Potential Innovations (New Ideas/Trends) Detection in Information Network

Dmitry Romanov, Mikhail Ponfilenok, and Nikolay Kazantsev

***Abstract*—This paper considers the task of semantic innovation (new idea or trend) detection possibility in corporate information field applying the analysis of dynamic changes in enterprise information flows. Here we consider such hypothesis that the sign of innovation appearance is the dramatic increase of new terminology employment (that has occurred rarely or has not been used at all), or the appearance of previously not occurred combinations of old "terms" (words) and "N-grams" (sequences of words). Presented results of this experiment are based on ENRON, TREC2005 databases, e-mail exchange among employees of Russian company. We have overseen conditions under which the innovation detection is statistically significant and made the conclusion about the future of above-listed hypothesis of innovation detection.**

***Index Terms*—Innovation detection; innovation sources; terms; innovation; managing technological change; computer-aided text analysis; knowledge management; enterprise content; unstructured information.**

I. Introduction

Under post-industrial society conditions and active usage of innovations, the identification of promising ideas on possible early stages of life cycle is a very important issue. Participating in large number of consulting and educationa projects, authors face real problems of companies, among which: impossibility to reflect in information systems changes of requirements rendered to business (compliance), high dependence on specific person on a certain place, high financial risks while changes or modernization of information systems. Enterprise information changes dynamically, the amount of data increases and if the corporate structure is bulky enough the question of competence search (who is an expert in this area?) becomes very urgent.

II. Paradigms of Innovation

Since "innovation" notion is widely interpreted by different authors and is now attached to several of topics we give the definition from our working area. We consider innovations in frames of knowledge management and therefore suggest the slightly modified definition of this notion:

Innovation is a new word ("term") or a new combination of used words ("N-gram") in the past that is widely used in corporate information field today;

"Terms" and "n-grams" are namely the blocks of corporate information content. N-gram contains a combination of terms, where n – shows the quantity of members. For instance in bigram there are two terms, it's the most frequent size of n-gram.

An enterprise which wants to implement innovations has to think of such questions like: uncovering, developing and growing of ideas which are highly required within the enterprise; instantly verifying and reconsidering its technological and intellectual potential.

A. Topic Significance

Innovation yields competitive advantage. Since the postindustrial economy is based on information spreading, management of innovations is one of the most important parts of information theory.

First step of innovation management is always a new idea appearance. But despite of brainstorming and other methods of creation activity stimulation ideas don't occur according to manager's schedule or an order. The traditional idea management approach is following: employees add their new ideas in some information system. This approach can work during crowd sourcing outside the corporation, when there is an abundance of "idea creators" and they have an evident motivation. Inside the corporation where special corporative culture is required, could be some psychological barriers and motivation could be not evident: new idea could occur, emerge and then die out, being postponed because of current operational tasks won't survive before recording into database.

Answers to these questions could lie in several scopes:

Possession of internal processes of high maturity, namely, having actual knowledge of process logic, internal risk, quality control and management;

Creation of expert competence centers, where knowledge can be accumulated, stored and distributed enterprise wide. Experts in such centers should be linked by (via) cross-functional principle, thus elaborating solutions for complex interdisciplinary problems;

Implementation of unified information field (UIF) support and management system, which becomes a basis for interdisciplinary links, deploying complex problems and external information environment, scientific and innovation trends monitoring.

In order to explore all the scopes we made a research through most cited papers on innovation detection, featuring: (1) innovation appearance in web communities; (2) practical application of the principles of open innovation (creating cultural innovative climate in organization); (3) using

Manuscript received June 25, 2012; revised July 19, 2012.

The authors are with S&E Center of Information Control Technologies National State University Higher School of Economics (e-mail: dromanov@it.ru, ponfilenok@gmail.com, nicolay.kazantsev@gmail.com)

DOI: 10.7763/IJFCC.2012.V2.122

innovations from business perspective.

III. Literature Review

Several research groups worked on this issue. Whilst collecting the research data we used didactical method for our literature review. We searched in open source scientific databases for recently published academia papers that contained following keywords: "innovation detection", "idea discovery", etc.

A. Nakatsuji Research Group

Original motivation behind research of Nakatsuji et. all (2007), (2009) was an increasing popularity of blogs which are used for publishing and sharing ideas among users. Therefore, the authors make an attempt to detect "innovative topics" in blog communities, stating that "information sharing systems for blogs could enable users to expand their interests by browsing the collections of blog entries published by other users." (Nakatsuji, Yoshida and Ishida 2009). They define "innovative topic" as "[…]a topic that includes new concepts that are likely be interesting to the user even though those concepts are not present in the user profile[…]" (Nakatsuji, Yoshida and Ishida 2009).

Nakatsuji et. all (2007) state that the idea of "innovation topics" detection in blog communities is actual and novel, because current information retrieval in blog services relies only on keyword searches of blogs using Google or based on simple metadata such as that of an RSS. Furthermore, there is no any special function to generate personalized searches easily, so users need to consider and enter search keywords that suit their own interests appropriately. Such a keyword search is time consuming and troublesome (Nakatsuji, Yoshida and Ishida 2009). Hence, in order to counteract aforementioned problems, Nakatsuji et. all (2007), (2009) in their research papers propose to construct user profile automatically as a user-interest ontology, which is a class hierarchy of user interests with interest weights. Then, the researchers suggest measuring the similarity of interest ontologies considering the degree of interest agreement to each class and instance. Afterwards, they apply the techniques to help users create a blog community by browsing innovative blog entries which include information unknown to users with a high probability of being interesting.

More precisely though their method can be performed in three steps. In the course of the research, Nakatsuji et. all (2007) (2009) approaches the goal of detecting innovative topics in three main steps: (1) extracting an interest ontology by classifying users' blog entries into a service domain ontology; (2) generating interest-sharing group by measuring similarity between ontologies; (3) detecting innovative instances by analyzing differences of class hierarchy between ontologies. As a result, the authors propose user-interest ontology generation algorithms that allow to detect "innovative topics" for a user by analyzing the ontologies of user group that have a high similarity to the ontology of the user (Nakatsuji, Yoshida and Ishida 2009).

B. VAN DER MEER Research Group

VAN DER MEER (2007) covers the subject of the practical application of the principles of open innovation. He determines "innovation is the total set of activities leading to the introduction of something new, resulting in strengthening the defendable competitive advantage of a company". In his research done in a form of case study he invistigates innovation from cultural (e.g. creation of an innovative climate) and structural (e.g. systematic use of innovation mechanisms) perspectives of organization. While performing the research, van der Meer (2007) comes to the conclusion that innovating requires a different way of thinking and creation a different culture that would nurture the desire to work with smart people inside and outside the company, to create external and internal R&D, and to build a better business model (van der Meer 2007). Moreover, innovation also requires different organizational structure which is closer and where mechanisms for importing and exporting knowledge, ideas and projects are more transparent.

C. RIEDERER AND BAIER Research Group

RIEDERER AND BAIER (2005) investigate the notion of innovation, the same as van der Meer (2007), from the business perspective. However, compare to van der Meer (2007), they do not keep the definition of innovation abstract, "something new" (van der Meer 2007), but they try to give an overview of different types of innovation. Thus, under company's perspective they define product and process innovations which can be categorized into sustaining and disruptive ones (Koudal 2004), (Riederer and Baier 2005). As Riederer and Baier (2005) say, "sustaining innovations are those which improve an already existing product or process". Then, disruptive innovations are innovation which are radically different from any process or product or service that previosly existed on the market (Riederer and Baier 2005). The same as van der Meer (2007), Riederer and Baier (2005) support an idea that in order to generate innovation within the company, this company has to maintain constant innovation culture which encourges creativity, evaluation, implementation and measurement if innovative ideas. Moreover, the company has to have different sources of innovation ideas. For example, such sources could be suppliers and customers of the company (Riederer and Baier 2005).

D. Wallace Research Group

Wallace (2012) in his dissertation on the relationship of organizational learning to knowledge management and its impact on innovation extends the definition of innovation by considering three main facet of it: products, processes and ideas which are "interplay to keep businesses competitive in the future by adding economic value and the growth necessary in the marketplace". He defines product component of idea as a flow of new products which plays a role of implementation of innovation. Processes are the resources for innovation which play a role of translation of innovation idea into something tangible. Finally, the third part is that of intelligence, ideas and knowledge which is the idea component of innovation (Wallace 2012). In his research he illuminates that organizational learning and knowledge management are important determinants of innovation emergence. According to Wallace (2012), "to manage ideas across an organization the organization needs

to have an idea of what it is searching for and why, a means to distribute that new idea or concept to leverage the collective wisdom of the organization and a way of reinforcing new behaviors in order to take advantage of what has been discovered". Thereby, there is a need to have a knowlegde of what has to be innovated and then when it appearce the learning process will help to realize the idea or concept of innovation.

E. Conclusion From the Review

Nakatsuji et.all defines innovation as new topic or interests that might be attaractive to internet users. Since from the business perspective blog users are just employees,

Both researches of van der Meer (2007) and of Riederer and Baier (2005) deal with innovation looking at it from the angel how it can be created within the company. They mention the idea component of innovation as if it is something which is naturally incorporated in innovation.

Wallace et al., made a decomposition of innovation within the organization, and detected what is need in order to accumulate the innovation for business purposes.

Hence after analizing the papers above, the problem of innovation detection has one of the best proprities in innovation management science nowadays and it is crucial to detect innovations automatically in information field.

IV. Research Aim And Conditions

Employees generate information (new terms and n-grams) every day in forms of e-mails, creating database records etc. The amount of information generated could be represented in graphical form depending on time where the distribution of each dot is a frequency of usage of term.

Our main task was to detect how behaves this distribution, and how the speed of change varies in time. In order to achieve our goal we applied an approximation function that draws a trend line using the method of the least squares.

To facilitate the processing we used the second degree polynomial function in the following form:

$$f(x) = Ax^2 + bx + c \qquad (1)$$

Tracking the distribution behavior we accessed critically the main hypothesis of semantic innovation detection:

It is possible to detect innovations in corporate information field observing unexpected increase in frequency of (1) new terms application (that either were used seldom or even not used before), or (2) new combinations application featuring "old" notions.

In order to approve or reject the hypothesis we've processed the corporate data using the created C# program that goes through following steps:

1) Finds terms and n-grams in corporate e-mail exchange that are used not less than several times a month and not less than 10 times a year;
2) Detects approximation against second order polynomial of the distribution function of relative frequencies of term and n-grams; also the modulo of approximation error is counted;
3) Draws a chart of distribution of polynomial's senior member.

To reduce the experimental error of this research we required a large corporate database with a long activity history. For this need we've selected an ambitious Russian software integrator I.T.C.o, a large IT company that has variety of businesses and operates on multinational markets.

Being granted to access the corporate database we started collecting the data. The information field observance period has been chosen long enough (not less than half a year), that enables the increase of statistical significance of measurement and excludes seasonal factors influence.

V. Description

After step 2 we stored the mined data into the Table I. For each popular term (mostly terms were in Russian language since the database belonged to Russian company), we counted the coefficient of highest degree (senior) member A from the polynomial function *f(x)* and the approximation error modulo R.

TABLE I: Mined Data From the Corporate Mail Exchange

Term (translation from Russian)	Coefficient of senior (highest degree) member of approximation from the f(x)	Approximation error modulo R
LEARN	-0,000122	2,292952
"ДАТА"(date)	0,000016	0,729728
PROGS	-0,000079	1,242225
"ЗАКАЗЧИК" (client)	0,000009	0, 683106
"РЕАКЦИЯ" (reaction)	-0,000015	1,357177

After processing the corporate information of I.T. Co we summarized the results in two most distinct charts (Fig. 1, Fig. 2).

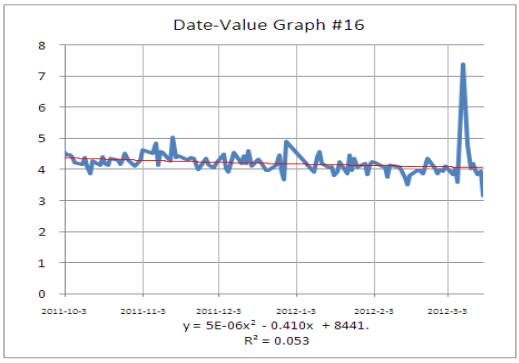

Fig. 1. The quantity of tern utilization according to dates

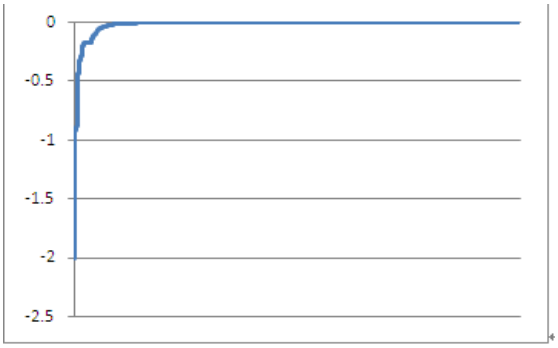

Fig. 2. Distribution of coefficients by the senior (highest degree) member of approximation polynomial

Fig. 1 shows the relative quantity of term utilization (bi-gramms) in certain moment of time.

The peak on 12/03 shows an increased term (bi-gram) application during this day.

Fig. 2 shows the distribution of A of an approximation polynomial f(x) (hint: function was approximated by the second order polynomial; for polynomial's coefficients detection *technique of least squares* was used).

We have observed only those coefficients for that the approximation error was less the chosen α. Chart 2 shows that the majority of A coefficient's values are located near zero, it actually means that, the increase or decline of term application frequency has a linear distribution and is not an innovation. For innovations relevant are terms and bi-gramms, whose senior coefficients match the top-right area and the increase of its usage is non- linear. On the opposite left-bottom area of the chart contains terms and bi-gramms that are excluded from corporate information field, for which decline in term usage inside the corporate information field is non-linear. There is also a variety of terms that belong to *semantical noise*. To exclude such terms we used a vocabulary that contains the full list of employee names and other popular notions. We compared each term that could be an innovation with those list and made the particular decision.

VI. Conclusion

According to Fig (1-2) the logical implication is easily seen: the majority of terms and bi-grams has a stable frequency of use in company that does not change significantly in time or has a small growth. Our research has shown that, in corporate information field there are terms and bi-grams that differ from the main trend these have a rapid growth of use, that clearly is a sign of corporate innovation appearance or a sign of those fact that, some part of companies business activity is abandoned or has a remarkable problems there (fires people that possess unique competences).

For instance, just after integration of german subject-oriented business process optimization system Metasonic such growth has shown the Russian term "МЕТАСОНИК".

We propose to use the described set of algorithms enabling the detection of terms with significant frequency of appearance in the information field fluctuation. Disclosed terms and bigrams, which are characterized by significant increase of utilization, can be referred to the innovation appearance in the frame of the company. We state that analyzing the information flows using the described method is vital step in enterprise innovation management.

Acknowledgments

The given research was held in a frame of the contract № 13.G25.31.0096 with the Ministry for Education and Science of Russian Federation «Creation of hi-tech manufacture of unstructured information processing in cross-platform system on the open source software basis in order to increase management efficiency of innovative activity of enterprises in modern Russian economics.

References

[1] A. Z. Spector, "Achieving application requirements," *In Distributed Systems, S. Mullender, Ed. ACM Press Frontier Series*, ACM, New York, NY, pp. 19-33, 1989.

[2] K. Peter, "Mastering Innovation – Exploiting Ideas for Profitable Growth," Delloite, 2004.

[3] N. Makoto, M. Yoshida, and T. Ishida, "Detecting innovative topics based on user-interest ontology," *Web Semantics: Science, Services and Agents on the World Wide Web*, pp. 107-120, 2009.

[4] N. Makoto, Y. Miyoshi, and Y. Otsuka, "Innovation Detection based on User-Interest Ontology of Blog Community," *Innovation*, pp. 1-14, 2007.

[5] P. R. John and M. Baier, "Innovation Management – An Overview and some Best Practices," *Communication*, 2005, pp.1-58.

[6] V. D. M. Han. "Open Innovation? The Dutch treat: Challenges in Thinking in Business Models," *Creativity and Innovation Management*, 2007, pp. 192-202.

[7] B. W. Richard, "The Relationship of Organizational Learning to Knowledge Management and its Impact on Innovation," *The Relationship of Organizational Learning to Knowledge Management and its Impact on Innovation by Richard B. Wallace. Colorado Springs: ProQuest LLC*, May 2012.

[8] K. Peter, "Mastering Innovation – Exploiting Ideas for Profitable Growth," *Delloite*, 2004.

[9] N. Makoto, M. Yoshida, and T. Ishida, "Detecting innovative topics based on user-interest ontology," *Web Semantics: Science, Services and Agents on the World Wide Web*, pp.107-120.

[10] N. Makoto, Y. Miyoshi, and Y. Otsuka, "Innovation Detection based on User-Interest Ontology of Blog Community," *Innovation*, pp.1-14.

[11] P. R. John and M. Baier, "Innovation Management – An Overview and some Best Practices," *Communication*, pp.1-58.

[12] V. D. M. Han, "Open Innovation? The Dutch treat: Challenges in Thinking in Business Models," *Creativity and Innovation Management*, pp.192-202.

[13] B. W. Richard, "The Relationship of Organizational Learning to Knowledge Management and its Impact on Innovation," *The Relationship of Organizational Learning to Knowledge Management and its Impact on Innovation by Richard B. Wallace. Colorado Springs: ProQuest LLC*, May 2012.

[14] P. A. Gloor, R. Laubacher, S. B. C. Dynes, Y. Zhao, "Visualization of Communication Patterns in Collaborative Innovation Networks - Analysis of Some W3C Working Groups, Proceedings of the twelfth international conference on Information and knowledge management," *Novembe*r 03-08, 2003, New Orleans, LA, USA