

RSE PUBLICATION MONITOR

Visualizing how much people write about software

05.03.2024 | JAKOB FRITZ, ROBERT SPECK

HOW MUCH SOFTWARE IS USED IN ACADEMIA

Survey in UK across universities (2014) [1]:

- **92%** use research software
- **56%** develop research software
- For **67%** it would not be practical to conduct their work without software

Survey in US across postdocs (2017) [2]:

- **95%** use research software
- **28%** develop research software
- For **63%** it would not be practical to conduct their work without software

[1]: Hettrick, S. et al. (2014). UK Research Software Survey 2014 [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.14809>

[2]: Udit Nangia, & Daniel S. Katz. (2017). Survey of National Postdoctoral Association - Dataset [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.843607>

WRITING ABOUT SOFTWARE

- Used software (sometimes/often?) is mentioned in text publications
- Citing an accompanying text publication (good) or the software itself (better)
- Idea: Count mentions of software in text publications from researchers
 - ➔ Can be done by scanning the full text for certain keywords
 - ➔ Find matches regardless if software is properly cited or only mentioned

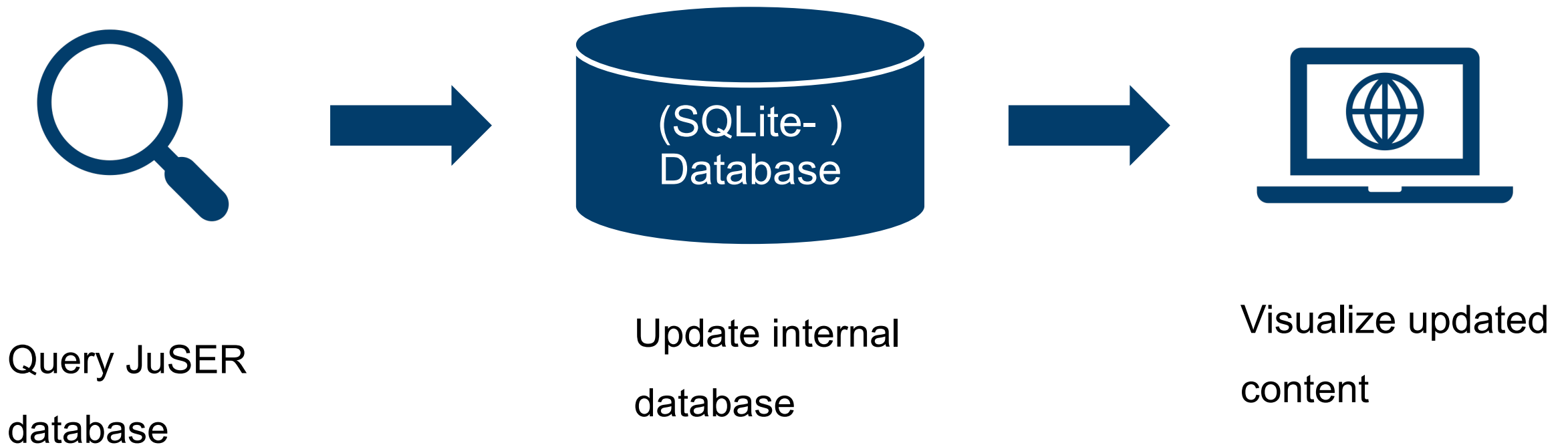
FINDING PUBLICATIONS FROM FZJ

An example from Jülich

- Publication database JuSER provided by ZB (Central Library)
- Own publications from employees need to be entered there
 - ➔ (rather) complete overview of publications from FZJ for the past years
- Used as source of data. Other sources could also be added (not yet done)

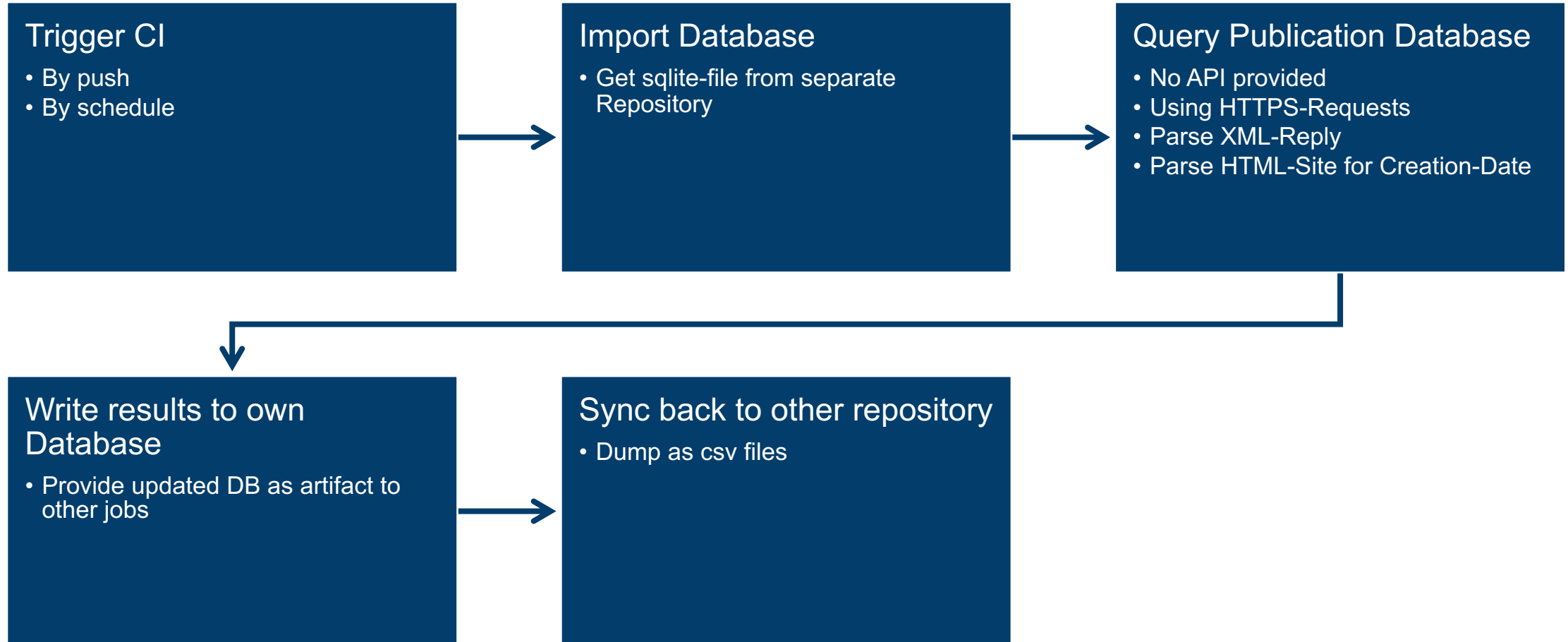
STRUCTURE OF RSE PUBLICATION MONITOR

All in CI; no webserver needed



STRUCTURE OF RSE PUBLICATION MONITOR

Crawler



FINDING PUBLICATIONS FROM FZJ

No full text available, but only search on full text

- Full text search returns result out of context
 - No certainty on whether Keywords are used in assumed way
 - E.g. "Code", "RSE"
 - No phrase search in full text; only words
 - No counting possible
 - Unknown, if a publications only briefly mentions it, or if it is central to the publication
 - Based on the indexed full text (done by JuSER); no direct access to the index
 - No check if software is properly cited
- ➔ Because of all those limitations, it is recommended to use the monitor for trends rather than precise numbers

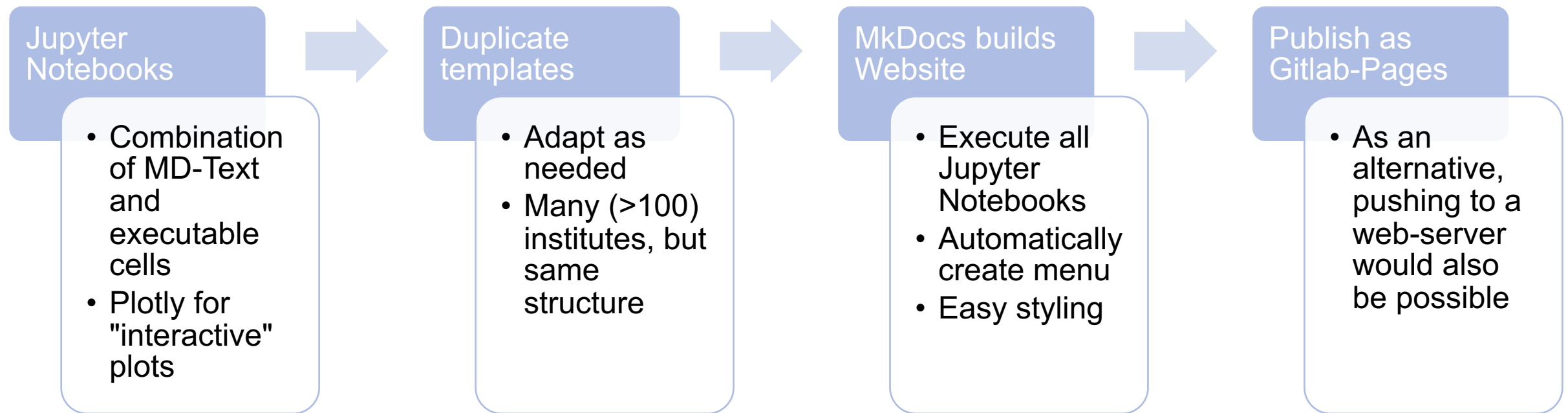
STRUCTURE OF RSE PUBLICATION MONITOR

Database

Institutes	Journals	Keywords	Involved Institutes	Publications
<ul style="list-style-type: none">• Institute-ID• Short Name• Long Name	<ul style="list-style-type: none">• Journal-ID• Journal Name	<ul style="list-style-type: none">• Publication-ID• Software• Code• Github• Gitlab• ...	<ul style="list-style-type: none">• Publication-ID• Institute-ID	<ul style="list-style-type: none">• Publication-ID• Journal-ID• Publication Year• Publication Type• Creation Date• Last modified

STRUCTURE OF RSE PUBLICATION MONITOR

Visualization



THE RSE PUBLICATION MONITOR

Questions

- How many publications are there in JuSER?
- Overall and per year
- How many of those mention any keyword?
- Are there keywords that often appear together?
 - Can we even find a “direction” in this correlation?
- Do some institutes mention certain keywords more often than others?

Following examples for JSC, but also available for other institutes and institute-sections

THE RSE PUBLICATION MONITOR



RSE Publication Monitor

Welcome

ER-C >

IAS >

IBG >

IBI >

IEK >

IKP >

INM >

INW >

JCNS >

PGI >

ZEA >

Overview

Own analysis

Disclaimer

Welcome

On the pages linked here, you can find multiple statistics on publications with RSE topics from the Forschungszentrum Jülich (FZJ)

Total number of publications included in this monitor: 130798

Total number of publications since 2012 that are included in this monitor: 64444

Last update of the RSE-Monitor: 01.03.2024

What is the RSE Publication Monitor about?

The RSE Publication Monitor is an online tool to show how much and in which context people at FZJ write about software. For this, the RSE Publication Monitor queries the [JuSER database](#) to find publications that mention certain keywords related to software. Based on this information, it is now possible to browse through correlations of keywords mentioned in these publications. It is also possible to find out how often different keywords are used in publications from different institutes. This tool can be easily extended to cover other keywords of interest and everyone is encouraged to design their own analysis. The code is publicly available and we are always open for collaborations and further ideas.

Table of contents

What is the RSE Publication Monitor about?

Recommended usage

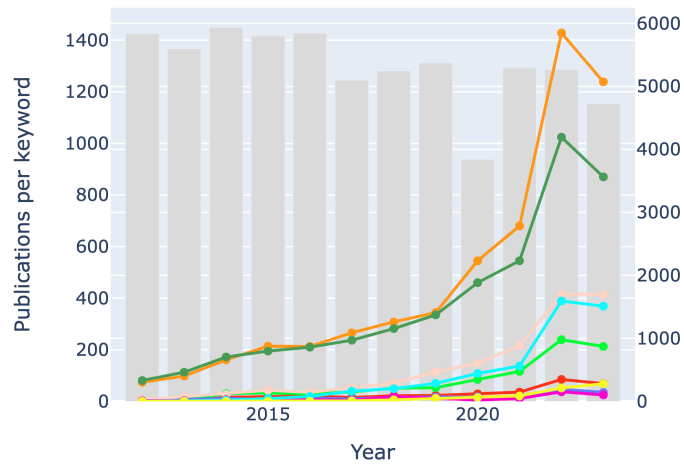
Limitations

Contact

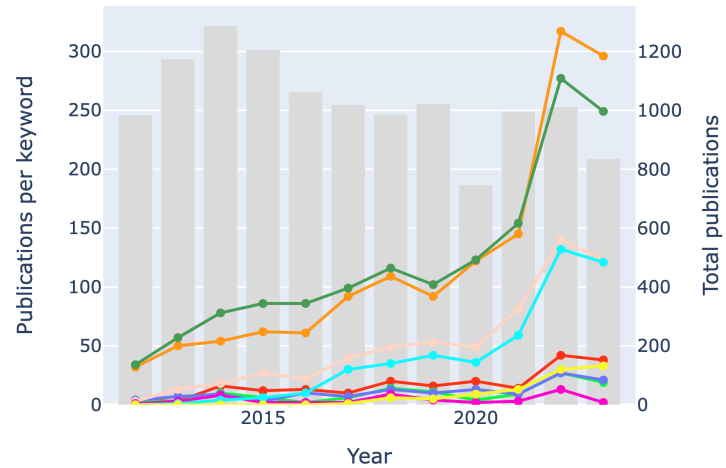
THE RSE PUBLICATION MONITOR

Publications over time

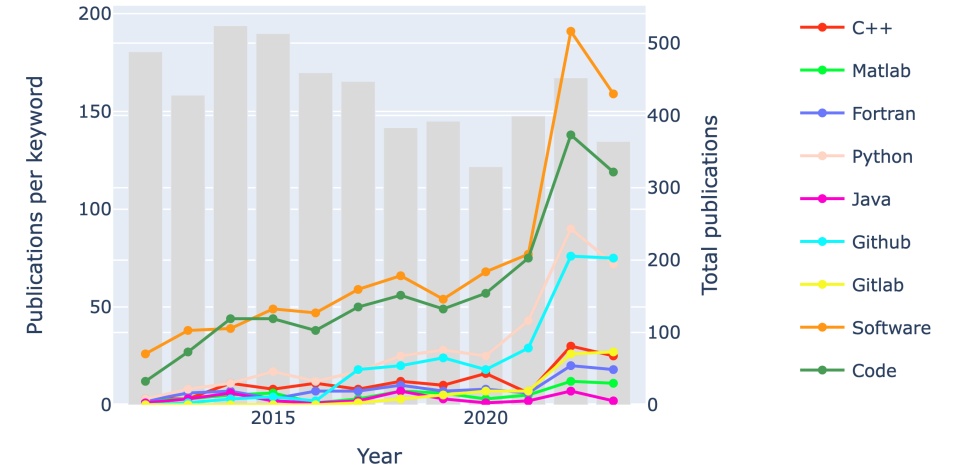
FZJ



IAS

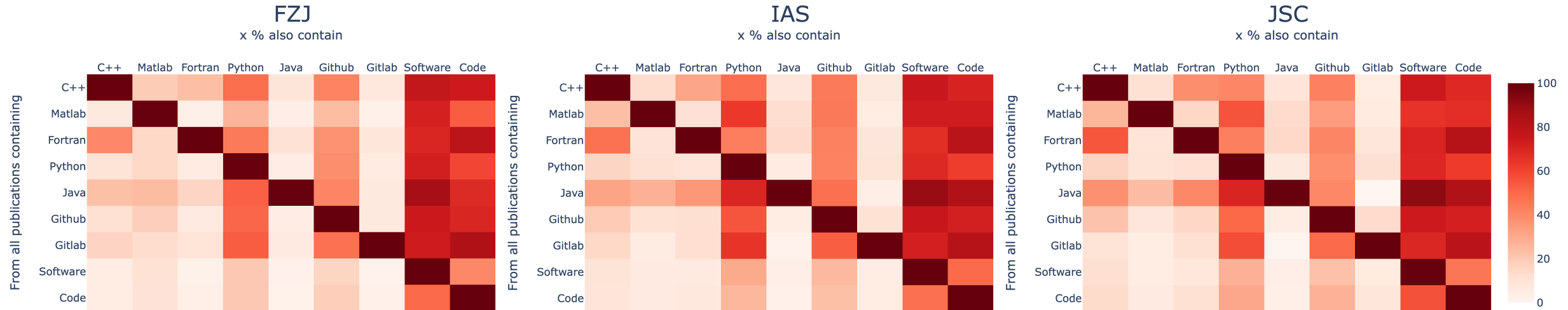


JSC



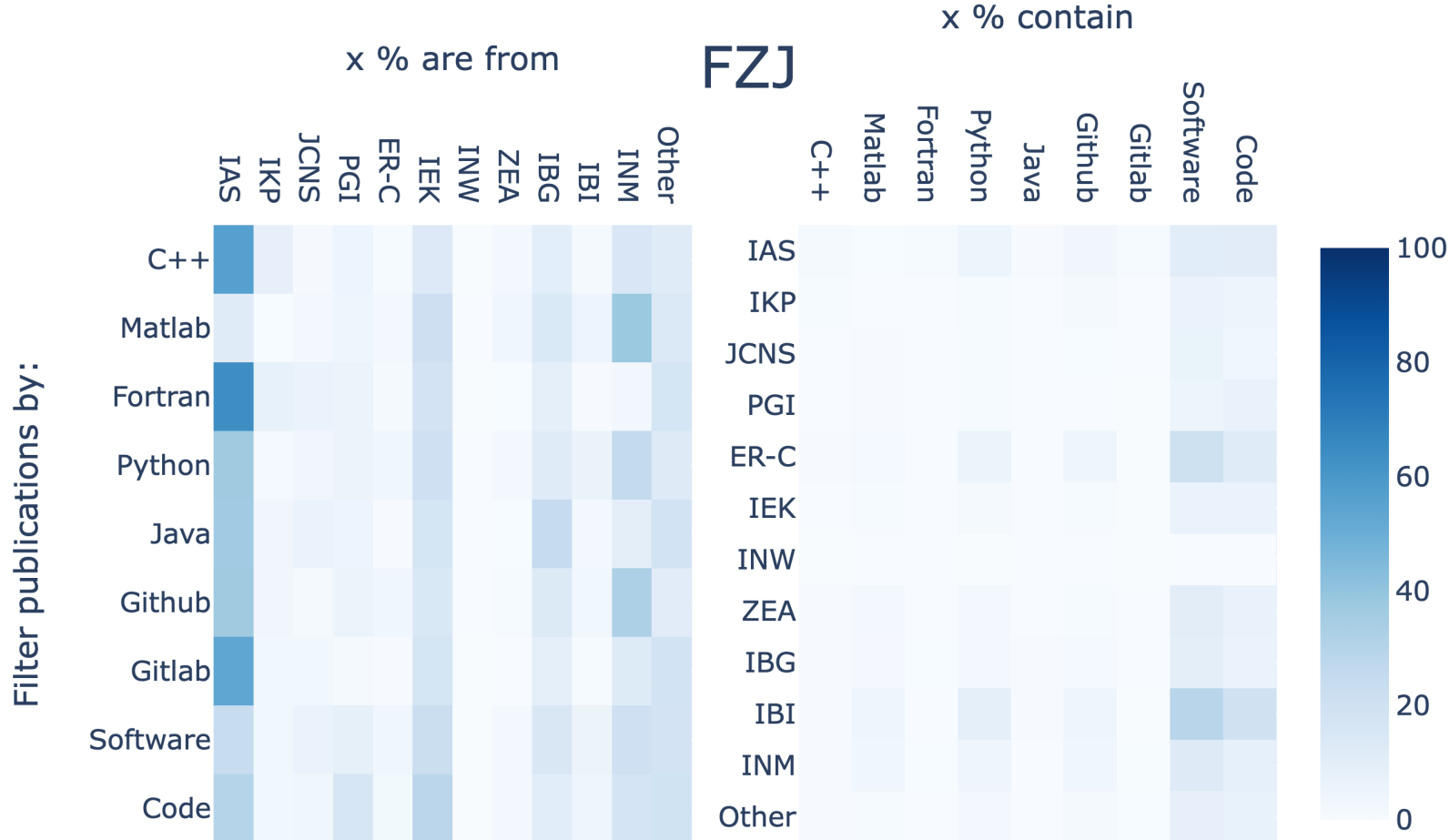
THE RSE PUBLICATION MONITOR

Correlation of keywords



THE RSE PUBLICATION MONITOR

Correlation of institutes and keywords



THE RSE PUBLICATION MONITOR

Current status

- Currently, only statistical examination
- Verification of results (for a few examples)
- No deep look into why trends look the way they look and if/how they are connected
- Data available for download and own analysis

IT IS OPEN SOURCE

Find code at

<https://gitlab.jsc.fz-juelich.de/rg-rse/rse-monitor>

Feel free to fork and adapt.

Feel invited to send Merge-Requests.



ENOUGH OF PRESENTING IT

Explore yourself!

Go to go.fzj.de/rsemonitor



Feel free to send questions & feedback
to me: j.fritz@fz-juelich.de