Türkiye’deki Yazılım Geliştirme İşleri İçin İstenen Becerilerin Metin Madenciliği Kullanılarak Analiz Edilmesi

Görkem Giray [0000-0002-7023-9469]
Bağımsız Araştırmacı, İzmir, Türkiye
gorkemgiray@gmail.com


An Analysis of Desired Skills for Software Development Jobs in Turkey Using Text Mining

Görkem Giray [0000-0002-7023-9469]
Independent Researcher, İzmir, Turkey
gorkemgiray@gmail.com

Abstract. Understanding the desired skills in software industry is critical in many aspects, including designing university curricula, organizing trainings, launching online courses, guiding software developers for self-development. To this end, we crawled and analyzed 1,597 job ads from kariyer.net to find most
desired technical and soft skills in Turkish software development industry. Our analysis reveals a substantial demand for experience in SQL, JavaScript and HTML/CSS languages in line with the recent trends worldwide. ASP.NET and MS SQL Server are the dominant web framework and database whose knowledge is needed most in job ads according to our dataset. Knowledge on Linux operating system is the most desired skill consistent with StackOverflow developer survey 2019 results. Visual Studio is the most desired development environment and .NET is the dominant framework. According to our dataset, experience on only a few software testing tools are sought by employers. The most desired soft skills are team work, communication and analytical/problem solving skills.

**Keywords:** Technical Skill, Soft Skill, Turkish Software Development Industry, Text Mining, Topic Modeling.

### 1 Introduction

Understanding the desired skills in software industry is critical in many aspects, including designing university curricula, organizing trainings, launching online courses, guiding software developers for self-development. To this end, we crawled and analyzed job ads from kariyer.net to answer the following research questions (RQs):

- **RQ1:** What are the most desired technical skills (experience on languages, web frameworks, databases, platforms, development environments, libraries, tools and software testing tools) in Turkish industry?
- **RQ2:** What are the most desired soft skills (non-technical skills) in Turkish software industry?
- **RQ3:** How can we classify job ads and analyze the various aspects of these clusters?

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 explains our research method in detail. In Section 4, we present our results and comment on these results. Section 5 includes the threads to validity. Finally, Section 6 concludes the paper.

### 2 Related Work

Table 1 displays the related work on analyzing desired skills by processing job ads. [1] has the biggest dataset by far compared to the other studies as well as our study. The rest of the studies have analyzed thousands of job ads and have a dataset of a similar scale with our study. Our study aims to analyze the desired skills for Turkish software industry and differs from the other studies in this respect. [2] presents a similar analysis for Thailand.
Table 1. Related work on analyzing desired skills by processing job ads.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Scope</th>
</tr>
</thead>
</table>
• analyzed approximately 210,000 job ads. |
• analyzed 500 ads for IT positions on the soft skills mentioned. |
• analyzed 1,736 job ads. |
| [5]  | 2018 | • gathered German information systems related job starter postings from four large online platforms: two general platforms (stepstone.de and monster.de) and two platforms specializing in entry-level jobs (get-in-it.de and absolventa.de).  
• analyzed 6,848 job ads. |
• analyzed 2,229 job ads. |

Besides the similar studies in the literature, we used the results of two surveys [6], [7] in discussing our results.

3 Research Method

Fig. 1 displays the research method we applied in this study. In the first step, we crawled 2,914 job ads from kariyer.net, a web site for job and employee search in Turkey. We crawled the job ads categorized under the main category named “informatics and telecom”. We included the sub-categories related to development, database/data warehouse, mobile programming, project management/business analysis, testing, product management and web-based application development. Even though we carefully selected the sub-categories to be included, there were many job ads that are irrelevant to software development and hence to our analysis. We manually filtered out the irrelevant job ads by checking job titles (step 3 in Fig. 1). There were many job ads that are not related to software development, such those seeking for networking experts, helpdesk technicians, etc. In addition, we also filtered out job ads whose working places are out of Turkey. After filtering out irrelevant job ads, we
ended up with a dataset consisting of 1,597 job ads. In the fourth step, we prepared our dataset for mining. We translated all job ads written in English to Turkish using Google Translation service.

In parallel with first, third and fourth steps, we also formed a list of skills to be searched in job ads (step 2 in Fig. 1). We used the items in StackOverflow’s survey [6] conducted this year for languages, web frameworks, databases, platforms, development environments, libraries and tools. Since this survey does not include testing tools, we extracted a list of software testing tools from Wikipedia. We formed a unified list for soft skills by using the earlier studies in the literature [3], [4], [8].

In the fifth step, we mined the skills in job ads. We tried to minimize the errors due to misspelling and different acronyms used. We tried to mine all possible strings for a skill. For instance, we used the following phrases for MS SQL Server: (1) “microsoft sql server”, (2) “ms sql”, (3) “mssql”, (4) “ms sql server”, (5) “mssql server”, (6) “sql-server”, (7) “ms.sql”. We obtained the frequencies of each skill as the result of fifth step.

In the sixth step, we formed document vectors for each job ad. These vectors only included the skills we obtained in the fifth step. We assumed that each job ad is represented with these skills. This type of approach is named as “bag of words” representation. We provided these document – word vectors to the seventh step as an input to build a topic model. We formed many topic models using Latent Dirichlet Allocation (LDA) [9] in the seventh step and finally came up with a reasonable model. This topic model consists of five topics. α and β parameters used to build this model was 0.85 and 0.76 respectively (for the details of these parameters refer to [10]). Using the topic model obtained, we clustered all job ads into five groups.

At the eighth step, we analyzed all of the findings and reported as presented in the next section.

We used Jupyter Notebook as a development environment. We developed our software using Python programming language. We used pandas library for data processing, nltk library for natural language processing and gensim library for building topic models. We also used Microsoft Excel for storing and analyzing data.
Fig. 1. The research method used in this study.
4 Results and Discussion

4.1 RQ1: Desired Technical Skills

Programming, Scripting, and Markup Languages. Fig. 2 (a) shows top 10 programming, scripting, and markup languages that are most frequently mentioned in computing job ads. Fig. 2 (b) displays top 10 programming, scripting, and markup languages that are used by developers according to [6]. SQL is by far the most desirable skill among languages and this finding is consistent with those reported in [7]. SQL is the third most used language according to [6]. JavaScript and Java are the two other languages desired by employers and used by developers as identified in this study, [7] and [6]. Python is the fastest-growing major programming language today [6], [7] also identified Python as the fifth most desired language. On the other hand, Python is desired relatively less according to our analysis. [7] also reported that the highest gap between demand and supply will be in Python language skill within five years in Turkey. HTML/CSS is by far the most desired skill among languages and this finding is consistent with those reported in [7]. HTML/CSS is the second most used language according to [6] and C# is reported as one of the most desired languages in top five by [7] and is ranked as seventh most used language by [6].

![Fig. 2. Top 10 programming, scripting, and markup languages that are most frequently (a) mentioned in computing job ads; (b) used by developers according to [6].](image)

Web Frameworks. Fig. 3 (a) shows top 10 web frameworks that are most frequently mentioned in computing job ads. Fig. 3 (b) displays top 10 web frameworks that are used by developers according to [6]. ASP.NET is the most desired web framework in Turkish job ads. jQuery and Angular follow ASP.NET with a slight difference. jQuery is the most broadly used of these web frameworks according to [6]. In addition, [6] reported that this year more developers say they use React.js than Angular, a switch from last year. According to our dataset, employers in Turkish software industry do not seek for React.js skill that much. Spring and Vue.js are the other two web frameworks sought by employers and these two frameworks are in top ten most used web frameworks according to [6].
Nearly one fifth of the employers look for professionals who can use MS SQL Server. Oracle, MySQL, and PostgreSQL are the other DBMSs whose knowhow is important to employers. According to [6] MySQL is the most commonly used database, like last year. PostgreSQL has taken the second spot this year, edging ahead of MS SQL Server, which is the most desired database based on our dataset. Another interesting remark is that SQLite, which is the fourth most commonly used database according to [6] is not included within top ten in our dataset.

Platforms. Fig. 5 (a) shows top 10 platforms that are most frequently mentioned in computing job ads. Fig. 5 (b) displays top 10 platforms that are used by developers according to [6]. Linux is the most desired platform by Turkish employers and most common platform used worldwide according to [6]. Employers in Turkey look for developers working with mobile operating systems, i.e. iOS and Android. Windows operating system is still among the most used operating systems. Docker is also gaining attraction both worldwide and among Turkish employers.
Development Environments. Fig. 6 (a) shows top 10 development environments that are most frequently mentioned in computing job ads. Fig. 6 (b) displays top 10 development environments that are used by developers according to [6]. The job ads in our dataset do not include too much information on development environments whose knowhow is important to employers. Visual Studio is the top development environment sought by Turkish employers and used worldwide [6]. On the other hand, Turkish employers do not ask for experience specifically on Visual Studio Code, which is the most commonly used development environment worldwide [6].

Other Frameworks, Libraries, and Tools. Fig. 7 (a) shows top 10 other frameworks, libraries, and tools that are most frequently mentioned in computing job ads. Fig. 7 (b) displays top 10 other frameworks, libraries, and tools that are used by developers according to [6]. While Node.js is the most commonly used tool worldwide [6], employers in Turkey do not look for professionals with Node.js experience. More developers say they use .NET than .NET Core worldwide [6] and the same is valid for our dataset. React Native knowhow is looked by one tenth of employers. While Pandas is popular library worldwide [6], experience in Pandas is not important to Turkish employers.
Testing Tools. The survey whose results are reported in [6] does not include any specific question addressing software testing tools. Due to the increasing importance of software testing and availability of testing tools, we mined our dataset to explore the frequencies of testing tools in computing job ads. We formed a list of testing tools from Wikipedia under “software testing tools category” (https://en.wikipedia.org/wiki/Category:Software_testing_tools). We included the tools under “free software testing tools”, “graphical user interface testing tools”, “load testing tools”, “security testing tools” and “unit testing frameworks”. Our final list included 116 tools. Surprisingly only three tools, i.e. Selenium, NUnit, and JUnit, have been mentioned 19, 16, and 16 times respectively in 1597 job ads. A few of the tools have been mentioned less than 10 times in computing job ads. According to our dataset, employers do not seek for experience on software testing tools.

4.2 RQ2: Desired Soft Skills

Besides technical skills, soft skills (non-technical skills) also play an important role in software development. Therefore, employers generally prefer to include the soft skills they require for their open positions. To find out the frequencies of soft skills in our dataset, we formed a list of soft skills by exploring the earlier studies [3], [4], [8].

Fig. 8 shows the frequencies of soft skills in our dataset. There is a high demand for team work, communication and analytical/problem solving skills. Since software development is generally conducted by a team, the ability to work in teams is very important. Moreover, communication skills are also important and becoming more and more important since agile software development methods suggest frequent and oral communication in teams.
4.3 RQ3: Clustering Job Ads

Positions in software development can be classified according to skill requirements. Different positions require different sets of skills. Therefore, we clustered the job ads into five groups using topic modeling, similar to the approach in [5]. In topic modeling, number of topics is an input and does not have an initial value [10]. After several iterations, we obtained five topics to cluster the job ads in our dataset. Table 2 shows these five clusters and the most frequently mentioned top five technical/soft skills in these clusters.

The job ads in the first cluster mainly expect experience in languages and tools launched by Microsoft. This cluster includes the highest number of job ads, i.e. 429, so, according to our dataset, we can infer that employers look for experience on Microsoft products at most. The second cluster includes SQL, Java, JavaScript languages; Spring application framework for Java and Oracle database. The third cluster includes C and C++ languages and three operating systems, i.e. iOS, Android and Linux. As we can see from the frequencies, the third cluster does not have dominant skills. In the fourth cluster, only SQL has a relatively high frequency and the rest of the items in top five have less than 10% frequencies. On the other hand, the soft skills are much more frequently mentioned in this cluster compared to the others. Software developer, as most of the other professions, should use their soft skills more as they gain more managerial responsibilities [11], [12]. Therefore, we can infer that this cluster mostly includes job ads that are seeking for managerial roles. The dominant theme for the fifth cluster is web application development. The top five languages and tools sought by employers in this cluster are HTML/CSS, JavaScript, jQuery, SQL and Angular.
Table 2. Five clusters of job ads and the most frequently mentioned technical/soft skills in these clusters.

<table>
<thead>
<tr>
<th>Cluster</th>
<th># of job ads</th>
<th>Skill</th>
<th>Technical Skills</th>
<th>Frequency (%)</th>
<th>Soft Skills</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>429</td>
<td>SQL</td>
<td>74</td>
<td>Team work</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C#</td>
<td>67</td>
<td>Problem solving</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.NET</td>
<td>57</td>
<td>Communication</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASP.NET</td>
<td>50</td>
<td>Learning</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MS SQL Server</td>
<td>47</td>
<td>Creativity</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL</td>
<td>77</td>
<td>Team work</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Java</td>
<td>69</td>
<td>Communication</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>258</td>
<td>Spring</td>
<td>41</td>
<td>Problem solving</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>40</td>
<td>Motivation</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>JavaScript</td>
<td>29</td>
<td>Learning</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>270</td>
<td>C</td>
<td>32</td>
<td>Communication</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C++</td>
<td>29</td>
<td>Team work</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>iOS</td>
<td>23</td>
<td>Creativity</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Android</td>
<td>22</td>
<td>Problem solving</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linux</td>
<td>21</td>
<td>Learning</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>352</td>
<td>SQL</td>
<td>36</td>
<td>Team work</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>JavaScript</td>
<td>7</td>
<td>Problem solving</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HTML/CSS</td>
<td>7</td>
<td>Communication</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>7</td>
<td>Learning</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Java</td>
<td>6</td>
<td>Creativity</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>288</td>
<td>HTML/CSS</td>
<td>84</td>
<td>Team work</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>JavaScript</td>
<td>78</td>
<td>Problem solving</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>jQuery</td>
<td>54</td>
<td>Communication</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQL</td>
<td>49</td>
<td>Learning</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Angular</td>
<td>45</td>
<td>Creativity</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

5 Threats to Validity

Our study has some threads to validity. First, we crawled only kariyer.net as the data source. There are other web sites and environments for job search, hence kariyer.net does not represent the whole software industry in Turkey. On the other hand, it is one of the most popular and widely-used online platforms for job search. Moreover, we crawled the web site during two days in May 2019. Therefore our dataset reflects only a snapshot for skill requirements in Turkey for software development. Consequently, the results of our analysis have limited generalizability. To check the generalizability
of our results, we compared our results with another survey in Turkey [7] and a survey done worldwide by StackOverflow web site [6] whenever comparable data are available in the results of these surveys. Second, we manually filtered out irrelevant job ads. Therefore, this step is subject to a researcher bias. We think that this would be limited since most of the eliminated job ads are clearly irrelevant to software development, such as job ads seeking for networking experts, help desk technicians etc.

6 Conclusions

In this paper, we provide an overview of the technical and soft skills desired by employers in Turkish software industry. Our analysis reveals a substantial demand for experience in SQL, JavaScript and HTML/CSS languages in line with the recent trends worldwide. ASP.NET and MS SQL Server are the dominant web framework and database whose knowledge is needed most in job ads according to our dataset. Knowledge on Linux operating system is the most desired consistent with StackOverflow developer survey 2019 results. Visual Studio is the most desired development environment and .NET is the dominant framework. According to our dataset, experience on only a few software testing tools are sought by employers. The most desired soft skills are team work, communication and analytical/problem solving skills.

References