# FORESIGHT PROCESS BASED ON TEXT ANALYTICS Nataliya Pankratova, Volodymyr Savastiyanov

**Abstract**: The significant increase in number of information sources unfavorable affects on traditional foresight techniques not directly adapted to big data era. Without automation of knowledge processing the quality of final foresight product is significally dependent on human (experts, analytics) abilities. In the article the new process workflow is proposed using text analytics tools to support all stages of foresight. The proposed advanced model of fact extraction with modified rules is based on new workflow, which includes marking data with additional metadata, using automated classification and sentiment extraction techniques, data quality improving steps in addition to quantitative and qualitative analysis of data. The modified rule based model of knowledge extraction adapted to used toolkit is presented. Given approach were tested on supporting of foresight process in domain of agricultural development of Crimea region.

*Keywords*: foresight, decision making, textual analytics, sentiment analysis, knowledge society, data mining, DSS.

**ACM Classification Keywords**: H.5.3 Group and Organization Interfaces - Computer-supported cooperative work

## Introduction

We are living in knowledge society now [Eurofound, 2003]. The knowledge about short and long-term trends is the major driving forces not only for technologies but society. The knowledge about markets, user preferences and requirements is vital for strategy decision making [Eurofound, 2003; Pillkahn, 2008]. The scale how to utilize this knowledge is very varied: from studies how to make short, middle and long-term future less unpredictable till how to get future more favorable to human, organization, region or even country initiatives.

The one of the importunes issue which is another side of knowledge basis - we have enough information to make any decision using all sources in the only case we are capable of to grab, separate, aggregate, clean and process all this available knowledge [Pillkahn, 2008]. Foresight study is the known approach how to partly deal with this problem [UNIDO, 2005]. There are numerous agencies well known for their foresight studies and various authors or scientific groups leading their own foresight management strategies. The most part of these studies was founded in the period of data availability, access and aggregation technologies development (i.e. storage, databases software, OLAP technologies). So, today all mentioned technologies lead to the problem of big data in the world scale, what is really insuperable obstacle even with foresight methodology on the way to make decision horizon both desirable and trustworthy [Zgurovsky and Pankratova, 2007].

The problem is even harder due to widespread of communication technologies and social networks, blogs technologies and data mining tools. All this technologies drives not only the rising speed of data and knowledge interchange and processing, but also the morbid dependency of even small people's actions from information flows. Receiving false (opinion spam), inequality or specially wrong formed (misinformation) knowledge from huge

number of sources forces key figures, decision makers and experts select week strategy or act without strategy under uncertain condition and time shortage due to panic or inadequate reaction of community [Gladwell, 2001; O'Reilly, 2013]. Also fixed-date absence of any key figures actions is encouraging instability of situation [Martin et al, 2006], causing money, time, resources, power or even human losses [Zgurovsky and Pankratova, 2007; Gladwell, 2001].

So, taking into account information above, the big supply to any decision support methodology is the strategy to stay as long as possible well informed using advanced knowledge representations and modern data mining techniques during the time flow before critical time threshold set in.

#### 1. Current information model of foresight process

In this paper the information model of foresight process implemented in the information platform of scenario analysis (IPSA) is using, which combines classical theory of foresight adopted to Ukrainian conditions and advanced mathematics in expert's opinion processing [UNIDO, 2005; Zgurovsky and Pankratova, 2007]. The analytical level of IPSA consists of numerous organizational stages, methods of quality and quantity analysis, and special mathematical methods of expert's opinion aggregation for all cases of typical input and combination of initial information for particular domain [Zgurovsky and Pankratova, 2007]. The process of current IPSA operation is shown on Figure 1.



Figure 1. Operation of current foresight information platform of scenario analysis

The process of operation is based on scenario analysis involving expert's knowledge processing through the mathematical instrumentality of qualitative and quantitative methods conforming expert judgements in a number of steps guided by organizational procedures. These judgements are reflecting researched subject area with trends and estimates [Zgurovsky and Pankratova, 2007]. A group of analysts involved into foresight on a regular basis are managing the plan of foresight process and supply expert's panels with initial information. The group of analysts also interacts with key figures that are the end users of foresight. The final product of foresight is a set of scenario alternatives with a priory chosen time scale in the future and concerning a priory chosen subject area.

All steps of foresight are performing with a help of IPSA, which is both an information portal for end users and interface to the processing libraries with qualitative and quantitate mathematical methods.

The integrity of knowledge base and researched subject domain coverage in foresight process is a question of the group of analysts' proficiency and is not supported by any computer aided technology on the mentioned above stage of IPSA development.

#### 2. Metadata used in current information model of foresight process

Current foresight process reflected in information platform of scenario analysis utilizes some special information entities for intersteps knowledge exchange. These entities are described in table below (Table 1).

| Nº | Metadata name     | Metadata description and source                                     |  |  |
|----|-------------------|---|--|--|
| 1  | Time scale        | On organizational stage   |  |  |
| 2  | Goal              | Goals of foresight  |  |  |
| 3  | Idea              | Sources: environmental scanning, brainstorm                         |  |  |
| 4  | Cluster of ideas  | Sources: environmental scanning                                     |  |  |
| 5  | Expert's estimate | Sources: Delphi, Saaty method, cross-impact, morphological analysis |  |  |
| 6  | Key technology    | Sources: environmental scanning, brainstorm, morphological analysis |  |  |
| 7  | Driving force     | Sources: environmental scanning, brainstorm, morphological analysis |  |  |
| 8  | Scenary           | Sources: scenarios, STEEPV (STEEEPVA)                               |  |  |
| 9  | Roadmap           | Sources: Roadmapping  |  |  |

 Table 1. Metadata of foresight process

The current metadata are descriptive by nature and allow only building knowledge storage system. Some of produced by qualitative methods entities are semistructured [Buneman, 1997] and physically are stored as textual blocks with numbers, facts, conclusions or other knowledge. Further in the article this metadata would be referred to as metadata of type I (or just *metadata-1*).

# 3. Big-data-ready information process of foresight

Big data readiness in foresight is mentioned in sense of how to deal with rapidly increasing number of new knowledge sources under time pressure. There is a strong need to enforce all participant of foresight process (i.e. analysts, experts, key figures and end users) to stay at equal level of being kept informed during foresight process. At the same time humans are limited in their decision making ability due to limitations in their ability to retain and process information [Simon, 1956]. So there is a strong need in the automated tools for relevance

analysis of new knowledge generated in every current period of time: previously generated scenario, extracted trends, facts, key figures, drivers and other knowledge; scenario alternatives which were not chosen by decision makers; hypothetical events from data scientists, future studies, science fiction and other sources.

We should take into account the text-based nature of mostly (up to 80%) current and future sources of qualitative business data (news, social networks, blogs, reviews, reports, transcripts, legal papers, emails, etc.) [Berry and Linoff, 2011; Ryan and Bernard, 2000] and currently existing experience storage types in information platform of scenario analysis (see Table 1). The most appropriate tools for processing that information sources is text analytics toolkit. Using advanced automated toolkit for knowledge mining brings new data and metadata into the foresight process for sake of improving scenario trustworthiness and quality.

Text analytics toolkit today is reach on various technics of data aggregation, classification, fact extraction, topic mining, ontology mining, features or aspects extraction and opinion mining [Berry and Linoff, 2011; Chakraborty et al, 2013]. Some of these tools could directly improve certain qualitative methods of foresight some involved into other methods could support the foresight process to improve the quality of results.

# 4. Advanced information model of foresight process

Advanced information model of foresight strategy additionally includes specially designed knowledge database, module of information quality assessment, and module of foresight process supporting (Figure 2).



Figure 2. Operation of advanced foresight information platform of scenario analysis

The knowledge base is supposed to store both primary and processed structured data, such as sets of knowledge and facts, trends, metadata, structural and functional relationships, etc.

Information quality assessment module provides an assessment of the consistency and relevancy of knowledge representation in database regarding to subject domains, which are formed by all gathered knowledge: data from real objects and systems, hypothetical objects, systems and notion about them in the representation of the

experts, also the signals from the external environment in certain time slices of dynamic changes in the behavior of the studied system and its environment during the process of foresight.

Foresight process supporting module combines forming of organizational procedures by schedule (queue of procedures) with automated notifications to the participants that would have been sent on a base of automated information quality assessment. Forming a queue of procedures is necessary: to enhance on regular basis the reliability and trustworthiness of scenarios by improving the quality of a priori mined information and knowledge; to adjust the foresight process to new knowledge utilizing during its course.

### 5. Strategy of foresight process support

Information platform of scenario analysis receive structured and semistructured data objects *Obj* reflecting real world on their input from the informational space (i.e. all available digital sources of knowledge):

# Obj=< Obj<sub>St<i></sub>, Obj<sub>SSt<j</sub>>, i=1,I, j=1,J.

On the data standardization step all data is storing into the relational database as raw data. Structured  $Obj_{St}$  data would be useful for various statistical processing and direct representation with OLAP or visual analytics technologies on later stages. All semistructured  $Obj_{SSt}$  data, mostly in text representation (or XML with metadata), is also storing as RAW data into the relational database with minimal metadata if available:

Obj<sub>SSt<j></sub> = < ID, Body, CrawIDte, Author, SourceID, SourceDte >,

where *ID* is unique document identifier, *Body* - is raw data, *CrawlDte* - is crawl date, *Author* is author of source information, *SourceID* is unique document identifier in source stream (i.e. URL, etc), *SourceDte* - date extracted from source.

In the next step stored semistructured data is processing by *information quality assessment module* to retrieve additional metadata of type I and II. The scheme of the strategy is shown on Figure 3.

According to the mentioned strategy on the first stage all data is processing by *information quality assessment module* for sake to enrich the mined knowledge with additional metadata could be extracted from raw data and combine it coupled with statistical data into frame based hierarchy which reflects the examining subject area(s) [Liu, 2012]. The *information quality assessment module* is combining text analytics framework based on classification, fact extraction and sentiment analysis tools. There are a lot of tools and approaches how to do this tasks with help of rule based algorithms or machine learning [Liu, 2012]. Currently the mentioned strategy had been tested with help of SAS(R) Text analytics toolkit [Chakraborty et al, 2013].

Two level classifications are used: based on general classifiers to expose the density of subject domains covering by sources and foresight advanced metadata classifiers with fact extraction step.

As a general classifier could be used IPTC classifier from the standard SAS(R) package. There are also a number of others classifiers according to the different subject areas available or could be built [Ryan and Bernard, 2000; Moldovan and Girju, 2001]. General classification is very important step to identify possible interrelations between examining subject areas in available sources. All extracted metadata is storing to the knowledge base.

A special approach to combining classification with fact extracting step is using to identify special metadata of type II (metadata-II) (Figure 4) in advanced foresight process. It is using to separate following entities from input data:

- Goal phrases;
- Structure and interrelations of entities;
- External environment influence and interaction points with external trends;
- Time horizon of trends and gathered knowledge;

- Problem identification;
- Effect detection with sorting on future effects and past facts, suggestion.



Figure 3. Scheme of foresight process support strategy

The mentioned approach is built on a base of general model of facts extraction from natural language texts [Simakov et al, 2006], supplemented by a set of rules for sentiment extraction and adopted to using SAS(R) text analytics toolkit. The modified model with extended set of rules is given below:

$$E = \langle T, V, a \rangle, \tag{1}$$

where T are all text objects (documents) from input data, V - all rules, a - logical function  $a(t_i, v_j)$  which is «True» if  $t_i$  satisfies  $v_j$ .

The difference of the proposed model from general model of facts extraction from natural language texts [Simakov et al, 2006] is that text is not a sequence of words, but sequence of paragraphs, sentences and then words:

$$\begin{split} t &= par_1 par_2 \dots par_{|N|}, \\ par &= sent_1 sent_2 \dots sent_{|M|}, \\ sent &= wrd_1 wrd_2 \dots wrd_{|K|}, \\ wrd_k &\in \{Wrds, Pnkt, POS\_tag, *, Aa\}, \end{split}$$

where *Wrds* are list of words, *Pnkt* is punctuation, *POS\_tag* are part-of-speech tags, \* is any single word, *Aa* is any word starting from capital letter. As in [Simakov et al, 2006] we could present a text fragment in a view of:

 $t=t_1 + t_2 + \dots + t_j$ .



Figure 4. The structure of additional metadata in advanced foresight process

Needed set of fragments  $T_{ri}^q = \{t\}$  in every text is covered by pattern  $r_i^q$  and all text is covered by all possible fragments  $T_r = U T_{ri}^q$ . Pattern  $r_i^q = \langle c, e, d \rangle$ , where *c* is lexical restriction, *e* is exception from lexical restriction, *d* is rule scope and  $d \in \mathbf{N}$ ,  $q \in \{1,2,3,4,5,6,7,8\}$ . For convenience we limit  $0 < d \le 170$ . Exception from lexical restriction *e* is implicitly equal empty set in some rules to make the patterns compatible to the SAS(R) text analytics toolkit.

$$\forall p \ s_{p} \in s \ \forall t \in T_{n's}^{5} \rightarrow \begin{cases} \exists w_{ij} \in \{w_{i}\}, \forall j \ w_{ij} \in \{t\}, j \geq 2 \\ \mid t \mid \geq 2, \forall a, b \in \{j\}, a < b \\ \forall j \ w_{ij} \in c, \forall j \ w_{ij} \notin e, e = \emptyset \end{cases}$$

$$\forall p \ s_{p} \in s \ \forall t \in T_{n's}^{6} \rightarrow \begin{cases} \exists w_{ij} \in \{w_{i}\}, \forall j \ w_{ij} \in \{t\}, j \geq 2 \\ \mid t \mid \geq 2, \forall a, b \in \{j\} \ dist(w_{ia}, w_{ib}) \leq d, a < b \\ \forall j \ w_{ij} \in c, \forall j \ w_{ij} \notin e, e = \emptyset \end{cases}$$

$$\forall p \ s_{p} \in s \ \forall t \in T_{n's}^{7} \rightarrow \begin{cases} \exists w_{ij} \in \{w_{i}\}, \forall j \ w_{ij} \in \{t\}, j \geq 3 \\ \mid t \mid \geq 3 \\ \forall a, b, c \in \{j\}, a < c < b, w_{ia}, w_{ib} \in c, w_{ic} \notin e \end{cases}$$

$$\forall p \ s_{p} \in s \ \forall t \in T_{n's}^{8} \rightarrow \begin{cases} \exists w_{ij} \in \{w_{i}\}, \forall j \ w_{ij} \in \{t\}, j \geq 1 \\ \mid t \mid \geq 1 \\ \forall j \ w_{ij} \notin e, e = \emptyset \end{cases}$$

 $T_{ri}$ <sup>8</sup> pattern currently not presented in the used toolkit (e is explicitely set to  $\emptyset$ ), but it is added to the model to theoretically fulfill also possible negation pattern. Also in other rules e is explicitely set to  $\emptyset$  to satisfy used software toolkit limitations. The final extraction rule in set of rules V has a form:

$$v_i = (\{\}, s, w), j \ge 1,$$

where argument name  $\arg_j \in \{\emptyset, \{a \neq z\}\}$ , s is sentiment, w is weight of the rule, s  $\in \{-1, 0, 1\}$ ,  $w \in \mathbb{R}$  (taken  $w \in \mathbb{R}$ )

(0;10]). Also, there could be modifications:  $s \in \{+,=,-\},\{-2;-1;0;1;2\}, \{-3;-2;-1;0;1;2;3\}$  according to human recognition ability [Miller, 1956]. When  $arg_j = \emptyset$  there is no facts to be extracted, only matching is needed. The main advantage of proposed model on [Simakov et al, 2006] is the principle how the sequence of fact could be extracted. According to the definition, sequence of facts could be returned into different arguments or sequentially extracted and concatenated into the one single argument. There could be present or absent prefixes and postfixes around any fact should be extracted.

With a help of sentiment colored rules it is possible to identify positive, negative or neutral trends in external environment, effects of possible action plans of authorities, identify problems. To identify them six predefined conceptual categories for sentiment rules classification with fact extraction ability could be used: simple sentiment word or phrase; decreased and increased quantity of an opinionated item; high, low, increased and decreased quantity of a positive or negative potential item; desirable or undesirable fact; deviation from the norm or a desired value range; produce and consume resource and waste [Liu, 2012]. This classification is also useful for automated opinion extraction from expert's conclusion or scenario essay in free form (text form).

The heart of any knowledge, sentiment, fact extraction and classification system is the set of available subject domain's feature hierarchies (product features) could be used with extraction rules. There is a known technique to extract product features, which imply opinions [Liu et al, 2011]. It allows automatically extract possible features

from subject domain, which are declaring desirable or undesirable facts. This advanced technique is not utilizing in current work, only automated extraction of noun phrases with manual sorting and filtering for Russian and Ukrainian languages was done instead. The words and phrases mined as products and features were used in sentiment rules for subject domains processing. The rules were written using SAS(R) sentiment analysis studio according to the proposed model (**100**) with wide range of special Boolean rule modifiers («OR», «AND», «SENT», «ORD», «DIST», «ORD\_DIST», «UNLESS») and special rules types available [Reckman et al, 2013].

Processed and stored into the knowledge base data then enters on the input of foresight process supporting (FPS) module. As a part of automated tools of FPS there are *database of rules describing foresight methods* usage and *database of knowledge conflict resolving with knowledge incompleteness filling rule set.* 

Database of rules describing foresight methods usage consist of hierarchy of frames, which slots are describing input, output and control of every method and rules of process activation which are required to fulfill the slot to complete every particular frame. In other words this database consists of workflow and all metadata of particular foresight process, in this case according to Ukrainian Foresight Program [Zgurovsky and Pankratova, 2007]. Currently all methods of that foresight approach get strong formalization, allowing doing the automation [Zgurovsky and Pankratova, 2007].

Database of knowledge conflict resolving with knowledge incompleteness filling rule set help to complete the frame hierarchy, resolve conflicts of knowledge and discover new knowledge and relations. One part are just «if ... else» rules which reflect the process to complete all slots of frames. If there is no possibility to find some knowledge the process ends forming the message to particular expert (experts group) to fulfill incomplete knowledge with help of analysts' personnel. It should be recalled that foresight is not isolated from human computational process, so recall to expert judgement or utilizing human brain resources is normal foresight workflow.

Another part of rules are knowledge conflict resolving rules. There could be extracted group of facts that consist of opinions, subjective information, not proven by authorities or government instances numbers, expectations, external trends and so on. All this fact could conflict or strengthen each other. This conflict could be resolved manually by experts or due to rule sets, in comparison with gualitative data and forecasts. Also there are known number of techniques to support or reduce expensive brainwork. In the work of Lerman and McDonald [Lerman and McDonald, 2009] shown the technique how to summarize mined opinion in case of different product features. that is mean we could summarize mined data to show the expert group «bird view» picture in form of colored radar chart with marked divergences, exclusive features, numbers and polarities of sentiments. Other technique to summarize mined opinion and resolve conflicts is direct usage of foresight methods: Delphi [Pankratova and Malafeeva, 2012], Saaty [Pankratova and Nedashkovskaya, 2013] and others [Zgurovsky and Pankratova, 2007]. In addition the mined by information quality assessment module into knowledge base products and features could be compared with product and features from the ontologies with structured information that are available online. Lu at al. [Lu et al, 2010] proposed methods for selecting a subset of aspects from the ontology that can best capture the major opinions, including size-based, opinion coverage-based, and conditional entropy-based methods. They also have done the approach to order aspects, give measures for quantitative evaluation of both aspect selection and ordering and give the way to discover new aspects. In this paper all mined products and features are only compared with industry ontology by set intersection to find out coverage of features with subject domain. In future work implementing of mentioned techniques and approaches is planned.

Step by step by discovering all black holes and eliminating conflicts in knowledge base the foresight process through sequences of foresight method usage is heuristically forming. In this algorithm two important cycles to form final schedule should be separated: sequence of foresight method usage is forming, interaction of foresight method depending on the input data is forming. To avoid infinite loops of knowledge mining with open set of

knowledge horizon expert methods are used to terminate the loop. Within every chosen method a priory relevancy level is given or could be accepted during the process.

## 6. Foresight of agricultural development of Crimea region

Given approach were tested on supporting of foresight process in domain of agricultural development of Crimea region. The source documents from Crimea authorities, expert opinions, interviews and other provided information sources were collected, structured and processed. In addition supplementary data was retrieved from reviews of agricultural development central parts of Ukraine, Russia and Belorussia, official programs of development from Russia and Belorussia, existing official programs of some agricultural branch development from Ukrainian ministry of agriculture, transcripts from Ukrainian parliament sessions, reviews of agricultural technologies and news sources. The CTEA-classifier<sup>1</sup> were used as a general classifier on a first step of classification. Initial rules for CTEA classifier were written by group of analysts manually. The processed data were put into knowledge base as hierarchy structured facts (Table 2).

| Parameter   | Q-ty  | Hierarchy type for storage |
|---|-------|----------------------------|
| All objects   | 25360 | Static (Structural)        |
| Objects relevant to CTEA                                      | 406   | Static (Structural)        |
| Objects in trends in agricultural domain (Ukraine + Crimea)   | 11991 | Functional                 |
| Objects in problems in agricultural domain (Ukraine + Crimea) | 2000  | Functional                 |
| Objects in goals in agricultural domain (Ukraine + Crimea)    | 1862  | Functional                 |
| Non-unique objects in agricultural domain                     | 7060  | Static (Structural)        |
| Technologies in agricultural domain                           | 378   | Functional                 |
| Problems in agricultural domain                               | 225   | Functional                 |
| Trends in agricultural domain                                 | 1385  | Functional                 |
| Goals in agricultural domain                                  | 112   | Functional                 |

 Table 2. Knowledge base statistic (domain of agricultural development of Crimea region)

After all stages in knowledge base were formed Trends group, Problems group, Goal group, Technologies group, Key Factors group with given statistic:

- Final Trends group consist of 12 major trends;
- Final Problems group consist of 143 significant problems;
- Final Goals group consist of 82 goals;
- Final Technologies group consist of 253 technologies;
- Key Factors group consist of 35 factors.

<sup>&</sup>lt;sup>1</sup> State Standard of Ukraine approved a new classifier DK 009:2010 "Classification of types of economic activities" (CTEA), which corresponds to the European standards and requirements.

Taking a part in similar foresight researches of the same scale as a member of foresight analysts group the author has noticed that new approach saves up to 35% of time on knowledge processing in comparison with traditional foresight process. Also the modified foresight process was given new metrics in mean of knowledge extraction dynamic progress. It is possible to introduce new foresight process's KPI based on knowledge quality and knowledge coverage in addition to traditional organizational KPIs of traditional foresight. The integrity of knowledge base and researched subject domain coverage in foresight process is a question of the group of analysts' proficiency and is not supported by any computer aided technology on the mentioned above stage of IPSA development.

#### Conclusion

The significant increase in number of information sources unfavorable affects on traditional foresight techniques not directly adapted to big data era and need to implement knowledge mining tools into the process. The organizational nature of foresight combined with utilizing of experts' opinion leads to unpredictable increase of time and costs needed to knowledge convergence in organization with foresight process in big data era. To contradict described problem modified structure of foresight process were proposed. Strategy of foresight process support using text analytics tools was implemented.

The proposed advanced model of fact extraction with modified rules is based on new strategy, which includes marking data with additional metadata, using automated classification techniques in addition to quantitative and qualitative analysis of data. Classification, fact extraction and sentiment analysis tools help to find new interrelations of available knowledge and significantly increase and equalize the level of awareness of foresight participants which leads to increasing confidence level of final foresight product.

The approaches were tested on supporting of foresight process in domain of agricultural development of Crimea region. Implementing foresight-supporting tools based on text analytics could save up to 35% of time on knowledge processing. In addition it is possible to introduce new foresight process's KPI based on knowledge quality and knowledge coverage complementary to traditional organizational KPIs of traditional foresight.

#### Bibliography

- [Berry and Linoff, 2011] Berry, Michael J. A., and Gordon S. Linoff. Data Mining Techniques For Marketing, Sales, and Customer Relationship Management, 3rd edition. Wiley Computer, 2011.
- [Buneman, 1997] Peter Buneman Semistructured data. In: Proc. ACM Symposium on Principles of Database Systems, pp. 117-121, Tucson, AZ., Abstract of invited tutorial, 1997.
- [Chakraborty et al, 2013] Goutam Chakraborty, Murali Pagolu, Satish Garla. Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS, SAS Institute Inc., 2013.
- [Eurofound, 2003] European Foundation for the Improvement of Living and Working Conditions. Handbook of Knowledge Society Foresight. 2003. <<u>http://www.eurofound.europa.eu/pubdocs/2003/50/en/1/ef0350en.pdf</u>>.
- [Gladwell, 2001] Gladwell, M.: The Tipping Point. How Little Things Can Make A Big Difference. Boston: Little, Brown & Company, 2001.
- [Lerman and McDonald, 2009] Lerman, Kevin and Ryan McDonald. Contrastive summarization: an experiment with consumer reviews. in Proceedings of NAACL HLT 2009: Short Papers. 2009.
- [Liu, 2012] Liu B., Sentiment Analysis and Opinion Mining, ISBN-10: 1608458849, ISBN-13: 978-1608458844, Morgan & Claypool Publishers, 2012
- [Liu et al, 2011] Zhang, Lei and Bing Liu. Identifying noun product features that imply opinions. in Proceedings of the Annual Meeting of the Association for Computational Linguistics (short paper) (ACL-2011). 2011b.

- [Lu et al, 2010] Lu, Yue, Huizhong Duan, Hongning Wang, and ChengXiang Zhai. Exploiting Structured Ontology to Organize Scattered Online Opinions. In Proceedings of Interntional Conference on Computational Linguistics (COLING-2010). 2010.
- [Martin et al, 2006] Martin, B., Cashel, C., Wagstaff, M., & Breunig, M. Outdoor Leadership: Theory & practice. Champaign, IL: Human Kinetics, 2006.
- [Miller, 1956] George A. Miller, "The Magical Number Seven," Psychological Review (March 1956), vol. 6 j , no. 2., 1956.
- [Moldovan and Girju, 2001] Dan Moldovan and Roxana Girju, An Interactive Tool For The Rapid Development of Knowledge Bases. In International Journal on Artificial Intelligence Tools (IJAIT), vol 10., no. 1-2, March 2001.
- [O'Reilly, 2013] O'Reilly O'Reilly Media, Inc., Big Data Now: Current Perspectives from O'Reilly O'Reilly Media, Inc., O'Reilly Media, 2013, ISBN: 978-1-449-37420-4.
- [Pankratova and Malafeeva, 2012] N.D. Pankratova, L.Y. Malafeeva Formalizing the consistency of experts' judgments in the Delphi method // Cybernetics and Systems Analysis: Volume 48, Issue 5 (2012), Page 711-721, 2012.
- [Pankratova and Nedashkovskaya, 2013] Pankratova N., Nedashkovskaya N. Estimation of Sensitivity of the DS/AHP Method While Solving Foresight Problems with Incomplete Data // Intelligent Control and Automation, v.4, №1. – 2013. -P. 80-86
- [Pillkahn, 2008] Ulf Pillkahn, 'Using Trends and Scenarios as Tools for Strategy Development', Siemens, 2008.
- [Reckman et al, 2013] Hilke Reckman, Cheyanne Baird, Jean Crawford, Richard Crowell, Linnea Micciulla, Saratendu Sethi, and Fruzsina Veress. Rule-based detection of sentiment phrases using SAS Sentiment Analysis, Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, pages 513-519, Atlanta, Georgia, June 14-15., 2013.
- [Ryan and Bernard, 2000] Ryan, G., & Bernard, R. Data management and analysis methods. In N. Denzin& Y. Lincoln (Eds.), Handbook of Qualitative Research (pp. 769–802). Thousand Oaks, CA: Sage, 2000.
- [Simakov et al, 2006] Andreev A, Berezkin D., Simakov K. The model of fact extraction from natural language texts and the learning method. RCDL 2006, <http://www.ixlab.ru/pub/docs/RCDL\_2006\_1.pdf>

[Simon, 1956] Simon, H. A. Rational choice and the structure of the environment. Psychological Review, 63,129-138.,1956.

[UNIDO, 2005] UNIDO, TECHNOLOGY FORESIGHT MANUAL, Vienna, 2005, ISBN 978-3-89578-304-3

[Zgurovsky and Pankratova, 2007] Zgurovsky, Mikhail Z., Pankratova, N.D., System Analysis: Theory and Applications, Springer, 2007, ISBN 978-3-540-48880-4.

## **Authors' Information**



**Nataliya Pankratova** – DTs, Professor, Deputy director at Institute for applied system analysis, National Technical University of Ukraine "KPI", Av. Pobedy 37, Kiev 03056, Ukraine; e-mail: <u>natalidmp@gmail.com</u>

Major Fields of Scientific Research: System analysis, Theory of risk, Applied mathematics, Applied mechanics, Foresight, Scenarios, Strategic planning, information technology



**Volodymyr Savastiyanov** – Researcher at Institute for applied system analysis, National Technical University of Ukraine "KPI", Av. Pobedy 37, Kiev 03056, Ukraine; e-mail: <u>vvs@iasa.com.ua</u>

Major Fields of Scientific Research: system analysis, foresight, future studies, text analytics, sentiment analysis, data mining, knowledge mining, statistical analysis, artificial intelligence.