# INFORMATION TECHNOLOGY EMPLOYERS ANALYSIS: THE CASE STUDY OF THE NORTH-WEST OF RUSSIA

**Eugene Pitukhin, Aleksandra Dyatlova, Anna Tulaeva, Aleksey Varfolomeyev**

*Petrozavodsk State University (RUSSIAN FEDERATION)*

## Abstract

The paper presents a methodology to analyse employers in some region by means of job advertisements gathering and processing. The problem of employers' evaluation and classification is important for job seekers' decision rules elaboration. A solution to this problem can be particularly useful for rapidly changing industries with a large number of small and medium-sized enterprises, e.g. IT (Information Technology) industry. The proposed methodology is tested in the paper on the example of the analysis of employers of the North-West of Russia.

Keywords: Job Advertisement, Cluster analysis, Information technology, Labor Market, Employers, Classification.

## 1 INTRODUCTION

Today information technology (IT) is one of the most actively developing industries. Many employers, regardless of the type of their activities, have staff who works in the area of information technology. Thus, employers have a need for personnel in this field, which they search for by placing job advertisements. One of the sources of filling such vacant positions is graduates of universities of IT programs. It should be noted that the demand for IT specialists is very high all over the World; therefore, a wide choice of relevant professions opens for IT job seekers.

For IT graduates there are a number of questions arising in the process of employment. One of the most important is the choice of an employer with suitable salary and working conditions. To help job seekers in making decisions it would be valuable to elaborate a methodology to classify employers, which offer IT jobs (to be short, IT employers). This methodology should be tested in terms of specific regions, since the IT job market in different regions has its own characteristics that are not fully understood.

The paper is dedicated to the methodology of classification of IT employers depending on offered vacancies, wage level, required competences and working conditions and testing of this methodology analyzing job offers of employers in the North-West of Russia. The offered methodology is based on creation of software system written in Python, which automatically collects data about vacancies from job advertisement websites, then formalizes text values, transferring it to numbers or text codes. The received tables are analyzed with methods of the cluster analysis in the suitable software environment. Results of the analysis should be interpreted by the expert.

The described methodology is applied to the analysis of IT employers in the North-West of Russia. As an example, we built a framework that formalizes data from such job websites as HeadHunter.ru (http://hh.ru). The received tables are analyzed with methods of the cluster analysis, which are a part of MS SQL Server Analysis Services. As a result, various categories of employers are marked out in dependences on the characteristics considered at a clustering. Results of the cluster analysis allow assuming some regular characteristics of this particular labor market. So, employers whose area of specialization is not connected to IT, as a rule, demand skilled experts and at the same time pay quite a lot, on the other hand, employers from the IT sphere pay less, but are less demanding to work experience, giving to employees the opportunities of additional training in the workplace.

Analysis and classification of employers in the North-West of Russia will provide new knowledge that can be used by job seekers when deciding on employment. The methodology of analysis proposed in the paper can be applied in different regions. In addition, the results of the analysis of identified competencies, which are really in demand by employers, can be used to update the curricula of the relevant bachelor and master programs

The rest of the paper is organized as follow. Part 2 presents a review of publications related to the topic. In the part 3, the methodology of employers' analysis through job advertisements processing is

presented and discussed. Part 4 describes case study experiment's design and results. Part 5 presents research conclusions and sums up the paper.

## 2 RELATED WORK

Personnel selection is an important task for any organization. The successful solution of that task allows achieving success in activities and gaining competitive advantages. The development of information technology has led to the emergence of the concept *e-recruiting* (or *online-recruiting*) [1], that is, the use of Internet technologies to publish and search for vacancies and resumes, as well as to link job seekers and employers to each other. E-recruiting platforms actively develop using a variety of advanced technologies [2, 3]. Often e-recruiting systems are equipped with the functions of personalized search and ranking [4, 5] which makes them similar to recommender systems. In addition to the obvious advantages, e-recruiting has drawbacks, in particular, the process of mutual choice can be difficult due to formal and non-transparent evaluation of indicators instead of live communication [6], but there is no alternative to that approach.

One of key parts of e-recruiting is a continuously changing set of job advertisements (ads, offers, vacancies) – as a rule, unstructured or poorly structured texts on the sites of e-recruiting platforms, containing information on the vacancy and the employer, as well as the requirements for the applicant. Job advertisements analysis by means of Data Mining and Text Mining methods has become increasingly popular in recent years [7, 8]. The same methods can be applied to the job seekers' resumes published on the Internet. That approach opens a way to create smart information services, automatically connecting suitable jobs and resumes to each other [9, 10]. Intelligence of those services can be significantly increased by using ontologies of competencies, allowing to implement semantic searching [11, 12].

The analysis of competencies derived from job advertisements leads to the ability to characterize the current demands of the labor market. It is actively used for curricula management in universities and competency evaluating in organizations [13, 14]. The papers [15, 16] are devoted to the development of curricula management models and information systems for decision support by program directors and university administrations in the process of modernization of educational programs. The paper [17] describes an example of job advertisement analysis for curricula management problem solving.

The analysis of employers is equally important, especially from the job applicants' point of view. The attractiveness of an employer for a job applicant is traditionally described by the qualitative feature *employer branding* [18]. There is big amount of publications about employer branding analysis and evaluation. As a rule, data for research are obtained by interviewing students, job seekers or employees of enterprises. In particular, the article [19] reveals determinants of students' perceptions of employers. The paper [20] is devoted to the study of the need for different employer branding strategies in different cultures. The papers [21, 22] explore the attractiveness of companies for their employees depending on various factors. The paper [23] addresses the problem of the lack of attractiveness of foreign-based companies in some countries. The paper [24] examines the attitude of India's IT industry employees to employers. The paper [25] represents a good example of quantitative analysis of employer branding and the use of social media to improve employer attractiveness.

The degree of attractiveness of employers is also demonstrated by numerous employer-rating sites, such as Glassdoor (https://www.glassdoor.com), Great Place To Work (http://reviews.greatplacetowork.com), Indeed (http://www.indeed.com/cmp/Employers/reviews), Kununu (http://www.kununu.com), etc. In Russia, the company HeadHunter (http://rating.hh.ru) annually makes the ratings of employers.

The credibility of such ratings is a separate problem [26], as those employees who are satisfied with the working conditions leave often feedback. In addition, not all companies fall into such ratings - small and medium-sized enterprises or newly opened companies remain outside the ratings. Therefore, employer-rating sites should be supplemented with data on employers obtained in other ways. Classification of employers founded on salaries and working conditions offered by them, would allow obtaining patterns useful for job seekers. Based on that patterns students could assume in advance, what types of employers can be attractive to them, regardless of whether those employers fell in ratings.

# 3 METHODOLOGY

We suppose to analyze employers through the analysis of job advertisements. As a rule, job advertisements contain in their text both direct information about the employer, such as title, city, and area of specialization (or have a link to such information) and other characteristics related to the employer (salary, working conditions, etc.). It is hard to hope to get an employer branding evaluation with the help of ads analysis, but you can try to break employers into groups according to similar characteristics of job advertisements, and then interpret the results with the help of experts. Thus, in the end, it becomes possible to obtain decision rules for future pretenders to vacant jobs, e.g. university students. Those rules should describe which types of employers are best suited for job search depending on the particular preferences.

Thus, the proposed methodology of employers' analysis consists of the following stages:

1 Choice of the region, industry branch and job ads sites for analysis

2 Developing or tuning software tools for gathering job features from sites

3 Data processing, making flat table of numbers or textual codes, describing the advertisements (part of these values characterizes the employer)

4 Application of classification or cluster analysis methods to the flat table of job advertisements features.

5 Interpretation of results, decision rules elaboration

To create a program that automatically collects information about job ads from the site, it is needed to turn to well-developed Web crawling and scraping tools. The task of Web scraping can be solved by means of traditional server side programming languages such as PHP or Java. However, the more popular decision is the development of Web scraping software by means of Python [27] and R [28] programming environments. Particularly, Python-programmers can use advanced libraries, such as requests and BeautifulSoup for HTTP-requests and HTML-parsing, as well as a powerful framework Scrapy.

Classification as a supervised (machine learning) approach to divide objects into groups of similar ones, and clustering as unsupervised classification are among the most popular directions in the area of data analysis. For the goals of job advertisement analysis clustering is more suitable because there is no learning set of advertisements or employers. There is a huge amount of clustering methods [29], the most famous of them are hierarchical clustering and the method of K-means, represented in Python and R. Most clustering methods are based on distance or similarity measures in a feature space; consequently, they are more suitable for quantitative data, because the transfer from nominal data to their distances loses a lot of valuable information about the values of features. Unlike well-known cluster analysis methods and tools, MS SQL Server Analysis Services [30] uses alternative approach to clustering which is based on factor influence evaluation. In our opinion that approach is more suitable for nominal data clustering then traditional methods. In addition, that approach facilitates the interpretation of the results of clustering.

# 4 CASE STUDY

Let us demonstrate the application of the methodology proposed above using the example of job advertisement analysis for IT employers in one of the regions of Russia.

## 4.1 Case study experiment design

To conduct the experiment, we need to determine the region and the job ads site, and to choose the way to get data from the site and to process them further.

### *4.1.1 Region and job advertisement site choosing*

In order to make the research results useful to students of our university, we chose Northwestern Federal District of Russian Federation as a research region, which includes Petrozavodsk. Obviously, IT-employers are concentrated in large cities, so for the experiment we chose the five largest cities - the centers of territorial units (subjects of federation) of Russia: Kaliningrad, Arkhangelsk, Vologda, Murmansk, and Petrozavodsk. Despite the fact that St. Petersburg is also part of the Northwestern Federal District, we did not include it in the study, since St. Petersburg and Moscow differ sharply from

other cities in Russia in terms of the level of development of the IT industry, the proposed salaries and working conditions.

In Russia, as in many other countries, there are a large number of websites publishing job advertisements. To solve the problem of automated data acquisition about vacancies, it is necessary to create specialized programs for each site, since the information structure of large sites is unique. According to the peculiarities of the information content, the job ads sites are quite different from each other. So, in a number of sites it is difficult to find job ads for individual cities and industries. It can be shown that the sites HeadHunter.ru (http://hh.ru) and Avito.ru (http://www.avito.ru) are among the most convenient by the interface and developed according to the information structure and content. However, the number of ads on HeadHunter.ru is noticeably larger (for example, for Petrozavodsk on September 26, 2017, the number of ads for work in the field of IT is 69, while on Avito.ru there are only 29), so for our experiment we chose the site HeadHunter.ru.

### *4.1.2 Web scraping, data processing and clustering*

For our experiment, we chose Python with the libraries requests and BeautifulSoup as a means of implementing Web scraping. There was written a Python-script, which collects data about job ads from the site HeadHunter.ru. In general (without taking into account the structure of the site), the pseudo-code of the script we created looks like this:

```
import requests, BeautifulSoup, csv libraries
for each city:
      define start_url for job ads searching
      get HTML from start_url
      find quantity of pages in the search result
      for each page:
            define url
            get HTML from url
            find texts about job title, job id, salary, employer title, etc.
            write texts and city name as a row in CSV-file
```

With the help of this script there were collected all the job ads in the field of IT, published on the site HeadHunter.ru on September 26, 2017 for the five cities listed in section 4.1.1. The distribution of ads by city is shown in Table 1.

***Table 1.*** *IT Job advertisements distribution by cities*

| | Population (01.01.2017, in thousands) | All IT job ads | IT job ads with described salary |
|---|---|---|---|
| Kaliningrad | 467 | 257 | 159 |
| Arkhangelsk | 351 | 71 | 53 |
| Vologda | 313 | 111 | 79 |
| Murmansk | 298 | 65 | 42 |
| Petrozavodsk | 278 | 69 | 44 |
| total | | 573 | 377 |

It is necessary to process gathered data because the fragments of texts received from the site should be reduced to a monotonous form and converted to numbers or textual codes that are convenient for cluster analysis. For this operation, the string conversion functions can be useful. They are available in many software environments, including Python. In our experiment, we used string functions of MS Excel, which allowed us to perform all the necessary operations for converting data into numerical codes.

For the clustering of ads in our experiment, we chose MS SQL Server Analysis Services because we already used this tool for job advertisement analysis in our previous work [17] and it proved to be very useful when presenting data in a form convenient for interpretation. To connect web crawling tasks, data cleaning up and clustering in a single software environment, it would be useful to implement all these tasks using Python.

## 4.2 Results of experiment

To classify employers that offer vacancies to IT job seekers, our starting point was a flat table in MS Excel format. Each row in the table relates to a job advertisement, and columns relate to advertisement features gathered from site and converted to numbers or codes. Each job advertisement is characterized by several features (factors), such as city, expected experience, proposed monthly salary, employer's area of specialization (IT or non-IT), type of employment (full, part time, etc.), job schedule, and title of vacant position.

### *4.2.1 Key influencers analyzing*

First, you need to determine what exactly we will analyze. To do this, we will use the Analyze Key Influencers tool of MS SQL Server Analysis Services to determine which of the factors have the greatest impact, for example, on the salary level that the employer indicates in each of job ads. By applying this analysis tool to the "Salary" factor, you can detect the effects of other columns on the values of the target column.

As input factors that could theoretically affect the "Salary" factor, the factors described in table 2 were chosen.

***Table 2.*** *Input factors and their values*

| Name | Description | Set of values |
|---|---|---|
| City | Name of city | Arkhangelsk, Kaliningrad, Murmansk, Petrozavodsk, Vologda |
| IT | Employer's area of specialization (IT or non-IT) | yes, no |
| Experience | Expected experience | no, 1_3, 3_6, MoreThan6 |
| Type_of_employment | Type of employment | full, part, project |
| Schedule | Job schedule | flexible, FlyInFlyOut, FullDay, remote, shift |
| Title | Title of vacant position | Android developer, Junior designer, etc. |

As a result of the analysis, monthly salary were divided into 5 levels: Very low, Low, Middle, High, Very high, where

- Very low: < $640;
- Low: $640 – $1143;
- Middle: $1143 – $1679;
- High: $1679 – $2176;
- Very high: >= $2176.

As follows from Fig. 1, the most influential factors are:

- IT;
- Experience;
- Type_of_employment.

| Column | Value | Favors | Relative Impact |
|---|---|---|---|
| experience | no | < 640 | |
| type_of_employment | part | < 640 | |
| experience | 1 _ 3 | 640 - 1143 | |
| experience | 3 _ 6 | 1143 - 1679 | |
| type_of_employment | project | 1143 - 1679 | |
| experience | moreThan6 | 1143 - 1679 | |
| experience | 3 _ 6 | 1679 - 2176 | |
| IT | yes | 1679 - 2176 | |
| experience | 3 _ 6 | >= 2176 | |

*Figure 1. The results of key influencers' analysis for salary*

From Fig. 1 you can see that for a very low salary the key is the lack of experience. For the low level of salary, experience of 1 to 3 years is required, and for the middle salary level, the requirements for experience increase (from 3 to 6 years), High level of salary is offered by employers from IT area of specialization for those applicants, whose work experience is from 3 to 6 years. In addition, the highest level of salary in all cases requires the experience of 3 to 6 years. The relative influence of the "Type_of_employment" factor is very small, so in the future analysis it will not be taken into account.

### *4.2.2 Job advertisements clustering*

MS SQL Server Analysis Services uses clustering method, which is based on analysis of influencing factors found before. There is a special tool «Detect Categories», which allows you to receive clusters by specifying the columns that will participate in cluster definition. Based on the analysis described in Section 4.2.1, three factors were selected: expected experience, level of salary, and employer's area of specialization. As a result, the following six categories or clusters were defined (see Fig. 2 and Fig. 3).

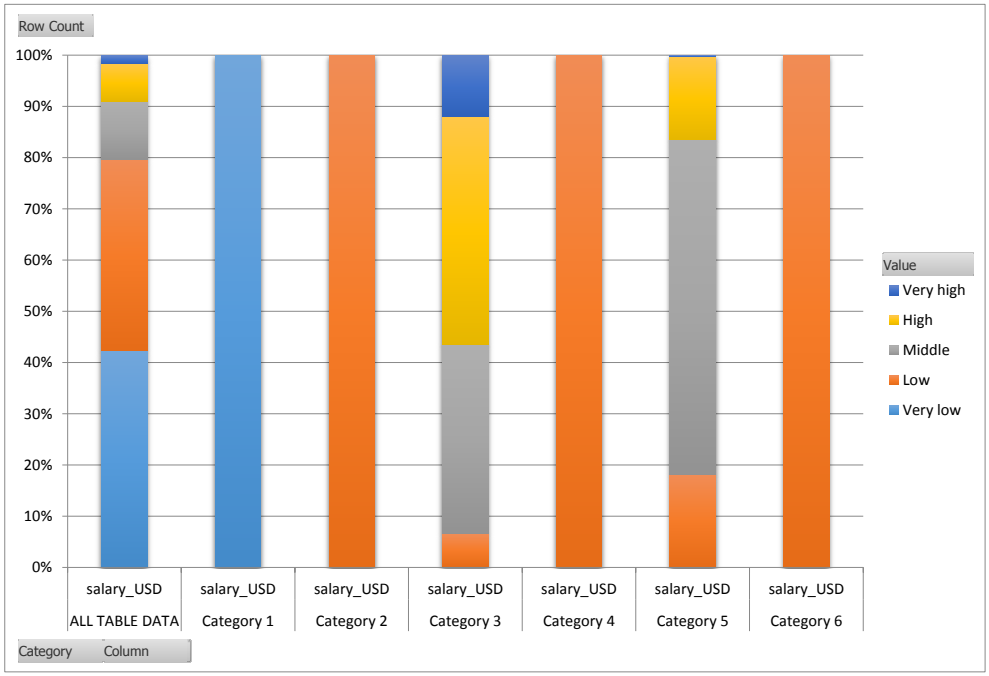


*Figure 2. Salary level distribution for the defined categories of employers*

On the basis of Fig. 2, the defined categories of employers can be characterized by the level of offered salaries:

- Category 1 is characterized by very low level of salary;
- Three categories (2, 4, and 6) are characterized by low level of salary;

– Category 5 is characterized by middle level of salary;
– Category 3 is characterized by high level of salary.

| Category | Column | Value | Relative Importance |
|---|---|---|---|
| Category 1 | salary_USD | Very low:< 640 | |
| Category 1 | experience | no | |
| Category 2 | salary_USD | Low:640 - 1143 | |
| Category 2 | IT | no | |
| Category 2 | experience | 1 _ 3 | |
| Category 2 | experience | moreThan6 | |
| Category 3 | salary_USD | High:1679 - 2176 | |
| Category 3 | experience | 3 _ 6 | |
| Category 3 | salary_USD | Middle:1143 - 1679 | |
| Category 3 | salary_USD | Very high:>= 2176 | |
| Category 3 | IT | yes | |
| Category 4 | salary_USD | Low:640 - 1143 | |
| Category 4 | IT | yes | |
| Category 4 | experience | 1 _ 3 | |
| Category 5 | salary_USD | Middle:1143 - 1679 | |
| Category 5 | experience | 3 _ 6 | |
| Category 5 | experience | moreThan6 | |
| Category 5 | salary_USD | High:1679 - 2176 | |
| Category 6 | salary_USD | Low:640 - 1143 | |
| Category 6 | experience | 3 _ 6 | |

*Figure 3. The profiles of defined categories of employers*

Let us describe categories of employers in more detail (see Fig. 3).

Category 1: Very low salary is offered to job seekers with no experience.

Category 2: The employers from non-IT areas of specialization, which offer low level of salary for IT-job seekers.

Category 3: The employers from IT area of specialization, which propose high level of salary for job seekers with substantial work experience in IT industry (of 3 to 6 years).

Category 4: The employers from IT area of specialization, which offer vacancies with low level of salary for job seekers with work experience of 1 to 3 years.

Category 5: Employers from all areas of specialization offer middle level of salary to specialists with substantial work experience in IT industry (of 3 to 6 years and more than 6 years).

Category 6: A small number of employers expect to hire specialists with work experience of 3 to 6 years at a low salary.

As a result, it can be concluded that the level of future salary is directly proportional to the level of experience at the applicant. However there are cases of exclusion, when, despite of experience, the salary remains low. It can also be concluded that for IT job seeker without experience or with little experience it would be better to contact those companies that are directly related to the field of information technology because they treat the lack of experience more loyal, apparently providing the opportunity for additional training within the organization.

## 5 CONCLUSIONS

Thus, in the course of this study, a methodology was proposed for classifying employers based on job advertisements analysis. The proposed methodology was applied for the analysis of employers providing IT job positions, in the five largest centers of territorial units in the North-West Federal District of the Russian Federation. For the analysis, a Python script was written, which gathered job ads from the site HeadHunter.ru, then the received data were translated into numbers and text codes, and analyzed using cluster analysis tools in MS SQL Server Analysis Services. As a result, groups of employers with similar job characteristics were obtained.

As a result, the following recommendations for employment in the IT field can be singled out, namely, when choosing an employer:

- the salary level is directly proportional to the work experience of the applicant;
- employers who are directly related to IT area of specialization, pay less to applicants who have little experience, but provide the opportunity for additional training directly in the organization;
- the key factors influencing the level of salary offered in vacancies are the applicant's work experience and the employer's attitude to the IT area of specialization;
- about 34% of employers did not indicate the level of salary in the job ads. Among them about 50% of employers are in the IT area if specialization. It is unclear whether this feature is related to an unofficial employment or to the fact that the salary is determined by the results of the interview.

This information can be useful for graduates of higher educational institutions of the North-West of Russia when deciding on employment.

## ACKNOWLEDGEMENTS

The paper is written in the framework of the strategic development program of Petrozavodsk State University as the Flagship University of the Republic of Karelia.

## REFERENCES

[1] I. Lee, “The evolution of e-recruiting: a content analysis of fortune 100 career Web sites,” *Journal of Electronic Commerce in Organizations*, vol. 3, no. 3, pp. 57-68, 2005.

[2] P. Meo, G. Quattrone, G. Terracina, D. Ursino, “An XML-based multiagent system for supporting online recruitment services,” *IEEE Transactions on Systems, Man, and Cybernetics*, *part A*, vol. 37, no. 4, pp. 464-480, 2007.

[3] E. Faliagka, A. K. Tsakalidis, G. Tzimas, “An integrated e-recruitment system for automated personality mining and applicant ranking,” *Internet Research*, vol. 22, no. 5, pp. 551-568, 2012.

[4] J. Malinowski, T. Keim, O. Wendt, T. Weitzel, “Matching people and jobs: a bilateral recommendation approach,” in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, vol. 6, IEEE Computer Society, 2006.

[5] U. Straccia, E. Tinelli, S. Colucci, T. Di Noia, E. Di Sciascio, “A system for retrieving top-k candidates to job positions,” in *Proceedings of the 22nd International Workshop on Description Logics,* 2009. Retrieved from http://ceur-ws.org/Vol-477/paper_7.pdf

[6] M. T. Thielsch, L. Traumer, L. Pytlik, “E-recruiting and fairness: the applicant's point of view,” *Information Technology and Management*, vol. 13, no. 2, pp. 59-67, 2012.

[7] I. A. Wowczko, “Skills and Vacancy Analysis with Data Mining Techniques,” *Informatics*, vol. 2, no. 4, pp. 31-49, 2015

[8] S. Debortoli, O. Müller, J. vom Brocke, “Comparing Business Intelligence and Big Data Skills — A Text Mining Study Using Job Advertisements,” *Business & Information Systems Engineering*, vol. 6, issue 5, pp. 289–300, 2014.

[9] J. Martinez-Gil, A. L. Paoletti, K.-D. Schewe, “A smart approach for matching, learning and querying information from the human resources domain,” in *ADBIS 2016: New Trends in Databases and Information Systems*, Springer (CCIS, vol. 637), pp. 157-167, 2016.

[10] G. Racz, A. Sali, K.-D. Schewe, “Semantic matching strategies for job recruitment: a comparison of new and known approaches,” in *FoIKS 2016: Foundations of Information and Knowledge Systems*, Springer (LNCS, vol. 9616), pp. 149-168, 2016.

[11] S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, M. Mongiello, M. Mottola, “A formal approach to ontology-based semantic match of skills descriptions,” *Journal of Universal Computer Science*, vol. 9, no. 12, pp. 1437-1454, 2003.

[12] F. Garcia Sanchez, R. Martinez-Bejar, L. Contreras, J. T. Fernandez-Breis, D. Castellanos Nieves, “An ontology-based intelligent system for recruitment,” *Expert Systems with Applications*, vol. 31, no. 2, pp. 248-263, 2006.

[13] F. Ahmed, L. F. Capretz, P. Campbell, “Evaluating the demand for soft skills in software development,” *IEEE IT Prof.*, vol. 14, pp. 44–49, 2012.

[14] Q. Yang, X. Zhang, X. Du, A. Bielefield, Y. Q. Liu, “Current market demand for core competencies of librarianship - a text mining study of American Library Association’s advertisements from 2009 through 2014,” *Applied Sciences*, vol. 6, no. 2, 48, 2016.

[15] J. Wang, Y. Li, “Design and implementation of curriculum resource management model based on domain ontology,” in *Proc. Computer Science and Information Technology Conf.*, pp. 217-221, 2010.

[16] A. Varfolomeyev, E. Pitukhin, M. Nasadkin, “Curriculum management information system,” in *ICERI2015 Proceedings*, pp. 8040-8046, 2015.

[17] E. Pitukhin, A. Varfolomeyev, A. Tulaeva, “Job advertisements analysis for curricula management: the competency approach,” in *ICERI2016 Proceedings*, pp. 2026-2035, 2016.

[18] K. Backhaus, S. Tikoo, “Conceptualizing and researching employer branding,” *Career Development International*, vol. 9, no. 5, pp. 501-517, 2004.

[19] B. Arachchige, A. Robertson, “Business student perceptions of a preferred employer: a study identifying determinants of employer branding,” *IUP Journal of Brand Management*, vol. 8, no. 3, pp. 25-46, 2011.

[20] E. Alniaçik, Ü. Alniaçik, S. Erat, K. Akçin, “Attracting talented employees to the company: do we need different employer branding strategies in different cultures?” *Procedia - Social and Behavioral Sciences*, vol. 150, pp. 336 – 344, 2014.

[21] L. Mundia, S. Mahalle, R. Matzin, G. A. N. Zakaria, N. Z. M. Abdullah, S. N. A. Latif, “Prediction of employer–employee relationships from sociodemographic variables and social values in Brunei public and private sector workers,” *Psychology Research and Behavior Management*, no. 10, pp. 257–269, 2017.

[22] I. Bakanauskiene, R. Bendaravičienė, L. Barkauskė, “Organizational attractiveness: an empirical study on employees attitudes in lithuanian business sector,” *Problems and Perspectives in Management*, vol. 15, no. 2, pp. 4-18, 2017.

[23] F. J. Froese, A. Vo, T. C. Garrett, “Organizational attractiveness of foreign-based companies: a country of origin perspective,” *International Journal of Selection and Assessment*, vol. 18, no. 3, pp. 271-281, 2010.

[24] N. Sharma,T. J. Kamalanabhan, “IT employee’s brand attitudes and the role of internal corporate communication: a survey of Indian IT industry,” *International Journal of Business Excellence*, vol. 7, no. 1, pp. 52 – 75, 2014.

[25] A.-M. Sivertzen, E. R. Nilsen, A. H. Olafsen, “Employer branding: employer attractiveness and the use of social media,” *Journal of Product & Brand Management*, vol. 22, no. 7, pp. 473-483, 2013.

[26] T. Sander, B. Sloka, H. Kalkis, “The trust of the information from employer rating platforms,” in *Advances in Human Factors, Business Management and Leadership. AHFE 2017*. Springer (Advances in Intelligent Systems and Computing, vol. 594), 2018.

[27] R. Mitchell, *Web Scraping with Python*. O'Reilly, 2015.

[28] S. Khalil, M. Fakir, “RCrawler: An R package for parallel web crawling and scraping,” *SoftwareX*, vol. 6, pp. 98–106, 2017.

[29] D. Xu, Y. Tian, “A Comprehensive Survey of Clustering Algorithms,” *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.

[30] J. MacLennan, Z. Tang, B. Crivat, *Data Mining with Microsoft SQL Server 2008*. Indianapolis: Wiley, 2009.