

What Skills do IT Companies look for in New Developers?

A Study with Stack Overflow Jobs

João Eduardo Montandon^{a,*}, Cristiano Politowski^b, Luciana Lourdes Silva^c, Marco Tulio Valente^a, Fabio Petrillo^d, Yann-Gaël Guéhéneuc^b

^a*Federal University of Minas Gerais*

^b*Concordia University*

^c*Federal Institute of Minas Gerais*

^d*Université du Québec à Chicoutimi*

Abstract

Context: There is a growing demand for information on how IT companies look for candidates to their open positions. **Objective:** This paper investigates which hard and soft skills are more required in IT companies by analyzing the description of 20,000 job opportunities. **Method:** We applied open card sorting to perform a high-level analysis on which types of hard skills are more requested. Further, we manually analyzed the most mentioned soft skills. **Results:** Programming languages are the most demanded hard skills. Communication, collaboration, and problem-solving are the most demanded soft skills. **Conclusion:** We recommend developers to organize their resumé according to the positions they are applying. We also highlight the importance of soft skills, as they appear in many job opportunities.

Keywords: Hiring, Soft Skills, Hard Skills, Stack Overflow

Software is “eating the world” and we everyday observe the rise of companies centered on software. At the same time, software is becoming more complex, with its development requiring highly qualified developers [1]. As a result, modern IT companies allocate developers to work in specific areas, such as databases, security, user interface (front-end design), core features (back-end design), mobile development, etc.

Therefore, there is a growing demand for information on how modern companies deal with this context when looking for developers to their open positions. To find out what are the skills required by companies when selecting new employees, we report in this article the first results of an analysis of more than 20,000 job opportunities available in Stack Overflow Jobs portal, a platform that allows companies to publish new opportunities for IT professionals. In particular, we leveraged the main characteristics of 14 IT roles, such as Backend, Frontend, and Mobile developers. We also analyze which soft skills are mostly requested by IT companies when selecting new candidates. Finally, we discuss the implications of this analysis for both recruiters and developers.

1. The Anatomy of a Job Opportunity

A job opportunity is a declaration of expectations. It describes what a company expects from its candidates, as well as what the candidates should expect from the company. For this, a job opportunity includes important company details, such as mission, culture, benefits, etc. At the same time, the company should provide enough details so candidates can decide whether they are qualified or not for the position. Lastly, an opportunity should list the job’s responsibilities, so that candidates can be aware of their duties.

Figure 1 shows an example of a job opportunity posted on Stack Overflow Jobs portal. As we can observe, some skills are technically-driven, e.g., *Python*, *REST & API*, *JavaScript*, etc. Skills in this group

*Corresponding author

URL: joao.montandon@dcc.ufmg.br (João Eduardo Montandon)

Senior Python Software Engineer

████████████████████ | No office location

██████████ Remote

Technologies

python javascript odoo

Job description

About the role:

Odoo is the ERP used by ██████████ and is written in Python. As the roll-out of Odoo through-out the organisation is nearing completion, it is now necessary to build out our own in-house Python team, so we are not reliant on Odoo consultants for making changes in our Odoo setup. The Python Developer will be **incredibly organized, enthusiastic, confident,** and a **great communicator.**

Your responsibilities include:

- Implement the required modules to operate Odoo for ██████████
- Support the integration according to stakeholders' requirements.
- Assist and train other stakeholders for their usage of Odoo.
- Mentor other Python developers within the team.

About you:

You are:

- **Initiative-taking**, you are self-motivated, a doer, and can drive projects from start to finish
- **A leader**, you have a proven ability to inspire, energize and mobilize a team towards a goal
- **A team-player**; you are comfortable working with different styles and believe (like us) that together we achieve much more than alone
- **Driven**; you are used to working hard to achieve a goal you care about and running several projects in parallel
- **A great communicator**; you are comfortable in communicating in English both written and oral, including leading meetings, selling your ideas and storytelling

You have:

- Minimum of 3 years as a software engineer
- 2+ years **Python** experience
- Knowledge of best engineering practices in agile software development: architectural paradigms, code reviews, branching, task management, documentation, testing
- Thorough experience of **REST & API** practices
- In-depth knowledge of database design and optimization
- **Excellent written and verbal communication skills in English**
- **Ability to work well with teams effectively without supervision**

We'd be extra impressed if you also have:

- **Javascript** and **Nodejs** knowledge
- ERP knowledge or willing to learn the **ERP Odoo**

We offer:

- Fully remote and highly talented distributed team
- Working on great tech stack with cutting edge technologies
- Product company with a long-term vision
- Competitive salary depending on the relative work experience.
- Project exposure and ownership that impacts our users, product, and business
- Challenging technical tasks, fast learning cycles, and meaningful feedback.

Location: remote. Flexible within timezone CET +/-4

We believe that the more inclusive we are, the better products we build and the better we are able to serve our customers. Women and other minorities under-represented in tech are strongly encouraged to apply.

Figure 1: An example of a post in Stack Overflow Jobs portal. Hard skills are annotated in blue, while soft skills are in yellow.

are known as hard skills [2]. By contrast, other skills denote behavioral characteristics of the candidates, such as *verbal communication*, *team player*, *leadership*, etc. They are known as soft skills [3]. In this article, we analyze both hard and soft skills required by IT job opportunities.

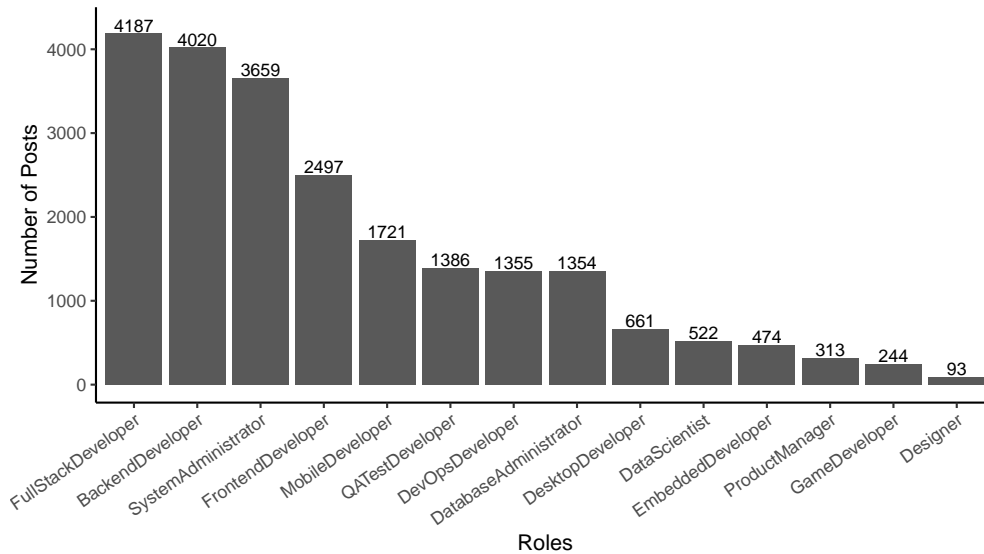


Figure 2: Distribution of job posts collected in this study.

2. Data Collection

We investigated the jobs posts available on Stack Overflow Jobs portal.¹ We collected the posts visible in the platform for three months, from March 25th to June 28th, 2019 by downloading the posts available in every weekday during this period. As a result, we retrieved a total of 20,968 job posts, including their titles, description, and tags. Furthermore, each post in Stack Overflow is automatically associated with at least one out of 14 predefined roles provided by the platform, such as `BACKENDDEVELOPER`, `FRONTENDDEVELOPER`, `MOBILEDEVELOPER`, etc. We also collected this information to perform a fine-grained analysis of each post.

Figure 2 shows the distribution of opportunities for each role. `FULLSTACKDEVELOPER`, `BACKENDDEVELOPER`, and `SYSTEMADMINISTRATOR` are the most demanded positions, with at least 3,600 posts each, representing 52.8% of all job posts. By contrast, six roles have less than 1,000 posts: `DESKTOPDEVELOPER`, `DATASCIENTIST`, `EMBEDDEDDEVELOPER`, `PRODUCTMANAGER`, `GAMEDEVELOPER`, and `DESIGNER`. Together they represent 10.2% of the analyzed posts.

3. Analyzing Hard Skills

On Stack Overflow Jobs, companies describe the hard skills required for a given position using the same tagging mechanism provided by Stack Overflow’s main platform. Tags play a central role in this ecosystem, as they are used to identify Q&A topics [4]. Therefore, we use the tags added to each post as an indication of the hard skills demanded by each one. For instance, the tags associated to the job post in Figure 1 (top) indicate that candidates should master *python*, *javascript*, *NodeJS*, and *odoo*.

In total, we identified 1,916 hard skills (i.e., tags) mentioned 70,680 times in the collected posts. The frequency of these skills follows a long tail shape, where few elements are generally responsible for almost all occurrences. For this reason, we removed the hard skills with less than 10 occurrences. We ended up with 282 hard skills, representing 67,062 occurrences (95%). Next, two authors used open card sorting [5] to group tags into abstract categories. After analyzing together a group of 50 tags, they leveraged six abstract hard skills: *Languages*, *Libs & Frameworks*, *OS & Infrastructure*, *Process & Methods*, *Data Systems*, and *Development Tools*. Then, they independently annotated the remaining 232 tags into these categories (i.e.,

¹<https://stackoverflow.com/jobs>

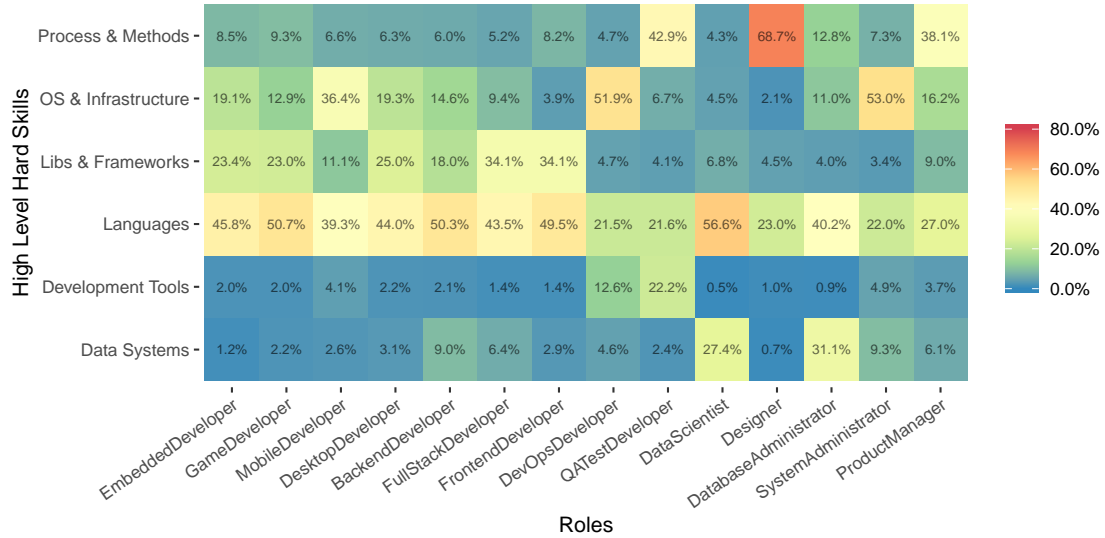


Figure 3: Heatmap indicating how hard skills are distributed among developers roles.

each tag was evaluated by two authors). The inter-rater agreement, calculated using Cohen’s Kappa [6], was 0.82. Lastly, both authors discussed the conflicting cases to reach a common ground.

Figure 3 shows a heatmap with the distribution of high-level hard skills (rows) per developer roles (columns). With this heatmap, we intend to provide a technology-agnostic analysis, i.e., one that focuses on high-level hard skills (e.g., *Languages*) instead of current technologies (e.g., *Java*). In this way, we also intend to increase the validity of our results against technological changes.

First, *Languages*-based skills are relevant for all roles, ranging from 21.5% (DEVOPSDVELOPER) to 56.6% (DATASCIENTIST). In fact, *Languages* is the most required hard skill for 9 out of 14 technical roles analyzed. Even designers must master some sort of languages to better provide their prototypes, such as *css*, *html*, and *javascript*. However, the concentration of *Languages* is higher for development-based roles, e.g., MOBILEDEVELOPER (39.3%), GAMEDEVELOPER (50.7%), FULLSTACKDEVELOPER (43.5%), and FRONTENDDEVELOPER (49.5%). These characteristics make *Languages* the only skills that are significantly mentioned in all roles.

Likewise, development-based roles also demand skills on *Libs & Frameworks*, which generally have more than 20% of participation in this group. They are specially required for FULLSTACKDEVELOPER and FRONTENDDEVELOPER (34.1%, both).

Regarding *OS & Infrastructure*, two roles stand out with more than 50%: SYSTEMADMINISTRATOR (53.0%) and DEVOPSDVELOPER (51.9%). Interestingly, MOBILEDEVELOPER appears next with 36.4%. This result is explained by the fact that such developers must master mobile operating systems, such as *ios* and *android*.

For *Process & Methods*, DESIGNER stands out with 68.7% of the skills coming from this group. Indeed, the most mentioned DESIGNER skills are related to user interfaces, e.g., *user-experience*, *user-interface*, and *design*. Next, QATESTDEVELOPER and PRODUCTMANAGER follow up with 42.9% and 38.1%, respectively. Differently, these roles require hard skills associated with software quality assurance. For instance, most QATESTDEVELOPER opportunities require knowledge on automated testing methods. PRODUCTMANAGER’s candidates should have experience in agile development processes such as *scrum*. By contrast, *Process & Methods*-based skills are not highly demanded for development-based jobs.

Skills on *Data Systems* are more required for DATABASEADMINISTRATOR and DATASCIENTIST, with 31.1% and 27.4%, respectively. Surprisingly, *Development Tools*-based skills are considerably mentioned only in QATESTDEVELOPER (22.2%) and DEVOPSDVELOPER (12.6%). In fact, tools like *selenium* and *jenkins* are in the top-5 most mentioned ones for this category.



Figure 4: Soft skills word cloud.

4. Analyzing Soft Skills

Unlike hard skills, soft skills are only available on jobs’ textual descriptions, i.e., Stack Overflow does not include tags for such skills. This limitation makes the act of collecting soft skills more complex, as we have to extract them from unstructured text. We handled this issue by first randomly selecting a sample of 376 opportunities from a subset of 17,756 posts.² We discarded the remaining 3,212 posts as they were not written in English and, therefore, do not fit into our analysis. Then, three authors manually annotated sentences that refer to soft skills, such as “good communication skills”, “ability to work independently”, etc. In total, we extracted 1,530 sentences from 315 opportunities. Next, we generated a word cloud to identify the most common terms among these sentences. As some soft skills can be described using multiple words (e.g., “verbal communication”), we adapted the word cloud to consider bigrams as well.

Figure 4 depicts the word cloud for the top-100 most frequent terms. We can observe that *communication* plays a central role among the most required soft skills. Our sample allows us to conclude that $32\% \pm 5\%$ of the posts in our population (i.e., about one in three posts) mention this skill. *Collaboration*-based skills also feature a special position in this ranking. For instance, the word “team” appears in at least 22% of the jobs posts (i.e., $27\% \pm 5\%$). Lastly, some companies also require experience in *problem-solving* skills, as the words “analytical”, “problem solving”, or “deliver” are present in at least one out of ten posts (i.e., $15\% \pm 5\%$).

5. How do Skills Change Over Time?

Litecky et. al. [7] leveraged the most demanded hard and soft skills in the IT industry between July 2007 and April 2008 through an analysis of more than 200,000 online job opportunities. When we compare their results with ours, we observe some changes regarding hard skills. For instance, languages such as *C/C++*, *SQL*, and *Java* were frequently mentioned in 2008, as they were listed as major ones in 14 out of 20 positions (70%). Likewise, in our study *Languages* is the most requested high-level hard skill in 9 out of 14 developer roles (64%). Interestingly, no library or framework appeared among the popular hard skills in 2008. This represents a contrast with our study, as *Libs & Frameworks* are the second most demanded hard skill for six roles. Finally, we observe minor changes in terms of soft skills. For instance, *leadership* is mentioned in 20% of the opportunities in 2008. Similarly, terms like *leadership* (7.6%), *lead* (4.0%), and *leading* (3.5%) are also featured among the most popular soft skills in our study. However, these numbers should be interpreted with caution, since the studies followed different methodologies.

²The sample size was determined after specifying a limit of 95% confidence level and 5% confidence interval in a well-known sample size calculator, available at <https://surveysystem.com/sscalc.htm>.

6. Conclusion

For candidates to IT jobs, our study reveals the distribution of the most demanded hard skills for 14 contemporary developer roles. For example, people who are applying to a frontend position should describe the *programming languages* and *libraries & frameworks* they master, since 49.5% and 34.1% of the tags used by companies when hiring such developers refer to these hard skills. Furthermore, the study confirms the importance companies give to soft skills. The most requested ones are related to communication, collaboration, and problem-solving. We plan to extend the study described in this paper in two major directions. Firstly, by including other jobs portals besides Stack Overflow in our analysis [7]. Secondly, by investigating machine learning approaches to identify soft skills automatically [8, 3]. The data used in this study is available at <https://doi.org/10.5281/zenodo.3906955>.

References

- [1] J. E. Montandon, L. L. Silva, M. T. Valente, Identifying Experts in Software Libraries and Frameworks among GitHub Users, in: 16th International Conference on Mining Software Repositories (MSR), 2019, pp. 276–287.
- [2] X. Xia, Z. Wan, P. S. Kochhar, D. Lo, How Practitioners Perceive Coding Proficiency, in: International Conference on Software Engineering (ICSE), 2019, pp. 1–12.
- [3] L. Sayfullina, E. Malmi, J. Kannala, Learning Representations for Soft Skill Matching, in: Lecture Notes in Computer Science, Springer International Publishing, 2018, pp. 141–152.
- [4] A. K. Saha, R. K. Saha, K. A. Schneider, A Discriminative Model Approach for Suggesting Tags Automatically for Stack Overflow Questions, in: Working Conference on Mining Software Repositories (MSR), 2013, pp. 73–76.
- [5] D. Spencer, Card sorting: Designing usable categories, Rosenfeld Media, 2009.
- [6] M. Warrens, Five Ways to Look at Cohen’s Kappa, Journal of Psychology & Psychotherapy 05 (2015) 1–4.
- [7] C. Litecky, A. Aken, A. Ahmad, H. J. Nelson, Mining for Computing Jobs, IEEE Software 27 (1) (2010) 78–85.
- [8] M. Bastian, M. Hayes, W. Vaughan, S. Shah, P. Skomoroch, H. Kim, S. Uryasev, C. Lloyd, LinkedIn skills: large-scale topic extraction and inference, in: ACM Conference on Recommender systems (RecSys), ACM Press, 2014.