Automatic generation of methodological materials in the study of technical and engineering texts in a foreign language

Sergej Bronov\textsuperscript{1,2,3} and Elena Stepanova\textsuperscript{4}

\textsuperscript{1}Siberian Federal University, Svobodny 79, Krasnoyarsk, 660041, Russia
\textsuperscript{2}Krasnoyarsk State Agrarian University, Mira 90, Krasnoyarsk, 660049, Russia
\textsuperscript{3}E-mail: nulapr@mail.ru

Abstract. Masters, postgraduate students, engineers and scientists in Russia should study new areas of knowledge in a foreign language, for example, in English. But they have a double task: 1) learning English for a specific professional field; 2) studying the professional field using the acquired knowledge of the English language. The main requirement here is the minimum time for training. The purpose of the training process under consideration is to develop knowledge and skills for reading texts in a specific area of knowledge. At this stage the problem of active English language proficiency, i.e., writing your own articles or speaking is not considered. The student does not have the similar material in Russian. The material studied contains new concepts that are not known to the student. In standard university textbooks simple material is used that is already known to the student in Russian. Therefore, the student can understand the meaning of the content which significantly simplifies the learning process. But in this case the situation is different. Automation of training material design within the framework of the situation is carried out in two stages: the analysis of the studied materials and the synthesis of training materials. At this stage the existing software such as Microsoft Word and Excel together with Visual Basic for Application (VBA) can be used. The corresponding algorithms are applied for each operation. The automated development of training materials' proposed technology for mastering a foreign language in a narrowly specialized field of knowledge enables to increase the effectiveness of subsequent training in a foreign language.

1. Introduction
Masters, postgraduate students, engineers and scientists in Russia should study new areas of knowledge in a foreign language, for example, in English \cite{1, 2, 3}. But they may not know English well. Then a double task arises: 1) learning English for a specific professional field; 2) studying the professional field using the acquired knowledge of the English language. The main requirement here is the minimum time for training. Therefore, it is necessary to master English only as much as it is required in order to understand the specific material being studied.

Any student usually translates the required text using a dictionary and thus learns a new area of knowledge. But if the student has a poor command of a foreign language it is required to simultaneously study this language. The universal training manuals are used for this purpose. They are not designed for studying a specific material but for teaching the language for a wide area of knowledge such as computer science, electrical engineering, mathematics, etc. This process is ineffective as the student should learn not only what corresponds to the topic being studied but vocabulary and grammar phrase structures as well which are not in the material planned for study.
The main idea of the work is to form the training material based on the texts studied combining the study of a new area of knowledge with the simultaneous study of English for this specific field. It is proposed to automate the process of preparing training materials based on modern information technologies. The article considers the general principles and some methods of solving the formulated problem.

2. General algorithm for the training material formation

The purpose of the training process under consideration is to develop knowledge and skills for reading texts in a specific area of knowledge. At this stage the problem of active English language proficiency, i.e. writing your own articles or speaking is not considered. Such problems will be solved in the future.

A new field of knowledge is often represented by some materials such as a monograph, a textbook, a set of scientific articles, etc. They consider new knowledge for the student, and the corresponding terms and set expressions are used [4, 5, 6].

The features of this process are as follows. The student does not have the similar material in Russian. The material studied contains new concepts that are not known to the student. In standard university textbooks simple material is used that is already known to the student in Russian. Therefore, the student can understand the meaning of the content which significantly simplifies the learning process [7, 8, 9]. But in this case the situation is different. English vocabulary in the new field of knowledge may not correspond to the Russian language, i.e. the formation of new Russian terms can be required. When studying the materials in a foreign language these terms shouldn't necessarily be turned into Russian words. However, they should be at least understood in the material studying process.

Grammar structures and set expressions as well as vocabulary can be new, not previously used in the scientific literature. If a fundamentally new field of knowledge is being studied, then they are not considered to be the established set phrases and their understanding requires additional efforts.

Automation of training material design within the framework of the situation is carried out in two stages: the analysis of the studied materials and the synthesis of training materials (figure 1).

![Diagram](https://example.com/diagram.png)

**Figure 1.** Generalized scheme for the methodological materials formation.

For teaching English, the studied text is used which is analyzed. The text structure is defined as a result of the analysis:
• the analysis of the entire text;
• the text is divided into sentences;
• the sentences are divided into words;
• the words are combined into set expressions.

At this stage the existing software such as Microsoft Word and Excel together with Visual Basic for Application (VBA) can be used. The corresponding algorithms are applied for each operation.

The results of the texts analysis allow to form the training materials. Each class includes the typical educational elements:
• texts;
• pre-text vocabulary;
• exercises to memorize;
• tests to control the studied material.

The basic principles used here are as follows.

The training material is formed based on the studied texts. These texts are fed to the system input which results in forming the classes in the text form. Therefore, for each set of texts will be formed its own training material. The gradual presentation of new material is as follows: at each new class it makes up 5-10% of new words and expressions. The system includes the dictionary for the study area which is constantly updated taking into consideration the texts used.

3. Analysis of the texts array for the studied subject area and the synthesis of training materials

To study the subject area, the texts (books, articles) in electronic form are selected. Their analysis is carried out as in figure 2.

Figure 2. Generalized text analysis scheme.

The system analyzes the texts, distinguishes lexical units (words and phrases) as well as grammar structures. After that the system compares the highlighted words and grammar structures with those
available in the dictionary and demonstrates the new lexical units and grammar structures and includes them in the dictionary.

The system finds synonyms and makes the reciprocal links to each other in vocabulary and dictionary (it reflects the words meaning taking into consideration their usage in the sentences).

As a result, the information model of a foreign language text on a subject area is formed. It contains the information about the elements (words) and their various mutual links. This model is the basis for building a functional model. The functional model shows the links between vocabulary and grammar structures. On its basis it is possible to identify the system properties of texts related to the subject area under consideration. For this purpose, the theory of artificial intelligence systems is used: fuzzy logic, neural networks, etc.

4. Training materials formation

Training materials are formed based on the analysis results. They use words and typical grammar structures contained in the texts studied which are supplemented by general vocabulary and general grammatical phrases (figure 3).

![Synthesis of training materials]

Figure 3. Generalized synthesis scheme for training materials.
For the educational material' synthesis the process of gradual introduction of new vocabulary and grammar structures against the background of previously studied material is used as well as repetition in various combinations, tests verification and fixing on of authentic text fragments.

For the educational material automated' formation the appropriate algorithms are used based on the information model usage obtained as a result of the texts array analysis.

5. Conclusion
The automated development of training materials' proposed technology for mastering a foreign language in a narrowly specialized field of knowledge enables to increase the effectiveness of subsequent training in a foreign language.

References
[1] Drozd O V, Russkikh P A, Chentsov S V and Kapulin D V 2019 Structural-dynamic approach to the formalization of information exchange objects under integrated information environment IOP Conference series: Materials science and engineering 537 032077
[2] Drozd O V and Kapulin D V 2017 The model of electronic design document as a part of integrated information environment of radioelectronic enterprise Dynamics of systems, mechanisms and machines
[3] Taryma A K, Shershneva V A and Vainshtein Y V 2019 Development of professional ICT competence for future teachers of the Tuva republic under the conditions of bilingualism Perspektivy nauki i obrazovania 4 77–90
[4] Vainshtein I V, Shershneva V A, Esin R V and Noskov M V 2019 Individualisation of education in terms of e-learning: experience and prospects Journal of siberian federal university: Humanities and social sciences 9 1753–70
[5] Zykova T V, Shershneva V A, Vainshtein Y V, Danilenko A S and Kytmanov A A 2018 E-learning courses in mathematics in higher education Perspektivy nauki i obrazovania 4 58–65
[6] Kukartsev V V, Chzhan E A, Tynchenko V S, Antamoshkin O A & Stupina A A 2018 Development of adaptive educational course in the SibFU E-learning system J. of Siber. Fed. Univer. Humanities and Social Sci 11(5) 740–52
[7] Lozitskaja E V, Chzhan E A, Kukartsev V V & Tynchenko V S 2019 Automation of university curriculum construction using didactic unit arrays Journal of Physics: Conference Series 1353 012131
[8] Bronov S, Stepanova E, Katsunova A, Pichkovskaya S, Volkov M, Puh M and Sokolov P 2019 The process of forming a tree of didactic units at the automated curriculum' synthesis Education transformation issues 3 97–102
[9] Bronov S, Stepanova E, Pichkovskaya S, Sheluhin A and Panikarova N 2019 Information technology in the educational program design Journal of physics: IOP Conference series: APITECH-2019 1399 1399