**Contemporary Engineering Sciences, Vol. 9, 2016, no. 14, 671 - 678**
**HIKARI Ltd, www.m-hikari.com**
**http://dx.doi.org/10.12988/ces.2016.6327**

# Monitoring of Information Space for Professional Skills Demand

**Evgeny Nikulchev**

Moscow Technological Institute, Moscow, Russia

**Dmitry Ilin**

Moscow Technological University MIREA, Moscow, Russia

**Dmitry Biryukov**

Moscow Technological Institute, Moscow, Russia

**Gregory Bubnov**

Moscow Technological Institute, Moscow, Russia



**Abstract**

At the present time the technology market is constantly growing, new technologies emerge every day so it makes it hard to track all of them and evaluate their potential. Therefore, it is difficult to decide if they are promising enough to be learned and integrated into a professional education system. It is necessary to take into account that the growing demand for professionals follows the growing demand for specific technology, while the education process, in general, takes significant time. As a consequence, the acute need to change the staff training approach from reactive to proactive arises. The article suggests methods of expertise demand monitoring for the IT sector, based on comprehensive monitoring of the Web, including employers' demands, company history and publications. Specialized software, implemented for solving forecasting problems, will be used as the main tool for market demands forecasting.

**Keywords**: skills, service for learning, professional expertise

## 1 Introduction

There are many different systems, which track and systematize various events that could be useful to help to comprehend technology evolution dynamics. But these systems usually do not have centralized data storage and do not estimate all required external factors. Precise trend prognosis requires collecting and involvement of maximal amount of available information.

Thus, it is expedient to develop a system in order to perform data collection from public sources and to analyze a company's statistics taking into account all available external factors to forecast expertise relevance rise and fall. A combination of public sources representing global trends and internal source (a company's statistical data) representing local trends should improve forecasting accuracy.

Such public sources are: the number of scientific paper publications per year (Google Scholar), the number of patent filings per year (Google Patents), data based on search requests (Google Trends), and also the number of available job positions based on data from internet-recruitment services.

The usage of such a solution could provide a number of competitive advantages, such as:

- Determination of the fastest growing technologies on the market.
- Timely retraining of company's highly qualified personnel.
- Precise choice of the most relevant classes for university studies.
- Improvement of educational programs based on market demands.

## 2 Materials and methods

The hypothesis is that having enough statistical data about a company's staff workload on specific expertise, and also based on the law of large numbers regarding public sources, one can forecast the market demand for a specific category of professionals with high precision.

The algorithmic base of the software solution should be built based on the intellectual time series analysis.

The software solution has to have a number of features such as: data collection, data storage, and statistics processing. It is beneficial to implement data collection using 2 different approaches: real time data collection and scheduled data collection on daily basis.

Data storage is a trivial task and does not apply any restrictions. It can be implemented using DBMS MySQL.

Data processing is essential for the software solution. Forecasting should be performed at 2 levels: forecasting of each public data source trend and forecasting of internal company data using public sources forecast results as overlay data.

The result of the algorithm execution will be a chart representing trends based

on the achieved forecast. Future releases of the program will include an expert system.

It has been decided to use a company's statistical data as the main data source for forecasting. The forecast will be based on the dynamics of the quantitative measure of expertise involved in a company's projects.

However, usage of a single data source could not be considered as fully representational. It was decided to use public sources as well. The choice is: Google Patents, Google Scholar, Google Trends, HeadHunter, and Indeed. These resources have information on different areas of the IT market, so it will help to gain a more complete picture about demanded competencies and about competencies, which are going to be in demand.

Google Trends is a public web-service, based on Google search, and it shows popularity index of words in search requests. Trend dynamics can be requested for a specified time span and it includes extra geolocation data.

Google Patents is a search engine indexing patents and patent filings from United States Patent and Trademark Office (USPTO), European Patent Office (EPO), World Intellectual Property Organization (WIPO), Deutsches Patent- und Markenamt (DPMA), Canadian Intellectual Property Office (CIPO), and China's State Intellectual Property Office (SIPO). The earliest patents from the United States date back to 1790, EPO and WIPO – to 1978. Optical character recognition has been applied to the oldest patents. All foreign patents are translated to English using Google Translate to be indexed and available in search results.

Google Scholar is a scientific publication search engine, which works with all available categories of publications and disciplines. It has been online since November 2004. It indexes most peer-reviewed online journals of Europe and America's largest scholarly publishers. Google Scholar indexes public scientific articles as well as papers published in commercial journals.

It is reasonable to assume that the number of references to a specified competence in scientific publications for a selected time span should be in direct proportion with this competence's relevance.

HeadHunter – a Russian internet-recruitment company. Its website contains relevant vacancies for numerous occupations. Indeed Inc. – an American internet-recruitment company.

Usage of global market information, as well as local market information, should lead to more precise forecasting results.

Linear regression and SMO regression from Weka library (implemented within the forecasting plugin) are being used in the software solution as forecasting methods, as well as the autoregressive model method. The library is licensed under GNU GPL [1].

## 3 Experimental research

Solution market analysis has shown that there is a wide number of software solutions with comparable functionality to collect, analyze and visualize forecast results. The closest system class in the applied sense is the marketing information

system class (MkIS). Its main aim is to identify, measure and forecast marketing [2–4]. This aim is very close to trend forecasting in the IT sector. Despite numerous theoretical publications, most of the software solutions available on the market are a combination of CRM and PRM systems with an addition of campaign success evaluation.

However, the market does not have any application software for the defined problem. Thus, it can be considered as an additional confirmation of the expediency of implementation.

Google Trends service offers information in a convenient format so it can be easily aggregated. It has its own data visualization. It offers information from 2007 to the current year, but statistics for the current year are incomplete, so the data for this year should not be included into calculations.

Google Patents offers information on a number of patents found using a specified keyword. Patent filings are considered as the most relevant value and it follows aggregation of search result count. As an advantage it offers a wider date range than Google Trends but it also has a number of disadvantages. Firstly, the data collection process takes significantly more time because each value can only be obtained by using a specific request. Secondly, the number of search results for 2014 and 2015 years together does not match the sum of results for 2014 and 2015 years retrieved separately. Thirdly, the charts built based on collected data on various competences are, in most cases, identical and differ only by amplitude as it is shown in fig. 1.

Google Scholar has mostly the same advantages and disadvantages as Google Patents. Unlike Google Patents, it does not offer search date ranges shorter than a year. As well as this, there is a crawler bot protection, which triggers after several requests. It requires the user to enter a CAPTCHA or to pass a similar test to prove that the requester is not a crawling bot.

The main goal of using HeadHunter is to collect number of available job positions found by a keyword. This service has following disadvantages:

– No functionality to request historical data by a specified date range.
– The data collection mechanism must provide data storage for request results.
– It requires the user to have a list of keywords long before the collecting process.
–

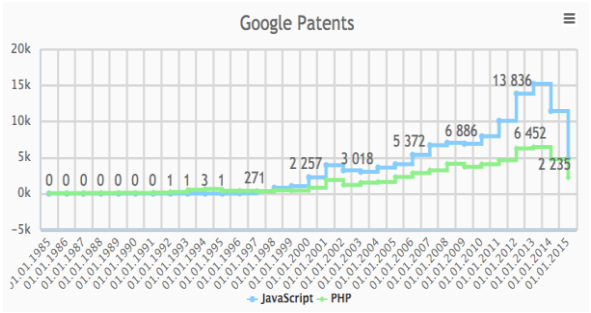


Fig. 1. Example of Google Patents statistics

However, this service provides open access to API and gives precise results.

Indeed this service is similar to HeadHunter in most cases. Yet it has a couple of differences:

– API requires the user to be registered.

– Searching using keywords such as LESS could give unnecessary information because it intersects with a commonly used word.

In general, the number of found vacancies in comparison to HeadHunter is ten times larger. Yet, some competences have a local market influence, as shown in fig. 2. The initial data (fig. 3.) has been processed using the aforementioned methods to forecast a known time span of 300 days. The results can be viewed in fig. 4.

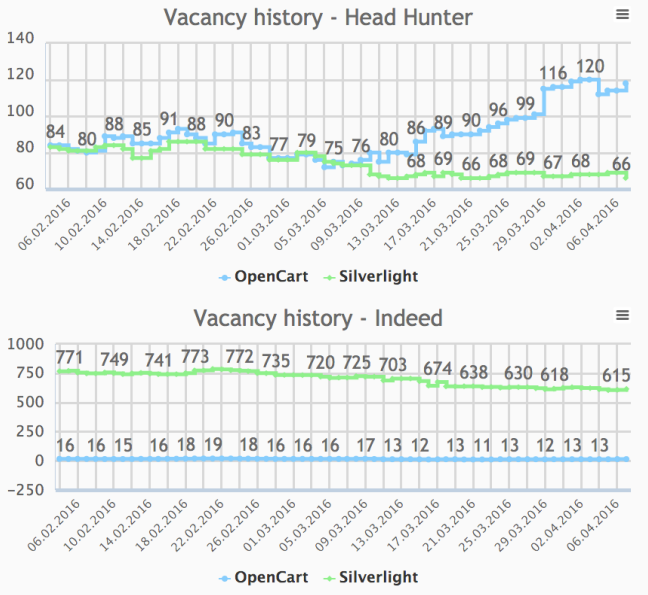


Fig. 2. Comparison of HeadHunter and Indeed statistics

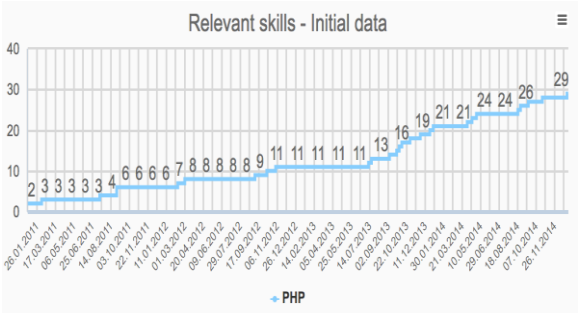


Fig. 3. Initial data

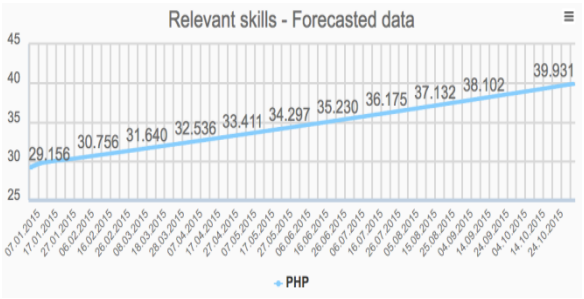


Fig. 4. Linear regression

The SMO regression method has the smallest mean absolute percentage error. Now it can be applied with the previously collected trend statistics from public sources. The result is shown in fig. 5.

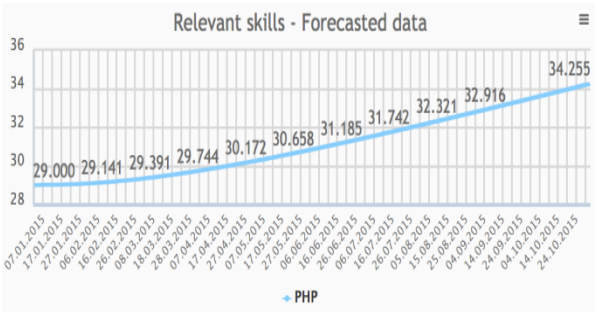


Fig. 5. SMO regression with overlay data

The mean absolute percentage error increased to 4,84%, which is presumably related to the lack of precise statistical data.

Furthermore, the results of the SMO algorithm differed depending on the operation system environment. The difference between calculations in Windows and Linux is approximately 3.7%. The other algorithm results were not affected by the execution environment.

Based on the results, a method was developed for the monitoring of relevant expertise in the IT sector. It consists of the following steps:

1. Data collection and identification of current knowledge and competences used in a company or a university.
2. Data collection from public sources based on a keyword list.
3. Analysis of the dynamics of trends in the Web.
4. Forecasting of local trends, taking into account the dynamics of the Web.
5. If demand for the expertise increases, then there is a need to expand the number of employees working in forecasted competence.
6. If a recession is forecasted, then there is a need for staff cutbacks or retraining.

Trial operations using provided data and studies have shown a growing demand for technologies such as PHP, Magento, JavaScript, CSS.

A slight increase in demand can be observed for the technologies: SCSS, C#, and Hybris.

The demand for Linux expertise has remained on the same level, while the demand for Windows Server expertise has shown a minor recession.

## 4 Discussion

It should be assumed that not all organizations collect statistics on their used competencies, which imposes a number of restrictions on the integration of the developed system into business processes.

At the moment we do not have sufficient statistical data to say with certainty that the addition of trending data increases the accuracy of the forecast.

By virtue of the domain specificity, artificial tests cannot confirm or deny the influence of external factors on the local trends, so it is planned to collect more up-to-date statistics. Therefore using the collected material it will be possible to increase the accuracy of forecast results and to get additional confirmation of software solution correctness and the method in general.

There is a plan to expand the software to the level of an expert system to reduce the qualification requirements for the user in future [5–6].

## 5 Conclusion

Usage of the developed system is best justified in major educational and commercial organizations, whose scale allows to diversify activities to better respond to current and forecasted market needs. Usage of the software solution is less appropriate for medium and small organizations since demand for their services is largely determined not by general trend of the market, but rather by the reputation of the organization earned by solving issues of specific classes.

## References

[1] X. Chen, Y. Ye, G. Williams, X. Xu, A survey of open source data mining systems, Chapter in *Emerging Technologies in Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg, 2007, 3 - 14. http://dx.doi.org/10.1007/978-3-540-77018-3_2

[2] S. M. S. Freihat, The Role of Marketing Information System in Marketing Decision-making in Jordanian Shareholding Medicines Production Companies, *International Journal of Research and Reviews in Applied Sciences*, **11** (2012), 326 - 336.

[3] M. S. Ezekiel, J. F. Eze, J. A. Anyadighibe, A Study of Marketing Information System (MIS) As a Contributory Factor in the Performance of Selected Transport Companies in Calabar Metropolis, *American Journal of Tourism Research*, **2** (2013) 154 - 159.

[4] S. Titov, E. Nikulchev, G. Bubnov, Learning Practices as a Tool forr Quality Costs Reduction in Construction Projects, *Quality - Access to Success*, **16** (2015), 68 – 70.

[5] V.N. Petrushin, E.V. Nikulchev, D.A. Korolev, Histogram arithmetic under uncertainty of probability density function, *Applied Mathematical Sciences*, **9** (2015), no. 141, 7043 – 7052.
http://dx.doi.org/10.12988/ams.2015.510644

[6] N.N. Astakhova, L.A. Demidova, E.V. Nikulchev, Forecasting method for grouped time series with the use of k-means algorithm, *Applied Mathematical Sciences*, **9** (2015), no. 97, 4813-4830.
http://dx.doi.org/10.12988/ams.2015.55391

**Received: April 14, 2016; Published: June 3, 2016**