A Study of Remote and On-site ICT Labor Market Demand using Job Offers from Stack Overflow

Conference Paper · September 2021

DOI: 10.1109/SEAA53835.2021.00039

4 authors, including:

Konstantinos Georgiou
Aristotle University of Thessaloniki

Lefteris Angelis
Aristotle University of Thessaloniki

A Study of Remote and On-site ICT Labor Market Demand using Job Offers from Stack Overflow

Ioannis Apatsidis
School of Informatics
Aristotle University
Thessaloniki, Greece
iapatsid@csd.auth.gr

Konstantinos Georgiou
School of Informatics
Aristotle University
Thessaloniki, Greece
konsgeor@csd.auth.gr

Nikolaos Mittas
Department of Chemistry
International Hellenic University,
Kavala, Greece
nmittas@chem.ihu.gr

Lefteris Angelis
School of Informatics
Aristotle University
Thessaloniki, Greece
lef@csd.auth.gr

***Abstract*— As the industry is moving towards digitalized solutions and practices, a growth in remote working has been observed with companies embracing flexibility for their workforce. Global crises, such as the coronavirus pandemic, have also accelerated this process, transforming the labor market. This trend is reflected in job portals, that contain an increasing number of remote job advertisements. Recognizing this evolving change, we perform a thorough study in Stack Overflow, to examine the main characteristics of remote working that discriminate it from its on-site counterpart. By collecting and analyzing 8514 job posts and leveraging text mining and graph theory methodologies, we attempt to pinpoint the primary elements that define each category, from dominant technologies to job positions and top seeking industries. The findings suggest that remote working presents differences from traditional working, being mainly associated with the software engineering sector and with well-known software development and data analytics technologies.**

Keywords*— *Remote Working, Job Analytics, Stack Overflow, Software Engineering

I. Introduction

The working landscape is a constantly evolving system requiring continuous monitoring regarding its evolution. The fast-moving technological advances have certainly transformed the way individuals perceive the labor market, while the digital transformation and the vision of Industry 4.0 have posed significant challenges to all key industries moving away from traditional, physical types of jobs and welcoming flexible, remote means of working. In some cases, this evolution has been natural, keeping pace with technological innovation [1], while in other cases, change was mandatory to conform to a global crisis [2]. Particularly in 2020, marked as the year of the coronavirus pandemic, the increase of remote working is more notable than ever.

Since the outbreak of the pandemic, and even before it, teleworking has become a staple for millions of employees worldwide. As a result, remote jobs have highly increased since businesses and organizations have been forced to adjust to these new circumstances and encourage working from home [3,4,5]. Apart from this crisis, the great expectations from the rising of the new digitalized era have contributed to the tremendous acceleration of flexible working practices. In addition, remote working is a mean of economic relief, as it provides organizations with the opportunity to cut down on infrastructure and physical expenses.

The motivation of this study is the need to explore and understand the dynamics of remote and on-site jobs from the scope of employment seekers, employers and employees. On-line portals constitute a great source for investigating these characteristics, since they provide a wealth of information about the required skills related to current trends of labor market digital needs. In the current study, we explore the rising phenomenon of remote working in jobs related to developers and digital specialists and attempt to shed light on their characteristics. As a primary data source, we used Stack Overflow (SO), a well-known question and answer site for professional and enthusiast programmers. Our study introduces a robust framework aiming (a) to contribute to the current scientific knowledge regarding ICT remote and on-site labor market, (b) to benefit a wide range of stakeholders involved in the job offer and seeking processes and (c) to act as a benchmark for further analysis.

II. Related Work

The area of job analytics refers to the study of the content of job posts to extract knowledge about the labor market and its characteristics [6]. Some indicative methodologies are the analysis of jobs about information systems [7], or the clustering of job titles based on the required skillsets [8]. Content analysis of advertisements [9,10] collected from HigherEdJobs.com and Monster.com, are conducted to understand the workplace, skills, and duties of the employees. Moreover, content studies are performed to point out the required competencies and knowledge for multiple areas [11], such as factory plants [12], big data [13] as well as jobs related to Industry 4.0 [14]. In addition, a multisource study has been conducted by Papoutsoglou et al. [15] concerning the evolution of the labor market. Hauff et al. [16] propose an automated way to match job advertisements with Github profiles using the ReadMe file. Another work by Daneva et al. [17] focuses on the demand of requirements engineering jobs in the Netherlands labor market. In another study [18], the authors analyze job advertisements and curriculum vitae to improve the quality of the recruitment process. Text mining is also used to classify candidates by processing resumes [19] and to predict future performance of employees [20]. Karakatsanis et al. [21] match job advertisements with occupation descriptions and identify the most in-demand occupations in the market.

Remote working has been also the subject of research activities focusing on the software engineering and ICT sector. As agile management plans facilitate the working environment [22], software specialists work remotely, reaping benefits. The COVID-19 pandemic accelerated this transition, indicating the potential of remote working practices regarding communication and interaction [23,24]. Finally, the prospects of inclusiveness and identity disclosure appeal to software developers, who opt to work from a distance rather than relying on physical presence [25]. In general, the concept of a "virtual worker" [26] who is characterized by teamwork and rapid problem solving is gaining attention as a reliable workforce model.

Regarding the exploration of developer interactions and technologies that are useful to employers in SO, Westwood et al. [27] study the relationships and co-occurrences of the tags and predict tag popularity in SO. A graph-based study [28] is also conducted for examining the associations and evolution of the software technologies based on tags. The activities of developers, particularly in the COVID-19 era, are highlighted by Georgiou et al. [29], along with the increased interest for seeking technological solutions to tackle the pandemic. Papoutsoglou et al. [30] emphasize on human factor and peopleware, detecting competencies and trends from job advertisements. Montadon et al. [31] recognize the most important competencies that companies are looking for in new developers and their evolution over the years.

III. Research Objectives and Research Questions

The main research objective behind the current paper is to thoroughly investigate the characteristics of traditional and remote working and profile the primary traits that differentiate this flexible mode of employment. This, in turn, will allow us to deduce, whether the adjustment of the industry in this new digitalized and remote climate is competent or whether on-site working is still considered the norm.

Considering the above aims, we formulate the following *Research Questions* (RQs):

[RQ$_1$] *What are the differences in the characteristics of on-site and remote job posts?*

Motivation: As remote job posts are increasingly gaining traction on SO, our inclination behind RQ$_1$ is to explore and portray the differences between on-site and remote posts by analyzing metadata and comparing these two categories.

[RQ$_2$] *Which positions are primarily sought in remote and on-site posts based on the job titles*?

Motivation: To provide insights on the labor market demand for specialists, RQ$_2$ aims to leverage the posts titles in order to trace job positions, that are mostly sought from recruiters and how are they differentiated in traditional and remote working.

[RQ$_3$] *Which are the most prominent technologies for remote and on-site positions and what are the interconnections between them*?

Motivation: In RQ$_3$, our purpose is to shed light on the technical skills sought by recruiters. Each post can be labelled with several tags as a general overview of the technological content of the job. In addition, as posts are significantly different based on the desired position, different technologies may be utilized in conjunction, to efficiently portray the required technological skillsets. Hence, our goal in this RQ is two-fold. Apart from detecting the most popular technologies for each category, we also wish to explore their co-occurrences in job posts, interpret their interconnections and identify potential differences in each category.

IV. Methodology

The methodology of our approach is illustrated in Figure 1 and can be described as a four-step process consisting of (*i*) *data collection*, (*ii*) *feature extraction*, (*iii*) *data cleaning and preprocessing* and (*iv*) *data analysis*, followed by the results and discussion of the findings.

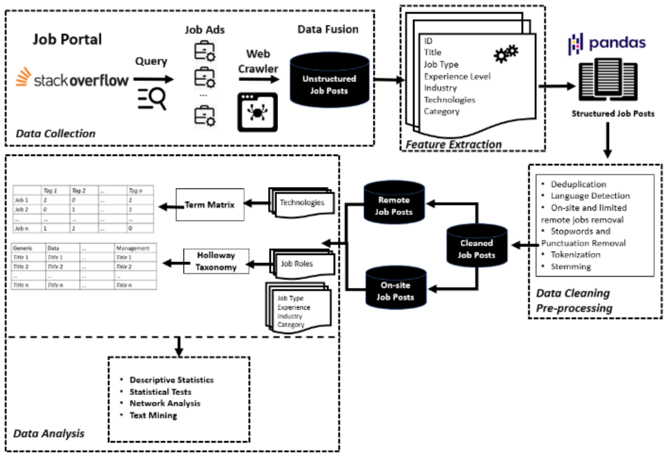

Fig. 1. Methodology of the study

A. Data Collection

We decided to retrieve relevant data from SO, which includes a dedicated portal for job posts. We chose SO and not another job platform since SO, primarily being a forum related to ICT [32,33] and contains a plethora of jobs related to this field. Moreover, SO is proven to be quite efficient in detecting technological demands [34] which can be reflected in job posts. As the objective of this study was focused on examining all the jobs belonging either to the on-site or remote categories, we retrieved all the available posts in a specific timeframe.

To retrieve the required job posts, we utilized a robust collecting strategy based on flexible scraping of web objects. More specifically, a data collector was developed using *Python* to scrape available posts and collect metadata. By applying this data collection process in SO, a dataset of 8514 job posts, published between November 1, 2020 and December 5, 2020, was collected. The period and length of the timeframe were specifically selected to cover one month, indicating the objective of the current study, which is to investigate the differences between two job categories in a short time span. To further clarify the extracted information, we provide, in Figure 2, an illustrative example of a post highlighting fields of interest.

As showcased, the title of a post provides a brief description of the role associated with it. The category of each job (on-site or remote) was extracted by the specific characterization gave by the recruiter, as indicated in Figure 2. The next section contains informative metadata about specific details such as the job type (full time, contract, or internship), experience level or the industry associated with the recruiting company. Finally, the tags field contains the technical skills required for a particular post. It should be emphasized that these tags can be referring to technologies (e.g. “python”), tasks (e.g. “web-scraping”) or products (e.g., “api”). It should be noted that the job role was discarded, in favor of a more robust taxonomy that will be described below and company data (size, type) were not examined as they were out of the scope of the current study.

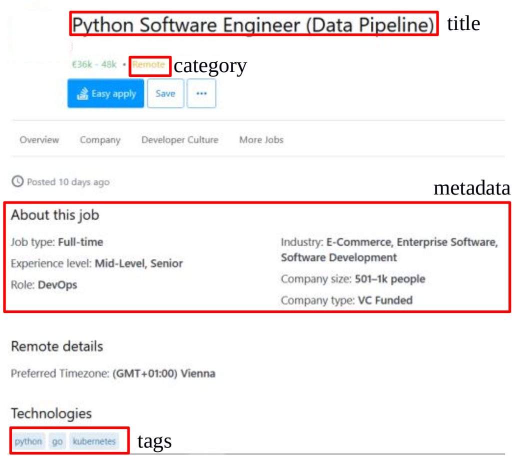

Fig. 2. Example of job post

B. Feature Extraction

The main unit of analysis is the *job post* (referred as post), represented by a web object containing multiple data and properties and can be expressed as a tuple J in the form of

$$J = (id, c, t, ty, ex, in, tech) \quad (1)$$

where *id* is the identification number of the post, *c* is the category (remote, on-site), *t* is the title, *ty* is the job and *ex* is the required experience level (e.g. *senior*). As far as the experienced level is concerned, for jobs that had more than one values, we kept the minimum degree of experience required. For example, a post with experience levels of “Senior, Lead” would return only the “Senior” value.

TABLE I. Extracted Features

Feature	Description
Category	Job category (Remote, On-Site)
Title	The title of the job
Type	The type of employment
Experience Level	The required experience
Industry	The industrial sector
Technologies	Required technical skills

Moreover, *in* represents the industry to which the company that posted the job belongs and *tech* are the technologies accompanying the post, expressing the required technical skills. In Table 1, we present the final extracted features for each post.

C. Data Cleaning and Preprocessing

The retrieved job posts were comprised of unstructured elements (e.g. *Title*) and thus, we had to employ *Natural Language Processing* (NLP) *and Text Mining* (TM) methodologies in order to transform them into interpretable form. However, for fields that had a distinct number of elements, such as the *Type* or *Experience Level*, we decided to keep them in their initial form.

After removing duplicate posts, we also excluded all non-English jobs. In addition, some posts categorized as on-site and limited remote, providing both physical and remote working schedules were removed too. These cleaning procedures reduced the number of job posts to 7070. *Titles* were then subjected to (*i*) stopwords and punctuation removal, (*ii*) tokenization and (*iii*) stemming.

Regarding the *Technologies* field, we initially gathered the set of unique tags from job posts in a list to perform a merging process. This was achieved by simplifying tags that referred to language or software versions. For example, the tag “python 3.x” was converted to “python”. However, as tags are determined individually by the recruiters, posts had varied terms that may not be relevant to technologies but rather to technological concepts. Thus, we utilized a pre-defined lexicon of 182 terms to filter the set of tags and keep only the ones relevant to technologies. The lexicon was extracted from the SO Developer Survey[1], that is launched by SO annually and attempts to pinpoint the main technologies used by developers in their questions and answers. Then, for each job, the filtered tags were formulated into Boolean variables with values of 0/1 indicating their absence or presence in the job post.

D. Data Analysis

For RQ_1, we accumulated the collected posts and conducted descriptive analysis to highlight the main characteristics and differences between the two categories, based on their metadata (*Type*, *Experience Level*, *Industry* fields). Statistical hypothesis tests (e.g. the *chi-square test of independence*) were also conducted to examine, whether there was noted a statistically significant association between the collected features.

Regarding the discovery of the primary job positions in RQ_2, we turn our attention to the preprocessed titles and employed advanced TM techniques. Each title was initially transformed into a set of n-grams representing a sequence of terms that were found into the text. After experimenting with several values for n, we decided to compute the bigrams as they provide a clearer picture of title’s content.

Our next step was to define a concise taxonomy of job positions that encapsulated titles under several domains expressing different duties and responsibilities for specialists. This approach allowed us to map the extracted titles to specific categories and examine, which categories are

[1] https://insights.stackoverflow.com/survey/2020

dominant in on-site and remote posts. In our study, we decided to follow the methodology proposed by [35], which gathers several positions and distributes titles to them, in an taxonomy. The proposed taxonomy covers a wide range of positions on demand, from data and software related positions (*Data*, *Generic*, *Product*) to technical or maintenance jobs (*Systems*, *Quality Assurance* (QA), *Operations*) while also including jobs regarding security (*Security and Compliance*, *IT*) and managerial or customer engagement positions (*Solutions or Sales*, *Management*) (Table 2).

TABLE II. TAXONOMY STRUCTURE [35]

Job Positions	**Indicative Job Titles**
Generic	Developer; Software Developer; Programmer; Engineer; Software Engineer; SDE (Software Development Engineer)
Systems	Systems Engineer, Systems Architect, Systems Analyst, Software Architect
Product	Product Engineer; Full stack Engineer; Backend Engineer; Frontend Engineer; Web Developer; Application Engineer
Data	Machine Learning Engineer; Data Scientist; Data Architect; Data Analyst; Data Engineer
Operations	DevOps Engineer; Site Reliability Engineer; System Administrator; Cloud Architect Infrastructure Engineer
Quality Assurance (QA)	QA Engineer; SDE in test (SDET); Test Engineer; Quality Engineer; Automation Engineer
Solutions or Sales	Solutions Engineer; Customer Support Engineer; Solutions Architect; Sales Engineer; Professional Services Engineer
IT	IT Administrator; System Administrator; Network Administrator; Database Administrator
Security and Compliance	Security Engineer; Security Architect; Information Security Analyst; Information Security Architect
Management	Engineering Manager; Development Manager; Software Engineering Lead; Senior Software Engineering Lead; Director

Based on this categorization system, we matched job posts to the categories by seeking the titles of each category into the bigrams of every post. Evidently, this categorization process is not mutually exclusive, and one job may be mapped to more than one category. In our analysis, we took that into consideration and decided to match the posts to all possible categories.

Concerning RQ_3, we initially calculated the frequency of each tag across the whole dataset to pinpoint technologies that are mostly preferred by recruiters. In parallel, we made use of graph theory to explore the interconnections between technologies. The analysis was based on the *Association Rule Graph* (ARG) proposed by Cui et al. [36]. We chose this approach as the labelling of job posts with technological skills consists a tagging system with inter-connected aspects.

An ARG network between tags requires the evaluation of three measures that can be computed between a tag_i and a tag_j. The frequency ($freq(tag_i)$) is the total number of appearances of a tag in the entirety of posts. The support ($supp(tag_i, tag_j)$) expresses the number of co-occurrences between two tags in a job post. Finally, the confidence metric $\left(conf\left(tag_i \rightarrow tag_j\right)\right)$ is the probability of detecting tag_j in a job post that has been labelled with tag_i given that $freq(tag_i) < freq(tag_j)$. The confidence metric is calculated by dividing the support metric between two tags with the frequency of tag_i.

$$conf\left(tag_i \rightarrow tag_j\right) = \frac{supp(tag_i, tag_j)}{freq(tag_i)} \qquad (2)$$

Based on the above formula, an ARG network is defined as a directed graph, where each tag_i is represented by a vertex associated with a weight that is equal to $freq(tag_i)$. Moreover, for each pair of tags $\{tag_i, tag_j\}$ presenting a number of co-occurrences in the job posts set, a directed edge capturing the strength of the association between the two tags is defined along with a weight computed by the $conf\left(tag_i \rightarrow tag_j\right)$ metric. It should be noted that the process of extracting an ARG network was implemented for both categories (remote, on-site).

V. Results

In this section, we present the findings of our study based on the RQs.

[RQ_1] *What are the differences in the characteristics of on-site and remote job posts?*

An initial inspection of the job distributions between the two examined categories, indicates that on-site posts are still the majority, with 81% being on-site and 19% being remote. This is not an unexpected finding, as remote working is still evolving, gaining popularity among enterprises and corporations. However, the mere fact that almost 20% of jobs, even in a limited timeframe, are remote is an encouraging sign for its growing presence. In addition, almost all posts concern full-time employment, as the market requires specialized professionals that can cope with the demands of the employers while providing extensive services. Full time jobs constitute 98.4% percent of on-site posts and 93.9% of remote posts. Remote job posts offer limited possibilities of internships, with only two posts being internships. This is expected because being trained as an intern is significantly harder when working from a distance. On-site jobs appear more accepting of internships, as an intern can easily be trained by seasoned members of a company.

The distribution of the required *Experience Level* for the two levels of *Category* (Figure 3) indicates a different approach when seeking individuals to fill a job post. Indeed, the chi-square test of independence revealed a statistically significant association between *Category* and *Experience Level*, $\chi^2(5) = 151.84, p < 0.001$. It seems that due to the rising demand for remote working and the opportunities it offers for employers, companies that incorporate remote opportunities to their business models actively prefer senior experience workers.

Mid-level positions are still highly posted but employees in such positions may not have the required efficient performance. Moreover, managerial and lead positions are notably decreased, possibly because individuals with such prestigious jobs would be required to be present in their organizations to fulfill their duties more efficiently and would be recruited for on-site working.

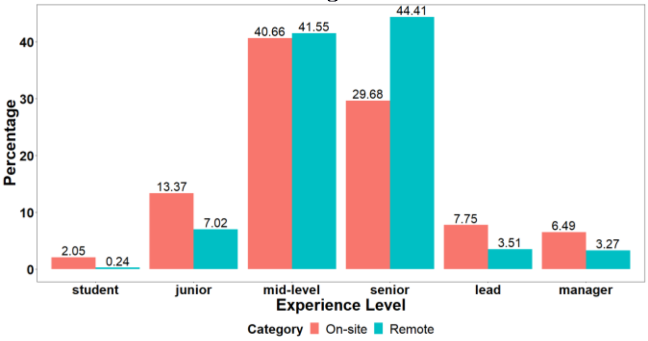

Fig. 3. Distribution of experience levels (on-site vs. remote)

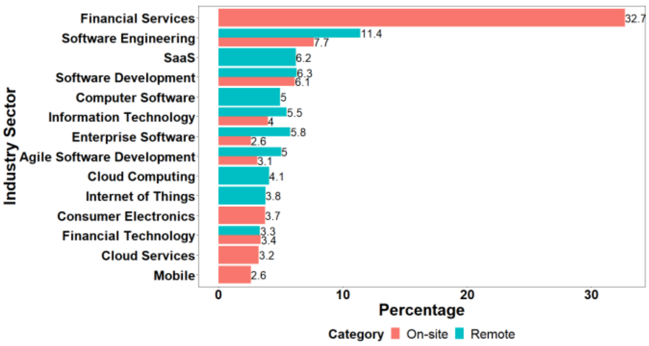

Fig. 4. Top ten industries for each category (on-site vs. remote)

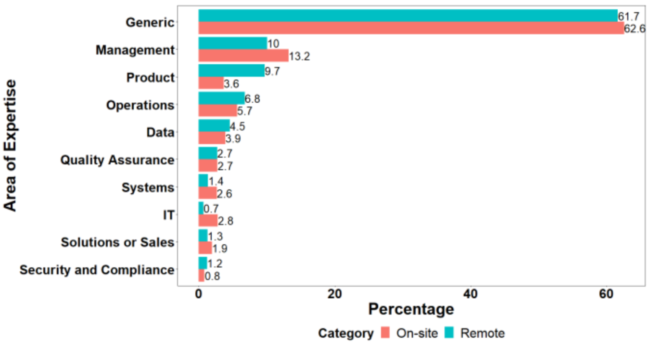

Fig. 5. Distribution of job positions (on-site vs. remote)

On-site jobs are addressed primarily to mid-level experience professionals, with senior experience posts closely following, as employers understand that having a workforce comprising entirely of senior specialists is a challenge. In addition, junior level jobs are also present, in an unexpected percentage, indicating the trend of the industry to offer some opportunities to emerging talents. It should be clarified though that on-site posts entail the risk of hiring an inexperienced individual while remote recruiters are more careful when selecting candidates. As expected, management and lead experience levels are present in higher percentages.

Regarding the top ten leading industries (Figure 4) that seek employees in SO, we observe a clear difference between the two categories. A statistically significant association, $\chi^2(13) = 1955.50, p < 0.001$, is also observed between *Category* and *Industry* fields. In general, remote job posts are primarily associated with software services and development and attract highly skilled IT specialists. This is made evident by the presence of Software Engineering and Software Development categories, as well as more generic categories such as Computer Software and Information Technology. Other sectors involve the distribution of software (SaaS, Enterprise Software), as well as cutting-edge practices such as Internet of Things and Cloud Computing. These trends indicate a tendency in the industry of preferring remote employees for IT services and software production when this is applicable.

On-site posts have a notable difference with Financial Services holding the lead. This advantage can be attributed to the fact that Financial Services require personal interactions and human interventions for the handling of economic data. Thus, any specialists employed in this sector would ideally be required to be physically present in the company. This is further proved by the existence of jobs relevant to Financial Services. In addition, we can discern some industries that heavily rely on physical working such as Consumer Electronics that cannot function in a remote environment. Apart from these categories, on-site posts are also related with IT industries, but in a lesser extent than the remote category that is exclusively comprised by IT sectors.

[RQ$_2$] *Which positions are primarily sought in each category based on the job titles*?

The mapping of the titles to the categories of the defined taxonomy allowed us to examine which job positions are comprising remote and on-site posts, as shown in Figure 5. Once more, the chi-square test of independence revealed a statistically significant association between *Category* and *Area of Expertise*, $\chi^2(9) = 138.41, p < 0.001$. An immediate observation is that both categories are mainly comprised by Generic positions as their most frequent, indicating the increased need for specialists that develop and maintain software for various purposes. However, remote posts appear to be more closely related to jobs that involve the maintenance of systems or the design of a system architecture. In addition, Operations related positions are quite popular, indicating the comfortability of conducting DevOps testing and acting as a reliability engineer or a cloud engineer while working remotely. Finally, high values in the Product and Data positions are clear indicators of the need for remote frontend, full-stack or web developers and data scientists that do not rely on company infrastructure and provide quality work from their own homes under flexible working hours. As for the Management related positions, its increased percentage showcases that remote positions are mainly targeted for senior developers that can handle pressure, but they cannot be attributed to managers, directors or CIOs, proven by the experience levels of RQ$_1$, as these type of positions are better handled when accompanied by physical presence.

The on-site category is characterized by a higher percentage of Generic positions, as developers are still sought to work in physical locations, as well as a notable percentage of Management related job positions, which expresses the demand for leading individuals to be present in their companies or departments and coordinate projects or initiatives. Apart from that, no notable differences are observed, even though Data positions slightly surpass the Product counterpart, possibly due to the overabundance of jobs that seek Data Scientists in the Financial Services Industry. This further proves that in sectors where human interaction is crucial for the fulfillment of duties, on-site working is still considered a staple for employers.

However, remote working remains a viable choice for recruiters as the working environment of on-site jobs can be easily adapted for remote circumstances.

[RQ3] *Which are the most requested technologies in each category and what are the interconnections between them*?

The exploration of tag frequencies among job posts revealed the different demands and requirements of recruiters between the two job categories. In Table 3, the top ten technologies for on-site and remote posts are presented.

TABLE III. TOP TEN TECHNOLOGIES

On-Site		Remote	
Tag	**Frequency (f_i %)**	*Tag*	**Frequency (f_i %)**
java	32%	javascript	22%
python	16%	reactjs	19.4%
javascript	12%	aws	18.8%
reactjs	11.3%	python	17.1%
aws	9.4%	node.js	12.3%
cloud	9%	java	11.4%
spring	8.7%	typescript	8.7%
c++	8%	kubernetes	8.1%
sql	7.8%	c#	7.8%
c#	7.5%	sql	7.2%

On-site posts are heavily investing in backend and software engineering tools, as showcased by the presence of tags such as *java*, *sql* and *c++*. In addition, *python* is an indicator of data handling skillsets combined with *cloud* and *aws* that point to data engineering or database administration. Finally, the presence of *javascript* and *reactjs* along with *c#* and frameworks such as *spring* relates to the construction and maintenance of software.

The remote counterpart is different, as tags related to frontend development (*javascript, reactjs, node,js, typescript*) are in the forefront, emphasizing the comfortability of being employed as a frontend developer while working from home. In addition, the remaining tags point to job positions related to data manipulation and cloud computing, with cloud and developer platforms (*aws*, *kubernetes*) and programming languages for data analysis (*python*) being present. Finally, *sql* can be attributed to database handling and administration.

The inspection of the ARG of the remote posts (Figure 6) reveals that *python*, representing data manipulation and handling is connected with *sql* for database administration. In addition, *amazon-web-services (aws)* is also connected with *python*, representing the usage of cloud technologies for data linkage and expressing the growing technological prowess of data science and cloud architectures. A separate component contains deployment platforms such as *kubernetes* and *docker* that are typically related with system and software maintenance.

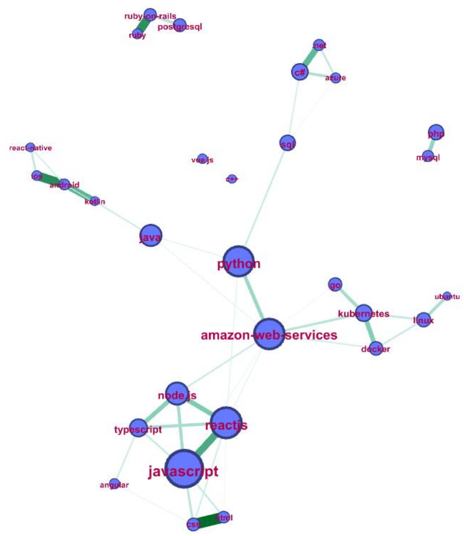

Fig. 6. Remote ARG network

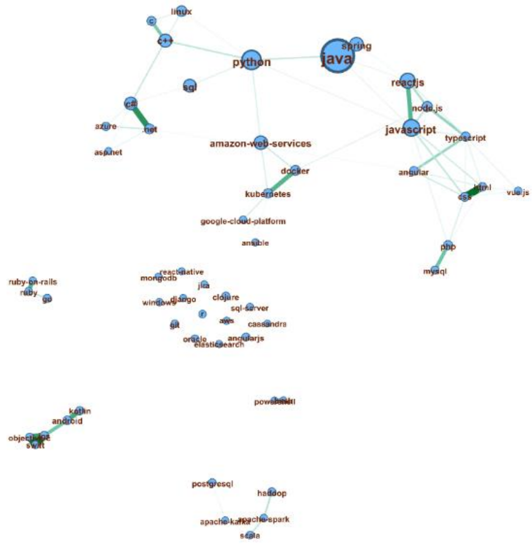

Fig. 7. On-Site ARG network

Regarding the *java* vertex, while it is linked with *python* and *aws*, indicating backend engineering procedures, it is also pointing to *android*, *ios* and *kotlin* which are technologies typically linked with mobile development. Finally, a sizeable component contains technologies adherent to software and frontend development such as *javascript*, *reactjs* ad *node.js*.

This further proves the increased need for frontend specialists that can not only design but also deploy software or act as system architects, as proven by the connection with *aws*. It should also be noted that the connection of *reactjs* with *python* can be attributed to the technologies related in software design and the pipelining of data and backend architectures.

The landscape of the ARG network of the on-site posts (Figure 7) is notably different. *java* is holding a leading position as the most frequent technology and is connected with *python* and several frameworks (*spring*) or deployment platforms (*docker*).

These connections, along with the connections with *aws* and *kubernetes* are relevant to the construction of backend architectures and the engineering of data. In addition, *python* is connected with OS oriented languages (*c*, *c++*) and *sql*, referring to the administration of systems such as *linux* or databases. An interesting finding is the existence of two isolated communities relevant to big data analysis (*apache-spark*, *hadoop*, *scala*, *apache-kafka*) and mobile development (*android*, *kotlin*, *ios*), which seem to be quite specific about the skillsets they express. In addition, the *r* programming language is also isolated, but this can be attributed by the difficulty of recruiters to efficiently pair it with other technological skills. Finally, software development is also present (*javascript*, *reactjs*) but in a lesser extent than the remote ARG.

VI. Discussion

In this section we briefly discuss the key findings for every RQ and their practical value.

RQ_1: *Characteristics of remote and on-site job posts*

While on-site posts are majority, remote posts hold a notable presence of almost 20%. Their experience levels differ, as remote working is clearly addressed to senior level and mature individuals while on-site posts offer some room for junior or mid-level specialists. As for the industries that seek employees, remote jobs are tightly linked with software development, software engineering and general ICT areas, while on-site jobs are prevalent in financial technologies and physical industries such as consumer electronics, while also retaining a relevance with software and ICT.

RQ_2: *Job positions based on titles*

By utilizing a predefined taxonomy and the *Title* field, the dominant positions for each *Category* are showcased. In general, remote working is mainly associated with DevOps, testing and system maintenance while an increased need for data handling specialists and software developers that do not rely on company locations is also notable. Management positions, while in a growing trajectory, are probably not associated with CIOs or directors but rather with senior developers.

On the other hand, on-site working is much more closely related with managerial staff that is required to be present in their respective companies to fulfill their duties. Moreover, there is a pressing need for data engineers and scientists in sectors that rely on human interaction (e.g. financial services) while traditional positions relevant with software development, software engineering and web development are still in high demand, despite the impressive rise of remote positions that concern the same roles.

RQ_3: *Most requested technologies and interconnections*

The examination of the tags of the posts clearly reveals the technological demands of each category. Remote posts are heavily relevant with software development, with specialized frameworks such as *reactjs* and *node.js* or javascript having prevalent positions. In parallel, an increased demand of data related skillsets is noted, evident by the presence of python and sql. On site posts seem to focus more on backend technologies such as java and system architecture with aws and python.

The ARG networks validate these findings, with remote posts mainly associated with software development (*javascript*, *reactjs*), the engineering and deployment of cloud architectures (*aws*, *cloud*) and data analytics (*python*) while the on-site ARG indicates a demand for backend related technologies (*java*, *c++*) and system administration skills. In parallel, both ARGs contain technologies related to mobile development while the on-site ARG also contains a component for big data analytics, which requires infrastructure not available in remote working.

VII. Threats to Validity

Regarding the internal validity, we based the categorization of posts in matching the bigrams of job titles with the titles of the taxonomy. However, this led to posts being assigned in more than two categories. While this can create some confusion, it was expected that this process would not be exclusive. Additionally, for our analysis we merged similar tags – technologies to reduce their number, as their initial number prevented an easy analysis. However, a further manual cross examination was added to the merging process by the authors, discussing emerging conflicts to ensure that the proposed synonyms were accurate and non-limiting.

Regarding external validity, one potential threat is the fact that the analysis and research was conducted only on job posts on SO. Even though SO contains a variety of jobs from multiple companies and organizations, there are ample opportunities for further research in order to compare our results with other well-known job portals (e.g. Monster, Indeed). Furthermore, we should note that the data collection process was conducted for a limited timeframe and thus, does not reflect the full picture of the job post activity on the site. Taking into consideration that new job advertisements are constantly getting posted, the extracted findings may vary for different timespans.

VII. Conclusions and Future Work

Remote working in the IT sector has known an ever-growing trajectory in recent years, as it offers an opportunity for employers to cut down on expenses while simultaneously hiring skilled individuals for their workforce. This increase is reflected in recruiting platforms, where many employers prefer remote workers from on-site personnel. Particularly in recent times, with the global crisis of COVID-19, remote working is slowly gaining more ground in the industry as a necessary measure. In this climate of increasing popularity, understanding the characteristics of remote jobs that set them apart from their on-site counterparts would be of great value for employers and employees that want to either adjust their recruiting strategies or hone their skillsets accordingly.

In this study, we emphasize on these observations and attempt to detect the technological aspects, areas of expertise and attributes that shape the profile of remote working. Our findings indicate that remote jobs are indeed attracting the interest of the industry, particularly in matters related to Software production and engineering or the manipulation and handling of data. The preferred areas of expertise often concern senior positions with leading responsibilities or generic practices such as software development. Moreover, the technical skills required for such positions support this statement, with *python*, *javascript* and *reactjs* being some indicative popular technologies.

Remote workers are expected to increase in a larger scale in these troubling times, with recruiting platforms posting a tremendous amount of job posts in this area. Thus, as future work we plan to expand this study by collecting data regarding job posts from SO and other platforms (e.g. Remotive) in different timespans, examining their temporal evolution and characteristics while comparing potential changes in technological demand.

REFERENCES

[1] Cascio, W. F., & Montealegre, R. (2016). How technology is changing work and organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 349-375.

[2] Bick, A., Blandin, A., & Mertens, K. (2020). Work from home after the COVID-19 Outbreak.

[3] Béland, L. P., Brodeur, A., & Wright, T. (2020). The short-term economic consequences of Covid-19: exposure to disease, remote work and government response.

[4] Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). Achieving effective remote working during the COVID-19 pandemic: A work design perspective. *Applied Psychology*, *70*(1), 16-59.

[5] Brynjolfsson, E., Horton, J., Ozimek, A., Rock, D., Sharma, G., & Tuye, H. (2020). COVID-19 and Remote Work: An Early Look at US Data. doi:10.3386/w27344

[6] Sackett, P. R., Walmsley, P. T., & Laczo, R. M. (2012). Job and work analysis. *Handbook of Psychology, Second Edition*, *12*.

[7] Ahsan, K., Ho, M., & Khan, S. (2013). Recruiting project managers: A comparative ana-lysis of competencies and recruitment signals from job advertisements.ProjectManagementJournal,44(5), 36–54.https://doi.org/10.1002/pmj.21366.

[8] Javed, F., McNair, M., Jacob, F., & Zhao, M. (2016). Towards a job title classification system. *arXiv preprint arXiv:1606.00917*.

[9] Todd, P. A., McKeen, J. D., & Gallupe, R. B. (1995). The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990.MISquarterly,1–27

[10] Litecky, C., Aken, A., Ahmad, A., & Nelson, H. (2010). Mining for Computing Jobs. IEEE Software,27(1), 78-85. doi:10.1109/ms.2009.150

[11] Addom, B. K., Kim, Y., & Stanton, J. M. (2011). EScience professional positions in the job market. Proceedings of the 2011 IConference on - IConference '11. doi:10.1145/1940761.1940846

[13] Debortoli, S., Müller, O., & Brocke, J. V. (2014). Comparing Business Intelligence and Big Data Skills. Business & Information Systems Engineering, 6(5), 289-300. doi:10.1007/s12599-014-0344-2

[14] Pejic-Bach, M., Bertoncel, T., Meško, M., & Krstić, Ž. (2020). Text mining of industry 4.0 job advertisements. International Journal of Information Management,50, 416-431. doi:10.1016/j.ijinfomgt.2019.07.014

[15] Papoutsoglou, M., Ampatzoglou, A., Mittas, N., & Angelis, L. (2019). Extracting knowledge from on-line sources for software engineering labor market: A mapping study. *IEEE Access*, *7*, 157595-157613.

[16] Hauff, C., & Gousios, G. (2015). Matching GitHub Developer Profiles to Job Advertisements. 2015 IEEE/ACM 12th Working Conference on Mining Software Repositories. doi:10.1109/msr.2015.41

[17] Daneva, M., Wang, C., & Hoener, P. (2017). What the Job Market Wants from Requirements Engineers? An Empirical Analysis of Online Job Ads from the Netherlands. 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). doi:10.1109/esem.2017.60

[18] Palshikar, G. K., Srivastava, R., Pawar, S., Hingmire, S., Jain, A., Chourasia, S., & Shah, M. (2018). Analytics-Led Talent Acquisition for Improving Efficiency and Effectiveness. Advances in Analytics and Applications Springer Proceedings in Business and Economics, 141-160. doi:10.1007/978-981-13-1208-3_13

[19] Gonzalez, T., Santos, P., Orozco, F., Alcaraz, M., Zaldivar, V., de Obeso, A., et al. (2012).Adaptive employee profile classification for resource planning tool. 2012 Service Research and Innovation Institute Global Conference, 544–553.

[20] Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world. Industrial and Organizational Psychology, 9(3), 621–640. https://doi.org/10.1017/iop.2016.6

[21] Karakatsanis, I., AlKhader, W., MacCrory, F., Alibasic, A., Omar, M. A., Aung, Z., et al.(2016). Data mining approach to monitoring the requirements of the job market: A case study. Information Systems, 65(4), 1–6. https://doi.org/10.1016/j.is.2016.10.009

[22] Deshpande, A., Sharp, H., Barroca, L., & Gregory, P. (2016). Remote working and collaboration in agile teams.

[23] Crowley, F., & Doran, J. (2020). COVID-19, occupational social distancing and remote working potential: An occupation, sector and regional perspective. *Regional Science Policy & Practice*, *12*(6), 1211-1234.

[24] Rodeghero, P., Zimmermann, T., Houck, B., & Ford, D. (2021, May). Please turn your cameras on: Remote onboarding of software developers during a pandemic. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 41-50). IEEE.

[25] Ford, D., Milewicz, R., & Serebrenik, A. (2019, May). How remote work can foster a more inclusive environment for transgender developers. In *2019 IEEE/ACM 2nd International Workshop on Gender Equality in Software Engineering (GE)* (pp. 9-12). IEEE.

[26] Schulze, J., & Krumm, S. (2017). The "virtual team player" A review and initial model of knowledge, skills, abilities, and other characteristics for virtual collaboration. *Organizational Psychology Review*, *7*(1), 66-95.

[27] Westwood, S., Johnson, M., & Bunge, B. Predicting Programming Community Popularity on Stack Overflow from Initial Affiliation Networks

[28] Georgiou, K., Papoutsoglou, M., Vakali, A., & Angelis, L. (2019, September). Software technologies skills: A graph-based study to capture their associations and dynamics. In Proc. of the 9th Balkan Conference on Informatics (pp. 1-7)

[29] Georgiou, K., Mittas, N., Angelis, L., & Chatzigeorgiou, A. (2020, August). A preliminary Study of Knowledge Sharing related to Covid-19 Pandemic in Stack Overflow. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* (pp. 517-520). IEEE.

[30] M. Papoutsoglou, N. Mittas and L. Angelis, "Mining people analytics from stackoverflow job advertisements" , Software Engineering and Advanced Applications (SEAA) 2017 43rd Euromicro Conference on, pp. 108-115, 2017.

[31] Montandon, J. E., Politowski, C., Silva, L. L., Valente, M. T., Petrillo, F., & Guéhéneuc, Y. (2021). What skills do IT companies look for in new developers? A study with Stack Overflow jobs. Information and Software Technology,129, 106429. doi:10.1016/j.infsof.2020.106429

[32] Barzilay, O., Treude, C., & Zagalsky, A. (2013). Facilitating crowd sourced software engineering via stack overflow. In *Finding Source Code on the Web for Remix and Reuse* (pp. 289-308). Springer, New York, NY.

[33] Barua, A., Thomas, S. W., & Hassan, A. E. (2014). What are developers talking about? an analysis of topics and trends in stack overflow. *Empirical Software Engineering*, *19*(3), 619-654.

[34] Chen, C., & Xing, Z. (2016, September). Mining technology landscape from stack overflow. In *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (pp. 1-10).

[35] Osman, O. (2020). *The Holloway Guide to Technical Recruiting and Hiring: Align your team to avoid expensive hiring mistakes*. Holloway.

[36] Cui, B., Yao, J., Cong, G., & Huang, Y. (2010, December). Evolutionary taxonomy construction from dynamic tag space. In International Conference on Web Information Systems Engineering (pp. 105-119). Springer, Berlin, Heidelberg.