Matching GitHub developer profiles to job advertisements

Claudia Hauff
Delft University of Technology
Delft, the Netherlands
c.hauff@tudelft.nl

Georgios Gousios
Radboud University Nijmegen
Nijmegen, the Netherlands
georgios@cs.ru.nl

ABSTRACT

GitHub is a social coding platform that enables developers to efficiently work on projects, connect with other developers, collaborate and generally "be seen" by the community. This visibility also extends to prospective employers and HR personnel who may use GitHub to learn more about a developer's skills and interests. We propose a pipeline that automatizes this process and automatically suggests matching job advertisements to developers, based on signals extracting from their activities on GitHub.

Categories and Subject Descriptors

D.2.7 [**Software Engineering**]: Distribution, Maintenance, and Enhancement—*Version control*; D.2.9 [**Software Engineering**]: Management—*Programming teams*

General Terms

Management

Keywords

github, jobs

1. INTRODUCTION

Today, social coding platforms such have become an important tool for developers to showcase their work and become visible in the developer community. GitHub[1] in particular has become an established way for developers to create a portfolio of their work to be considered during the hiring process by potential employers [**?**]. In order to find potential employers, developers search for job openings in various online job portals and compare their experiences and activities with the described position. This is a cumbersome process as many job advertisements are lengthy, mentioning a plethora of programming languages, libraries and techniques that the perfect candidate should be familiar with. The Moreover, each

---

[1]https://github.com/

of these items is usually conditioned on the number of years of experience or the level of expertise and may fall into the "required" or "preferred" category. Over the years, job advertisements have required more and more skills from prospective employees. This has led to a situation where a developer matching 50% to 60% of the described requirements may actually be a very well qualified candidate for the advertised position. Another complicating factor is the fact that job advertisements' writing style may be influenced by the numerous people involved in the creation of a job profile (managers, developers, HR personnel, etc.). Here a "semantic gap" may exist between search terms a developer is using to find suitable advertisements in job portals and the terms that actually appear in an advertisement.

Additionally, is usually either required or preferred, asking for n years of experience GitHub provides several user-based summary statistics such as *Contributions in the last year*, *Number of forked projects*, *Number of followers* that provide basic insights into developer activities. The usefulness of this information is limited, as it does not offer immediate insights into the developer's programming abilities, the particular languages the developer is regularly using or the type of development toolchain the developer is using.

We conclude that considering the vast amounts of job advertisements in the IT sector, finding a job advertisement that is a good match with one's own abilities and desires is currently an inefficient and complex process.

Business-oriented social networks such as LinkedIn[2] are using recommender algorithms to *push* job advertisements to its users (in addition to the traditional *pull*-based model where users are actively searching among the available advertisements). Recommender algorithms determine the *similarity* score $s(U, A)$ between pairs of user profiles U and advertisement profiles A; if s exceeds a pre-set threshold the advertisement is shown to the user. While this process moves the burden of determining the degree of matching away from the user, it is limited in its abilities due to the lack of detailed user profile data as statements such as "Experienced Java developer" or "Embedded Software Engineer" contain relatively little information.

Portal such as GitHub on the other hand allow us to extract very detailed information about a developer's profile. This information is not just limited to the developer's main programming languages but also includes information about the use of particular software libraries, development tool-chains, the expertise and proficiency in algorithms, data structures and design patterns, etc.

Based on this insight, in this paper we propose a proof-of-concept to (i) make use of GitHub to *automatically* mine highly-detailed

---

[2]https://www.linkedin.com/

developer profiles, (ii) *automatically* extract relevant information from natural language job advertisements and (iii) derive a similarity function based on features discussed in the literature.

2. DATA

3. PRELIMINARY WORK

4. RELATED WORK

5. CONCLUSIONS