Construction of Semantic Networks for Natural Language Texts

Tetyana Korotyeyeva[1], Ruslan Tushnytskyy[2]

1. Software Department, Lviv Polytechnic National University,
12 S. Bandera street, Lviv, UKRAINE;
e-mail: 1. tat@polynet.lviv.ua,
2. ruslan.tushnytskyy@gmail.com

Abstract – Paper conducted a review of existing types of semantic networks, analyzes the existing algorithms semantic analysis of natural language texts, and proposes a new approach to analyze the semantics of the text in the Ukrainian language.

Keywords – semantic networks, semantic analysis, intellectual data processing, natural language, texts processing.

I. Introduction

Permanent and continuous development of computer technology provides the opportunity to develop such demanding computational resources areas as research in the field of artificial intelligence. Efficient acquisition of knowledge from texts is an ancient order field of artificial intelligence. The first studies in this field have created quite logical approaches, but the results obtained by the algorithms are quite weak in terms of correct knowledge. Development of the Internet was the impetus for the development of the field of machine learning.

II. Semantic Analysis Problem

Semantic analysis of natural language texts is a complex and non-trivial problem, which needs considerable computing resources and complex algorithmic approaches. Also, this problem is unique to each language, as well as all the languages differ in their syntactic and morphological characteristics. That is what makes the task very challenging and difficult-adaptive for each language.

Throughout the history of the development of artificial intelligence and knowledge extraction from natural language text has been developed a set of algorithms and approaches to learning from this little structured kind of knowledge. After each sentence and its syntactic component of a kind and unique, especially in the Ukrainian language and extremely diverse. It is important to distinguish the following approaches for the analysis of natural language texts: *heuristic approach*; *clustering*; *an approach based on building information*; *parsing*.

III. Semantic Networks Investigation

Proposed algorithm is based on parsing a single sentence, and based on this analysis, the selection of entities and relationships between them. This algorithm also serves as a basis for developing algorithms for comparing two texts of the degree of their similarity and to generate a set of keywords for the analyzed text. With the generated semantic networks text of any dimension can be represented in a convenient end-user representative form – as the graph whose vertices are the essence of the text (the main characters, geographical location, basic definitions etc.) and the ratio between them, which is described as one entity affects another.

To investigate the construction of a semantic network based natural language text, was chosen three texts the same size but different subjects. Paper size is approximately thirty pages of A4. These texts differ stylistically and type of their origin – was considered a scientific text type, literary text, and text built from a combination of information taken from the internet. Table 1 shows a comparative description of different types of texts.

TABLE 1

TEXT ANALYSIS OF DIFFERENT TYPES

Type the text	Entity of the total number of words, %	Relations between entities of the total number of words, %	Obtained information, %
Research paper	18 %	11 %	29 %
Artistic text	13 %	19 %	31 %
Data from Internet	5 %	3 %	4 %

Conclusions

A new algorithm for the selection of ontologies from texts, based on their parsing has developed. To analyze the effectiveness and applicability of the algorithm software was designed and implemented.

The results indicate the possibility of obtaining correct data by parsing texts. This approach is language-oriented, as it is based on dictionary words in a certain language and meaning of the language of words. This achieves a best analysis of texts.

References

1. Y. Wilks, D. Fass, “The preference semantics family”, Computers & Mathematics with Applications, vol.23, issues 2-5, pp.205-222, 1992.
2. E.A. Feigenbaum, “The art of artificial intelligence. Themes and case studies of knowledge engineering”, Proc. of 5th IJCAI, pp.1014-1029, 1977.
3. E. Agichtein, L. Gravano, “Snowball: Extracting relations from large plaintext collections”, Proc. of the 5th ACM Intern. Conf. on Digital Libraries (DL-00), pp.85-94, 2000.