UNCOVERING SECRETS OF THE MAVEN REPOSITORY: *MAVEN PACKAGING*

Maven, a widely adopted software ecosystem for Java libraries, plays a critical role in the development and deployment of software applications. The research aims to address this knowledge gap by conducting a comprehensive analysis of Maven packaging and informing developers, library maintainers, and more about Maven library practices.

INTRODUCTION

- The understanding of the composition and characteristics of the Maven repository is very limited, leaving users and contributors unaware of the contents they interact with.
- The research will help to make informed decisions for the optimization of library usage and contribute to the advancement of software engineering practices, especially software distribution.

OBJECTIVE

- RQ1 - How common are different packaging types on Maven?
- RQ2 - Which checksums are commonly used for bytecode and how do they change over time?
- RQ3 - Which qualifiers are widely used on Maven?
- RQ4 - What kinds of files are packaged in the libraries' executables?

METHODOLOGY

- Data Selection - Simple Random Sampling - One version per package - 479,915 sample packages
- Core Infrastructure
- RQ1
  - POM File, Maven Index, Maven Repository
- RQ2
  - Maven Repository - Parsing names of artifacts
- RQ3
  - Maven Repository - Parsing names of artifacts
- RQ4
  - Maven Repository - Resolving the primary executable and saving extensions of files

WORKFLOW

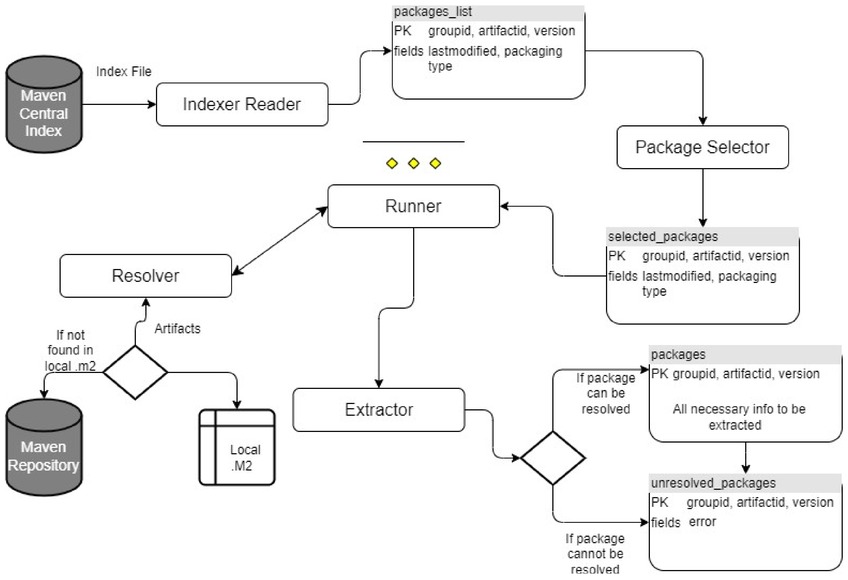


Figure 1: Overview of Core Infrastructure

ANALYSIS

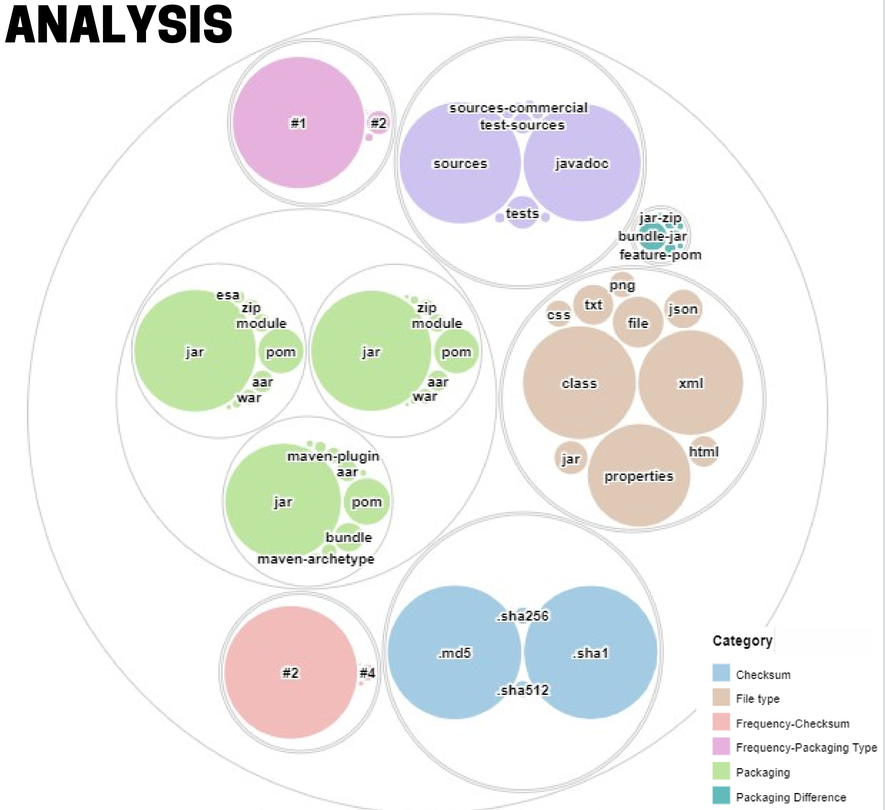


Figure 2: Summary of all the results

AUTHORS

Priyam Rungta
P.Rungta@student.tudelft.nl

AFFILIATIONS

Software Engineering Research Group (SERG) - Delft University of Technology

RESPONSIBLE STAFF

Supervisor - Mehdi Keshani
Responsible Professor - Sebastian Proksch

RESULTS

- 1.4% of the packages were excluded from the analysis due to the unavailability of POM files.
- Inconsistency between sources
  - 9% of packages - POM file and the index file.
  - 4% of packages - Index file and the repository.
  - 12% of packages - POM file and the repository.
- 0.08% packages lacked any checksum algorithm.
- Sources and Javadoc comprise of 90% of total qualifiers present in repository.
- 80% of packages encompassed file types beyond the top 10 categories.

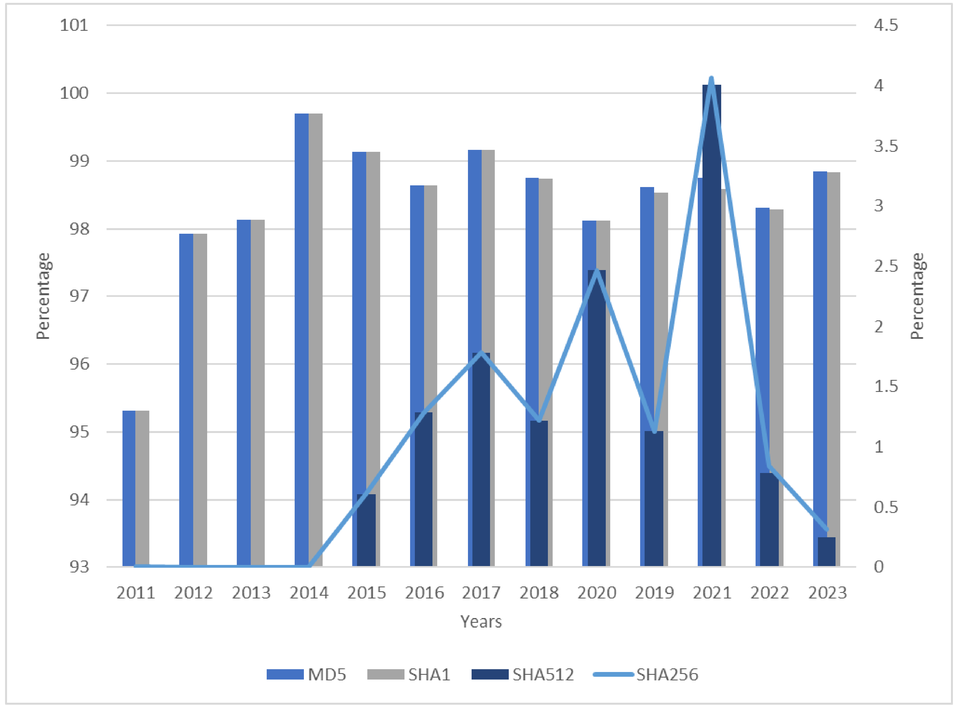
CONCLUSION

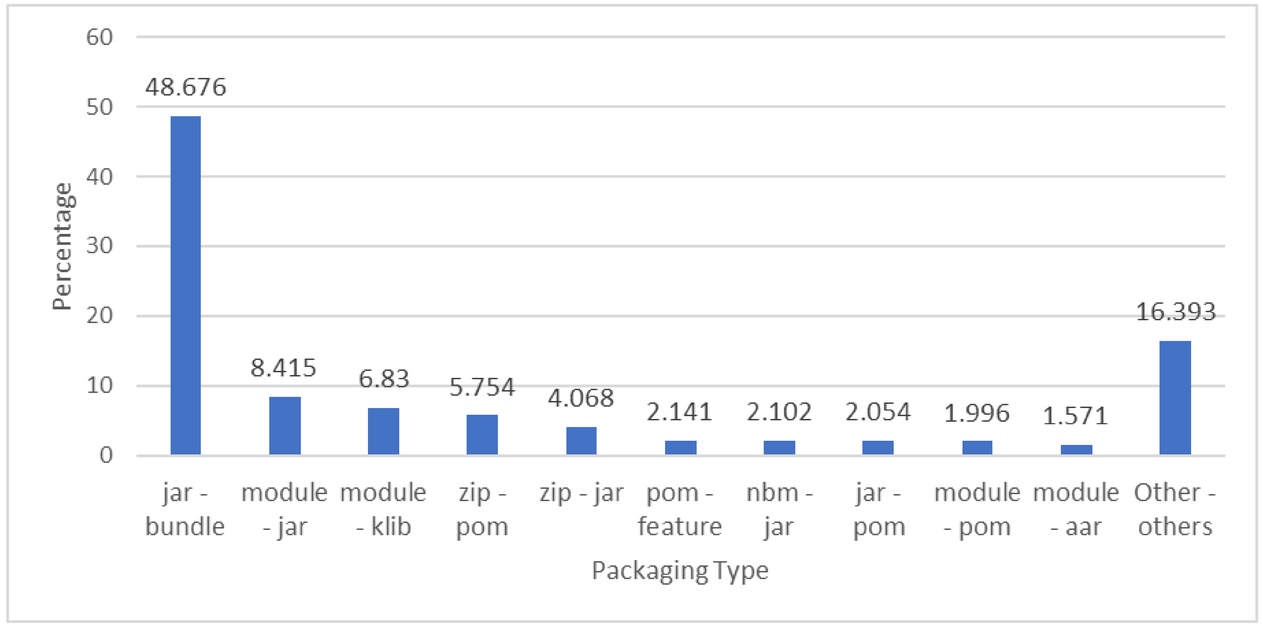
- Inconsistencies identified among different data sources, emphasize the need for improved data consistency and reliability within the Maven ecosystem.
- Encouraging wider adoption of secure hash functions, as only 1.4% packages utilize them.
- Develop adaptable approaches to optimize Maven library utilization and interplay between characteristics.

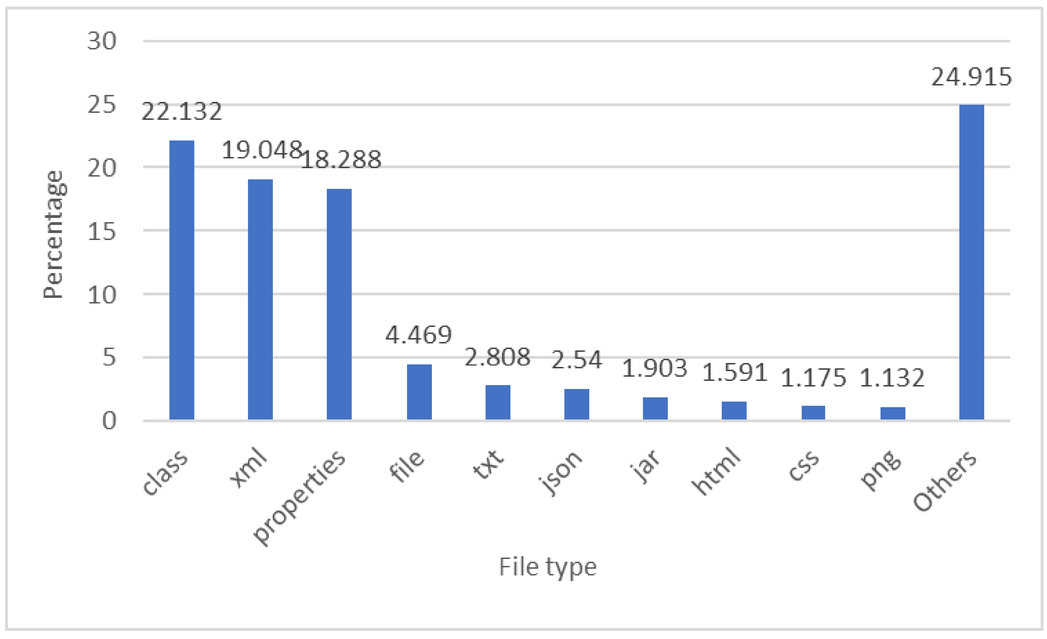
RELATED LITERATURE

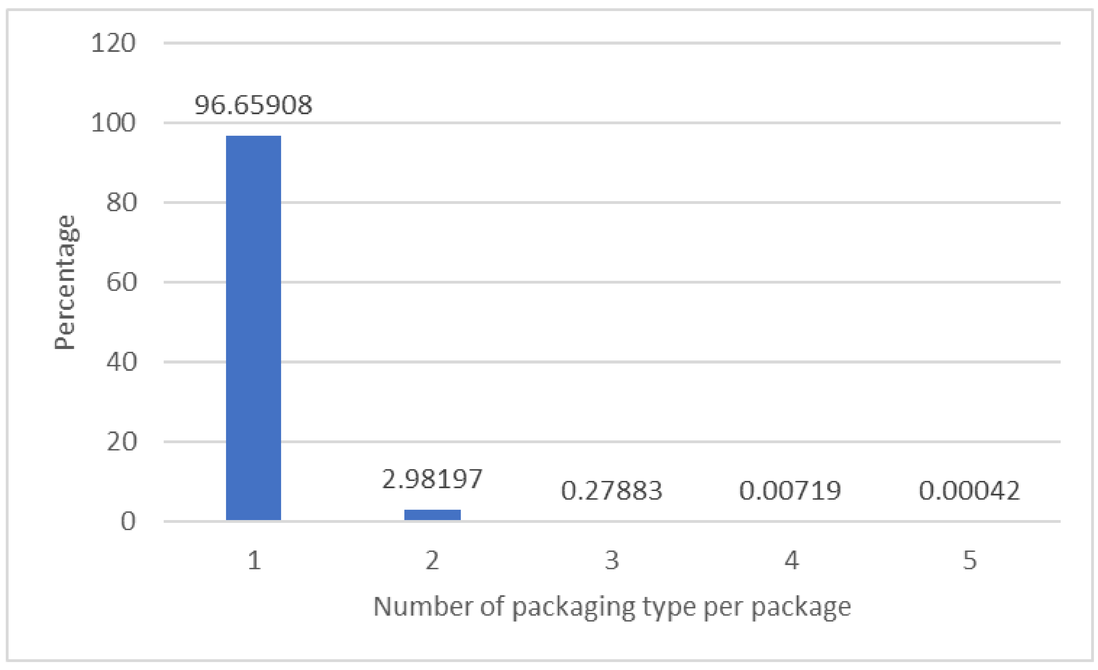
- S. Raemaekers, A. Van Deursen, and J. Visser, "The maven repository dataset of metrics, changes, and dependencies," in 2013 10th Working Conference on Mining Software Repositories (MSR), pp. 221–224, IEEE,2013
- T. Kanda, D. M. German, T. Ishio, and K. Inoue, "Measuring copying of Java archives," Electronic Communications of the EASST, vol. 63, 2014
- A. Benelallam, N. Harrand, C. Soto-Valero, B. Baudry, and O. Barais, "The maven dependency graph: a temporal graph-based representation of maven central," in 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR), pp. 344–348, IEEE, 2019

CHARTS


101
100
99
98
97
96
95
94
93
Percentage
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Percentage
2011
2012
2013
2014
2015
2016
2017
2018
2020
2019
2021
2022
2023
Years
MD5
SHA1
SHA512
SHA256


60
50
40
30
20
10
0
Percentage
48.676
8.415
6.83
5.754
4.068
2.141
2.102
2.054
1.996
1.571
16.393
jar - bundle
module - jar
module - klib
zip - pom
zip - jar
pom - feature
nbm - jar
jar - pom
module - pom
module - aar
Other - others
Packaging Type


30
25
20
15
10
5
0
Percentage
22.132
19.048
18.288
4.469
2.808
2.54
1.903
1.591
1.175
1.132
24.915
class
xml
properties
file
txt
json
jar
html
css
png
Others
File type


120
100
80
60
40
20
0
Percentage
96.65908
2.98197
0.27883
0.00719
0.00042
1
2
3
4
5
Number of packaging type per package