

Zubchenko A.

National Aerospace University "Kharkiv Aviation Institute"

ALGORITHMS AND SOFTWARE

FOR ABSTRACTING TEXTUAL INFORMATION

The modern world has about 7151 languages, which belong to 142 most diverse parts of their manifestation. Hence the problem of their abstracting and

annotating - their number, variety of features and complexity of each language are unique in their own way. Therefore, in our scientific work, we study a topic that can radically simplify and optimize our future work - algorithmization and software solutions for abstracting and annotating information.

The relevance of scientific research is based on the difficulty of recording and translating information in large volumes, if referring to a language that was not used before, research in existing databases.

The scientific novelty consists in the discovery of new possibilities for the translator, their expansion, acquisition of new skills in the use of application programs for annotation and abstracting of data for further optimization and acceleration of work, without loss in the quality of the work performed.

The object of research is software tools and algorithms of actions of a potential user.

The subject of research is algorithms and software tools for abstracting and annotating textual information.

The hypothesis of this research is to reveal the fact that when using PZ and PP it is possible to achieve greater efficiency, two cases can serve as a comparison: Cambridge spent 50 years writing three volumes of the English dictionary, when as for the translation of volumes 1-4 of Leo Tolstoy's "War and Peace" 2 years were spent on English.

The purpose of this work is to study techniques such as annotation and abstracting, their operation using algorithms, programming languages, etc. The study of such a concept as quasi-referencing and the ability to write abstracts in ways that exclude the superfluous from the context.

The main tasks to achieve the goal of the work are:

1. Familiarize yourself with the types of documents and essays.
2. You need to be able to clearly understand the difference and what functions annotation and abstracting perform.
3. To know what bibliography is, to know algorithms for simplified compilation of abstracts.
4. Consider how to quickly edit text using a computer, thanks to commands in the C# programming language.
5. Familiarize yourself with such a concept as quasi-referencing.
6. To form an understanding of the pros and cons of the technical method of obtaining document characteristics.
7. See an example of shortened abstracting, thanks to new algorithms on real text.

The methods of this work were methods of algorithmic compilation of abstracts and annotations in a certain programming language.

The material for this work was taken as an example from the BBC website on the Internet to show exactly the algorithmic technique of abstracting.

The theoretical significance of the work is that the conducted research is a contribution to linguistics, especially for the future, where all the work for humans will be performed by machines, with the help of algorithms.

The practical value is the possibility of using research results in modifying the technical use of software and algorithms with the aim of further optimizing work.

C# language and its' algorithms of text information processing:

It is possible to implement the connection of databases with further user intervention, such as the selection of the profile database "military, scientific, oil, journalistic, household, etc. An example can be the C# language, developed specifically for such cases:

Case 1 (selection of filters, in the program it will be possible to configure which case the user needs);

Console.WriteLine(); (for the text field for entering information required by the user);

Static.Void.LibraryAdd Example; (adding a database to implement the automatic translation function).

Other methods of processing text information without using computers:

1. **The hint method.** According to this method, sentences for the extract were selected from the text on the basis of a dictionary of "positive", "negative" and "zero" words previously compiled by a person. Words with zero weight were not taken into account when selecting sentences.

2. **Keyword method.** It was based on statistical criteria for extracting keywords from the text using the method of H. P. Lun.

3. **The header method** is one of the types of positional methods proposed by the developers. It was believed that the main content of the document is expressed by its title and subtitles. The dictionary was made up of the words included in the headings and subheadings.

4. **Method of localization (location).** This is the second method, which also refers to the type of positional. It is based on research, which shows that the most informative are sentences containing information about the purpose,

methods, conclusions and results of the study. These sentences can be at the beginning or end of a paragraph, a paragraph, a passage of text. Depending on the location, they receive the appropriate weight.