Peculiarities of an Information System Development for Studying Ukrainian Language and Carrying out an Emotional and Content Analysis

Taras Basyuk, Andrii Vasyliuk

Lviv Polytechnic National University, Bandera str.12, Lviv, 79013, Ukraine

Abstract

This article analyzes the existing methods and well-known systems that provide tools for learning the Ukrainian language and describes the mechanisms for evaluating these skills. The technologies and software tools for conducting emotional content analysis were analyzed, which made it possible to identify the main shortcomings of the existing approaches and showed the relevance of the research. Models of structuring and formalization of knowledge and information, which are presented in the content of the educational environment, have been developed in order to provide a basis for the development and programmatic implementation of individualized user access to the requested educational materials methods. The design of the software system was carried out using a structural approach and displaying the created diagrams in accordance with the IDEF0 standard. The study presents a functional model and its decomposition, which created a basis for understanding the peculiarities of the Ukrainian language learning system functioning and the implementation of emotional and content analysis. The mobile application is built using the Flutter framework and written in the Dart language, providing fast compile times and dynamic reloading without the need for a restart. Natural language processing is implemented by a separate module that provides tokenization and parsing, lemmatization/stemming, part-of-speech tagging, and identification of semantic relationships. The process of implementing emotional content analysis is described, which includes the following stages: selection of a working algorithm, selection/creation of the sentiments' dictionary, creation of rules, collection, cleaning and storage of data, emotional content analysis, output and visualization of results. A software tool has been created that works in prototype mode and implements the described functionality.

1. Introduction

Learning any language is relevant in the context of business, literature, culture, or politics. Its knowledge enables people to follow events from relevant primary sources, the opportunity to conclude, and form a personal opinion and position regarding the chosen topic. In the modern pace of life, the world requires a person to have global knowledge, the best and most innovative of which can be obtained only by using various primary sources of information. At the same time, mastering any language makes it possible to communicate with a broader circle of people and to learn about culture, customs, and values from native speakers [1].

As for the Ukrainian language, the problem of learning it today is widespread, both among foreign citizens traveling for business, and among residents of our country who, due to objective circumstances, want to improve their language skills. The Ukrainian language ranks 44th in the world in terms of the

number of speakers, which is 45 million people who use it in everyday life or consider it their native language [2]. As you know, the most popular way to master it is to attend language courses or classes with tutors. However, these methods are financially and time-consuming and require changing schedules, which is not always convenient. Because of that, given the popularity of information technologies, the decision to make language learning accessible to anyone using software applications [3,4] is promising.

Another aspect of the research is the automated analysis of emotions expressed in the form of Internet comments about goods and services in the Ukrainian segment of the global network. The mentioned problem is relevant since more and more Ukrainian-language resources are being created on the territory of Ukraine, which do not include means of automated analysis for the analysis of reviews. Today, thanks to the Internet, customers share their experiences and opinions about products and services much more openly, and companies, in order to receive feedback, create a significant number of means of communication: telegram channels, hotlines, forms for evaluating customer satisfaction, chatbots, etc. However, often dissatisfied customers do not use these resources but share their experience on social networks, online resources with reviews, or personal blogs. These reviews can significantly affect the company's reputation, therefore it is essential to track various mentions of the brand and determine customer satisfaction depending on the emotional and meaningful analysis [5]. Therefore, an urgent task is to develop an information system for learning the Ukrainian language and introduction of means of emotional and meaningful analysis into it, which will provide additional means of overcoming the language barrier and analyzing user satisfaction.

1.1 Analysis of recent research and publications

1.1.1 Known methods of language learning

After analyzing the literary sources, we can conclude that there are many ways of learning a language, and everyone can choose the most convenient one, but there are three main approaches: structuralist (represents the language as a system of structurally interconnected elements with some hidden meaning), functional (presupposes language as a mechanism for performing various functions, for example, expressing one's opinion or requesting a certain service) and interactive (considers language as a means for creating and maintaining social connections, focusing around negotiations, activities, everyday communication, etc.). Among the most famous structuralist methods, we can single out [6]:

● Grammatical translation method. The main thing in this method is that listeners focus on learning the rules of grammar and increase their vocabulary by learning literal translations by heart. The main purpose of the application is to study classical languages. At the same time, one of the tasks of the method is that after completing the study, the students should be able to freely use language techniques such as spelling, grammar, and vocabulary, for comfortable reading, writing, and understanding of texts in various contexts. Learning grammatical rules increases the understanding that language can be represented as a system to be analyzed.

● Audio-language method. The purpose of this method is to develop language skills in a strictly defined order: listening-speaking-reading-writing. Learning consists of certain skills formation with the help of multiple, mechanical material repetitions. The task is the study of standard structures and the ability to apply them according to the situation. The result of language learning is monitored by using specially created test tasks. The advantage of this method is the development of situational techniques for the new grammatical material presentation and a specific set of tasks.

As for functional methods, the oral approach to situational language learning has become the most widespread among them. This method was developed by applied linguists from Britain, Harold E. Palmer and A. S. Hornby, who conducted significant research on language learning and found that the most attention should be paid to reading skills and defined such a concept as "vocabulary control". According to this, to learn a language, it is necessary to memorize a vocabulary of about 2,000 words that are most often found in written texts, and it was assumed that understanding these words allows you to create a basis for understanding texts in the process of reading. At the same time, the concept of

"grammatical control" appeared, which focused attention on grammatical constructions that are most often encountered in the process of conversational technique. After that, these constructions were added to dictionaries and guides for listeners. A significant difference between the direct method and the oral approach was that the methods formed based on this approach became theoretical foundations that indicate how to choose content, structure the complexity of tasks, and present materials. It is believed that this approach allows students to acquire certain skills that are repeated in certain situations: presentation (introduction of new material in context), practice (controlled practical phase), and production (classes are built on less controlled practice)) [7].

As for interactive methods, it is worth noting among them:

● The direct method, which is often called the natural method, consists of the listeners avoiding their native language and communicating only in the language being studied [8]. This method is based on the idea that learning another language should be similar to learning a native language: a child never pays attention to another language in order to learn the native language, and in the same way, the native language is not necessary when learning a foreign language. The method emphasizes correct pronunciation. According to it, listeners should focus on conversations and avoid text until they acquire speaking skills. Anything related to grammar and writing should be avoided, as it slows down the acquisition of speaking skills. Classes begin with the student learning simple words such as window, pen, table, etc. This allows you to motivate the listener because he thinks that he has mastered the language immediately. Eventually, the lessons progress to verb forms and other grammatical constructions, with the goal of learning about twenty new words in each lesson.

● The scenario method is a type of direct method when teaching is directly related to the language being studied. According to François Gohen, listeners learn the language faster when it is presented in chronological order. Namely, individual expressions are studied based on activities, in the order in which they occur, for example, leaving the house, going to the bus stop, getting on public transport. He noticed that if this sequence is broken, memorizing sentences becomes almost impossible. Another of his observations regarding memory was the concept of "incubation", which determines the time of "memorization", namely, the period of time during which linguistic ideas should be stored in the memory under the condition of their use in speech [9].

● The language "immersion" method is used in the process of teaching other disciplines in the language being studied. The term "immersion" in the context of learning foreign languages for a professional direction with the use of integrated special courses has two main definitions: first, it is a method of learning a foreign language by teaching one or more disciplines in this language. Secondly, it is a special type of integrated foreign language learning, the purpose of which is to master a foreign language for special purposes. Full immersion in a foreign language, and translation, in particular, occurs gradually. To achieve the effect of language immersion, the teacher carefully selects the educational material and arranges the tasks, and clearly plans the work at seminars and practical classes. The teacher should direct his activities guided by the following principles: the content of the analyzed authentic texts should include information that expands professional knowledge; tasks for professionally oriented texts are grouped according to the principles of gradual complexity; the nature of tasks changes from illustrative and reproductive to constructive and creative [10, 11].

The given list of methods used in learning foreign languages is not exhaustive. But the conducted analysis shows that an ideal method does not exist and is unlikely to exist [12]. Because each person, as an individual, is characterized by different skills that should be taken into account in the learning process. Therefore, it can be concluded that the specified mechanisms can be provided through an adaptive approach, which is planned to be implemented in the constructed system.

1.1.2. Peculiarities of emotional content analysis

Emotional content analysis is one of the areas of natural language processing, the purpose of which is to distinguish the thoughts and moods of people from written texts. Different types of analysis try to get different results: feelings, emotions, personal experience, and relation to a product or service. Features of emotional content analysis are the need to filter a significant amount of data; data analysis is performed in real-time; the application of the "sequence of criteria", namely the labeling of text

depending on emotions, is subjective, influenced by personal experiences, thoughts and beliefs. By using a centralized sentiment analysis system, companies can apply the same criteria to all their data, which helps to increase the accuracy of the results obtained [13].

Sentiment analysis is extremely important in many fields, from business to politics. Today, it is widely used both for processing reviews about brands, and for monitoring opinions in social networks about upcoming elections, attitudes towards certain events, or following certain processes or activities. The main areas of application of sentiment analysis are monitoring of social networks, monitoring of brand position, determination of the voice of the customer (Voice of the customer), and customer service [14]:

● Monitoring of social networks. With the proliferation of social media, people's opinions have become more accessible than ever. With the use of social networks people most often share their emotions about services, products and express their opinion about certain social problems. If earlier it was enough for brands to create a good advertisement and a convenient website with a few dozen reviews, thus forming a positive image of their company, now the quick processing of feedback from customers comes first. At the same time, a significant number of mentions and reviews do not always indicate the quality of the company's work. It is important to understand what exactly customers are saying. Using the analysis of emotions in social networks, you can find potential customers, monitor the work of competitors, and identify urgent problems before they get out of control.

● Brand monitoring. Brands receive feedback not only from social networks, but also from news sites, blogs, forums, product reviews, and more. Analysis of moods and opinions provides an opportunity to collect feedback and additionally obtain information about the geographical location, gender, and age of customers. This helps to correctly form the target audience and choose the correct development strategy. Analysis of these factors helps to understand how the image of the brand changes over time and to compare it with the image of competitors. It allows you to identify potential PR crises in real-time and take action before they become serious problems.

● Voice of the customer. Social media and brand monitoring provide general information about customer sentiment that needs to be further structured. At the same time, this information can be collected both in the process of extracting (parsing) data from resources and using the survey mechanism. Net Promoter Score (NPS) surveys are one of the most popular ways for companies to get feedback by asking a simple question: Would you recommend a company, product and/or service to a friend or family member? These surveys receive a single score on a scale of numbers. Numerical survey data are easy to summarize and evaluate, but they do not explain the reasons for choosing a particular score. Adding the following question, such as "Why did you leave such a rating?", will make it possible to clarify possible problems both in the company itself and in its external policy. These questions provide more information, but it is more difficult to process them manually, given that the task of their emotional and content analysis using automated means is urgent.

● Customer service. A high level of service is as important as the product itself. Customers expect that working with the company will be fast, intuitive, customer-oriented, and hassle-free. If the experience turns out negative, the customer will switch to a competitor. Sometimes a single negative experience is enough for a customer decide to change companies. Analyzing customer interactions allows you to ensure that employees follow the appropriate protocol, are loyal, and do not make mistakes during service.

1.1.3. Analysis of known systems

The main problem in learning any language is that there are many systems of analogs, which in most cases do not have the characteristics of complexity. Since the given task consists of two subtasks, the analysis of the existing solutions will be carried out in two directions: systems that specialize in learning the Ukrainian language and systems that are used for emotional and content analysis. Well-known systems for learning the Ukrainian language:

● A national platform for studying the Ukrainian language of the Ministry of Culture and Information Policy of Ukraine [15]. This resource contains online and offline resources for learning the Ukrainian language, both for citizens of Ukraine who want to improve their Ukrainian language skills and for foreigners who want to learn it. Among the disadvantages, it is possible to identify

significant bulkiness and excessive informativeness, and therefore the difficulty of choosing the necessary tool.

● The project "Language is the DNA of the Nation" [16]. An educational project for those who want to improve their knowledge of the Ukrainian language. Learning takes place in a game form. All the rules, exercises, and other things are presented by the fictional hero - Lepetun. It is depicted on each card with rules and tasks, which allows the associative memory to work and learn the language faster. The application has the following main functions: learning the basic rules of spelling; description of phraseological units and synonyms; correct accentuation of words; exercises for mastering the rules; instructions on how to get rid of Russianisms. Among the shortcomings can be identified: a small theoretical base; only spelling rules are present; a small number of exercises; there is no possibility to determine the level.

● R.I.D. mobile application [17] – the application was created to improve and deepen knowledge of the Ukrainian language and culture in the country and beyond. A feature of the application is the presence of a set of tasks and exercises that have certain restrictions on use. In general, the application is characterized by functional stability and a developed interface. The lack of learning progress and infrequent updates can be identified among the disadvantages.

● Language Course S.L. [18] - a lexical simulator for learning the Ukrainian language. The main emphasis of the system is related to the study of the spoken Ukrainian language for use in the process of travel, business, and education. The application contains a Ukrainian dictionary with flashcards and a translation of 10,000 words and is characterized by free use. Unstable operation and the presence of errors during operation can be identified among the disadvantages.

● Learn languages - Mondly [19] is an application that provides tools for learning about 30 languages, including Ukrainian. A feature of the system is the presence of daily lessons, the main emphasis of which is focused on memorizing words, constructing sentences, and practicing dialogues. The app uses a special learning technique that focuses on memorizing keywords to build sentences and phrases. Among the disadvantages can be defined as the presence of superficial language constructions and the commerciality of the application.

● Let's Remember And Learn Ukrainian [20] is a developing application that is a mobile tutor for the independent study of vocabulary and phonetics at the elementary level. The list of words is specially selected from various topics that are used in everyday life. This self-tutor allows you to effectively learn the correct pronunciation and writing, due to the presence of visual and audio support. Properly organized educational material will help you learn words quickly and easily, and the learning process itself is built in several stages: training (allows you to remember nouns, adjectives, verbs, and the alphabet); testing and verification of learned words are organized in the form of gamification methods: reading and association, visualization and spelling. Among the shortcomings, one can identify the commerciality of the application, the presence of advertising, and the lack of detailed results demonstrating the level of language knowledge.

Regarding the systems that are used for emotional content analysis, among the most popular, the following were highlighted:

● Awario [21] is a social network monitoring and analysis tool. It covers major social media networks, news, blogs, and forums. Built-in sentiment analysis sorts brand mentions into positive, negative, and neutral. Next, the program creates a graph that shows how mentions of the brand have changed over time and how sentiment has changed. In addition to emotions and sentiments, you can also see the most frequently used topics related to the company. Among the shortcomings can be identified: lack of support for the Ukrainian language and the commercial nature of the application.

● Talkwalker [22] is one of the leading American software products for analyzing social data, which helps companies and agencies collect and interpret information for reputation management, as well as find ideas for marketing and PR actions. The list of key features of the system includes team communications, channel management, content analysis, automated alerts, and intelligent media processing. Using this software, brand managers can monitor online forum discussions, receive alerts on social activities, access trend forecasting charts, and respond to comments in real-time. The disadvantages of this service are determined by the work in the English-speaking segment of the global network and the significant cost of use.

● Critical Mention [23] - differs from other options in that it analyzes news and other publications containing references to a specific company. With its help, you can quickly find out about all new

mentions and react in time. Since news production is now a 24/7 job, such software can help monitor feedback on online resources and social networks.

● Lexalytics [24] is a business analytics solution that analyzes different types of text. Lexalytics works with social media comments, polls, reviews, and other text content. In addition to emotional content analysis, the tool performs categorization, topic extraction, and intent detection, which can make it easier for companies to track extended context and understand business strengths and weaknesses.

● Social Searcher [25] is a social network monitoring platform that includes a sentiment analysis tool. To get started, you need to enter a keyword, hashtag or username, and Social Searcher will tell you what sentiments characterize the specified keyword. This platform forms reports in accordance with the source of communication, which allows you to get detailed information depending on the resource.

The peculiarity of these programs is the work with English-language content and the commerciality of the application. As for the Ukrainian language, today there is no industrial model that would carry out an emotional and content analysis of reviews, which is primarily due to the small base of corpora for conducting research. The main projects include:

● The corpus of the Ukrainian language GRAK is a representative, structured collection of texts in the Ukrainian language, accompanied by a program that allows you to build your own subcorpora based on the corpus, search for words, grammatical forms, and their combinations, as well as process search results, sort, make balanced samples and receive various statistical information. The corpus is intended for linguistic research on grammar, vocabulary, and history of the Ukrainian literary language, as well as for use when compiling dictionaries and grammar [26].

● The Lang-uk project [27] is an open community of people (software developers, linguists, and researchers) who are passionate about natural language processing and intelligent text processing. The Lang-uk community is focused on the development of new projects for the collection of Ukrainian corpora and other text resources. The Lang-uk project website features open-source corpora and dictionaries. Among them: vector representations of words (Word2Vec, LexVec, and GloVe 300d vectors), a lemmatized version of the vectors was built using the Ukrainian POS tag dictionary.

● Linguistic portal Mova.info [28]. A corpus of the Ukrainian language that provides means of searching by subcorpus and by morphological features. In this resource, frequency dictionaries for multiple sections and authors, grammar dictionaries, and dictionaries of comparisons have been created separately.

The conducted analysis showed that the available software is characterized by significant shortcomings, which makes the task of creating an information system for studying the Ukrainian language with the possibility of emotional and meaningful content analysis an urgent task.

1.2 The main tasks of the research and their significance

The purpose of the research is the development of an information system for learning the Ukrainian language and the implementation of emotional content analysis. The conducted research will provide means for creating on its basis software for managing informational and educational content, generating an individual learning environment for each user, and providing functions for emotional and meaningful analysis of user feedback. To achieve the goal, the following tasks must be solved: analyze the existing approaches, methods and software tools used in the field of learning and evaluating foreign language skills; determine the main tasks that arise at the same time; develop models of structuring and formalization of knowledge and information presented in the content of the educational environment, in order to provide a basis for the development and programmatic implementation of methods of individualized user access to the requested educational materials; to implement a mobile application for learning the Ukrainian language and providing tools for emotional content analysis.

The results of the study solve the actual scientific and practical task of activating the meaningful and motivational side of the educational process and will provide the means to open additional opportunities for individualization and differentiation of Ukrainian language learning and monitoring of emotional content analysis.

2. Major research results

In order to present the main aspects of the studied subject area, its conceptual scheme was built, which is shown in Fig. 1. As can be seen from the figure, the key tasks in the creation process are the construction of a database model, the development of methodological and algorithmic foundations of the subsystem for the construction of an individual learning and communication environment, the informatization of the learning management process, and the provision of mechanisms for emotional and meaningful content analysis. The user receives individualized access to the system's resources with the help of the subsystems of organizing the educational request, building an individual learning environment, the learning management subsystem, and the emotional content analysis subsystem.
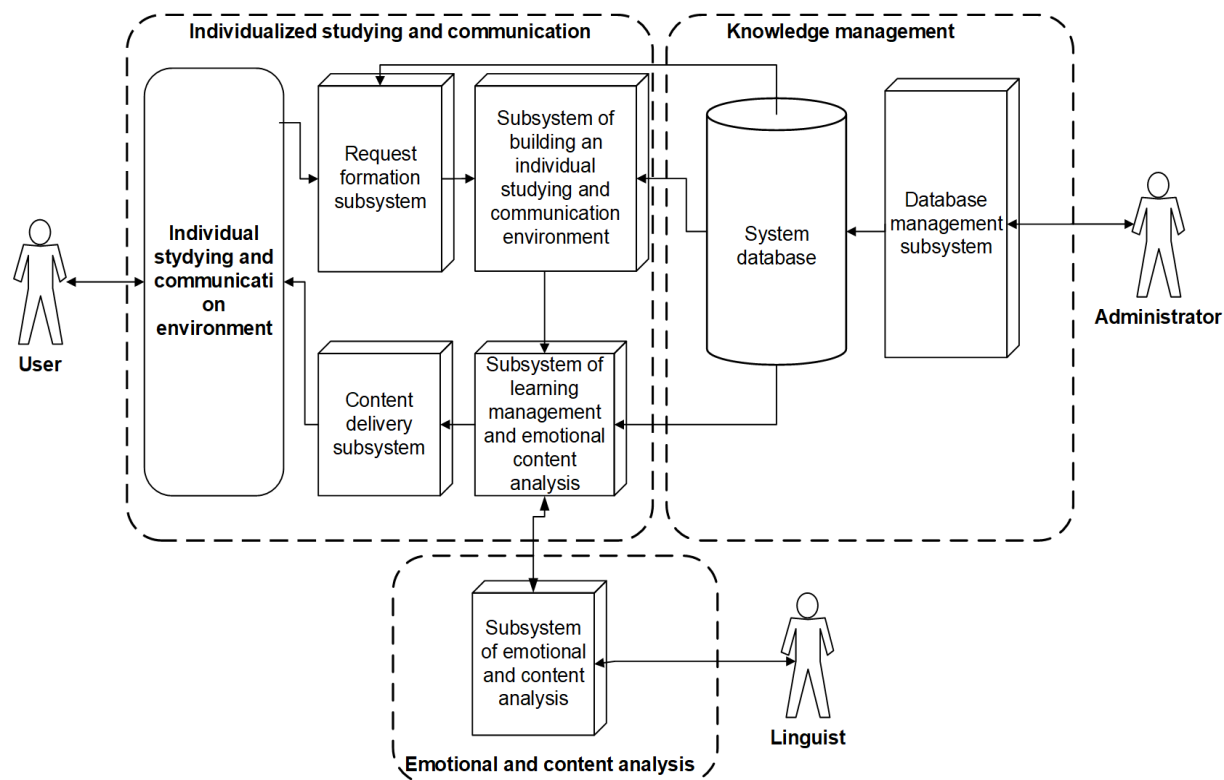
**Figure 1:** Conceptual diagram of the information system

Further work was aimed at conducting a systematic analysis of the subject area using the methodology of functional modeling and graphic description of processes. For these purposes, a structural approach and the IDEF0 standard, which is intended for the formalization and description of business processes, were used. A context diagram reflecting the process of forming an individual learning environment for learning the Ukrainian language is presented in Fig. 2.

In the specified model, information is received from the listener, which contains individual educational goals, interests, the purpose of using the system, etc., and knowledge of the subject area and educational materials. With the help of internal models, the educational system forms an individual learning environment. Support for individualization and adaptability of the educational process should be implemented on the basis of content intellectualization, which should be ensured at the stage of its creation [29]. In turn, intellectual content is the central entity for knowledge management in the context of the synthesis of an information system for learning the Ukrainian language and the implementation of emotional and substantive analysis in the process of communication.
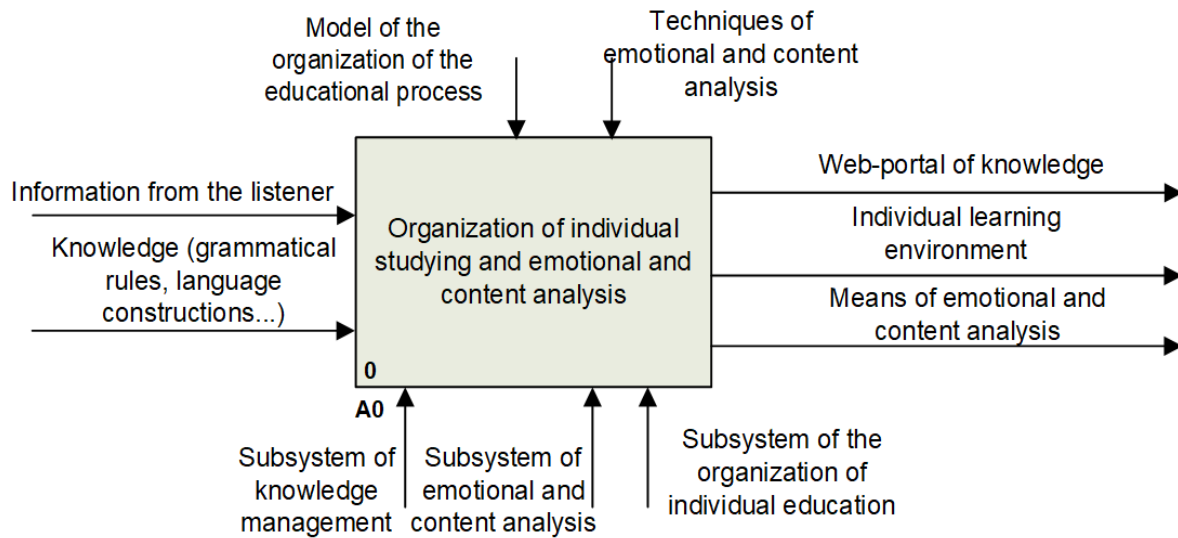
**Figure 2:** Functional model of the designed system

The created database of formalized educational content at the stage of knowledge management can be used as a knowledge portal, which presents informational and educational resources in a structured form for familiarization with the grammatical features of the Ukrainian language. On the other hand, such a database, thanks to the application of methods of generating an individual educational environment, serves as a basis for building relevant educational courses that meet the individual educational goals and interests of the student, thereby ensuring individualized access for users to the requested information [30]. A separate function of the system is emotional content analysis, which will be carried out using algorithms and language processing methods. In order to detail the described functionality, we will perform the process of decomposition of the created model (Fig. 2). The decomposition process was carried out taking into account the demarcation of knowledge management processes, the organization of individualized training, and the implementation of emotional and meaningful analysis. As a result, the functional model presented in Fig. 3 was obtained.
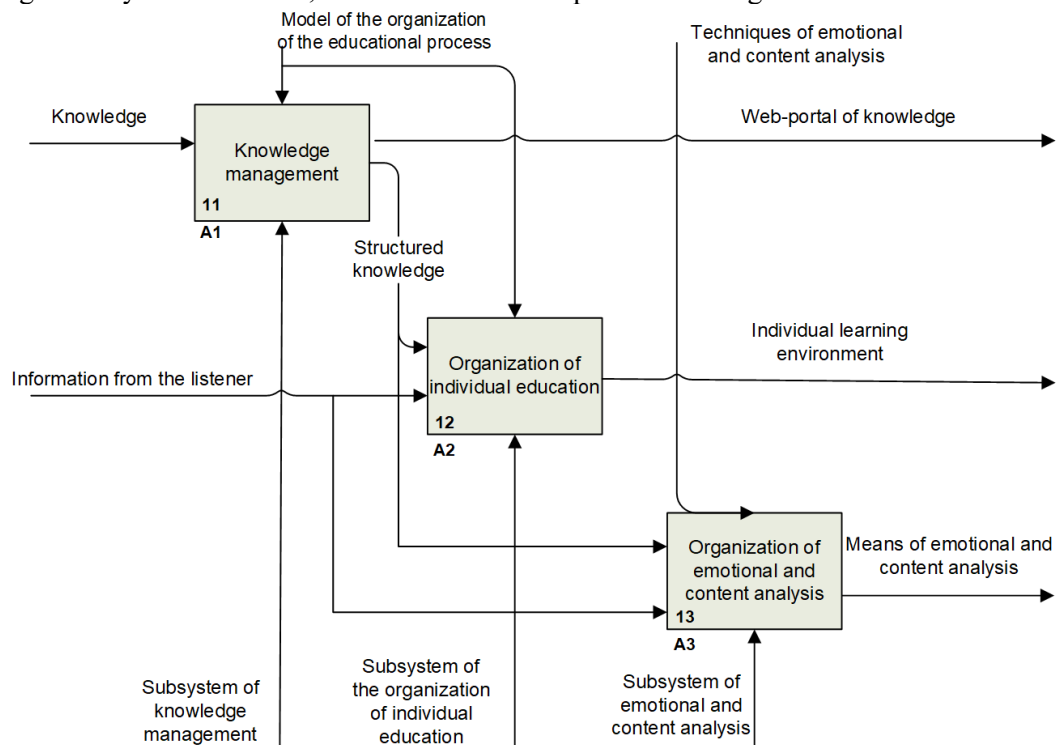
Figure 3: Functional model decomposition

To implement the system based on the presented functional model, it is proposed to use a set of models and methods as the basis of software tools for managing educational content. This complex of models contains such components as a complex content model, a model of educational content organization, a model of knowledge control and diagnostics, a model of educational inquiry, a model of the subsystem of generating an individual learning environment, and a model of conducting emotional content analysis. Thanks to this, the functioning of the learning environment is ensured according to the concept in which knowledge management plays the role of preparing a repository or knowledge portal, and the organization of learning is based on the methods of using this repository as a generator of an individual learning environment. In this way, the implementation of this approach is achieved by dividing the work of the system into three levels: the level of knowledge management; the level of organization of individual training; level of emotional and meaningful analysis. Knowledge management is aimed at forming a didactic-oriented knowledge base of the subject area in which training takes place. The organization of individual training is based on the use of formalized knowledge, which is obtained at the first level of working with the system, to build an individual learning environment. The organization of emotional content analysis will be based on natural language processing and will be implemented by a separate module that will provide tokenization and syntactic analysis, lemmatization/stemming, part-of-speech tagging, and identification of semantic relationships. After further decomposition of the presented model (Fig. 3), we will present a detailed functional model of the system (Fig. 4).
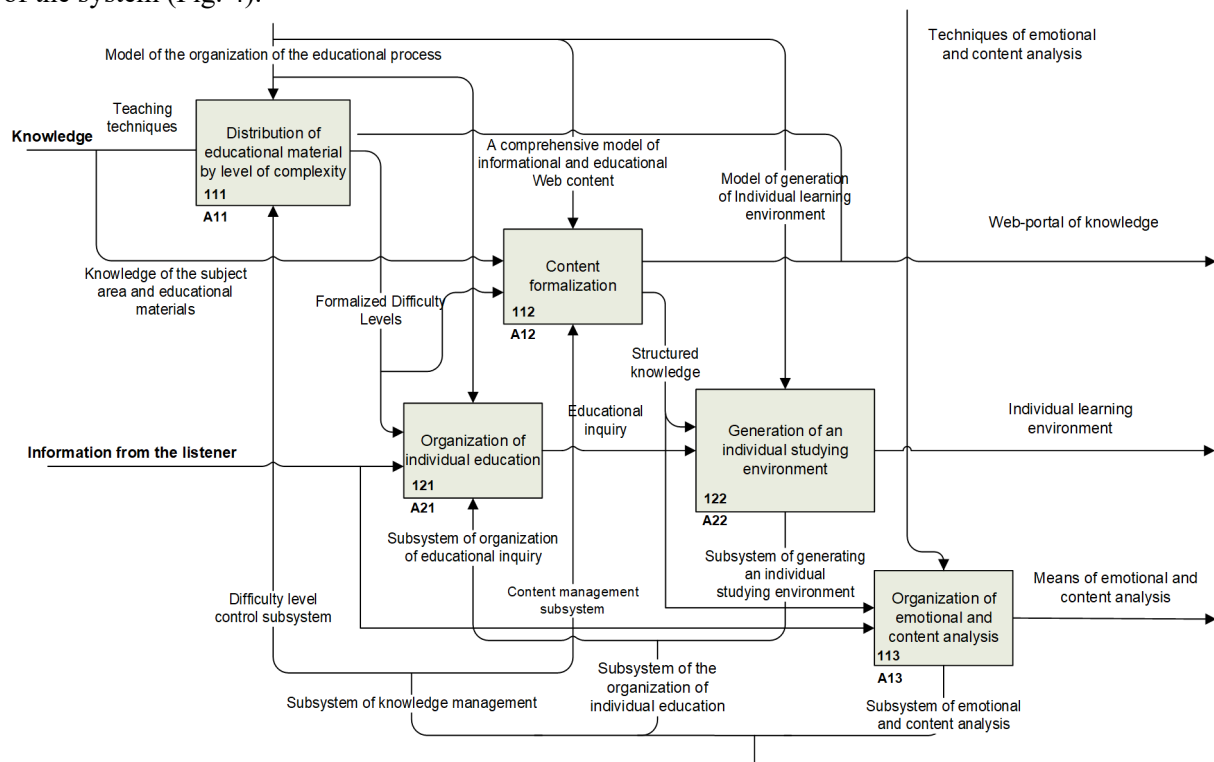
Figure 4: Detailed functional model

The level of knowledge management involves the performance of two key functions: the distribution of educational material by level of complexity and the formalization of content. This happens on the basis of the educational process organization model and the comprehensive educational Web-content model, respectively. In this way, knowledge of the subject area is formalized with the help of educational content and its semantic component [31].

The emotional content analysis level uses a set of created rules to determine the main context. These rules will include the use of emotion dictionaries, which define lists of words with positive or negative connotations. The operation of the subsystem will include the following stages:

- identifying lists of polarized words (negative words, such as bad, terrible, etc., and positive words, such as good, kind, etc.) and assigning numerical values to them;

- adding rules that have the greatest impact on the meaning of the sentence, such as the use of "not" and other negative constructions.
- splitting the input text into tokens and counting the number of positive and negative words in the source text.
- conducting statistics on user sentiments.

The process of implementing emotional content analysis can be divided into several stages: selection of the algorithm for the system, selection/creation of a dictionary of sentiments, creation of rules, collection, cleaning and storage of data, emotional content analysis, output and visualization of results.

Selection of the system operation algorithm. There are many variations on how a sentiment analysis system based on rules and sentiment dictionaries can work. The simplest and least effective systems use two sentiment dictionaries, positive and negative, to break the text into words and count the number of polar words in the text. Which words are more - such is the result. This approach has little accuracy because it does not take into account the order of words in the sentence and negative constructions. Another approach also uses dictionaries of sentiments, but they are built according to a different principle: each word is assigned a value within certain limits (most often from -1 to 1) [33]. Neutral words receive 0, words with a negative connotation receive negative values, and words with a positive connotation receive positive values. The value itself is chosen according to the intensity of the emotional coloring. In such systems, the polarity of words, and their order in the sentence, to which the rules apply are simultaneously analyzed (for example, "not" before a word changes its meaning to the opposite). The second option was used in the research process.

Selection/Creation of sentiment dictionary. There are several dictionaries of sentiments for the Ukrainian language, but most of them are not distinguished, but simply divided into positive and negative [33]. In addition, large dictionaries often contain words that have different meanings in different contexts. For example, the word "high" can be positive in the context of "high quality" and negative in the context of "high price". Such ambiguous words can negatively affect the results of the analysis. Therefore, in this study, it is proposed to create a prototype of a designated dictionary with words with a clearly expressed polarity, which are most often used when writing reviews about goods or services. The stages of creating such a dictionary are based on research [33]. To highlight the most frequently used words, 500 reviews from various sources were analyzed: Internet resources, and social networks. On the basis of this, 200 words were singled out, which determine meaningful coloring. Each word is tagged pos for positive words and neg for negative and is rated for polarity from -1 to 1. Words with a clearly defined positive color have the highest value, and strongly negative words have the lowest value. Since the evaluation of words is a rather subjective phenomenon, during the creation of the dictionary, an analysis of existing dictionaries for the English and Ukrainian languages was carried out to establish evaluation criteria, according to which words have the following values: 0.3, 0.5, 0.7 and 1 (with positive and from "captive signs"). The words were entered in the dictionary without endings, for this the stemming operation was used. Today, the literature describes many approaches that perform lexical analysis (Stemka, MyStem) or shorten words (Porter stemmer, Paice/Husk Stemmer), but in most cases, they do not have Ukrainian localization. In view of this, as a stemmer in this study, it is proposed to use the method described by T. Holub [34], which is based on a modification of Porter's algorithm [35] and does not require the use of a generated database, which reduces equipment requirements and the number of performed calculations [31].

Creating rules. As a result of the conducted research, a set of rules was formed, the description of which is given below. In the process of content analysis, one of the main aspects that significantly changes the meaning of a sentence in the Ukrainian language is the use of objections. The negation before the word must change the sign of the number to the opposite. However, it should be noted that "not bad" does not always equal "good". The polarity reverses, but the intensity decreases. Therefore, it is advisable not only to change the sign but also to reduce the intensity. Namely, if there is a negative particle in front of the word to be evaluated, the value of this word must be multiplied by a certain coefficient, which is set at the level of -0.7. Example: sent_analysis ('good'): 0.5, sent_analysis('not good'): -0.35. In the process of using reinforcing adverbs (very, extremely, incredibly, etc.), the intensity of the word that is located further increases. Therefore, if the word being analyzed is preceded by a strengthening word, its value must be multiplied by a certain coefficient, which is determined at the level of 1.5. Example: sent_ analysis('benign'): 0.5, sent_analysis('very benign'): 0.75. If adverbs that

have a strengthening meaning are in front of a negative construction, then they strengthen the entire construction. Example: sent_analysis('good'): 0.5, sent_analysis('very bad') : -0.525.

Often in a sentence, the reinforcing word separates "not" and an adjective, for example, "not very kind." The adverb "very" refers to the particle "not", so they must be analyzed together. "Not very good" has a higher degree of positive value than "not good", so if the word being analyzed is preceded by "not" together with a reinforcing word, the value of the word must be multiplied by a certain coefficient, which is set at the level of -0.5. The display of the structure of the rules is given in the form of a binary tree of decisions (Fig. 5).
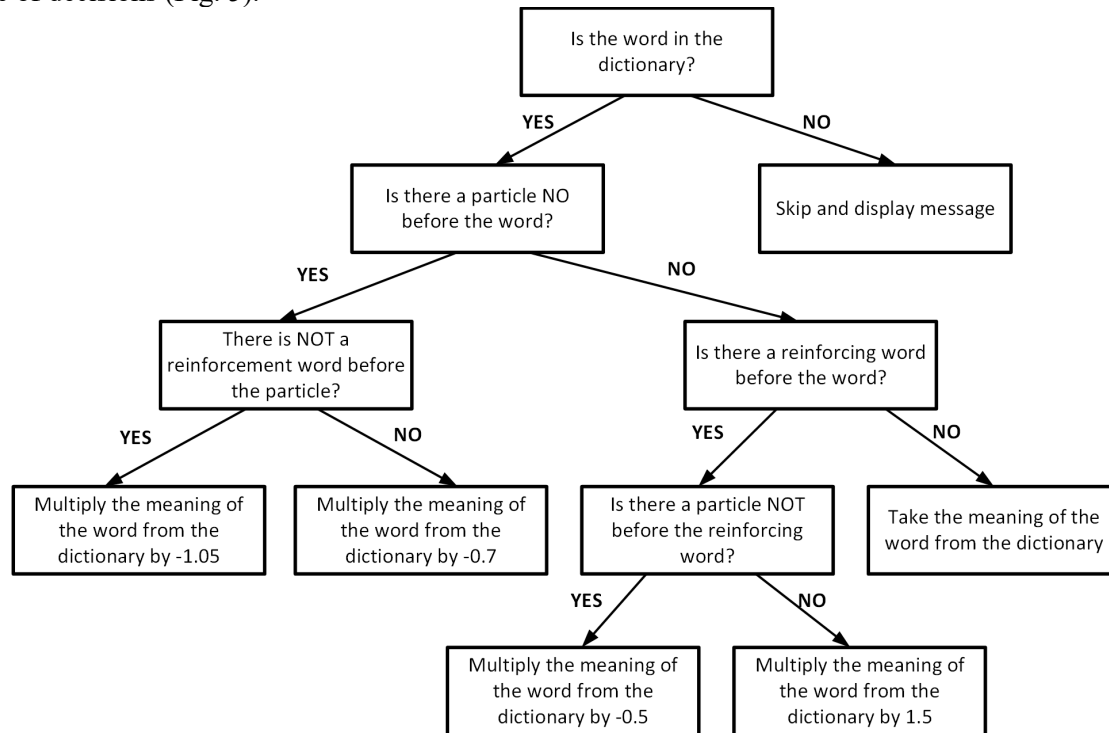
Figure 5: Decision tree

Collection, cleaning and data storage. The conducted analysis showed that the data must be obtained from specific Internet resources of websites and social networks. As for Internet resources, a number of software libraries are used for the automated uploading of reviews, among the most common is the Requests library. However, the format of the downloaded content, in most cases, contains a lot of markup elements, which requires a data cleaning operation, which should be implemented using the *BeautifulSoup library*. Cleaning or pre-processing of data, which is defined by a set of operations: lowercase text, removal of punctuation marks and semantically irrelevant text, which is often found in documents from HTML pages, and stemming operations. In this study, checking the language of the text and highlighting only Ukrainian-language texts is implemented using the *langdetect library*. Next, the described operations are performed and the result is entered into the storage.

After cleaning the data, an e*motional and content analysis* is carried out, for which a set of rules and a corresponding decision tree are used (Fig. 5). The result of the specified operation will be the formation of a python dictionary, the key of which is feedback, and the value is its emotional and content value expressed as a numerical value from -1 to 1.

Derivation and visualization of results. Text data can be graphically represented in various ways, depending on the purpose [36]. Pie charts will be used to display the results of the emotional content analysis, which will show the ratio of positive, neutral and negative reviews, histograms will show the distribution of the polarity of reviews during a certain time, and linear graphs will be used to compare the emotional coloring of reviews in different periods (comparison of data by months, years etc).

The level of organization of individualized training contains a sequence of two key functions: the organization of an educational request and the generation of an individual learning environment. Information from the listener (individual learning goals of the student, interests, language constructions) is received at the input of the educational request organization process. The generator of an individual

learning environment, manipulated with the help of a suitable model structured knowledge and educational request, outputs a set of individualized content, which, together with a set of learning support functions, represents an individual learning environment. From the administrator's point of view, he mainly performs mentoring functions, the linguist performs system training.

The next stage was the development of the system using modern software tools. The developed system consists of two parts - a mobile application and a web platform for organizing the educational process and creating a comprehensive model of educational content. The mobile application was created using the Flutter framework [37], which is a set of open-source software created by Google Corporation. Flutter applications are written in the Dart language [38] and run in a virtual machine during writing and debugging, providing fast compile times and dynamic reloading without the need for a restart. The graphical interface is implemented using Flutter tools using widget mechanisms. In addition to the mobile application, a web platform has also been developed, the client and server parts of which are written using frameworks and libraries of the Python language [39]. To create and manage the database of the information system, the MySQL database management system was used in combination with the MySQL Workbench tool [40].

The developed system is characterized by an intuitive interface and described functionality. When entering the system from the mobile application, the user is offered to choose the functionality that needs to be obtained: learning the Ukrainian language or emotional content analysis. If you choose to study the Ukrainian language, the main window with the available main menu is displayed (Fig. 6).
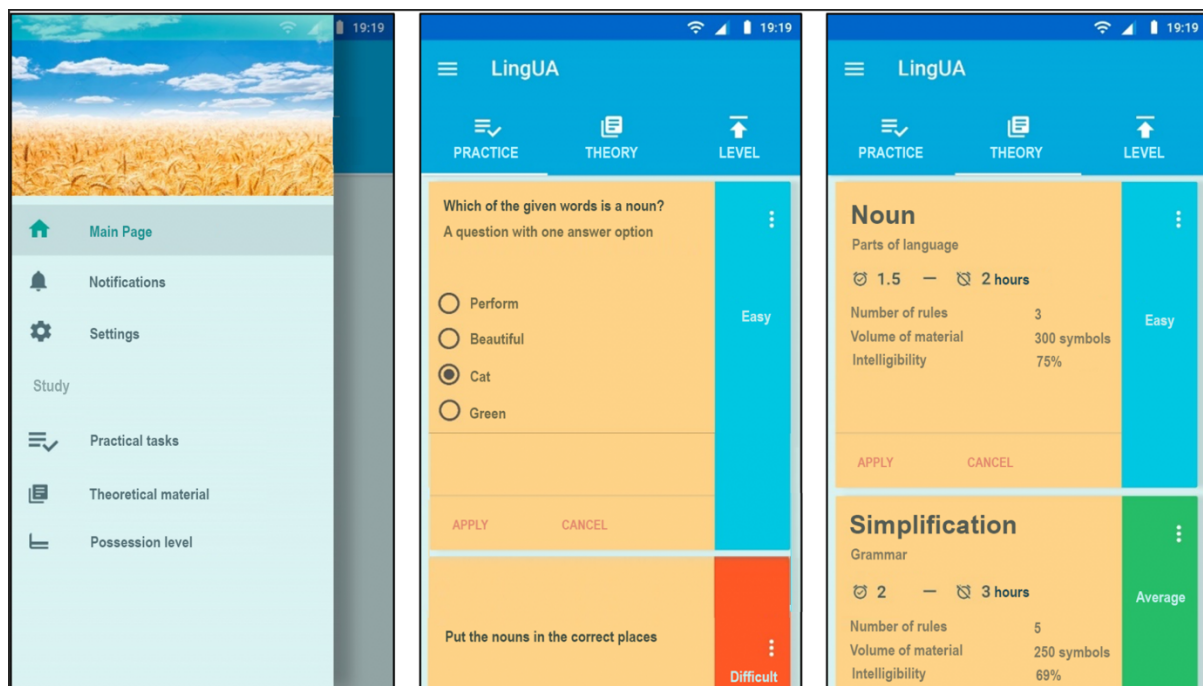
Figure 6: Screenshots of system windows: main menu, nesting of practical and theoretical classes

This page shows a pull-down menu that allows you to go to settings and notifications, or use the main functions, that is, open practical tasks, theory or level determination. After exiting this menu, the main interaction interface with three attachments is displayed. The first attachment contains practical tasks. The system forms them based on the passed theoretical materials, and they correspond to the level of the performer. The following attachment contains theoretical material.

If you select the appropriate tab, in particular "Noun", a new card will open with information about the theoretical material according to the topic (Fig. 7). There is also a tab with information about the user's level, namely: how much material remains to be completed to reach the next level, how many practical tasks have been completed at this level, and how much theoretical material has been completed. In the settings, you can learn more about all the levels that can be achieved.
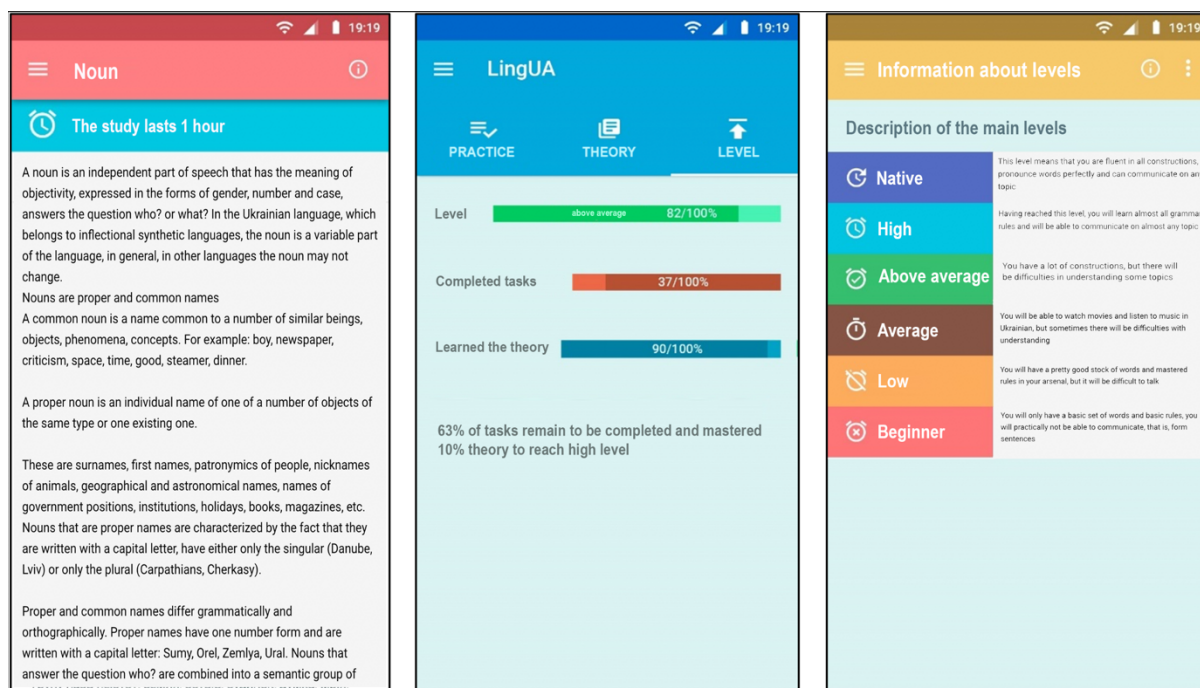
Figure 7: Windows that show system functionality

A feature of the system is the presence of various types of interactive tasks (Fig. 8), with the help of which you can comprehensively improve skills in mastering the Ukrainian language. The main categories, in the form of a list of cards, in the application are "Categories of words", "Learn words" and "Pronunciation of words" (Fig. 8)
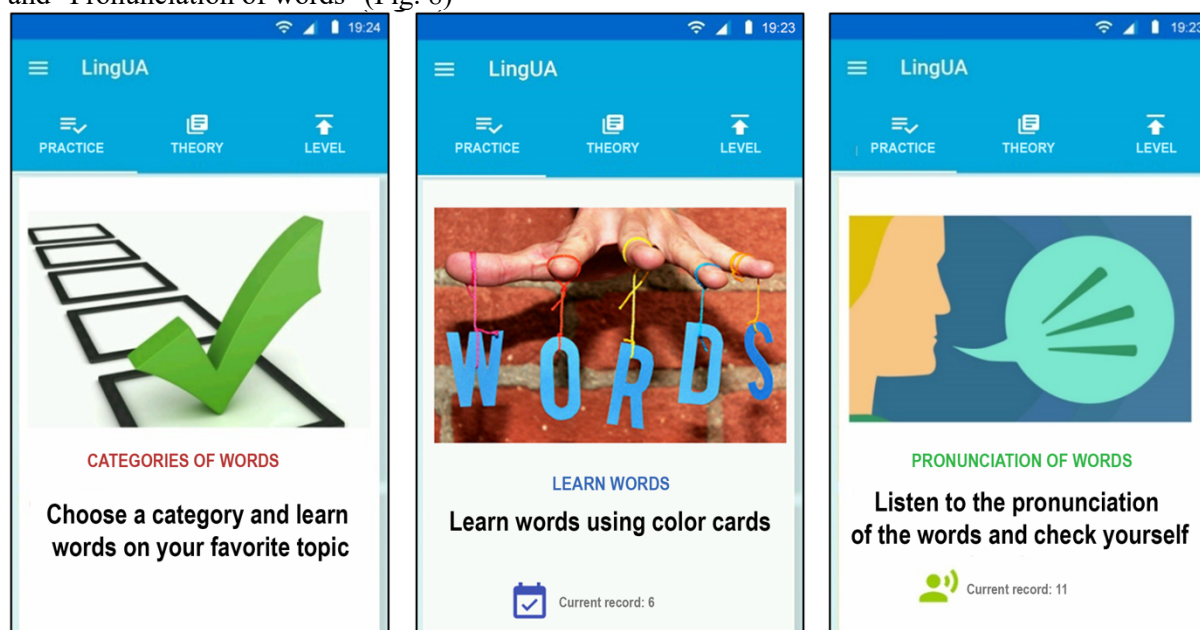

Figure 8: Windows that show advanced system functionality

As for emotional content analysis, when working with the system, it is necessary to enter the URL address of the corresponding resource containing reviews of the product, service or brand, and check the possibility of accessing it. If everything is fine, the system downloads/cleans them and analyzes them according to the described work algorithm. The result of the work can be presented in the form of a pie chart, a histogram and a line graph. As an example, we will present the results of the emotional content analysis of 50 reviews of product_X from resource_Y during two periods (September-June) (Fig. 9).
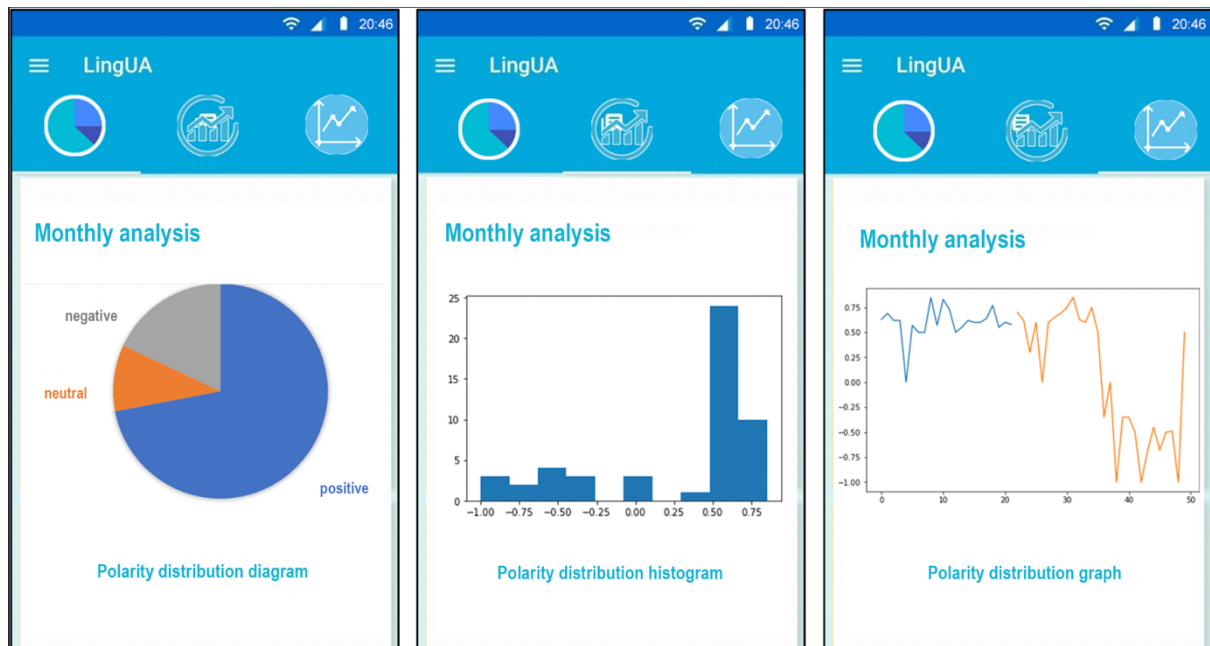
Figure 9: The result of emotional content analysis. From a privacy point of view, the resource and product name is not specified.

The information generated on the basis of the download is stored in the database and has a structure: the serial number of the feedback, the field of the cleared text of the feedback, the numerical value of the analysis, and the data of writing the feedback. On the polarity distribution diagram, we can see the ratio between positive, negative and neutral feedback. The graph of polarity distribution needs special attention, which shows that in the second half of October, the polarity of reviews changed sharply from positive to negative, which may indicate the existing products or an open conflict between the company and its customers. In any case, such a situation requires attention from the company.

3. Conclusion

As a result of the conducted research, the existing methods and known systems that provide means of learning the Ukrainian language and describe the mechanisms for evaluating the specified skills were analyzed. The technologies and software tools for conducting emotional content analysis were analyzed, which made it possible to determine the features of existing approaches. As the analysis showed, today there are many software systems, but all of them are characterized by certain shortcomings, from the commerciality of the application to limited functionality, which makes the task of developing an information system for learning the Ukrainian language and meaningful content analysis urgent. Models of structuring and formalization of knowledge and information presented in the context of the educational environment have been developed, providing the basis for the development and software implementation of methods for individualized user access to the requested educational materials. The next stage was the design of the software system using a structural approach and displaying the created diagrams in accordance with the IDEF0 standard. The study presents a functional model and its decomposition, which created a basis for understanding the peculiarities of the functioning of the Ukrainian language learning system and the implementation of emotional and meaningful analysis. An applied software system has been developed, which provides methodological and algorithmic foundations for building an individual learning and communication environment, provides informatization of the learning management process, and conducts meaningful content analysis. During the development of the software product, the modular programming principle was used. This structure of the system allows modifying its individual parts in the future, without impairing performance and loss of functionality. The constructed prototype of the software system can be useful as an additional tool for anyone who is interested in learning the Ukrainian language and emotional content analysis.

Further research will be aimed at testing and improving systems, eliminating conflicts and expanding functionality in accordance with defined requirements.

4. References

[1] R. Sushko, The voice and sounds of the native language, Apriori 2020, (In Ukrainian)
[2] Wordtips - The 100 Most-Spoken Languages in the World, URL. https://word.tips/100-most-spoken-languages/
[3] S. Dixon, 100 Ways to Teach Language Online: Powerful Tools for the Online and Flipped Classroom Language Teacher, Wayzgoose Press, 2020.
[4] S. Kosslyn, Active Learning Online: Five Principles that Make Online Courses Come Alive, Alinea Learning, 2021.
[5] K. Ramandeep, Sentiment Analysis - From Theory to Practice, LAP LAMBERT Academic Publication, 2017
[6] C. Chapelle, The Handbook of Technology and Second Language Teaching and Learning (Blackwell Handbooks in Linguistics), Wiley-Blackwell, 1st edition, 2017.
[7] J. Algeo, British or American English?: A Handbook of Word and Grammar Patterns, Cambridge University Press, 2006.
[8] E. Spooner, Interactive Student Centered Learning, Rowman & Littlefield, 2015.
[9] K. Conrad, The Language Teaching Controversy. Rowley, Massachusetts: Newbury House, 1978.
[10] O. Kanishcheva, V. Vysotska, L. Chyrun, A. Gozhyj, Method of Integration and Content Management of the Information Resources Network. In: Advances in Intelligent Systems and Computing, 689, Springer, 2018, pp. 204-216.
[11] T. McConachy, I. Golubeva, M. Wagner, Intercultural Learning in Language Education and Beyond: Evolving Concepts, Perspectives and Practices: 38 (Languages for Intercultural Communication and Education), Multilingual Matters, 2022.
[12] A. Vasyliuk, T. Basyuk, Construction Features of the Industrial Environment Control System, Proceedings of the 5rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2021). Volume I: Main Conference, Kharkiv, Ukraine, April 22-23, 2021, Vol-2870: pp.1011-1025.
[13] W. Pedrycz, S. Chen, Machine Learning Sentiment Analysis and Ontology Engineering, Springer International Publishing, 2016.
[14] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2020.
[15] National platform for studying the Ukrainian language of the Ministry of Culture and Information Policy of Ukraine, URL. https://speakukraine.net/
[16] "Language - DNA of the Nation", URL. https://ukr-mova.in.ua/
[17] R.I.D. it's time to become familiar with the language, URL. http://rid.ck.ua/
[18] Language Course S.L., URL. https://www.languagecourse.net/
[19] Learn languages – Mondly, URL. https://www.mondly.com/
[20] Let's Remember And Learn Ukrainian, URL. http://www.domosoft.biz/ua/gnlu.html
[21] Awario, URL. https://awario.com/
[22] Talkwalker, URL. https://www.talkwalker.com/
[23] Critical Mention, URL. https://www.criticalmention.com/
[24] Lexalytics, URL. https://www.lexalytics.com/
[25] Social Searcher, URL. https://www.social-searcher.com/
[26] M. Shvedova, R. von Waldenfels, S. Yarygin, A. Rysin, V. Starko, T. Nikolayenko, General regionally annotated corpus of the Ukrainian language (GRAC), Jena, 2017.
[27] Project Lang-uk, URL. https://www.lang.org.ua/uk/
[28] Linguistic portal Mova.info, URL. http://www.mova.info/
[29] D. Nayab, English Teachers' Attitudes in acquiring Grammatical competence: By using Grammar Translation Method and Communicative Language Teaching at Graduate Level (In The Context of Punjab), LAMBERT Academic Publishing, 2020.

[30] O. Naum, L. Chyrun, O. Kanishcheva, V. Vysotska, Intellectual System Design for Content Formation. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 2017, pp. 131-138.

[31] T. Basyuk, A. Vasyliuk, V. Lytvyn, Mathematical Model of Semantic Search and Search Optimization, Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2019). Volume I: Main Conference, Kharkiv, Ukraine, April 18-19, 2019, Vol-2362: pp.96-105.

[32] L. Manika, M. Margam, Application of sentiment analysis in libraries to provide temporal information service: a case study on various facets of productivity. Social Network Analysis and Mining. 8 (1), 2018, pp.1–12.

[33] A. Romanyuk, M. Romanishyn, Tonal dictionary of the Ukrainian language based on the sentiment-annotated corpus, URL. http://nbuv.gov.ua/UJRN/Um_2013_43_10

[34] T. Golub, Yu. Tyagunova, The method of Ukrainian language stitemming for the classification of documents based on Porter's algorithm. In: Scientific works of the Donetsk National Technical University, vol. 1, 2017, pp. 59-63 (In Ukrainian).

[35] M. Porter, An algorithm for suffix stripping Program. In: Data technologies and application, vol. 40(3), 2006, pp. 211-218.

[36] T. Basyuk, A. Vasyliuk, Approach to a Subject Area Ontology Visualization System Creating, Proceedings of the 5rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2021). Volume I: Main Conference, Kharkiv, Ukraine, April 22-23, 2021, Vol-2870, pp. 528–540.

[37] F. Zammetti, Practical Flutter: Improve Your Mobile Development, Apress L. P., 2019.

[38] S. Ladd, K. Walrath, Dart: Up and Running: a New, Tool-Friendly Language for Structured Web Apps, O'reilly, Incorporated, 2012.

[39] W. Mckinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, 2nd edition, 2017.

[40] D. Nichter, Efficient MySQL Performance: Best Practices and Techniques, O'Reilly Media, 2022.