

---

## **A Novel Graph Decomposition Approach to the Automatic Processing of Poorly Formalized Data: Innovative Ideas – A Management Case Study**

---

N.V. Patrikeeva<sup>1</sup>, V.N. Babeshko<sup>2</sup>, E.A. Kamenev<sup>3</sup>, V.E. Podolskiy<sup>4</sup>

**Abstract:**

*In the following paper we present a novel approach to unstructured data processing by imposing a hierarchical graph-based structure on the data and decomposing it into separate subgraphs according to optimization criteria. In the scope of the paper we also consider the problem of automatic classification of textual data for the synthesizing the hierarchical data structure. The proposed approach uses textual information on the first stage to classify ideas, innovations, and objects of intellectual property (OIPs) to construct a multilayered graph. Numerical criteria are used to decompose constructed graph into separate subgraphs. In the scope of the research we apply the developed approach to the innovative ideas in a management case study. The research has been conducted in the scope of a joint research project with financial aid of Ministry of Education and Science of Russian Federation RFMEFI57314X0007.*

**Key Words:** *Poorly Formalized Data, Graph Theory, Management of Ideas, Data Structures*

---

<sup>1</sup> Consulting Director, IBS, ibs@ibs.ru

<sup>2</sup> Senior Consultant of the Department of Education, PhD (Technical Sciences), Associate Professor National University of Science and Technology MISiS; IBS, min@misis.ru

<sup>3</sup> Deputy Director of the Innovation and Technology Centre, Moscow Institute of Physics and Technology, kamenev.ea@mipt.ru

<sup>4</sup> 5th-Year Master's Degree, Bauman Moscow State Technical University, v.e.podolskiy@gmail.com

---

## **Introduction**

In most cases, automatic processing can only be conducted on a structured and normalized data set. If the data set is unstructured, then most of data processing algorithms will fail to provide adequate results. The selection of a formalized model for the data is a process that is conducted manually. The main reason behind using manual approach is that even the state-of-the-art text mining algorithms are incapable of processing complex semantics of non-formalized and unstructured textual data.

On the other hand, important patterns can be extracted only from the structured data. There are varieties of methods that provide basic data cleansing and structuring services: clusterization algorithms, normalization techniques, data cleansing, etc. A general automatic formalized model construction can be realized for more complex cases. This becomes possible due to high abstraction power of graph theory. However, the graph theory is to be used carefully – trying to apply this theory without considering the research field features can lead to models that lack needed level of details for further processing. Common approaches to the processing of connected and poorly structured data are based on graphs theory.

Poorly formalized and unstructured data is one of the largest segments in the data processing. Experts agree that almost 80 to 85 percent of business-relevant information originates in unstructured form. The manual processing of unstructured data is costly and time-consuming. To interpret unstructured information such techniques as natural language processing (NLP), data mining, text analytics are used. Further, the patterns found can be organized and structured using mathematical graphs. Different kinds of graphs can be used to represent different features of data.

The field of innovation management was selected to apply the results of the research, because the development of this novel approach to data structuration is a part of work on universities' innovation life cycle model (ILCM) and innovation management system. The basic components of innovation management are ideas and innovations.

For example, the idea can be represented by unstructured textual descriptions and a group of illustrations. It is important to classify the idea based on its description. The manual classification becomes difficult in case of large number of ideas. The idea can be provided in the form of terms-rich description. Such description contains specific key words. There are many techniques to identify such words and phrases, however these approaches are leaving aside the problems of textual structure and semantics. There are other attributes that must be used in order to structure ideas. Considering all these attributes, the process of ideas management becomes fairly complicated, thus automatic tools become highly important for efficient innovations management.

Back in 1986, Andrew H. Van de Ven has outlined four central problems in innovation management (Van de Ven, 1986): managing attention; managing ideas into good currency; managing part-whole relationships; institutional leadership. According to Van de Ven's 2004 paper, these cornerstone problems remain intact and continue to heavily influence the performance of innovation management process (Van de Ven and Engleman, 2004). In our work we will concentrate on ideas and innovations automatic structuration tasks which are in the scope of outlined problem of managing ideas into good currency. One of the major parts of this problem is to correctly define the most promising and optimal (in respect to given resources) idea. One of the best ways to estimate an idea would be to compare it with implemented analogues. The estimation can become rather complicated when considering ideas interconnections and dependencies on each other or on specific innovation.

In the second section of the paper we will provide a literature review of automatic and semi-automatic graph-based decision-support approaches dealing with the processing of poorly formalized data as well as a basic review of innovation management literature. The third section will be dedicated to the description of innovation management formalization using ideas, innovations and objects of intellectual property (OIP) passports. In the fourth section we consider a novel hypergraph automatic construction approach using text mining and machine learning algorithms for the purpose of imposing a structure on unstructured innovation management data. Finally, this section encompasses the description of novel approach to constructed hypergraph decomposition using value by resource unit approach. The fifth section of the paper includes a case study – the modified approach to the automatic structuration of unstructured textual data in hypergraph is used on a test set of more than 6 000 ideas. The sixth section covers discussion of the proposed approach and its applications. The last section of the paper discusses the future work.

## **Literature review**

The international experience in the field of innovation management is wide enough to cover most of the innovation management models that are used today in commercial organizations as well as in non-profit organizations (Hulla and Liob, 2006). The standardization in this field started with the development of Manual Frascati by OECD (OECD, 2002). The innovation activity estimation on the macro-level is commonly conducted based on the systems of indicators developed by the Commission of the European Communities (CEC). This methodology was adopted by the members of European Union to compare their KPIs with USA and Japan (Sitenko, 2010). CEC indicators are based on Manual Frascati. At the beginning, the CEC system included 20 indicators divided into four groups (European Commission, 2005): human resources, new knowledge generation, transfer and use of knowledge, financing of innovation and results of innovation. In 2005, the CEC indicators system expanded up to 26 indicators. All these indicators form a complex indicator -

---

Summary Innovation Index (SII). This index represents the innovation activity estimation in a specific country. In order to represent the results of science and innovations development, OECD introduced scientific and innovation profile for each country (OECD, 2008).

There are two basic models that were outlined in research of American institutes: diffuse model and intra-organizational model (Batrutdinov and Fedoseev, 2008). While the diffuse model characterized the development of innovations management on a macro level, the intra-organizational model described features of innovations management in a separate company or in an authority. The modern research of innovation processes is based on two hypotheses – the hypothesis of “technology push” and the hypothesis of “the pressure of market demand”. The first model is directed from science to market, while the second one is its opposite. The model of technology push is concerned with a chain of transformations of idea into a commercial product (Dodgson and Rothwell, 1994). The second model is also linear but is based on initial market research. However, both models are highly deterministic and thus they do not consider the probabilistic nature of innovative process.

The second generation of innovation management models was oriented towards market demand. However, these models had some drawbacks based on their reactivity - such models were able to describe incremental innovations only. In 1970, linear models were still in use as a special case of general process which in turn unified science, technology and market. The importance of all three components was outlined by Rothwell (Rothwell, 1994) and Freeman (Freeman, 1995). According to Rothwell, innovation process includes feedbacks while the model mostly remains linear. Kline and Rosenberg noted that the innovation development is a complex and mostly chaotic process, so linear models cannot describe innovation process well. They proposed a chain model and noticed the influence of knowledge management (Kline and Rosenberg, 1986). Iterative models were developed by Japanese researchers in order to describe the innovative process as a kind of activity that penetrates functional structures of organization (Imai et al., 1985)

The network of interactions is a foundation of fourth generation of innovative models. Functions of innovative system are distributed across this network. The fifth generation model is an integrated network model that aims to describe the complexity of the innovative process. This model is based on knowledge that is contained in the organizations as well as in the inter-organizational processes. The main feature of the network model is its ability to comprehend the environmental influence and effective communications with the environment. The fifth model is a direct successor to the fourth model. The main difference is that the technology of technological change is dependent on information technologies (Trott, 1998).

The sixth generation of innovative models encompasses open innovations paradigm. According to this paradigm, innovation is seen as a distributed network process which focuses not only on creation of innovations inside a company but on assimilating outside ideas too. The research field becomes wide, and thus it becomes easier to find and implement ideas. Gassmann, Enkel, and Chesbrough proposed a model named “Open innovation with three core processes archetypes” (Enkel et al., 2009). This model consists of three stages: outside-in process, inside-out process, and a coupled process. In order to embed these processes, the company must demonstrate following characteristics: absorptive capability, multiplicative capability, and relational capability. Evolution of the proposed model also encompasses nine perspectives: spatial perspective, structural perspective, user perspective, supplier perspective, leveraging perspective, process perspective, tool perspective, institutional perspective, and cultural perspective (Gassman et al., 2010; Sieg et al., 2010).

Despite a large number of developed models, there is still a doubt of whether specific model can efficiently describe innovative process or not. E.g. Mahdi outlined that most of the described models are deterministic (Mahdi, 2002). Thus, most of the companies tend to use their own simplified models. One of the examples of such models is a stage-gate process that determines the sequence of stages. One of the most prominent models is the Cooper’s selection model which is based on the ideas selection process (Cooper, 2001). Cooper also outlined major characteristics of the next generation of models which include (Cooper, 1994): adaptability and conditional decisions, focus, agility. One of the most recent models is a strategic networking model which is based on innovative communications processes between innovators and consumers. Overall, it can be shown that it is impossible to construct a single universal model which will cover every feature of innovation management systems.

While networks models are becoming a solid instrument of innovations processes research and engineering, it is important to gain an instrument, suitable for the analysis of these models. It appears that graph theoretical and matrices models are well-suited for this task.

One of the first works in the field of application of graph theory to decision support dates back to 1983 – H.-D. Haustein and M. Weber proposed a methodology based on a decision tree model (a subclass of graph) and quantification of risks in order to select the most promising innovative project (Haustein and Weber, 1983). The next widely known case of application of graph theory to the analysis of innovations occurred in 1995 – E. Santarelli outlined that directed graphs and adjacency matrices can be used to provide the economic analysis of innovations (Santarelli, 1995). Author described a technological regime as a unipathic graph (a tree) consisting of vertices, representing innovations on different levels of development. Similar graph concept has been proposed as a part of a framework for analysis based on cost of innovation processes (Zygiaris, 2009).

Another solid graph-based approach to innovations estimation and management was introduced by M. D. König and S. Battiston. According to the proposed approach, innovations are formed through so-called dynamic innovations networks in which agents compete for the most valuable knowledge for production, while knowledge can only be created through collaborations and knowledge exchange (König and Battiston, 2009). The dynamic innovations network is represented by undirected graph where every node is considered as an agent having its own utility. Authors also proposed a method for utility estimation based on agents' connectivity as well as the cost parameter (König et al., 2008).

R. V. Rao proposed a decision-making methodology for evaluation of alternative flexible manufacturing systems based on digraph and matrix methods (Rao, 2006). The digraph and its matrix representation are used to construct the index that is used to select a specific configuration of flexible manufacturing system. M. Darvish, M. Yasaei, and A. Saeedi applied similar digraph and matrix methods to solve a contractor selection problem (Darvish et al., 2009). As authors stated, the contractor selection problem includes several interconnected attributes (example model includes 9 attributes), thus making graphs a reasonable instrument for its solution. S. Grover et al. proposed graphs and matrices as decision-making support tools in total quality management (TQM) (Grover et al., 2004). Considering connectivity between the factors, authors proposed variable permanent TQM matrix (VPM-TQM) corresponding to the five-critical element TQM digraph and a TQM index as a permanent value of VPM-TQM.

A graph-based multi-agent decision making (GMADM) model to cope with multi-criteria decision making (MCDM) problems with the interrelated criteria was first introduced in (Xiaohan and Zeshui, 2012). The GMADM's basic application is to formalize relations between distinct criteria. According to the authors' research results, GMADM can be used to prioritize projects or decisions based on the aggregate criteria estimation. It is important to note that in these methods graphs are primarily used only as an illustration of criteria interaction.

H. Safari et al. introduced a combined methodology consisting of Analytic Hierarchy Process (AHP), graph theory, and matrix approach for the problem of equipment selection (Safari et al., 2013). The graph theory is used to outline interconnected criteria of CNC machines in order to construct fuzzy matrices to compare different options. Another application of graph theory and matrices approach was demonstrated in regard to financial management. Authors used digraph to model interactions between 7 financial factors thus introducing complex characteristic values of graph matrix (Hu, 2010; Na, 2011).

A. Kuyumcua and A. Garcia-Diaz discussed a problem of revenue management in airline industry (Kuyumcua and Garcia-Diaz, 2000). Researchers proposed three computational models that help to decide how to increase revenues of an air

---

transportation company. The third model is based on detecting cliques in a split graph according to some numeric constraints.

An empirical study by A. Keller encompassed another application of graph theory to economics – author proposed to use graphs as representations of economic models, i.e. vertices are considered to be model's variables while the edges represent dependence between variables in model equations (Keller, 2007).

Application of graph theory to innovations analysis and management by N. J. McCullen et al. used the concept of innovation diffusion constructing the model of social networks with circulating innovations based on graphs (McCullen et al., 2013). In such model innovations are spread across the network of actors only in case of utility being higher than specified boundary value. There are also examples of application of graph theory to a patent data analysis (Sunghae, 2015; Valverde et al., 2007).

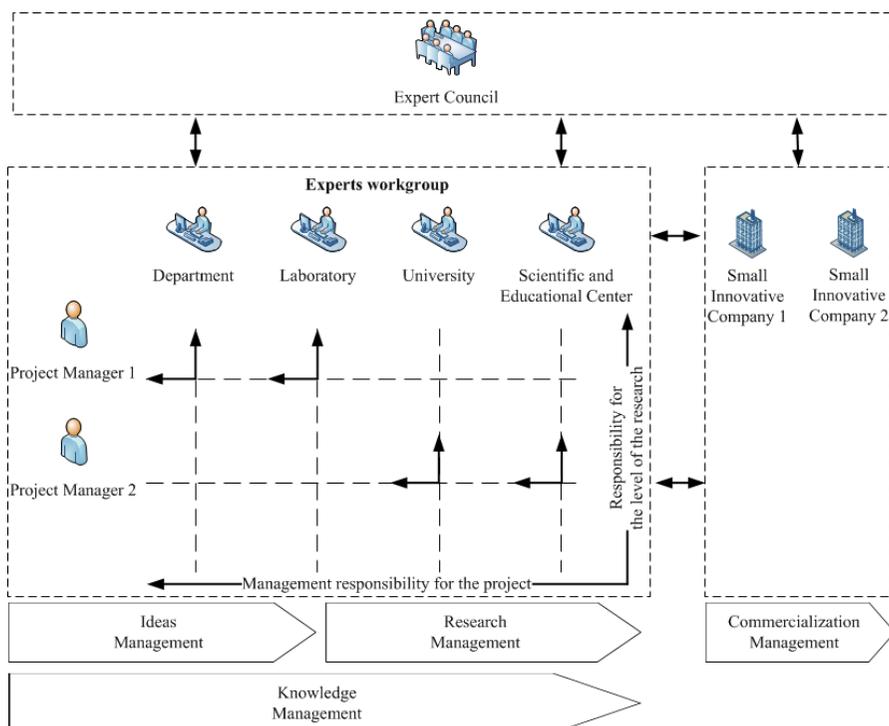
There are different approaches to use graphs in order to estimate innovations. The type of graph used depends on a kind of task that is to be solved by applying the specific model. A common pattern in applying graph models to the tasks of innovation management and decision making is application of adjacency matrix to construct a transition from graphical representation to some form of quantification.

Despite the innovation orientation of described approaches, there is no unified solution that encompasses ideas, innovations, and patents at the same time. The approach proposed in the third section of this work comprises some of the basic ideas behind application of graphs to innovation management and decision making while at the same time concerning specific research goals and user-oriented requirements.

### **Methodology for Ideas, Innovations, and Objects of Intellectual Property Formalization**

In the market economy the innovative activity is based not on operational approach but rather on a project management methodology. In the foundation of project management methodology lies a matrix structure of ideas and innovation management. An example of this structure is provided in figure 1.

**Figure 1. The matrix model of innovations management in universities**



The matrix organization of ideas and innovations management helps to escape disadvantages of the simplistic project management architecture by strictly decoupling management and professional responsibility for project results. A new actor – project manager – is responsible for the result of the project. The matrix structure implemented in university will orient the organization and teams towards practical results acquisition. There is also a part of the model responsible for project's results commercialization. If the project considered to be successful, a small innovative company can be established.

In the scope of the matrix model, a continuous process of knowledge management must be established. In case of large scientific institution with a large number of researchers as well as with a large quantity of ongoing projects, it is necessary to organize and somehow formalize ideas and innovations management. The future of each innovation depends on the presence of financial support and resources needed to develop a product. In this regard a complex approach to innovative projects estimation was proposed (Archipov and Pishko, 2012). The key point of this approach is that innovations must be investigated consequently on different levels of detalization: innovation idea, innovation proposal, innovation project or in our case – ideas, innovations, and objects of intellectual property (OIPs).

Commonly, ideas, innovations, and OIPs originate in unstructured form as a set of textual descriptions, illustrations, formulas, etc. When dealing with ideas, innovations, and OIPs, it becomes rather important to impose a formal structure on a set of intellectual objects to see a big picture. Formalized methodology and a set of algorithms can be used to solve the task of data structuration. The automatic approach can be used mainly as a first step in the process of structuration. The final structure must be specified using experts knowledge.

Each of the proposed kinds of intellectual objects has its own formalized description. We have combined descriptive fields into a single table. Due to the restrictions on paper volume, three sections were omitted: economical characteristics of idea; the estimate of commercialization and success of idea; research and development results (innovation). However, remaining indicators give a glimpse at the approach for ideas, innovations and OIPs representation that was used.

**Table 1. A brief structure of ideas, innovations, and OIPs passports**

| IDEA  | INNOVATION  | OBJECT OF INTELLECTUAL PROPERTY (OIP)          |
|---|---|--|
| <b>General Information</b>  |   |  |
| Idea ID   | Innovation ID   | OIP ID   |
| Date of passport creation   | Date of passport creation   | Date of passport creation                      |
| Idea name   | Innovation (research work) name                                       | OIP name                                       |
| Full idea name  | Brief annotation  | OIP kind                                       |
| Field of scientific research  | Stage in the life cycle of innovation                                 | Supposed method for securing rights            |
| Priority direction of science, technologies, and technics development | List of ideas related to innovation                                   | Document for securing rights                   |
| Critical technology   | Direction of technical development of the results                     | Serial number of document for securing rights  |
| The supposed innovation   | Priority direction of science, technologies, and technics development | Date of issue for document for securing rights |
| The possible application field  | Critical technology   | Starting date for document for securing rights |
| The target consumers  | Possible application field  | Ending date for document for securing rights   |
| The socio-economic task being solved                                  | Target consumers  | Registering authority                          |
| The assumed time for market entry                                     | Amount of financing   | Brief annotation                               |
| Author rating   |   | List of innovations related to OIP             |

| IDEA   | INNOVATION  | OBJECT OF INTELLECTUAL PROPERTY (OIP)                                 |
|--|---|---|
|  |   | Priority direction of science, technologies, and technics development |
|  |   | Critical technology   |
|  |   | Target consumers  |
|  |   | OIP's cost  |
| <b>Author (Team)</b>   |   |   |
| <i>Data on author</i>  |   |   |
| Name and surname   | Name and surname  | Name and surname  |
| Academic degree  | Academic degree   | Academic degree   |
| Position   | Position  | Position  |
| Field of scientific interests                                  | Field of scientific interests   | Field of scientific interests   |
| Is a team leader?  | Is a team leader?   | Is a team leader?   |
| <i>Data on co-authors</i>                                      |   |   |
| Name and surname   | Name and surname  | Name and surname  |
| Academic degree  | Academic degree   | Academic degree   |
| Position   | Position  | Position  |
| Field of scientific interests                                  | Field of scientific interests   | Field of scientific interests   |
| Organization   | Organization  | Organization  |
| <b>The Qualitative characteristics of the idea</b>             | <b>Research and Development Information</b>                           |   |
| The justification of the idea                                  | <i>Research and Development I</i>                                     |   |
| The justification of the importance of the supposed innovation | Research and Development theme  |   |
| The life cycle stage   | Priority direction of science, technologies, and technics development |   |
| The degree of novelty  | Critical technology   |   |
| The scale of importance  | Brief annotation (awaited results)                                    |   |
| The coverage of the supposed usage                             | Type of research and development work                                 |   |
| The level of competitiveness                                   | List of ideas related to research and development work                |   |
| The conditions of usage  | Theme code  |   |
| The awaited effect   | International classification code                                     |   |
| Technological impact   | Federal program   |   |
| Technical impact   | Source of financing   |   |
| Social impact  | Amount of financing   |   |
| Ecological impact  | Contract number   |   |
| Economical impact  | Starting date   |   |
| Integral impact  | Ending date   |   |

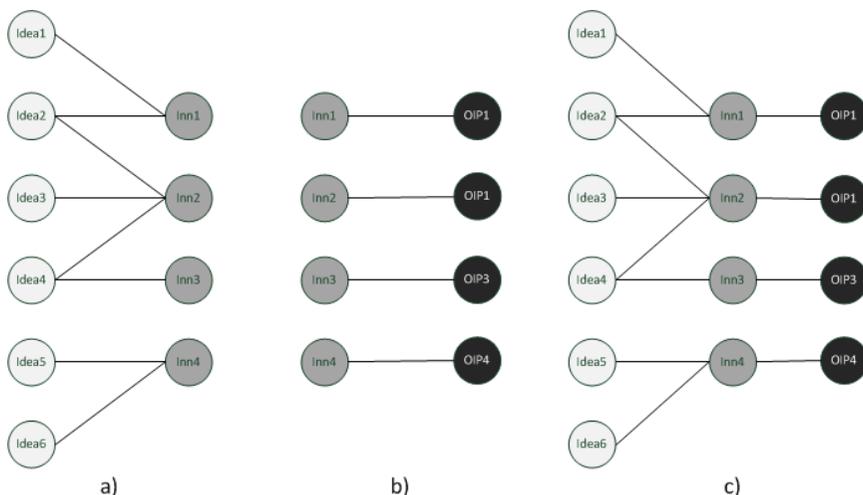
| <b>IDEA</b>  | <b>INNOVATION</b>   | <b>OBJECT OF INTELLECTUAL PROPERTY (OIP)</b> |
|--|---|--|
|  | Base of conducting the research<br>Client organization name<br>Code of legal form for client organization<br>Contractor organization name<br>Code of legal form for contractor organization<br>Subcontractor organization name<br>Code of legal form for subcontractor organization<br><i>Research and Development 2...</i> |  |
| <b>Economical Characteristics of Idea</b>                            |   |  |
| <b>Evaluation of Commercialization Potential and Success of Idea</b> | <b>Research and Development Results</b>   |  |

Innovation life cycle is usually considered as a chain of following processes – idea generation, idea implementation, development of the product, product usage to product commissioning and its replacement by a new product. The innovative product life cycle can be decomposed into several stages: research; development; production; consummation. The proposed graph model mainly covers research, development, and production stages of the innovation life cycle.

### **Graph based approach to automatic prospective research programs design**

The developed graph model for representation, described in previous section, encompasses two types of graphs – hypergraphs and bipartite graphs. According to our model, ideas can only have relations with innovations; ideas and innovations nodes can be combined in a single bipartite graph as presented in figure 2a. Also, the relation between innovations and objects of intellectual property (OIP) can be represented as a bipartite graph as well (see figure 2b). A union of two graphs makes transitions from idea to OIP much clearer (see figure 2c). Despite the fact that presented bipartite graph model is clear and easy to understand, it does not address the question of relations between the nodes of the same type. However, this representation has high illustrative power.

**Figure 2. Set of graphs for innovation management graphs: a) ideas-to-innovations graph; b) innovations-to-objects of intellectual property; c) whole innovation management graph**



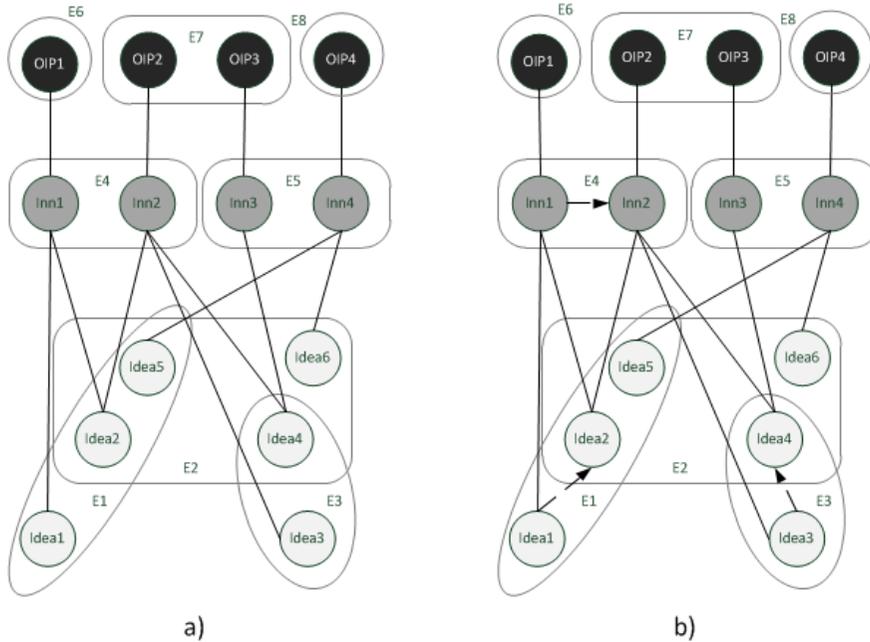
The first outlined problem can be solved by introducing a new dimension to our model. This new dimension concerns the detalization of relations on each level of the previous model. These levels are: ideas level, innovations level, OIPs level. The need for the second dimension arises from three facts about the ideas, innovations and OIPs: first, when the number of ideas rises it is easy to get to a problem of repetitive ideas; second, ideas (as well as innovations and OIPs) can be in relations of part-whole and predecessor-descendant; and, finally, ideas (as well as innovations and OIPs) can represent a single class of objects.

In order to incorporate these features into our model, we introduce hypergraph structures on each level of the model. The main property of hypergraph is that each edge can be adjacent to more than two nodes, thus making connections between groups of ideas and innovations clearer. An example of this model can be seen in figure 3a. Evolution relations between nodes of the same type can be shown by the means of directed edges on each level. The modified graph is presented in figure 3b.

Edges of hypergraph on each level of complex graph can be defined using the machine learning. Let us consider the ideas level. Two primary classification schemes that can be imposed on ideas are: classification according to the field of knowledge, classification according to the field of application. The superposition of these two classified spaces can further be transformed into hypergraph edges based on some criteria. As a next step we will define a solid method to construct idea-to-innovation-to-OIP graph. This algorithm needs two kinds of dictionaries: a dictionary of the most common words in English and a dictionary of highly specialized words (such as scientific terms).

Initial data: graph as shown in figure 2c with each node having its own passport. The algorithm steps will be as follows:

**Figure 3. Bipartite graph model with hypergraphs structures: a) simple model; b) model with “part-of” and preceding relations**



- 1) Textual descriptions of ideas are being preprocessed with following steps:
  - a. convert text to lower case;
  - b. remove common words (e.g. “the”, “is”, “it” etc.);
  - c. remove numbers and punctuation marks;
  - d. stem words (i.e. extract the base of the word);
  - e. remove all the words except for highly specialized words;
  - f. construct a term-to-document<sup>5</sup> matrix of the following form

$$TDM = \left( \begin{array}{c|ccc} term_1 & a_{11} & \dots & a_{1m} \\ \dots & \vdots & \ddots & \vdots \\ term_n & a_{n1} & \dots & a_{nm} \end{array} \right)$$

where  $a_{ij}$  is a number of occurrences of  $term_i$  in document (idea)  $j$ .

- 2) Converting TDM into term-to-document frequency matrix (TDFM)

<sup>5</sup> Idea is considered to be a document.

$$TDFM = \begin{pmatrix} term_1 & \left| \begin{array}{ccc} a_{11} & \dots & a_{1m} \\ \hline \sum_{i=1}^n a_{i1} & \dots & \sum_{i=1}^n a_{im} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \\ \hline \sum_{i=1}^n a_{i1} & \dots & \sum_{i=1}^n a_{im} \end{array} \right. \\ \dots & \\ term_n & \end{pmatrix}$$

- 3) In order to reduce dimensionality, a frequency filter can be used on TDFM, i.e. we calculate each term frequency in the corpus and solve the following optimization problem giving the priority to the most frequent terms:

$$\begin{cases} n \rightarrow \min \\ \sum_{i=1}^n a_{ij} > 0, j = \overline{1:m} \end{cases}$$

- 4) Constructing n-dimensional space where each axis corresponds to one keyword term in a corpus. The value on the axis is a frequency of the term.  
 5) Placing the ideas into constructed n-dimensional space as data points

$$Idea1 = \left\{ \frac{a_{11}}{\sum_{i=1}^n a_{i1}}, \frac{a_{21}}{\sum_{i=1}^n a_{i1}}, \dots, \frac{a_{n1}}{\sum_{i=1}^n a_{i1}} \right\}$$

- 6) Selecting one of m pre-defined class labels and constructing a binary SVM classifier for classifying ideas into two categories: the labeled class and other classes.  
 7) Repeat step 6 for other pre-defined class labels. As an output of this step we will have m binary classifiers that can intersect, i.e. the same idea can be classified into distinct classes by different classifiers (a case of interdisciplinary idea).  
 8) Classifying each idea using each binary classifier, thus acquiring a vector of possible classes for each idea

$$cl_i = \begin{pmatrix} p_1^i \\ p_2^i \\ p_3^i \\ \dots \\ p_m^i \end{pmatrix}$$

where  $p_j^i$  denotes the probability of idea  $i$  being a part of the specific  $j^{th}$  class.

The same idea can be a part of several classes if they are general enough.

- 9) Constructing hypergraph's edges based on ideas' probabilities of being a part of a specific cluster, i.e. each hypergraph edge corresponds to a class of ideas. Two hypergraph edges are adjacent to the same idea node only if the idea has

probability of being a part of each of those classes higher than a specific threshold.

- 10) Repeat steps 1-9 for innovations and OIPs.
- 11) Connect ideas nodes to innovations nodes with directed edges using “List of ideas related to innovation” field of passport of each innovation.
- 12) Connect innovation nodes to OIP nodes with directed edges using “List of innovations related to OIP” field of passport of each OIP.

As a result of the algorithm, a multilevel graph with ideas, innovations and OIPs layers will be constructed. This graph is further used as an input for research and innovation programs defining algorithm.

In order to estimate ideas and innovations numerically, we propose to use some criteria. These criteria cover different areas of ideas and innovations estimation.

Ideas can be evaluated by using two approaches: a) author’s evaluation; b) expert’s evaluation. These two approaches must be combined in order to provide the most adequate estimate of the idea. The author’s evaluation of the idea is a preliminary requirement for further expert’s evaluation. The author’s rating for idea  $I$  can be computed using the formula:

$$R(I) = \sum_i (u_i \times \sum_j X_j(I))$$

where  $u_i$  is a weight depicting the impact of the  $i^{\text{th}}$  criteria on the indicator value;  $X_j$  is a value of the indicator for estimation of potential for commercialization.

Expert’s evaluation of idea’s potential can be represented as:

$$R = r_1 + r_2 + r_3 + r_4$$

where  $r_1$  is a coefficient that takes into account the content and scientific value of the project;  $r_2$  is a coefficient that determines the team’s potential and feasibility of the project;  $r_3$  is a coefficient to estimate the starting conditions;  $r_4$  is a coefficient that estimates market prospects of the project.

The second part to estimating the idea is to define the cost of implementation. The cost is defined in money needed to implement the project,  $C$ . The overall estimate of the  $i^{\text{th}}$  idea is computed as follows

$$Ef_i = \frac{R_i}{C_i}$$

The ideas with highest  $Ef$  value are considered to be the most prospective for implementation.

In order to estimate innovation and research work we need to use different formula:

$$R(Np) = \sum_i (u_i \times \sum_j X_j(Np))$$

$R(Np)$  is the final rating of scientific potential of the research;  $u_i$  stands for coefficients characterizing the degree of influence of  $i^{\text{th}}$  criterion on the indicator's value, also each such coefficient should be in range from 0 to 1. Lastly,  $X_j$  is a value of scientific potential indicator for each research.

In the proposed framework we estimate OIPs based on the cost of OIPs. If research institution can create large amount of OIPs which can be sold to other institutions, companies, and individuals, then their work can be considered efficient. In the scope of our research we use the following formula to estimate OIPs cost:

$$V_p = \sum_{i=1}^n (P_i - E_i) \cdot (1 + d)^{-i}$$

$V_p$  is a joint value of OIPs computed using DCF method;  $P_i$  is a net profit generated by OIPs in  $i^{\text{th}}$  period (year);  $E_i$  are expenditures connected with supporting OIPs (PR, marketing, etc.);  $i$  is a serial number of period (year) of getting the income from OIPs;  $d$  is a discount rate;  $n$  is an amount of predicted periods.

Combining three variables ( $E_f$ ,  $R(Np)$ ,  $V_p$ ) from all three levels of the constructed graph, we can introduce a criteria for innovative programs construction based on graph nodes connections and values of criteria for each level of nodes.

The main approach to construct innovative programs consisting of chains of transformations of ideas into innovations and OIPs is based on searching for a set of separate<sup>6</sup> subgraphs of constructed multi-level graph by maximizing the value of a cost function  $f$  with respect to a set of resource constraints  $C$ . The following formula illustrates the idea:

$$decompose(G) = \begin{cases} \arg \max(f(SG_1, SG_2, \dots, SG_n)) \\ \sum_{i=1}^n r(SG_i) \leq C \end{cases}$$

The cost function  $f$  as well as the resource function  $r$  and resource constraints  $C$  must be defined according to the developed criteria and knowledge about the structure of the graph  $G$ . The specified criteria for ideas, innovations, and OIPs estimation leads to the following system:

<sup>6</sup> On the level of innovations and OIPs.

$$\left\{ \begin{array}{l} f_1(SG_1^I, SG_2^I, \dots, SG_m^I) \rightarrow \max \\ f_2(SG_1^{II}, SG_2^{II}, \dots, SG_m^{II}) \rightarrow \max \\ f_3(SG_1^{III}, SG_2^{III}, \dots, SG_m^{III}) \rightarrow \max \\ \sum_{i=1}^m r(SG_i) \leq C \\ (A^{I-II} \times A^{II-III})_j \geq 1 \end{array} \right.$$

The most important task is to construct optimal research programs based on estimation of quantified parameters on each level of multilevel graph, i.e. on each level of graph we clusterize intellectual objects into several groups that have to obey certain conditions: a) the number of groups on each level must be equal to the number of groups on other layers; b) at least one transient connection between an idea and OIP must exist; c) a research program cannot include innovations and OIPs from other research program (however, ideas from different research programs can become a basis for several programs simultaneously); d) research programs' cumulative demand for resources must not exceed organization's available resource base.

The task of construction of groups (clusters) of intellectual objects on each level of a graph is solved by maximizing a specific value for each kind of intellectual objects. It is important to explicitly outline the connectivity condition. We do this by multiplying adjacency matrices ( $A^{I-II}$  is a binary matrix that represents connections between level of ideas and level of innovations,  $A^{II-III}$  is a binary matrix that represents connections between level of innovations and level of OIPs) for each subgraph (research program) and determining whether the resulting matrix has zero-sum rows or not.

First, we will define graph levels as a set of graph nodes of the same type (idea, innovation or OIP), i.e.:

- Level of ideas:

$$G^I = \{x_1, \dots, x_{n_I}\} : \text{type}(x_i, "idea") = \text{true}, i = \overline{1:n_I}$$

- Level of innovations:

$$G^{II} = \{x_1, \dots, x_{n_{II}}\} : \text{type}(x_i, "innovation") = \text{true}, i = \overline{1:n_{II}}$$

- Level of OIPs:

$$G^{III} = \{x_1, \dots, x_{n_{III}}\} : \text{type}(x_i, "OIP") = \text{true}, i = \overline{1:n_{III}}$$

The predicate  $\text{type}(x_i, "typelabel")$  returns true value only if the node  $x_i \in G$  has type label "typelabel".

In order to decompose the multilayer graph we use the following decomposition algorithm (*Algorithm-1*):

1. Construct the space of possible decomposed graph configurations (

$$SGSpace = \{SGConf', \dots, SGConf''\} = \{\{SG'_1, \dots, SG'_N\}, \dots, \{SG''_1, \dots, SG''_M\}\}, \forall SGConf^{(i)}: G = SG_1^{(i)} \cup \dots \cup SG_N^{(i)} = \{V, E\}$$

) using the following rules (heuristics) in order to minimize the size of a search space:

a. each subgraph must contain the nodes from every level of the original graph, i.e.

$$\forall SG \in SGConf: (\exists x_i \in G^I: x_i \in SG) \wedge (\exists x_j \in G^{II}: x_j \in SG) \wedge (\exists x_k \in G^{III}: x_k \in SG)$$

;

b. there must be no disconnected edges in the subgraph  $\forall u_i \in E: (\exists x_j \in V: u_i \in adj(x_j)) \wedge (adj(u_i) \neq \emptyset)$ ;

c. each subgraph must contain all preceding nodes for OIPs' and innovations' levels (i.e. nodes that have edges directed towards the OIP or innovation component of the graph so that  $x_i \in SG_j^{II} \rightarrow \forall x_k \in SG_j^I: x_i \in adj(x_k)$  and  $x_i \in SG_j^{III} \rightarrow \forall x_k \in SG_j^{II}: x_i \in adj(x_k)$ );

d. two subgraphs can be unified into single subgraph if their ideas sets have non-empty intersection, i.e.  $\exists SG = SG_1 \cup SG_2: SG_1^I \cap SG_2^I \neq \emptyset$ .

2. Assign to each decomposition configuration specific value according to criteria computed on each level of the model.

3. Search the graph decomposition configuration space for optimal configuration with respect to constraints on resources.

The proposed algorithm can be modified in order to further reduce the search space and time needed by the algorithm (*Algorithm-2*):

1. A few interconnected ideas  $x_1^I, x_2^I, \dots, x_N^I \in u_i^I$  (where  $u_i^I$  is a hypergraph edge) are randomly selected in  $G$  as a foundation for further construction of a subgraph  $SG^{(0)}$ . The initial subgraph  $SG^{(0)}$  is then constructed according to rules 1a-1d of Algorithm 1. The constructed graph  $SG^{(0)}$  will be used as a basis for iterative process of finding the optimal subgraph in respect to joint ideas, innovations, and OIPs estimation criteria discussed above. As it is an initial estimation, there is no need for the subgraph to represent specific research program,

relevant to our objectives. The estimation of effectiveness of  $SG^{(0)}$  as a research program is saved as  $f^{(0)}$ .

2. **The inclusion step.** A new idea  $x_j^I$  is introduced to a subgraph  $SG^{(0)}$  if and only if this idea is adjacent to at least one hypergraph edge that in respect is adjacent to one of the ideas nodes in  $SG^{(0)}$ , i.e.  $SG^{(0)} \leftarrow SG^{(0)} \cup x_j^I: \exists x_k^I \in SG^{(0)}, x_j^I \in adj(adj(x_k^I))$ . Among all such ideas we select an idea with the highest  $E_f$  value. Among all adjacent innovations we select innovation with the highest  $R(Np)$  score:

$$SG^{(0)} \leftarrow SG^{(0)} \cup x_i^{II}: \exists x_j^I \in SG^{(0)}, x_i^{II} \in adj(adj(x_j^I)), x_i^{II} = arg(max(R)).$$

Similar approach is used to add an OIP node to subgraph:  $SG^{(0)} \leftarrow SG^{(0)} \cup x_p^{III}: \exists x_i^{II} \in SG^{(0)}, x_p^{III} \in adj(adj(x_i^{II})), x_p^{III} =$

$$arg(max(V_p))$$

. So, the resulting subgraph  $SG^{(0)}$  has expanded. If the resulting subgraph resource consumption is larger than existing resource limitations, i.e.  $C - r(SG^{(i)}) \leq 0$  then results of the step are discarded  $SG^{(0)} \leftarrow SG^{(0)} \setminus \{x_j^I, x_i^{II}, x_p^{III}\}$ , and we continue with the step 3. Otherwise, if there is no such idea  $x_j^I$  in  $G$  that is close to initial set of ideas then a new subgraph is formed (go to step 1 and start a new instance of algorithm).

3. **The exclusion step.** This step is needed to exclude previously selected nodes in order to give new opportunities to subgraph construction in case when  $|C - r(SG^{(i)})| \leq \delta, \delta \rightarrow 0$ . This step is similar in effect to the step of random modification of data point in simulated annealing optimization algorithm. On this step an idea node  $x_j^I$  is randomly selected in  $SG^{(0)}$  (however, usually it is best to penalize the selection of newly introduced nodes on step 2). Then we exclude this node from subgraph, i.e.  $SG^{(0)} \leftarrow SG^{(0)} \setminus x_j^I$ . Keeping in mind that the edges in the graph are not oriented, we must exclude innovation and then OIP nodes that have hanging edges. In case of innovation nodes,  $SG^{(0)} \leftarrow SG^{(0)} \setminus x_i^{II}: \exists u_q^{I-II} \in SG^{(0)}, u_q^{I-II} \in adj(x_i^{II}), \exists! x_i^{II} \in adj(u_q^{I-II})$ . In case of OIP nodes,

$$SG^{(0)} \leftarrow SG^{(0)} \setminus x_p^{III} : \exists u_w^{II-III} \in SG^{(0)}, u_w^{II-III} \in adj(x_p^{III}), \exists! x_p^{III} \in adj(u_w^{II-III})$$

. Also, another penalization condition can be added – the selected nodes are not excluded from the graph if their exclusion will cause the loss in  $f(SG^{(0)})$  of more than specified amount in percent (e.g. 10%).

4. We mark the resulting graph as  $SG^{(1)} \leftarrow SG^{(0)}$ .

5. We compare overall estimation  $f^{(1)}$  of  $SG^{(1)}$  with overall estimation  $f^{(0)}$  of  $SG^{(0)}$ . More generally, we compare  $f^{(i)}$  of  $SG^{(i)}$  with  $f^{(i-1)}$  of  $SG^{(i-1)}$ . If  $|f^{(i)} - f^{(i-1)}| \leq \varepsilon$  (where  $\varepsilon \rightarrow 0$ ) and  $C - r(SG^{(i)}) > 0$  then we mark research program  $SG^{(i)}$  as final and repeat the whole algorithm for  $G \leftarrow G \setminus SG^{(i)}$ . Otherwise, if  $|f^{(i)} - f^{(i-1)}| > \varepsilon$  then we repeat steps 2-5 for  $SG^{(0)}$  replaced by  $SG^{(i)}$ . But if  $|f^{(i)} - f^{(i-1)}| \leq \varepsilon$  and  $C - r(SG^{(i)}) = 0$  or  $G \setminus SG^{(i)} = \emptyset$  then the set of optimal research programs is formed.

In summary, Algorithm-2 tries to maximize the value of  $f(SG)$  with respect to resource limitations  $C$ . The state search space for specific  $SG$  can be graphically represented as an  $n$ -dimensional space, where  $n \leftarrow |V|$  (each axis has two values – node is included in subgraph, node is not in subgraph). In this space the possible values of  $f(SG)$  form a kind of surface. It is, however, important to note its discrete character. But it can be interpolated to represent a continuous surface. Then the task of finding the optimal configuration of subgraph-research program can be restated as a problem of finding the global maxima (or minima, depending on  $f(SG)$  structure) and selecting according nodes of  $G$ . In such formulation, the problem can be solved using common optimization methods and techniques.

### Case study

The initial data set consists of **6148** textual descriptions of ideas publically available on site <http://www.whynot.net/> in English up to 18<sup>th</sup> September, 2013. Originally, there are more ideas on the site (7281 ideas), but ideas with empty description were not considered in case study. Each idea in the data set is ranked by site users with the following grade system: very weak, weak, average, strong, very strong. We convert these rankings into numerical form for the purpose of simplification: very weak - 1, weak - 2, average - 3, strong - 4, very strong - 5. 0 is reserved for possible non-ranked ideas.

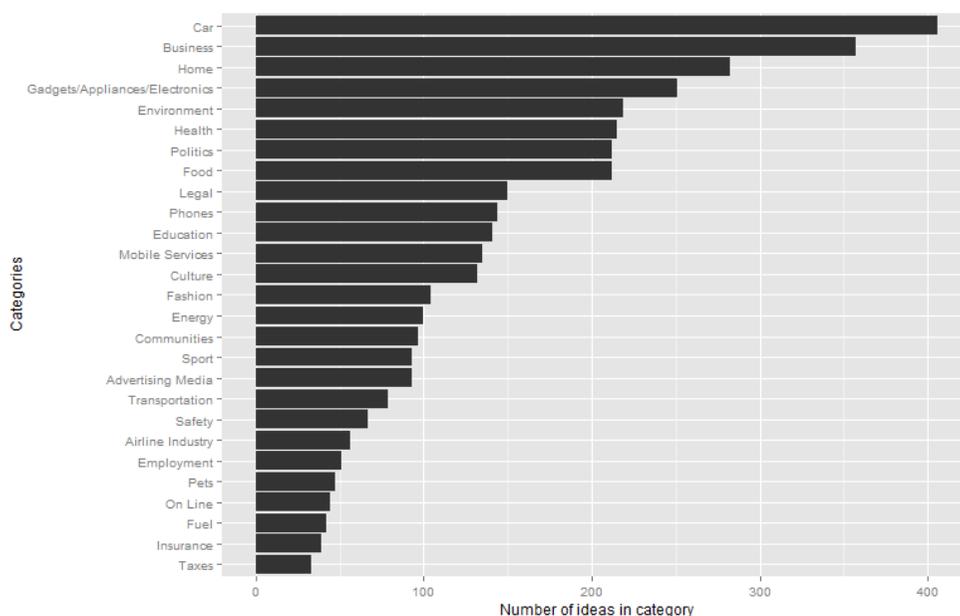
The data acquisition is done automatically by accessing the ideas descriptions on separate pages of the site: <http://www.whynot.net/ideas/>. The downloaded record for each idea contains its description, the category of idea set up by users, and the numerical ranking of the idea. The initial number of categories (205) was reduced down to 27 by statistical and basic contents estimation of the data set with the following hand relabeling.

First, we have defined general categories of ideas. We have selected categories that are in 75%-quantile range by number of ideas (categories with 31 or more ideas in each). It is important to note that for machine learning purposes we are limiting ourselves with relatively large categories that are not always can be described as general in the scientific manner. For example, despite the fact that “Agriculture” is indeed a general category, it contains only 2 ideas, and thus it renders itself useless for machine learning and further classifiers construction.

Further, we have added ideas from the similar categories into the general categories. This process was conducted by hand relabeling (e.g. “Cars” and “Car”). The remaining ideas from small categories were transferred into uncategorized group. Some of categories with non-informative description were also uncategorized – ideas were transferred into uncategorized group (e.g. “Inventing” category is too general). As an outcome, around 38% of all ideas are uncategorized and thus they can be used to validate the results of classification.

The resulting figure shows distribution of ideas among remaining categories (Figure 4).

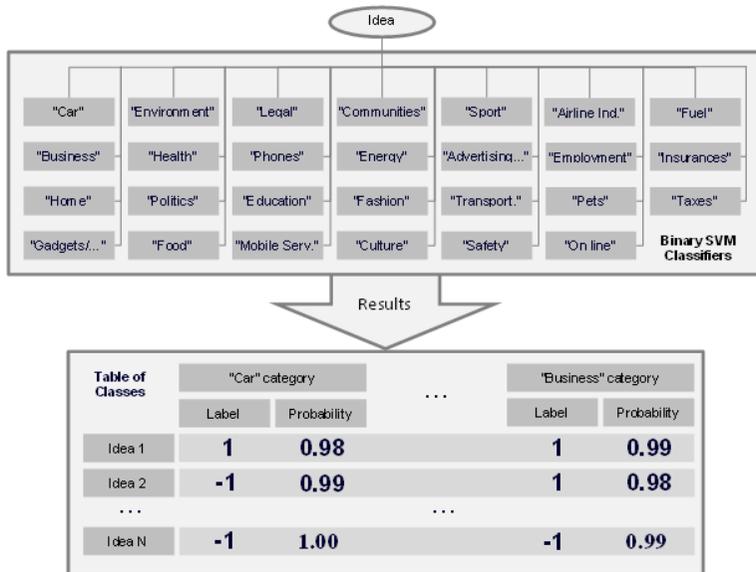
**Figure 4. Distribution of ideas among selected general categories**



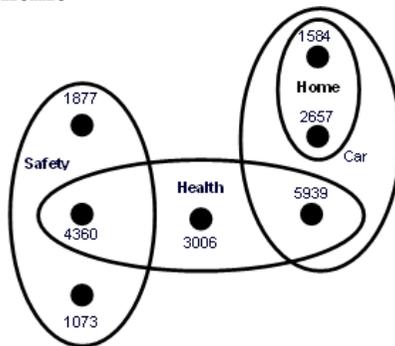
According to our approach, we have constructed 27 binary SVM classifiers (one for each category) thus dividing the term space into a set of regions. Each idea can be classified into several categories simultaneously if its comparison against several categories will yield a positive value. Ideas of the similar category are considered to be hypergraph nodes adjacent to the same edge. If an idea is categorized into several classes then it is considered as a joint between two edges of the hypergraph.

The classifiers structure is represented in figure 5. The results of classification for a small number of ideas are depicted in figure 6 in form of hypergraph.

**Figure 5. Example ideas hypergraph representation**



**Figure 6. Classifying scheme**



The remaining two layers (innovations and OIPs) can be constructed in the similar manner. The connections between adjacent layers are done using undirected edges in order to emphasize that each class of intellectual objects can become a basis for objects of another class. Due to the space restrictions, we will consider an example of the construction and partitioning of full graph in the next paper.

**Discussion**

While we have presented a graph-based approach for innovations and ideas estimation and research and innovations programs construction, only the results of graph construction methodology are provided in this work. The results of the test have shown that textual descriptions and numeric parameters can be used to classify the unstructured data such as ideas and innovations. The results of the classification can be visually and mathematically represented in the form of multilayer graph with each layer being a hypergraph with intellectual objects as nodes. This representation can both be used for the purpose of mathematical processing to find out optimal research programs and their characteristics as well as visually represent the ideas transformation process. Multilayer graph representation can support decision-making process (in research and investments) because of its simplicity and ability to represent relations between different intellectual objects.

The existing approaches to innovations management usually do not deal with ideas, the basic stage of intellectual object development. Usually, a large number of ideas is generated and implemented in educational and research institutions. Thus, the process of collection and facilitation of new ideas becomes an essential part for university's innovative activity supporting the establishment of innovative students and teachers community, whereas for companies it remains unnecessary. Also, the approaches discussed in literature mostly demonstrate little to no automation as well as solving the task of innovations estimation in a simplistic manner that does not actually require graphs or other complex data structures. The proposed approach heavily relies on text mining techniques, optimization algorithms, and graph structures to introduce the connections between intellectual objects.

The main impact of the proposed approach is its ability to automatically classify unstructured data in order to impose a structure on it (hypergraph structure) and its ability to compose research and innovative programs automatically using existing optimization algorithms based on complex criteria. The resulting research programs can further be estimated by the experts in order to narrow down to the most promising and relevant innovations and ideas for financing and support. Overall, the proposed approach for innovative programs automatic selection can become a foundation to form a smart policy for choosing ideas and research projects to invest money.

The proposed approach is also a part of developing innovation management information system. The main goal of the innovation management system is to provide the organization (for example, the university) with innovative programs consisting of the most promising projects. The information system is planned for deployment in leading Russian universities. It is expected that this system will become a necessary tool for students to present their ideas to wide community of peers and teachers. The main goal of the system is to help decision makers to correctly identify the most promising projects and ideas. Another goal is to support the development of students' research and development community. Combined, the

achievement of these goals can largely support educational institution in the field of facilitation of innovations development.

### **Future work**

The proposed approach is used in ideas and innovations management system developed for automatization of innovation management processes in universities, thus the further development of the proposed approach mainly depends on the quality of content provided by authors. There are several prospective directions for the evolution of the proposed approach.

The wide functionality of the system covers interests of every group of potential users. There are at least three major categories of users for the system under development. These groups are:

- students and PhD students:
  - participation in contests for grants (easily generated grant documentation based on the description of idea or innovation that is present in the system);
  - ability to estimate student's idea and get the feedback on ideas;
  - ideas and innovations generation using an algorithm of inventive problem solving;
  - ideas publication, search for teammates, discussion.
- the university management:
  - decision making on resource distribution;
  - research programs formation, research topics diversification;
  - decision making on plans for research teams establishment and support;
  - decision making on plans to commercialize the objects of intellectual property.
- teaching staff and research workers:
  - full-scale students and PhD students involvement in research activities (based on project-oriented education);
  - research teams formation;
  - ideas and innovations development tracking on all stages of its life cycle;
  - university's database of ideas and innovation creation.

All of the listed directions are top priority for system's further development and thus they must be studied thoroughly in our future works. Below, we will discuss a few prospective directions.

Users' feedback and comments is a source of additional information on ideas and innovations. This information in the form of comments can further be incorporated in the estimation of ideas and innovations. Sentiment analysis algorithms and opinion mining methods can be used to extract the information on users' opinions

and include it in the ideas and innovations estimation process. Detailed comments can also be used to establish additional connections between ideas that seem to be not connected directly.

Another research direction can evolve around authors' or teams' interactions as well as the individual achievements of students. Usually, prospective ideas and innovations are provided by the same individual or team. It is thus important to develop an approach to identify such teams or individuals in order to combine their efforts on the basis of similarity of their projects and ideas. Such expansion for our approach can lead to introducing yet another concept – the concept of author's passport. This concept is functionally independent of other passports and introduces the new category of building blocks in innovations management system.

The third research direction can be based on an introduction of additional information for the purpose of benchmarking of ideas and innovations based on best samples of ideas and innovations. These additional ideas and innovations libraries can also be used as a retrospective material to construct trend lines to predict evolution directions for ideas and innovations.

Finally, there is a highly prospective research direction which suggests introduction of scenario construction methodology to provide different options of research and innovation according to the objectives of organization, its resources and the timeline. Each scenario can be represented as a set of specific objectives, resource base and determined timeline, and as a set of research programs constructed based on ideas and innovations provided to the innovation management system. In the scope of this direction, it is promising to develop methods that can dynamically adjust scenarios to changing conditions.

Further work in these research directions will help us to expand the number of parameters under consideration as well as the accuracy of research and innovative programs composing. Moreover, some of these research directions can provide additional results that can help authors generate more detailed and novel ideas.

## **Conclusion**

In the scope of this paper we have proposed the novel approach to automatic unstructured information processing. We have also presented means of formalization for this kind of information using concepts of idea, innovation, and OIP passports for the case of innovation management. The proposed approach consists mainly of two stages. The first stage was introduced to solve the problem of primary structuration of unstructured data based on unstructured textual description of data items. The second stage is devoted to finding substructures in structured information in respect to some criteria. The results of the second stage as a set of substructures can then be used to organize the unstructured information in the planned manner. In the scope of our research we have used graph structures and machine learning in

order to organize the unstructured data automatically. The results of application of the approach to the case of unstructured ideas descriptions have shown that the approach can easily achieve the objectives and provide valuable information on ideas, innovations and OIPs groupings. The proposed approach will be implemented in innovation management system for higher education institutions to improve innovation management processes and provide a solid support for innovative ideas implementation.

### **Acknowledgement**

This work was supported by the Ministry of Education and Science of Russian Federation grant RFMEFI57314X0007.

### **References**

- Archipov, A., Pishko (2012), “N. Innovative enterprise development: objectives, criteria, features of the planning”, Proceedings of the Higher Educational Institutions, The Technology of Light Industry, No. 1, p. 79-83.
- Batrutdinov, A., Fedoseev (2008), “Basic models of the innovation process and innovation classification features”, Issues in Modern Economics, N 2 (26), URL: <http://www.m-economy.ru/art.php?nArtId=1988>.
- Cooper R.G (2001), “Winning at new products. Accelerating the process from idea to launch”, Cambridge (MA): Perseus Publishing.
- Cooper R.G. (1994), “Third-Generation New Product Processes”, Journal of Product Innovation Management, Vol. 11, No. 1, p. 3-14.
- Darvish M., Yasaei, M., Saeedi, A. (2009), “Application of the graph theory and matrix methods to contractor ranking”, International Journal of Project Management, Issue 27, p. 610–619.
- Dodgson M., Rothwell R. (Eds.) (1994). “The Handbook of Industrial Innovations”, Aldershot: Brookfield.
- Enkel E., Gassmann O., Chesbrough H. (2009), “Open R&D and Open Innovation: Exploring the Phenomenon”, R&D Management, Vol. 39, No. 4., p 311-316.
- European Commission. European Innovation Scoreboard 2005, “Comparative analysis of innovation performance”, p 46.
- Freeman, C. (1995), “The National System of Innovation”, Historical Perspective, Cambridge Journal of Economics, London, Pinter.
- Gassmann, O., Enkel, E. and Chesbrough, H. (2010), “The future of Open Innovation”, R&D Management 40, 3, p 213-222.
- Grover, S., Agrawal, V. P., Khan, I. A. (2004), “A Digraph Approach to TQM Evaluation of an Industry”, International Journal of Production Research, 42:19, p 4031-4053.
- Haustein, H.D., Weber, M. (1983), “Decision Support for Innovation Management: Application to the Lighting Industry”, RR-83-29, p 66.
- Hu, G. (2010), “Research on Stability of Financial Management System Based on Graph Theory”, 2nd International Conference on Industrial Mechatronics and Automation, pp 461-464.

- Hulla, C. E., Liob, B. H. (2006), "Innovation in Non-Profit and For-Profit Organizations: Visionary, Strategic, and Financial Considerations", *Journal of Change Management*, Volume 6, Issue 1, p 53-65.
- Imai K., Nonaka I., Takeuchi H. (1985), "Managing the New Product Development Game - The Uneasy Alliance", K.Clark and R.Hayes, Boston, Harvard Business School Press.
- Keller, Andre A. (2007), "Graph Theory and Economic Models: from Small to Large Size Applications", *Electronic Notes in Discrete Mathematics*, Issue 28, p 469-476.
- Kline S.J., Rosenberg N. (1986), "An overview of innovation // The positive sum strategy: Harnessing technology for economic growth", edited by R.Landau and N.Rosenberg, Washington: National Academy Press.
- König, M. D., Battiston, S. (2009), "From Graph Theory to Models of Economic Networks. A Tutorial", *Networks, Topology and Dynamics*, *Lecture Notes in Economics and Mathematical Systems*, Volume 613, p 23-63.
- König, M. D., Battiston, S., Napoletano, M., Schweitzer, F. (2008), "On Algebraic Graph Theory and the Dynamics of Innovation Networks", *Networks and Heterogeneous Media*, Volume 3, Number 2, p 201-219.
- Kuyumcua, A., Garcia-Diaz, A. (2000), "A polyhedral graph theory approach to revenue management in the airline industry", *Computers & Industrial Engineering*, Issue 38, p 375-396.
- Mahdi S. (2002), "Search strategy in product innovation process: theory and evidence from the evolution of agrochemical lead discovery process", DPhil Thesis, Unpublished, SPRU. – University of Sussex, UK.
- McCullen, N. J., Rucklidge, A. M., Bale, C. S. E., Foxon, T. J., Gale, W. F. (2013), "Multiparameter Models of Innovation Diffusion on Complex Networks", *Siam J. Applied Dynamical Systems*, Vol. 12, No. 1, p 515-532.
- Na, L. (2011), "The Analysis on Financial Management System based on Graph Theory", *International Conference on Computer Science and Network Technology*, p 2482-2485.
- OECD (2002), "Frascati Manual: Proposed Standard Practice for Surveys of Research and Experimental Development", OECD, Paris.
- OECD Science (2008), "Technology and Industry Outlook 2008", OECD publications, p 258.
- Rao, R. V. (2006), "A Decision-making Framework Model for Evaluating Flexible Manufacturing Systems using Digraph and Matrix methods", *The International Journal of Advanced Manufacturing Technology*, Volume 30, Issue 11-12, p 1101-1110.
- Rothwell R. (1994), "Towards the Fifth-generation Innovation Process", *International Marketing Review*, Vol. 11, No. 1, p 7-31.
- Safari, H., Faghih, A., Fathi, M. R. (2013), "Integration of Graph Theory and Matrix Approach with Fuzzy AHP for Equipment Selection", *Journal of Industrial Engineering and Management*, *JTEM Issue 6(2)*, p 477-494.
- Santarelli, E. (1995), "Directed Graph Theory and the Economic Analysis of innovation", *Metroeconomica* 46:2, p 111-126.
- Sieg, J., Wallin, M. and von Krogh, G. (2010), "Managerial Challenges in Open Innovation: a Study of Innovation Intermediation in the Chemical Industry", *R&D Management* 40, 3, 2010, p 281-292.
- Sitenko, D. (2010), "Macroeconomic indicators to assess innovation: the European experience", *Journal of the Military University*, No 3 (23), p 149-154.
- Sunghae, J. (2015), "Patent Big Data Analysis using Graph Theory", *Advanced Science and Technology Letters*, Vol.85, *Information Technology and Computer Science*, p 25-28.

- 
- Trott, P. (1998), “Innovation Management and New Product Development”, Essex, UK, Prentice Hall.
- Valverde, S., Solé, R. V., Bedau, M. A., Packard, N. (2007), “Topology and Evolution of Technology Innovation Networks”, *Physical Review E* 76, p 7.
- Van de Ven, A. and Engleman, R. (2004), “Central Problems in Managing Corporate Innovation and Entrepreneurship”, *Advances in Entrepreneurship, Firm Emergence and Growth*, Vol. 7, p 47-72.
- Van de Ven, Andrew H. (1986), “Central Problems in the Management of Innovation”, *Management Science*, Vol. 32, No. 5, Organization Design, p 590-607.
- Xiaohan Yu, Zeshui Xu. (2012), “Graph-based Multi-agent Decision Making”, *International Journal of Approximate Reasoning*, Issue 53, p 502–512.
- Zygiaris, S. (2009), “Regional Innovation System Failures and Highlights”, *Romanian Journal of Regional Science*, Vol. 3, No. 2, p 54-76.

