NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

*Leonid Gokhberg, Ilya Kuzminov, Pavel Bakhtin,
Elena Tochilina, Alexander Chulok, Anton Timofeev,
Alina Lavrinenko*

BIG-DATA-AUGMENTED APPROACH TO EMERGING TECHNOLOGIES IDENTIFICATION: CASE OF AGRICULTURE AND FOOD SECTOR

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: SCIENCE, TECHNOLOGY AND INNOVATION

WP BRP 76/STI/2017

Leonid Gokhberg[1], Ilya Kuzminov[2], Pavel Bakhtin,[3] Elena Tochilina[4], Alexander Chulok[5], Anton Timofeev[6], Alina Lavrinenko[7]

BIG-DATA-AUGMENTED APPROACH TO EMERGING TECHNOLOGIES IDENTIFICATION: CASE OF AGRICULTURE AND FOOD SECTOR[8]

The paper discloses a new approach to emerging technologies identification, which strongly relies on capacity of big data analysis, namely text mining augmented by syntactic analysis techniques. It discusses the wide context of the task of identifying emerging technologies in a systemic and timely manner, including its place in the methodology of foresight and future-oriented technology analysis, its use in horizon scanning exercises, as well as its relation to the field of technology landscape mapping and tech mining. The concepts of technology, emerging technology, disruptive technology and other related terms are assessed from the semantic point of view. Existing approaches to technology identification and technology landscape mapping (in wide sense, including entity linking and ontology-building for the purposes of effective STI policy) are discussed, and shortcomings of currently available studies on emerging technologies in agriculture and food sector (A&F) are analyzed. The opportunities of the new big-data-augmented methodology are shown in comparison to existing results, both globally and in Russia. As one of the practical results of the study, the integrated ontology of currently emerging technologies in A&F sector is introduced. The directions and possible criteria of further enhancement and refinement of proposed methodology are contemplated, with special attention to use of bigger volumes of data, machine learning and ontology-mining / entity linking techniques for the maximum possible automation of the analytical work in the discussed field. The practical implication of the new approach in terms of its effectiveness and efficiency for evidence-based STI policy and corporate strategic planning are shortly summed up as well.

Keywords: Emerging technologies, foresight, strategic planning, STI policy, Russian Federation, agriculture, food sector, text mining, tech mining, STI landscape mapping, horizon scanning

JEL: C55, O1, O3

[1] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: lgokhberg@hse.ru
[2] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: ikuzminov@hse.ru
[3] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: pbakhtin@hse.ru
[4] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: etochilina@hse.ru
[5] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: achulok@hse.ru
[6] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: Aatimofeev@hse.ru
[7] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: alavrinenko@hse.ru
[8] The Working Paper was prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) and supported within the framework of the subsidy granted to the HSE by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

Introduction and literature review

Global development relies more and more on the introduction of new technologies and refinement of the existing ones (*Nelson & Winter, 1982*). The impact of labor and capital among the productivity factors is becoming less dominant, whereas the attention is shifting towards the significance of multi-factor productivity (*Arnaud B. et al., 2011*). The latter, in turn, consists largely of systemic effects of science and technology (S&T) development (*Guellec & van Pottelsberghe De la Potterie, 2001*). One of the main results of S&T development are technologies, meaning, the codified and applicable pieces of knowledge that can be used for production and distribution of goods and services, other purposeful economic and non-economic, but socially impactful activities (*Oxford English Dictionary*). A significant layer of scientific literature deals with the semantics of the word "technology" and finding distinction between processes, technologies, products, markets and other closely connected concepts (*MacKenzie & Wajcman, 1999; Schatzberg, 2006; Brian, 2009*). However, the problem of conceptual distinguishing of technologies lies beyond the scope of the paper.

In conditions of global challenges for sustainable development and attempts to reduce global threats driven by complex issues (such as climate change, ageing population, natural resource scarcity, water security, human health and wellbeing) (*Kallhauge et al., 2005; Keenan et al., 2012*), the global and national governance systems are faced with extremely difficult missions. The solutions to the existing global challenges, as many researchers and international think tanks see it, lie in wide-scale adoption of new technologies (*Omenn, 2006*). At the same time many organizational innovations, such as new business models, citizen vigilance schemes, and governance mechanisms, which could possibly alleviate some of the global issues, are also becoming feasible solely because of development of some universal, or platform, or enabling technologies (*Gokhberg et al., 2013*).

Therefore, governance and management systems have to acquire technology-awareness capabilities (*Spitsberg et al., 2013; Momeni & Rost, 2016; Bildosola et al. 2017*). This means that effective S&T and innovation policy becomes more and more crucial success factor for governance both on global and national levels, as well as for corporate strategic management. For S&T and innovation policies to become effective they need to be both evidence-based (*Smith & Haux, 2017*) and proactive (*Mani, 2004a; Aghion & Griffith, 2008*). This, in turn, formulates the necessity of constant monitoring of *the technologies that are emerging*, as they are important drivers of efficiency of the human activity. One of the purposes of such activity is an early-on seed support of socially-oriented high-tech initiatives[9] (*Parahina et al., 2014; Komatsu et al., 2016*) and new promising businesses[10] (*Vishnevskiy & Yaroslavtsev, 2017*) for them to be able to get over the "valley of death" (*Wessner, 2005*). Another case is preemptive building of necessary enabling infrastructures (*Candorin, Klofsten & Johansson, 2016*), or proactive regulatory initiatives protecting the vulnerable strata of population from potentially disruptive effects of some technology-driven innovations (*Brocker, Dohse & Soltwedel, 2012*).

One of the examples of direct implementation of the concept of emerging technology into the governance practice is the FET (Future and Emerging Technologies) grant mechanism of the EU[11]. Important role in fostering the development of emerging technologies lies with the large institutional granting authorities such as NIH[12] and DARPA in the United States, RFFI[13] and RNF[14] in Russia, etc. Others include initiatives, less directly referencing to the "emerging" attribute and thus dealing with some of the established technologies. Examples include critical

---

[9] In Russian case: Agency for Strategic Initiatives/ Social Innovation Center. URL: https://asi.ru/eng/social/business/ (date last accessed 23.11.17).

[10] In Russian case: Agency for Strategic Initiatives/ National Technology Initiative. URL: https://asi.ru/eng/nti/ (date last accessed 23.11.17).

[11] URL: http://ec.europa.eu/programmes/horizon2020/en/h2020-section/future-and-emerging-technologies (date last accessed 23.11.17).

[12] For example, URL: https://grants.nih.gov/grants/guide/rfa-files/RFA-CA-14-006.html (date last accessed 23.11.17).

[13] URL: http://www.rfbr.ru/rffi/eng (date last accessed 23.11.17).

[14] URL: http://rscf.ru/en/ (date last accessed 23.11.17).

materials and critical technologies processes in Russia (*Sokolov, 2011*), national technology initiatives (United States[15], Russia[16]), technology development stimulating activities in Japan, South Korea, China, India, South Africa, Brazil (*Kojima et al., 2012; Park & Leydesdorff, 2010; Fan & Watanabe, 2006; Mani, 2004b; Van Zyl, 2011; Viotti, 2002*).

The concept of emerging technology, its scope and definition is a highly discussed topic in social sciences. While a number of impactful publications concentrate on conceptualizing the term "emerging technology" (*Rotolo, Hicks & Martin, 2015; Halaweh, 2013*), it is sufficient for the scope of this paper to understand the emerging technology as a new technology that might have a significant impact on the economic activity in certain sectors of the economy. The closely interconnected, and fuzzily (or non-hierarchically) interrelated fields of technology foresight (*Lucheng et al., 2010*), future-oriented technology analysis (*Joung & Kim, 2017*), STI horizon scanning (*Furukawa et al., 2015*), trend spotting (*Jermann et al., 2015*) and other deal with the challenge of identifying and mapping the emerging technologies.

The atomicity of the technology is an adjustable parameter. For instance, each branded industrial equipment system, with its inherent know-hows and unique strengths can be seen as a technology, and there are tens of thousands of such technologies in any industry, or even narrow sub-industries. However, each industrial equipment item can be further decomposed by functional parts constituting it, and each part can be seen as both technology and the result of a number of production processes being technologies themselves. Such level of detail, as a rule, is not needed for global or national governance. At the same time, a distribute understanding by the expert community of the building blocks of each technology grouping, strengths, limitations and other parameters of each narrow, indivisible technology is needed for the final high-level decisions (*Shen et al., 2010*). Therefore, ontology building is required with ontology being a hierarchical taxonomy with many-to-many relations (*Tsui et al., 2010*).

Further step is entity linking (*Rao, McNamee & Dredze, 2013*), for instance, attributing each technology and each technology grouping on all hierarchical levels the centers of competence, intellectual property rights limitations, technical prerequisite (such as needed materials), applications, potential effects on environments, economy, and society, etc. (*Russia 2030: Scientific and Technology Foresight*).

In the era of explosive growth of diversity of S&T (*Klein et al., 2012*) and of quantity of available information (*Hidalgo, 2015*), technology identification and mapping becomes less and less feasible without the use of modern data science techniques. This necessity is caused not only by the "information push", but also by significant drawbacks of human-performed analytics caused by a number of basic biological and psychological limitations (*Tversky & Kahneman, 1975; Simon, 1982, etc.*). Reconciliation of ideas among the large expert groups could lead to overextended periods of foresight – one of the main sources of evidence for modern STI policy. This mean that some technologies might already transform from the emerging stage to the stage of commercially viable products before a dedicated foresight report on the emerging technology landscape is published. All these factors along with budget limitations drive governments and private companies towards at least partial automation of foresight and strategic planning activities.

As the example of early attempts at solving this problem, the tech mining (*Porter, 2004; Madani, 2015; Huang et al., 2015; Bakhtin & Saritas, 2016*). However, the technical basis powering the tech mining is not sufficiently scalable and strongly relies on large expert validation and manual filtering and cleaning of data outputs. Therefore, it is locked to the field of narrow sectoral case studies where analysis of hundreds of documents is a sufficient data sample.

Another attempt in this field is creating and regularly updating the ontologies specific for future-oriented studies, such as, for example, ontology of weak signals (*Popper, 2010*). They allow further automatic structural analysis and comparison. The problem with such approaches is the necessity to manually or, in best case, semi-manually enrich the dedicated database, which

[15] URL: https://www.nano.gov/ (date last accessed 23.11.17).
[16] Agency for Strategic Initiatives/ National Technology Initiative. URL: https://asi.ru/eng/nti/ (date last accessed 23.11.17).

makes the approach prone to subjectivity, human errors, and bad scalability. Therefore, such specific ontologies rarely include more than several hundreds of entities (such as weak signals, trends, or technologies). They also tend to marginalize becoming a skewed and outdated subsets of latent ontologies conjured by live collective information-generation processes, such as the entirety of newsfeeds (*Yoon, 2012*), global blogosphere (*Melville, Gryc, Lawrence, 2009*) or social networks (*Simon & Leker, 2016*). Clearly, more generalizable, scalable, human-independent, and big-data-oriented approaches and models are necessary in this field.

Fortunately, technology identification and mapping can be made less human-dependent due to recent developments in computational power and broadband digital communications (*Wu et al., 2014*). They have given rise to such powerful tools as cloud infrastructures (*Dikaiakos et al., 2009*): distributed low-latency big data repositories (*Simmhan et al. 2013*). They provide great opportunities for applying newest machine learning tools, such as deep learning with multilayer neural networks, random forests with gradient boosting, support vector machines and many other (*Deng & Yu, 2014; Rojas, 2013; Breiman, 2001; Ben-Hur et al., 2001; Ding et al., 2002; Xu, Liu, Gong, 2003; Banerjee et al., 2007; Mikolov et al., 2013; Ge & Mooney, 2009; Dohrn & Riehle, 2011; Ermilov et al., 2013; Google Books Ngram Viewer; Abadi et al., 2016; Singhal, 2012*). These computationally intensive models requiring large sets of heterogeneous statistical data for training, can be applied to a variety of knowledge-work automation tasks, including those in the area of technology foresight and horizon scanning. In addition, the integrated environment, which the Internet[17] provides today, allows to partially automate the machine learning itself through automatic generation of big data learning samples, and even ontologies for the purpose of entity linking (*Fortuna, Grobelnik, Mladenic, 2006; Lee et al., 2007; Dehab, Hassan, Rafea, 2008; Sanchez & Moreno, 2008*).

For the purposes of emerging technologies identification, the text mining / semantic analysis tools seem to be most appropriate, as a task of identification of new man-made phenomena of known nature (technologies in this case) can be reduced to identification of new syntactic constructions signifying them. The fact that man-made artifacts tend to be explicitly named, described and discussed with the use of written language makes the problem well-posed.

There are certain issues that complicate the analysis of emerging technologies. One of the main problems is the commercial secrecy implemented by innovative companies in order to protect new ideas and inventions for the purpose of gaining competitive advantages in the future. To some extent, this issue may not be taken into account due to undisclosed technologies developed within the corporations being out of the scope of stimulating S&T policies. At the same time, when these new technologies are rolled out for use in commercial application, they fall under the various regulative policies, and certain disclosure of information becomes mandatory. However, it is necessary to mention that alternative analytical approaches based solely on expert activities also cannot solve the task of identifying the existence of undisclosed proprietary technologies. So, this is not the issue of text mining, but of future oriented technology analysis in general.

Moreover, the cost of analyzing big amounts of textual documents may be much lower than that of expert activities due to the advent of open science and open innovation paradigms and fast development of large open-access datasets suitable for S&T monitoring. In fact, the amount of openly available metadata today (such as summaries of various full-text sources, specialized Internet user discussions, and the like) raises the importance of metamining and dynamic ontology mining for such high-level tasks as technology landscape mapping for international and national governance.

To demonstrate the power of the text-mining-augmented techniques for technology identification and mapping we use the case of emerging technologies in the agriculture and food (A&F sector). Our choice is dictated by the fact that large proportion of most vicious global challenges are directly related to A&F sector (Godfray et al., 2010), and seemingly cannot be solved without radical technology innovation across the globe (*Royal Society, 2009*).

[17] URL: https://datahub.io (date last accessed 23.11.17).

Thus, A&F sector is one of the largest greenhouse gas emitters, globally, and the situation grows worse because of global shift towards consumption of animal products, more resource intensive and environmentally unsustainable. The global food problem is far from solution, as several hundred million people (*FAO, 2009*) in less developed countries face undernourishment and even famine, while global population growth (*United Nations, 2015*), which is far from plateauing puts additional demand pressure on the global food production-distribution systems.

The global food problem is aggravated by clearly expressed negative environmental trends threatening to decrease gross A&F output in the future (*World Bank, 2007*). These negative trends include degradation of bioproductivity of agricultural land, namely soil erosion (*Montgomery, 2007*), soil compaction (*Hamza & Anderson, 2005*), negative net nutrients flaw and fertility fall (*Pimentel, 2006*). They also include World Ocean bioproductivity loss due to overfishing (*Srinivasan et al., 2010*), contamination with harmful anthropogenic substances (*Aarkrog, 2003*), climate change and acidification (*Hoegh-Guldberg & Bruno, 2010*).

Understanding the severity of global A&F-related challenges and potentially significant role of the Russian Federation in overcoming them, the Government of Russia commissioned the development of A&F Foresight as a high-level national strategic document. Among the task of this study conducted in 2015 and 2016, the emerging technologies identification and mapping was conducted based on a synthesis of expert-based and digital-data-driven methods. The Russia A&F Foresight 2030 was officially endorsed in January 2017[18].

This paper discusses one of the aspects of technology foresight methodology enhancements that was introduced in that study, with considerations on further augmentation and automation of S&T evidence gathering based on big data, semantic analysis and machine learning.

Methodology

The main hypothesis that is tested in this paper is that "emerging technology" as a signifying syntactic construction has not, to a large extent, lost its semantic utility despite easily visible hype around this concept. Along with other closely semantically associated terms (synonymous and quasi-synonymous) and with the proper use of modern automatic syntactic analysis techniques it can be effectively used as a highly-informative anchor term for bulk identification of the phenomena signified by it, namely, *technologies that are currently emerging* from the ever-intensifying global science and technology development process.

After analyzing this hypothesis based on the ample material of 2-year foresight study of the A&F sector and with the use of functionality of National Research University Higher School of Economics's (NRU HSE) Text Mining System, we proceed to discuss the wider context of big-data-augmented technology identification and mapping and, more generally, the prospects of big-data-driven automation of STI policy evidence gathering.

The technical infrastructure and methodology of identification and mapping of emerging technologies via text mining, applied in the study, is described below.

The system is built completely on open source code libraries as well as proprietary code of the NRU HSE. The architecture of the system provide hybrid SQL/noSQL database allowing for distributed computations and high-intensity in-memory data operations. The pattern of data storage developed by the authors allows optimal combination of data normalization and array manipulation principles and provides the ability to get most of data extraction and integration responses within seconds to minute for queries over the whole data storage.

The composition of data sources of our text mining system at the time of the exercise included:

---

[18] URL: https://issek.hse.ru/data/2017/02/06/1167349282/%D0%9F%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7%20%D0%BD%D0%B0%D1%83%D1%87%D0%BD%D0%BE-%D1%82%D0%B5%D1%85%D0%BD%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B9%20%D1%81%D1%84%D0%B5%D1%80%D1%8B.pdf (date last accessed 23.11.17).

- stratified random sample of summaries and metadata of around 2 million of internationally top cited research papers for 10 years period, acquired from the open citation indexes and other open data sources;
- stratified random sample of summaries and metadata of around 2 million international patents for 10 years period, acquired through open access sources of WIPO PCT patents;
- 5 million newsfeeds items from Alexa and SimilarWeb tops of global news portals with science and technology flavor, for the period since the inception of the WWW;
- more than 200 000 analysis and forecast reports, declarations, proceeding and other documents in PDF format, openly accessible through web search engines and institutional web sites, including the web sites of UN organizations; of them 30 thousand directly related to agriculture, with around 12 thousand documents by FAO, 8 thousand by USDA and from other organizations.

At the time of the study, the system featured more than 12 million individual documents, several hundred million individual sentences, of which up to 3 million documents were at least partially relevant to A&F sector and adjoining sectors, such as biotechnology and bioenergy, more than a billion terms, of which more than one hundred million were object signifiers (see below).

The principal steps in data extraction, transformation and loading for the purpose of filling the integrated database of the system include the following:

- data extraction from heterogeneous documents of both structured (publication, patent metadata) and unstructured (full text reports, declarations and other documents) formats,
- sentence segmentation and word tokenization,
- word lemmatization and part-of-speech tagging,
- syntactic constructions formation based on universal dependencies standard of syntactic analysis with focus on *properties* and *functions* (Yoon & Kim, 2012) of text objects,
- weighting the syntactic collocations by the probability of their high information content (semantic role of signifying the objects, processes, concepts, and other phenomena of similar nature),
- named entities extraction and classification by types (persons, companies, geographies, etc.),
- splitting documents into pseudodocuments based on semantic similarity, clustering and biclustering of documents, pseudodocuments and syntactic collocations,
- compaction of sparse term-(pseudo)document matrices into graph structures and clustering of graphs,
- by-term calculation of absolute and relative frequency, dynamics of relative frequency, cross-relevance, monopolism, specificity, cross-specificity and other metrics derived from text statistics and augmented text-syntactic statistics (over 50 attributes of occurrence, graph and syntactic metrics),
- application of semi-supervised (including, bootstrapped though automatically generated labelled sets) and supervised machine learning techniques aimed at classification and ranking of terms and pseudodocuments.

The following three approaches have been used in this paper for extraction of technologies:

1. Cascade identification of words being governors within terms allows to identify unigrams – universal signifiers of semantic field of *"techologicality"*, i.e. words that radically increase the probability of an ngram containing them to be a name of certain technology. Examples of such words are *technology*, *method*, *system*, *platform*, *model*, *tool*, *layer*, *enzyme* and many other (fig. 1). Extraction of all object-signifying words allows to then get hundreds of thousands of terms – candidates for being names of technologies (for instance, DNA sequencing *technology*, or recirculating aquaculture *system*, etc.). These long lists are

then filtered with the use of author-built machine learning algorithms dealing with "information-richness" of terms, their monopolism and specificity and other attributes. The resulting lists containing just thousands of terms are than linked to entities of existing ontologies to acquire information about their expansions, definitions, and Russian translations.

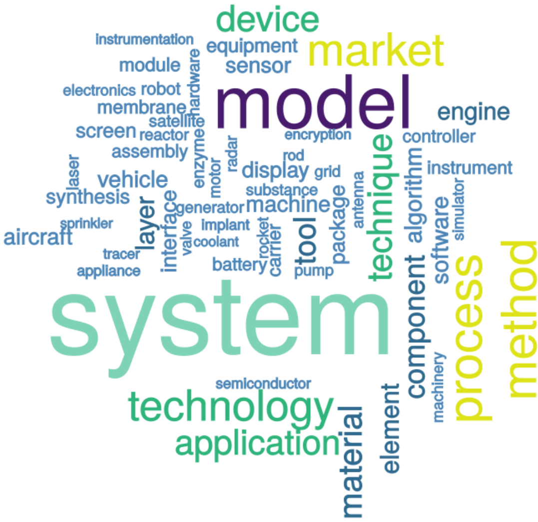

Figure 1. Proposed universal signifiers of semantic field of "techologicality"

Source: National Research University Higher School of Economics

2. Identification of statements aligning with certain syntactical patterns and containing anchor technology terms or their quasi-synonyms acquired through *word2vec*-like approaches (Mikolov et al., 2013). Examples of such statements are sentences containing the closely semantically interrelated (as shown by term clustering in our text mining system over the whole range of available data) anchor terms of "emerging technology" and "disruptive technology" together with enumeration, definition, or declaration syntactic patterns. Elements of such statements are mentions of technologies that are deemed emerging by the authors of the statements. The list tokenization and following disambiguation of the meaning is done in a manner similar to described above (iterative multidimensional filtering and entity linking).

3. Identification of wires (press release headlines) on issuance of market reports by key global players in this field (each of which have up to a million or even more market reports) is done on media database segment of the system. Extraction of press releases lexically relevant to agriculture is done (by cosine similarity of documents by vectors of significant syntactic constructions, and other, more advanced methods based on pseudodocument topic modelling, term-document biclustering, etc.). Persistence analysis (neologism identification) is performed over the significant terms extracted from such press release headlines.

Analysis of dynamics of intensity of the presence in the discourse during the last years is done for the candidate technology-signifying terms. Interpretation of the results of this analysis is done based on the assumption that currently unfolding technology trends (including the development and adoption of emerging technologies) are characterized by the growth of interest towards them at least in one of the corpora of documents (science, patents, news and blogs, analytical reports). It is suggested that emerging technologies with strong potential of surviving and upscaling to the global production systems have a signature of ever-increasing public awareness of them.

The mapping of the technologies is inseparably linked to their identification and is the last stage of filtering of the candidate technology-signifying terms acquired through three identification methods described above. The mapping is executed through the use of combination of machine learning methods, in the core of which lies the ensemble of co-occurrence term graph clusterizers with regularizators responsible for control of mutual information and topic monopolism of terms. The principal results of mapping are clusters of terms (or organisations and persons, or documents) that give insights into optimal groupings on technologies by semantic similarity.

Findings

1. Technology-signifying terms identification

Taking into consideration scope limits of our paper, only one proposed approach for extraction of technologies – with universal signifiers of semantic field of "techologicality" – could be fully illustrated. Agriculture and food technologies identified from the full texts on sectoral topics are listed below:

3D cell culture
active composting technologies
active packaging technologies
adaptive agricultural technologies
aeration technologies
aeroponics technologies
agricultural conservation technologies
agricultural drones
agricultural harvester
agricultural inoculants
agriculture sprayer
agrobacterium technologies
agrochemical application technologies
agroecological technologies
agroforestry technologies
agrometeorological technologies
agropastoral technologies
agroprocessing technologies
agrosensor technologies
agrovoltaic technologies
algae technologies
algal biofuel technologies
alternative aquaculture technologies
animal breeding technologies
animal cloning technologies
animal genetics
animal growth promoters & performance enhancers
anti-erosion technologies
antifouling technologies
aquaponic technologies
aquasilviculture technologies
artificial catfish breeding technologies
artificial insemination technologies

artificial meat technologies
artificial placenta technologies
azoxystrobins
bio stimulants
biochip technologies
biocides
biocontrol agents
bioconversion technologies
bio-engineering technologies
biofertilizer technologies
biofuels technologies
bioherbicides
biological seed treatment
biomarker technologies
biomass densification technologies
bionematicides
bioprinting technologies
biorefinery technologies
brooding box technologies
bt crop technologies
BT technologies
bycatch reduction technologies
canning technologies
cattle health technologies
cellular technologies
chlorine free technologies
cloning technologies
composting technologies
conventional breeding technologies
cotton seed treatment
CRISPR
crop disease molecular diagnostics technologies
crop technologies
cultivation technologies
cultured meat technologies
culturing technologies
dairy alternatives
dairy herd management
dairy technologies

DNA micro array technologies
DNA recombination technologies
DNA sequencing technologies
drip technologies
dry food technologies
elevator technologies
enzyme technologies
farm equipment
farm technologies
feed phytogenics
feed prebiotics
feed probiotics
feed technologies
feedlot technologies
fermentation technologies
fertigation technologies
fertilisation technologies
fertilizer deep placement technologies
fish farming technologies
fish processing technologies
fishpond technologies
food & beverage cold chain logistics
food acidulants
food flavors
food safety testing
food stabilizer systems
fruit juice packaging
functional non-meat ingredients
garden tractors
genetic modification technologies
genetic technologies
genetically modified organisms
genomic technologies
genomics technologies
germplasm technologies
gluten free food

green revolution technologies
greenhouse technologies
hatchery technologies
haying & forage machinery
horticultural technologies
hybrid rice technologies
hybridoma technologies
hydroponic greenhouse technologies
hydroponic technologies
inoculant technologies
integrated agriculture technologies
integrated soil fertility management technologies
intensive maize technologies
irrigation technologies
LEISA technologies
livestock technologies
mariculture technologies
marine hatchery technologies
meat processing technologies
meat substitutes
microalgae technologies
microbial technologies
micro-garden technologies
micro-irrigation technologies
micronutrient technologies
mini-hatchery technologies
minimum tillage technologies
moisture save technologies
molecular breeding technologies
molecular marker technologies
molluscicides
mutagenesis technologies
next-generation sequencing technologies
non-GM breeding technologies
offshore aquaculture
organic dairy
pasteurization technologies
pelleting technologies
pest control services
pesticide technologies
platform tray feeder technologies
polyculture technologies
post-harvest technologies
poultry processing technologies
precision agriculture
precision irrigation
pruning technologies
rapid composting technologies
recirculating aquaculture technologies

resistant crop technologies
restriction fragment length polymorphism technologies
rice technologies
sanger technologies
scale processing technologies
seafood technologies
seeding technologies
smart irrigation
smart packaging technologies
soil conservation technologies
soil management technologies
sustainable aquaculture technologies
synthetic biology
synthetic seed technologies
system of rice intensification technologies
transgenic cotton technologies
transgenic crop technologies
transgenic technologies
trawl technologies
tree trimmers
urea-molasses multinutrient manufacturing technologies
veterinary technologies
weed technologies
zinc-finger nucleases technologies

2. Technology-signifying terms mapping

Further step of technology-signifying terms analysis involve their mapping on the bases of semantic proximity and discourse presence intensity. Two main approaches for subsequent visualization were developed: semantic and trend maps.

Semantic map (fig.2) is a clustered co-occurrence graph, which consists of thousands of connected vertices, each of which represents some important term (notion). Unfortunately, in the illustrative visualization few of vertices can be labelled without overlap. In our text mining system semantic map is interactive, with ability to zoom in and drill down. However, for the purposes of compiling the reports a table of co-occurrences describing the whole semantic graph can be downloaded by the user. Semantic map demonstrates dynamic classification, which is essential task for technology landscape mapping. Candidate technology-signifying terms aggregate in the following large colored clusters: urban agriculture technologies (robotic greenhouses, vertical farms, recirculating aquaculture systems, aquaponics, hydro/aeroponics, artificial lighting, sensors and control systems etc.), precision agriculture (geographic information systems, autonomous vehicles, GNSS and remote sensing, etc.), environmental management (waste processing, soil management, organic agriculture, etc.), synthetic and molecular biology (genetic engineering, vaccines, antibiotics, probiotics production technologies, embryo transfer, DNA sequencing), advanced food technologies (biochemical, enzyme technologies, nanotechnologies for food industry, active packaging, nutrient additives), local bioenergy and smart grid (biofuel production, solar energy etc.). Dynamic classification based on clustering mechanisms provides mapping of the identified technologies, thematic linkage among them and with non-technological concepts and topics in the sphere. It provides an ad hoc ontology for fast human learning and for expert discussions, even in the absence of official taxonomies. This allows to map even emerging fields which haven't yet been categorized for the purposes of official statistics.

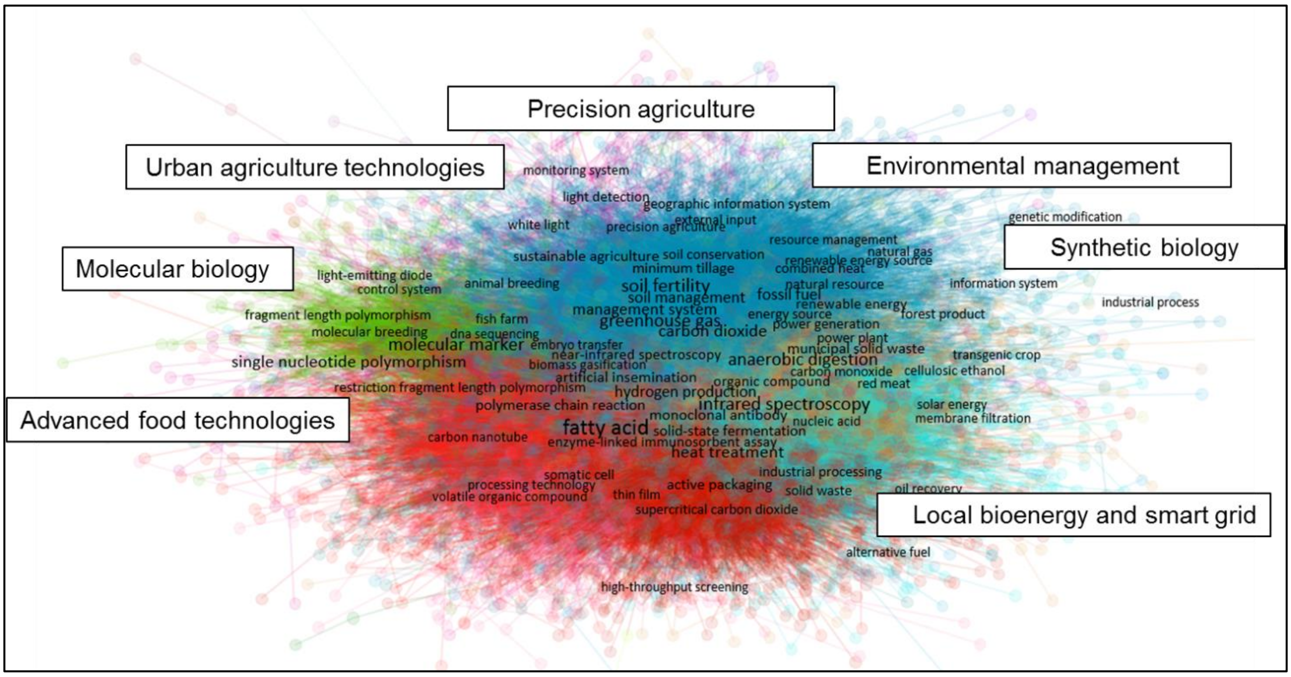

Figure 2. Semantic map of agriculture and food sector technologies

Source: National Research University Higher School of Economics's Text Mining System

3. Emerging technologies identification

Identified technologies can also be distinguished from one another by dynamics of intensity of their presence in the discourse during the last years. It can be visualized as trend maps: 2-dimensional plots with one axis representing the popularity of a term and the other showing the year-by-year dynamics of the normalized popularity (relative frequency of use). For the trend maps of technologies in agriculture on media resources and patent applications see fig. 3 and 4 respectively). *The upper-right quadrant* consists of the strongest topics shaping the future agenda of the sector, they are popular and gaining traction: in media they exemplified by CRISPR technologies, agroforestry and aquaponic technologies, precision agriculture and microalgae technologies etc. In patents this group also consists of several genetic technologies but moreover includes fertigation and hatchery technologies. *The lower-right quadrant* contains the so-called "weak signals": they are highly trending but underrepresented in discourse yet. They can contain the *emerging technologies*. This group presented in media by smart irrigation technologies, molecular breeding and zinc-finger nucleases technologies etc. The presented show, for instances, that in patents, at the same time, bionematicides, mutagenesis and synthetic seed technologies are increasingly gaining popularity. Among the popular topics losing their significance in media are fertilisation, pruning, antifouling technologies and many more. In patents, topics with declining popularity are composting and horticultural technologies, among other.

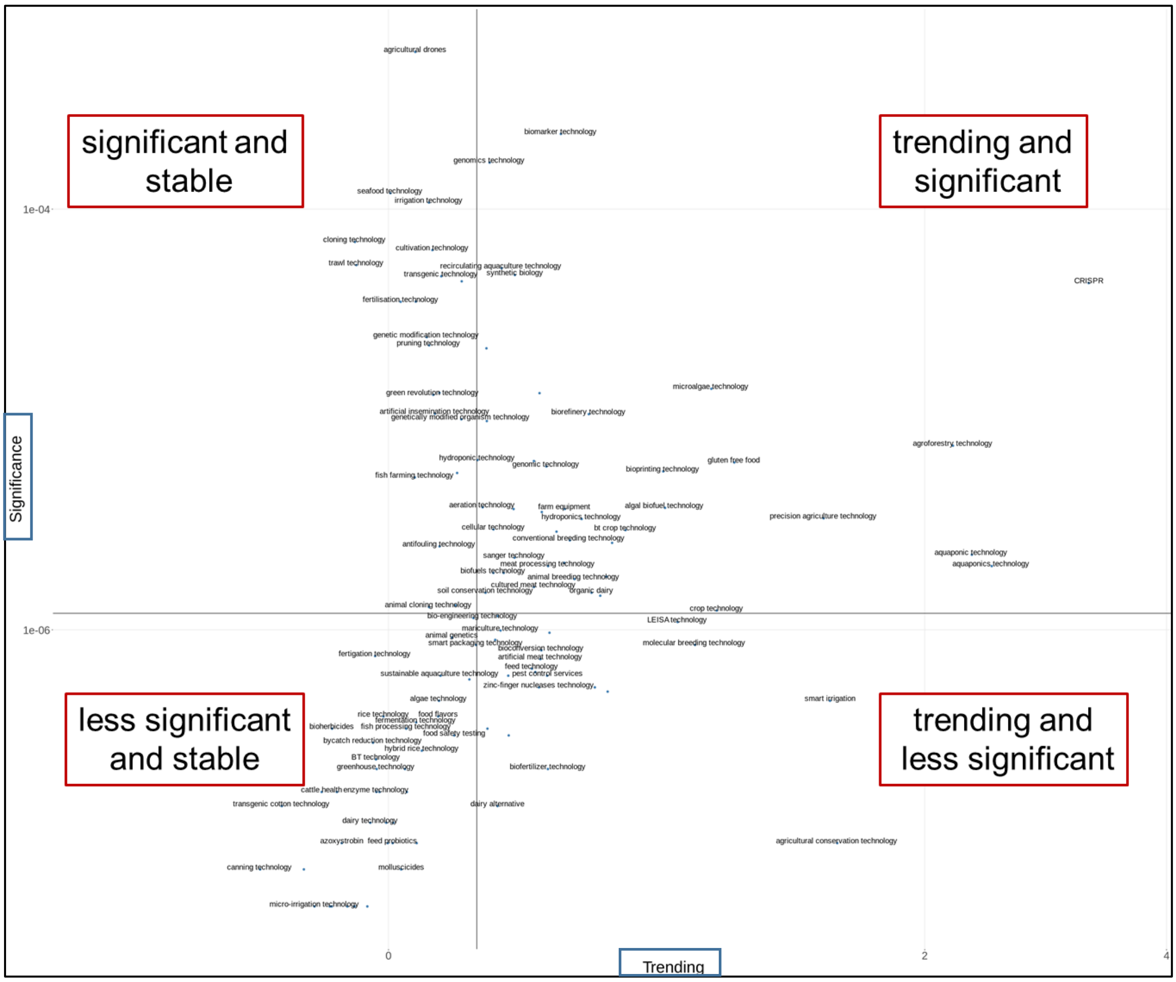

Figure 3. Trend map of agriculture and food sector technologies on media resources

Source: National Research University Higher School of Economics's Text Mining System

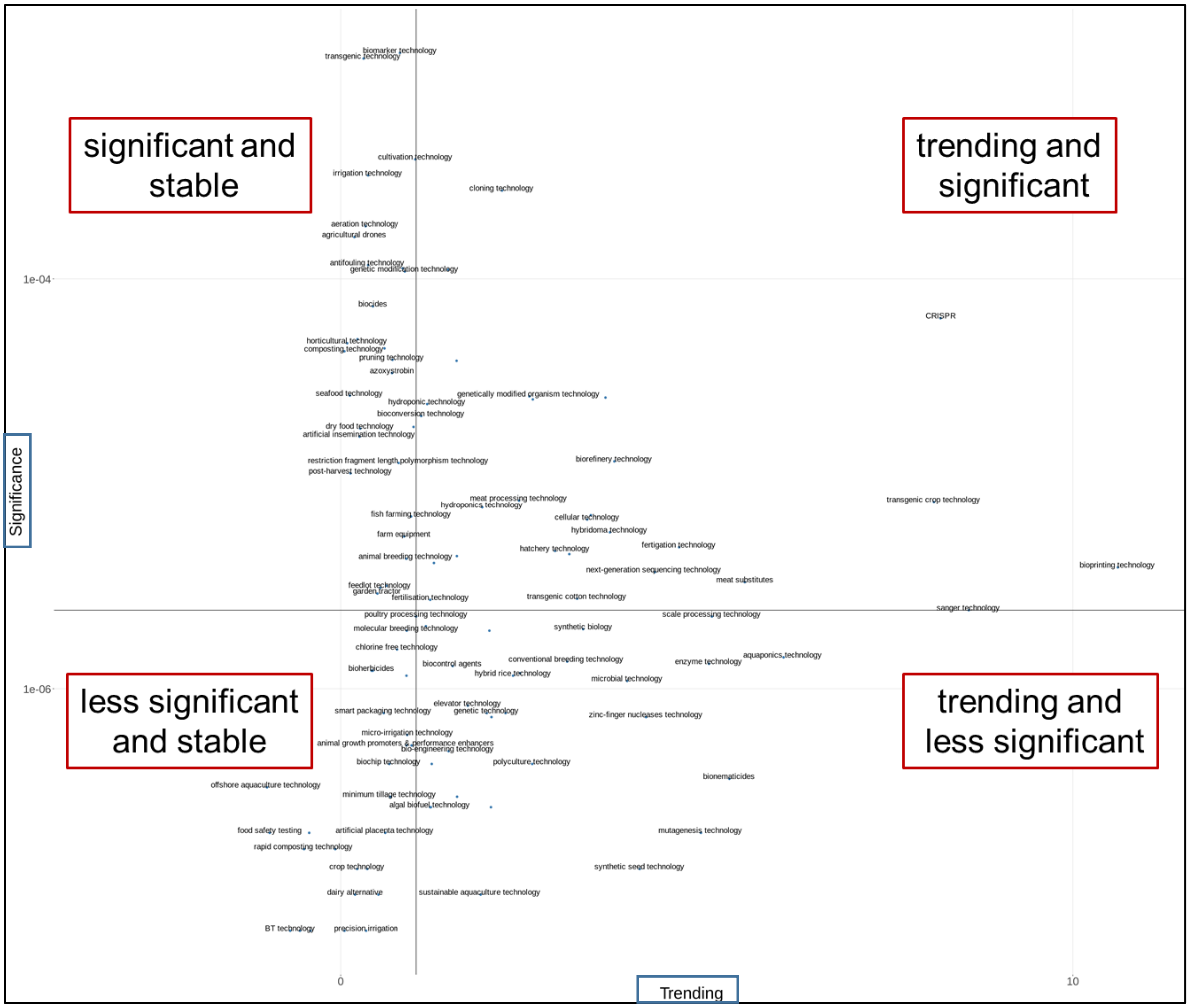

Figure 4. Trend map of agriculture and food sector technologies on patent resources

Source: National Research University Higher School of Economics's Text Mining System

To make more apparent the differences in normalized popularity of topics in the media and, for example, in patents special visualization instrument was generated – hype map (fig. 5).

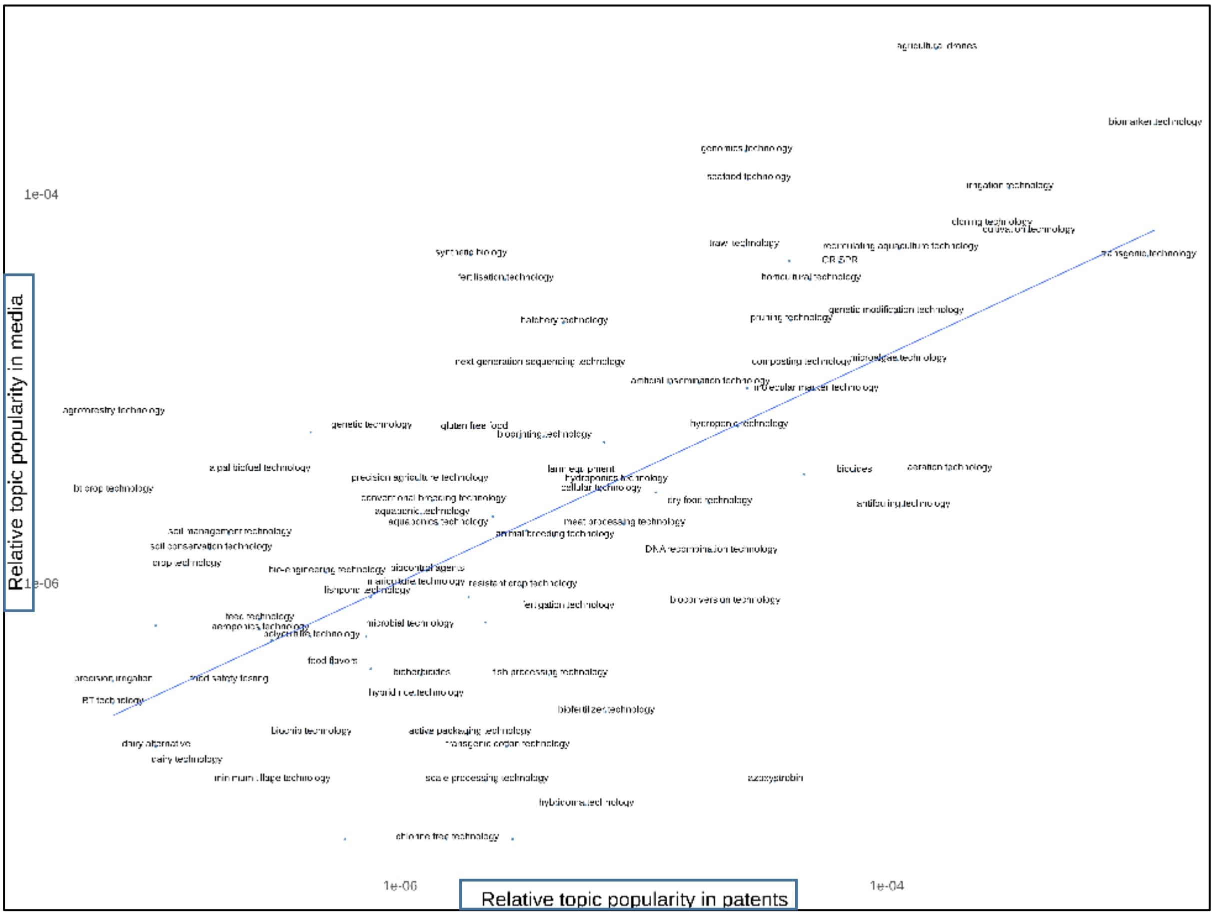

Figure 5. Hype map of agriculture and food sector technologies popularity in media resources vs patents

Source: National Research University Higher School of Economics's Text Mining System

The blue line on the map shows the "watershed" between the topics more popular in the media and more popular in patents. In this case following topics attract remarkably more attention in media: agroforestry, agricultural drones, gluten free food production etc. It can be interpreted as the high level of related technology packages' commercialization and public awareness rise.

Based on the presented text mining results the structured table of emerging technology trends affecting agriculture and food sector of the Russian Federation was formed through the subsequent expert interviews and panel discussions during foresight study of the sector (see Annex).

Conclusions and discussion

The raw results of the emerging technologies identification and mapping (before the final integration of outputs of three different algorithms) show that the three algorithms used (syntactic constructions extraction, anchor enumerations extraction, and relevant neologisms extraction) are only partially mutually redundant. The first and second methods having common anchor terms as seeds might be expected to yield similar results, but this is not the case. The explanation is that the first method filters out any individual terms without an anchor term within, while the second method requires the presences of the anchor term only in the statement, but not in the members of enumeration. One of the consequences is that branded, trademarked and other proprietary technologies are almost not present in the output of the first method. (With some exception, such as Round-Up pesticide, which name has gone almost denominative in the GMO application discourse, so that there are such phrases in the analyzed texts as "roundup technology"). At the same time, concentration of such technologies in the output of the second method is quite high, reaching up to 67% due to concentration of technology enumeration, to significant extent, in media (news and blogs), which are oriented to covering the products existing on the market.

The last method is less applicable for the identification of emerging technologies with low technology readiness levels, as the mere existence of market of the report on the market made by the new technology shows that the technology is relatively mature. However, concentrating on the segment of the market reports with neologisms in their headings (which correspond to headlines of the connected press releases) helps to overcome this issue, as many market reports seem to be released while the markets themselves exist only as anticipations, not real product segments. Also concentrating on the press releases without ballpark value estimates of the markets (these estimates are disclosed in press releases on market reports covering more or less mature markets) can further help. In addition, markets with quite high-declared growth rates (some reach more than 100% per year, however, not necessarily in agriculture) are the good candidates. Finally, mining of the publicly available tables of contents of these market reports can help identify more concrete and less mature technologies. This is because each market report tend to look into tens to hundreds of product segments. Most of these segments are technology-driven, and reports covering relatively mature markets often cover the yet virtually inexistent segments. The latter are pointers to the emerging technologies.

The three methods of identification have different performance in terms of the number of significant and relevant entities extracted. The first method operates relatively good on all the sources of data, with better results on research papers and especially patents abstracts, which contain less low-informative terms than, for instance, general reports of international organizations, and discuss technologies in more concrete terms than the latter. The second method has low performance on science because research papers abstracts tend not to list the technologies, instead listing the research methods used. The research methods in some fields (such as x-ray diffraction, scanning electron microscopy and many other in materials sciences) could easily be interpreted as technologies, while the share of such methods in agriculture science is much lower. The enumerations of technologies are not performing well on patents too, as patents describe inventions’ details rather than list technologies. The method performs well on analytical reports and best on media, where enumerations are very popular. The third method works only on a very specialized segment of media, but for correct work (principal terms identification, neologisms detection) requires automatically built ontologies based on other sources of data.

The inner circle tools of mapping (aimed at preliminary filtering of data and at technology classification suggestions) help analysts, in our experience, to rise work productivity. It takes analysts much less time to review hundreds instead of thousands of terms due to filtering, choose most important technologies and categorize them with the help of machine-generated suggestions on places of technologies in sectoral ontology. Furthermore, productivity is improved due to the "single window" effects of automatic attachment of expansions (of abbreviations), definitions and translations of technology-signifying terms. These effects decrease the negative consequences of cognitive flow disruptions produced by switching between data sources in search of information, which is dominant in traditional analysts working settings.

The outer circle mapping tools (aimed at helping to present the results easily and effectively), such as semantic map and trend map shown above, can be used by both the internal analysts of a company conducting a foresight study and the invited experts. These tools provide very useful information products for structuring expert interviews and panel discussions (for the example of results usage in the A&F Foresight 2030 see fig. 6). They also give quite usable products out-of-the-box, which are easily understandable by the decision makers. Therefore, there is a capacity for transversal automation of some aspects of the production of evidence on emerging technologies for supporting decision-making.

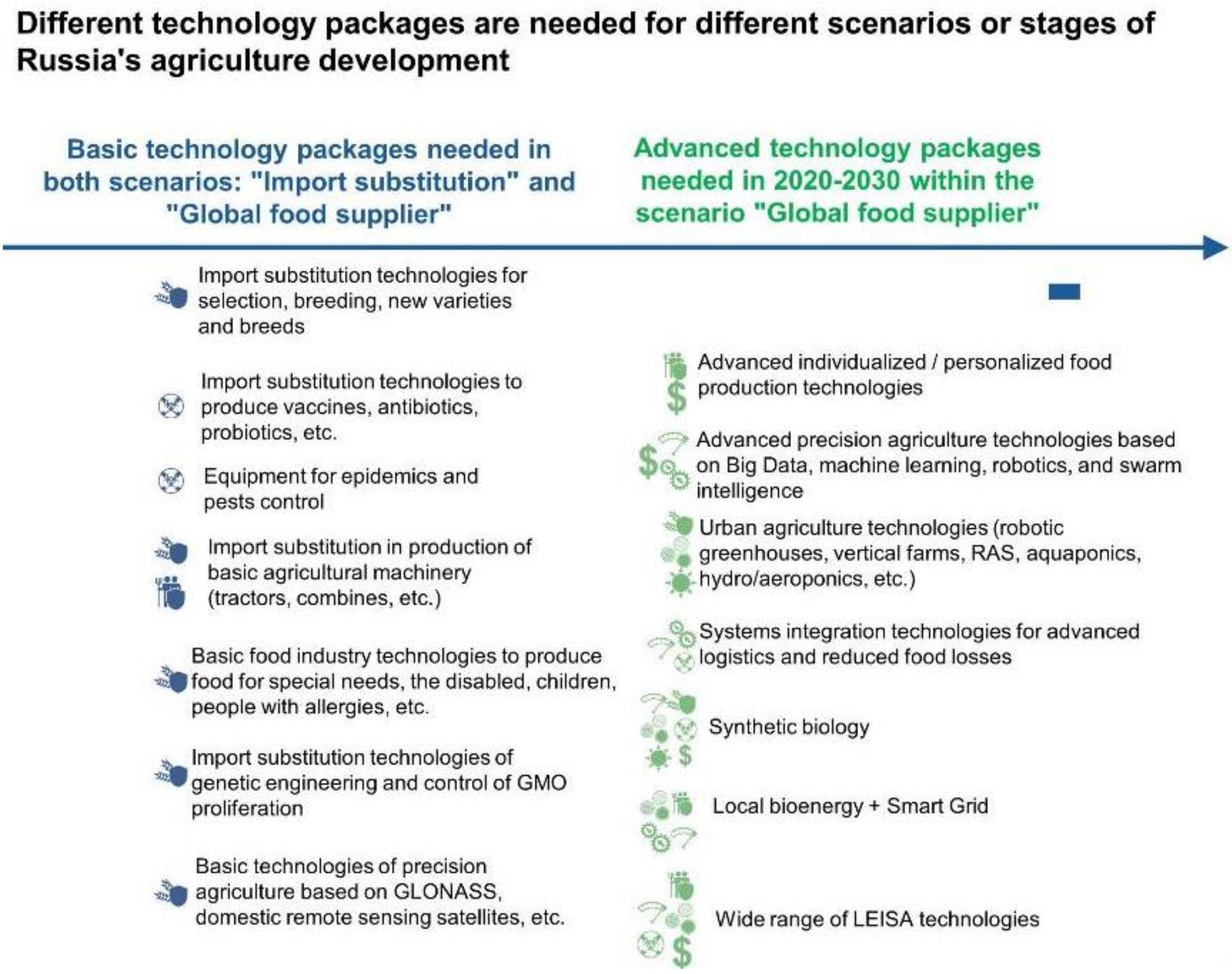

Figure 6. Example of technology grouping discussed during expert panels

Source: National Research University Higher School of Economics

Complete automation of the process is not a desired outcome, as the continuous dialogue among decision makers and expert community is of vital importance for the sustainable governance. However, the demonstrated tools of big data analysis give decision makers the leverage in the negotiation with the sectoral stakeholders, scientists and consultants. The latter will have to address practical questions about the technologies they advertise for government or corporate support based on objectively collected data on the parameters of these technologies in the current semantic field.

Also fostering the new generation of analysts who are strong in both the domain field and science, technology, engineering and mathematics (STEM) / data science is needed. The process of production of analytical results by big data systems should be transparent and understandable, so that the results are interpreted correctly and fully, where no data manipulation is possible. The demonstrated text mining approach will not replace the domain experts as foresight analysts in the foreseeable future, but will transcend the foresight exercises from local ontology building to the high-level ontology interpretation. Many foresight studies have been restricted to collective mapping of trends in some domain in hope that the mapping is precise and full. If mapping is automated, resources are available for what foresight is really about: building an integrated and balanced vision of the future based on an intense interpersonal communication of domain experts.

Annex. Hierarchical table of emerging technology trends that can affect agriculture and food sector of the Russian Federation

	Biotechnology
1.	Substituting conventional mainline crop varieties with genetically modified ones, more resistant to pests, diseases, droughts, herbicides; and making economic use of wild plants and animals, breeding new varieties and species on their basis
2.	Substituting conventional forest plantations with plantations of fast-growing genetically modified trees
3.	Increased application of animal cloning for specific purposes (producing biologically active preparations, veterinary research, etc.)
4.	Increased number of high-tech R&D projects to clone and mass-breed extinct animal species (banteng, mammoth, etc.).
	Smart efficiency
5.	Abandoning flood irrigation in favour of drip underground irrigation, to significantly reduce water consumption
6.	Increased demand by businesses for high-tech precision soil fertility diagnostics solutions, to abandon standardised uniform application of fertilisers in favour of dynamic one, differentiated to match specific indicators of nutrients' content in soil
7.	Moving on from manually operated agricultural machinery to driverless machines, based on micro-geopositioning and self-learning robots technologies
8.	Abandoning conventional fertilisers in favour of composite and slow release fertilisers (capsules' shells degrade under specific weather conditions, with capsule layers containing various nutrients subsequently being released into the soil in coordination with plants' life cycle stages)
9.	Reaching the commercialisation stage of technologies allowing to detect nutrient shortages (macro- and microelements) in crops' nutrients in real time
10.	Reaching the commercialisation stage of technologies to monitor the state of health and specific needs of individual farm animals in real time
	Substituting chemical-based solutions with biological ones
11.	Increasingly large-scale substitution of agricultural chemicals with organic fertilisers – by-products of agricultural activities, leading to reduced costs and reduced negative impact on the environment
12.	Increasingly large-scale application of integrated pest protection technologies; abandoning pesticides in favour of biological weed and pest killers
13.	Discontinuation of the use of antibiotics in animal farming in favour of innovative immunomodulatory techniques
	Environmental sustainability
14.	Increased demand by agricultural companies and environmentally conscious consumers for integrated remote monitoring of good practices of agriculture production and tracking product supply chains (including RFID/GNSS labelling)
15.	Increasingly large-scale application of water filtration and prior treatment in irrigation systems, to efficiently prevent soil salination
16.	Abandoning conventional mechanical-based agricultural waste water filtration technologies (capable of removing organic compounds) in favour of nanotechnology- and microbiology-based fine filtration solutions, for complete water treatment
17.	Supplementing conventional ploughing techniques by no-till farming and other soil conservation technologies
18.	Developing integrated technological solutions and equipment to create industrial-scale artificial agroecosystems

19.	Sustainable and rapid reduction of costs of technological solutions for mass production of second-generation biofuel, including gaseous fuels from organic materials, liquid organic fuels from cellulose
20.	Stepping up exploratory research to mass-produce inexpensive low-energy fertilisers for algae, to make third-generation biofuel production technologies competitive.
21.	Increased demand, in developed and developing countries alike, for microbiology-based technologies to recover degraded and polluted soils, make lands ruined by bad irrigation practices, overgrazing, and pollution by industrial, communal, and radioactive waste suitable for agricultural use again
	Intensification – compacting
22.	Active application, by advanced aquaculture complexes, of technologies to combine fisheries and agriculture (aquaponics), allowing to process fish excreta to make plant nutrients *in situ*, in a fully closed water cycle
23.	Abandoning conventional food packaging materials in favour of nanocellulose, with bactericide properties
24.	More efficient aquaculture technologies for fresh and sea water (fish farming, growing shellfish and crayfish); abandoning fishing in favour of aquaculture
25.	Developing climate-independent agricultural infrastructure including closed artificial ecosystems for agricultural purposes
26.	Developing commercial solutions for super-intensive plant growing based on hydroponics, aeroponics, robotics, and “verticalisation” (vertical farms)
27.	Technologies for growing farm animals’ nutritious tissues in artificial nutrient solutions (in-vitro meat, laboratory meat) moving on from conceptual development stage to demonstration of feasibility
28.	Increased number of projects to build fully robotic greenhouses in adverse climate areas
	Reducing food industry waste
29.	Application of technologies for instant low-temperature (shock) freezing to replace conventional freezing technologies, to better preserve organoleptic and nutritious properties of agricultural products
30.	Application of new preserving agent types, programmed to self-destruct after a certain period of time, for safer preservation of agricultural products
31.	Abandoning practices of food waste disposal at garbage dumps in favour of smart recycling technological solutions, using food waste to produce energy and biochemical products

References

Aarkrog, A. (2003). Input of anthropogenic radionuclides into the World Ocean. Deep Sea Research Part II: Topical Studies in Oceanography, 50(17), 2597-2606.

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., & Ghemawat, S. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Aghion, P., & Griffith, R. (2008). Competition and growth: reconciling theory and evidence. MIT press.

Arnaud, B., Dupont, J., Koh, S. H., & Schreyer, P. (2011). Measuring Multi-Factor Productivity by Industry: Methodology and First Results from the OECD Productivity Database. OECD. Paris.

Bakhtin, P., & Saritas, O. (2016). Tech Mining for Emerging STI Trends Through Dynamic Term Clustering and Semantic Analysis: The Case of Photonics. In Anticipating Future Innovation Pathways Through Large Data Analysis (pp. 341-360). Springer International Publishing.

Banerjee, A., Dhillon, I., Ghosh, J., Merugu, S., & Modha, D. S. (2007). A generalized maximum entropy approach to bregman co-clustering and matrix approximation. Journal of Machine Learning Research, 8, 1919-1986.

Ben-Hur, A., Horn, D., Siegelmann, H. T., Vapnik, V. (2001). Support vector clustering. Journal of machine learning research, 2, 125-137.

Bildosola, I., Rio-Belver, R.M., Garechana, G., Cilleruelo, E. (2017). TeknoRoadmap, an approach for depicting emerging technologies. Technological Forecasting and Social Change, 117, 25-37.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Brian, A.W. (2009). The Nature of Technology. New York: Free Press.

Brocker, J., Dohse, D., & Soltwedel, R. (Eds.). (2012). Innovation clusters and interregional competition. Springer Science & Business Media.

Candorin, E., Klofsten, M., & Johansson, S. G. (2016). Future Challenges for Science Parks-Attractiveness and Recruitment of Talents. In High Technology Small Firms (HTSF) Conference June,(13), 14 & 15, 2016, University of Liverpool Management School, UK.

Dahab, M. Y., Hassan, H. A., & Rafea, A. (2008). TextOntoEx: Automatic ontology construction from natural English text. Expert Systems with Applications, 34(2), 1474-1480.

Deng, L., Yu, D. (2014). Deep learning: methods and applications. Foundations and Trends in Signal Processing, 7(3–4), 197-387.

Dikaiakos, M. D., Katsaros, D., Mehra, P., Pallis, G., & Vakali, A. (2009). Cloud computing: Distributed internet computing for IT and scientific research. IEEE Internet computing, 13(5).

Ding, C., He, X., Zha, H., Simon, H. D. (2002). Adaptive dimension reduction for clustering high dimensional data. In Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on (pp. 147-154). IEEE.

Dohrn, H., & Riehle, D. (2011). Design and implementation of the sweble wikitext parser: unlocking the structured data of wikipedia. In Proceedings of the 7th International Symposium on Wikis and Open Collaboration (pp. 72-81). ACM.

Ermilov, I., Martin, M., Lehmann, J., & Auer, S. (2013). Linked open data statistics: Collection and exploitation. In International Conference on Knowledge Engineering and the Semantic Web (pp. 242-249). Springer Berlin Heidelberg.

Fan, P., Watanabe, C. (2006). Promoting industrial development through technology policy: Lessons from Japan and China. Technology in Society, 28(3), 303-320.

FAO (2009). The State of Food Insecurity in the World, FAO, Rome.

Fortuna, B., Grobelnik, M., & Mladenic, D. (2006, October). Semi-automatic data-driven ontology construction system. In Proceedings of the 9th International multi-conference Information Society IS-2006, Ljubljana, Slovenia (pp. 223-226).

Furukawa, T., Mori, K., Arino, K., Hayashi, K., Shirakawa, N. (2015). Identifying the evolutionary process of emerging technologies: A chronological network analysis of World Wide Web conference sessions. Technological Forecasting and Social Change, 91, 280-294.

Ge, R., & Mooney, R. J. (2009). Learning a compositional semantic parser using an existing syntactic parser. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2 (pp. 611-619). Association for Computational Linguistics.

Godfray, H., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, Muir, Pretty, J. Robinson, S., Thomas, S., Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people, Science Express, 327(5967), 812–818.

Gokhberg, L., Fursov, K., Miles, I., & Perani, G. (2013). 15 Developing and using indicators of emerging and enabling technologies1. Handbook of innovation indicators and measurement, 349.

Google Books Ngram Viewer info page: https://books.google.com/ngrams/info

Guellec, D., B. van Pottelsberghe De la Potterie (2001). R&D and productivity growth: A panel analysis of 16 OECD countries. STI Working Paper, 2001/3. OECD. Paris.

Halaweh, M. (2013). Emerging technology: What is it. Journal of technology management & innovation, 8(3), 108-115.

Hamza, M. A., & Anderson, W. K. (2005). Soil compaction in cropping systems: a review of the nature, causes and possible solutions. Soil and tillage research, 82(2), 121-145.

Hidalgo, C. (2015). Why information grows. The evolution of Order, from Atoms to Economies.(Ebook) New York: Basic Books.

Hoegh-Guldberg, O., Bruno, J. F. (2010). The impact of climate change on the world’s marine ecosystems. Science, 328(5985), 1523-1528.

Huang, Y., Zhang, Y., Ma, J., Porter, A., Wang, X. (2015). Tracing Technology Evolution Pathways by Combining Tech Mining and Patent Citation Analysis. In 2015 Portland International Conference on Management of Engineering & Technology.

Jermann, C., Koutchma, T., Margas, E., Leadley, C., Ros-Polski, V. (2015). Mapping trends in novel and emerging food processing technologies around the world. Innovative Food Science & Emerging Technologies, 31, 14-27.

Joung, J., Kim, K. (2017). Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. Technological Forecasting and Social Change, 114, 281-292.

Kallhauge, A. C., Sjöstedt, G., & Corell, E. (2005). Global challenges: Furthering the multilateral process for sustainable development. Greenleaf Publishing.

Keenan, M., Cutler, P., Marks, J., Meylan, R., Smith, C., & Koivisto, E. (2012). Orienting international science cooperation to meet global 'grand challenges'. Science and Public Policy, 39(2), 166-177.

Kiel, D., Arnold, C., Baccarella, C., Voigt, K. I., & Hoffmann, D. (2015). Technology Identification in Relation to Embedded Systems.

Klein, J. T., Grossenbacher-Mansuy, W., Häberli, R., Bill, A., Scholz, R. W., & Welti, M. (Eds.). (2012). Transdisciplinarity: joint problem solving among science, technology, and society: an effective way for managing complexity. Birkhäuser.

Kojima, S., Ohmori, T., Watanabe, S., & Maeda, Y. (2012). Japan's Industrial Technology Development: The Role of Cooperative Learning and Institutions. Y. Okada (Ed.). Springer Science & Business Media.

Komatsu, T., Deserti, A., Rizzo, F., Celi, M., & Alijani, S. (2016). Social Innovation Business Models: Coping with Antagonistic Objectives and Assets. In Finance and Economy for Society: Integrating Sustainability (pp. 315-347). Emerald Group Publishing Limited.

Lee, C. S., Kao, Y. F., Kuo, Y. H., & Wang, M. H. (2007). Automated ontology construction for unstructured text documents. Data & Knowledge Engineering, 60(3), 547-566.

Lucheng, H., Xin, L., Wenguang, L. (2010). Research on emerging technology selection and assessment by technology foresight and fuzzy consistent matrix. Foresight, 12(2), 77-89.

MacKenzie, D. A., Wajcman, J. (1999). Introductory Essay. The Social Shaping of Technology (2nd ed.). Buckingham: Open University Press.

Madani, F. (2015). 'Technology Mining' bibliometrics analysis: applying network analysis and cluster analysis. Scientometrics, 105(1), 323-335.

Mani, S. (2004a). Government, innovation and technology policy: an international comparative analysis. International Journal of Technology and Globalisation, 1(1), 29-44.

Mani, S. (2004b). Institutional support for investment in domestic technologies: An analysis of the role of government in India. Technological Forecasting and Social Change, 71(8), 855-863.

Melville, P., Gryc, W., Lawrence, R. D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1275-1284). ACM.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Momeni, A., Rost, K. (2016). Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling. Technological Forecasting and Social Change, 104, 16-29.

Montgomery, D.R. (2007). Soil erosion and agricultural sustainability. Proceedings of the National Academy of Sciences, 104(33), 13268-13272.

Nelson, R.R., Winter, S.G. (1982). An Evolutionary Theory of Economic Change. Harvard University Press. Cambridge.

Omenn, G. S. (2006). Grand challenges and great opportunities in science, technology, and public policy. SCIENCE-NEW YORK THEN WASHINGTON-, 314(5806), 1696.

Oxford English Dictionary. Definition of Technology. URL: http://www.oed.com/view/Entry/198469?redirectedFrom=technology#eid

Parahina, V. N., Boris, O. A., Bezrukova, T. L., & Shanin, I. I. (2014). State support for creation and development of socially-oriented innovative enterprises. Asian Social Science, 10(23), 215.

Park, H. W., Leydesdorff, L. (2010). Longitudinal trends in networks of university–industry–government relations in South Korea: The role of programmatic incentives. Research policy, 39(5), 640-649.

Pimentel, D. (2006). Soil erosion: a food and environmental threat. Environment, development and sustainability, 8(1), 119-137.

Popper, R. (2010). iKNOW project. URL: http://www.foresight-platform.eu/wp-content/uploads/2010/06/5.4-Popper_iKNOW_EFP_final.pdf

Porter, A., Cunningham, S. (2004). Tech Mining: Exploiting New Technologies for Competitive Advantage. John Wiley & Sons, Inc.

Rao, D., McNamee, P., Dredze, M. (2013). Entity linking: Finding extracted entities in a knowledge base. In Multi-source, multilingual information extraction and summarization (pp. 93-115). Springer Berlin Heidelberg.

Rojas, R. (2013). Neural networks: a systematic introduction. Springer Science & Business Media.

Rotolo, D., Hicks, D., Martin, B. R. (2015). What is an emerging technology?. Research Policy, 44(10), 1827-1843.

Rotolo, D., Rafols, I., Hopkins, M. M., Leydesdorff, L. (2014). Scientometric Mapping as a Strategic Intelligence Tool for the Governance of Emerging Technologies. SPRU Working Paper SWPS 2014-10.

Royal Society (2009). Reaping the benefits: Science and the sustainable intensification of global agriculture, RS Policy Document 11/09, The Royal Society, London.

Russia 2030: Scientific and Technology Foresight. URL: https://prognoz2030.hse.ru/data/2014/12/25/1103939133/Prognoz_2030_final.pdf

Sanchez, D., & Moreno, A. (2008). Learning non-taxonomic relationships from web documents for domain ontology construction. Data & Knowledge Engineering, 64(3), 600-623.

Schatzberg, E. (2006). Technik Comes to America: Changing Meanings of Technology before 1930. Technology and Culture 47(3), 486-512.

Shen, Y. C., Chang, S. H., Lin, G. T., Yu, H. C. (2010). A hybrid selection model for emerging technology. Technological Forecasting and Social Change, 77(1), 151-166.

Simmhan, Y., Aman, S., Kumbhare, A., Liu, R., Stevens, S., Zhou, Q., & Prasanna, V. (2013). Cloud-based software platform for big data analytics in smart grids. Computing in Science & Engineering, 15(4), 38-47.

Simon, H. A. (1982). Models of bounded rationality: Empirically grounded economic reason (Vol. 3). MIT press.

Simon, H., Leker, J. (2016). Using Startup Communication For Opportunity Recognition—An Approach To Identify Future Product Trends. International Journal of Innovation Management, 20(08), 1640016.

Singhal, A. (2012). Introducing the knowledge graph: things, not strings. Official google blog.

Smith, K., & Haux, T. (2017). 8. Evidence-based policy-making (EBPM). Handbook of Social Policy Evaluation, 141.

Sokolov, A. (2011). Russian Critical Technologies 2015. European Foresight Platform. Foresight Brief No. 079. URL: http://www.foresight-platform.eu/wp-content/uploads/2011/04/EFMN-Brief-No.-79-Russian-Critical-Technologies-2015.pdf

Spitsberg, I., Brahmandam, S., Verti, M. J., & Coulston, G. W. (2013). Technology landscape mapping: At the heart of open innovation. Research-Technology Management, 56(4), 27-35.

Spitsberg, I., Verti, M. J., Brahmandam, S., & Coulston, G. W. (2015). Capitalizing on Emerging Technologies: A Path to Creating Opportunities in New Markets. Research-Technology Management, 58(4), 17-27.

Srinivasan, U. T., Cheung, W. W., Watson, R., Sumaila, U. R. (2010). Food security implications of global marine catch losses due to overfishing. Journal of Bioeconomics, 12(3), 183-200.

Tait, J. I., & Diallo, B. (2011). Future patent search. In Current challenges in patent information retrieval (pp. 389-407). Springer Berlin Heidelberg.

Tsui, E., Wang, W. M., Cheung, C. F., Lau, A. S. (2010). A concept–relationship acquisition and inference approach for hierarchical taxonomy construction from tags. Information processing & management, 46(1), 44-57.

Tversky, A., Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In Utility, probability, and human decision making (pp. 141-162). Springer Netherlands.

United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision, Data Booklet. ST/ESA/SER.A/377.

Van Zyl, A. (2011). Innovation in South Africa: the role of the Technological Innovation Agency. South African Journal of Science, 107(1-2), 1-2.

Viotti, E. B. (2002). National learning systems: a new approach on technological change in late industrializing economies and evidences from the cases of Brazil and South Korea. Technological Forecasting and Social Change, 69(7), 653-680.

Vishnevskiy, K., & Yaroslavtsev, A. (2017). Russian S & T Foresight 2030: case of nanotechnologies and new materials. Foresight, 19(2).

Wessner, C. W. (2005). Driving innovations across the valley of death. Research-Technology Management, 48(1), 9-12.

World Bank (2007). World Development Report 2008: Agriculture for Development, World Bank, Washington.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. IEEE transactions on knowledge and data engineering, 26(1), 97-107.

Xu, W., Liu, X., & Gong, Y. (2003). Document clustering based on non-negative matrix factorization. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 267-273). ACM.

Yoon, J. (2012). Detecting weak signals for long-term business opportunities using text mining of Web news. Expert Systems with Applications, 39(16), 12543-12550.

Yoon, J., Kim, K. (2012). TrendPerceptor: A property–function based technology intelligence system for identifying technology trends from patents. Expert Systems with Applications 39, 2927–2938.

Elena Tochilina
Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Moscow, Russian Federation. E-mail: etochilina@hse.ru