Journal of Physics: Conference Series

PAPER • OPEN ACCESS

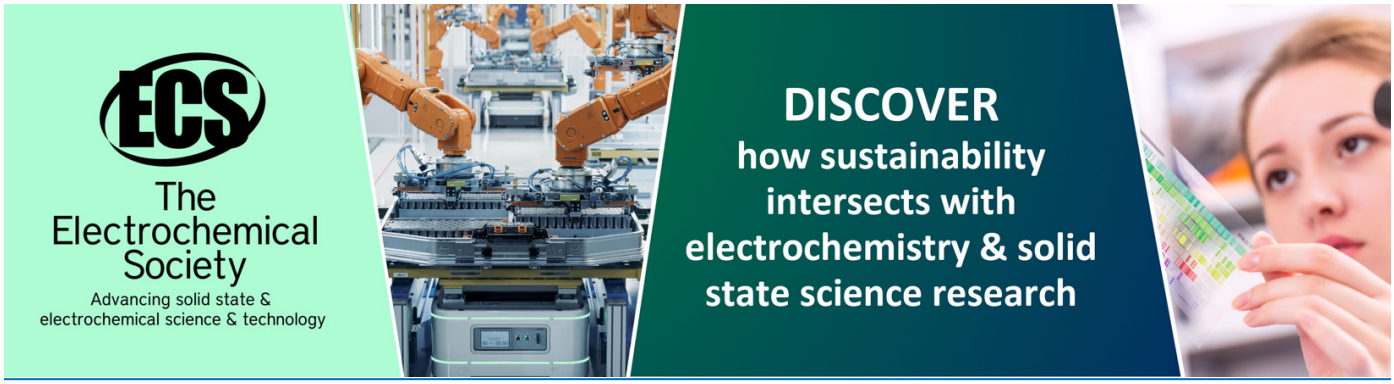
Method of identification of patent trends based on descriptions of technical functions

To cite this article: D M Korobkin *et al* 2018 *J. Phys.: Conf. Ser.* **1015** 032065

View the article online for updates and enhancements.

You may also like

- On the Question of Reducing Technogenic Risks of Accidents in the Technical Systems of Buildings and Structures
A S Glukhanov and E A Biryuzova
- Application of fast singular spectrum decomposition method based on order statistic filter in rolling bearing fault diagnosis
Yonggang Xu, Jinxin Cao, Jiyuan Zhao et al.
- Mathematical modelling of risk and safe regions of technical systems and surviving trajectories
A N Rogalev, A A Rogalev and N A Feodorova



This content was downloaded from IP address 18.224.214.215 on 30/04/2024 at 04:29

IOP Conf. Series: Journal of Physics: Conf. Series **1015** (2018) 032065 doi:10.1088/1742-6596/1015/3/032065

Method of identification of patent trends based on descriptions of technical functions

D M Korobkin, S A Fomenkov, A B Golovanchikov

Volograd State Technical University, 28, Lenina Ave., Volgograd, 400005, Russia

E-mail: dkorobkin80@mail.ru

Abstract. The use of the global patent space to determine the scientific and technological priorities for the technical systems development (identifying patent trends) allows one to forecast the direction of the technical systems development and, accordingly, select patents of priority technical subjects as a source for updating the technical functions database and physical effects database. The authors propose an original method that uses as trend terms not individual unigrams or n-gram (usually for existing methods and systems), but structured descriptions of technical functions in the form “Subject-Action-Object” (SAO), which in the authors' opinion are the basis of the invention.

1. Introduction

Due to a modern quick change of the global market and a widespread occurrence of innovations, the life cycle of a new technical solution can be very short. In this situation, it is actual to forecast the direction of technical systems development (trend) based on monitoring technology changes in the global patent space (patent trend analysis). The scientists are addressing the problem of identifying the patent trends: R. Frietsch [1], H. Bronwyn [2], B. Yoon [3], Y. Park, A. Kontostathis [4], etc. Based on the analysis of patent trends, it is useful to select the priority technical directions as a source of information for the actualization of physical effects and technical functions databases.

At present, the rejection of using the International Patent Classification (IPC) in the process of patent trends determination is actual. As the basis of the developed method, it is proposed to use descriptions of technical functions in the SAO (“Subject”, “Action”, “Object”) form to identify both interclass patent trends and trends within IPC classes.

SAO is a semantic structure that is used to represent the technical functions extracted from textual data, including the patent texts. “Object” and “Subject” are the words or phrases that are related to the text semantics. “Actions” are verbs, which represent the operations that connect the objects and subjects. The development of natural language processing technologies allows SAO structures to contain complete semantic information about the technical functions described in the text.

Currently, the problem of the patent databases analysis for the extraction of technical functions is actively studied by scientists H. Park [5], D.U. Yufeng [6], S. Choi [7], K. Kim, J.Y. Lee, J.Hu, J. Yoon, J. Guo, S. Fang, and others. They apply the technical functions at the initial stages of new technical systems design and use the representation of technical functions in the SAO form (Subject-Action-Object).

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. Method for determining scientific and technological priorities of technical systems development

The authors propose an original method that uses as trend terms not individual unigrams or n-gram (usually for existing methods and systems), but structured descriptions of technical functions in the form “Subject-Action-Object” (SAO) [8,9], which in the authors' opinion are the basis of the patent.

The method of patent trend detection consists of the following stages:

1) identification of “key” SAOs for each time interval;
2) determination of trend SAOs;
3) search of patents that contain the trend SAOs;
4) cluster formation for the patents that contain trend information.

One of the main tasks of the developed method of patent trends searching is to present the trends in the form of informative, user-friendly terms that reflect the core of trends. “Key” SAOs are very suitable for such terms.

3. Extraction of the SAOs from the patent texts

To extract the technical functions in SAO presentation from patent texts, one used the previously developed procedures [10] for segmentation of complex sentences of patent texts, morphological and semantic analysis with the construction of dependency trees and building the deep-syntactic structures based on the Meaning-Text Theory for reduced Stanford dependencies.

With the correct work of the semantic analyzer, the root node of the deep-syntactic structure must be a verb (Figure 1) (it is extracted as an “Action”), then its child nodes are extracted with actant relations I (“Subject”) and II (“Object”), and then for each node “Subject”, “Object” and “Action” the child nodes are extracted with attributive relation according to the Meaning-Text Theory. The result of extracting the SAO from the sentence “The super-capacitor electrode further comprising a silane coupling agent” is:

- action: comprises (root node - “comprises”, attributive relation – “further”);
- subject: super-capacitor electrode (I actant relation - “electrode”, attributive relation – “super-capacitor”);
- object: silane coupling agent (II actant relation - “agent”, attributive relation – “silane”, “coupling”).

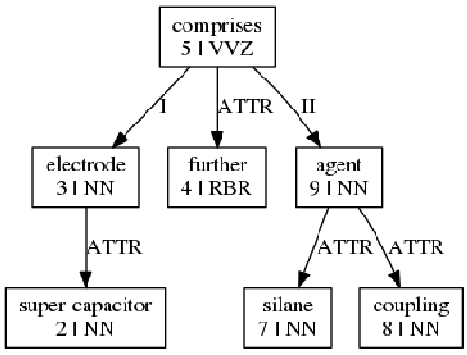


Figure 1. The example of the technical function

To increase the information content of the “Subject-Action-Object” structures, several SAOs are combined into the common structure according to the developed algorithm of grouping (comparison) SAOs [11].

4. Identification of trend SAOs

To search for the patent trends, a method of searching for trend SAOs has been developed. The trend SAO implies a textual description of a technical function for which the frequency of its occurrence in patent texts increases with a time. However, there are random non-informative SAOs, which are characterized by the miss of their occurrence in the previous time period and the beginning of their use

in the current time period. In order to filter out such SAOs, let us use the hypothesis that in a short period of time the frequency of use of SAO can not hardly change (decrease).

$$SAOvec_i = \{(SAO_1, TF_1), (SAO_2, TF_2), ...(SAO_k, TF_k)\} \quad (1)$$

$$Tr(SAOvec_{i+1}, SAOvec_i) = \{(SAO_1, Ch(TF_1^{i+1}, TF_1^i)), ...(SAO_k, Ch(TF_k^{i+1}, TF_k^i))\}$$

$$Ch(TF_j^{i+1}, TF_j^i) = \frac{TF_j^{i+1} - TF_j^i}{TF_j^i + Kr},\ Ch(TF_j^{i+1}, TF_j^i) \geq L_T$$

where $SAOvec_i$ is the vector of SAOs occurrences in the patents that are granted for the time interval i;

TF_j - frequency of the j-th SAO in the patent texts;

$Tr(SAOvec_{i+1}, SAOvec_i)$ – the trend vector (trend) of occurrence of all SAOs between two neighboring time intervals;

$Ch(TF_j^{i+1}, TF_j^i)$ - the function that determines the trend (change of the occurrence of the j-th SAO) between two neighboring time interval;

Kr - the coefficient of decreasing the importance of the "rare" SAO;

L_T - the limit of the minimum allowable trend change for the time interval.

Coefficient Kr is meant for excluding from the list of the trend SAOs, which rarely occur in the patent texts. With an insignificant change of the absolute frequency of occurrence of such SAOs, there is a significant increase of their relative change Ch between two neighboring time intervals. The experiments carried out showed that the optimum value is Kr = 150.

The L_T parameter is designed to regulate the number of finding trend SAOs. Its value is inversely proportional to the number of SAOs that can be considered as trend and directly proportional to their average significance as trends.

The set of trend patents for a certain period of time is defined as follows:

$$T_{SAO} = \{SAO_1, SAO_2, ...SAO_{N_{SAO}}\} \quad (2)$$

$$SAO_j \in T_{SAO} : \forall i\ Ch(TF_j^{i+1}, TF_j^i) \geq L_T$$

where T_{SAO} – the set of the trend SAO;

N_{SAO} - the number of the trend SAO.

5. Search of patents that contain trend SAOs

The next step is to search for patents containing trend SAOs:

$$T_{PAT} = \{PAT_1, PAT_2, ...PAT_{N_{PAT}}\} \quad (3)$$

where T_{PAT} – set of trend patents

N_{PAT} - the number of trend patents.

The authors have implemented two approaches to the search for trend patents:

- The search for the trend SAO is carried out only in the patent claims. The main idea of this approach is that all SAOs found in the claims are more significant than the SAOs contained in the other part of the patent text.

$$PAT_i \in T_{PAT} : \exists SAO_j\ SAO_j \in Claim_i,\ SAO_j \in T_{SAO} \quad (4)$$

  where $Claim_i$ – patent claim PAT_i.
- The search for the trend SAO is carried out in all parts of the patent. The main idea of this approach is that trends can also occur in patents, the main description (claim) of which may

not be relevant to the subject of the trend. Thus, the description of the trend can be contained in any part of the patent text (for example in the field "Desc" – full patent description).

$$PAT_i \in T_{PAT} : \exists SAO_j \;\; SAO_j \in \{Claim_i, Abst_i, Desc_i\}, \;\; SAO_j \in T_{SAO} \qquad (5)$$

where $Abst_i$ – the field "Abstract" of the PAT_i patent;
$Desc_i$ – the field "Description" of the PAT_i patent.

6. Cluster formation for patents containing trend information

To realize the grouping of trend patents into clusters, it is necessary to implement the following stages of analysis:

- building a term-document matrix;
- clustering based on the Latent Dirichlet Allocation (LDA) [12,13] model (Figure 2) and using the constructed model to obtain the distribution of vectors by the clusters (unnamed topics).

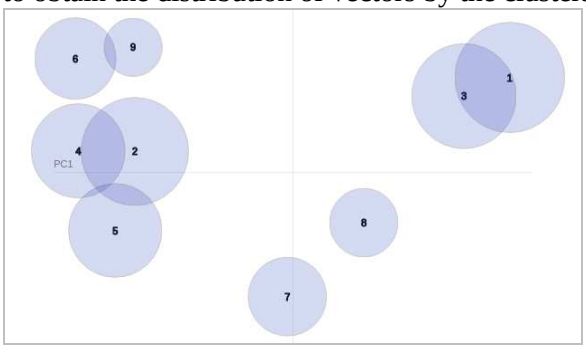


Figure 2. Distribution of vectors from unnamed topics

A term-document matrix is a table, whose rows are documents and columns are terms. This matrix describes the frequency of occurrence of each term in all documents. Terms are trend SAOs, the documents –trend patents.

To construct the LDA model, a TF-based matrix should be generated. TF (term frequency) is the frequency of occurrence of the trend SAO in the texts of trend patents. To increase the information content of the trend SAOs, several similar SAOs are joined into one unique SAO.

First, it is necessary to obtain a thesaurus of all trend SAOs of the patent database voc = $\{t_1,t_2,t_3,..,t_n\}$,, where t_i is the next unique trend SAO. Then each i-th row of the term-document matrix will be the resulting thesaurus, and each column will represent the number of occurrences of the unique trend SAO in the i-th patent.

The framework Spark and its MLlib library for machine learning allow obtaining the thesaurus of all the terms of a set of documents and building a term-document matrix.

Based on the LDA model, it is possible to obtain synchronous clustering of trend patents and trend SAOs on the same set of clusters (topics). As a result, a "soft" cauterization is constructed - a trend patent may refer to several topics with varying degrees of ownership:

$$SAO_Cl_i = \{SAO_t\}, \; Pat_Cl_i = \{PAT_k\} \qquad (6)$$

where i = 1,..,cl – index of the cluster;
cl – the predetermined number of clusters;
SAO_Cl_i - i cluster of the trend SAOs;
$SAO_t \in T_{SAO}$ — t-th trend SAO in cluster SAO_Cl_i;
Pat_Cl_i – i-th cluster of the trend patents;
$PAT_k \in T_{PAT}$ — k-th trend patent in cluster Pat_Cl_i.

Further, let us rank the obtained topics by the number of trend patents and correlate trend patents and their trend SAO.

7. Conclusions

The developed method of identifying technology changes (detection of patent trends) is focused on using as trend terms not individual unigrams or n-grams (usually for existing methods and systems), but structured descriptions of technical functions in the SAO form (Subject, Action, Object).

A trend SAO is a textual description of a technical function, for which the frequency of its occurrence in patent texts increases with time. For the filtering of random non-informative SAOs, the hypothesis is made that in a short time period, the frequency of SAO use cannot change sharply (decrease).

Two approaches to the search for trend patents have been implemented: searching only in the patent claim and searching in full text of patent (trends can occur in patents, the claims of which may not match with the trending topic).

For the grouping of trend patents into clusters, the following stages of analysis are realized: construction of the term-document matrix and clustering based on the LDA model. The term-document matrix contains the frequencies of occurrence of all trend SAOs in all trend patents. Based on the LDA model, it is possible to obtain a synchronous clustering of trend patents and trend SAOs on the same set of clusters. The authors rank the obtained topics by the number of trend patents and correlate with trend patents and their trend SAO

8. Acknowledgments

The reported study was funded by the Russian Foundation for Basic Research (research project No. 18-07-01086 A, research project No. 16-07-00534 A).

References

[1] Frietsch R, Neuhäusler P, Rothengatter O 2012 Patent applications - structures, trends and recent developments *Studien zum deutschen Innovationssystem* **8** 68
[2] Bronwyn H H, Jaffe A B, Trajtenberg M 2001 *NBER Working Paper* **8498** 60-77
[3] Yoon B, Park Y 2004 *The Journal of High Technology Management Research* **1(15)** 37–50
[4] Kontostathis A, Galitsky L, et al 2004 *Survey of Text Mining* 185-224
[5] Park H, Yoon J, Kim K 2012 Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics* **90** 515
[6] Yufeng D U, Duo J I, Lixue J, et al. 2016 Patent Similarity Measure Based on SAO Structure *Chinese Sentence and Clause for Text Information Processing* **30(1)** 30-36
[7] Choi S, Park H, Kang D, Lee J Y, Kim K 2012 An SAO-based text mining approach to building a technology tree for technology planning *Expert Systems with Applications* **39** 11443–11455
[8] Korobkin D M, Fomenkov S A, Kolesnikov S G, Golovanchikov A B 2016 Technical Function Discovery in Patent Databases for Generating Innovative Solutions *Multi Conference on Computer Science and Information Systems 2016* ed P Kommers 241-245.
[9] Finogeev A G, Parygin D S, Finogeev A A 2017 The convergence computing model for big sensor data mining and knowledge discovery *Human-centric Computing and Information Sciences* **7 (11)**
[10] Korobkin D M, Fomenkov S A, Kravets A G, Kolesnikov S G 2017 Methods of Statistical and Semantic Patent Analysis CIT&DS 2017 ed A Kravets, M Shcherbakov, et al (Germany, Springer) 48-61
[11] Korobkin D M, Fomenkov S A, Kravets A G, Kolesnikov S G 2017 Prior art candidate search on base of statistical and semantic patent analysis *Multi Conference on Computer Science and Information Systems 2017* ed Y Xiao, A P Abraham 231-238

[12] Korobkin D M, Fomenkov S A, Kravets A G, Golovanchikov A B 2016 Patent data analysis system for information extraction tasks *13th International Conference on Applied Computing (AC) 2016* ed H Weghorn 215-219
[13] Sadovnikova N, Parygin D, Kalinkina M, Sanzhapov D, Trieu N N 2017 Models and methods for the urban transit system research *CIT&DS 2015* 488–499