Method of detection of technical functions performed by physical effects

D M Korobkin, S A Fomenkov

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

E-mail: dkorobkin80@mail.ru

Abstract. For the synthesis of the physical operation principle of a technical system, the authors used the physical effects realizing the technical functions that compose the selected functional structure of the designed technical system. Authors developed the method for extraction of the descriptions of physical effects from the patents of USPTO and RosPatent databases, and the method for extracting of technical functions from Natural Language documents including patent texts. The method of automated construction of a matrix of physical functions, performed by physical effects, is based on the detection of latent dependencies in the consolidated matrix "Physical Effects – Technical Functions". Developed software has been tested for tasks of extraction of the physical effects and technical functions from patent documents.

1. Introduction

Analysis of the computer-aided innovation systems (CAI) [1], such as Goldfire Innovator, TechOptimizer, Innovation Workbench, Idea Generator, Pro / Innovator, etc., shows that these CAI systems do not solve the fundamental problem of updating the information component of the new technical systems generation. One should use the world patent database (more than 20 million patents) and open databases of scientific and technical information as a global knowledge base most correctly. The engineer/inventor realizes the required functions of the designed technical system on the basis of heuristic morphological synthesis or the physical operation principle (POP) [2], which is a sequence (network) of physical effects (PhE) [3]. Therefore, it is required to extract the following data from the global information space: physical and technical effects, morphological features and their alternatives (elements of functional structure, technical realizations of objects), i.e. information necessary to solve the problems of information support for the synthesis of new technical solutions.

2. Method of automated filling of the database of technical functions performed by physical effects on the basis of patent database analysis

The task of determining the most effective physical operation principle of a technical system with a selected functional structure that was synthesized on the basis of a morphological matrix is solved by 2 methods: 1) the method for automating the procedures for synthesizing the physical operation principle based on the database of technical functions performed by physical effects, and 2) the developed method for verification of the practical realizability [4] of the synthesized POP based on the criteria parameters of the physical effects that are a part of a POP structure.

The procedure for synthesizing the physical operation principle structures uses the physical effects realizing the technical functions that compose the selected functional structure of the designed technical system.

It is required to develop: a) the method for extraction of descriptions of the physical effects from patent of USPTO and RosPatent with textual analytics; b) the method for extracting of technical functions from Natural Language (NL) documents on the basis of a developed model for representing the technical function; c) the method of automated construction of a matrix of physical functions performed by physical effects on the basis of latent dependencies detection.

Authors analyzed the software for fact extraction from the unstructured text documents: ABBYY InfoExtractor, Compreno, RCO Fact Extractor, SAS Text Analytics, RapidMiner Text Mining. These systems can not extract descriptions of physical effects from an NL text. The authors compared the statistical Natural Language Processing (NLP) [5], deep learning NLP [6], and rule-based NLP [7] tools: Stanford NLP, MaltParser [8], Link Grammar Parser [9], AGFL, Tomita Parser.

3. The method for extraction of descriptions of the physical effects from patent of USPTO and RosPatent

For the task of extraction of descriptions of the physical effects from patent texts, the previously developed procedures are used [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 [11] for reduced Stanford dependencies.

An example of deep-syntactic structures from the sentence of patent US20130307109A1 (“Solid-state image sensor and imaging system”) “When light enters the semiconductor junction of such an element, the incident light causes photoelectric conversion, generating charges” is shown in Figure 1.

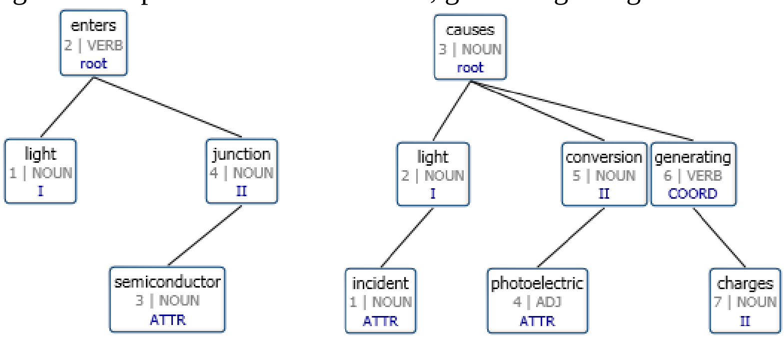


Figure 1. Deep-syntactic structures

According to the PhE model [12], developed at the CAD Department of Volgograd State Technical University, input cause-action produces an output effect-action on the environment or object. For extraction of the physical effect from text, it is necessary to find the predicates (verbs) such as “change”, “increase”, “decrease”, “dependence”, “change”, “generate”, “act”, “cause”, etc. that show some “action” with arguments. For these predicates, authors correlated the arguments: actant relations “I” (something that acts), “II” (something that has an effect where an action is realized) with elements of the PhE description: input cause-action (A), output effect-action (C) and object (B).

The authors developed a new PhE model based on the developed ontology “Physical Effect” [3]:

$$M_{PhE}=<C, D, B, R_C, R_B>, \quad (1)$$

where C – set of predicates (verbs), c_i C;

D - set of actant relations of arguments $\{I, II\}$, D_i D – set of actant relations of predicate c_i, d_j D_i;

B – set of elements of the PhE description (A, B, C), B_k B;

$c_i\ C\ \ d_j\ D_i\ d_j \xrightarrow{def} B_k$,

R_C – relation $C\ D$, (c_i,d_j) R_C determines an element(s) of the PhE description with actant relation d_j of predicate c_i;

R_B – relation R_C B, $((c_i,d_j),B_k)R_B$ determines a set of ontology concepts that correlated with an element of PhE description b_k , b_k B_k.

The authors have developed a set of patterns for extraction of the PhE descriptions from NL patent text (104 patterns in Russian, 36 patterns in English) on the base of model M_{PhE},

The authors also improved the method of building a semantic network of the PhE description that was extracted from the NL text:

a) node $V_i = (T_i, B_i)$, where T_i – argument (Natural Language terms) of predicate c_j, $T_i\, R_B$; B_i – PhE element(s) which is represented as T_i in the NL text, $B_i \in R_C$.

b) arc C_j – the predicate that was extracted from the NL text.

Figure 2 shows the example of the semantic network from the sentence: «When light enters the semiconductor junction of such an element, the incident light causes photoelectric conversion, generating charges».

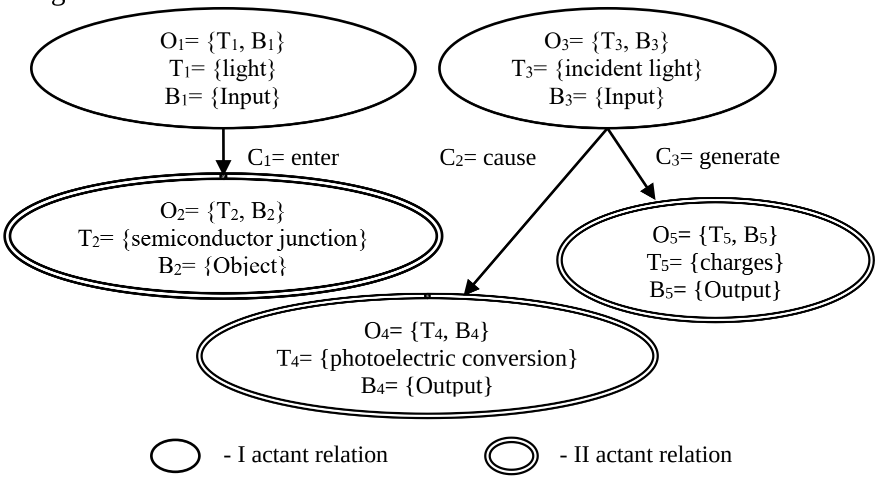


Figure 2. The constructed semantic network

The authors developed the algorithm of PhE extraction from NL patent texts on the basis of model M_{PhE} and the method of building the semantic networks that consist of the next steps:

- The semantic analysis of patent text with procedures of segmentation of complex sentences, morphological analysis, construction of the dependency trees, building the deep-syntactic structures based on the Meaning-Text Theory for reduced Stanford dependencies.
- Search in the NL patent text for the predicates (verbs) from a thesaurus of database “Physical Effects”.

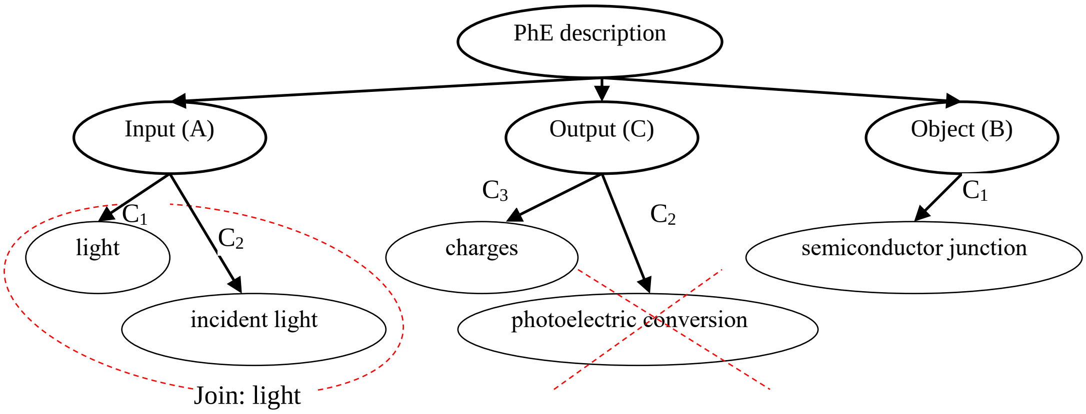


Figure 3. Construction of PhE description

- Detecting the important arguments (with actant relations I, II) of found predicates from received deep-syntactic structures. Arguments (NL terms) must be found in ontology thesaurus “Physical Effect Input”, “Physical Effect Output”, “Physical Effect Objects”.
- Building a semantic network.
- Construction of PhE description (Figure 3) using the procedures: a) concept joining, and b) exception of the concepts that are not found in their corresponding concept classes “Physical Effect Input”, “Physical Effect Output”, “Physical Effect Objects”.

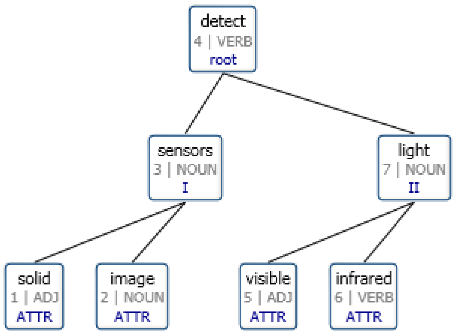
For this example (Figure 3) the authors obtained a description of PhE “Photoelectric effect” which has “light” as input cause-action, “charge” – as the object and a “semiconductor” – as output effect-action.

4. The method for extracting of technical functions from patent texts

Currently, the problem of the patent databases analysis for the extraction of technical functions in the SAO form (Subject-Action-Object) is actively studied by scientists H. Park [13], D.U. Yufeng [14], S. Choi, K. Kim, JY Lee, J. Hu, J. Yoon, J. Guo, S. Fang, and others.

For extraction of the technical functions in the SAO form from NL, patent texts are also used by the previously developed procedures [15] for building the deep-syntactic structures (Meaning-Text Theory). The root node of the deep-syntactic structure is a verb (Figure 1) (“Action”), its child nodes are extracted with actant relations I (“Subject”) and II (“Object”), and for each node “Subject”, “Object” and “Action”, the child nodes are extracted with attributive relation.

The description of the PhE “Photoelectric effect” was found in the patent US20130307109A1 “Solid-state image sensor and imaging system”. Then in this patent, it is necessary to search for descriptions of technical functions. An example of SAO extracted from the sentence “Most solid image sensors are used to detect visible or infrared light” is shown in Figure 4:



- action: “detect” (root node - “detect”);
- subject: “solid image sensors” (I actant relation - “sensors”, attributive relation – “solid”, “image”);
- object: “visible or infrared light” (II actant relation - “light”, attributive relation – “visible”, “infrared”).

Figure 4. An example of the extracted technical function

The result of SAO extracted from the sentence of patent US20130307109A1: “The element isolation may achieve by impurity diffusion using a PD junction”:

- action: “perform” (root node - “achieve”);
- subject: “impurity diffusion” (I actant relation - “diffusion”, attributive relation – “impurity”);
- object: “element isolation” (II actant relation - “isolation”, attributive relation – “element”).

Structuring of the set of SAO components values was made on the basis of a set of terms, most often encountered in the technical function descriptions (for example, in the International Patent Classification and in the field “Practical application” of the PhE database (1200 descriptions). Using this structuring of information, the class of the verb “achieve” is defined as “perform”.

The obtained components of the technical function representation according to the SAO model recognized in the patent sentences are listed in Table 1.

Table 1. Technical functions from patent US20130307109A1

No	Action	Subject	Object	Action class
1	detect	solid image sensors	visible or infrared light	DETECT
2	perform	impurity diffusion	element isolation	PERFORM

The algorithm of grouping (comparison) SAOs [16] is used to combine several "Subject-Action-Object" structures into one common structure.

5. Results

The basic criterions used for evaluation of extraction of PhE and technical functions (TechFun): a) recall - ratio of the number of retrieved relevant PhE / TechFun to the total number of relevant PhE / TechFun, b) precision - ratio of the number of retrieved relevant PhE / TechFun to the total number of retrieved irrelevant and relevant PhE / TechFun:

$$\text{Recall} = \frac{R_{rel}}{R_{rel} + NR_{rel}}, \quad (2)$$

$$\text{Precision} = \frac{R_{rel}}{R_{rel} + R_{nrel}}, \quad (3)$$

where R_{rel} – the number of retrieved relevant PhE / TechFun;
NR_{rel} – the number of not retrieved relevant PhE / TechFun;
R_{nrel} – the number of retrieved irrelevant PhE / TechFun.

Table 2. The evaluation criterions of technical functions extraction

No exp	R_{rel}	NR_{rel}	R_{nrel}	Recall	Precision
1	82	23	14	0.78	0.85
2	133	42	31	0.76	0.81
3	301	86	96	0.78	0.76

Table 3. The evaluation criterions of physical effects extraction

No exp	R_{rel}	NR_{rel}	R_{nrel}	Recall	Precision
1	12	6	8	0.67	0.6
2	18	17	20	0.51	0.47
3	25	14	23	0.64	0.52

6. Conclusion

The authors developed the method for extraction of descriptions of the physical effects from the patents of USPTO and RosPatent databases, and the method for extracting of technical functions from Natural Language documents including patent texts.

The method of automated construction of a matrix of physical functions performed by physical effects is based on the detection of latent dependencies in the consolidated matrix "Physical Effects – Technical Functions". The consolidated matrix is formed from two term-document matrices: a) the first matrix describes the frequency of occurrence (TF/IDF) of each term (technical functions in SAO form) in all patent documents, b) the second matrix describes the frequency of occurrence of physical effects in all patens.

7. 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] Orloff M A 2006 Inventive Thinking through TRIZ: A Practical Guide *Springer Science & Business Media* 352
[2] Vayngolts I I, Korobkin D M, Fomenkov S A, Golovanchikov A B 2017 Synthesis of the Physical Operation Principles of Technical System *CIT&DS 2017* ed A Kravets, M Shcherbakov **754** 575-588
[3] Korobkin D M, Fomenkov S A, Kamaev V A, Fomenkova M A 2016 Multi-agent model of ontology-based extraction of physical effects descriptions from natural language text *J. of Physics: Conference Series* 498-501
[4] Korobkin D M, Kolesnikov S G, Fomenkov S A, Golovanchikov A B 2015 Evaluating an operation principle practicability of new technical system *IISA 2015* DOI: 10.1109/IISA.2015.7388068
[5] Manning C D, Raghavan P, Schütze H 2008 Introduction to Information Retrieval *Cambridge University Press*
[6] Young T, Hazarika D, Poria S, Cambria E 2017 Recent Trends in Deep Learning Based Natural Language Processing *arXiv:1708.02709*
[7] Pilan I, Volodina E, Johansson R 2014 Rule-based and machine learning approaches for second language sentence-level readability *Building Educational Applications Using NLP*
[8] Nivre J, Hall J, Nilsson J 2006 MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. *LREC 2006* 2216-2219.
[9] Introduction to the Link Grammar Parser 2016
[10] Fomenkova M A, Korobkin D M, Fomenkov S A 2017 Extraction of Physical Effects Based on the Semantic Analysis of the Patent Texts *CIT&DS 2017* ed A Kravets, M Shcherbakov, et al **754** 73-87
[11] Mel'čuk I A 1988 Dependency Syntax Theory and Practice SUNY Publ NY
[12] Korobkin D M, Fomenkov S A, Kolesnikov S G, Lobeyko V I, Golovanchikov A B 2015 Modification of Physical Effect Model for the Synthesis of the Physical Operation Principles of Technical System *CIT&DS 2015* ed A Kravets, M Shcherbakov, et al **535** 368-378
[13] Park H, Yoon J, Kim K 2012 Identifying patent infringement using SAO based semantic technological similarities *Scientometrics* 90 515
[14] 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
[15] 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
[16] 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