ABSTRACT
of the dissertation work Smaiyl Assel Maralbaikyzy
on the topic: «Models and methods of intellectual information and training system», submitted for the degree of doctor of philosophy (PhD) in the specialty 6D070300 – Information systems

Relevance of the research topic: Development of an educational resource that allows to organize an effective learning process, taking into account the semantic representation of resources within the Smart University.

The experience of many years of work of universities and postgraduate education centers shows that in the system of continuing professional education, individual training is the main requirement, and the main limitation in solving the problems of individual education is time. Another important aspect of this issue is the content of the declared training programs, which do not take into account the individual needs of future specialists. The learning process is the same for each student; the material for each course is also the same. For a strong user, the content may be too simple and insufficient and, therefore, ineffective, and for a weak user, it may be complex and incomprehensible. Thus, the search for alternative ways to individualize learning remains an urgent problem. Qualitatively new opportunities for self-learning and improving professional knowledge are provided by new information technologies of distance learning using local and global networks. Unlike the traditionally constructed courses of full-time and even more so correspondence training, the use of information technologies opens the way for training directly at the workplace, which, if properly organized, allows you to individualize the process and allocate the necessary amount of time for training users without any noticeable interruptions in work. The concept of computer-based learning is based on the principles of autonomy (self-management) of the process. Its implementation involves a new combination of training and control programs with a developed component of mutual moral responsibility of teachers and users. Autonomy in the educational process implies not only the independence of the educational institution, but also the right of the student to choose an individual path of study within a multi-level education system.

Aim of research. System development using models and methods of an intellectual information system based on semantic data representation, taking into account the features of the trajectory and individualization of the learning process.

Main goals. To achieve the research goal, the following tasks were set and solved:

1) Review and analysis of existing information systems for educational resources;
2) Models and methods for building an educational resource;
   2.1) Formation of the content of the educational resource;
   2.2) Compilation of content based on semantic analysis;
   2.3) Determining the sequence of content;
   2.4) Management of the number of training elements in accordance with Microlearning;

3) Development of the system in the form of a web portal;
4) Software testing;
5) Software implementation at Astana IT University.

The main provisions of the dissertation submitted for defense:

1) The methodology for constructing an educational resource is based on semantic analysis;
2) Establishing the coefficient of semantic proximity of a pair of topics based on the proposed hypothesis;
3) The location of expressions in the content (on slides, texts) is determined by the microlearning method;
4) The result of an experimental study of the developed system, showing an increase in the quality of education and efficiency due to the semantic representation of resources, also creates conditions for the digitalization of the educational process.

Scientific novelty of the work:

The use of models and methods of content formation, which are designed for any student, take into account the user's characteristics, is scientifically justified.

For the first time, methods of semantic analysis were used to create learning content.

A hypothesis is put forward using methods for determining the semantic proximity of words using a contextual set, subsequently, the scientific hypothesis was proved using the Pearson correlation method to determine the sequence of content, and the rules for distance learning using ontology are also given in the scientific work.

It is proved that the use of the developed system improves the quality of education and shows efficiency.

Practical significance of the work:

The theoretical and applied results obtained in this work can be used in the design and development of a system for information and training organizations. And also, the obtained scientific results can be developed as a theoretical basis and an applied basis for automating the learning process by taking into account the individual characteristics of each student (intellectualization).

The developed system with the use of models and methods of the intelligent information and training system is implemented in the university, "Astana IT University" LLP, which is currently actively used in training.

Research methods:

Semantic analysis (semantic proximity of a pair of topics), a level decomposition algorithm, local reference methods such as the Dice measure, the Jaccard index and the Kulchinsky measure, as well as the Pearson correlation method, Microlearning methods, a nonparametric method of statistical analysis using the Mann-Whitney criterion.

Approbation of research results and publications

The main provisions and scientific results of the work were reported and discussed at seminars of the Department "Information Systems" of the International University of Information Technologies and at international scientific

and practical conferences: V All-Russian Congress of Young Scientists at the National Research University ITMO (Russia, St. Petersburg, 2016); International Scientific and Practical Conference "INNOVATIONS IN EDUCATION AND SCIENCE" dedicated to the 25th anniversary of Independence of the Republic of Kazakhstan and the 20th anniversary of the Suleiman Demirel University (Kazakhstan, Kaskelen, 2016); International scientific and practical Conference "The Path to Independence through Decades" dedicated to the 25th anniversary of Kazakhstan's Independence (Kazakhstan, Almaty, IITC, 2016); The 2nd International Conference "Information Technologies in Science and Industry 2016" (Kazakhstan, Almaty, IITC, 2016); International Scientific and Practical Conference "5th International Conference Science and Society – Methods and problems of practical application" (Canada, Vancouver, 2018); in the scientific and practical ONLINE conference "Digital Kazakhstan-the introduction of the mechanism of IT technologies in Technical and Vocational Education: features of the use of modular and credit learning technologies in the educational process" organized by the Republican Educational and Methodological Association (Kazakhstan, Almaty, 2018); "3rd International Conference on System Reliability and Safety (ICSRS 2018)" (Spain, Barcelona, 2018); 5th International Conference on "Digital technologies in science and industry-2019" (Kazakhstan, Almaty, 2019); 2021 IEEE Smart Information Systems and Technologies (SIST) (Kazakhstan, Nur-Sultan, 2021) and Dortmund International Research Conference 2021 (IRC 2021).

Structure and scope of work

The structure of the work is determined by the set goal and the sequence of solving the formulated tasks and is built on the problem-thematic principle. The dissertation consists of an introduction, four chapters, a conclusion and a list of references, including 77 names of the sources used.

Publication of research results

The main results obtained during the dissertation work are published in 12 printed works, of which 4 articles are published in publications recommended by the Committee for Control in the Field of Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan, 1 article is published in a publication indexed by the Scopus and Web of Science Core Collection databases, with an impact factor of 1.635, quartile for computer science and information system-Q3, percentile for computer science applications – 51, 4 articles are published in the collections of international foreign conferences (Canada, Spain, Ukraine, Germany), 6 articles are published in the collections of international scientific and practical conferences (Kazakhstan) and 1 copyright certificate of entering information into the state register of rights to objects protected by copyright, No. 19399 dated July 14, 2021 for the computer program "Intelligent information system using models and methods of semantic data representation".

The introduction provides a rationale for the relevance of the chosen topic of the dissertation. The purpose, object, subject, and tasks of the dissertation are formulated. The results of conducted research are described, their scientific novelty

and practical significance are shown. Data on approbation of the main results of the dissertation work are given.

In the first section of the dissertation examines the importance and role of an intelligent information system for educational institutions. A review and analysis of modern approaches to the automation of educational processes is carried out, as well as a literature review. The features of the content of the discipline, as well as the sequence and number of educational elements are considered.

Information technologies occupy a central place in the process of intellectualization of society, the development of its educational system and culture. In addition, the use of training information tools has proved to be a very effective method for both self-education systems and systems of advanced training and retraining of personnel.

The processes of informatization of modern society and the closely related processes of informatization of all forms of educational activity are characterized by the processes of improvement and mass dissemination of modern information and communication technologies (ICT). Such technologies are actively used to transmit information and ensure interaction between the teacher and the student in modern systems of open and distance education. A modern teacher should not only have knowledge in the field of ICT, but also be a specialist in their application in their professional activities. The concept of technology includes the application of scientific and engineering knowledge to solve a practical problem. Then information technology can be considered the process of turning knowledge into an information resource. The purpose of information technology is the production of information for its subsequent analysis and decision-making on the basis of it to perform an action. Information and communication technologies (ICT) is a general concept that describes various devices, mechanisms, methods, and algorithms for processing information. With the help of network means of ICT, it becomes possible to have wide access to educational and methodological and scientific information, to organize operational consulting assistance, to model research activities, to conduct virtual training sessions (seminars, lectures) in real time.

Analyzing the social significance of information technologies, it seems appropriate to state the following:

- Modern information and telecommunications technologies make it possible to activate and effectively use the information resources of society, which are the most important strategic factor in its development.
- The development of civilization is taking place in the direction of the information society, in which the objects and results of the work of the majority of the employed population are no longer material values, but mainly information and scientific knowledge. At the same time, information technologies make it possible to optimize and, in many cases, automate the information processes taking place in society.
- Information processes are important elements of other more complex industrial or social processes. In this regard, information technologies are considered as components of the corresponding production or public technologies.

- Telecommunications technologies, being a part of information technologies, play an extremely important role in ensuring information interaction between people and organizations, as well as in the systems of preparation and dissemination of mass information.

7 higher educational institutions of Kazakhstan were selected, such as: Astana IT University, International University of Information Technologies, Narkhoz University, AlmaU, UIB, Al-Farabi Kazakh National University and Astana Medical University. Since March 16, 2020, all universities have switched completely to distance education in connection with the pandemic worldwide. All of the above universities use the "Moodle" SDO, as there are advantages of the system: a modular web-oriented structure, support for multimedia data, it is possible to work with courses created according to international standards SCORM and IMS, as well as an open source system.

The analysis of distance education revealed the following advantages: flexible schedule of the educational process, as well as the availability of materials and mobility. Disadvantages: self-discipline, responsibility, tolerance, commitment and perseverance.

Adaptive intelligent systems assume: flexibility of learning in an interactive educational environment, personalization and adaptation of learning, free access to content regardless of geography. Such technologies make it possible to develop educational and methodological materials, as well as to form individual learning trajectories.

A survey was conducted among 94 respondents in russian language about the use of an adaptive intelligent learning system. Questions were devoted to distance learning. Since the adaptive system in the first place can be used in distance learning. In one of the questions it was necessary to indicate from which city user filling out a questionnaire, or user would like to study. The results of the questionnaire showed that respondents want to learn using an adaptive intellectual system and how relevant it is now. Based on the results of the questionnaire, it can be concluded that students do not want to spend time on subjects that they already know, they also want an individual learning trajectory. The system must be adaptive intelligent for users.

The review and analysis showed that the problem with the content sequence exists and has not been solved, in this regard, this paper examines the features of the relationship of differences between matrials, determines the content sequence based on semantic data analysis, and also implements a system with an individual learning process.

In traditional training, there are shortcomings in the formalization of processes, since the training is the same for all students. The content is the same, and the types of perception of users are different, it is also possible that the user knows the entire course, but he must teach it. The traditional training system is not flexible, not individual for each user. The existing methods of formalization of training are weak, and it is desirable to increase the effectiveness of training by individualizing training. The development of the system using methods and models will increase the effectiveness of training. The section also mentions a multi-

criteria learning model, which also gives good learning efficiency. The literature review showed that there is a trend of individual training in the world. Many scientists have developed adaptive systems for users. An adaptive intelligent system will help all users to effectively spend time on training, while gaining the necessary knowledge.

The second section of the dissertation, the methodology of creating an intelligent information and training system is considered. In particular, the content of the learning process should depend on the user's perceptual modality. And semantic content analysis with the help of ontology will provide an individual trajectory for the student.

In connection with the informatization of education, ideas of adaptive learning were also used in computer training. The main requirement that the information training system should be designed to take into account the principles of adaptive learning is to ensure the learning process (both within the university and in distance education) in accordance with the individual characteristics of the learner. This task can be solved by implementing in the training system various techniques and methods related to different variants of the training system functionality and various ways of its implementation.

Psychophysiological abilities by type of thinking are divided into visuals, audials and kinesthetics. Diagnostics of the dominant perceptual modality of S.Efremtsev is a technique of the leading channel of perception. Visitors perceive information through images, memory, imagination. Visuals have increased gesticulation in accordance with Figure 2.1. Audials perceive information through hearing in accordance with Figure 2.2. It is possible to identify audials by asking a simple question: what did he do yesterday? If he looks to the side and to the left, he remembers truthful information; if he looks to the side and to the right, he comes up. Kinesthetic perceive information through feelings, emotions. It should be noted that a person completely belonging to one type of thinking does not exist. In each of us there is a piece from each of them, but at the same time we have a dominant type of perception. According to statistics, in 35% of people the dominant type of perception is visual, in 25% of people the dominant type of perception is audial, and in 40% of people the dominant type of perception is kinesthetics.

For visuals, techniques for visualizing content: diagrams, graphs, illustrations, demonstration models, experiments and experiments.

For audials: Lecture material, audio courses, video, audio books.

In the course “Database in Information Systems” at the International University of Information Technologies, a test was conducted to determine the type of thinking for 3 groups of students according to the method of the leading perception channel.

Intellectual system for education can help users to learn effectively. The information system assumes:

- flexibility of learning in an interactive educational environment;
- personalization and adaptation of training;
- diverse content by type of perception;
- free access to content regardless of geography.

It is the intellectual system that allows the development of revolutionary teaching materials, as well as the formation of individual learning paths.

The adaptability of the system is that each user will have their own learning path. When entering the system, users pass 2 tests that will determine the type of perception and level of knowledge for a specific course.

Intellectual information systems are a natural result of the development of conventional information systems. They have concentrated the most high-tech technologies with a high level of automation not only of the processes of preparing information for decision-making, but also of the processes of developing solutions based on the data obtained by the information system.

The development of the ontology includes:

1) defining classes in ontology;

2) the arrangement of classes in a taxonomic hierarchy (subclass-superclass);

3) definition of slots and description of the allowed values of these slots;

4) filling in the values of the instance slots.

The following 16 classes with attributes: «departments» class has id, name, created_at, updated_at; «users» class has id, name, email, password, remember_token, created_at, updated_at; «students» has id, character_type, user_id, created_at, updated_at; «teachers» has id, name, user_id, department_id, created_at, updated_at; «subjects» has id, name, department_id, created_at, updated_at; «teacher_subjects» has id, teacher_id, subject_id, created_at, updated_at; «quizzes» has id, title, teacher_id, subject_id, isPsychological, created_at, updated_at; «question_types» has id, name, created_at, updated_at; «student_quizzes» has id, student_id, quiz_id, accepted, result, created_at, updated_at; «questions» has id, title, question_value, quiz_id, question_type_id, created_at, updated_at; «themes» has id, name, subject_id, created_at, updated_at; «materials» has id, title, created_at, updated_at; «answers» has id, right, content, question_id, created_at, updated_at; «question_themes» has id, question_id, theme_id, created_at, updated_at; «theme_materials» has id, material_id, theme_id, character_type, created_at, updated_at and «student_quiz_results» has id, question_id, answer_id**.**

The second section discusses the methodology for creating content using ontology. The development of the ontology begins with the definition of classes, then need to distribute the classes in the taxonomic hierarchy, then the definition of slots and the description of the acceptable values of these slots and at the end filling in the values of the instance slots. Thus, it is possible to create ontology of content for educational institutions. The distribution of content by the type of user perception is also determined. In education, take into account 2 types of user perception: audial and visual. For the first type of users, more audio materials are needed, and for the second type of perception, video materials are needed. Creating an ontology will help to create conditions for individualizing learning. Using various methods of semantic data analysis can create an individual learning path for users.

The third section of the dissertation, various methods and models for calculating the semantic proximity of words and topics are considered. The

problem is solved with a sequence of topics in one discipline, as well as with the number of training elements. A hypothesis is proposed, as well as a proof of the hypothesis using various scientific methods. A mathematical model for the information system is developed. The assessment of the effectiveness of training with an intelligent system is also determined.

The adaptability of the system is that each user will have their own learning path. Users log on to the system pass 2 tests that will determine the type of perception and level of knowledge for a particular course. The system will provide personalized content each time.

Consider the content generation algorithm. The relationship between the T_N topics of any course can be represented as a directed graph G_T, the vertices of the graph are topics, and the arcs define the relationship between the topics. Properties of the topic graph G_T:

1) There is an initial vertex that corresponds to the beginning of learning process, and an ending vertex that corresponds to the end of learning process.
2) G_T is an infinite graph, since having mastered the sequence of topics, the beginning of which is some topic T_N, the student cannot begin to study it again.
3) Since the graph G_T is infinite, it can be decomposed into levels, that is, represented as a hierarchy.

In this case, at the upper level there will be a fictitious initial vertex, and at the lower level, a fictitious final vertex, which corresponds to the end of the learning process. The final vertex is determined on the basis of the analysis of the competencies that the student must have after graduation.

Consider the algorithm decomposition into levels of an infinite graph G_T:

1) Find a vertex without incoming arcs and assign it a rank r = 0. Delete arcs from this vertex.
2) Suppose that at some stage in the graph there are no vertices without incoming arcs. Assign the following value of rank r to these vertices and cross out the arcs that go out of them.
3) Stage 2 is repeated until all the vertices are ranked.

Remark 1. The infiniteness property is hereditary, that is, if any vertex of the graph is deleted along with incident arcs, the remaining subgraph is also an infinite.

Remark 2. Non-contour graphs have the following properties:

- There is at least one vertex without incoming arcs (initial vertex);
- There is at least one vertex without outgoing arcs (called finite);
- The graph can be represented as a hierarchy, that is, decomposed into levels, while the level number is the length of the maximum path from the initial vertex to the vertex of this level.

Use method for calculating semantic proximity by professor Bondarchuk. Suppose that w_1 and w_2 are words for which semantic proximity must be calculated. The Bondarchuk's method can be divided into several steps:

1. Formation of contextual sets of words w_1 and w_2 (these sets contain words with which the words w_1 and w_2 are often used in the same context). These sets contain words with which the words w_1 and w_2 are often used in the same context. Then we form a common contextual set of words. Obviously, the cardinality of this set will be equal to n+m.
2. Calculating normalized affinities between the common determinant and each of the words $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$
3. The calculation of semantic proximity. Consider the calculation of semantic proximity between the words w_1 and w_2. To do this, we calculate the coefficients R_i for all words from the context set C.

Also need to calculate the semantic proximity of all topics with each other, shown in Table 1:

Table 1. Semantic proximity between topics

Topics	$\boldsymbol{w_1}$	$\boldsymbol{w_2}$	**Semantic proximity**
Topic 1 - Introduction to the Information Systems	Topic 1	Topic 2	0,3
Topic 2 - Software Requirement Specification (part 1)	Topic 2	Topic 3	0,8
Topic 3 - Software Requirement Specification (part 2)	Topic 3	Topic 4	0,77
Topic 4 - Information System modeling	Topic 4	Topic 5	0,483
Topic 5 - UML modeling	Topic 5	Topic 6	0,81
Topic 6 - Design of IS	Topic 6	Topic 7	0,44
Topic 7 - Database	Topic 7	Topic 8	0,21
Topic 8 - Search Engine	Topic 8	Topic 9	0,37
Topic 9 - Software Testing Information System	Topic 9	Topic 10	0,48
Topic 10 - Feasibility study of the IS	Topic 10	Topic 11	0,91
Topic 11 - Criteria for IT projects	Topic 11	Topic 12	0,27
Topic 12 - Development of IS	Topic 12	Topic 13	0,67
Topic 13 - ADO technology	Topic 13	Topic 14	0,34
Topic 14 - Final presentation, Topic 15 - Examination	Topic 14	Topic 15	0,94

Put-forward hypothesis: if the semantic proximity is greater than 0.75 (as a result of expert surveys and an experiment), then the topics are dependent, namely, a large semantic relationship between the topics is determined. In other words, the

content of these topics is linked and dependent on each other. And of course, the themes are consistent. If we stick to the hypothesis, then the dependence appears in the themes:

1. dependency - T2, T3;

2. dependency – T3, T4;

3. dependency – T5, T6;

4. dependency – T10, T11;

5. dependency – T14, T15.

Rules:

- Rule 1: If the student has not mastered the 2nd topic, they will not be able to go to the 3rd topic;
- Rule 2: If a student has mastered topic 2 but has not mastered topic 3, they will not be able to go to topic 4;
- Rule 3: If a student has not mastered topic 5, they will not be able to go to topic 6;
- Rule 4: If a student has not mastered topic 10, they will not be able to go to topic 11;

Rule 5: If a student has not mastered topic 14, they will not be able to go to topic 15.

Using semantic representation of data, as well as semantic analysis of the text, can develop a sequence of topics in a single subject. And in the future, can use this method not only in the subject, but also in the course, or in drawing up a plan for academic years.

The experiment was conducted at Astana IT University during the first trimester of 2019 year among first-year students. This experiment involved 115 users divided into 5 groups. Students first took a test to determine the level of knowledge, then a test to determine the type of perception. There are only 2 types of perception for the educational system: visual and audial. The type of perception determined the type of content. For visuals, the system provided more videos, presentations, for audials, only audiobooks and lectures. After each lecture, tasks were given to complete, as shown in Table 2. During for the first trimester users can pass all asignments, include practical or laboratory tasks and quizzes. Quizzes was multiple questions choice. All asignments given by order.

Table 2 – all points for one group

		type	level kn	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15
1	1_student	audial	22	95	49	64	70	72	78	85	93	87,5	95	90	80	81	87	82,5
2	2_student	visual	20	95	63	73	55	74	76	85	85	75	85	80	70	74	75	75
3	3_student	visual	20	95	43	50	60	57	72	85	92	87,5	100	75	80	80	90	81,6
4	4_student	audial	23	70	43	53	65	32	53	85	85	37,5	75	70	50	62	74	64,1
5	5_student	audial	22	95	60	60	70	88	83	80	90	56	85	85	70	73	84	80,4
6	6_student	audial	16	95	76	73	69	75	85	90	90	75	75	77	60	75	100	88

7	7_student	visual	15	55	76	70	59	30	44	60	74	62,5	85	80	60	64	50	52,4
8	8_student	audial	15	95	76	65	50	90	84	70	85	87,5	100	90	70	75	100	87,7
9	9_student	audial	15	95	43	33	75	63	73	50	50	62,5	80	85	60	60	65	65,9
10	10_student	audial	7	0	66	70	83	71	75	80	75	0	85	80	60	63	77	72,2
11	11_student	audial	13	80	76	70	59	57	61	75	70	50	70	75	70	69	64	64,6
12	12_student	visual	17	85	76	64	60	85	85	90	90	87,5	90	85	85	83	85	84,4
13	13_student	audial	13	0	43	53	49	52	54	75	80	0	0	75	70	59	60	57,9
14	14_student	audial	21	95	46	57	83	78	77	60	90	0	100	75	55	90	99	89,7
15	15_student	visual	29	95	76	70	60	73	77	85	85	75	85	90	80	79	85	80,8
16	16_student	audial	17	85	60	53	70	73	78	85	90	87,5	90	90	80	79	79	78,7
17	17_student	visual	18	80	63	67	60	85	75	80	90	0	0	60	70	90	90	85,5
18	18_student	audial	28	85	49	50	66	93	83	70	84	87,5	85	85	85	78	80	80,3
19	19_student	audial	7	95	49	57	64	56	77	75	75	75	85	90	75	75	79	77,2
	Average			78,42	59,63	60,63	64,6	68,63	73,15	77,1	82,78	60,75	81,66	80,89	70	74,15	80,15	76,25

Proof of the hypothesis. H_o - null hypothesis, all topics are unrelated. H_A - alternative hypothesis, topics are related, where the semantic proximity is greater than 0.75. Using local reference methods such as the Dice measure, the Jaccard index, and the Kulchinsky measure, and the Pearson correlation method, one can characterize the existence of a linear relationship between two quantities.

Teachers gave an expert assessment of the relationship between the topics of each lecture. We need to calculate of semantic proximity using the Bondarchuk's method, the Jaccard's index, the Kulchinsky's measure, the Dice's measure, expert evaluation and the Pearson's correlation. Students passed the tasks and average values were used for all methods. A summary of the full text was taken from each topic. By the number of words between topics, we can calculate the proximity of words using different methods.

For expert evaluation 3 teachers of the same discipline evaluated the proximity of topics. Each topic was evaluated and the final score was taken as an average value. Substituting for the method formulas we get the values for all methods. The Table 3 shows pairs of topics and also calculated semantic proximity using the Bondarchuk's method, the Jaccard's index, the Kulchinsky's measure, the Dice's measure, expert evaluation and the Pearson's correlation.

Table 3 - Semantic proximity of topics using different methods

w_1	w_2	**Bondarchuk's semantic proximity**	**Jaccard' index**	**Kulchinsky' measure**	**Dice' measure**	**Expert evaluation**	**Pearson's correlation**
Topic 1	Topic 2	0,3	0,1379	0,2244	0,2666	0,3	0,05246
Topic 2	Topic 3	0,8	0,0638	0,1428	0,0857	0,9666	0,77802
Topic 3	Topic 4	0,77	0,0862	0,1607	0,1785	0,75	0,7516
Topic 4	Topic 5	0,483	0,0416	0,0881	0,0909	0,4666	0,06282

Topic 5	Topic 6	0,81	0,025	0,049	0,0526	0,9	0,89231
Topic 6	Topic 7	0,44	0,0363	0,0789	0,0526	0,45	0,28259
Topic 7	Topic 8	0,21	0,0338	0,0697	0,0869	0,3166	0,68234
Topic 8	Topic 9	0,37	0,0172	0,0356	0,0277	0,3166	0,12597
Topic 9	Topic 10	0,48	0,0172	0,0356	0,0434	0,55	0,55783
Topic 10	Topic 11	0,91	0,0192	0,0384	0,0333	0,883	0,82699
Topic 11	Topic 12	0,27	0,0174	0,1333	0,0133	0,25	0,46811
Topic 12	Topic 13	0,67	0.0185	0,0366	0,04	0,5333	0,44746
Topic 13	Topic 14	0,34	0,0232	0,0463	0,0526	0,3666	0,73213
Topic 14	Topic 15	0,94	0,0243	0,048	0,0434	0,9166	0,95878

The correlation and semantic proximity show that they are closer to the expert assessment of teachers. Thus, we can conclude that where there is a semantic proximity greater than 0.75, the topics are dependent and consistent. For this subject, there are 5 pairs of topics that are interrelated, as described in the hypothesis. When the semantic proximity is greater than 0.75 according to the Bondarchuk's method and the Pearson's correlation there is a similarity of data.

During testing, users used an intelligent system for training. Through the system, we received completely content based on the type of perception. Students also took tests and assignments. Thus, we can see the initial level of knowledge and the final knowledge. Grades for assignments are also similar to the semantic proximity of topics.

Semantic analysis, as well as methods of semantic proximity of topics, helps to solve many problems in the educational sphere: the sequence of topics, the number of educational elements and the content.

Mathematical model has been developed for information and training system, which includes 4 different coefficients (1) and has condition (2).

$$W = \frac{1}{4}(\alpha w_1 + \beta w_2 + \gamma w_3 + \delta w_4) \qquad (1)$$

where $\alpha -$ coefficient of Bondarchuk′ semantic proximity,
β – coefficient of Jaccard' index,
γ - coefficient of Kulchinsky' measure,
δ - coefficient of Dice' measure.

The condition characterizes the put-forward hypothesis: if any x(i) word respond in Text T1, y(j) respond in Text T2 and semantic proximity of two topics is equal to or exceeds 0,75, then these two topics are similar in meaning, otherwise the two topics are different.

$$\text{If } \forall(x(i) \in T1)\ and\ ((y(j) \in T2)\ and\ (\rho(x(i), y(j))) \geq 0{,}75$$
$$\Rightarrow\ (x(i), y(j) \subseteq z) = 1$$
$$else\ \ (x(i), y(j) \not\subseteq z) = 0 \qquad (2)$$

where x(i) – i a word in the text T1,
y(j) – j a word in the text T2,
ρ (·) – distance calculated using the Bondarchuk method,
$\rho(\cdot) \geq 0{,}75$.

Microlearning can help people to expand their knowledge and not waste a lot of time. To create the content of education in microlearning, it is necessary to extract knowledge and use annotations, thus compressing the main facts and obtaining the necessary information.

According to Microsoft research, the average attention span is now just 8 minutes. Microlearning should contain content, especially the duration of video or audio materials should be 5-10 minutes, number of slides in the presentations should be 5-10, and number of pages in the text materials should be 1-2 pages, test and practical tasks. The structure of Microlearning is similar to the structure of the traditional teaching method, but there are differences, such as testing of the initial level of knowledge, lesson (audio/video and text material), test or practical tasks and final test/ practice task. Each user category has its own mini-course. The total number of the category is 10.

This section discusses algorithms, models and methods for creating an intellectual information system for education. The level decomposition algorithm is used to solve the content sequence. A hypothesis is put forward and proofs are given using various methods, including the calculation of the semantic proximity of Bondarchuk's method, Jaccard index, the Kulchinsky measure, the Dice measure, expert evaluation and the Pearson’s correlation. A mathematical model of the learning system based on the semantic proximity of topics is proposed. It is used to extract knowledge for the microlearning method. Microlerning allows users to perceive and assimilate information faster and easier. The evaluation of the efficiency of using the system using a nonparametric method of statistical analysis using the Man-Whitney test is calculated. The effectiveness of training with the use of IS is 1.15 times higher.

The fourth section describes the technical implementation of an intelligent information and training system “edu-elt.kz”. The software is implemented as a website using the open-source general-purpose programming language PHP, the MySQL integration DBMS, and a web framework, Laravel, designed for development using the MVC architectural model. The choice of the PHP programming language is due to the fact that it is a scripting language created for sites with dynamic content. User can fee course features: duration of course, quantity of lectures, quizzes, test result and type of perception or character. In that page also user can see description of the course, indivial curriculum and members of the course in accordance with Figure 1.

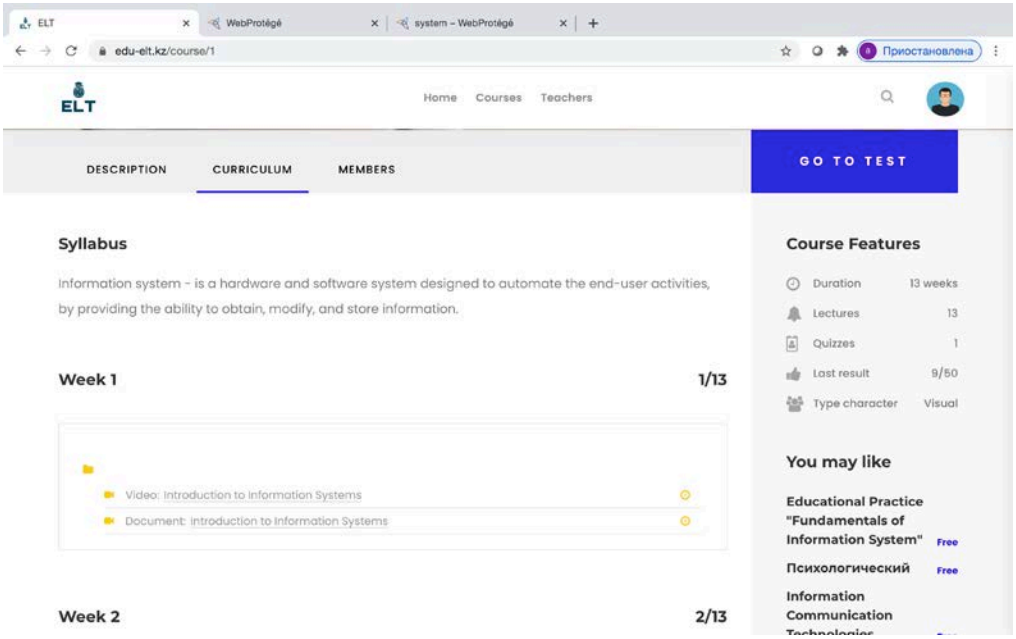

Figure 1 – Curriculum of the course

Conclusions

In this dissertation research, the development of an intelligent information and training system using models and methods of semantic analysis is carried out. The theoretical research carried out within the framework of the dissertation allowed us to obtain the following results of practical and scientific significance:

1. The analysis and review of the existing systems of distance learning and adaptive intelligent systems. The requirements for the developed system are formulated.
2. Used scientific methods for the formation of content intended for any student, which takes into account the characteristics of the user.
3. Developed a methodology for building an educational resource based on semantic analysis.
4. A hypothesis is put forward about the semantic links between content-dependent and proved in the course of the study.
5. The sequence and management of the number of educational elements of an intelligent system for distance learning in the field of information technology are proposed. All components of this system are described in detail. A mathematical model is developed using semantic data analysis.
6. The result of the experimental study of the developed system, which demonstrates an increase in the quality of education and efficiency due to the semantic representation of resources, also creates conditions for the digitalization of the educational process.

The reliability of the theoretical results of the study is confirmed by scientific evidence, experimental studies, as well as experimental and industrial tests in real conditions.

Conducted of calculation of the degree of assimilation of educational material. It was found that in the experimental group the degree of assimilation increased by 0.62 points, and in the control group it fell by 0.13 points. The effectiveness of training with the use of IS is 1.15 times higher. The significance of differences between the control and experimental groups was assessed by a nonparametric method of statistical analysis using the Mann-Whitney test. The economic efficiency of the implemented work results is confirmed by the implementation act.

In the future works, can consider recommendations for using and creating new content for training and consider extracting knowledge from other types of data, such as audio and video, used in courses.