NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Nadezhda Mikova, Anna Sokolova

SELECTION OF INFORMATION SOURCES FOR IDENTIFYING TECHNOLOGY TRENDS: A COMPARATIVE ANALYSIS

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: SCIENCE, TECHNOLOGY AND INNOVATION

WP BRP 25/STI/2014

Nadezhda Mikova[1], Anna Sokolova[2]

SELECTION OF INFORMATION SOURCES FOR IDENTIFYING TECHNOLOGY TRENDS: A COMPARATIVE ANALYSIS[3]

Technology foresight 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, and others). 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. In connection with this, the purpose of this paper is to conduct a comparative analysis of the results of technology monitoring by using various information sources (scientific publications, patents, media, foresight-projects, conferences, international projects, dissertations, and presentations). The proposed approach is tested on the area of green energy and the results are described and analyzed. Possible factors that can affect the results of data processing are considered and discussed in order to more efficiently use the comparative analysis of quantitative and qualitative procedures for identifying, correcting, and updating global technology trends on a regular basis.

JEL Classification: O31, O32, O33, O38.

Keywords: technology foresight, technology trends monitoring, evidence-based studies, quantitative methods, information sources, data clustering, green energy

[1] National Research University Higher School of Economics (Russia), Institute of Statistical Studies and Economics of Knowledge, Centre for Statistics and Monitoring of Science, Technology, and Innovation, Research Intern, nmikova@hse.ru

[2] National Research University Higher School of Economics (Russia), Institute of Statistical Studies and Economics of Knowledge, Laboratory for Science and Technology Studies, Senior Research Fellow, avsokolova@hse.ru

[3] The authors would like to thank Segrey P. Filippov for his great expert support of this research.

Introduction

As technological progress speeds up and the time needed for implementing innovation shortens, it is important to identify and regularly monitor global technology 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 because of the need to react quickly to technological changes and to make strategic decisions in a timely manner. The theory and practice of foresight has experienced a high interest in evidence-based forecasting in the past decades. This approach implies covering a broad range of information sources, as well as actively applying quantitative processing methods. Because of information overload, key tasks in identifying technology trends include choosing the right data sources, extracting core information from them, and interpreting it correctly.

The selection of information sources depends on several factors: the purpose of a study, its methodology, available resources, and other parameters of the project. In addition, this choice can be determined by what is understood as “technology trend”. Different authors propose different interpretations of the term and use other notions associated with it (Table 1).

Table 1 – Definitions associated with the notion of a “technology trend”

Term	**Definition**	**Example**	**Source**
Technology applications	These technology trends and applications will be the focus of continued research and development consideration, market forces, and debate. Many of these technologies will be applied in some guise or other, and the effects will be significant and astonishing, changing lives across the globe.	Pervasive sensors	[Silberglitt et al., 2006]
Disruptive technologies (innovations)	Disruptive technologies bring to a market a very different value proposition than had been available previously. Generally, disruptive technologies underperform established products in mainstream markets. But they have other features that a few fringe (and generally new) customers value. Products based on disruptive technologies are typically cheaper, simpler, smaller, and, frequently, more convenient to use.	Personal desktop computers	[Christensen, 1997]
Game-changers	These are specific technologies that are potential “game-changers” in the hands of enemies of the United States. Emerging biotechnologies that may enable functional brain imaging, covert communications, the spread of disabling infections, and sensor spoofing are likely to affect the conduct of military operations and the status of	Gene therapy	[NRC, 2005]

	national security in the future.		
Key items	Items that need a focused approach for the resolution of global and national challenges.	Efficient power storage systems	[NISTEP, 2010]
Emerging technologies	The notion of "emerging technologies" is used here to focus on developments that: (a) result from contemporary advances in a given field of knowledge; (b) are rapidly evolving; and (c) have high potential to result in inventions and innovations with significant societal and economic impacts.	Medical scanners	[Gokhberg et al., 2013]
Key enabling technologies	Key enabling technologies (KETs) are knowledge intensive and associated with high R&D intensity, rapid innovation cycles, high capital expenditure and highly-skilled employment. They enable process, goods and service innovation throughout the economy and are of systemic relevance. They are multidisciplinary, cutting across many technology areas with a trend towards convergence and integration. KETs can assist technology leaders in other fields to capitalise on their research efforts.	Photonics	[European Commission, 2009]
Research fronts	Research fronts represent the most dynamic areas of science and technology and the areas that attract the most scientific interest.	Organic thin-film transistors	[Upham and Small, 2010]
Hot research areas	Hot research areas are characterised by a high level of citation activity.	Post-genome research	[OECD, 2010]
Mega-trends	Megatrends sweep through the industry in the context of an uncertain global economy – leaving an ever-widening gap between winners and losers.	Smart mobility	[Ernst & Young, 2012]
Micro-trends	Cutting-edge ideas and patterns of opportunity in specific thematic areas.	Magnetic chargers for smartphones	[TrendHunter, 2013]

It is not difficult to see that the notions listed in Table 1 do not contradict each other, but only differ in what they put an emphasis on from the point of view of:

- *effects* (technology applications focus on solving socio-economic problems; disruptive innovations focus on the ability to change the structure of markets; game-changers are connected with the ability to undermine the global status of a country; and key technologies focus on a general high impact on the development of business, state, and society);

- *the life-cycle stage* (emerging technologies are at an early stage of development and have a high potential for impact);
- *scale* (mega-trends span a number of generations and are global in scale, while micro-trends are separate innovative ideas that can rapidly increase the competitiveness of businesses);
- *the way in which they can be identified* (research fronts are defined as clusters of documents detected on the basis of co-citation analysis).

If we were to study the differences in emphasis contained in technology trends based on the stage of their life cycle, then we could divide sources of data into five categories [Martino, 2003] (Table 2). For instance, while looking for information about emerging technologies and research fronts, it is advisable to use databases of scientific publications (Science Citation Index); while analysing applied research, databases in engineering sciences (Compendex) can be of help; when scanning for technology applications, then one might consult patent databases (USPTO patent database); when searching for disruptive innovations, news resources (Newspapers Abstracts Daily) are useful; and while analysing key future technologies and evaluating their social and economic impact, the business and popular press can be a reliable resource.

Table 2 – Choosing information sources based on the technology life-cycle stage

	R&D stage	**Typical source**
1	Basic research	Science Citation Index
2	Applied research	Engineering Index (Compendex)
3	Development	USPTO patents database
4	Application	Newspapers Abstracts Daily
5	Social Impacts	Business and popular press

Source: [Martino, 2003].

However, if we were to study technology trends in more general terms, such as topical, cutting-edge, and quickly developing technology areas that can significantly affect the future development of the economy and society, then the correct choice of information sources is not so unambiguous.

In theoretical works that aim to discover global technology trends, the most widely used sources of data are scientific publications [Chen, 2006; Cobo et al., 2011; Daim et al., 2006; Guo et al., 2011; Kajikawa et al., 2008; Kostoff, 2008; Morris et al., 2002; Porter and Cunningham, 2005; Shibata et al., 2008; Smalheiser, 2001; Upham and Small, 2010] and patents [Campbell, 1983; Corrocher, 2003; Daim et al., 2006; Dereli and Durmusoglu, 2005; Fattori et al., 2003; Kim et al., 2008; Kim et al., 2009; Lee et al., 2009; Lee et al., 2011; Li et al., 2009; Porter and Cunningham, 2005; Trappey et al., 2006; Tseng et al., 2007; Wang et al., 2010; Yoon and Park, 2004]. There are also studies that propose relying on other sources of information, such as newspapers [Daim et al., 2006]; business related resources, such as Lexis-Nexis [Porter and Cunningham, 2005] and information about venture capital funds and start-ups [Cozzens et al., 2010]; data from specialized conferences [Porter and Cunningham, 2005]; and others. In addition, there is widespread and increasing interest in the development of indicators for measuring technology trends. Such indicators can be used for monitoring their arrival, development, and subsequent diffusion, as well as their key social and economic impacts [Gokhberg et al., 2013; Meissner and Sokolov, 2013].

However, the issue of applicability and comparison of core and extra sources of information for the purposes of technology monitoring has yet to receive sufficient coverage in the literature. Thus, the purpose of this work is a comparative analysis of the results of identifying technology trends based on various sources of data, using the area of green energy as an example.

1. Research methodology

To achieve this purpose, the following tasks are resolved in this study:

- selecting sources for analysis;
- identifying technology trends for each source using automated data processing tools in the field of green energy;
- expert analysis of the final lists of trends in order to detect relevant trends;
- independent expert research aimed at identifying technology trends in the given subject area;
- comparative analysis of the lists of trends obtained from different information sources and through expert analysis;
- discussion of results.

At the start of our research, we chose the following sources of information for detecting technology trends: scientific publications, patents, media, foresight projects, conferences, European Commission (EC) projects, dissertations, and presentations. Identifying technology trends based on these sources is done using the following methodology:

- *Stage 1:* “Creating a list of keywords”;
- *Stage 2:* “Collecting data”;
- *Stage 3:* “Clustering data using Vantage Point software”;
- *Stage 4:* “Compiling a final list of trends”.

At the first stage, a list of key phrases is created for a definite identification of the subject field, which can include anything from one phrase to combinations of 30 to 50 keywords. The second stage focuses on collecting data from various sources of information, which can be carried out using the following methods (Table 3).

Table 3 – Information sources and methods

№	Collections	Information sources	Methods
1	Scientific publications	ISI Web of Science	Bibliometric analysis
2	Patents	Derwent Innovations Index	Text mining
3	Media	Factiva	Bibliometric analysis
4	Foresight projects	EFMN, EFP	Text mining
5	Conferences	Official websites of conferences	Text mining
6	EC projects	CORDIS Europe	Text mining
7	Dissertations	ProQuest	Bibliometric analysis
8	Presentations	SlideShare	Text mining

The third stage is clustering, which includes the following procedures: 1) importing collections into Vantage Point software; 2) pre-preparing data; and, 3) clustering keywords (factor analysis). Vantage Point software uses customized import filters for uploading collections, which makes it possible to import data from most of the popular electronic databases (Web of Science, Scopus, Derwent Innovations Index, Factiva, etc.). Next, at the stage of pre-preparing data, it is necessary to carry out the following procedures: remove duplicate documents, carry out a linguistic analysis and stemming[4] of the text, and exclude stop-words, which are words that carry no particular meaning on their own, such as prepositions, conjunctions, and pronouns. Clustering and visualising connections between selected keywords is carried out within Vantage Point using factor analysis, which allows us to show only the most important technology clusters on the map (main components). For instance, in order to identify the most promising technology trends, it is necessary to divide the resulting total of keywords into sub-groups and create a factor map. In this case, the main factor for clustering is co-occurrence of terms in the texts of documents. For instance, if two key phrases occur together in a large volume of documents, they will be added to the same cluster. Different clusters can be placed on the map in isolation from each other or be interlinked if they contain the same keywords. Figure 1 shows part of a cluster map.

[4] Many words have the same base (stem), but carry out different semantic functions (for instance, “computation” and “computing”). The process of stemming helps find the stem of several similar words.

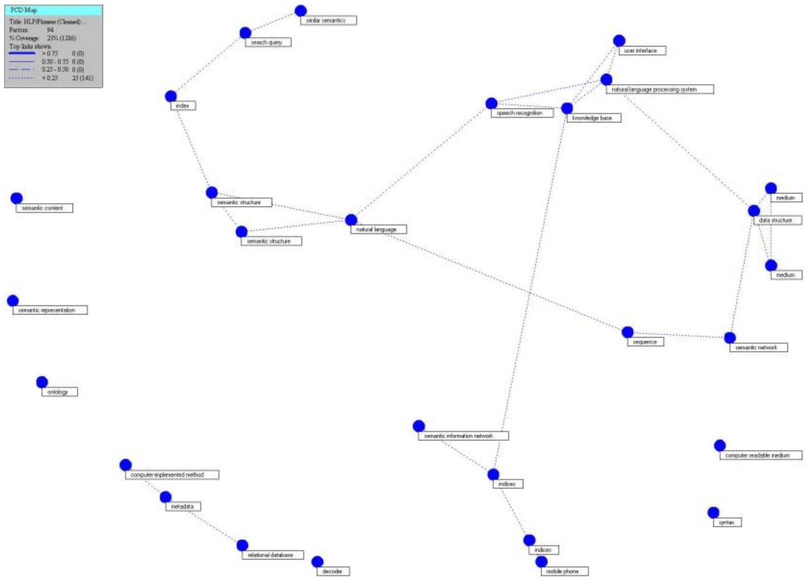

Figure 1: Part of a cluster map generated by Vantage Point

Vantage Point software names the resulting clusters based on the most frequent keywords. Each cluster is characterized by a vector of key phrases (descriptors), which it includes and which are listed on the map in receding order according to the frequency in which they occur. The number of clusters in Vantage Point can be pre-set (for instance, 5 factors, 10 factors, etc.), or be calculated by the software itself, based on the number of documents that are being processed. Forming the final list of technology trends includes a preliminary analysis of clusters, visualisation of links between clusters, and also expert procedures (interviews, surveys, etc.), which are carried out to select relevant technology trends on this basis.

2. Analysis of the possibilities of using different sources for identifying technology trends

Most studies on monitoring technology trends extensively use electronic databases while gathering information. For instance, bibliometric databases of scientific publications (Web of Science, Scopus and others) can be useful from the point of view of tracking research fronts and emerging technologies, while patent databases (USPTO[5], EPO[6], JPO[7] etc.) can be helpful for finding information on specific technology solutions and applications in a particular field. Data in electronic databases can be both structured (title, author, year, etc.) and unstructured (abstracts, full texts of documents). While choosing a database, it may be necessary to consider the characteristics of the database as a whole, ensuring that it meets a set of minimum standards in line with information objectives. Considerations may include suitable coverage,

[5] United States Patent and Trademark Office.
[6] European Patent Office.
[7] Japanese Patent Office.

comprehensiveness of coverage, biases, content quality, record structure, and keyword availability [Porter and Cunningham, 2005].

Choosing the method for collecting data from selected sources is a separate task, which is often resolved by different ways of compiling a list of keywords that define the subject area. For example, these can be general words that describe the subject area, a list of keywords selected through consulting with experts [Lee et al., 2009; Morris et al., 2002], keywords from the most important documents [Kim et al., 2008]; or a combination of these methods [Kim et al., 2008; Porter and Cunningham, 2005]. An alternative search strategy is to compile a list of articles or patents based on a certain criterion, such as articles from specialized journals [Cobo et al., 2011; Guo et al., 2011; Kajikawa et al., 2008; Kostoff et al., 2008], the most highly cited publications [Upham and Small, 2010], patents from specific classes of the International Patent Classification (IPC) [Corrocher et al., 2003; Lee et al., 2011], or National Science Council (NSC) patents [Tseng et al., 2007].

What follows are brief descriptions of information sources that can be used to identify global technology trends.

Databases of scientific publications

Databases of publications that contain peer-reviewed papers from scientific and technical journals are among the most popular information sources used for detecting technology trends. Many of them offer access not only to abstracts, but also to full texts of articles, reviews, conference papers, and so on. Electronic databases of scientific publications can be divided into general (interdisciplinary), such as Web of Science (including Science Citation Index and Social Science Citation Index), Scopus, Pascal, and others; and specialized, such as MEDLINE (medicine), Chem Abstracts (chemistry), and CiteSeerX (information technology). In addition, there are separate engineering databases, like Engineering Village, EI Compendex (Engineering Index) and INSPEC, as well as specialized resources for full-text searching of scientific publications in all formats and disciplines (Scirus and Google Scholar, among others).

As part of technology monitoring, data obtained from scientific publications can be used for:

- studying research fronts [Chen, 2006; Kajikawa et al., 2008; Morris et al., 2002; Shibata et al., 2008; Upham and Small, 2010];
- scanning emerging technologies [Guo et al., 2011; Porter and Cunningham, 2005];
- searching for new technology solutions to existing problems [Kostoff, 2008]
- defining “blank” technology areas that are not yet developed but are potentially significant [Morris et al., 2002];

- tracking the evolution of technology trends [Cobo et al., 2011; Shibata et al., 2008]; and
- forecasting the development of technology [Daim et al., 2006].

Citation analysis is a popular method of processing *structured data* while monitoring scientific publications [Chen, 2006; Kajikawa et al., 2008; Morris et al., 2002; Shibata et al., 2008; Upham and Small, 2010]. Text mining is generally used for processing *unstructured data* of such publications [Chen, 2006; Guo et al., 2011; Kostoff, 2008; Porter and Cunningham, 2005].

Patent databases

Patent analysis plays a significant role in identifying technology trends because inventions that are actually patented point to a scientific and engineering interest in technology in the given subject area. Patent analysis involves collecting data about patent applications stored in relevant databases (such as USPTO, EPO, JPO, WIPO). Certain database providers also offer paid access to generally available patent information – such as from the USPTO database – that is structured in a way that makes its use more efficient (for example, the Derwent Innovations Index, QuestelOrbit (QPat), MicroPatent and others).

As part of technology monitoring, patent data can be used to:

- identify technology trends and emerging technologies [Corrocher et al., 2003; Fattori et al., 2003; Porter and Cunningham, 2005; Trappey et al., 2006];
- search for information on technology problems and solutions in a subject area [Kim et al., 2009];
- analyse patent trends [Campbell, 1983] and monitor patent activity over time [Dereli and Durmusoglu, 2005];
- define the maturity level and life-cycle stage of technologies based on comparing their characteristics with universal evolutionary patterns [Wang et al., 2010];
- visualize data on the development of technology areas [Kim et al., 2008; Yoon and Park, 2004];
- group topics in self-organizing systems of technology classification [Tseng et al., 2007];
- define “blank” but promising technology areas [Lee et al., 2009];
- validate the results of technology monitoring [Kostoff et al., 2008];
- detect evolutionary patterns of technology development [Lee et al., 2011]; and,
- forecast the future development of technology [Daim et al., 2006].

As with the case of scientific publications, citation analysis [Morris et al., 2002] can be used for processing *structured patent data*. In addition to citation information, patent analysis can make use of structured data from other bibliometric fields, such as keywords [Kim et al., 2008], or simply the number of patents [Campbell, 1983; Dereli and Durmusoglu, 2005; Lee et al., 2011]. While analysing patent information, many authors deal with *unstructured data* (abstracts, full texts of patents, etc.) and use text mining for their analysis[8]. This method largely reduces processing unstructured patent information to extracting keywords from the text and then processing them with the help of linguistic and statistical means.

Even though many authors tend to choose one key information source – either scientific publications or patents – while processing data on technology trends, there have also been attempts to use a combination of these two resources. For instance, Shibata et al. [2010] compare the structure of citation networks of publications and patents in order to discover the differences between the two. The main purpose of this work is to identify non-commercialized gaps between science (publications) and technology (patents). However, when combining publications and patent analyses, it is necessary to take account of a number of factors: 1) publications are more likely to reflect academic activity, whereas patents tap industrial R&D efforts; 2) company policies on publishing and patenting vary greatly; 3) the content of publication and patent abstract records are partly comparable, but not identical; and, 4) publication abstracts usually try to explain what the researchers did and why they did it, while patent abstracts often try not to convey corporate intentions [Porter and Cunningham, 2005].

Additional sources (media, foresight projects, conferences, projects of international organizations, dissertations, presentations, web resources, etc.)

The main advantage offered by electronic databases is the fact that they contain large volumes of information in a structured format. In addition to scientific publications and patents, other structured or semi-structured databases can be used to identify technology trends. Examples include databases of *media outlets* (Factiva), *foresight projects* (EFMN), *projects of international organizations and national governments* (CORDIS Europa, NTIS, RaDiUS), *dissertations* (ProQuest), and *presentations* (SlideShare). The Internet is becoming an important repository of science and technology information, most of which is openly available to researchers, engineers, and the public. Internet resources are different from electronic databases, above all in that most online data exist in an unstructured (full-text) format. The best results can be achieved when using a combination of electronic databases and online resources [Porter and

[8] Text mining is a rather new technique that has been proposed to perform knowledge discovery from collections of unstructured text. In short, text mining puts a set of labels on each document and discovery operations are performed on the labels. The usual practice is to put labels to words in the document. Then, the document in text format can be featured by keywords that are extracted through a text mining algorithm [Yoon and Park, 2004].

Cunningham, 2005]. From the point of view of monitoring global technology trends, useful online resources can include materials presented at key *conferences* in the subject area, as well as textual content from relevant *websites*.

Additional sources can be useful in numerous ways. News sources can be helpful for identifying emerging technologies [Porter and Cunningham, 2005]; business resources can be used while searching for the demands of the business community, such as Lexis-Nexis [Porter and Cunningham, 2005] and information on venture capital funds and start-ups [Cozzens et al., 2010]; and conference materials are good for tracking the most promising technology directions [Porter and Cunningham, 2005]. Additional sources of information – for example, specialized web resources – can be useful when scanning for risks and perspective technologies in the subject areas of knowledge [Palomino et al., 2013].

Web-scanning, or web-scraping, can be used for analysing *unstructured online information*. This method allows one to mine large volumes of textual information because it ensures a constant inflow of additional data from websites. For example, Lucene[9], a full-text information retrieval tool, and Nutch[10], an open-source search engine built on it, can be used as a basis for creating a web collection. These well-known and well-documented open-source software products implement basic search infrastructure: reverse indexing, search bots, parsers for different types of documents (HTML, PDF and others), and a user-friendly web interface. In addition, web-scraping can make use of meta-search systems, which consecutively compare results obtained from different sources (for example, credit card fees, insurance premiums, real estate prices, etc.). The method of web-scraping, which is the constant monitoring of news, markets, and so forth, makes it possible to collect data for further technology analysis and evaluations of markets, competitiveness, etc.

3. Compiling the trends lists based on different sources: an example of green energy

The following parameters were used in most of eight electronic databases to collect and filter related documents:

- Query: green energy;
- Timespan: from 2002 to 2012;
- Language: English.

[9] http://lucene.apache.org.
[10] http://nutch.apache.org.

Table 4 shows aggregate data that represent collections created for the subject area of green energy: data sources, the number of documents, the names of biometric fields that are being processed, and also the format of collections before and after pre-preparation of data.

Table 4 – Aggregated data on collections formed for the subject area of green energy

№	**Name**	**Data sources**	**Number of documents**	**Field for processing**	**Format before processing**	**Format after processing**
1	Scientific publications	Web of Science	18,912	keywords	*.txt	*.txt
2	Patents	Derwent Innovations Index	5,000	abstract	*.txt	*.txt
3	Media	Factiva	44,310	title	*.html	*.smart XML
4	Foresight projects	EFMN, EFP	40	text	*.doc	*.smart XML
5	Conferences	Official websites	19	text	*.doc or *.pdf	*.smart XML
6	EC projects	CORDIS Europe	129	text	*.doc	*.smart XML
7	Dissertations	ProQuest	1,189	keywords	*.html	*.smart XML
8	Presentations	SlideShare	100	text	*.ppt	*.smart XML

A total of 69,699 documents in different formats (*.txt, *.html, *.doc, *.pdf, *.ppt) were collected, most of which were converted into *.smart XML format for further processing in Vantage Point software.

While clustering documents in Vantage Point, the number of clusters (factors) can be set by the user. Therefore, this study analysed links between clusters that were created by dividing data in each collection into 5, 10, 15, and 20 factors. Links between clusters are visualized in order to single out families of trends that have similar content in common. Figure 2 is a visual map of links between clusters (5, 10, 15, and 20 factors), which were obtained as a result of processing the collection “Conferences”. Clusters shown in one colour contain similar key phrases and form “horizontal” families of trends.

Figure 2 – Map of links with a different number of clusters
for the collection “Conferences”

Therefore, while choosing a small number of factors (e.g. 5 factors), clusters represent large numbers of keywords that reflect major technology trends in the field of green energy. If the number of factors is increased (to 10, 15, or 20 factors), these technology areas are subdivided into smaller areas (subareas), creating families of trends. Furthermore, their further subdivision produces clusters that contain more detailed information about technology applications, solutions, and products (for example, production of sun-powered dryers, production of asynchronous double-winding electric generators, creating energy-efficient heating, ventilation, and air conditioning, and developing wind-energy generators). Most clusters, which were identified by dividing the collection into 20 factors, are linked to clusters that were obtained by dividing the collection into 5 factors, but the map also shows “stand-alone” clusters (e.g., using microbiology to solve environmental problems, the need to monitor the efficiency of various power generators based on meteorological and other data, the development of space-based solar power, developing hydropower energy, etc.). Therefore, if the main task of technology monitoring is to single out 3 to 5 major research directions in the subject area, then it is advisable to conduct analysis on the basis of 5 factors. If, however, the focus is on finding smaller areas (subareas) or specific technology applications, then an analysis based on a division into 20 factors may be more appropriate. Similar to the collection “Conferences”, maps of clusters were created for the remaining seven collections in the field of green energy.

4. Comparing the results of expert and analytical procedures

After analysing visual maps, final lists of clusters (potential trends) were compiled. An analytical structure of trends in green energy that was created on this basis represents all collections of data (Table 5).

Table 5 – An analytical structure of trends in the field of green energy

№	Groups of trends	Collections represented[11]
1	Solar energy	(1,2,3,5,6,8)
2	Wind energy	(5,8)
3	Hydropower	(5,8)
4	Bioenergy and waste processing	(1,2,3,4,5,7,8)
5	Hydrogen economy	(1,2,4,5)
6	Space-based solar power	(5)
7	Chemical engineering and nanotechnology	(1,2,6,7)
8	Power electronics	(8)
9	Green information technology	(3,5,6,7,8)
10	Microbiology	(5,6,8)
11	Transport	(1,3,4,5,6,7,8)
12	Green manufacturing	(8)
13	Green construction and design	(3,5,6,7,8)
14	Green supply and logistics systems	(5,6)
15	Reducing CO_2 emissions	(3,4)
16	Complex monitoring and modelling	(5,7,8)
17	Security	(7)
18	Green management	(7)
19	Green business	(3,5,8)
20	Education and the labour market	(4,5,8)
21	Food services	(7)
22	The EU's efforts to develop green technology	(6)
23	International standards and licensing	(3,4,5,8)

It follows that the resulting clusters represent both technology areas and social and economic challenges related to green energy. All of the lists of clusters were then validated by experts in this field. The validated clusters were divided by experts into *direct* clusters, which are directly connected to the subject area of green energy, and *adjacent* clusters, which are connected to other areas (Table 6).

Table 6 – Aggregated data on the selected clusters in the area of green energy (based on dividing collections into 20 factors)

№	Name	Direct clusters	Adjacent clusters	Total

[11] Collections (numbers):
(1) Scientific publications
(2) Patents
(3) Media
(4) Foresight projects
(5) Conferences
(6) European Commission projects
(7) Dissertations
(8) Presentations

1	Academic papers	2	0	2
2	Patents	3	1	4
3	Media	7	4	11
4	Foresight projects	4	1	5
5	Conferences	7	5	12
6	EC projects	4	5	9
7	Dissertations	3	3	6
8	Presentations	6	2	8
	Total:	36 (63%)	21 (37%)	57 (100%)

It becomes clear that most of the clusters selected as a result of expert procedures were classified by experts as direct trends (63% from the total), while 37% of the clusters were classified as adjacent trends. The most informative collections were “Media” and “Conferences”. On the other hand, an *expert list of trends* was compiled as a result of expert procedures (Table 7).

Table 7 – Expert list of trends in the area of green energy

№	**Subarea**	**Trends**
1	Solar energy	Solar cells
2		Solar thermal power stations
3		Solar thermal collectors
4	Wind power	On-shore wind farms
5		Off-shore wind farms
6	Bioenergy	Biogas production
7		Biodiesel production
8		Bioethanol production
9	Hydropower	Large hydropower stations
10		Small hydropower stations
11	Geothermal energy	Geothermal heat pumps
12		Geothermal power stations
13		Hot dry rock geothermal power stations
14	Energy storage	Electrochemical cells
15		Pumped-storage power plants
16	Fuel cells	Proton exchange membrane fuel cells (PEMFC)
17		Solid oxide fuel cells (SOFC)
18		Molten carbonate fuel cell (MCFC)
19	Hydrogen energy	Electrolysis of water
20		Hydrogen storage

After analysing, comparing, and integrating the expert and analytical lists with one another, a table of correlations was created (Table 8).

Table 8 – Table of correlations between the expert and analytical lists of trends

№	**Trends from the expert list**	**Number of corresponding trends**	**Represented collections[12]**

[12] Collections (numbers):
(1) Academic papers
(2) Patents
(3) Media

		from the analytical list	
1	Solar cells	8	(1,2,3,7,8)
2	Solar thermal power stations	3	(8)
3	Solar thermal collectors	1	(5)
4	On-shore wind farms	3	(5,8)
5	Off-shore wind farms	0	-
6	Biogas production	1	(8)
7	Biodiesel production	8	(2,3,4,5,6,7,8)
8	Bioethanol production	5	(7,8)
9	Large hydropower stations	1	(8)
10	Small hydropower stations	1	(8)
11	Geothermal heat pumps	1	(8)
12	Geothermal power stations	1	(8)
13	Hot dry rock geothermal power stations	0	-
14	Electrochemical cells	5	(3,5,6)
15	Pumped-storage power plants	1	(8)
16	Proton exchange membrane fuel cells (PEMFC)	1	(6)
17	Solid oxide fuel cells (SOFC)	1	(6)
18	Molten carbonate fuel cell (MCFC)	1	(6)
19	Electrolysis of water	3	(1,2,4)
20	Hydrogen storage	0	-
	Total:	45	(1,2,3,4,5,6,7,8)

Therefore, as a result of comparing the analytical and expert lists, it follows that the clusters (trends) that were discovered using quantitative methods were reflected in 17 out of the 20 trends from the expert list (85%). As part of our research, we analysed those clusters from the analytical lists that are not part of any trends in the expert list. These clusters can be divided into the following groups:

1. Technical devices, technologies, and models.

This group includes smaller-scale technology applications (at the level of sub-trends), specific technologies, and models from green energy and adjacent fields, which can be used to study the process of energy transformation in different systems. Examples of technical devices, technologies, and models include the transmission electron microscope, self-propagating high-temperature synthesis, X-ray crystallography, Fourier infrared spectroscopy, the use of the Langmuir model (adsorption model) in an isothermal process, laser lithography, and others. Clusters from this group are most widely represented in collections “Scientific publications”, “Patents”, and “Dissertations”.

(4) Foresight projects
(5) Conferences
(6) European Commission projects
(7) Dissertations
(8) Presentations

2. Green energy-related technology areas from the wider area of energy (energy efficiency and energy saving), as well as other related areas (such as the rational use of nature, nanotechnology, and transport systems).

The fact that the analytical list includes not only clusters from the area of green energy and the wider area of energy, but also clusters from related disciplines can point to strong links between different areas of technology. For instance, trends from the area of green energy can be linked to technologies in the area of energy efficiency and energy saving, such as new sources of light (including light-emitting diodes) and smart lighting systems, green nuclear energy technology, and so forth. Examples of areas that are developing between adjacent disciplines include interactive mapping technologies for connecting manufacturers and potential consumers of organic waste in order to produce biogas (green energy and the rational use of nature), using environmentally-friendly heavy-duty plastics (green energy and nanotechnology), producing hydrogen fuel (green energy and transport), using environmentally-friendly multipurpose devices (green energy and information and communications technology), biodiesel production technology (green energy and biotechnology), and others. Clusters from adjacent areas occur most frequently in the collections "Media", "Conferences", "European Commission Projects", "Dissertations", and "Presentations".

3. Social, economic, and political challenges.

This group includes analytical clusters that involve information about global problems (social, economic, political, etc.), whose technology solutions can be found in the area of green energy and its related fields. Examples of such social and economic trends include the need to make life in the city more comfortable by using green technology, developing green-energy education and making it more accessible in different parts of the world, the need to change society's attitude towards green technology (attention to the environment, using new materials, etc.) with the purpose of overcoming social barriers hampering the development of green energy, and the fight waged by social groups for protecting and improving air quality. Social and economic clusters are reflected in the collections "Foresight Projects", "European Commission Projects", and "Presentations".

4. Irrelevant clusters.

When analysing the analytical list of clusters across all of the collections, a small number of clusters (around 2.5%) were discovered that are neither directly nor indirectly related to the area of green energy. Examples of irrelevant clusters include the increased attention to weight loss and low-calorie foods (such as green tea); developing ways of diagnosing and treating Alzheimer's disease, Parkinson's disease, cancer, diabetes, atherosclerosis, arthritis, asthma, and other diseases; the increasing popularity of organic food (for example, chia seeds); and others.

Such clusters were discovered while processing the collections “Scientific publications”, “Patents”, and “Media”.

Therefore, information which was retrieved as a result of studying those analytical clusters that were not included in the list of expert trends can be used for further improving the methodology for monitoring global technology trends, studying the specificity of separate collections, and analysing conditions for selecting the best information sources depending on the subject area and life-cycle stage of the technology in question, as well as for updating the list of trends in the future. After analysing these results, it can be concluded that such collections as “Scientific Publications”, “Patents”, and “Dissertations”, which contain the most information about specific technical devices, technologies, and models, can be effective from the point of view of analysing fundamental and applied research. The following collections can be most suited for discovering areas of technology application: “Media”, which contains information about the most-reported areas of application; “Conferences”, which includes potentially successful commercial solutions; “European Commission projects”, which reflects the contents of projects aimed at developing interdisciplinary areas; and “Presentations”, which includes a description of the general technology landscape and links between technologies. While describing the social and economic impact of technology development, the following collections can be helpful: “Foresight Projects”, which contains information about social, economic, and political challenges; “European Commission projects”, which includes a description of political initiatives aimed at resolving global problems; and “Presentations”, which reflects the most topical problems of technology development.

Discussion and conclusion

By comparing the results of the analytical (quantitative) and expert (qualitative) approaches to identifying global technology trends in the area of green energy, quantitative results were reflected in 85% of the trends from the expert list. The opportunities offered and the results obtained by using any of the eight represented collections can depend on the following:

- the specificity of a subject area (for example, political factors can play an important role in technology development in the area of green energy, and therefore a large portion of useful information can be contained in projects and programmes for developing the energy industry in individual countries and also across the globe);

- the life-cycle stage of the technology in question (the collections “Scientific publications”, “Patents”, and “Dissertations” can be used to analyse fundamental and applied research; “Media”, “Conferences”, and “Presentations” can be helpful in discovering areas of applications; “Foresight projects”, “European Commission projects”, and “Presentations” can be suited for evaluating the social and economic impact of technology development;

- selection of information resources (for example, while using electronic databases, it is necessary to take into account the time required to reflect information in a structured format; whole online content analysis makes it possible to process information that is being obtained online);

- search strategy (general search requests may be used while studying social and economic reasons for the development of technology, but a search using a keywords list or a selection of more relevant documents can be more effective while analysing fundamental and applied studies);

- selection of fields to be processed (information from some fields, such as the title or abstract, may be enough to search for general technological trends, but processing full texts of documents may be required for detecting specific technology solutions, devices, and applications).

As a result of comparing expert (qualitative) and analytical (quantitative) approaches towards identifying global technology trends in the subject area of green energy, and in order to make sure that the goals that were set match the results of research, we propose that one account for the following important factors while collecting and processing data:

1. A detailed discussion with key experts that formalises the scope of the subject area before quantitative analysis is started is an important factor that makes the results more efficient. This formalisation should result in the creation of a list of keywords that unambiguously determine the subject area under consideration.

2. In some cases, collecting data using the full expert list of keywords that describe the subject area can be more efficient for creating more consistent collections, rather than collecting data using general search requests (such as “green energy”).

3. Collecting data from specialized journals on relevant topics or from appropriate classes of selected patent classification may be useful for forming collections for a narrowly focused subject area.

4. The presence of adjacent trends in the analytical list (37% of the total number of selected clusters) can point to technological links between the subject area (green energy) and other areas, such as the rational use of nature, nanotechnology, transport systems, information and communications technologies, and so on. The share of discovered adjacent trends can point to how interdisciplinary the area of technology in question is. In the future, information about technology links may be used for detecting convergence trends, which develop where disciplines overlap, as well as for describing the resulting technology trends.

5. After comparing expert and analytical lists, it is possible to single out clusters (trends) that do not overlap. These trends can be of significant interest from the point of view of technology monitoring. A detailed analysis of such clusters can make it possible to identify specific technology solutions (sub-trends) and links between one subject area and other disciplines, and can also reveal social and economic issues that bring about the need for the development of relevant areas of technology.

6. An important step in interpreting the results of data processing is a division of technology trends (supply side) and socio-economic trends (demand side). In order to identify them, different collections of information can be used, which are formed based on specific lists of keywords. For instance, collections that are more suitable for searching for socio-economic trends (demand side) can be “Foresight projects”, “European Commission projects”, and “Presentations”. For discovering technology trends (supply side), “Scientific publications”, “Patents”, “Media”, “Conferences”, and “Dissertations” can be useful.

Overall, it can be concluded that quantitative methods of data analysis using software (such as Vantage Point) do considerably simplify the process of identifying global technology trends. However, practical experience shows that using such methods on their own is not always the most effective approach. In order to make the results of technology monitoring as relevant as possible, it is suggested to use an analytical approach for detecting trends combined with expert procedures. Also, account should be taken of technology areas that are reflected in one list of trends, but not the other. Comprehensive coverage of information sources makes it possible to discover emerging technologies at the earliest stage of their development, for example, using the latest news published online and materials from recent conferences. In addition, the analytical approach is particularly important because it allows for the fully automated collecting, processing, visualization, and interpreting of data. Therefore, provided that the factors which affect the results of data processing are taken into account as much as possible, a comparative analysis of the results of both analytical and expert procedures can be used as an important tool for identifying, correcting, and updating information about global technology trends on a regular basis.

References

Campbell R.S. (1983) Patent Trends as a Technological Forecasting Tool, World Patent Information, 5(3), pp. 137-143.

Chen Ch. (2006) CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature, Journal of the American Society for Information Science & Technology, 57(3), pp. 359-377.

Christensen C.M. (1997) The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail, Harvard Business School Press.

Cobo M.J., Lopez-Herrera A.G., Herrera-Viedma E., Herrera F. (2011) An Approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field, Journal for Informetrics, 5(1), pp. 146-166.

Corrocher N., Malerba F., Montobbio F. (2003) The emergence of new technologies in the ICT field: main actors, geographical distribution and knowledge sources, TENIA project.

Cozzens S., Gatchair S., Kang J., Kim K.-S., Lee H.J., Ordóñez G., Porter A. (2010) Emerging technologies: quantitative identification and measurement, Technology Analysis and Strategic Management, 22(3), pp. 361-376.

Daim T.U., Rueda G., Martin H., Gerdsri P. (2006) Forecasting Emerging Technologies: Use of Bibliometrics and Patent Analysis, Technological Forecasting and Social Change, 73(8), pp. 981-1012.

Dereli T., Durmusolgu A. (2009) A trend-based patent alert system for technology watch, Journal of Scientific and Industrial Research, 68(8), pp. 674-679.

Ernst & Young (2012) View from the top: global technology trends and performance.

European Commission (2009) Preparing for Our Future: Developing a Common Strategy for Key Enabling Technologies. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussels,COM (2009), 512/3.

Fattori M., Pedrazzi G., Turra R. (2003) Text mining applied to patent mapping: a practical business case, World Patent Information, 25, pp. 335-342.

Gokhberg L., Fursov K., Miles I.D., Perani G. (2013) Developing and Using Indicators of Emerging and Enabling Technologies, in: Handbook of Innovation Indicators and Measurement, Cheltenham, pp. 349-380.

Guo H., Weingart S., Borner K. (2011) Mixed-indicators model for identifying emerging research areas, Scientometrics, 89(1), pp. 421-435.

Kajikawa Y., Yoshikawa J., Takeda Y., Matsushima K. (2008) Tracking emerging technologies in energy research: toward a roadmap for sustainable energy, Technological Forecasting and Social Change, 75(6), pp. 771–782.

Kim Y.G., Suh J.H., Park S.C. (2008) Visualization of patent analysis for emerging technology, Expert Systems with Applications, 34(3), pp. 1804-1812.

Kim Y., Jeong Y., Jihee R., Myaeng S.-H. (2009) Automatic discovery of technology trends from patent text, Proceedings of the 2009 ACM symposium on Applied Computing, pp. 1480-1487.

Kostoff R.N., Briggs M.B., Solka J.L., Rushenberg R.L. (2008) Literature-related discovery (LRD): Methodology, Technological Forecasting and Social Change, 75(2), pp. 186-202.

Lee H., Lee S., Yoon B. (2011) Technology clustering based on evolutionary patterns: the case of information and communications technologies, Technological Forecasting and Social Change, 78(6), pp. 953-967.

Lee S., Yoon B., Park Y. (2009) An approach to discovering new technology opportunities: Keyword-based patent map approach, Technovation, 29(6-7), pp. 481-497.

Li Y.-R., Wang L.-H., Hong Ch.-F. (2009) Extracting the significant-rare keywords for patent analysis, Expert Systems with Applications, 36(3), pp. 5200-5204.

Martino J. (2003) A review of selected recent advances in technological forecasting, Technological Forecasting and Social Change, 70(8), pp. 719–733.

Meissner D., Sokolov A. (2013) Foresight and Science, Technology and Innovation Indicators , in: Handbook of Innovation Indicators and Measurement, Cheltenham, pp. 381-402.

Morris S., DeYong C., Wu Z., Salman S., Yemenu D. (2002) DIVA: a visualization system for exploring document databases for technology forecasting, Computers and Industrial Engineering, 43(4), pp. 841-862.

NISTEP (2010) The 9th Science and Technology Foresight, National Institute of Science and Technology Policy, NISTEP report №140 “The 9th Delphi Survey”, March 2010, available at: http://www.nistep.go.jp/achiev/sum/eng/rep140e/pdf/rep140se.pdf.

NRC (2005) Avoiding Surprise in an Era of Global Technology Advances, National Academy of Sciences.

OECD (2010) Measuring Innovation: A New Perspective, Paris.

Palomino M.A., Vincenti A., Owen R. (2013) Optimising web-based information retrieval methods for horizon scanning, Foresight, 15(3), pp. 159-176.

Porter A.L., Cunningham S.W. (2005) Tech mining: Exploiting new technologies for competitive advantage, John Wiley & Sons, Inc.

Shibata N., Kajikawa Y., Sakata I. (2008) Detecting emerging research fronts based on topological measures in citation networks of scientific publications, Technovation, 28(11), pp. 758-775.

Shibata N., Kajikawa Y., Sakata I. (2010) Extracting the commercialization gap between science and technology – case study of a solar cell, Technological Forecasting and Social Change, 77(7), pp. 1147-1155.

Silberglitt, R., Antón, Ph.S., Howell, D.R., Wong, A. (2006). The Global Technology Revolution 2020, In-depth Analysis: Bio/Nano/Materials/Information Trends, Drivers, Barriers and Social Applications.

Smalheiser N.R. (2001) Predicting emerging technologies with the aid of text-based data mining: the micro approach, Technovation, 21(10), pp. 689-693.

Trappey A.J.C., Hsu F.-Ch., Trappey Ch.V., Lin Ch.-I. (2006) Development of a patent document classification and search platform using a back-propagation network, Expert Systems with Applications, 31(4), pp. 755-765.

TrendHunter (2013) TrendHunter database – http://www.trendhunter.com.

Tseng Y.-H., Lin C.-J., Lin Y.-I. (2007) Text mining techniques for patent analysis, Information Processing and Management, 43(5), pp. 1216-1247.

Upham S.P., Small H. (2010) Emerging research fronts in science and technology: patterns of new knowledge development, Scientometrics, 83(1), pp. 15-38.

Wang M.-Y., Chang D.-S., Kao C.-H. (2010) Identifying technology trends for R&D planning using TRIZ and text mining, R&D Management, 40(5), pp. 491-509.

Yoon B., Park Y. (2004) A text-mining-based patent network: Analytical tool for high-technology trend, Journal of High Technology Management Research, 15(1), pp. 37-50.

Nadezhda Mikova
National Research University Higher School of Economics (Russia), Institute of Statistical Studies and Economics of Knowledge, Centre for Statistics and Monitoring of Science, Technology and Innovation, Research Intern, nmikova@hse.ru

Anna Sokolova
National Research University Higher School of Economics (Russia), Institute of Statistical Studies and Economics of Knowledge, Laboratory for Science and Technology Studies, Senior Research Fellow, avsokolova@hse.ru