A Characterization Methodology for Candidates and Recruiters Interaction in Online Recruitment Services*

Alan Cardoso[1], Fernando Mourão[2], Leonardo Rocha[1]
[1]Universidade Federal de São João del-Rei and [2]Seek AI Labs
(alanc,lcrocha)@ufsj.edu.br
fernando.mourao@catho.com

ABSTRACT

Online recruitment services have attracted an increasing number of candidates and recruiters who are looking for better job opportunities and the best professionals in their respective areas. These services, through search and recommendation systems, explore candidates and job profiles to identify the ideal candidates for each job vacancy. There are many challenges when we follow this scenario, such as reciprocal matches between vacancies and candidates, temporal dynamics (candidate/vacancy relationship varies over time) and imbalances between demand and supply between areas. Modeling the preferences and behavior of candidates and recruiters is an essential task for which improvements can be proposed by these services to mitigate their challenges. We present in this work a methodology that aims to help answer questions that may be asked about users preferences and behavior, extracting information that leads to improvements in existing functionality and the creation of new ones. We applied our methodology to actual data and questions, which were provided by Catho, the leading Latin American market company within this segment. In the analysis of results, we present opportunities for improvement in online recruitment services, such as the creation of a tool to help register job vacancies and resumes.

CCS CONCEPTS

• **Information Systems → Recommender Systems**.

KEYWORDS

Preference Elicitation, Job Recommender Systems; Characterization

ACM Reference Format:
Alan Cardoso[1], Fernando Mourão[2], Leonardo Rocha[1]. 2019. A Characterization Methodology for Candidates and Recruiters Interaction in Online Recruitment Services. In *Brazilian Symposium on Multimedia and the Web (WebMedia '19), October 29-November 1, 2019, Rio de Janeiro , Brazil.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3323503.3349541

1 INTRODUCTION

Recently, so-called online recruitment services have attracted to the Web an increasing number of users and companies with urgent demands [1, 18]. Such services are specialized in managing two main actors: (1) candidates, people with specific characteristics and competencies looking for new employment opportunities; and (2) recruiters, companies that offer job vacancies and look for the best professionals in the market. In general, the goal is to identify the ideal candidates for each job, using search and recommendation algorithms that explore candidates' profiles (i.e., curriculum) and the skills required by different job vacancies [3].

Given this context, an essential task of these services is to model the profile of candidates and job vacancies offered under the same guidelines. It is very common to observe in these platforms resumes and job vacancies described in a very specific (or even very generic) way, making it difficult to match the candidate and the job vacancy. Moreover, in these scenarios, **(1) the needs are reciprocal** because mutual satisfaction is required for both, the candidate and the recruiter of the job vacancy [13]. The market is very dynamic and, consequently, **(2) supply and demand for job vacancies can vary significantly over time**. In addition, this dynamism can lead to **(3) imbalances between demand and supply** between areas, so that there are areas with more candidates than vacancies or vice versa.

Considering the challenges previously mentioned, **we propose in this work a methodology of characterization that aims to elucidate issues related to the preferences and behaviors of the main actors involved in online recruitment**. We expect to be able to provide the companies and researchers with the necessary incentives to mitigate the aforementioned challenges, defining a starting point for professionals in the field to structure analysis and characterization of data in this domain, enhancing personalization of job searches and recommendations. Our methodology has eighteen metrics, divided into five dimensions. The first dimension refers to **(i) Area Dimension (AD):** which answers questions related to the area of action of the candidates and the area in which the job vacancy is offered; **(ii) Benefits Dimension (BD):** that considers a series of characteristics that can be described in a job vacancy offered by a recruiter, or desired by a candidate in a vacancy to be occupied, such as salary or working hours; **(iii) Competence Dimension (CD):** which considers issues related to the professional skills described in the curriculum or required for the job vacancy; **(iv) Geographic Dimension (GD):** which evaluates characteristics regarding the geographical location of users and job vacancies; and **(v) Temporal Dimension (TD):** which analyzes how demands and offers behave over time.

We apply our methodology to actual data, from a sample provided by CATHO[1], the market leading company in Latin America. Although there are a lot of questions, we look for useful ones that have been identified in Catho's actual use cases and which clarify

*This work is partially supported by SEEK, Catho Online, CNPq and FAPEMIG.

WebMedia '19, October 29-November 1, 2019, Rio de Janeiro , Brazil

ACM ISBN 978-1-4503-6763-9/19/10…$15.00
https://doi.org/10.1145/3323503.3349541

[1]www.catho.com.br

the behavioral preferences of system users. We separate the results into three distinct views, the first being **(i) Catho's Point of View:** where we address practical issues that help the company understand the dynamics of its data collection; **(ii) Candidate's Point of View:** which clarifies how candidates behave and aims to maximize the chances of being recruited; and **(iii) Recruiter's Point of View:** which answers questions regarding the behavior of recruiters, it is intended to help them secure suitable candidates for their job vacancies. Among the results obtained by the application of our methodology in this sample, we observed, for example, that 60% of the vacancies are related to only 20% of the different professional areas, which tends to pose a challenge for candidates seeking a return to the market in areas of low demand. This result shows that the Recommendation Systems (RSs) should operate differently by area, according to the demand-to-supply ratio. As regards the evaluation of the characteristics of resumes and job vacancies, we observed, for example, that the higher the salary, the higher the candidate/vacancy ratio; and that 60% of the vacancies offered require very specific skills, significantly restricting potential candidates (one candidate/vacancy). From this result, we must consider, for example, the creation of a tool to help register resumes and job vacancies. This tool may recommend the skills and requirements to be registered to maximize the chances of a future match between candidate and job vacancy.

In summary, the main contributions of this article are: *(i) the proposal of a new characterization methodology capable of monitoring variations in the demand and supply of employment, correlating different aspects and helping to build more effective online recruitment services; (ii) the evaluation of the methodology proposed in real data, from Catho, eliciting user's preferences (information); and (iii) identification of a number of opportunities for improvement that can be applied to these services (knowledge).* One can observe that the proposed methodology is a guide to be used to investigate the data through classic Business Intelligence (BI) or Datawarehouse tools, for example, and that we did not identify any systematic study similar to ours in online recruitment domains.

2 RELATED WORKS

Recently, online recruitment services have established as one of the leading recruitment forms around the world [3]. In addition to the practicality of such services, studies also highlight the speed at which the recruitment process occurs [15] and the resource economy provided to candidates and recruiters [9]. However, the effectiveness of these systems is still limited due to numerous challenges [16]. Among them, we highlight the need for a deep understanding of the interaction of candidates and recruiters with this type of system. Another challenge is the quality and usefulness of the information entered by candidates and recruiters in the system [7]. For example, it is common to see extremely detailed and specific (or even generic, non-specific) resumes and vacancies that significantly hamper tasks such as search and recommendation. Thus, as in other domains, the characterization and analysis of data related to online recruitment services are one of the primary requirements for the success of this type of service. [8].

In general, this characterization and analysis are made considering three different perspectives: (1) of the candidate, who is describing his/her resume; (2) of the recruiter, who is describing his/her position; and (3) in relation to the task of a match between resumes and vacancies, that is, finding candidates that match the needs of a job opportunity. *Spina et al* [18] analyze the characteristics of the queries performed by the candidates in job search systems to interpret how they look for job vacancies. Among the advantages of understanding how each user describes their main interests, the authors also conclude what metrics and search strategies are (or are not) effective in the recruitment scenario. In turn, *Salehi et al* [17] analyze how poorly formulated job searches can bring low-quality results and impact user's satisfaction.

Considering work related to the second line of analysis, we identify efforts that are concerned with categorizing some areas of action through latent topic models, to facilitate the matching task. *Turrell et al* [20], for example, use topic modeling to understand which sectors and segments each of the job vacancies are related to. In the same direction, in [22], the authors present a characterization focused on extracting the most important skills from seven different positions (administrator, analyst, developer, engineer, leader, support and tester), highlighting the most desirable skills in each position.

Finally, we review papers that are concerned with identifying possible insights to be explored in the course of the assignment of resumes and vacancies. In this case, we highlight work similar to the one presented in [11], in which the authors focus on the recommendation task for newly graduated candidates and still have no real work experience, a problem similar to Cold-Start in traditional RSs scenarios [10]. Taking into account the traditional scenarios, several authors propose the combination of different recommendation strategies to achieve a more robust and adaptable RSs [2, 5, 12, 23]. In turn, in [4], the authors use advanced strategies, with concepts of deep learning, to help in the task of finding potential talent for the vacancies. In [6], the authors try to understand the career development of candidates, to anticipate this perspective, providing RSs that are able to detect the next vacancies of interest of each candidate. Similarly, in *M. L. Tran et al* [19], authors make a detailed assessment of the RSs of existing jobs, pointing out their advantages and disadvantages. In [21], the authors also perform tests with eight different classifiers and compare the results of their application in the classification of job vacancies.

In this work, we present an iterative methodology of characterization, capable of providing analyzes under different aspects, with generic and specific questions, which can be defined by the user of the methodology (e.g., Recruitment Company). Our goal is that the proposed methodology serves as a starting point for professionals in the area to structure and systematize analyzes and characterizations about data in this domain. In this way, we are also able to provide information that improves the service as a whole, attacking these three lines simultaneously.

3 PROPOSED METHODOLOGY

Specialized online recruitment services use Search or Recommendation Systems to present the best available job opportunities for candidates and the best candidates for the job vacancies. This process needs to be fast and effective, as delays represent economic losses for both parties. Such systems need to be improved and always adjusted to various scenarios automatically. For example, we have scenarios of a shortage of vacancies, the absence of a candidate profile, imbalance between opportunities and seasonality of offers or demand for job vacancies. Thus, to achieve quality services, a continuous process of data analysis is required, maintaining an up-to-date view of the behavior of all actors and information. As an example, assume that a recruitment company, aiming to understand how the demand and supply of jobs is in 2018, interprets that the economic crisis in Brazil has not affected the IT sector, which continues to present demand above the supply of professionals. In this perspective, the RSs used must be adjusted to meet this market need by increasing the coverage of professionals in related areas. Understanding and adjusting recruiting services to keep up with these market fluctuations is not an easy task.

The proposed characterization methodology aims to assist companies and professionals specialized in recruitment to characterize and analyze the various factors that are in this domain. We expect that this methodology will serve as a starting point for professionals in the market to structure and systematize relevant data issues, assisting in making decisions about techniques and strategies to be applied or improved over time. In this sense, the methodology allows us to (1) organize the types of features existing in the domain into categories; (2) define and explore basic metrics that can be associated with each feature; and (3) systematize the process of answering varied domain questions, dividing each one into several simple questions. One can see that this is a guide that will be used to investigate the data through classic BI or Datawarehouse tools. It should also be noted that our proposal is particularly relevant given the scarcity and difficulty of obtaining real data for the characterization/investigation of recruitment services in the published works. We did not identify any systematic study of such data.

In the proposed methodology, the case begins with a question of interest raised by the online recruiting professional. This question must be embedded in one of the five dimensions supported by our methodology. Once the analysis dimensions are defined, the next step is to apply the metrics (described in detail in the following sections). These metrics generate behavioral profiles and can be used to (i) answer the listed question and solve the problem or (ii) refine the question raised through a new iteration in the methodology, resulting in a more detailed profile. This new iteration can be done by incorporating a new dimension and can use the knowledge previously generated. It is important to emphasize that our methodology is a guide and the online recruitment company should analyze the knowledge generated by the metrics, deciding if it is sufficient to answer the question or whether further improvement is required.

Considering the previous example, in which the company realizes that the economic crisis has not affected the IT sector w.r.t. the number of vacancies (Area Dimension), one may now want to iterate again in the methodology to understand which positions are generating more opportunities and thus generate recommendations even more appropriate to the market. Table 1 summarizes the metrics defined by our methodology. Finally, we emphasize that it is not a closed set of dimensions and metrics, and it is possible to expand them if necessary.

3.1 Area Dimension (AD)

The objective of this dimension is to evaluate the data from the perspective of the areas of activity. The goal is to understand which sectors of the market have more job vacancies, which ones have more candidates looking for a job vacancy, or which ones have excess or scarcity of opportunities. This information is important so that recruitment companies can establish, for example, different recommendation strategies according to the characteristics of the areas.

(AD-1) Distribution of job vacancies by area: provides a view on the professional areas with more and fewer job vacancies.

(AD-2) Distribution of demands by area: it presents the professional areas with more and fewer professionals in search of a job vacancy in the market.

(AD-3) Application distribution by area: this metric aims to highlight which are the professional areas most requested by candidates.

(AD-4) Candidate/vacancy distribution by area: seeks to understand if there are professional areas with many vacancies and few demands or vice versa.

3.2 Benefits Dimension (BD)

This dimension aims to evaluate the characteristics that make up a job offer. From the candidate's point of view, it refers to the items which he/she expects to find in an advertised job vacancy. Features such as *work schedules*, *contract type*, *salary*, *benefits*, *certificates*, *hierarchical level* (e.g., junior/senior programmer) are part of this dimension. From the information provided by this dimension, along with CD (Competence Dimension) analyzes, recruitment companies can assist both candidates and recruiters in registering jobs vacancy and curricula that are most likely to be successful.

(BD-1) Distribution of job vacancies by item: shows how the job vacancies are for a specific topic.

(BD-2) Requirement distribution by item: shows how the demands for a specific topic are.

(BD-3) Application distribution by item: shows how candidates are interested according to a topic.

(BD-4) Candidate/vacancy distribution per item: shows how the distribution between candidates' resumes and vacancies is according to a topic.

3.3 Competencies Dimension (CD)

This dimension considers issues related to the practical skills described in the candidate's curriculum or required for the specific job vacancy. As some examples, we have *customer service, machining*, *computer programming*, or even a very specific skill such as *computer programming in the Python language*.

(CD-1) Distribution of vacancies by competence: aims to understand the job vacancies according to the required competencies.

(CD-2) Distribution of competency requests: it shows the more and less common skills among professionals.

(CD-3) Distribution of applications by competency: it presents in which competences there are more applications to job vacancies to identify which ones are of greater interest of the candidates.

(CD-4) Candidate/vacancy distribution by competency: shows which are the most common and least common competencies in the market. Identifying rare skills can help recruiters better describe their vacancies.

Table 1: Metrics described by dimension (AD, BD, CD, DG and TD)

Dimension	Features Exemple	Metrics	Description
Area (AD)	general area	AD-1	For each professional area, we add the number of job vacancies available.
	subarea	AD-2	For each professional area, we add the number of applicants.
	expertise	AD-3	We add the number of applications separated by professional area.
		AD-4	The number of candidates is divided by the number of job vacancies in each professional area.
Benefits (BD)	salary	BD-1	For each item, we add the number of job vacancies available.
	working hours	BD-2	For each item, we add the number of applicants.
	food benefit	BD-3	We add the number of applications separated per item.
		BD-4	We divide the number of candidates by the number of job vacancies in each item.
Competence (CD)	skills	CD-1	For each competency, we add the number of vacancies available.
	spoken languages	CD-2	For each professional area, we add the number of applicants.
	previous jobs	CD-3	We add the number of applications separated by competence.
		CD-4	The number of candidates is divided by the number of job vacancies in each competence.
Geographic (GD)	city	GD-1	For each geographical region, we add the number of available job vacancies.
	state	GD-2	For each geographic region, we add the number of applicants.
	country	GD-3	We add the number of applications separated by geographic region.
	region	GD-4	The number of candidates is divided by the number of job vacancies in each geographical region.
Temporal (TD)	day	TD-1	Sum of the number of job vacancies in each temporal unit.
	month	TD-2	Sum of the number of applications in each temporal unit.

3.4 Geographical Dimension (GD)

For this dimension, we consider the geographic region of the candidates and the job vacancies. One can consider different geographic granularities, from a regional, state and municipal scale, or even any other granularity that the dataset supports. From this information, an online recruitment company can, for example, identify regions with opportunities for expansion of activities.

(GD-1) Distribution of job vacancies by geographic region: this metric presents a visualization of the distribution of job vacancies according to geographic region.

(GD-2) Distribution of requests by geographic region: this metric presents a view of the distribution of requests according to geographic region.

(GD-3) Application distribution by geographic region: This metric shows a view of the distribution of applications according to geographic region.

(GD-4) Candidate/vacancy distribution by geographic region: this metric aims to identify regions with the largest and the smallest relationships between candidate and vacancy.

3.5 Temporal Dimension (TD)

This dimension assesses the temporal dynamics of both job vacancies and job demands.

(TD-1) Vacancies over time: describes the variation of job vacancies over time.

(TD-2) Demands Over Time: describes the variation of job demands over time.

4 EXPERIMENTAL EVALUATION

In this section, we evaluate the proposed methodology using a data sample provided by Catho. We divide the analysis into three subsections: (4.2) *Catho point of view*; (4.3) the *Candidate's point of view*; and (4.4) *Recruiter's point of view*. First, for the analysis from Catho's point of view, we raise questions for three dimensions of analysis and guide our study by selecting some of the relevant attributes in each of them. Our goal is to have a broad understanding of the data. In a second moment, for the candidate's point of view, we apply an iterative process to the methodology, answering questions of real interest that help the candidates. Finally, we apply the iterative process to answer questions of interest of the recruiters where, again, we show that combining the proposed dimensions we can extract relevant information and, consequently, prove the applicability of the methodology.

4.1 Catho Database

Catho, located in Brazil, is an online recruitment company, which is part of the Seek[2] group, the largest multinational company in the online recruitment industry in the world. The data sample is divided into two main entities: resumes created by the candidates; and job vacancies created by recruiters. The sample has 376,762 resumes and 115,955 job vacancies, obtained for 13 months (January to January). For reasons of confidentiality, we cannot disclose the data used in these analyses.

4.2 Catho's point of view

In this analysis, our goal is to provide a more global view of the data sample evaluated, characterizing the demands by vacancy and candidate. For sake of space, we restricted this analysis to the area, benefits and temporal dimensions.

Question of Area Dimension (AD): *How are the demands, vacancies and candidate applications distributed according to the different areas in the sample?*

Analysis: We evaluated and compared the distributions of vacancies (AD-1), demands (AD-2) and vacancies application (AD-3) by professional areas, shown in Figure 1(a). In Figure 1(b) we present the candidate/vacancy distribution (AD-4).

Information: One can see that there is an imbalance of demands and job vacancies in relation to the applications in the areas. In our sample, 20% of the areas are responsible for 60% of the vacancies, a

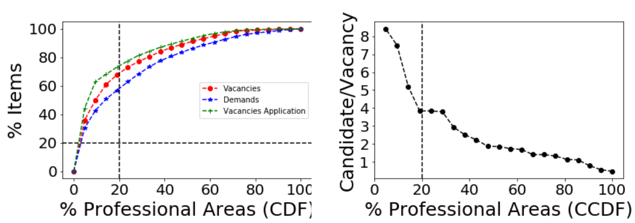
[2]www.seek.com.au

little less than 60% of the demands and 70% of the applications. We also have 80% of the areas of activity less active, both by candidates and by recruiters. This imbalance also causes differences in the candidate/vacancy relationship, as can be seen in Figure 1(b) and Table 2. In them, we can notice that 20% of the professional areas have a greater number of candidates/vacancy varying between eight (8) and four (4) candidates. The other professional areas present a more balanced number of candidates/vacancy, varying between four (4) and a little less than one (1) candidate.

Knowledge: Scenarios similar to those observed in our sample present potential for growth, through strategies that attract new users (i.e. candidates or recruiters), to cover less popular areas. Marketing strategies, for example, should act actively to narrow the gap in this distribution, preventing some areas from having a few vacancies and a few other candidates. There is also a need to treat different areas in different ways. This analysis shows that there are distinct parts within the sample used and one can even work with several RSs specific to the problems faced by each of these parts.

Table 2: Profissional Areas

Area	Candidate/ Vacancy	Area	Candidate/ Vacancy
IT	0,4732438	Marketing	1,8715728
Business	0,5635589	Industrial	2,2311015
Telemarketing	0,7829192	Financial	2,5044687
Administration	1,1037627	Telecommunication	2,9278688
Hotel Business	1,1369294	Engineering	3,7973395
Health	1,3223934	Agriculture	3,8365758
Arts	1,4028185	Veterinary	3,8365758
Architecture	1,4028185	Social Services	5,1929824
Supplies	1,6877085	International Business	7,5000000
Technical	1,7356205	Law	8,4136125
Educational	1.8529023		



(a) Metrics AD-1, AD-2 and AD-3 (b) Metrics AD-4

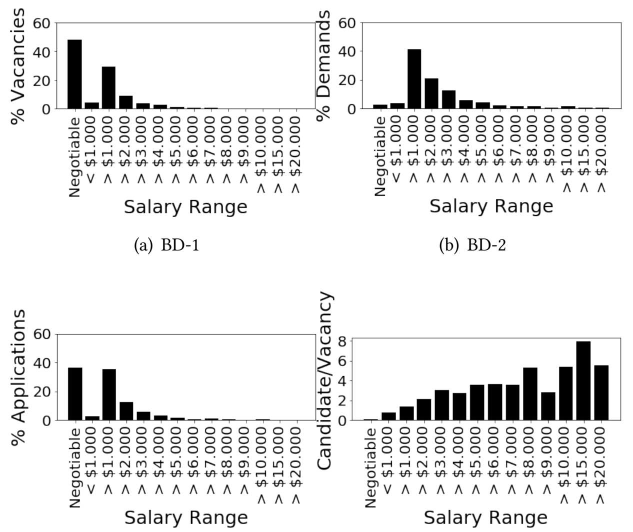
Figure 1: Area Dimension (AD)

Question Benefits Dimension (BD): *How is the distribution of salaries for the vacancies, demands, applications, and candidates/vacancy?*

Analysis: Among the various descriptive characteristics of a position and curriculum, we restrict our analysis to the item SALARY. We started our analyzes by evaluating and comparing the metrics BD-1, BD-2 and BD-2 (Figures 2(a), 2(b) and 2(c)).

Information: We can see that 50% of the job vacancies do not inform the salary and treat this as something to be defined after the first contacts with candidates. Despite this, only 35% of the applications are made in job vacancies with the salary question to be later defined. This is relevant information to be considered by recruiters since in such cases they do not call the candidates attention to the position. As wages rise higher, vacancies and demands decrease, showing that as values increase, there are fewer opportunities, but fewer candidates who are projected to qualify for such wages. However, evaluating the metric BD-4 (Distribution of candidate/vacancy ratio by salary range - Figure 2(d)), we can see that this decrease in vacancies and demands is not proportional: for higher salary ranges the candidate/vacancy ratio increases significantly.

Knowledge: These observations point to the difficulty in relocating candidates looking for high salaries. This is due to the increased competition for these job vacancies. This is an opportunity to adapt their RSs to work with the desired salary by the candidates. From the curriculum of a candidate and his/her competencies, it would be interesting to rank this candidate within the specific area and to focus on salary range compatible with their competencies (the higher, the better). This way we would improve the quality of recruitment.



(a) BD-1 (b) BD-2

(c) BD-3 (d) BD-4

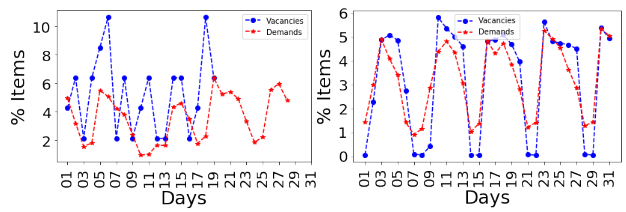
Figure 2: Benefits Dimension (BD).

Temporal Dimension (TD): *How is the distribution of vacancies and demands over a month?*

Analysis: To perform the analyses of the Temporal Dimension, we consider two distinct months, February and July of 2018, using daily granularity.

Information: In Figures 3(a) and 3(b), we present the distribution of job vacancies referring to the TD-1 and TD-2 metrics for the months of February and July, respectively. As one can perceive, in TD-1 for the month of July there are some inactivity periods, where it is clear the behavior of recruiting employees, who tend not to work during the weekends. In February, however, where the Brazilian party known as carnival took place, it is possible to notice that between the days 07 and 15 the number of job registrations was lower than on other days. we can also see the results about the demands, referring to the TD-2 metric. Unlike recruiters, candidates tend to remain lightly active on weekends and on holidays, such as the carnival. In both cases, both the job vacancies offers and the demands, present a seasonal decrease due to weekends and holidays.

Knowledge: It is clear to see that the recruitment company has challenges related to the seasonality of offers and demands. In the overall evaluation, it was clear that in the evaluated festive period, both demands and job vacancies offer, tend to be lower. However, in other periods, such as December, because of Christmas, there may be a hiring increase, especially in the sales and service sector. More iterations, analyzing other periods and other areas, one can understand how the behavior of users change over time. Evaluating which sectors suffer from this seasonality would allow the RS to deal with the increase or decrease of vacancies and offers, preventing the RS from maintaining the same strategy used during the rest of the year.



(a) 2018 February (b) 2018 July

Figure 3: Temporal Dimension (TD).

4.3 The candidate's point of view

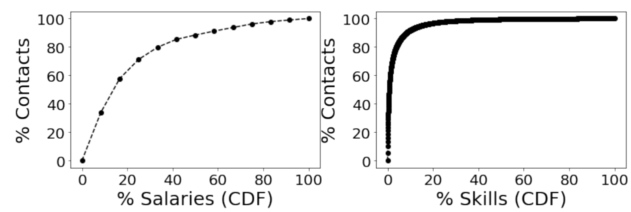
In this subsection, we performed some iterations about the methodology showing the importance of information that can be obtained through its application. All the questions discussed and presented are intended to identify strategies on how candidates should submit their resumes to increase their chances of contacting in an online job search system.

Question: *How should candidates in the IT industry meet the salary and skills requirement to have more contacts from recruiters?*

Analysis: For this analysis, we firstly iterate through the methodology using Area Dimension 3.1 and considering the IT sector. Next, we performed two different analyzes: one focused on salary and another focused on skills. Therefore, we rehearse once more on Benefits Dimension 3.2 considering the Salary item and analyze the number of contacts that the IT candidate receives, according to each desired salary range. We also did a second analysis, iterating a second time by Competencies Dimension 3.3 and evaluating the contacts according to the skills described in the curriculum.

Information: Figure 4(a) shows that 20% of the salary ranges account for approximately 60% of the recruiters' contacts. In Figure 4(b) presents a pronounced curve with 5% of the possible skills responding for 80% of all contacts. This shows that although one can observe several salary ranges and several distinct skills, the contacts are held around a few very specific sets. These data show that to increase the recruiters' chances for contacts the curricula should contain at least some of the most important skills or salary range.

Knowledge: The main application would be to use this information in the creation of a helping system that makes recommendations during the registration of the curricula (pre-registration recommendation). This system can indicate to a candidate what the chances of being contacted by a recruiter according to the salary and skills he/she has, for example. The system could also ask what relevant skills the candidate has. Let us assume the following scenario, the candidate adds to his/her curriculum the *J2EE specialist* skill. This is a very specific skill and, it reaches a smaller set of potential job positions. This pre-registration recommendation system could suggest the exchange of this skill by *Java programmer*. At the same time, it would advise the candidate, what percentage of vacancies would be reached with the suggested change, presenting a type of feedback to the candidate in real time. This tool could also suggest changes in job descriptions. Market sectors that have many technical skills, such as IT, can benefit from this tool.



(a) Contacts by Salary Range (b) Contacts by Skill

Figure 4: Analysis of contacts by salary range and by skills

Question: *How do the required skills and salaries offered vary from region to region for the IT sector?*

Analysis: In this analysis, we initially iterate through the Area Dimension 3.1 selecting the data according to the IT area. From that point, we did two different analyzes. For the first iterated by Geographic Dimension 3.4, we considered the cities of São Paulo, Rio de Janeiro and Belo Horizonte (the three largest cities in Brazil). Next, we iterate through Competencies Dimension 3.3 and extract the top 10 most wanted skills from recruiters. The second analysis was done in a similar way, iterating by the Benefits Dimension 3.2 considering the Salary item. As a final result, we present the list with the top 10 skills most wanted by region and the top 10 most wanted skills according to salary range. Finally, we apply the Jaccard distance [14] between these items to find out how similar skills are sought between regions and between the salary range.

Information: In the distance of Jaccard, the more similar two sets of items are sorted, the closer to 1 is the result. Evaluating the results shown in Tables 3 and 4, we observe a great similarity between the requirements of different regions. Considering the results for an adjacent salary range, the requirements in the lower salary range are very similar. The similarity increases as we move towards the intermediate salary range, falling back into the higher salary range.

Knowledge: These results may help in the tool mentioned previously for curriculum suggestions. For example, according to the skills registered by the user, it is possible to suggest to him/her that, select different cities for job search. In addition, observing that there are specific skill sets for a different salary range, it would be interesting to use the skills inserted by users to categorize them into specific salary range, and the Recommendation System (RS) could recommend vacancies from this categorization. These data, therefore, serve as both a subsidy for the creation of a pre-registration system as for the improvement of RS. Another interesting tool that could be added to these online recruitment services would be applications that can help candidates evolve in their careers, presenting learning and skills enhancements related to the salary range.

Table 3: Distance Between Skills for Places

Capital 1	Capital 2	Jaccard Distance
São Paulo	Rio de Janeiro	0,7
São Paulo	Belo Horizonte	1,0
Rio de Janeiro	Belo Horizonte	0,7

Table 4: Distance Between Skills for Salaries

Salary Range 1	Salary Range 2	Jaccard Distance
< 1.000,00	> 1.000,00	0,2
> 1.000,00	> 2.000,00	0,2
> 2.000,00	> 3.000,00	0,5
> 3.000,00	> 4.000,00	0,7
> 4.000,00	> 5.000,00	0,1
> 5.000,00	> 6.000,00	0,1
> 6.000,00	> 7.000,00	0,7
> 7.000,00	> 8.000,00	0,6
> 8.000,00	> 9.000,00	0,4
> 9.000,00	> 10.000,00	0,2
> 10.000,00	> 15.000,00	0,4
> 15.000,00	> 20.000,00	0,0

4.4 Recruiter's Point of View

In this subsection, we aim to provide an analysis that can help recruiters fill their job vacancies to increase the chances of selecting good candidates in an online recruitment application.

Question: *How should recruiters meet the salary requirement for the IT department to increase the number of candidates interested in the job vacancy?*

Analysis: To answer this question, we use several iterations through our methodology. The first iteration was done through the Area Dimension 3.1 using only the IT area again. Secondly, we iterate through Benefits Dimension 3.2 twice in a row, the first one using the Salary attribute and the second, the Position attribute. The positions selected for analysis were those of junior developer, full-time developer, senior developer, and manager.

Information: In Figure 5, we observe the percentage of applications of candidates, according to the position. There is a high percentage of applications in jobs which the salary range is not described. This data is directly related to the number of open positions, which do not have a described salary 2(a). Recruiters prefer not to fill the salary requirement. We believe it is for one of the two reasons: to try to maximize the number of candidates by encouraging interest in a likely salary negotiation; or, to be able to better evaluate the candidate to subsequently make an appropriate salary offer. Graphics 6(a),6(b),6(c) and 6(d) represent the mean number of applicants applying for each position according to each salary range. In those graphics, we have that the candidates are more likely to apply for places with declared salaries. Considering only the positions of developers, we have that positions that require less experience (junior developer and full-time developer) having a higher mean of applications compared to the senior developer position. These initial positions also tend to have more applications in initial/intermediate salary ranges. Senior developers tend to apply more for high salaries. On the other hand, the manager positions are the ones with the highest mean of applications, especially for the highest salary ranges.

Knowledge: These data show that a helping system can also be created for recruiters, giving useful information when registering the vacancy (pre-registration recommendation). Let us suppose a recruiter urgently needs to get a candidate for a senior developer vacancy. What salary should the recruiter offer so that the position/vacancy has the maximum number of applications? From the analyzed data, one can observe that to obtain a higher number of applications it is necessary to offer a salary around $8,000.00. Providing a real-time estimate of job interest according to the salary range reported is a potentially useful feedback mechanism for recruiters. The RSs can also benefit from this information. It is possible to quantify how many different users the same job can be recommended taking the mean number of applications per salary range.

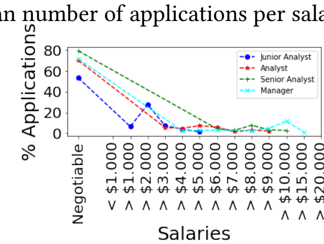
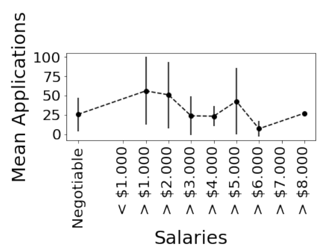
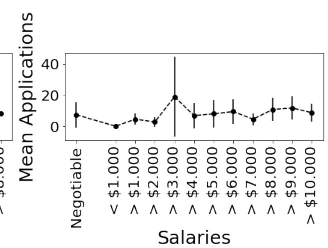


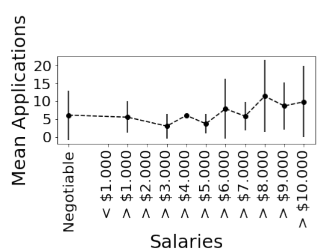
Figure 5: Percentage of applications per salary range



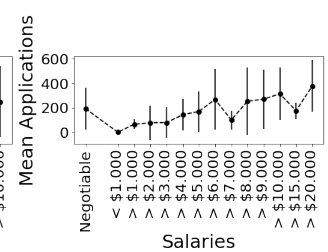
(a) Junior Analyst



(b) Analyst



(c) Senior Analyst



(d) Manager

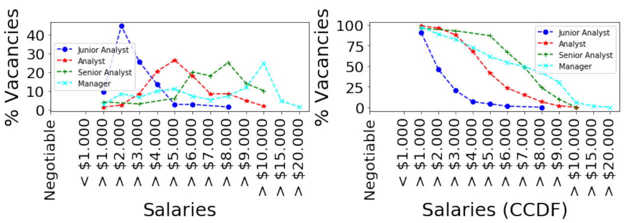
Figure 6: Mean number of applications per salary range

Question: *How attractive is a job advertisement for a candidate when compared to other ads for the same position in the IT department?*

Analysis: In this study, we used three iterations in the methodology. The first iteration was done through the Area Dimension 3.1 using the Information area. Subsequently, we iterate through Benefit Dimension 3.2 twice, in the first one using the Salary attribute and in the second, the Position/Vacancy attribute. The selected positions for analysis were those of junior developer, full-time developer, senior developer, and manager.

Information: Figure 7(a) shows the percentage of job vacancies for each position. In each of the curves in the graph, we can see that there is a peak. For the positions of developer this peak occurs in the salary range above $2,000. In a full-time developer, it is above $5,000. In the senior developer, the peak occurs about $8,000. As for manager positions, the peak of vacancy offers appears in the salary range above $10,000. The higher the salary, the more attractive the job becomes. This attractiveness can be measured through Figure 7(b), where we can see for each one of the salaries offered, how much ahead of the other vacancies the opportunity in question is located. For example, considering the position of Analyst, a salary offered above $7,000.00, is ahead of 75% of the other vacancies, since it is in the top 25% of the curve shown in the graph.

Knowledge: Together, these graphs help the system tell the recruiter how good the benefit is. Higher benefits tend to attract more applicants, especially to positions that require higher skills such as senior developer and managers. Therefore, the system may indicate that to stand out among the junior developer vacancies, a salary higher than $3,000 is enough. For a manager, however, it will be necessary to offer a salary greater than $15,000 a month. On the other hand, this information can be useful for RSs to serve more demanding users and to apply to a few vacancies. More attractive advertisements could be strictly recommended to very selective users who search for interesting positions according to their demands. This would increase the customer's satisfaction and at the same time keep him/her from getting bored with uninteresting vacancies.



(a) (b)

Figure 7: Vacancies by salary range

In addition to all the issues raised, many more could have been brought up into a discussion in this section. The set of questions addressed here are examples for the application of our methodology and how it answers real questions that can be very useful in identifying opportunities for improvement in different points of views.

5 CONCLUSIONS & FUTURE WORK

In this work, we present an iterative methodology of characterization about online recruitment services. Knowing, highlighting and understanding how to handle the key features of this type of system is of great importance to candidates and recruiters. In a way, this is the main contribution of this work: to point out the main points to be improved in the system as a whole to achieve a high level of efficiency in the match between candidates and jobs. Through this methodology divided into five dimensions of analysis (Area, Competence, Benefits, Geographical and Temporal), we are able to analyze and understand the characteristics of this domain in a sample of real data provided by Catho. We observe that there can be a great imbalance between supply and demand, considering several characteristics, such as the area of performance, different geographic regions, required skills, among others. In addition, we present details about the practical application of this methodology in real systems, through iterations that demonstrate the various levels of information that can be obtained from the correlation between the several dimensions. Despite being naturally biased conclusions about a single sample of a specific company, the proposed analyses are valid for distinct recruitment scenarios.

As future works, we intend to apply our methodology considering different periods and provide an assessment of the temporal evolution of these applications. Our goal is to propose new RSs of jobs capable of reaching a much larger number of people, placing them in the best possible positions, even in scenarios with a shortage of opportunities.

REFERENCES

[1] K. G. Coffman and A. M. Odlyzko. 2002. Internet growth: is there a "Moore's law" for data traffic? (2002), 47–93.
[2] J. O. Daramola, O. O. Oladipupo, and A. G. Musa. 2010. A Fuzzy Expert System Tool for Online Personnel Recruitments. *Int. J. Bus. Inf. Syst.* (2010).
[3] Maryam Fazel-Zarandi and Mark S. Fox. 2009. Semantic Matchmaking for Job Recruitment: An Ontology-Based Hybrid Approach.
[4] Sahin Cem Geyik, Qi Guo, Bo Hu, Cagri Ozcaglar, Ketan Thakkar, Xianren Wu, and Krishnaram Kenthapadi. 2018. Talent Search and Recommendation Systems at LinkedIn: Practical Challenges and Lessons Learned *(SIGIR '18)*.
[5] Wenxing Hong, Siting Zheng, Huan Wang, and Jianchao Shi. 2013. A Job Recommender System Based on User Clustering. (08 2013).
[6] Ioannis K. Paparrizos, Berkant Cambazoglu, and Aristides Gionis. 2011. Machine learned job recommendation. (10 2011), 325–328.
[7] Lucia Mýtna Kureková, Miroslav Beblavý, and Anna Thum-Thysen. 2015. Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. (2015).
[8] Stefan Lang, Sven Laumer, Christian Maier, and Andreas Eckhardt. 2011. Drivers, Challenges and Consequences of E-recruiting: A Literature Review. In *Proceedings of the 49th SIGMIS-CPR*. ACM, New York, NY, USA.
[9] In Lee. 2007. An Architecture for a Next-generation Holistic e-Recruiting System. *Commun. ACM* (2007), 81–85.
[10] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4 (2014), 2065–2073.
[11] Rui Liu, Wenge Rong, Yuanxin Ouyang, and Zhang Xiong. 2017. A Hierarchical Similarity Based Job Recommendation Service Framework for University Students. *Front. Comput. Sci.* (2017).
[12] Yao Lu, Sandy El Helou, and Denis Gillet. [n. d.]. A Recommender System for Job Seeking and Recruiting Website. In *Proceedings of the WWW 2013*.
[13] Jochen Malinowski, Tobias Keim, Oliver Wendt, and Tim Weitzel. 2006. Matching People and Jobs: A Bilateral Recommendation Approach.
[14] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu. 2013. Using of Jaccard Coefficient for Keywords Similarity.
[15] Kuhn Peter and Mansour Hani. 2014. Is Internet Job Search Still Ineffective? *The Economic Journal* 124, 581 (2014), 1213–1233.
[16] Luiz Pizzato, Tomek Rej, Thomas Chung, Irena Koprinska, and Judy Kay. [n. d.]. RECON: A Reciprocal Recommender for Online Dating. In *Proc. of the ACM RecSys 2010*.
[17] Bahar Salehi, Damiano Spina, Alistair Moffat, Sargol Sadeghi, Falk Scholer, Timothy Baldwin, Lawrence Cavedon, Mark Sanderson, Wilson Wong, and Justin Zobel. [n. d.]. A Living Lab Study of Query Amendment in Job Search. In *SIGIR 2018*.
[18] Damiano Spina, Maria Maistro, Yongli Ren, Sargol Sadeghi, Wilson Wong, Timothy Baldwin, Lawrence Cavedon, Alistair Moffat, Mark Sanderson, Falk Scholer, and Justin Zobel. [n. d.]. Understanding User Behavior in Job and Talent Search: An Initial Investigation. In *SIGIR 2017*.
[19] M. L. Tran, A. T. Nguyen, Q. D. Nguyen, and T. Huynh. [n. d.]. A comparison study for job recommendation. In *ICIC 2017*.
[20] Arthur Turrell, James Thurgood, Jyldyz Djumalieva, David Copple, and Bradley Speigner. 2018. *Using online job vacancies to understand the UK labour market from the bottom-up.* Technical Report.
[21] E. ul haq Dar and J. Dorn. 2018. Classification of job offers of the World Wide Web.
[22] Izabela A. Wowczko. 2015. Skills and Vacancy Analysis with Data Mining Techniques. (2015).
[23] Shuo Yang, Mohammed Korayem, Khalifeh Aljadda, Trey Grainger, and Sriraam Natarajan. 2017. Combining Content-based and Collaborative Filtering for Job Recommendation System: A Cost-sensitive Statistical Relational Learning Approach. *Knowledge-Based Systems* (09 2017).