Análise da Contribuição de Código entre Repositórios do GitHub

Laís M. A. Rocha, Thiago H. P. Silva, Mirella M. Moro
{laismota,thps,mirella}@dcc.ufmg.br

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Motivation and Goals

Is there **external contribution** of users on repositories of different Programming Languages?

*A developer who has greater ability to Java contributes with repositories of other programming languages besides Java?*
Motivation and Goals

Is there correlation among Programming Languages?

Is Java more related to Python than Assembly?
Related Work

● In the context of academic social networks
  ○ 3c-index
  ○ To measure the influence of researchers according to their specialities to other contexts.

● GitHub - Collaborative Social Network
  ○ Network structure, formation of the participants and relationships
  ○ Study of the types and uses of programming languages
  ○ How they relate to software quality
Methodology

1st step of indicator (3c-index)

Base Community of the contributor:

Volume → Score

Base Community = Contributor’s Largest Rank Position
**Stars**: Starring a repository allows you to keep track of projects that you find interesting, even if you aren't associated with the project.

**Forks**: is a copy of a repository that you manage. Forks let you make changes to a project without affecting the original repository.
Methodology

- Skill languages: JavaScript, Java, Python
- Base community: Java

GitHub repositories

- JavaScript: 207, 704
- Java: 269, 563
- Python: 263, 151

GitHub repositories

- Ruby: 262, 516
- Java: 200, 83
- Python: 446, 684

Contributors

- Java: 200, 83
- Python: 391, 127
Methodology

Tiobe Index: Top 12 PLs: *(March 2016)*

- Assembly
- C
- C++
- C#
- Java
- JavaScript
- Pascal
- Perl
- PHP
- Python
- Ruby
- Visual Basic
Most of the contributors have from 1 to 5 skill programming languages.

Less than 100 contributors have more than 20 skill languages.
External contribution per PL

![Bar chart showing the proportion of contributors for different programming languages.](chart.png)
External contribution per PL
Correlation among PLs
Correlation among PLs

Visual Basic  
Ruby  
Python  
PHP  
Perl  
Pascal  
JavaScript  
Java  
C++  
C#  
C  
Assembly

<table>
<thead>
<tr>
<th>Assembly</th>
<th>C</th>
<th>C#</th>
<th>C++</th>
<th>Java</th>
<th>JavaScript</th>
<th>Pascal</th>
<th>Perl</th>
<th>PHP</th>
<th>Python</th>
<th>Ruby</th>
<th>Visual Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Java</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JavaScript</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pascal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Python</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruby</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Basic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Correlation among PLs
Correlation among PLs
Correlation among PLs
Classification of PLs types

- **More Collaborative languages**
  - Ruby
  - Python
  - JavaScript

- **Very collaborative languages**
  - Java
  - PHP
  - C

- **Medium collaborative languages**
  - Assembly
  - Perl
  - VisualBasic
  - Pascal

- **Less collaborative languages**
  - C++
  - C#
Characterization of network contributors

VisualBasic
Representing less collaborative Programming Languages

JavaScript
Representing more collaborative Programming Languages
Characterization of network contributors

- Many components with few connections and few components with many
- Assembly and Python have more components than others
Characterization of network contributors

Graph ranking

- Even with the variation around the values, the overall performance is power law
- Many repositories with few contributors and few repositories with many
Conclusions

- Significant code contribution among programming languages

- There is correlation among contributors of repositories, especially among more collaborative Programming Languages

- Programming Languages with different profiles, some PLs with greater similarity than others and some with more collaborations than others
Future Work

- Application of other metrics of social networks
- To analyze the contributions when users do not give fork on repositories and also when despite giving fork in a repository they do not contribute to the project
Análise da Contribuição de Código entre Repositórios do GitHub

Laís M. A. Rocha, Thiago H. P. Silva, Mirella M. Moro
{laismota,thps,mirella}@dcc.ufmg.br

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais