Extracting Functional Job Roles From Professional Social Networking Sites Profiles

Anastasiia Nesterenko

National Research University Higher School of Economics
St. Petersburg, Russia,
nastyanestor@gmail.com

1 Introduction

Despite the employment crisis on the Russian job market, demand for the IT specialists is not stagnating. Growth of flexible specialisation on the market leaves the trace in skill descriptions[8, p. 119]. Using modern Network Science and Machine Learning methods, we analysed profile data from business social networking site MoiKrug.ru and were able to extract skill map and patterns, characterising functional job roles. This paper is a part of the project is aimed at comparing signals on two sides of the Russian IT job market: in requirements extracted from job advertisements and in skills extracted from profiles of potential employees. At this stage we make an attempt to understand how the supply is represented, which functional job roles exist, and how they are connected with each other.

Todd and McKeen[7] performed the analysis of roles and their dynamics during 1970 - 1990 in the Information Systems area, uncovering three main roles: computer programmers, systems analysts, information systems managers. The research was based on the analysis of job advertisements from newspapers. They showed a growing role of communication and business skills during that period, compared to knowledge of several programming languages and other technical skills. As for systems analysts, they needed to grow in both directions, although the requirement of technical skills had increased dramatically in the mid-80s.

Later Byrd and Turner[1] divided management skills into technology management skills, business functional skills, interpersonal skills, while classification of technical skills were stayed unchanged. Noll and Wilkins[4] chose for analysis of future demand for skills the following occupations: programmers, analysts, and end-user support. "Soft skills" continued to play a significant role, while in technical skills there were some changes toward the web-based languages. In process of time, more and more attention was paid to technical skills. Litecky et al.[3], analyzing the job advertisements with the help of statistical tools and clustering, identified 20 professional categories and their respective skill sets. Assessing the similarity of skills, the researchers combined more general occupations: web developers, software developers, database developers, managers (the largest area), and analysts.

Changes in the demand for skills contribute to the emergence of new professions. Debortoli [2] studies sharp rise of Big Data jobs compared to more traditional "business intelligence" using Latent Semantic Analysis on job advertisements devoted to these areas. They found some similarities and differences in application areas (about 15), and in required skills. The methods and concepts, which are specific to BI were: database administration, software engineering, BI architecture, whereas for Big Data quantitative analysis, machine learning, database administration, software engineering, software testing and data warehousing were more salient.

Another study that has the similar goals and objectives, was the research of Wowczko I. A. [9]. Based on the selected frequency terms in job titles, were identified professional subsets: Administrator (keyword: Administrator), Analyst (keyword: Analyst), Support (keywords: Engineer), Lead (keywords: Lead, Manager), Test,Tester, Quality, QA). Using these general categories, 4755 jobs were classified. During the analysis of their description, was constructed matrix terms based on ngramms, which to some extent, are similar to skills, although they are not so clear in comparison with the previous study. In general, the categories included both technical and managerial disciplines, similarly to our work.

While these studies show emergence of new specialisations, demonstrating development of IT area, there is a lack of up-to-date comprehensive skill map of Russian IT job market, and the proposed paper is a step in this direction.

2 Data and Methods

Using rvest package, we downloaded all available at that moment (11.2015) user profiles from business social networking site MoiKrug.ru. In total there were 11000 profiles, containing more than 1000 unique skill tags. After preprocessing was done: removing punctuation, making DocumentTermMatrix, correlation matrix, we extracted the hierarchy of tags[6], using the co-occurrence of the tags in profiles, and their network characteristics, built a hierarchical skill map.

Hierarchical clusterization was made, based on this distance matrix, to analyze skill areas underrepresented in the dataset. To analyze the quality of clusterization, we used silhouette[5] plot (Fig. 1), which shows the distribution of observations in a cluster and their fit. Silhouette width calculates how close the object is to other objects within the cluster in comparison with objects from the other clusters, the higher it is, the better is alignment of the elements in the cluster[0:1].

In addition, using an association rule learning algorithm Apriori agrawal:imieliski and transactions between users and skills, we extracted frequent combination of skills, characterising job profiles for the largest clusters.

3 Results

Analysis of the CVs tag hierarchy allowed us to identify two large skill clusters, containing areas of general purpose Web development, High-load systems, and Web- and UI- and graphics design, Project management, Internet marketing.

Clusterization based on the Jaccard index allowed us to extract 9 professional fields (Table 1), which we described as general web development, design, backend, mobile application development, infrastructure, frontend, systems administration, testing, administrative cluster.

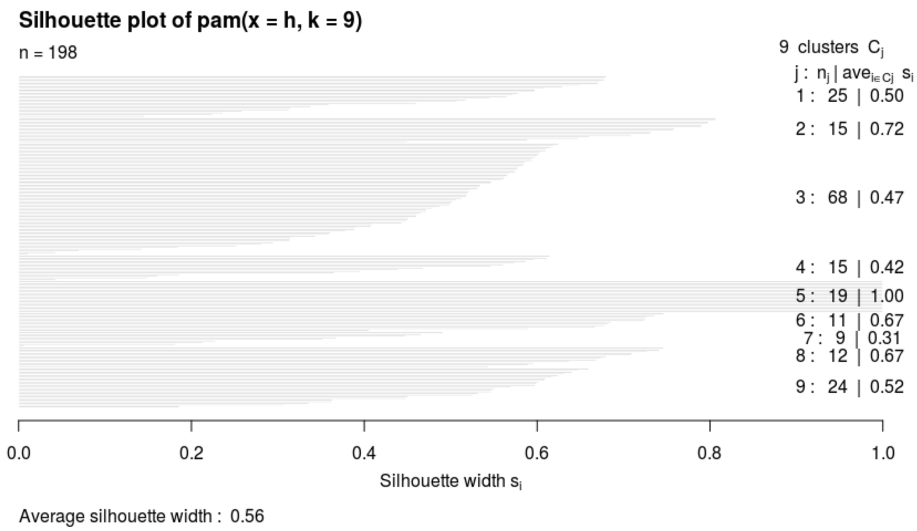

Fig. 1. Distribution of observations in clusters

Further comparison of the clusters, using the dissimilarity metric, made it possible to identify two large areas that are similar to the results of tag hierarchies, but in addition allowed to analyze smaller skill clusters.(Fig. 2).

Administrative cluster includes: marketing, analytics, sales, content and some part of management. All of these areas are quite close to each other and less presented in comparison with other sectors. Association rule analysis showed that the most common combinations of skills in administrative cluster are: SMM, Sales, Internet Marketing, Human Resource management, Project management. Therefore, roles here are quite mixed. The most common skills in the cluster of designers were UX-design, Adobe product family, Web design and Design of mobile applications. Regarding the development of mobile applications, here we have three main programming languages: Java, Objective-C and C++ and some links to them from skills like: development for Android/ for IOS, XML, Qt, SVN. In the sphere of backend the most common skills: Python and PHP. With Python we usually can find: Django, Linux, PostgreSQL, while PHP is linked with Redis, Laravel, MongoDB, Git, Symfony 2, Zend framework, Yii framework, Node.js. MySQL is connected with both of them. In the area of software development, we found trivial associations between HTML, CSS, Git, Javascript and JQuery being prevalent. A more detailed study of the links between the clusters revealed

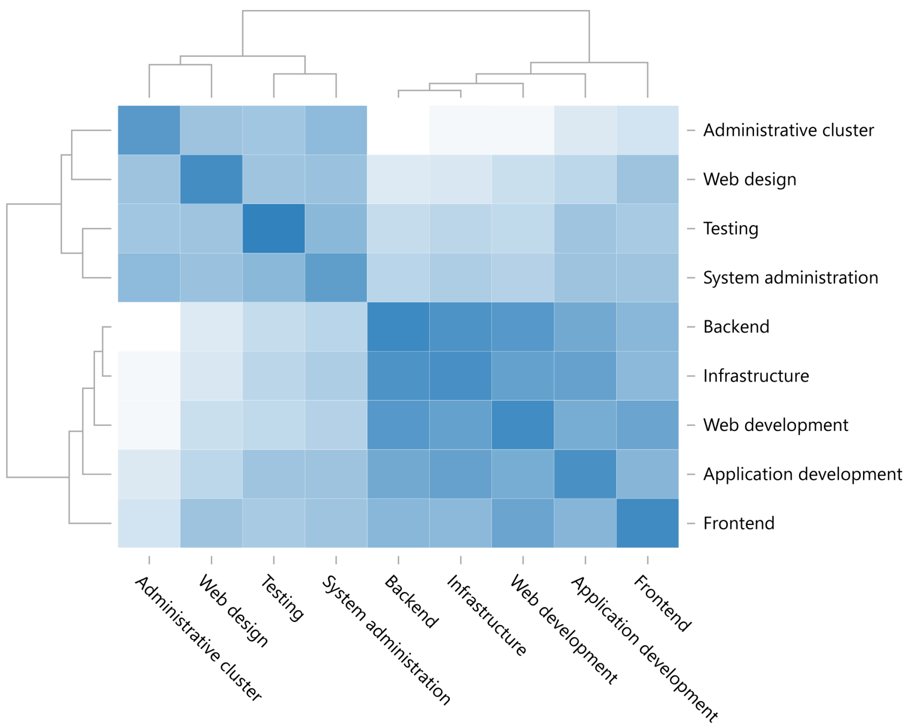

Fig. 2. Hierarchical clusterization

that system administration, software development, and machine learning are closely linked because of multipurpose programming language Python.

Unfortunately, algorithm did not reveal association rules for testing, frontend and system administration because of the lack of data, CVs in these sectors.

Table 1. Professional fields and related skills

cluster	skills	n
1	Angular.js, JQuery, HTML, CSS, Node.js Wordpress, Javascript, Web development, Git	25
2	Adobe Illustrator, UX design, Web design UI design, Adobe Indesign, Graphic design	15
3	Python, Java, SQL, C#, Ruby, XML MongoDB, MySQL, Yii framework, PHP	68
4	Swift, Development for iOS, Objective- Unity3d, Development for Android, Jira, Shell	15
5	C, C++, Delphi, Linux, Microsoft SQL server SVN, Unix, WCF, Wpf, Software Development	19
6	Grunt, Jade, Gulp bower, Less, Stylus Adaptive layout, Cross-browser layout, Sass	11
7	System administration, Network Administration Linux Administration, Project Management	9
8	Functional testing, Manual testing Software Testing, Testing Websites	12
9	Sales, Internet marketing Smm, Product Management	24

4 Conclusion and Future work

This paper presents the results of exploratory analysis of the Russian IT market, based on the data from the business social network MoiKrug.ru. Skills clusters, detected by the methods of social network analysis, revealed two large groups of functional roles. Hierarchical clustering and association rules allowed us to form nine clusters, which are closer to the professional fields. In addition there is an idea of connectedness (common skills) and separateness of areas.

Although current results dont allow us to make a direct comparison with the results of the previous studies of the IT market due the sampling bias, we underline some contemporary trends, in particular – the mixing of roles and an emergence of a large cluster of design jobs, interlinked with other IT areas, compared with previous research.

While this sample is not representative to the whole Russian IT-industry with a bias towards web-development and IT startup roles, and administrative sector jobs being underrepresented, we still consider the results interesting as they allow to discover flexible data-grounded job roles and skill patterns. Our current task is to improve our skill matching approach to allow comparisons taking into account skills on different generalisation hierarchy levels and compare these results with the structure, based on the job advertisements skills.

Acknowledgements

We would like to express our gratitude to Ekaterina Mekhnetsova, Stanislav Pozdniakov, Daria Kharkina, Vadim Voskresenskii, Paul Okopny, and Viktor Karepin.

References

1. T. A. Byrd and D. E. Turner. An exploratory analysis of the value of the skills of IT personnel: Their relationship to IS infrastructure and competitive advantage. *Decision Sciences*, 32(1):21, 2001.
2. S. Debortoli, O. Mller, and J. vom Brocke. Comparing Business Intelligence and Big Data Skills. *Business & Information Systems Engineering*, 6(5):289–300, 2014.
3. C. Litecky, A. Aken, A. Ahmad, and H. J. Nelson. Mining for computing jobs. *Software, IEEE*, 27(1):78–85, 2010.
4. W. M. Noll C. L. Critical skills of IS professionals: A model for curriculum development. *Journal of information technology education*, 1(3):143–154, 2002.
5. P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, Nov. 1987.
6. G. Tibly, P. Pollner, T. Vicsek, and G. Palla. Extracting tag hierarchies. *PloS one*, 8(12), 2013.
7. P. A. Todd, J. D. McKeen, and R. B. Gallupe. The evolution of IS job skills: a content analysis of IS job advertisements from 1970 to 1990. *MIS quarterly*, pages 1–27, 1995.
8. T. J. Watson. *Sociology, Work and Industry.* Routledge, Fourth edition, London, UK, 2003.
9. I. A. Wowczko. Skills and Vacancy Analysis with Data Mining Techniques. In *Informatics*, volume 2, pages 31–49. Multidisciplinary Digital Publishing Institute, 2015.