APPROVE
Director of the Institute of Applied
Mathematics and Computer Science
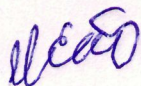_____ A.V. Zamyatin
« 16 » 05 2022
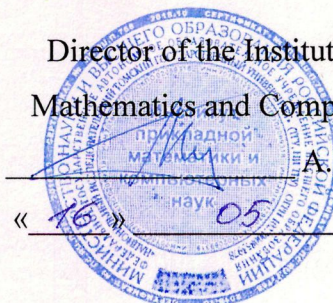
Evaluation materials of the current control and intermediate certification in the discipline
(Evaluation tools by discipline)
Natural Language Processing - II

in the major of training
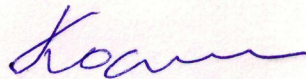**01.04.02 Applied mathematics and informatics**

Orientation (profile) of training:
Big Data and Data Science

Tomsk – 2022

ET was implemented:
cand. tech. Sciences, Associate Professor,
Associate Professor of the Department
of Theoretical Foundations of Informatics

M.S. Pozhidaev

Reviewer:
Dr. tech. sciences, professor,
Professor of the Department
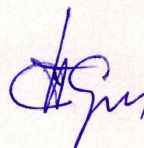of Theoretical Foundations of Informatics

Yu.L. Kostyuk

Evaluation tools were approved at a meeting of the educational and methodological commission of the Institute of Applied Mathematics and Computer Science (EMC IAMCS).

Protocol dated 12.05.2022 № 4

Chairman of the EMC IAMCS,
Dr. tech. Sciences, Professor

S.P. Sushchenko

Evaluation tools (ET) are an element of the system for assessing the formation of competencies among students in general or at a certain stage of its formation.

The ET is developed in accordance with the work program (WP) of the discipline.

1. Competencies and training outcomes, obtained upon the discipline mastery

Competencies	Competence indicator	Code and name of planned training outcomes that characterize the stages of competency formation	Criteria for evaluating training outcomes			
			Excellent	Good	Satisfactory	Unsatisfactory
UK-1. Able to carry out a critical analysis of problem situations based on a systematic approach, develop an action strategy	IUK-1.1 Identifies a problem situation, on the basis of a systematic approach, carries out its multifactorial analysis and diagnostics. IUK-1.2 Carries out the search, selection and systematization of information to determine alternative options for strategic solutions in a problem situation.	MR-1.1.1. The student will be able to: - find and use sources of additional information to improve the level of general and professional knowledge; - to select and process information on the chosen research topic; correctly quote and make references to the sources used in written works; - be able to apply natural science and mathematical knowledge to solve scientific and engineering problems in the field of informatics and computer technology.	90-100 points	70-89 points	55-69 points	0-54 points

	IUK-1.3 Suggests and justifies the strategy of action, taking into account the limitations, risks and possible consequences.					
GPC-1. Able to solve actual problems of fundamental and applied mathematics.	IGPC -1.1 Analyzes problems in the field of fundamental and applied mathematics.	MR-1.1.2. The student will be able to: - develop means of implementing information technologies (methodological, informational, mathematical, algorithmic, technical and software) - to conduct experimental studies according to a given methodology and analyze the results. - perform processing and analysis of data obtained in theoretical and experimental studies.				

1. Competencies and learning outcomes formed as a result of mastering the discipline

Competence	Competence Indicator	Code and name of learning outcomes (planned learning outcomes that characterize the stages of competence formation)	Criteria for evaluating learning outcomes			
			Excellent	Good	Satisfactory	Unsatisfactory
UK-1. He is able to carry out a critical analysis of problem situations based on a systematic approach, to develop a strategy of actions	IUK-1.1 Identifies a problematic situation, carries out its multifactorial analysis and diagnostics on the basis of a systematic approach.	OP-1.1.1. The ability to apply the basics of applied system analysis to identify problem situations and their causes.	The student is fluent in the material passed, answers correctly to the vast majority of teachers, is able to independently develop solutions to the proposed problems and shows the presence of knowledge acquired outside of classes.	The student has the necessary material and answers some of the teacher's questions. Able to solve tasks with a moderate number of hints.	The student has a poor command of the material and is practically unable to answer the teacher's questions. Copes with the solution of the tasks set only if there is intensive help from a teacher. There is no knowledge gained outside of lessons.	There is no knowledge of the subject. When formulating questions and tasks, he does not orient himself in the nature of the problem being solved. I am not able to answer the teachers' questions.
	IUK-1.2 Searches, selects and systematizes information to determine alternative options for strategic solutions in a problematic situation.	OP-1.2.1. Skills in constructing various strategies to eliminate the detected problem and choosing the most promising of them.				
	IUK-1.3 Proposes and justifies an action strategy taking into account limitations, risks and possible consequences.	OP-1.3.1. The ability to perform the necessary actions on the basis of the preparatory measures taken to eliminate the current problematic situation.				

UK-3. Able to organize and manage the work of the team, developing a team strategy to achieve the goal	IUK-3.1 Forms a strategy of teamwork based on a joint discussion of goals and activities for their implementation.	OP-3.1.1. Skills of distribution of roles in the team based on the identified strengths and weaknesses of team members.				
	IUK-3.2 Organizes the work of the team taking into account objective conditions (technology, external factors, limitations) and individual capabilities of team members.	OP-3.2.1. The ability to plan work with the construction of a schedule of deadlines and estimates of the completion of each stage.				
	IUK-3.3 Ensures the fulfillment of assigned tasks on the basis of monitoring teamwork and timely response to significant deviations.	OP-3.3.1. The ability to control the passage of work stages in accordance with the accepted schedule.				
PC-6. Able to manage the receipt, storage, transmission, processing of big data	IPK-6.1 Monitors and evaluates the performance of big data processing.	OR-6.1.1. Knowledge of the basics of microservice architecture for building horizontally scalable systems.				
	IPK-6.2 Uses methods and tools for receiving, storing, transmitting, and processing big data.	OR-6.2.1. Skills in configuring network file storage with support for atomic operations.				
	IPK-6.3 Develops proposals to improve the performance of big data processing.	OR-6.3.1. The ability to control the integrity of stored information.				

2. Stages of competency formation and types of evaluation tools

№	Stages of competence formation(sections of the discipline)	Code and description of training results	Type of evaluation tool (tests, assignments, cases, questions, etc.)
1.	Section 1. Introduction, history of discipline development, solved tasks, approaches, methods and tools	OR-1.1.1, OR-1.2.1, OR-1.3.1, OR-3.1.1.	Implementation of practical work No. 1
2.	Section 2. Text data preprocessing	OR-1.1.1, OR-3.2.1, OR-6.1.1, OR-6.3.1.	Implementation of practical work No. 2
3.	Section 3. Probabilistic algorithms	OR-1.2.1, OR-1.3.1, OR-6.1.1, OR-6.3.1.	Implementation of practical work No. 3
4.	Section 4. Formal Grammars	OR-1.1.1, OR-1.3.1, OR-3.2.1, OR-3.3.1.	Implementation of practical work No. 4
5.	Section 5. Vector representation of words	OR-1.1.1, OR-1.2.1, OR-3.2.1, OR-3.3.1.	Implementation of practical work No. 5
6.	Section 6. Seq2seq Model	OR-1.3.1,OR-3.2.1, OR-6.1.1, OR-6.3.1.	Implementation of practical work No. 6
7.	Section 7. Self-attention and Transformer	OR-1.1.1, OR-3.3.1, OR-6.1.1, OR-6.3.1.	Implementation of practical work No. 7
8.	Section 8. BERT and GPT-3	OR-1.2.1, OR-1.3.1, OR-6.2.1, OR-6.3.1.	Implementation of practical work No. 8

3. Typical control tasks or other materials necessary for the assessment of educational training outcomes

3.1. Typical tasks for conducting ongoing monitoring of progress in the discipline:

List of practical works:

Practical work No. 1. Parsing sites / using the api to get text data.

Practical work No. 2. Implementation of Porter stemmer.

Practical work No. 3. Using libraries for morphological analysis, solving the problem of partial markup.

Practical work No. 4. Vector representation of text, word2vec, skip-gram and CBOW models.

Practical work No. 5. Thematic modeling using the gensim library.

Practical work No. 6. Analysis of the tonality of text data. Deploying a trained model on the web.

Practical work No. 7. Building a language model, generating text.
Practical work No. 8. Generation of image caption.

3.2. Typical tasks for conducting intermediate certification in the discipline.
1. Compare the work of analytical formal grammars and probabilistic models.

2. Name the strengths and weaknesses of models based on artificial neural networks compared to analytical algorithms.

3. What operations become possible on words if you make the transition from their symbolic representation to vector representation?

4. What are the most common ways to obtain a vector representation of words can be distinguished?

5. What problems of recurrent models can be identified to justify the use of long-term short-term memory?

6. What is the key advantage in the mechanism of internal attention compared to the traditional mechanism of attention?

7. What task is the Transformer focused on in its original form?

8. What is the idea of fine tuning?

9. For what reasons can the BERT and GPT models be called complementary to each other?

10. What is the key idea of generative models?

4. Methodological materials that determine the procedures for evaluating training outcomes
4.1. Methodological materials for assessing the current control of academic performance in the discipline.
The current monitoring of progress is carried out during the delivery of practical work. Each work is evaluated according to the following parameters, according to the table in section 1:
• completeness of the implementation of practical work,
• answers to questions on practical work,
• answers to questions on theory from the relevant section of the course,
• ability to correct mistakes and promptly make changes in practical work.

4.2. Methodological materials for conducting intermediate certification in the discipline.

"Credited" – the student completed all practical work, answered all questions on practical work;

"Not credited" – the student did not pass any practical work, did not answer questions on practical work.

During the test, the student can increase his grade by re-submitting the relevant practical work.