

Service for Professional Predictive Learning of Skills based on the Patent Analysis of Technologies

Evgeny Nikulchev
Moscow Technological Institute
Moscow, Russia
nikulchev@mail.ru

Dmitry Ilin
Moscow Technological University
Moscow, Russia
i@dmitryilin.com

Gregory Bubnov
Moscow Technological Institute
Moscow, Russia
gbubnov@gmail.com

Egor Mateshuk
Moscow Institute of Physics and Technology
Dolgoprudny, Russia
e.mateshuk@gmail.com

Abstract—In the high-tech and intensively developing sectors of the economy, including the IT industry, the composition of the professional skills and competencies necessary for successful work of specialists is rapidly changing. This is happening due to the dynamics of the composition and capabilities of key and applied technologies. Experts working in the IT industry have to constantly monitor changes in approaches of building architectures and the computational/software systems themselves. Under such conditions, the information-computing service is important, allowing in an automated mode to identify trends in the development of technologies and their corresponding professional skills. Taking into account these trends, it is possible to specifically improve the skills of specialists, adjust the curriculum of universities, centers for professional advancement, etc. As the basis for identifying trends in technology development, the authors of the article proposed to use information from international patent databases. The reason is that large IT companies execute many patents before releasing a new high-tech solution (hardware or software) to the market. In the future the use of these patents in the output leads to changes in the composition of the professional skills of specialists in demand in the labor market. The article analyzes the current demand for IT professionals by employers in the labor market on request in the relevant information systems. The possibilities of using professional social networking sites for recruiting personnel by organizations, searching for jobs by specialists are considered. A prototype of the predictive learning service for the professional social networking site was developed. It provides the following: monitoring of the demand for professional skills in the labor market; the analysis of patents on technologies that are the basis for each of the existing and projected professional skills. The developed service will allow determining the levels of demand for professional skills; actualizing and improving the job seekers' professional skills; organizing the professional social networking sites; to form personal programs of training on promising technologies.

Keywords—Online social networks; social networking sites; technology life cycle; predictive learning; patent activity analysis; professional skills; topic detection; LinkedIn; ResearchGate; labor market; scalable services; decision making support; Node.JS

I. INTRODUCTION

Experts of high-tech specialties must constantly improve their skills to maintain their professional competence. Usually it involves mastering new competencies based on new technologies. This process takes place at different levels: large companies that are concerned about the competence of their employees organize advanced training, or the professionals themselves independently choose educational trajectories to remain sought-after specialists. In all cases, the most important question is - what exactly to learn. The main criteria for choosing an educational trajectory of professional development are: the skill should be popular in the labor market; it is desirable that the skill should be in demand in the future.

It is difficult to determine the most suitable sources for the decision support when choosing a professional trajectory. Companies, when choosing skills for the professional development of their employees, are guided by the technologies used in their activity and research on their development from common, often conjunctural, considerations and drawing information from analytical reports on the development of professions. Sources of information for individual choice of a specific person, which answer the question what should be learned in order not to lag in the profession, are: interviewing acquaintances, reading analytical materials. Here the social networking sites take a significant place.

Skills can spread in social networking sites not only as a result of technology development, but as a result of advertising activities. For example, it was observed regarding the skill of "deep learning technology", when the popularity of the term was ahead of the proposed algorithmical and technological solutions. Another example is "cloud computing": a set of previously known technologies turned into a new term and became an established professional skill.

In contrast to social networking sites focused on communication, in professional social networking sites, such as LinkedIn and ResearchGate [1], [2], the skill is one of the tools

of communication. The search for thematic publications, the formation of professional groups, the creation of topics for discussions are made with the use of the skills. The skills are also a tool for finding employers. Employers actively use resume search tools provided by social networking sites [3]-[5]. One of the reasons is a smaller proportion of untruthful information in user profiles [6]. There are also less common variations of social networking sites known as decentralized social networking sites [7], [8]. There is the qualitative difference in the audience [9]: users of professional social networking sites are mostly middle-aged people who are interested in building a network of professional relations.

It is possible to extend the use of such a tool as the skills for professional social networking sites, adding informativity through provision of information about innovations coming from scientific environment. Mainly it is patent activity reflecting the importance of technologies and corresponding skills [10]-[12].

II. CONCEPT

It seems right to us that a set of skills determines the professionalism of each specialist at a particular time. For example, a sought-after programmer in the labor market is the one that has a set of desirable skills and knowledge of required technologies, specific operation systems, programming languages and software development methodologies, which are currently used. Therefore, the programmer can be considered a professional. Exactly these skills reflect professional fitness. Obsolete programming languages, as well as cross-stitching, cannot be an actual professional skill. This applies to many high-tech, innovative professions, for example, a cardiac surgeon must possess both surgery methods and technical means that are recognized as relevant and effective at the present time and not in the past.

Thus, we consider professional fitness as the expertise in a set of relevant skills. Given the representation of the skills in social networking sites, our concept fits well into them.

The aim of our work is to create a service for actualization of the skills of employees of high-tech occupations, as well as for forecasting the relevance of skills related to the professions of the future.

In this paper we present the data-computing service designed to build educational trajectories based on professional predictive learning. Saying professional predictive learning we mean training the skills that will be in demand in some professions in the future. In other words, training is conducted before the impact of the technology on the market.

The following key conceptual theses were defined in the construction of the service:

- A set of skills determines the state of professional fitness at a given time. Thus, professional fitness is a dynamic characteristic. Even a very good knowledge of legacy technologies will not allow you to immediately begin to perform the functional duties and the success of the basketball team in college is not able to add any surgery skills. The need for certain skills is changing

with time even in scope of one profession. Therefore, it is possible to assess the timeliness of the skill.

- The importance of skills is determined by their relevance to the labor market. After all, in professional activity knowledge of current technologies takes the first place. The forecast of the development of the demand for technology in the future is also a question of interest. Based on the monitoring of vacancies, one can thereby assess the current demand for the skill.
- Skills are of importance in combinations with each other.
- The service should be built into social networking sites and use the connections and concepts of professional communication. Here we set the goal to use the existing terminology and principles of construction. The service should perform an analytical function, for example, for recommendations on the trajectory of advanced training.

Consider the use case of the service for an individual user (Fig. 1). First, the user's skills are analyzed. Based on the vacancies, the service identifies skills that can complement existing ones and increase the value of the specialist in the labor market. Further steps are optional. The user can request the forecast of the demand for the skills based on scientific publications and patents. After that, the user can request a set of training materials. This part can be implemented with the use of e-learning systems [13] or by integration with massive open online courses, like Coursera [14].

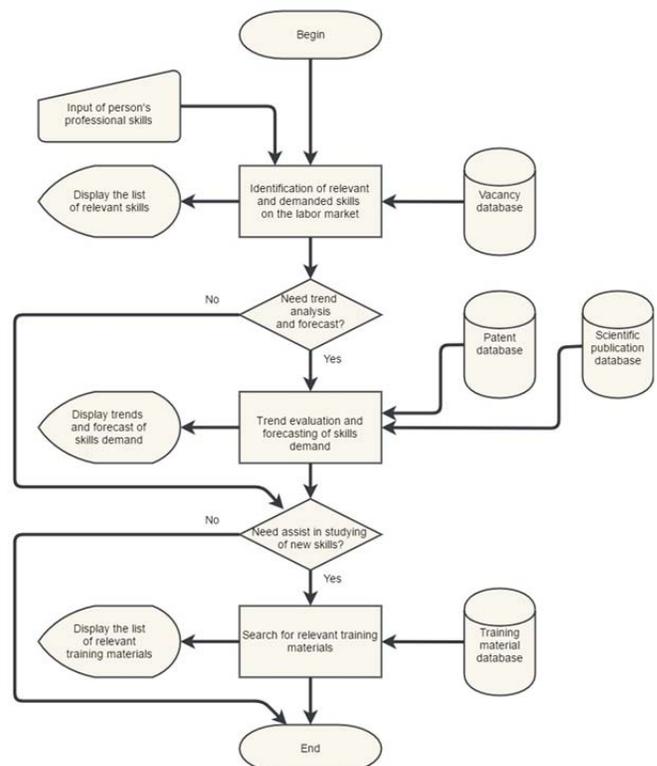


Fig. 1. System flowchart of the predictive learning service usage.

III. DESIGN

The software architecture of the service, as well as the stack of the applied software solutions is described in the paper [15]. In this section of the paper we focus on the functional component organization.

Fig. 2 shows a conceptual structure of the service that contains five major groups that implement the functionality.

- Initialization scripts – a set of software components that implement detection, startup, initial configuration and integration of all other parts of the project.
- Modules – these are the main components of the software project, divided into groups. They contain all the main functionality that allows the service to collect, store, analyze and display information for the user.
- Shell scripts – this is a set of components independent of the main application process, which are required for deployment or experimental research that does not require a graphical interface.
- Configurations – this is a set of settings with the possibility of overriding them depending on the execution environment. For example, it is possible to

store two different configurations for the development and production environment.

- 3rd party modules – these are collections of ready-made libraries available under open source software licenses. They are connected using package managers. This role is performed by the NPM and Bower utilities for server-side and user interface components respectively.

Modules have a uniform directory structure:

- Config – contains module configurations, lists of implemented database entities, lists of cron jobs.
- Cronjobs – contains cron job files for scheduling.
- Entities – contains database entity files, as a special type of models.
- Models – contains models that implement the business logic of the service.
- Routes – contains HTTP-request handlers that are mapped to the URL patterns.
- Static – contains CSS files and front-end software components.

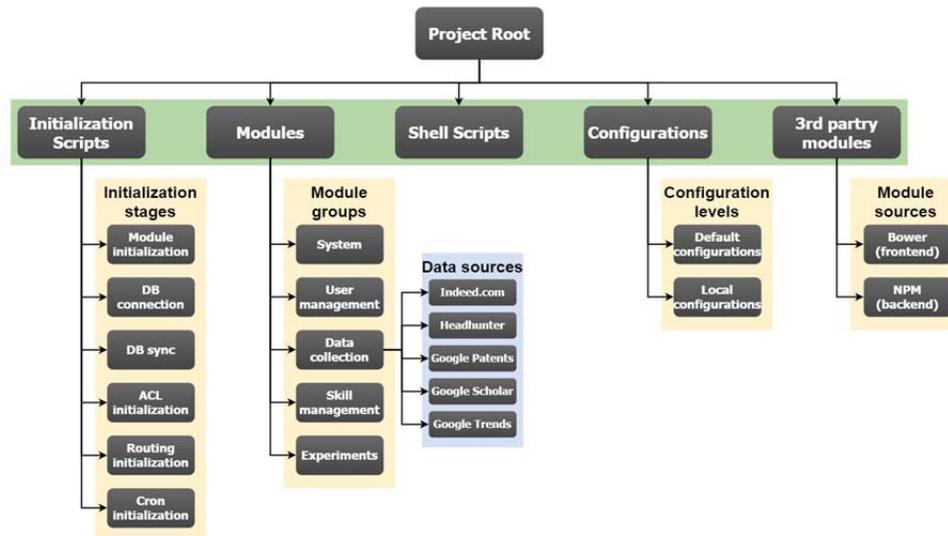


Fig. 2. Predictive learning service project structure.

The possibility of partial system initialization is very important. In the prototype, there are three launch modes: the data collection server, the HTTP server, and the combined mode. The first mode initializes the components necessary to aggregate data from external sources on a schedule. The second initializes the components responsible for the user interface and access to the service via the HTTP protocol, which includes the REST API, access control list and the user interface. Combined mode involves the launch of all components, which ensures the operation of the service on one computing node. In addition to this separation, a limited number of modules can be launched, which allows the system to be divided into many independent services, each of them implements part of the functionality. Thanks to this, with increasing the number of modules, the service retains the

opportunities for scaling. It can be further developed to interact with a large number of data sources, thereby increasing the value of information for users. In addition, based on the idea of the proposed service, specialized professional social networking sites can be built, focused on specific areas of knowledge (for example, in the IT sphere).

IV. USAGE EXAMPLE

A prototype of predictive learning service for the professional social networking site was developed in scope of the experiment, the collection. It performs homogenization and analysis of data from five external sources. Some of them provide data for more than one indicator that reflects the labor market state. Consider the current features of the service.

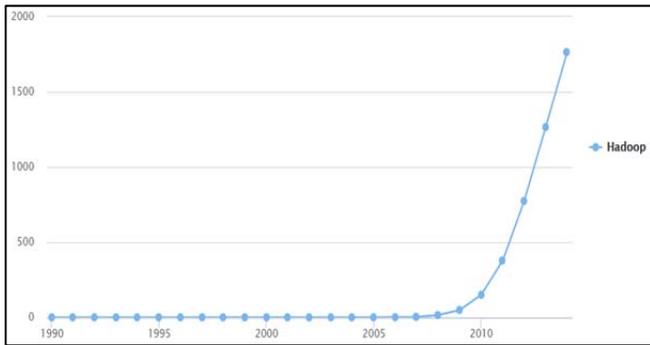


Fig. 3. Annual patent activity – Hadoop (Google Patents).

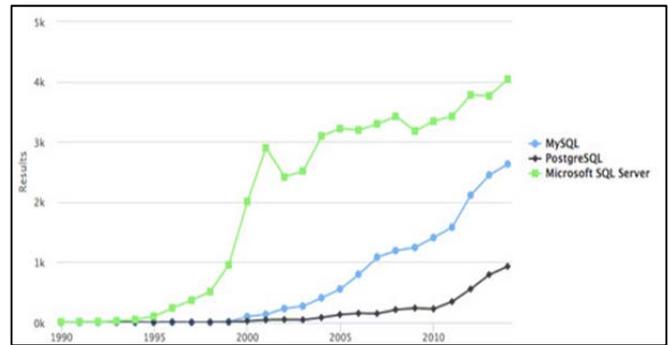


Fig. 5. Annual patent activity – RDBMS (Google Patents).

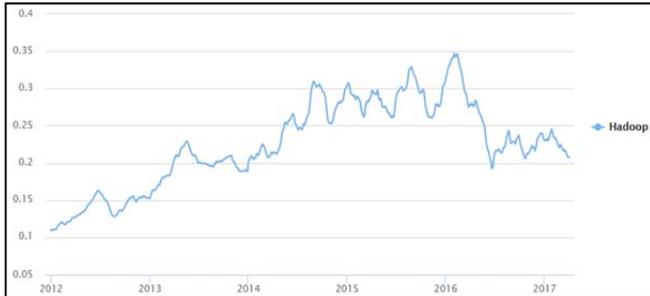


Fig. 4. Vacancies in the United States – Hadoop (Indeed.com, values in percent).

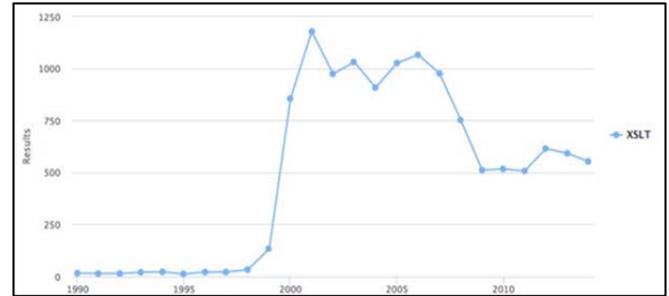


Fig. 6. Annual patent activity – XSLT language (Google Patents).

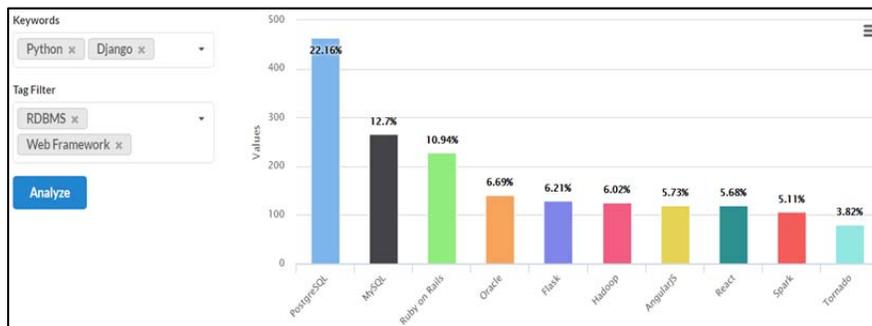


Fig. 7. Relevant RDBMS and web-frameworks for Python and Django (Headhunter).

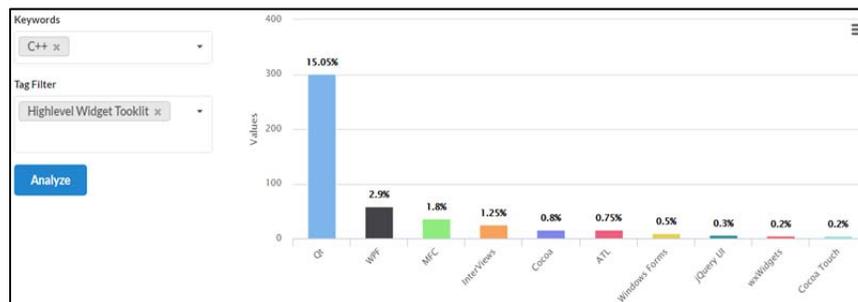


Fig. 8. Relevant widget toolkits for C++ (Headhunter).

Rapid growth of patents activity (Fig. 3) in 2010-2014 indicated growing interest in Hadoop technology. It led to growing demand on experts working with this technology (Fig. 4) in 2012-2015. Lower number of vacancies in 2016-2017 might be related to the well-established Big Data segment in the labor market which results in better balance between vacancies and CVs.

The service allows tracking the current stage of technology life cycle on the basis of patent activity. Data on several well-known RDBMS can be examined as an example (Fig. 5). Also, one can note that the XSLT, a known XML-document processing language, becomes obsolete (Fig. 6).

The predictive learning service solves the problem of selection of the most relevant skills for the studying. Let's

assume that the developer knows two main skills: Python and related web-framework Django. In order to increase the developer's own value in the labor market, it is relevant to study one or more RDBMS and another web-framework, with which the developer will have to interact. The results of labor market analysis (Fig. 7) show the relevant result: the most demanded RDBMS are the PostgreSQL and MySQL while the most demanded web-frameworks are the Ruby on Rails, Flask and AngularJS. Ruby on Rails is in demand because main Python's alternative is programming language named Ruby and vacancies frequently contain both titles.

A similar study can be performed for C++ programming language. In this case, it is reasonable to determine the most relevant widget toolkits. As it is shown in Fig. 8, the most demanded toolkit is Qt while others are less common.

V. DISCUSSION

The use of the developed service is promising for developers as a feature provided by social networking sites. The service can be applied for corporate training as well as for self-improvement of professional skills.

Indeed, every professional considers it important to know which of professional skills will be in demand in the future and what technologies need to be learned to keep being competitive in the profession.

There are several computational difficulties regarding the service, one of which is the big data. It is necessary to analyze the whole volume of patents and to identify the word combinations which frequency of meeting begins to grow. It is computationally complex task.

As research has shown, the patent business is highly developed and the annual number of patents for key technologies can be measured in thousands. Each individual patent is difficult to assess. Often patents mutually complement each other. In connection with this, companies are trying to create and buy patents packages around one technology they are interested in.

The company's production capabilities can be highly dependent on a single patent among many related. This led to the emergence of entire organizations that deal with the purchase and sale of patents (patent trolls, protective aggregators, super-aggregators), which further confirms the consistency of the emerging market.

The aforesaid allows us to state that when a valuable technology arises, one should expect a noticeable increase in patent activity. Due to this, patents can be considered a reliable source of information on trends in the development of the market. You can improve the reliability of the forecast and check its validity by supplementing the patent data with current indicators of the labor market.

Existing solutions integrate the idea of skills superficially. The well-known Coursera service is focused more on the professions (or entire expert areas), rather than on individual skills. The system of recommendations in it is based on global areas of knowledge (for example, Computer Science), which are further divided only into directions (algorithms, software

development). Service from Microsoft - Microsoft Education Community - also provides a platform for education and professional development. However, as in Coursera, it lacks specific skills (which can be found in professional social networking sites, such as ResearchGate). For this reason, there is no mechanism for organizing thematic discussions that are valuable for professional communities.

VI. CONCLUSION

The review of scientific works dedicated to social networking sites, professional social networking sites and analysis of patent activity was carried out. It was determined that there is a correlation between the development of technology and the annual number of patents.

The conception of the predictive learning service has been formed. The service can become a link between the professional social networking site, recruitment agencies and direct participants in the labor market.

The conceptual structure of the software solution is presented. The work of the prototype is shown on the example of the analysis of the demand for knowledge of programming languages, widget toolkits, RDBMS and Big Data instruments.

The analysis revealed the main difficulties to be solved: the lack of a structured description of skills, varying level of detail, lack of standards. Services that could implement the concept of skills for better organization of professional communities and training of specialists are considered. The conclusion is drawn that the use of patents for the analysis of trends in the labor market is possible and the addition of patent data with information regarding the labor market will increase the reliability of the results.

REFERENCES

- [1] M. Bastian, M. Hayes, W. Vaughan, S. Shah, P. Skomoroch, H. Kim, S. Uryasev and C. Lloyd, "LinkedIn Skills: Large-Scale Topic Extraction and Inference Mathieu," in Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14, 2014.
- [2] D. Nicholas, D. Clark and E. Herman, "ResearchGate: Reputation uncovered," *Learned Publishing*, vol. 29, no. 3, pp. 173-182, 2016.
- [3] D. Aguado, R. Rico, V. J. Rubio and L. Fernández, "Applicant reactions to social network web use in personnel selection and assessment," *Revista de Psicología del Trabajo y de las Organizaciones*, vol. 32, no. 3, pp. 183-190, 2016.
- [4] J. K. H. Chiang and H. Y. Suen, "Self-presentation and hiring recommendations in online communities: Lessons from LinkedIn," *Computers in Human Behavior*, vol. 48, pp. 516-524, 2015.
- [5] J. van Dijk, "You have one identity: performing the self on Facebook and LinkedIn," *Media, Culture & Society*, vol. 35, no. 2, pp. 199-215, 2013.
- [6] J. Guillory and J. T. Hancock, "The Effect of LinkedIn on Deception in Resumes," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 3, pp. 135-140, 2012.
- [7] Datta, S. Buchegger, L.-H. Vu, T. Strufe and K. Rzadca, "Decentralized Online Social Networks," in *Handbook of Social Network Technologies and Applications*, Springer Science & Business Media, 2010, pp. 349-378.
- [8] Y. Wang, L. Wei, A. V. Vasilakos and Q. Jin, "Device-to-Device based mobile social networking in proximity (MSNP) on smartphones: Framework, challenges and prototype," *Future Generation Computer Systems*, 2015.
- [9] M. M. Skeels and J. Grudin, "When Social Networks Cross Boundaries: A Case Study of Workplace Use of Facebook and LinkedIn," *Group '09*, vol. 10, no. 3, pp. 95-103, 2009.

- [10] H. M. Järvenpää, S. J. Mäkinen and M. Seppänen, "Patent and publishing activity sequence over a technology's life cycle," *Technological Forecasting and Social Change*, vol. 78, no. 2, pp. 283-293, 2011.
- [11] L. Gao, A. L. Porter, J. Wang, S. Fang, X. Zhang, T. Ma, W. Wang and L. Huang, "Technology life cycle analysis method based on patent documents," *Technological Forecasting and Social Change*, vol. 80, no. 3, pp. 398-407, 2013.
- [12] H. Ernst, "Patent information for strategic technology management," *World Patent Information*, vol. 25, no. 3, pp. 233-242, 2003.
- [13] P. C. Sun, R. J. Tsai, G. Finger, Y. Y. Chen and D. Yeh, "What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction," *Computers and Education*, vol. 50, no. 4, pp. 1183-1202, 2008.
- [14] J. Knox, "Digital culture clash: "Massive" education in the e-learning and digital cultures MOOC," *Distance Education*, vol. 35, no. 2, p. 164-177, 2014.
- [15] E. Nikulchev, D. Ilin, G. Bubnov and E. Mateshuk, "Scalable Service for Predictive Learning Based on the Professional Social Networking Sites," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 5, 2017.