A Time Series-Based Dataset of Open-Source Software Evolution

Bruno L. Sousa (DCC - UFMG)

Mariza A. S. Bigonha (DCC - UFMG)

Kecia A. M. Ferreira (DECOM - CEFET-MG)

Glaura C. Franco (DEST - UFMG)

Introduction

❏ Software evolution has been extensively studied in the last years

❏ Need for software evolution data

❏ **Problem:** lack of comprehensive, public and updated software evolution datasets

❏ There are some public software evolution datasets, but they are not comprehensive and updated

Goal

Build a comprehensive software evolution dataset with temporal data of open-source Java systems

Process of Building the Dataset

| Selecting the systems | Extracting the systems' releases | Collecting metrics | Generating time series |

❏ Nine systems from another dataset were included

❏ Criteria

1. Programming language
2. Popularity
3. Activity
4. Lifetime
5. Not be deprecated

Process of Building the Dataset

| Selecting the systems | Extracting the systems' releases | Collecting metrics | Generating time series |

❏ Interval of 14 days between the releases

❏ Automatic approach
  - ❏ Python
  - ❏ REST API from GitHub

Process of Building the Dataset

| Selecting the systems | Extracting the systems' releases | Collecting metrics | Generating time series |

❏Total of 46 software metrics computed

❏Used tool
- ❏ CK Tool
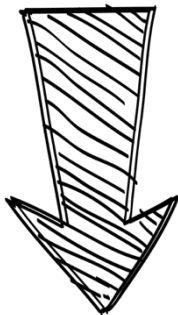
Process of Building the Dataset

Selecting the systems	Extracting the systems' releases	Collecting metrics	Generating time series

Time series file format

CSV

Conclusion

❏ We proposed and built a software evolution dataset

❏ It contains 46 systems with data of 46 software metrics

❏ Dataset website

  ❏ https://brunolsousa.github.io/software-evolution-dataset/index.html

THANK YOU!

CONTACT

❏ **Email:**
  ❏ bruno.luan.sousa@dcc.ufmg.br

❏ **LinkedIn:**
  ❏ https://www.linkedin.com/in/bruno-sousa-65a410aa/

A Time Series-Based Dataset of Open-Source Software Evolution

Bruno L. Sousa (DCC - UFMG)

Mariza A. S. Bigonha (DCC - UFMG)

Kecia A. M. Ferreira (DECOM - CEFET-MG)

Glaura C. Franco (DEST - UFMG)