**Victor I. Tishchenko**

Institute for Systems Analysis of Federal Research Center “Informatics and Control” of Russian Academy of Sciences

tishchenkovictor@gmail.com

**Web-based method for monitoring and identifying science and technology trends**

**Abstract**

The task of identifying new S&T trends can be solved in different ways. For the most part expert methods and text mining methods on scientific publications and patents are used to carry out perspectives of science and technology progress. The paper presents the method of valuation public interest in the promising S&T areas in Russian scientific space by means of statistical analysis of trends in web searches of technological terms selected as a result of the experts’ choice in Russian-language search engine Yandex (https://wordstat.yandex.ru).

**Introduction**

The solution of the problem of funding scientific research in resource-limited settings needs to identify and regularly monitor science and technology (S&T) trends, which can have a key impact on social and economic development in the long term. Key science and technology trends should be monitored regularly also because of flexible and timely strategic decision-making in response to technological changes. The results of such studies are in wide demand across a broad spectrum of stakeholders (government, business, research institutes, and the general public) involved in the development and practical use of long-term forecasts (Salmenkaita J.-P. , Salo A., 2002).

The task of identifying new S&T trends can be solved in different ways, using different techniques and different methodological approaches. For the most part, the studies of new S&T trends based on the integration of traditional methods of forecasting, such as expert methods (interviews, surveys, seminars, etc.) and automatic techniques using statistical methods and data mining techniques from texts such as scientific publications and patents (Kim, et al., 2009; Wang et al., 2010).

Foresight [technology] is mainly conducted by applying a combination of qualitative and quantitative methods. An evidence-based approach implies covering a wide range of information sources, as well as the active application of quantitative methods for processing. Therefore, it is very important to select the right sources of data, extract core information from them, and interpret the results correctly. In theoretical works devoted to identifying technology trends, the most widely used information sources are scientific publications and patents. There are also authors who propose relying on additional sources of data (media, conferences, business related

resources, etc.). However, the issue of applicability and comparison of core and extra sources of information for monitoring technology trends has not received sufficient coverage in the literature (Mikova N., Sokolova A., 2014).

So, in connection with this, it was considered the possibility to use the Internet as a possible source of information for identifying of promising S&T trends in Russian scientific space by means of statistical analysis of frequency of search queries of technological terms. In our research we followed the idea that an increase in requests corresponding the surge of public interest in a certain topic. Usually to search for S&T trends using Google or Twitter engines. But Russian-speaking users much more often used as a search engine Yandex than Google (56,6 to 37,0 %, <http://www.liveinternet.ru/stat/ru/searches.html?slice=ru;period=week>). So, in connection of our goal, identifying promising S&T areas in Russian scientific space, we analysed the queries in Yandex. The study based on aggregated Yandex Search query data (https://wordstat.yandex.ru). As a result, we got a few test queries, which may indicate that the proposed screening technology works. This approach implies covering a broad range of information sources, as well as actively applying quantitative processing methods.

**Methodology and main results.**

A growing body of research now shows that Internet search statistics can systematically predict important social and political behavior. So it was conjectured the existence of correlation between the number of search queries in certain subject (S&T terms) and public concernment in getting information about disruptive innovations in scientific challenges.

The idea of using new data type (Wordstat.Yandex) representing search queries statistics and providing the analysis of how this method can be used for identifying S&T in Russian scientific space based on the following provisions. Search engines are the primary means of access to information in the Internet. Statistics requests contain the number of user requests for each keyword for each period of time (up to a week or a month). By analyzing these statistics, you can not only just identify any feature requests are popular at the moment, but also compare the history of changes in these indicators. So you can evaluate the prospects of a trend in public interest using this type of analysis; predict changes in such an assessment in the near future; highlight areas of heightened interest of participants of networks; determine the events that influence the formation of public interest.

Using this method you should know that one of the most challenging issues in Internet query analysis is the selection of diagnostic search terms. Ensuring validity — i.e., that the Internet terms putatively representing a certain attitude or behavior actually represent them—is one major challenge. So we went through several stages of sophisticated maximum likelihood estimations starting with hundreds of search terms, which were then broken down into meaningful diagnostic clusters. And as a result a statistical analysis of trends in web searches of S&T terms was conducted containing the term - *Nanotechnology* and affiliated terms (such as *Fullerenes*,

*Nanotubes*, etc.). It was marked 50 popular search queries containing the word - Nanotechnology, and obtained data on them.

In order to eliminate seasonal factors (seasonal dependence on educational needs) impact on results it was developed and tested an algorithm “with seasonal filtering” requests:

1) For each of the inquiries necessary to obtain data on the frequency of search

engine Yandex.

2) Normalize this distribution for the month of maximum popularity this request.

3) To determine the relative importance of the popularity of the query in a given

month by dividing the maximum value.

4) Determine the "average" frequency curve.

5) Smooth seasonality factor (divide our normalized frequency on average for each

month) to make it possible to identify queries that are allocated to this general trend.

As a result, we got a few test queries and built several charts, which may indicate that the proposed screening technology is working. On the implementation of the sample can be a determined subject area, which is determined by the surge of public interest in this or that trend. Applying the algorithm on the general "broad" queries, we were able to identify the "exceptional" request that suggests in future, with the improvement of the algorithm, the results can be achieved on narrow topics, including scientific terms. In particular, scheduled data analysis and comparison data search trend in recent years.

**References**

Youngho Kim, Yingshi Tian, Yoonjae Jeong, Ryu Jihee, Sung-Hyon Myaeng.

Automatic Discovery of Technology Trends from Patent Text. In: Proc, SAC’09,

March 8-12, 2009, Honolulu, Hawaii, U.S.A., 2009.

Mikova N., Sokolova A. (2014) Global Technology Trends Monitoring: Theoretical

Frameworks and Best Practices. *Foresight-Russia*, vol. 8, no 4, pp. 64–83.

Oral M., Kettani O., Lang P. (1991) A methodology for collective evaluation and

selection of industrial R&D projects // Management Science. Vol. 37. No 7. Р. 871–

885.

Poh K. L., Ang B. W., Bai F. (2001) A comparative analysis of R&D project

evaluation methods // R&D Management. Vol. 31. No 1. P. 63–75.

Salmenkaita J.-P. , Salo A. (2002) Rationales for government intervention in the

commercialisation of new technologies // Technology Analysis and Strategic

Management. Vol. 14. No 2. P. 183–200.

Stewart T.J. (1991) A multi-criteria decision support system for R&D project selection

// Journal of the Operational Research Society. Vol. 42. No 1. P. 17–26.

Wang et al. Identifying technology trends for RD planning using TRIZ and text

mining, RD Management, vol. 40, N 5,2010