Interpretation of School Geometry Text Using Applied Ontology

Kurbatov S.S.
Scientific Research Centre for Electronic Computer Technology
Moscow, Russia
Email: curbatow.serg@yandex.ru

Vorobyev A.B.
Moscow Power Engineering Institute
Moscow, Russia
Email: cool-t2007@mail.ru

***Abstract*—The paper proposed approach to mathematical text interpretation using applied ontology. An experiment on partially automated creating of ontological conception of geometry texts for school was described (Including axioms, theorems and tasks).**

***Keywords*—geometry; ontology; natural language.**

I. INTRODUCING

The Problem of text interpretation in natural language (NL) using AI methods has a wide range of solutions from machine-generated translation to knowledge extraction using data mining (after natural language text processing). The Approach to a problem of mathematical texts interpretation using applied ontology is analyzed in this paper. The task is to obtain ontological representation of school geometry texts. Such representation must identify couplings between axioms and theorems, and realize ontology-based structure by geometry task NL-text for solver module.

Knowledge extraction from texts and moving from task NL-description to semantic representation have been described in number of papers [1, 2, 3], but above described task is different by its essential novelty. In this paper a full mathematical text translation is supposed to include the final representation in integral system that is able to solve geometric tasks automatically using its natural language description and display obtained solution in graphical representation. The graphic contains also NL- comments about solution steps and its description.

Thus stated task solution combines NLP achievements, ontology-oriented solution and computer-generated graphics, resulting in a fresh look on to the use of ontologies in partitioning NL-texts. In the applied aspect the solution opens prospects of its use to gain qualitatively new education level. The above mentioned information is the basis of the task solving relevance.

II. OVERVIEW

A. Ontology

General characterization of ontology arrangement software is given in [7]. The original ontology implemented using DBMS Progress software is used. The ontology is based on semantic web and uses a hypergraph framework to organize knowledge.

B. Subject Area

School geometry was selected as a subject area, but ontology representation illustrates the possibility of using Hilbert’s and Tarski’s axioms elements [4, 5].

C. Operation Logic

General logic of text interpretation implies linguistic processing of main array of text to create an ontological representation. Later the representation could be edited and checked visually or using solver. This process is described in more details in section III.

III. EXPERIMENT

Text for the experiment was taken from middle-school geometry textbook. Text fragments were extracted using Microsoft Word macros. Then these fragments were linguistically processed. As a result, an ontological structure was obtained which created semantic network of axioms, theorems and tasks.

A. Linguistic Processing

Text processing consisted of two steps: 1) revision and correcting; 2) linguistic translation. Within the first step text fragments were selected for the second step, where these fragments were translated to ontology using linguistic features. Especially most important fragments were marked with tags “theorem”, “proving” and “the theorem is proved”.

Morphologic and syntax steps of linguistic processing are based on knowledge about language and subject area. Semantic analysis is based on paraphrase conception [6], adapted for paper objects in view.

The full automation of Linguistic processing for real textbook is rather complicating. At this point of research the processing was executed in semi-automated mode, while the programming features were used as tools. It assisted the researcher in marking important text fragments and editing the automatically obtained ontological structures.

B. Ontological Representation

A graphic representation of a specialized syntactic structure of one of the first theorems is shown on fig. 1. The vertices’ connections generally conform to the NL-interrogatives (“who?”, “what?”, “where?” and etc.) reflecting the plurality and logic. Russian words on fig. 1 mean: “если” – “if”,

This paper was supported by grant RFBR project № 15-07-03847 "Interpretation of technically-oriented texts using application ontologies."

"равны" – "equal", "то" – "then", "и" – "and", "соответственно" – "conforming to" and etc. These words are concepts and relations of ontological representation, so there is no need to translate them all on English.

In terms of the subject area, the ontological structure is similar to the table representation on fig.2 (figure shows only small part of the table). Names, types and relations are self-explanatory, l_part and r_part are the left and the right parts of a theorem (an axiom or a rule). Also like on fig.1 words on fig. 2 are no need in translation on English for the same reason.

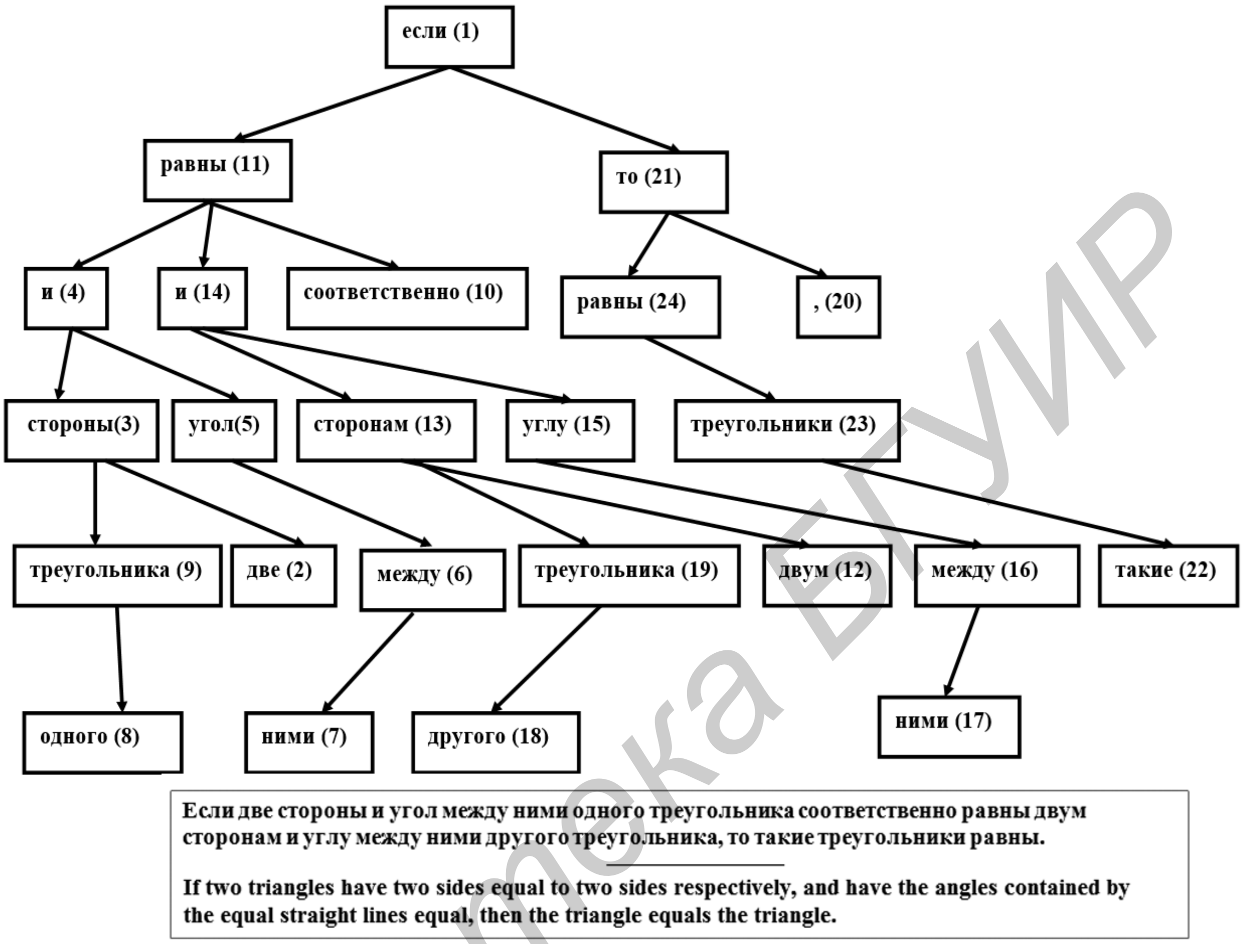

Fig. 1. The theorem text syntactic representation.

Object types and relation names act as links to the according geometry concepts (a point, a straight line, to belong, between, to equal). It allows to use the ontology heuristics, which descriptions contain the according concepts.

C. Examples

The ontological representation shall be shaped as a result of interpretation reflecting connections between of axioms and theorems, as well as the ontological structures reflecting tasks from the text. Examples of number of syntactic and ontological structures obtained in experiment are shown on HTML-page [7].

IV. Entire system

Let us briefly run through the general pattern of the system based on the ontological representation descried above.

A. General pattern

General pattern of the system is shown on fig. 3. The system includes linguistic translator, ontology, solver and graphic module. Syntactic structures of theorems similar to that shown on Fig.1 are compared with left sides of paraphrasing rules.

In case of successful comparison in the ontological representation the structures corresponding to the right part are formed. This process includes a lot of detail is not a part of the topic of this work. These details are related to the rule choosing (productiveness and cycling protection), to the anaphoric and elliptic constructions solving and etc.

B. Solver

A solver receives the ontological structure of task and forms a chain of basic operations using knowledge of the subject area. Further this chain is converted to text and sent to the graphics module to be executed.

Редактирование Сервис Выход

Код	Область	№ акс	Тип объекта-1	Имя объекта-1	Отношение	Тип объекта-2	Имя объекта-2	Левая/правая
1		1	точка	T-A	различны	точка	T-B	l-part
2	Геометрия	1	точка	T-A	на	прямая	Pr-1	r-part
3	Геометрия	1	точка	T-B	на	прямая	Pr-1	r-part
4	Геометрия	2	точка	T-A	различны	точка	T-B	l-part
5	Геометрия	2	точка	T-A	на	прямая	Pr-1	l-part
6	Геометрия	2	точка	T-B	на	прямая	Pr-1	l-part
7	Геометрия	2	точка	T-A	на	прямая	Pr-2	l-part
8	Геометрия	2	точка	T-B	на	прямая	Pr-2	l-part
9	Геометрия	2	прямая	Pr-1	совпадают	прямая	Pr-2	r-part
10	Геометрия	3	точка	T-A	различны	точка	T-B	l-part
11	Геометрия	3	точка	T-A	различны	точка	T-C	l-part
12	Геометрия	3	точка	T-C	различны	точка	T-B	l-part
13	Геометрия	3	точка	T-A	на	прямая	Pr-1	l-part
14	Геометрия	3	точка	T-B	на	прямая	Pr-1	l-part
15	Геометрия	3	точка	T-C	вне	прямая	Pr-1	r-part

Fig. 2. Screenshot of table of axioms, theorems and rules.

The ontological structure of solution is an acyclic graph. The theorem structure is also acyclic including direct and inverse theorems (e.g. Pythagoras theorem and its inverse theorem). Theorem organization suggests the possibility of certain restrictions on their use. This is important for solving the tasks in conditions of restrictions on an algorithmic basis.

C. Interface and graphics

A graphic module displays the result and comments, describing the actions. In current version graphic is realized using Microsoft Word macros that is determined by the applied aspects. Macros generated drawing allows observing solution steps (forward and backward) and keeping communication with ontology.

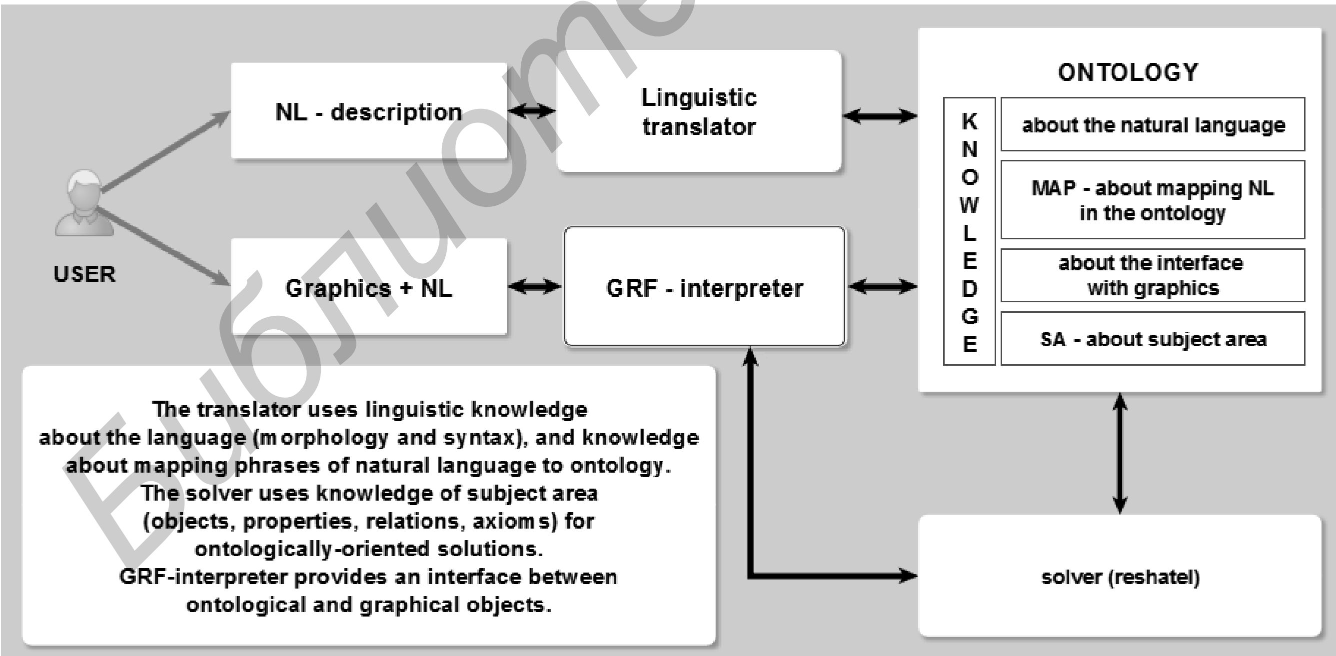

Fig. 3. System general pattern.

Only syntactic structures are contained in left side of paraphrasing rules. In the right side could be contained both syntactic groups and subject area oriented structures. In the latter case so called canonical NL-description of ontological structure is included in left side. This description is most clearly fixed NL-content in subject area terms.

For example phrases “point passes through line”, “point in the line”, “line passes through point” and etc. should be reduced to canonic description “**point contained in line**”.

For chronologically corrected theorem ontology forming was important to have regarded that the basic geometry axioms are listed in the end of the textbook. This is useful for pedagogical purposes, but it makes difficult the formation of the text ontological structure. It is the reason for partially automated representation creating, it means need for researcher intervention into the process.

V. Conclusion

The experiments have generally confirmed the prospects of the proposed approach. Its development intends increasing the level of automation in the ontology formation, improving solver capabilities and testing on a huge set of tasks.

References

[1] Хорошевский В.Ф. OntosMiner: семейство систем извлечения информации из мультиязычных коллекций документов // Девятая Национальная конференция по искусственному интеллекту с международным участием КИИ-2004. Т. 2. – М.: Физматлит, 2004, с.573-581.

[2] Осипов Г.С., Приобретение знаний интеллектуальными системами. Основы теории и технологии, М.: Наука, 1997.

[3] Симаков, К. В., Модели и методы извлечения знаний из текстов на естественном языке, Москва, МГТУ имени Н Э Баумана), Диссертации в Техносфере: http://tekhnosfera.com/modeli-i-metody-izvlecheniya-znaniy-iz-tekstov-na-estestvennom-yazyke#ixzz4QXlOO2Pv.

[4] David Hilbert: Grundlagen der Geometrie. 14. Auflage. Teubner, Stuttgart 1999.

[5] Беклемишев Л.Д., Введение в математическую логику Мех-мат МГУ, курс лекций, 2008, с. 44-48.

[6] Apresjan Ju.D. et al. (1989), Lingvisticheskoe obespechenie sistemy ETAP-2, Science, Moscow.

[7] http://www.eia--dostup.ru/onto_geom.htm.

ИНТЕРПРЕТАЦИЯ ТЕКСТОВ ШКОЛЬНОЙ ГЕОМЕТРИИ С ИСПОЛЬЗОВАНИЕМ ПРИКЛАДНОЙ ОНТОЛОГИИ

Курбатов С.С., Воробьев А.Б.

В работе предложен подход к интерпретации математических текстов с помощью прикладной онтологии. Описан эксперимент по полуавтоматической генерации онтологического представления текстов школьной геометрии. Разработана структура представления в онтологии аксиом, теорем и задач. Представлена схема целостной системы, включающей лингвистическую обработку текста, онтологический решатель и средства визуализации.