METHODS OF ANALYSIS AND SYNTHESIS OF ARTISTIC TEXT

Sevara Abbos Kizi Khaydarova
Lecturer, Journalism and Mass Communications University

Abstract

The article examines automatic text analysis, a number of complex operations performed by text on computer text, and its closeness to human language according to the given algorithm. The development of the theory of automatic text analysis and synthesis (development of linguistic foundations for the creation of artificial intelligence) and the implementation of practical needs are important tasks of computational linguistics.

Keywords: syntactic and semantic text analysis, formal grammar, automatic text synthesis, morphological analysis.

Introduction. Currently, information technology is an integral part of any field of professional activity, including linguistics. Studying the possibilities of using information technologies in linguistics implies knowing the basic concepts of the relevant field of knowledge, including linguistics (language, linguistics, computational linguistics, etc.) and informatics (information, algorithms, models, etc.).

Automatic text analysis involves several steps.

1) graphic analysis: highlighting the boundaries of words, sentences, paragraphs and other text elements (for example, sidebars in newspaper text);

2) morphological analysis: determining the initial form of each word used in the text and the set of morphological signs of this word;

3) analysis: determining the grammatical structure of text sentences;

4) semantic analysis: determining the meaning of phrases. Graphical analysis is also defined as tokenization (noun English token = a separate word, phrase or other significant element of the text). Different types of delimiters act as formal signals of the boundaries of text elements: spaces, boundaries between words, capital letters and punctuation marks, boundaries between sentences and sentence components, denoting boundaries between groups of sentences related to meaning paragraphs etc.

However, the formal way of defining word boundaries is not always used. For example, Chinese has no formal language boundaries. In addition, common European languages also have space-separated fixed word combinations that should be treated as a single lexeme, e.g. New York, such cases should obviously be accounted for in systems. Graphematic analysis, for example, by creating a list of detailed tokens, each of which is used in morphological analysis, and the word in the text is raised to its original form and determines a set of morphological features of the text form of the word: these are the part of speech; gender, number and case for nouns, number of words for verbs, etc.

Each word used in a text is called a form of speech (or word usage). Repetition of the same words is required to ensure the coherence of the text, so often different word forms of one or more sentences of the text are raised to the same original form: This is my village; Here is my house. Here I am rolling on a sled along a steep mountain.

Our usual dictionaries usually list not the word forms, but words shortened to a known and similar form.

Main part: Depending on the type of language, a lemma (the dictionary form of a lexeme) or a root (the core part of a word without a word that changes morphemes) can be used as the initial form of word forms used in the text. For example, the English words swim, swims, swam, and refer back to the swimming lemma. In word-rich inflectional and agglutinative languages, the transformation to store all possible word forms requires large memory resources. For example, Russian nouns that modify numbers (2 numbers) and cases (6 cases) have 12 word forms. The Russian verb is more characteristic, has a complex set of grammatical features and, accordingly, a large number of plural forms. In this case, it is more convenient to use its base as the initial form of the rising word. True, the term "base" in morphological analysis does not always have the same meaning as included in canonical (school) grammar. For example, if there is an exchange of letters in the word (sitting - I am sitting, friend - friends, etc.), in these cases the base (more precisely, quasi base or machine base) is part of the word, not only morphemes o A 'modifying word may be without changing letters, i.e. si # and so on # respectively.

This type of accentuation of bases is called tourismem, i.e. the formation of different word forms on one quasi base. Stemming is well suited for solving some automated tasks, such as searching the Internet. Thus, the noun photograph and photographic adjectives correspond to the user's request for a photograph as complete or incomplete quasi-bases. As a result of the search, the user receives a list of documents with the phrase photographic portrait and portrait photography.

For morphological analysis, the concept of machine base, which is understood as a sequence of letters from the beginning of the word form common to all word forms that are part of the formative paradigm of a given word, is important. The next step is to determine whether the word belongs to the part of speech (part-of-speech tagging) and its morphological features, which often occurs with the help of layering of modifying elements of the word (machine ending).

The end of the computer - the elements describing the form will be expressed in the form of understanding and paradigm of a particular lexeme. All possible sets of computer suffixes are anchored in a typical lexeme paradigm. In this, on the one hand, it is possible to observe that the typical paradigms of different lexeme forms (hence, computer windows), for example, pen and stroke, and on the other hand, the correspondence of lexeme forms to computer bases, having different typical paradigms, typical paradigms of the machine base of false # lexemes can be compared.

A complete morphological description of each word form is carried out at the end of the computer included in certain typical paradigms. For example, such rules are as follows: If the form of the word is also a verb, it can be a homonym, and there is an article in front of it, then this form of the word is a noun.

If a word form can be both a main clause and a subordinate clause, and there is no verb after it until the end of the sentence, then this word form is a main clause. Parsers are used for automatic morphological analysis - special computer programs work for automatic analysis of words. In addition to morphological parsers, there are also syntactic parsers. used for automatic analysis of syntactic structures of sentences. General morphological analysis includes the following steps:

1) normalization of word forms with the form of lemmatization, that is, bringing different word forms into some unified image - the original form (lemma) or root form, or to reduce different word forms to one quasi-stem;

2) determining the part of speech, i.e. determining the part of speech for each word, forms in the text;

3) complete morphological analysis - giving grammatical signs to the word form. In the analysis, it is necessary to determine the role and words in the sentence and their relationship with each other. The result of this stage of automatic analysis is to present the syntactic relations of each sentence in the form of models, for example, in the form of a dependency tree. The problem of parsing is the presence of alternatives for parsing (syntactic ambiguity), compare: three coats— (how many?) three (what?) coats three coats— (what to do?) three (what?) coats.

The appearance of syntactic ambiguity is determined by the lexical-morphological semi-ambiguity of word forms (the same word form can return to different initial forms or different morphological forms of the same lexeme), as well as the uncertainty of the analysis rules.

It can be aimed at representing a syntactic structure, for example directly in the form of a tree of components or a tree of dependencies.

Thus, the sentence "The girl washes the floor" is depicted in the first case with the model shown in Figure 1, and in the second - in Figure 2.[4] In the model of direct components, it is important to divide the syntactic structure into its paired elements: a sentence (S) is divided into a subject group (NP), in this case represented by a single noun (N) and a predicate . group (VP). Second variable verb is divided into (V) and adverb (O). In the dependency tree, the starting point of the analysis is the predicate (V) located at the top of the graph, to which the subject (N) and the complement (O) depend. As a result, x in both types of analysis is the starting point of the analysis. the same syntactic units differ - N, V and O - the syntactic relations between them are different. S Soap V N VP Girl gender V O Fig. 2 Dependency tree Girl soap gender Fig. 1. Direct structure tree It is true that in some cases the same syntactic structure requires the construction of different syntactic models.

He wears his coat on the street He wears a fur coat In order to choose the correct reflective model, syntactic relations in a particular sentence, in such cases, semantics should be involved. Semantic analysis is perhaps the most complex branch of automatic text analysis. In this case, it is necessary to establish semantic relations between the words in the text, to combine different linguistic expressions related to the same concept. Case grammars and semantic cases (valence) are used for semantic analysis of sentences. In this case, the semantics of the sentence is described by the connection of the main word (verb) with its semantic actives.

For example, the verb transmit is characterized by the semantic states of the giver (agent), receiver and object of transmission. Semantic analysis is based on the idea that the meaning of a word is not an elementary semantic unit. It is divided into more elementary meanings - vocabulary units of the semantic language. These units of semantic language are a kind of atoms, and from their various combinations "molecules" - the meanings of real words of natural language - are formed. For example, if there are elementary meanings "myself", "someone", "to have", "force", "stop", "start" and "no", then with their help we can o We can identify the whole group of words in the Uzbek language. In addition to the seven named words, which are both elements of the semantic language and words of the Uzbek language, there are the following words among them: 1) to own = "to own", 2) to own = "to be", 3) take = "have yourself", 4) give = have someone ". It is semantic analysis that allows solving the problems of semi-family (homonymy) arising from automatic analysis at all linguistic levels.

Lexical homonymy: words that do not have common elements of meaning in sound or spelling, such as scarlet, face, and type of disease. Lexical-morphological homonymy (the most common type of homonymy). Syntactic homonymy: a structure in which syntactic ambiguity has multiple interpretations. Automatic synthesis is a process of creating a coherent text, the separate steps of which are the same as morphological analysis, but in the reverse order: semantic synthesis is performed first, then syntactic, morphological and graphic.

Semantic synthesis - transition from semantic sign and word combination to its syntactic structure; syntactic - transition from the syntactic structure of word forms representing a word combination to a chain of lexical and grammatical signs; lexico-morphological - transition from lexical and grammatical features to real word form. During the morphological synthesis according to the normal form of the word and its parameters, the program finds a suitable word form.

Conclusion: Graphematic synthesis combines words into a single text, checking whether input text parts match output parts. One of the first computer programs to synthesize written dialogue in English is the Eliza program of the American scientist Joseph Weisenbaum. The program includes minimal linguistic information:

1) a set of keywords that implement some stable speech formulas;

2) the ability to turn the previous sentence into a general question.

The work of the Eliza program is illustrated by the following excerpt from a conversation between a young woman and its creator, the Eliza program, published in 1966 by D. Weizenbaum.

REFERENCES:

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. М.: МИЭМ, 2011.С.91—97,106—111.

2. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие.3-еизд. М.: ЛКИ, 2007.С.20—25.

3. Всеволодова А.В.Компьютерная обработка лингвистических данных: учеб. пособие.2-еизд., испр. М.: Флинта: Наука,2007.С.50—51,66—67.

4. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие.М.: Академия,2004.С.33— 34.

5. Тешабоева З. Қ. Бобурнома” нинг инглизча таржималаридаги фразеологик бирликларнинг когнитив ва лексикографик тадқиқи. Филол. Фан. доктори (DSc) дисс. автореф. – 2017.

6. A Cognitive Study of "Baburname"'S Translations and Principle of Compiling a Textual Dictionary. (2022). *Journal of Pharmaceutical Negative Results*, 1994-2006. https://doi.org/10.47750/pnr.2022.13.S08.244

7. Teshaboyeva Z. TRANSLATIONS OF PHRASEOLOGICAL UNITS AND PROVERBS OF" BOBURNOMA" INTO ENGLISH //Paradigmata poznani. – 2016. – №. 3. – С. 74-78.

8. Kamola Abdujabarova. IMPROVING LITERACY SKILLS THROUGH LEARNING READING. Журнал Наука и образование. 4/2021. ISSN 2414-5718 p.78

9. Abdullayeva, Markhabo Raxmonkulovna TRANSLATION PROBLEMS OF VERB PHRASEOLOGISMS EXPRESSING NATIONAL COLOR // ORIENSS. 2022. №Special Issue 25. URL: https://cyberleninka.ru/article/n/translation-problems-of-verb-phraseologisms-expressing-national-color (дата обращения: 15.12.2023).

10. ABDULLAYEVA MARXABO RAXMONKULOVNA. Milliy koloritni ifodalovchi frazeologizmlar tarjımasi muammolari (Agata Kristi asarlarining o'zbekcha tarjimalari misolida). Fil. fan bo'yicha falsafa.dok.. (PhD) Diss.. date. 2023/11