



XIIth International Symposium «Intelligent Systems», INTELS'16, 5-7 October 2016, Moscow, Russia

Usage of algorithms for the analysis natural languages texts based on intensional logic of Montague in area of finding answers on questions

Anatoliy Shvetcov*, Denis Gorbunov

Vologda State University, 15 Lenina str., Vologda 160000, Russia

Abstract

The issue of analysis and extraction of information from big data in natural language is one of the most difficult and important ones today. The search based on a meaning is one of the most perspective ways in this issue. The method of text analysis in natural languages, based on an intensional logic of Montague, which is adapted for Russian language, is being explored in this work. For this issue, a special algorithm had been made and tested on the corpus of text from a social network.

© 2017 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the XIIth International Symposium “Intelligent Systems”

Keywords: natural language processing; intensional logic of Montague; information extraction; big data.

1. Introduction

In the time of a global networks and the Internet, full of various data and information, the extraction of the useful knowledge from this data is an important thing. The traditional search systems, which are based on the similarity of words, synonyms, reference ranking, sometimes are not as effective as we need. One of the reasons – these systems aren't trying to understand the meaning of the texts that they analyze. That's why the information systems based on the meaning analysis would be more suitable.

* Corresponding author.

E-mail address: isit@mh.vstu.edu.ru

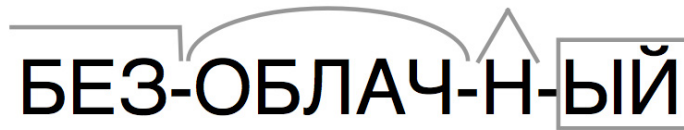


Fig. 1. Morphological analysis of a Russian word.

2. Method of Montague

The method of intensional logic of R. Montague is shown in this article. This approach to the semantic search was modified for Russian language and the area question-answering.

One of the issues you need to solve in the question/answer system is an understanding of natural language question from a user. In his main work, “English as formal language”, Montague says that every natural language we can view as a formal one. According to this approach, the meaning of sentence is a logic formula. This formula can be produced from analysis of a natural language sentence. In his work Montague show 2 ways of formalizing the natural language sentence:

Direct - describing and interpretation of a natural language in models

Two-step - in this way Montague uses the intermediate language of intensional logic. It means that at first, a natural language sentence is translated to the intermediate language. Afterwards the sentence in the intermediate language is transformed to models.

For question-answering problem, the second approach is more relevant.

Main principle of the formal semantic says that the relation between syntax and semantic is composite. That means that if we know the syntax structure of a sentence, we can build the logic formula.

But for Russian language this concept needs to be changed a little bit. That is because English and Russian languages are different in structure. English language is an analytic language, in which syntax is very important to show the meaning. Russian language is a flecnal language. This means the form of a word is very important to see the structure of a sentence. So even just knowing words of the sentence, we know a lot about the whole sentence.

It’s an adjective, masculine, singular, subjective case.

Here is shown that a word in Russian has complex structure. Each element of this structure can say something about the word - number, case, time, etc.

Montague implemented the concept of categories of the formal semantic to make a parity for the syntax structure of sentence. According this table, every syntax element in the natural language sentence has a match in this classification.

For example, we want to analyze this sentence in Russian natural language: “Каждый охотник желает знать, где сидит фазан”. At first, we analyze each word separately and then make a structure.

- “Каждый” ('ADJF,Apro inan,masc,sing,accs') - adjective, pronominal inanimate, masculine, singular, accusative case
Category - DET.
- “охотник” ('ADJF,Apro inan,masc,sing,accs') - noun, animated, masculine, singular, the nominative case
Category CN.
- “желает” ('VERB, impf, tran sing,3per,pres,indc') - verb, imperfect, transitional, singular, 3rd person, present time, indicative
Category - TV.
- “знать” ('INFN,impf,tran') - infinitive, imperfect, transitional
Category - TV.
- “где” ('ADV,B,Ques') - adverb, questional
Category - IAV.

Download English Version:

<https://daneshyari.com/en/article/4961458>

Download Persian Version:

<https://daneshyari.com/article/4961458>

[Daneshyari.com](https://daneshyari.com)