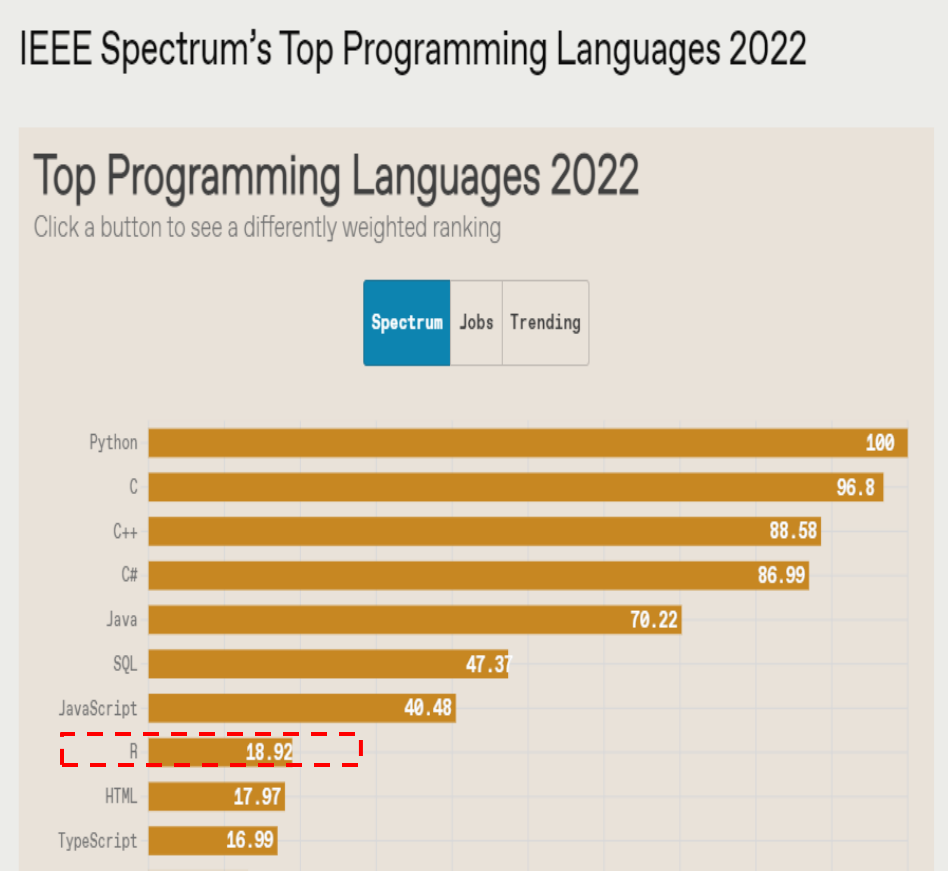
✉ zadiacodabux@ieee.org

@zadiacodabux

https://www.cs.usask.ca/faculty/zadiacodabux/



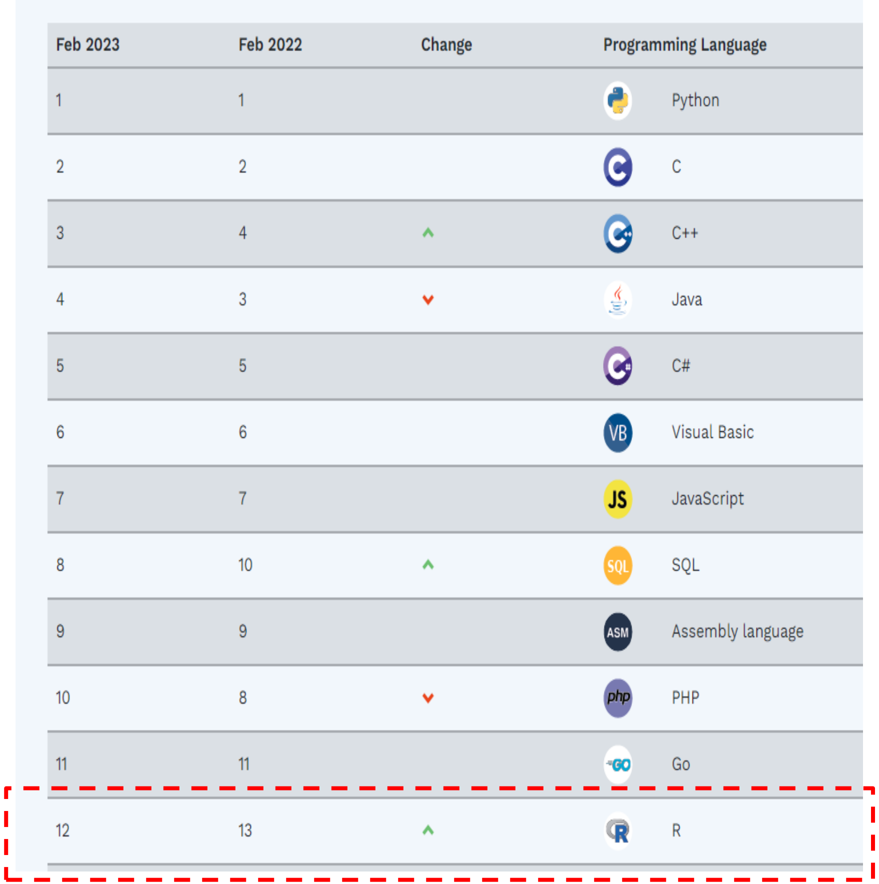
Source: Google Images



IEEE Spectrum's Top Programming Languages 2022

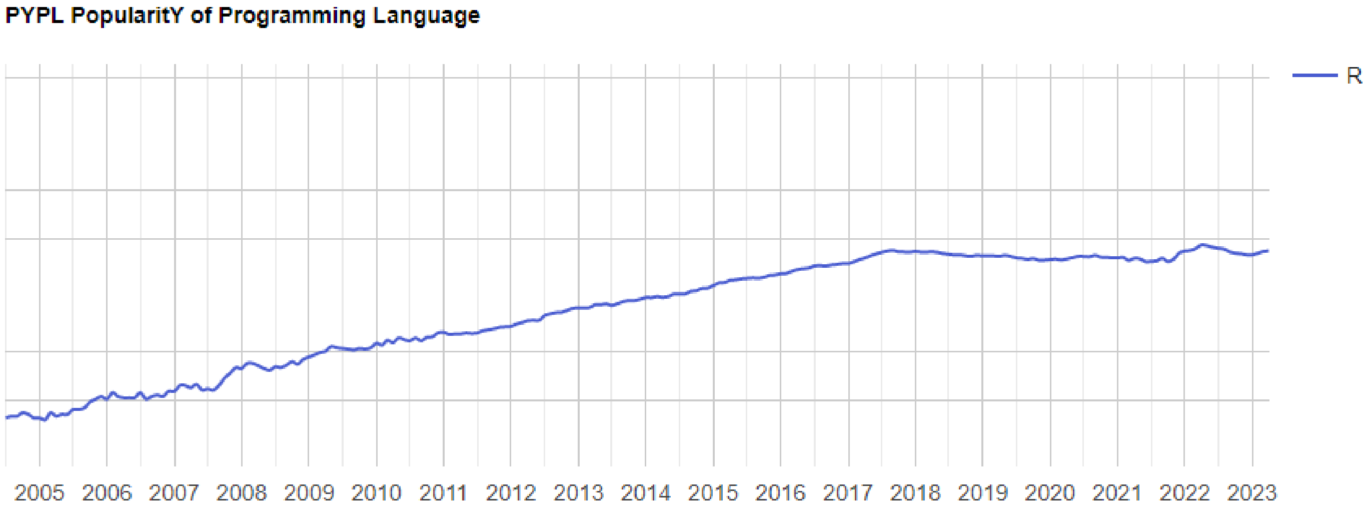
Top Programming Languages 2022

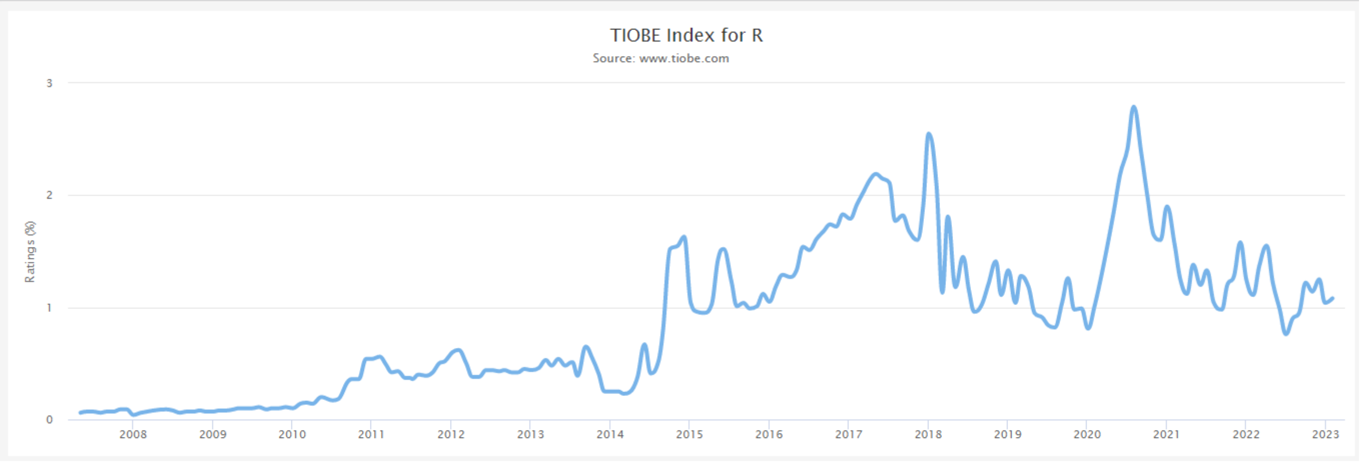
Click a button to see a differently weighted ranking



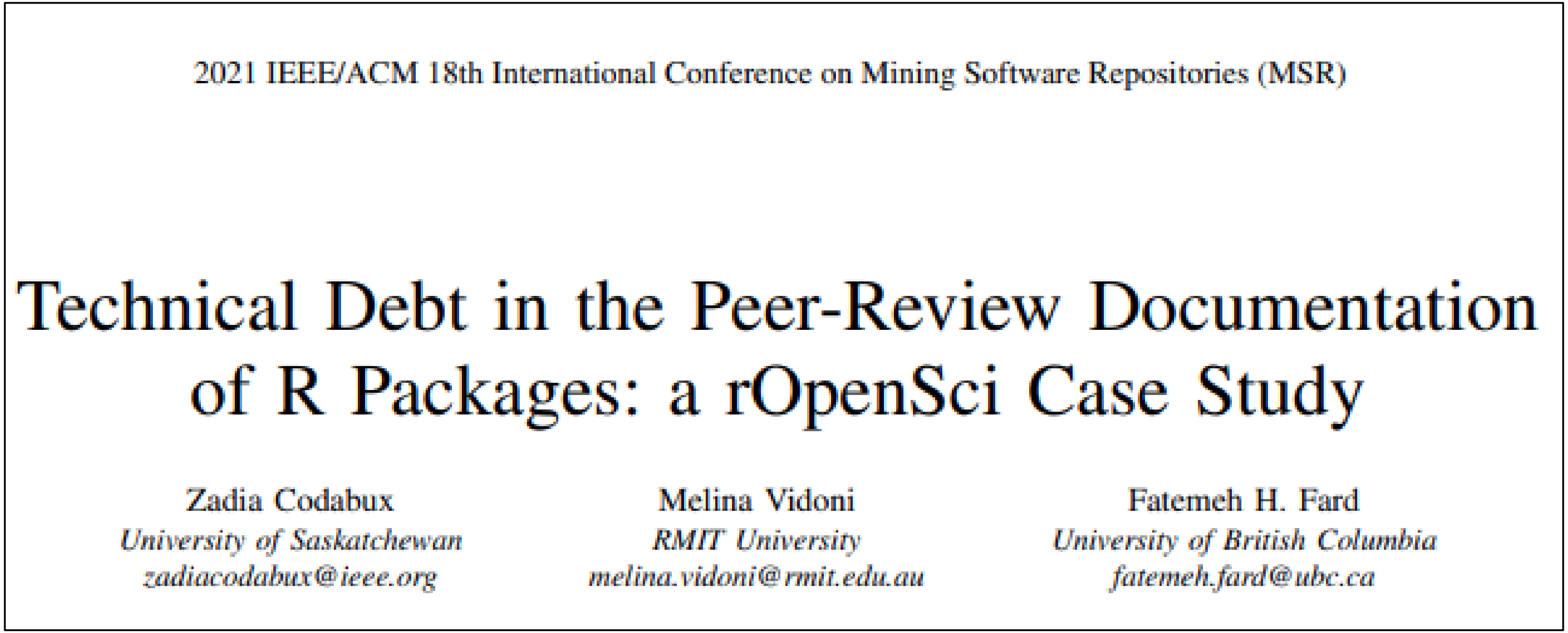
Feb 2023	Feb 2022	Change	Programming Language
1	1		Python
2	2		C
3	4	▲	C++
4	3	▼	Java
5	5		C#
6	6		Visual Basic
7	7		JavaScript
8	10	▲	SQL
9	9		Assembly language
10	8	▼	PHP
11	11		Go
12	13	▲	R

https://www.tiobe.com/tiobe-index/
https://pypl.github.io/PYPL.html
https://spectrum.ieee.org/top-programming-languages-2022





https://www.tiobe.com/tiobe-index/
https://pypl.github.io/PYPL.html
https://spectrum.ieee.org/top-programming-languages-2022



2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)

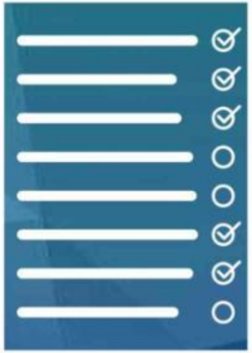
Technical Debt in the Peer-Review Documentation of R Packages: a rOpenSci Case Study

Zadia Codabux
University of Saskatchewan
zadiacodabux@ieee.org

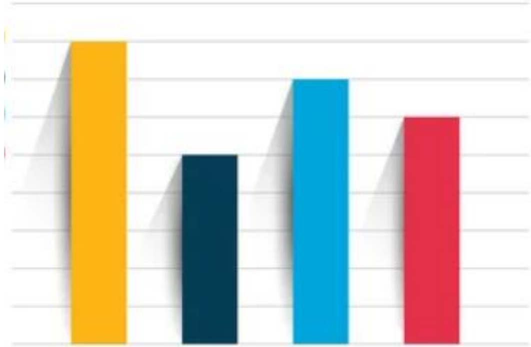
Melina Vidoni
RMIT University
melina.vidoni@rmit.edu.au

Fatemeh H. Fard
University of British Columbia
fatemeh.fard@ubc.ca

Goal: To investigate Technical Debt (TD) in the documentation of the peer-review process of R packages.



TD Types



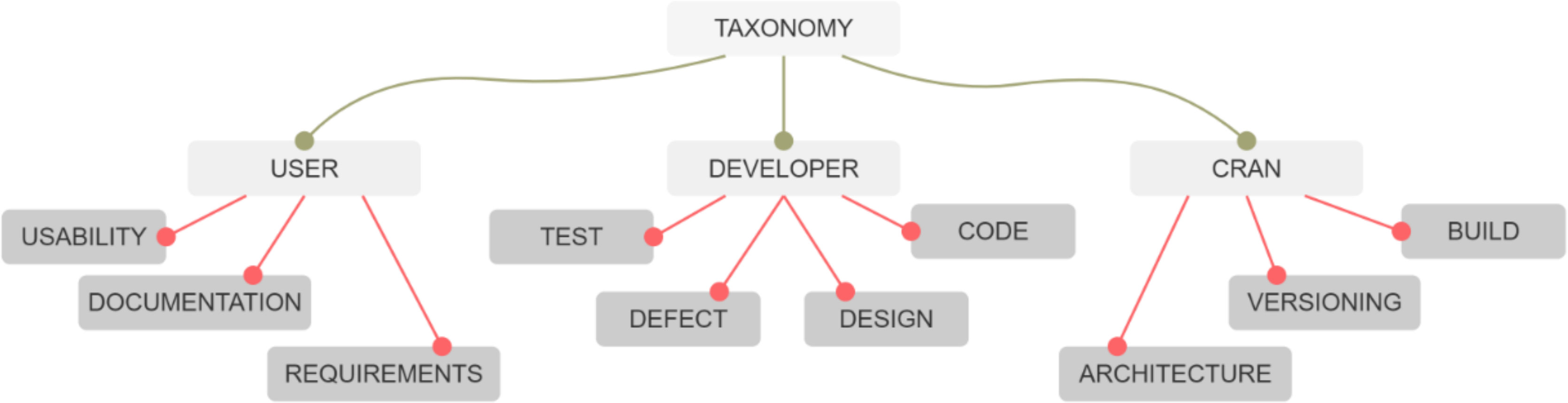
TD Types Distribution

TD Types based on User Roles

Zadia Codabux, Melina Vidoni, Fatemeh H. Fard, Technical Debt in the Peer-Review Documentation of R Packages: a rOpenSci Case Study, Mining Software Repositories Conference, 2021

Results

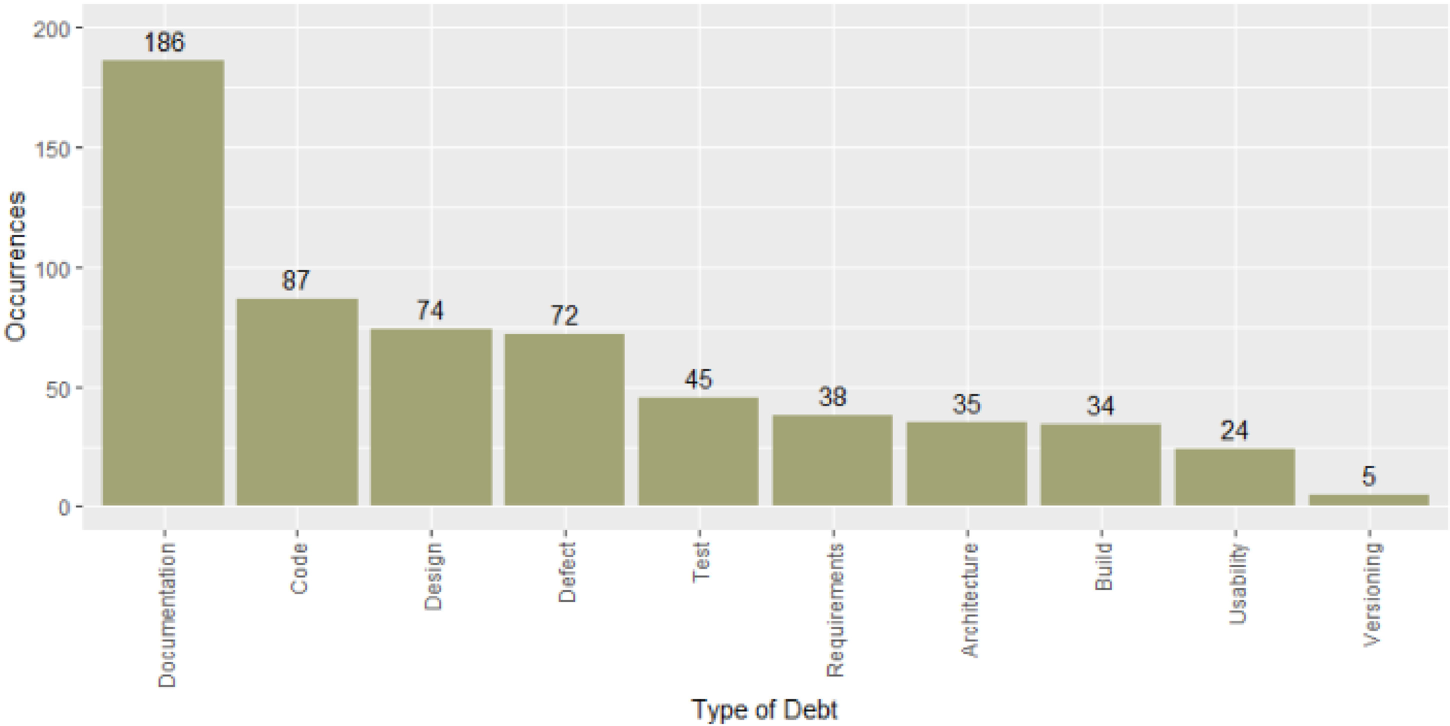
TD Types



Zadia Codabux, Melina Vidoni, Fatemeh H. Fard, Technical Debt in the Peer-Review Documentation of R Packages: a rOpenSci Case Study, Mining Software Repositories Conference, 2021

Results

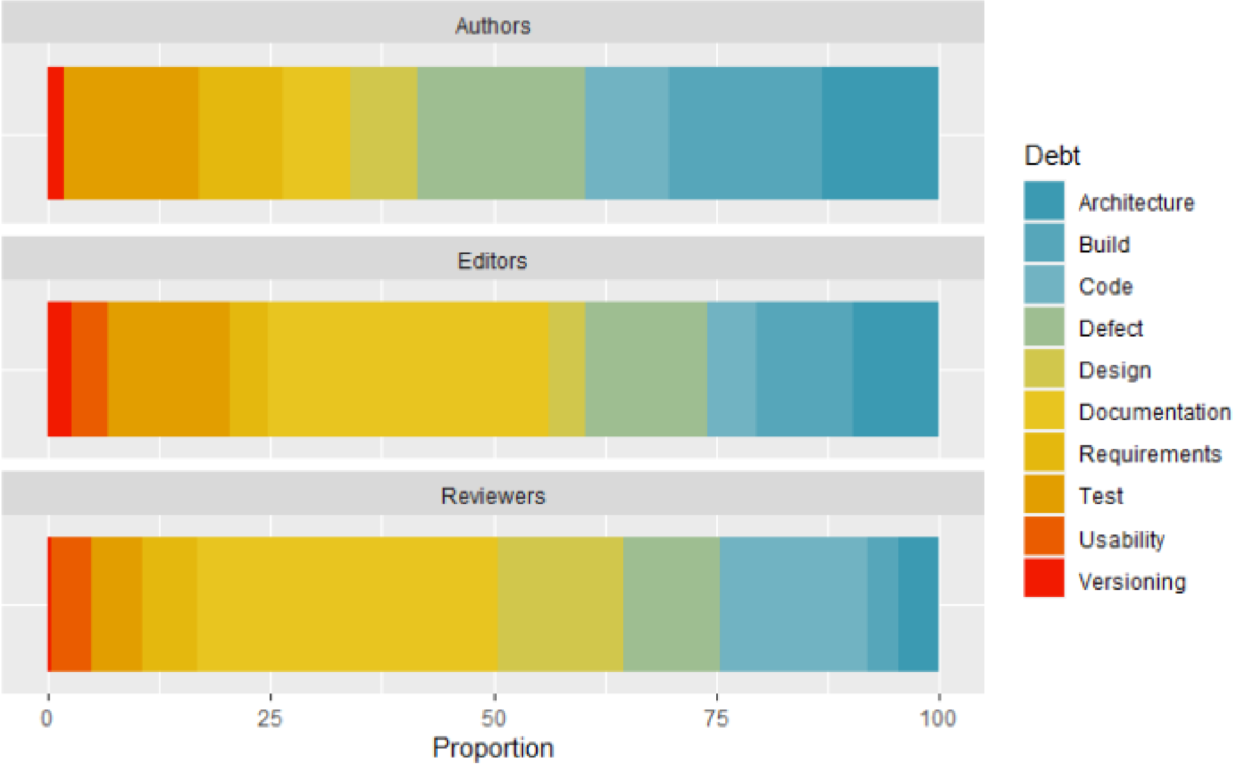
TD Types Distribution



Zadia Codabux, Melina Vidoni, Fatemeh H. Fard, Technical Debt in the Peer-Review Documentation of R Packages: a rOpenSci Case Study, Mining Software Repositories Conference, 2021

Results

TD Types based on User Roles



Zadia Codabux, Melina Vidoni, Fatemeh H. Fard, Technical Debt in the Peer-Review Documentation of R Packages: a rOpenSci Case Study, Mining Software Repositories Conference, 2021

Automated Software Engineering (2022) 29:53
https://doi.org/10.1007/s10515-022-00358-6

Self-admitted technical debt in R: detection and causes

Rishab Sharma[1] · Ramin Shahbazi[1] · Fatemeh H. Fard[1] · Zadia Codabux[2] ·
Melina Vidoni[3]

Goal: Automatic detection and causes of Self Admitted TD (SATD) in R Packages

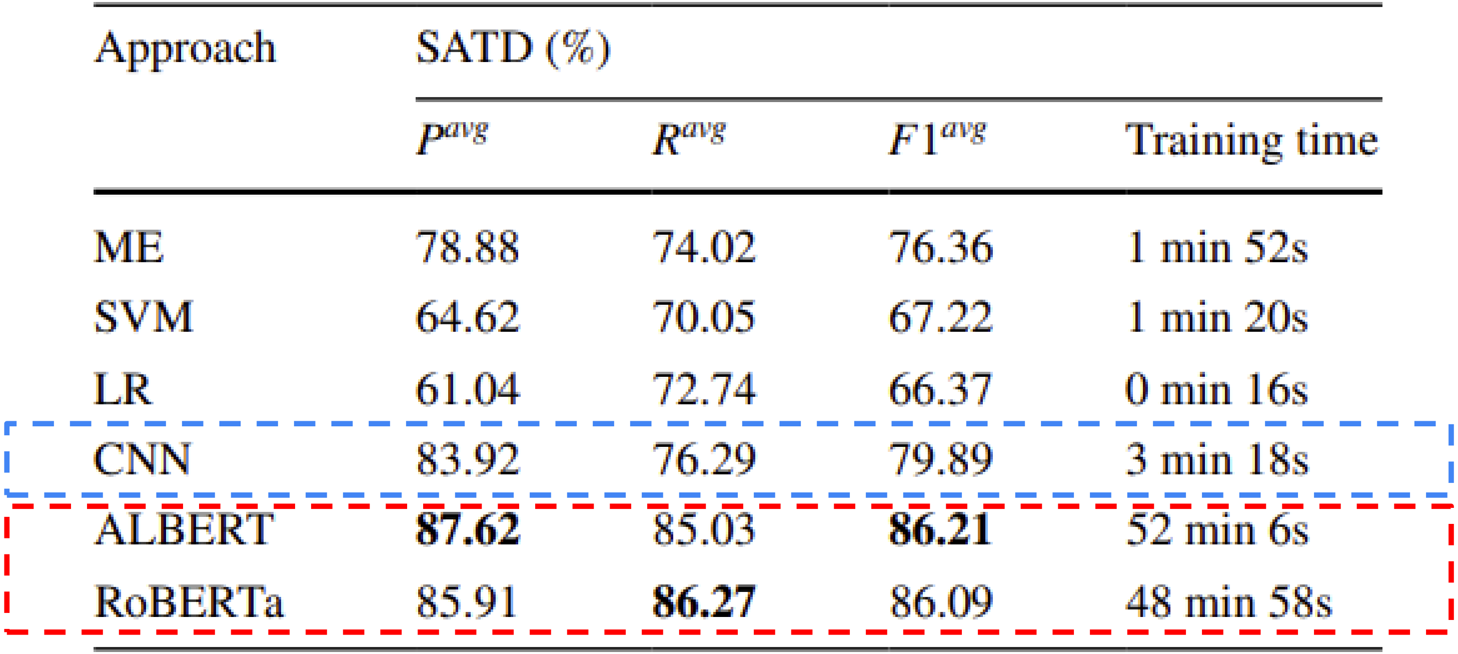
Best Performing Model:
SATD & SATD Types
Detection

Causes of SATD

Rishab Sharma, Ramin Shahbazi, Fatemeh H. Fard, Zadia Codabux, Melina Vidoni, Self-Admitted Technical Debt in R: Detection and Causes, Automated Software Engineering, 2022

Results

SATD Detection

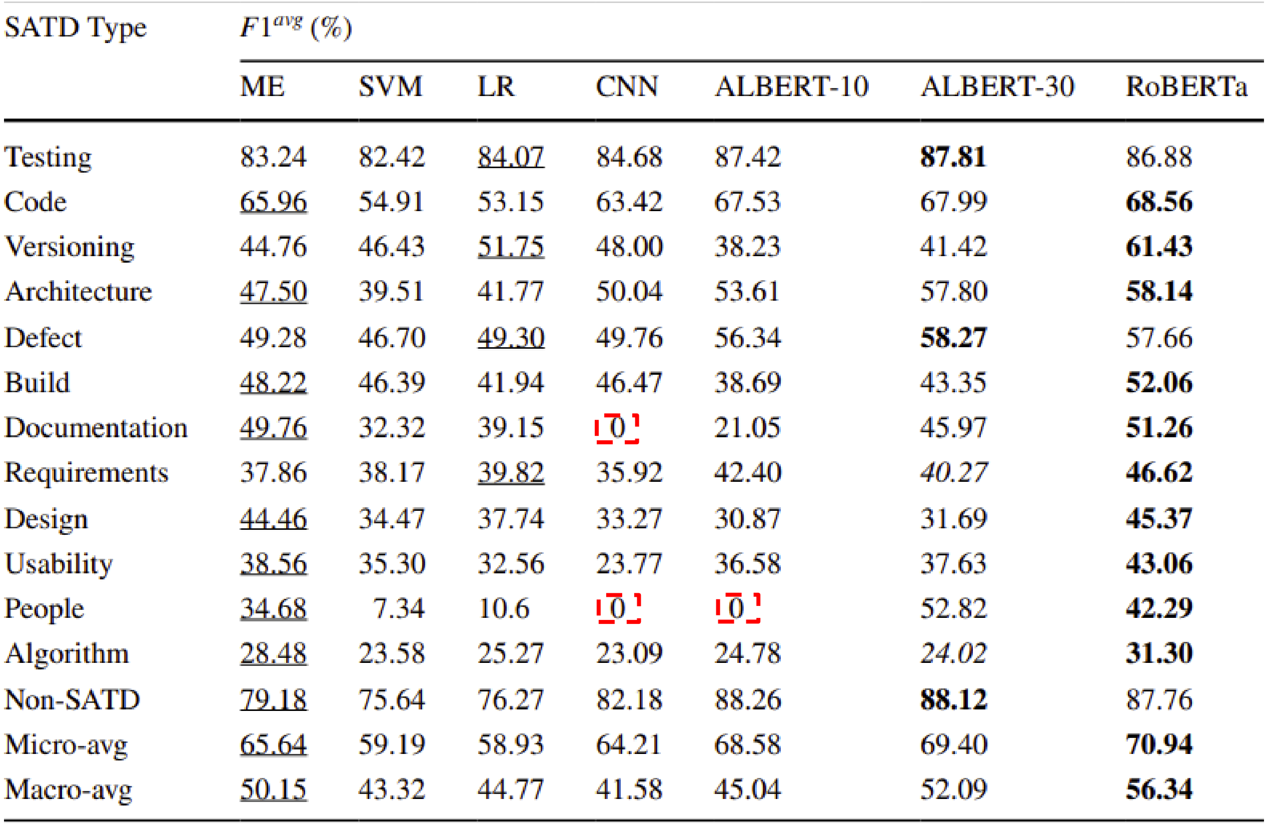


Approach	SATD (%)			
	P^{avg}	R^{avg}	$F1^{avg}$	Training time
ME	78.88	74.02	76.36	1 min 52s
SVM	64.62	70.05	67.22	1 min 20s
LR	61.04	72.74	66.37	0 min 16s
CNN	83.92	76.29	79.89	3 min 18s
ALBERT	**87.62**	85.03	**86.21**	52 min 6s
RoBERTa	85.91	**86.27**	86.09	48 min 58s

Rishab Sharma, Ramin Shahbazi, Fatemeh H. Fard, Zadia Codabux, Melina Vidoni, Self-Admitted Technical Debt in R: Detection and Causes, Automated Software Engineering, 2022

Results

SATD Types Detection



SATD Type	$F1^{avg}$ (%)						
	ME	SVM	LR	CNN	ALBERT-10	ALBERT-30	RoBERTa
Testing	83.24	82.42	84.07	84.68	87.42	**87.81**	86.88
Code	65.96	54.91	53.15	63.42	67.53	67.99	**68.56**
Versioning	44.76	46.43	51.75	48.00	38.23	41.42	**61.43**
Architecture	47.50	39.51	41.77	50.04	53.61	57.80	**58.14**
Defect	49.28	46.70	49.30	49.76	56.34	**58.27**	57.66
Build	48.22	46.39	41.94	46.47	38.69	43.35	**52.06**
Documentation	49.76	32.32	39.15	0	21.05	45.97	**51.26**
Requirements	37.86	38.17	39.82	35.92	42.40	*40.27*	**46.62**
Design	44.46	34.47	37.74	33.27	30.87	31.69	**45.37**
Usability	38.56	35.30	32.56	23.77	36.58	37.63	**43.06**
People	34.68	7.34	10.6	0	0	52.82	**42.29**
Algorithm	28.48	23.58	25.27	23.09	24.78	*24.02*	**31.30**
Non-SATD	79.18	75.64	76.27	82.18	88.26	**88.12**	87.76
Micro-avg	65.64	59.19	58.93	64.21	68.58	69.40	**70.94**
Macro-avg	50.15	43.32	44.77	41.58	45.04	52.09	**56.34**

Rishab Sharma, Ramin Shahbazi, Fatemeh H. Fard, Zadia Codabux, Melina Vidoni, Self-Admitted Technical Debt in R: Detection and Causes, Automated Software Engineering, 2022

Takeaways

- Documentation debt is the most recurrent, yet the least valued
- Not all users give the same importance to the different debt types
- Challenging debt types to detect: Requirement & Algorithm

✉ zadiacodabux@ieee.org
@zadiacodabux
https://www.cs.usask.ca/faculty/zadiacodabux/