Generating the Blueprints of the Java Ecosystem

Vassilios Karakoidas, Dimtris Mitropoulos, ***Panos Louridas****, Georgios Gousios, Diomidis Spinellis

Athens University of Economics and Business

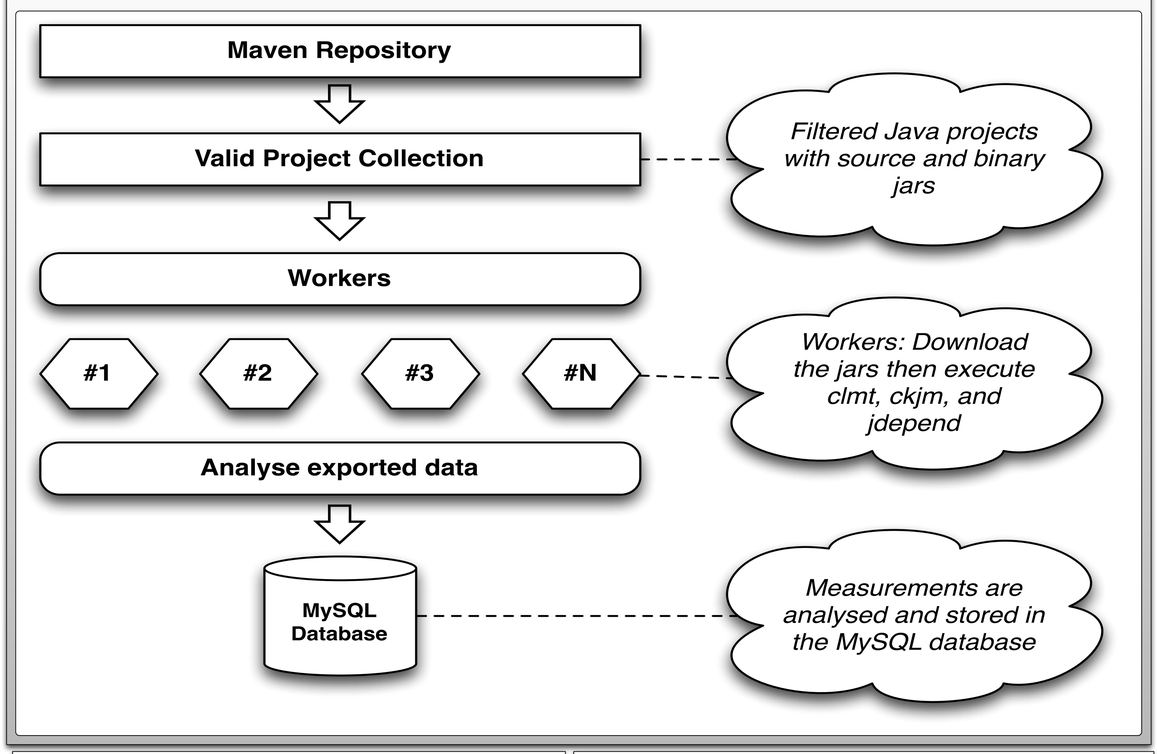
Department of Management Science and Technology

*louridas@aueb.gr

This work presents the dataset obtained by statically analysing a set of projects (*11,365 projects*) of the Maven Central Repository by three static analysis tools; Cross-Lanugage Metric Tool (CLMT), Chidamber and Kemerrer Java Metrics Tool (CKJM), and JDepend. These tools cover four aspects of a software project; class design, method design, package design and program size.

11,365 projects	22,730 Jars	74,565,772 LoC	446,749 Artifacts	43 Metrics	32,844,836 Measurements

Dataset Construction Process



Detecting Domain-specific Language Usage in Open Source Projects

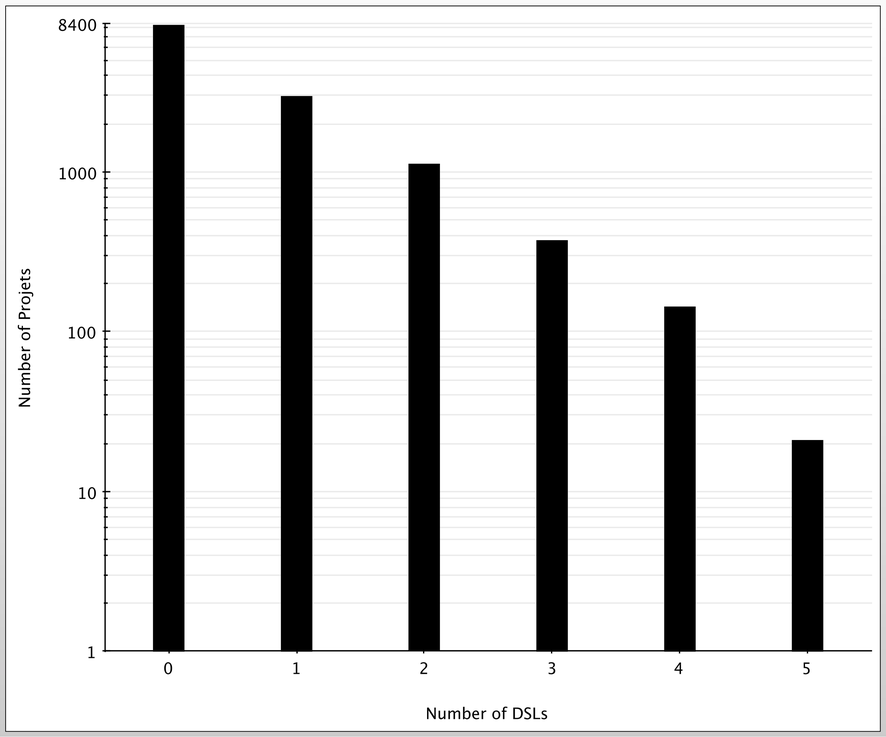
XML | SQL | HTML | RTF

Regular Expressions | XSLT | XPATH

DSLs

The detection process was easy, the source code was statically analysed and the usage of specific packages were detected e.g. java.util.regex (regular expressions), java.sql and javax.sql for SQL.

How many DSLs are used per project?



#1 XML with 3094 uses

Regex, 1751

SQL, 1035

XSLT, 888

XPath, 190

HTML, 68

RTF, 7

Facts

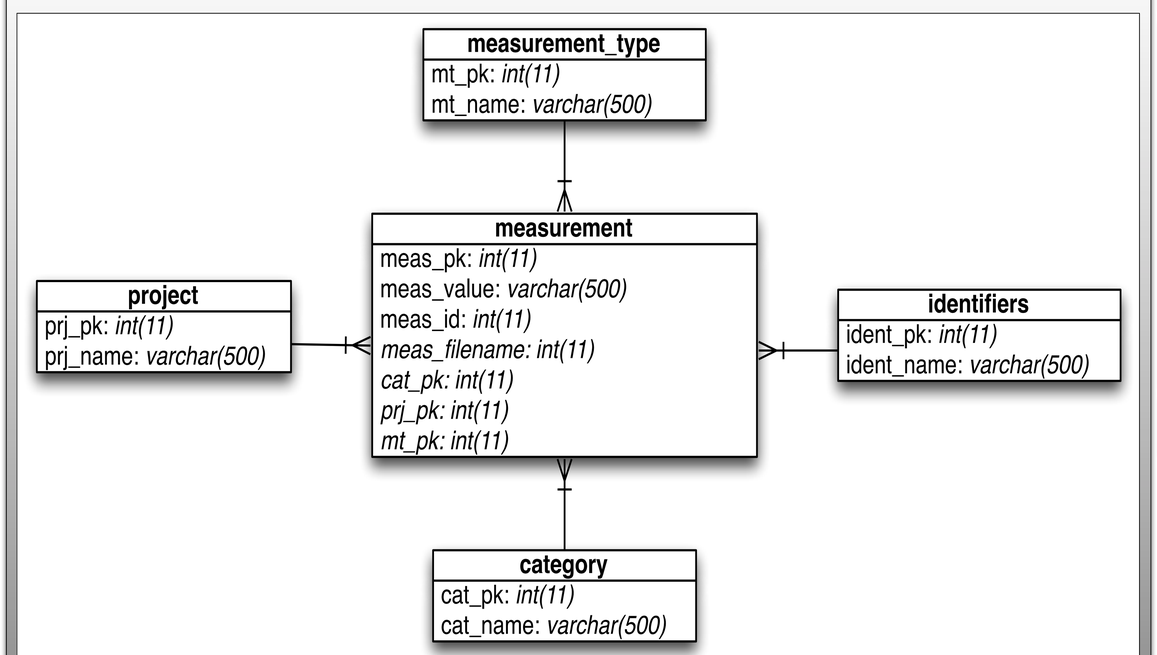
- *~35% of the projects are using at least one DSL*
- *547 projects are using four DSLs*
- *8 projects are using 7 DSLs!*

Metric Categories

- **Class** *17 metrics*
- **Method** *3 Metrics*
- **Program Size** *17 metrics*
- **Package** *6 metrics*

Github Dataset Repository

Database Schema



Popular DSL Combinations

XML, XSLT (475)	Regex, SQL, XML, XSLT (80)
Regex, XML (303)	Regex, SQL, XML (71)
Regex, XML, XSLT (158)	SQL, XML, XSLT (54)
SQL, XML (162)	XML, XPath (50)
Regex, SQL (116)	...

Research Opportunities

The dataset can be used by researchers to test their models and theories against a large set of emprical data e.g. fine tune software quality models that are based on metrics.

Practisioners can test their tools and validate their calculations against CKJM and JDepend.

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (nsrf) - Research Funding Program: Thalis - Athens University of Economics and Business - Software Engineering Research Platform.

European Union, European Social Fund. OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING. MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS MANAGING AUTHORITY. NSRF 2007-2013. Co-financed by Greece and the European Union