Multi-agent model of ontology-based extraction of physical effects descriptions from natural language text

Dmitriy M. Korobkin, Sergey A. Fomenkov, Valeriy A. Kamaev, Marina A. Fomenkova
CAD Department
Volgograd State Technical University
Volgograd, Russia
saf@vstu.ru

***Abstract*— Authors developed a representation model of physical effects knowledge based on an ontological approach. Authors proposed a model of multi-agent system for updating a physical effects database, including a model of agents, the model of interaction between agents, the role of agents in interaction. The developed system's efficiency was tested on a special documents corpus. Created multi-agent system can significantly reduce search time and increase efficiency in comparison with software "IOFFE".**

Keywords*— Domain ontology, fact extraction, multi-agent system, text mining, structured physical knowledge, physical effect.

I. Introduction

Physical Effect (PE) [1] is an objective relation between two or more physical phenomenon, each of which is described appropriate physical quantity. As any physical phenomenon realize in material medium so a representation diagram of PE in the form of "black box" is visual and useful: A→B→C where A – input cause-action, B - object, C – output effect-action.

Example: Physical Effect No 37. "Ohm's Law"

- Input cause-action: Electric field. Weak. Electric field strength.
- Object: Conductor, semiconductor.
- Output effect-action: Electric current. DC, AC. Electron, ionic, mixed. Current density.

The analysis of the studies [2-8] related to the processing of physical knowledge in the form of Physical Effects for the synthesis and choice of physical operating principle of new technical systems has shown that a problem of PE database update [1] (the basic procedures are search and extract PE descriptions) is very actual. The method of modifying the physical effects database created by employees of CAD department (Volgograd State Technical University [9, 10]) and realized as software "IOFFE" [11] also has serious drawbacks:

- thesaurus as the knowledge base
- significant time of new physical effects searching

Authors analyzed text mining' systems of (fact extraction from the unstructured text arrays): AeroAeroText, TextAnalyst, WordStat, Attensity, Carrot2, GATE, OpenNLP, Natural Language Toolkit, RapidMiner. None of the systems are couldn't extract physical effects from Natural Language text. We reviewed the principles of English language' ontologies: FrameNet, VerbNet, PropBank и NomBank. We made a comparison of systems built on the base of these ontologies and provided Semantic Role Labeling: EP4IR parser, Link Grammar, SENNA, SwiRL. We chose MaltParser [12] as basic system.

That's why the actual task is an automation of the process of extracting descriptions of physical effects from the Natural Language text at the expense of new, more productive approaches.

II. Domain ontology "Physical Effect"

We define the components of "Physical Effect" formal domain description for automation the process of extracting physical effect descriptions [9-11] from Natural Language (NL) texts.

- Domain ontology including concepts and their relations in "Physical Effect" domain;
- Thesaurus including terms that represent ontology concepts and relationships in NL text.

Describing the domain concepts is used a taxonomy and relations "IS-A" and "HAS-PART". We can make an example to show a relation "HAS-PART": a presence in the text a physical quantity "Magnetic Induction" definitely shows concept "Magnetic Field". The domain concepts thesaurus is formed from terms and their synonyms, for example, concept "Solid body" and synonym "Solid material".

Domain ontology was chosen as representation model of knowledge about physical effects. The domain ontology based on knowledge from physical effects database (developed by employees of CAD department of Volgograd State Technical University) and also based on model of the physical effect description [1].

The physical effect in the developed ontology is presented according to the Physical Effect model: {Input, Output,

Object}. Relations for connect the PE model elements together were introduced in the ontology:

- “CauseAction” used to determine the concept class “Physical Effect Output”.
- “EffectAction” - to determine the concept class “PE Input”.
- “ActionOject” - to determine the concept class “PE Object”.

Also the relations characterizing fundamental concepts of the domain ontology - a matter and properties of elements of a matter were entered:

- “FieldProperty” used to connect concept “Field” and its possible properties;
- “SubstanceProperty”. Respectively - for “Substance”;
- “Unit”. Allows to characterize physical quantity.

Other relations reflect the class hierarchy of the domain ontology and have “SuperclassOf” designation.

Ontology was designed using Protégé tool [13] and is presented in Fig. 1. The language used for description ontology - OWL [14]. Developed ontology structure allows to use queries to the knowledge base, the purpose of which is to identify elements of physical effect descriptions presented in the form of ontology concepts. We define a common physical effect’ description in accordance with the selected language SPARQL [15] and model:

```
SELECT DISTINCT? result WHERE {
? resultSuperClass ns:CauseAction ns: PhysicalEffectInput
? resultSuperClass ns:EffectAction ns: PhysicalEffectOutput
? resultSuperClass ns:ActionObject ns: PhysicalEffectObject
? result rdfs:subClassOf? resultSuperClass. }
```

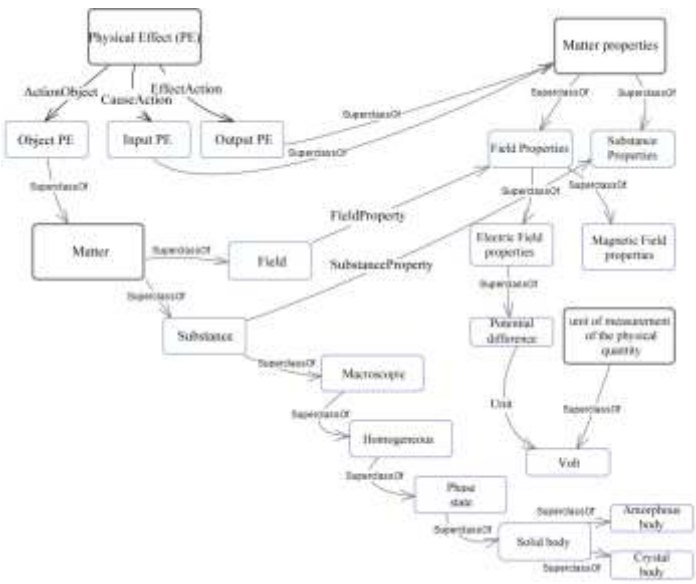

Fig. 1. Fragment of domain ontology

According to a model of Physical Effect [1] an “input cause-action” on the Physical Effect’ object produces “output effect-action” on the environment or this object. Therefore it is necessary to extract from NL text the predicates that show some “action” over arguments (“objects”) with defined semantic roles. We have identified a single class of all the predicates that are typical for a physical effect’ description in the NL text, such as “change (increase, decrease”), “dependence (depend, be directly proportional, be inversely proportional)”, “influence (relate, cause)”, etc.

For extraction of Physical Effect description it is necessary to define the three model components: input cause-action, object of action and output effect-action. We will present a specific physical effect’ description as a subclass of the common PE descriptions. Example:

```
Class ({ (electric field lines within a dielectric material
are weakened due to polarization) })
partial {
restriction (CauseAction allValuesFrom (Polarization)),
restriction (EffectAction allValuesFrom (ElectricField)),
restriction (ActionObject allValuesFrom (Dielectric))
})
```

Ontology physical schema is shown in Fig. 2. In ontology added physical effect description: “Temperature dependence of viscosity of liquid metal”.

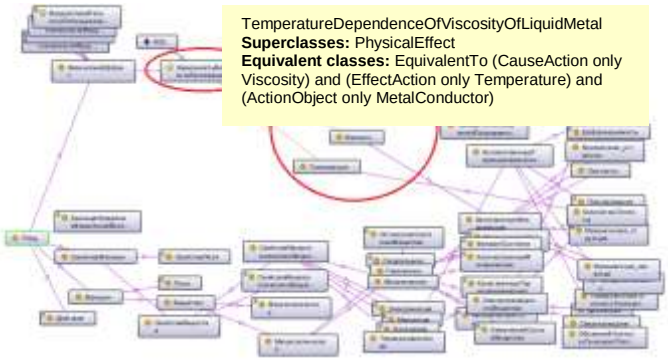

Fig. 2. Fragment of ontology physical schema

Query is generated from the results of the semantic analysis of text sentence with the expected physical effect description. We introduce the rules set that associate MTT relations [16] and queries to knowledge base, for example, sentence “electric field lines within a dielectric material are weakened due to polarization” corresponds to the following queries (Table 1).

TABLE I. MTT RELATIONS AND CORRESPONDING QUERY

MTT relations	Transformed relations	Query
SUBJECT (Weakened, Polarization)	Input (Weakened, Polarization)	? resultSuperClass ns: CauseAction ns: Polarization
ATTR (Lines, Material) + ATTR (Material, Dielectric)	Object (Electric field lines, Dielectric material)	? resultSuperClass ns: ActionOject ns: Dielectric
OBJECT (Weakened, Lines) + ATTR (Lines, Field) + ATTR (Field, Electric)	Output (Weakened, Electric field lines)	? resultSuperClass ns: EffectAction ns: ElectricField

The presence of correlations between MTT relations and queries to the knowledge base for the physical effects included in the ontology allows to extract the links between semantic roles and PE elements’ description that presented in the form of

PE domain' concepts for each predicate. Example of the physical effect description:

Input data:

The effect of temperature on the volume of gaseous bodies

Query to knowledge base:
SELECT DISTINCT? result WHERE
{? resultSuperClass ns: CauseAction ns: Temperature
? resultSuperClass ns: ActionObject ns: Gaseous body
? resultSuperClass ns: EffectAction ns: Volume }

Fig. 3. Semantic analysis

Result: Physical Effect No 29 "Thermal expansion of liquid and gaseous bodies".

These links have been presented in the form of patterns [1] of representations of physical effects' descriptions in NL text (104 patterns in Russian, 36 patterns in English).

C = Produce, I $\xrightarrow{def}$ PhysicalEffectInput, II $\xrightarrow{def}$ (PhysicalEffectOutput, PhysicalEffectObject).

C = Cause, I $\xrightarrow{def}$ PhysicalEffectInput, II $\xrightarrow{def}$ PhysicalEffectOutput.

III. Multi-agent system architecture

The approach based on the application of multi-agent systems MAS [17] used to solve the problem of searching and extracting PE description from NL text into Internet.

The agents operate in an environment in which agents includes in addition the following components: PE database; External sources of information (file system, the Internet); Users. After analyzing the multiple protocols of cooperation of agents authors settled on using a well-known model of contractual networks. In the FIPA [18] that protocol named Contract Net Protocol [19]. The ACL (Agent Communication Language) [20] language used to implement of interaction.

The developed MAS model allows:

- expand the functionality of the system through the introduction of new agents;
- using a single PE database in MAS;
- registration several agents of the same type in the MAS. For example, a search agents that can create a new instance of the agent by increasing the MAS load;
- change the composition of executing agents due to interaction on the auction basis.

The Visual C# chosen as the main programming language. As the database has been selected Microsoft SQL Server 2008. To access the data used T-SQL language. After the analysis was chosen AgentService platform, which allows realizing the MAS in the program agents' area. The messaging transport service (MTS) allowed develop mechanism "login" action agent, conservation agent states, working with streams. Create a set of classes that will simplify the agent's behavior: the creation of a connection request from other agents waiting for connection, communication agents, and the search agent.

System consists of several modules. MAS module manages the multi-agent system and provides methods for controlling the operation of the MAS operation control agents, display information about the MAS ongoing work. Query module converts the user requests to query the MAS and displays the query results. The PE module contains database administration tools for managing the PE database. In addition to these modules architecture includes PE database.

The strategy of searching physical knowledge onto Internet is based on two approaches:

- Operate with the initial array of hyperlinks from resources with physics content (for example, journals: The Success of Physical Science, Journal of Applied Physics, Physics of Solid etc.) defined by the system administrator.
- Use of search engine indexes (Google, Yandex etc).

The text analysis system realized in the form of the system with a hierarchical organization of search agents interaction (Fig. 4).

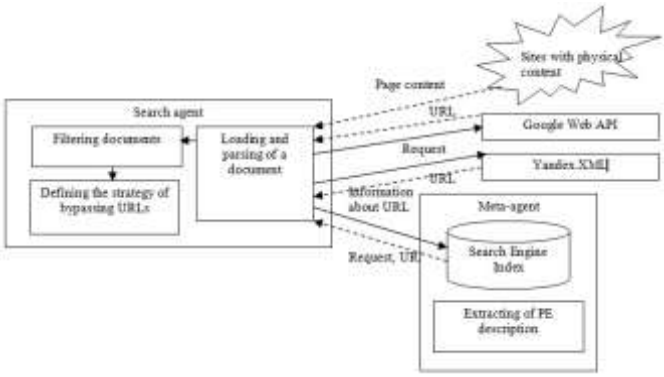

Fig. 4. Architecture of multi-agent system

Search agent extracts internal URLs (the same host with parsed document) and external URLs and passes them to Meta-agent. Meta-agent distributes the URLs between search agents on the base of the algorithm of bypassing the tree of extracting URLs [9]. A meta-agent works with a search engine index compiled on the basis of information transmitted by a search agent. The meta-agent passes initial URLs to a search agent and sends requests made on the basis of a modified PE description. The meta-agent performs the extraction of PE descriptions from initial source text. The search agent loads the documents using URLs passed by the Meta-agent or by Google Web API or Yandex.XML. The search agent performs html parsing, filtering documents and recursive bypass of URLs.

IV. Evaluating the effectiveness of developed system of of physical effects extraction

The system efficiency was tested on a special documents corpus which consists of 50 documents with 80 physical effects description (D^{rel} = 80 PE).

D^{rel}_{retr} – a number of relevant primary PE descriptions,

D^{nrel}_{retr} – a number of irrelevant primary PE descriptions,

D_{retr} – a number of constructed PE descriptions.

Average results of the program' test shown in Table 2. Comparison of efficiency indexes of the developed system and "IOFFE" system [11] are presented in Table 3.

TABLE II. RESULTS OF SYSTEM EFFICIENCY VERIFICATION

Efficiency index	Value
D^{rel}_{retr}	50
D^{nrel}_{retr}	64
D_{retr}	114
Precision	0,412
Recall	0,625

TABLE III. COMPARISON OF EFFICIENCY INDEXES

Efficiency index	IOFFE	Our software
Precision	0,306	0,412
Recall	0,518	0,625

V. CONCLUSION

The model of knowledge representation about physical effects based on an ontological approach is developed. Logical and physical representation of domain ontological model is designed. It allowed to get the following advantages:

- decision of synonymy problem by use objects classes. Synonyms are added in the ontology as objects of classes;
- possibility of knowledge extraction from the domain ontology. The presence of tools for automatic analysis and visualization;
- flexibility to ontology update and support.

A model of multi-agent system, including a model of agents, agent interaction model, and roles of agents in the interaction, is developed. The model allows to integrate MAS system through the development of open distributed system. System formed by several interacting agents to extend the functionality through the introduction of new agents and to implement a unified PE database.

The developed system efficiency was tested on a special documents corpus. Multi-agent system can significantly reduce search time and increase efficiency in comparison with software "IOFFE" developed by employees of CAD department of Volgograd State Technical University.

ACKNOWLEDGMENT

This research was partially supported by the Russian Fund of Basic Research (grant No. 15-07-09142 A, grant No. 16-07-00534 A).

REFERENCES

[1] Dvoryankin A.M., Polovinkin A.I., Sobolev A.N. Automating the search for operation principles of technical systems on the basis of a bank of physical phenomena. // Cybernetics. January–February, 1978, Volume 14, Issue 1, pp 79-86.

[2] Arel, E.: Goldfire Innovator. Volume II: Patents and Innovation Trend Analysis User Guide. Invention Machine Corporation, Boston, MA, pp. 263 (2004).

[3] Glazunov, V.N.: Technology of ideas: expert systems "Innovator" and "Edison", http://www.trizland.ru/trizba/pdf-articles/system_novator.pdf.

[4] Korobkin D.M., Fomenkov S.A., Kolesnikov S.G., Kamaev V.A. Synthesis of the Physical Principle of Operation of Engineering Systems in the Software Environment CPN Tools. //Research Journal of Applied Sciences, 2014, v.9(11), pp. 749-752.

[5] Zaripova, V., Petrova, I. Knowledge-Based Support for Innovative Design on Basis of Energy-Information Method of Circuits. In: Kravets, A., Shcherbakov, M., Kultsova, M., Iijima, T. (eds.) Knowledge-Based Software Engineering, Communications in Computer and Information Science, vol. 466, Springer International Publishing (2014).

[6] Zlotin, B., Zusman, A.: Directed Evolution: Philosophy, Theory and Practice. Ideation International, pp. 103 (2001).

[7] Fey, V., Rivin, E.: Innovation on Demand: New Product Development Using TRIZ. Cambridge University Press, pp. 242 (2005).

[8] Fomenkov S.A., Korobkin D.M., Kolesnikov S.G., Dvoryankin A.M., Kamaev V.A. Procedure of Integration of the Systems of Representation and Application of the Structured Physical Knowledge. //Research Journal of Applied Sciences, 2014, v.9(10), pp. 700-703.

[9] Korobkin, D.M., Fomenkov, S.A., Kolesnikov, S.G. Semantic network of physical effects descriptions in natural language context. WWW/Internet 2013 : proceedings of the IADIS International Conference (Fort Worth, Texas, USA, October 22-25, 2013) / IADIS (International Association for Development of the Information Society).

[10] Korobkin, D.M., Fomenkov, S.A., Kolesnikov, S.G., Orlova, Y.A. A multi-stage algorithm for text documents filtering based on physical knowledge. (2013) World Applied Sciences Journal 24 (24) PP. 91 – 97.

[11] Korobkin D.M., Fomenkov S.A., Kolesnikov S.G., Voronin Y.F. System of Physical Effects Extraction from Natural Language Text in the Internet. //World Applied Sciences Journal, 2013, v.24 (Information Technologies in Modern Industry, Education & Society), pp. 55-61.

[12] Hall, Johan.: MaltParser – An Architecture for Inductive Labeled Dependency Parsing, http://hall.maltparser.org/cv/pub/msireport06050_johan_hall_lic_final.pdf.

[13] Musen, M. Domain Ontologies in Software Engineering: Use of Protege with the EON Architecture//Methods of Inform. in Medicine, pages 540-550, 2010.

[14] OWL Web Ontology Language. Overview:W3C Recommendation 10 February 2004 Electronic resource. / eds. Deborah L. McGuinness, Frank van Harmelen. Access mode: http://www.w3.org/TR/owl-features/.

[15] Marcelo Arenas, Claudio Gutierrez, Carlos Hurtado, Jorge Pérez. Semantics and complexity of SPARQL. ACM Transactions on Database Systems (TODS) TODS Homepage archive. Volume 34 Issue 3, August 2009

[16] Korobkin, D., Fomenkov, S., Kravets, A., Kolesnikov, S., Dykov, M. Three-Steps Methodology for Patents Prior-Art Retrieval and Structured Physical Knowledge Extracting // Creativity in Intelligent Technologies and Data Science. CIT&DS 2015 Proceedings / ed. by A. Kravets, M. Shcherbakov, M. Kultsova, O. Shabalina. – [Switzerland] : Springer International Publishing, 2015. – P. 124-136.

[17] Tarasov, V.B, 1998. Artificial life and fuzzy evolution multiagent systems: the main theoretical approaches to the design of intelligent organizations. Journal of computer and systems sciences international, v.37(5), pp. 674-683.

[18] The Foundation for Intelligent Physical Agents, http://www.fipa.org.

[19] FIPA Contract Net Interaction Protocol Specification, http://www.fipa.org/specs/fipa00029.

[20] Agent Communication Language, http://www.fipa.org/specs/fipa00018/OC00018.pdf.