Interrelating Products through Properties via Patent Analysis

P.-A. Verhaegen[1], J. D'hondt[1], J. Vertommen[1], S. Dewulf[2], J. R. Duflou[1]
[1]Centre for Industrial Management, Dept. Mechanical Engineering, K.U.Leuven, Celestijnenlaan 300A Bus 2422, B-3001 Heverlee (Leuven), Belgium
Corresponding author: paularmand.verhaegen@cib.kuleuven.be
[2]CREAX N.V., Ieper, Maarschalk Plumerlaan 113, 8900 Ieper, Belgium
simon.dewulf@creax.com

Abstract
TRIZ emerged from systematic analysis of patents, a process involving the mapping of innovative patents to extracted generic problems and generic inventive principles. During problem solving, TRIZ users, relying on their TRIZ skills, map their specific problem to a generic problem, solve it via TRIZ tools, and map back to a specific solution. A methodology and algorithm are proposed that, through identification of specific word categories in patents, analysis of term-term correlation data, and data mining techniques, automatically identify similar products, and properties relating or differentiating products. This algorithm can quantifiably guide creativity efforts and aid in patent portfolio management.

Keywords:
Systematic Innovation, TRIZ, Creativity, Patent Analysis, Data mining

1 INTRODUCTION

1.1 Need for patent mining

In 2007, the United States Patent and Trademark Office (USPTO) granted 157283 patents of which most to corporations [1]. Patent offices assign each patent to one or multiple classes, categorizing them in a hierarchical system based on topic or technological area, such as the US or IPC patent classification schemes, through which patents related to an application area can be searched, e.g. IPC groups A45B 11 to A45B 19, A45B23 and A45B25 all relate to patents covering umbrellas.

Hundreds of companies' patent portfolio increases by 40 to over 3000 patents per year [1], making portfolio management ever more complex, and demanding tools and techniques enabling the discovery of market opportunities outside the application area of the organization's own technology as well as the identification of possibilities to license in complementary technology.

While commercially available patent databases offer full text and more specific search features based on different patent fields, such as citations, applicant, inventor or issue date fields, these functionalities do not allow a company developing umbrellas to directly search for market opportunities, or to identify complementary technology.

This research facilitates reaching these two objectives by proposing automatically identified similar products, and the properties relating or differentiating these products.

1.2 TRIZ

TRIZ is the Russian acronym for the Theory of Inventive Problem Solving, and encompasses a series of tools and a methodology for generating innovative ideas and solutions for problem solving. It was formed through the systematic interactive analysis of what TRIZ practitioners estimate to be one and a half to three million patents, from which forty thousand innovative patents were withheld and their applied innovative solutions were mapped onto a small number of extracted inventive principles.

TRIZ is based on three postulates [2] [3]:

- The Postulate of Existing Objective Laws states that engineering systems evolve according to a set of laws;
- The Postulate of Contradictions states that, in order to evolve, an engineering system has to overcome one or more contradictions; and
- The Postulate of the Specific Situation states that the problem solving process should take into account the specific problem peculiarities.

Derived from this patent analysis and based on the postulates, a set of TRIZ tools was conceived, of which the most popular are [4]:

- The Contradiction Matrix to solve technical contradictions;
- The Separations Principles to solve physical contradictions;
- Substance-Field (SU-Field) modeling and the Inventive Standards to transform technical systems;
- ARIZ as a list of logical procedures for eliminating contradictions; and
- TRIZ Trends as a system of laws that govern engineering system evolution.

TRIZ incorporates the idea of mapping a specific problem to a more general problem specification, solving this generic problem via the TRIZ toolset, and mapping back the generic solution to the specific problem. This enables TRIZ users to benefit from the generalized inventive solutions outside their fields of knowledge, but also relies heavily on the user's TRIZ skills.

As TRIZ users are interested in analogous inventions in other fields, or technological areas, that solve the same contradictions, the analyses can not easily be automated by simple search functionalities for other patents based on the IPC classes or patents fields.

Instead of requiring the user to map to and from the generic TRIZ solution space, the proposed algorithm directly relates products with other products from other technological areas with similar product properties, and assumes that the observed contradictions may already be solved in products with similar properties.

The following section gives an overview of related research on data mining of the structured and unstructured fields of patents, and on innovative idea generation. The third section describes the proposed methodology, while the fourth illustrates this methodology with a case study. The final section formulates the conclusions.

2 RELATED RESEARCH

Research has been conducted to automatically infer structure from non-text patent fields. Citation analysis permits functionalities such as the identification of major competitors, the construction of technology indicators, and documentary search possibly identifying related technologies and applications [5]. This analysis is based on references given by the applicant, which is optionally for the European Patent Office, and which are screened by the patent office, thus basing any further analyses on already known product or technology relations, and excluding e.g. new applications domains. In addition, most patents never get cited or only begin to get cited after several years [6].

Patent text fields, such as the title, summary, description or claims fields, contain vital information about the patent, and can be subjected to text mining techniques, which extract relevant information from less structured textual data through use of keyword extraction, pattern recognition, linguistic analysis, and statistical techniques. A series of text mining techniques for patent analysis is presented and evaluated in [7].

In [8], Yoon and Park propose a network-based analysis as an alternative to citation analysis. This methodology is based on keyword extraction and linking patents based on the occurrences of these keywords, instead of citations between patents. It allows users to visually identify patent network structure, such as central patents, or disjoint groups.

[9] proposes a case-based reasoning methodology and product innovation retrieval system (PIRS) for retrieving similar products based on 87 user-centered design (UCD) attribute dimensions. The techniques relies on a large database of products scored on these attributes, a manual process performed by UCD experts. The functions of the identified products are candidate ideas for the product under investigation. Compared to the methodology proposed in this research, the PIRS system can not retrieve products which are not manually analyzed and inserted in to database. The use of the UCD attributes furthermore causes products similar only in these attributes to be retrieved.

In [10], Yoon and Park describe a morphological analysis methodology based keyword dictionary developed by text mining patent and factor analysis on the terms. The morphology of all patents is identified, and technology gaps within a product or technology can be identified.

The commercially available Goldfire Innovator™ from Invention Machine [11] has a semantic engine to infer Subject-Action-Object (SAO) from plain text sentences in patents and queries and offers several TRIZ inspired idea generating functionalities based on an indexed database of these SAOs.

Other research by Cascini [12] [13] describes algorithms to automatically analyze patent text fields revealing the invention's components, architecture, and positional and functional interrelations, and aiding in identifying the solved TRIZ contradictions.

Research by He Cong and Loh Hang Tang [14] proposes a text based expert system which allows classifying patents according to TRIZ inventive principles. Similar research by the same authors [15] proposes an automatic patent classification system based on clustering to categorize patents in TRIZ inventive principles, and evaluates the performance of different clustering algorithms on the selected text features.

Other research by Cavallucci [16] proposed and validated the possibility to incorporate the eight original Altshuller's laws of development in the design process on a manifold case study. Based on TRIZ and domain knowledge, the conclusions concerning the development potential can be translated into specific directions for future improvements of the manifold.

In [17] and [18], Mann and Dewulf propose the concept of evolutionary potential, which is similar to the approach proposed by Cavalucci [16], but using more specific TRIZ trends or lines of evolution allows for a more actual and specific categorization.

Later research in Directed Variation by Dewulf [19] suggests depicting the product on a radar plot of property spectra, instead of trends or lines of evolution, e.g. rigid, jointed and flexible are all properties of the spectrum flexibility. While classical TRIZ assumes the evolution usually occurs in a certain direction along the trends, Directed Variation regards changes of properties towards both directions in a spectrum as variations to ensure certain functionalities of a product, e.g. for the surface spectrum, evolving towards the flat property can decrease resistance, while evolving towards the protruded side of the surface spectrum can increase grip or allow faster cooling. The radar plot of property spectra of a product, or product DNA, can be compared to the DNA of other products to find similar products. Figure 1, copied from this research, compares the product DNA of sugar and dish washing tablets, graphically illustrating the similarity and dissimilarity among the products, potentially inspiring the creativity of engineers.

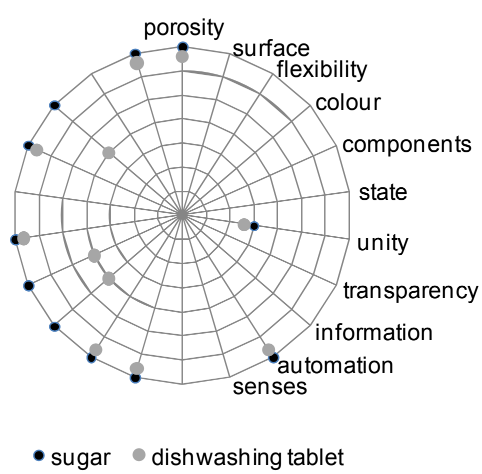

Figure 1: Product DNA comparison of sugar and dishwashing tablets [19]

Dewulf also identifies the link between adjectives and product properties, and between verbs and product functions. This research builds further upon this idea, proposing a method for automatic extraction of product properties and automatic comparison of products, and suggesting directions for creative efforts. This enables the discovery of market opportunities outside the application area of the organization's own technology as well as the identification of possibilities to license in complementary technology.

3 METHODOLOGY

This research proposes an algorithm and framework that, through patent analysis and identification of word categories, can extract information concerning the properties of a given product or product family, which in turn allows to identify properties relating or differentiating two products. Other functionalities based hereon are the finding of similar products. These algorithms can assist in steering the creative efforts of the R&D department in a formalized and quantifiable manner, and aid searching for market opportunities, or identifying complementary technology in the context of patent portfolio management.

3.1 Gathering properties

Currently several modules of a test platform have been implemented, some of which have been graphically depicted in Figure 2.

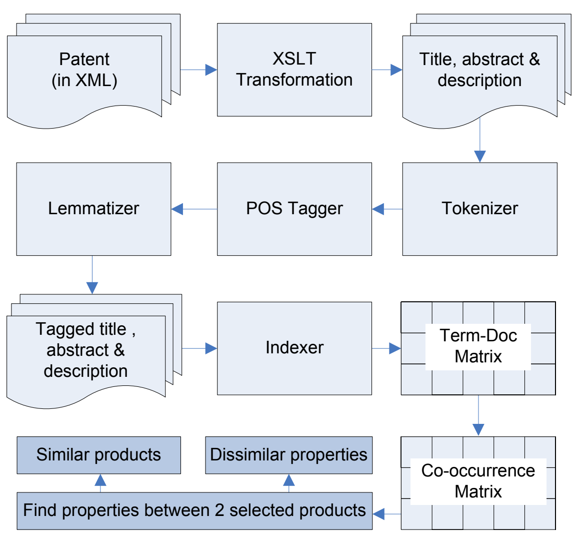

Figure 2: Modules of the platform

Patents written in English are converted into structured XML files, which are fed into a XSLT transformation module retaining only certain patent sections for further processing. Some patent sections contain specific numerical or textual information, such as patent number, date of application and authors. Other, more narrative, patent sections are:

- The title of the invention;
- The abstract;
- The claims section;
- The background section;
- The summary section;
- The description section;

[15] indicates the importance of including the titles and abstracts in the automatic classification of patents, while the summary section gives only marginal improvements. Other research [20] [21] shows that the inclusion of a certain number of words or lines of the description, applications and/or claims can be beneficial to patent classification. In the proposed approach, only the title, the abstract and the description sections are retained, although the additional benefit of processing the claims section too will be analyzed at a later stage. For most patents, the title and the abstract are available in English, which is not always the case for the other patent sections, such as the description section.

The XSLT transformation concatenates the text contained in the title, abstract and description fields and pipes this text on a per patent basis to the tokenizer module, which splits the text into a set of tokens to be interpreted by a Part-Of-Speech (POS) tagger. The tokenizer recognizes a manually assembled list of multiwords, e.g. 'de facto', which are then regarded by all subsequent step as being one word.

A TnT Tagger [22] is used to POS tag the text to the CLAWS5 tag set. This tagger is trained on a different set of patents in order to adapt the configuration files to the specific language used in patents. The trained tagger proves to correctly tag a word in more than 95 % of the cases. This 5 % error includes a number of words incorrectly tagged as adjectives which should have been identified as nouns that modify other nouns, or attributive nouns or noun adjuncts [23]. This misclassification occurs when the tagger encounters constructions such as 'loudspeaker system', 'textile cover', 'volume control', or 'earphone jack'.

It should be noted that [7] describes a method 'Keyword and phrase extracting' which allows for the identification of multiword phrases, based on the assumption that these multiwords would occur several times in the document. However, currently such a functionality is not implemented and further research will evaluate the usefulness of further decreasing the number of errors due to such misclassification.

The stream of tokens is then run through a rule based lemmatizer described in [24], which allows to normalize these words based on the given POS tag to the form used as the headword in a dictionary, e.g. 'cooled' and 'cooling' both map to 'cooled'. This step maps some misspelled word suffixes to a common lemma, but the main advantage of this strategy over the use of a Porter stemmer [25], which removes the word's suffixes, is that the resulting terms and further analyses are easier to interpret by humans.

3.2 Property selection

Currently, only nouns and adjectives are withheld for further processing. However, no filtering is done to extract only relevant adjectives or nouns. In a later stage of this research only selected adjectives will be processed through a property selection phase explained below. The result of this step is a list of adjectives and nouns as input for the indexer.

[19] defines a property as 'what a product is or has', its attributes. This is mainly expressed in adjectives and is related to physical parameters. Examples of properties are hollow, smooth, transparent, strong, and flexible. These are all generic, in contrast to product specific attributes, for example light weight or inspectable. These product specific attributes are related to functional requirements. A generic property such as hollow can lead to a product specific property as light weight, just as transparent can lead to inspectable. The link between adjectives and properties was further examined by the authors in [26].

Currently, only adjectives are processed in the property selection phase as it is assumed that adjectives can express system properties [19] and that adjectives are less domain specific than nouns. Further research will investigate the effects of including other word categories, such as verbs.

As a prerequisite for this research, the authors validated the possibility to identify clusters of adjectives which relate to the same generic product property. From a random sample of 22684 non-chemical patents, the process described in previous section produces a list of 81750 adjectives, of which 69260 only occurred in a single patent, and are discarded for further processing. The remaining 12490 adjectives are run though a Porter stemmer [25] resulting in 10361 different stems. A Porter stemmer was preferred over a lemmatizer because this analysis is only performed once by TRIZ experts, and the results are never interpreted by the users of the system.

A 10361 by 22684 term document matrix is constructed, weighted with a Term Frequency Inverse Document Frequency (tf-idf) scheme [27], and normalized to account for different patent text lengths. A singular value decomposition (svd) step [28] is performed to reduce the number of dimensions before clustering. Most related research uses a value around 300 as a rule of thumb for the number of reduced dimensions for a similar sized collection [29]. Through experimentation, this value was set to 1000 to ensure enough discriminatory power through the explained variance, possibly leading to overfitting the model, which is less an issue as the results are manually analyzed by TRIZ experts as explained below. The terms are then grouped by clustering to 700 clusters, a value experimentally determined through sweeping this variable.

Table 1 shows the first 5 clusters with the contained adjectives. These results can be used as an aid to manually identify adjectives that describe system properties. This manual step, currently being performed by TRIZ experts, is still needed because the results are too noisy for full automatic extraction.

Cluster Number	**Clustered adjectives (or noun adjuncts).** Related terms are in *Italic*	**Relationship**
1	Teletypewrite, *prefinished, preparatory, preassembled*	Preliminary action
2	Degaussed, *activated, deactivated, deactived*	Activation, time segmentation
3	Boss, *multilayered, layered*	Layered
4	Dislodged, bumper, cental[1], radiated	(Cars)
5	*Seam, ring, weak*	Segmenting, or attaching

Table 1: Examples of adjective clusters, manual identification of related adjectives and description of the relationship

It can be seen that cluster 4 relates to certain car parts, and not to a generic property, which can be detected in the manual step. The adjectives or noun adjuncts from this cluster should therefore not be withheld for further processing.

As stated in the first paragraph of this section, the research on this adjective or property filtering is not yet concluded. Therefore, no filter is currently applied and all adjectives are input to the processing in further steps. In a later stage the results from the clustering and manual verification will be used to identify the adjectives relevant for further processing.

Besides adjectives, and to allow product or technology identification, mostly expressed by nouns, this word category is also processed in further steps. Further research will examine the possibility to only retain nouns occurring in the headings of the IPC class hierarchy, such that nouns directly indicate a product, or technology.

3.3 Information from term-term correlation data

A slightly modified open source program Lucene [30], that only indexes a given list of adjectives and nouns and outputs results in human readable format, implements the indexer module of the test platform. This data is then read into a term document matrix A, in which each element A_{ij} represent the number of times term i occurs in patent j. From this, the term-term correlation matrix C is calculated as AA^T [27], in which C_{ij} is the sum, over all patents, of the product of the number of times a term i and the number of times a term j occurs in a patent.

Given two nouns that characterize two different products or product families, the term-term correlation matrix allows looking up adjectives which co-occur with these two nouns. These adjectives directly interrelate these two products, or product families. This methodology is illustrated by the term-term correlation matrix in Figure 3. The matrix elements 'X' indicate that the adjective co-occurs with noun 1 and noun 2, linking the product 1 with product 2 through this shared adjective.

It is at least equally important to find the property dimensions differentiating the two products. By looking at the differences in term-term correlations of an adjective with the product nouns, this methodology can be used to highlight the differences between the two products. This information can be used to transfer knowledge from one product to the other.

Given the assumption that adjectives relate to generic product properties, this technique allows to automatically calculate the degree of similarity of two products along these property dimensions. As rough indication the sum of the term-term correlations of all adjectives with the two nouns or products can be calculated.

As can be seen from the case study in the next section, some identified adjectives, or noun adjuncts, do not relate to generic product properties and should not be included in the analysis. This adjective filter, or property selection, explained in section 3.2, should allow filtering the adjectives based on relevance to generic product properties.

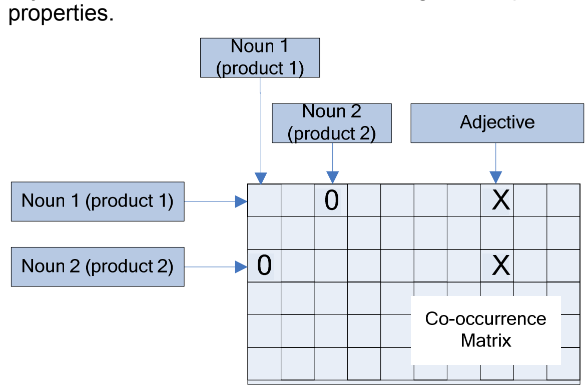

Figure 3: Co-occurrence matrix used to calculate the similarity between products

[1] Analysis of the patents reveals that «cental» does not refer to the weight unit, but should be interpreted as a misspelled version of the word «central».

This methodology can also be used to find products related to noun 1 or product 1. Looping over all different nouns, or products, permits comparing the closeness figures of different products, and finding related products in large patent databases. Adding the constraint that the two selected nouns may not co-occur in any of the patents, which is illustrated by the ‘0’ elements in the term-term correlation matrix of Figure 3, this allows to find directly unrelated products in different technological areas having similar in product properties. Depending on the size of the patent database this constraint can be implemented by a threshold value different of zero.

The methodology infers a link between two not directly related products. Such a higher order co-occurrence can also be found by techniques such as singular value decomposition, but these techniques complicate the interpretation of the property dimensions as these are linear combinations of the adjectives. In this light, section 3.2 can be seen as a manual step to ease this interpretation.

4 CASE STUDY

To illustrate the proposed methodology, the title, abstract and description sections of a random set of 64529 patents were tokenized, lemmatized and POS tagged.

The identified adjectives and nouns, collectively called terms, are mapped onto their lemma through a rule based lemmatizer. In a next step, specific chemical terms are identified and the patents in which a certain number of these terms occur were discarded from further processing. This was primarily done because the results from the chemical domain are less easily interpreted by the researchers due to their background. It can be envisaged to also exclude a list of words commonly found in patents [7].

The lemmatized terms are stored in an index file associating them with the patents in which they are found. This data is imported in a term-document matrix A, from which the term-term correlation matrix C is calculated.

The nouns ‘umbrella’ and ‘windscreen’ are selected for analysis, and based on the correlation matrix a list of adjectives co-occurring with these two nouns is retrieved. Under the assumption that these adjectives related to product properties, these adjectives directly link the two products.

Figure 4 presents the list of adjectives co-occurring with the nouns ‘umbrella’ and ‘windscreen’. The values in the bar chart represent the minimum of term-term correlation values of the adjectives with each of the nouns. It is noteworthy that the noun ‘windscreen’ is used to indicate ‘a screen for protecting something from wind’ and ‘a windshield of a motor vehicle’ [31], which explains the high occurrence of some adjectives, e.g. aerodynamic. These figures indicate that windscreens and umbrellas are both foldable and collapsible, and both products’ patents cover aerodynamic properties. Not all resulting adjectives relate to different product properties, and some are similar, e.g. foldable and collapsible. To facilitate the interpretation of the figures, the adjectives could be grouped in meaningful clusters, as explained in section 3.2.

Figure 5, a bar chart indicating individual term-term correlation values of the same adjectives with the umbrella and windscreen nouns, illustrates how the methodology can be used to highlight the differences between the products umbrella and windscreen. A designer can use this information to transfer knowledge from one product to the other, e.g. making a foldable windscreen based on the knowledge from the umbrella product family.

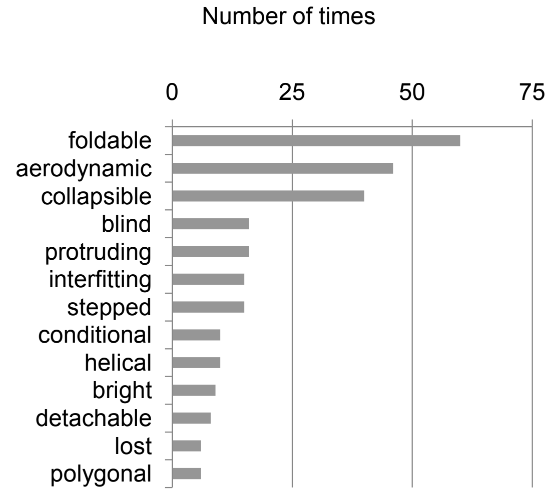

Figure 4: Minimum of term-term correlation values of the adjectives with the windscreen and umbrella

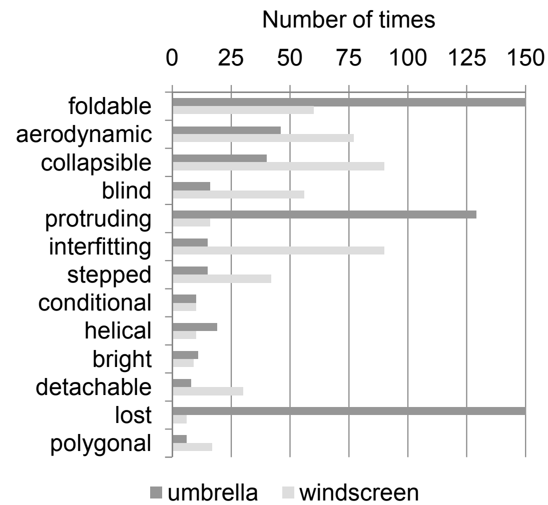

Figure 5: Term-term correlation values of the adjectives indicating the differentiating properties between umbrella and windscreen

An indication of the degree of closeness of two products along their property dimensions is given by the sum of the term-term correlation values, e.g. for the umbrella and windscreen products, this value is 384. This figure currently does not account for the length or number of patents, and further research will study the necessity of a normalization step.

Comparing the closeness figures of different products with the umbrella product, allows finding related products in large patent databases, e.g. products which are similar to umbrellas in terms of product properties. Adding the constraint that the two selected nouns may not co-occur in any of the 64529 patents in our database, the closest noun to the umbrella noun is ‘slider’, indicating that the designer could be inspired by looking at the slider product family.

The relevance of this result is verified by the fact that in our patent database no patent contains both the words umbrella and slider. An online search on a global patent database [32] reveals that of the 35388 patents covering umbrellas, 805 contain the word ‘slider’. This indicates that the proposed algorithm can find products with similar properties to the product under investigation, and allows to steer creative efforts.

5 SUMMARY

By means of a case study comparing the umbrella product category with windscreen products, it was shown that based on term-term correlation data between adjectives and nouns, the proposed methodology allows to automatically find product properties related to both products, and list these in order of relevance.

It was also shown that further analysis of the term-term correlation matrix permits finding properties which co-occur more often with one of the two nouns, enabling the extraction of properties differentiating the products.

By looping over different extracted nouns, the proposed methodology furthermore allows to automatically search for related products. This was demonstrated by the identification of the slider product, which is closely related to the given umbrella product, but not occurring with the umbrella product in database applied.

6 REFERENCES

[1] United States Patent and Trademark Office, 2008, Patenting by Organizations 2007.

[2] Khomenko, N., Ashtiani, M., 2007, Classical TRIZ and OTSM as a Scientific Theoretical Background for Non-Typical Problem Solving Instruments, Proceedings of the TRIZ-Future Conference, 73-80.

[3] Altschuller, G S., 1984, Creativity as an Exact Science - The Theory of the Solution of Inventive Problems, Gordon and Breach Science Publishers, New York.

[4] Savransky, S. D., 2000, Engineering of Creativity: Introduction to TRIZ Methodology of Inventive Problem Solving, Boca Raton, Florida.

[5] Michel, J., Bettels, B., 2001, Patent citation analysis: A closer look at the basic input data from patent search reports, Scientometrics, 51/1: 185-201.

[6] Karki M.M.S., 1997, Patent Citation Analysis: A Policy Analysis Tool, World Patent Information, 19/4: 269-272.

[7] Tseng, Y.-H., Lin, C.-J., Lin Y.-I., 2007, Text mining techniques for patent analysis, Information Processing and Management: an International Journal, 43/5: 1216-1247.

[8] Yoon, B., Park, Y., 2003, A text-mining based patent network: Analytical tool for high-technology trend, Journal of High Technology Management Research, 15: 37-50.

[9] Wu, M.-C, Lo, Y.-F., Hsu, S.-H., 2006, A case-based reasoning approach to generating new product ideas, The International Journal of Advanced Manufacturing Technology, 30: 166-173.

[10] Yoon, B., Park, Y., 2005, A systematic approach for identifying technology opportunities: Keyword-based morphology analysis, Technological Forecasting & Social Change, 72: 145–160.

[11] Goldfire Innovator, Invention-machine.com. Available from http://www.invention-machine.com

[12] Cascini, G., Russo, D., Zini, M., 2007, Computer-Aided Patent Analysis: Finding Invention Peculiarities, Monterrey Nuevo León, Mexico.

[13] Cascini, G., Russo, D., 2007, Computer-Aided Analysis of Patents and Search for TRIZ Contradictions, International Journal of Product Development, 4/1-2: 52-67.

[14] He, C., Loh H. T., 2008, Grouping of TRIZ Inventive Principles to facilitate automatic patent classification. Expert Systems with Applications, 34/1:788-795.

[15] Loh H. T., He, C., Shen, L., 2006, Automatic classification of patent documents for TRIZ users, World Patent Information, 28/1:6-13.

[16] Cavallucci, D., 2002, Integrating Altshuller’s Development Laws for Technical Systems into the Design Process, Annals of the CIRP, 50/1:115-120.

[17] Mann, D., 2002, Better Technology Forecasting using Systematic Innovation Methods, Technological Forecasting and Social Change, 70/8:779-95.

[18] Mann, D., Dewulf, S., 2002, Evolutionary-Potential in Technical and Business Systems, The TRIZ Journal.

[19] Dewulf, S., 2006, Directed Variation: Variation of Properties for New or Improved Function Product DNA, a Base for 'Connect and Develop', Proceedings of the ETRIA TRIZ Future Conference, Kortrijk, Belgium.

[20] Larkey L.S., 1999, A patent search and classification system, Proceedings of the fourth ACM conference on Digital libraries, ACM Press, New York, US., 179-187.

[21] Fall, C J, et al., 2003, Automated Categorization in the International Patent Classification, SIGIR Forum, ACM Press, 37/1:10-25.

[22] Brants, T., 2000, TnT - A Statistical Part-Of-Speech Tagger, Proceedings of the Sixth Applied Natural Language Processing Conference ANLP-2000, Seattle.

[23] Noun Adjunct., Wikipedia, Available from http://en.wikipedia.org/wiki/Noun_adjunct

[24] Carl, M., Schmidt, P., and Schütz, J., 2005, Reversible Template-based Shake & Bake Generation, Proceedings of Workshop Example-Based Machine Translation, MT Summit X, Phuket, Thailand.

[25] van Rijsbergen, C. J., Robertson, S.E., Porter, M.F., 1980, New models in probabilistic information retrieval, British Library Research and Development Report, 5587.

[26] Verhaegen, P.-A., Vertommen, J., D'hondt, J., Dewulf S., Duflou, J., 2008, Relating Properties and Functions from Patents to TRIZ trends, CIRP Design Conference 2008: Design Synthesis, Twente.

[27] Baeza-Yates, R. A., Ribeiro-Neto, B., 1999, Modern Information Retrieval, Addison-Wesley Longman Publishing Co., Inc., Boston, MA.

[28] Berry, M. W., Dumais, S. T., O’Brien, G. W., 1995, Using Linear Algebra for Intelligent Information Retrieval, SIAM Review, 37/4: 573-595.

[29] Berry, M. W., Dumais, S. T., Shippy, A.T., 1995, A Case Study of Latent Semantic Indexing, Technical report, UT-CS-95-271, Knoxville, TN, USA.

[30] Apache Lucene, available from http://lucene.apache.org.

[31] The Free Dictionary, available from http://www.thefreedictionary.com/windscreen.

[32] Micropatent, Thomson Reuters, available from http://www.micropatent.com.