Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Computer Systems Software
Department

ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR INFORMATION RETRIEVAL SYSTEMS

Syllabus

Requisites of the Course	
Cycle of Higher Education	***Second cycle of higher education (Master's degree)***
Field of Study	*12 Information technologies*
Speciality	*121 Software Engineering*
Education Program	*Software Engineering of Multimedia and Information Retrieval Systems*
Type of Course	*Normative*
Mode of Studies	*full-time*
Year of studies, semester	*1 year, 1 semester*
ECTS workload	*36 hours for lectures, 18 hours for practicals, 66 hours for self-study*
Testing and assessment	*Final test, modular control work, calendar control*
Course Schedule	*According to http://roz.kpi.ua/*
Language of Instruction	*English*
Course Instructors	*Senior lecturer, Shkurat Oksana, PhD, shkurat@pzks.fpm.kpi.ua*
Access to the course	Google classroom

Outline of the Course

1. Course description, goals, objectives, and learning outcomes

The study of the discipline «Artificial intelligence technologies for information retrieval systems» allows students to acquire competencies necessary for solving practical problems related to the developing software for intelligence processing of text, graphic and audio data and also the using technologies of artificial intelligence in software.

***The purpose** of studying «Artificial intelligence technologies for information retrieval systems» is to form abilities independently to develop software that implements artificial intelligence technologies and to use third-party software tools for developing intelligent information systems.*

***The subject** of «Artificial intelligence technologies for information retrieval systems» is information technologies of presentation, transformation, recognition and retrieval of text, graphic and audio data.*

*The study of the discipline «Artificial intelligence technologies for information retrieval systems» to form **professional competences (PC)** in students, necessary for solving practical tasks of professional activity related to the development, improvement and support of intelligent information systems for processing multimedia data:*
***PC07** Ability to critically comprehend problems in the field of information technology and at the frontiers of knowledge, to integrate relevant knowledge and solve complex problems in broad or multidisciplinary contexts.*

PC11 *Ability to apply and develop fundamental and interdisciplinary knowledge to successfully solve scientific problems of software engineering.*
PC13 *Ability to design and construct, implement and maintain web-based software systems to implement new information retrieval methods.*
PC16 *Ability to apply artificial intelligence technologies for designing information and search systems.*

The study of the discipline «Artificial intelligence technologies for information retrieval systems» to form ***program learning outcomes (PLO)*** *according to the educational program:*
PLO03 *Build and research models of information processes in the application field.*
PLO04 *Identify information needs and classify data for software design.*
PLO07 *Analyze, evaluate and apply at the system level modern software and hardware platforms to solve complex problems of software engineering.*
PLO16 *Plan, organize and perform software testing, verification and validation.*
PLO17 *Collect, analyze, evaluate the information needed to solve scientific and applied problems, using scientific and technical literature, databases and other sources.*
PLO19 *Formulate, experimentally test, substantiate and apply in practice in the process of software development innovative methods and competitive technologies for solving professional, scientific and technical problems in multidisciplinary contexts.*
PLO24 *Be able to modify existing and develop new methods and algorithms for classification and clustering of data, taking into account the characteristics of the subject area.*
PLO34 *Know the approaches, directions, models and methods of artificial intelligence, including machine learning; know the technology of software development of artificial intelligence systems, apply artificial intelligence methods in research and to solve applied problems.*

2. Prerequisites and post-requisites of the course (the place of the course in the scheme of studies in accordance with curriculum)

Successful study of the discipline «Artificial intelligence technologies for information retrieval systems» is preceded by the study of the disciplines «Mathematical Support of Multimedia and Information Retrieval Systems», «Algorithmic Support of Multimedia and Information Retrieval Systems», «Programming», «Information Retrieval Systems Software», «Components of Software Engineering» of the curriculum for training bachelors in the specialty 121 Software engineering.

The theoretical knowledge and practical skills obtained during the stugy of the discipline «Artificial intelligence technologies for information retrieval systems» can be useful for conducting scientific research on the topic of the dissertation and course projects in the specialty 121 Software Engineering.

3. Content of the course

The discipline «Artificial intelligence technologies for information retrieval systems» includes the study of the following topics:

Topic 1. Introduction to artificial intelligence technologies

Topic 2. Text recognition in information retrieval systems

Topic 3. Image recognition in information retrieval systems

Topic 4. Sound recognition in information retrieval systems

Modular control work

Final test

4. Coursebooks and teaching resources

Main literature:

1. Artificial intelligence: How does it work, why does it matter, and what can we do about it? / P. Boucher – EU, Brussels: European Parliamentary Research Service, 2020 – 76 p. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf

2. The technological elements of artifical intelligence / M. Taddy – NBER Working Paper, No. 24301. URL: https://www.nber.org/system/files/working_papers/w24301/w24301.pdf

3. Artificial intelligence: A Modern Approach / S.J. Russel, P. Norving – Prentice Hall, 2006 – 1409 p.

4. Methods and information technology of processing archival medical images: PhD Thesis: 05.13.06 / O.S. Shkurat. – Kyiv: 2020. – 224 p.

Additional literature:

5. Shkurat O. Composed Approach to Image Object Recognition. Débats scientifiques et orientations prospectives du développement scientifique: proceeding III International Scientific and Practical Conference (Paris, July 8, 2022). Paris, 2022. P. 171-173. ISBN: 978-617-8037-79-6

6. Shkurat O. Overview of Text Retrieval Algorithms. "Grundlagen der Modernen Wissenschaftlichen Forschung": proceeding III International Scientific and Practical Conference (Zurich, Switzerland, August 12, 2022). Zurich, 2022. P. 111- 112. ISBN: 978-617-8037-84-0

7. Artificial Intelligence: A moderm Approach / S.J. Russel, P.Norving. – Pearson Education, Inc., 2010, 1132 p.

8. Artificial Intelligence – URL: https://www.dcpehvpm.org/E-Content/BCA/BCA-III/artificial_intelligence_tutorial.pdf

9. Expert Systems: Principles and Programming / J.C. Giarratano, G.D. Riley – Australia; Boston, Mass.: Thomson Course Technology, 2005 – 842 p.

10. Principles of Database and Knowledge – Base Systems / J.D. Ullman – USA : Maryland, Computer Science Press, 1988 – 654 p.

11. Information System Design and Development – URL: https://egyankosh.ac.in/bitstream/123456789/25561/1/Unit-2.pdf

12. Software Testing Tutorial – URL: https://digitalpoint.tech/admin/uploads/4346d933bcfa1d59b368d121f6747980.pdf

13. Introduction to Database and Knowledge-Base Systems / S. Krishna – World Scientific Publishing Company, 1992 – 344 p.

Educational content

5. Methodology

No	*Type of a class*	*Description of a class*
Topic 1. Introduction to artificial intelligence technologies		
1	*Lecture 1. Fundamentals of artificial intelligence technologies*	*Basic definitions of artificial intelligence technologies. The history of the development of artificial intelligence technologies. Functional capabilities of modern intelligent systems. Tasks which are solved by modern artificial intelligence systems. The role of artificial intelligence systems in modern life. Key companies in the world in the field of artificial intelligence technologies.* *Materials for self-studying: p.6 №1.*

2	*Computer class 1. Designing the technical task for the information retrieval system*	*Task: using the Trello platform, develop the technical task for the information system of text/image/audio retrieval.* *Materials for self-studying: p.6 №2.*
3	*Lecture 2. Intelligent systems*	*The main problems and directions of developing artificial intelligence technologies. Speech recognition technologies. Computer vision. Biometrics and biometric recognition. Text processing and analysis. Research directions of artificial intelligence. Prospects for the development of artificial intelligence technologies.* *Materials for self-studying: p.6 №3, 29.*
4	*Lecture 3. Expert systems. Features of developing expert systems*	*Approaches to building artificial intelligence systems. Expert systems and areas of their application. The main purposes of expert systems. Characteristics of the expert system. Features of expert systems. Problems of using expert systems. Expert system architecture. Classification of expert systems. Dynamic and static expert systems. Acquisition of knowledge by expert systems. Technology of building expert systems. Tools of design and development of expert systems.* *Materials for self-studying: p.6 №4, 30-31.*
5	*Computer class 2. Multimedia data recognition software tools (part 1)*	*Task: Analyze and justify the choice of software tools (frameworks, Application Programming Interface and program libraries) for speech/text/image recognition.* *Materials for self-studying: p.6 №5.*
6	*Lecture 4. Neural networks as tools artificial intelligence systems*	*Artificial neural networks. Architecture of artificial neural network. Single-layer and multi-layer artificial neural networks. Classification of artificial neural networks. Training of artificial neural networks.* *Machine learning. Problems solved by machine learning methods. The main ingredients of machine learning. Data, features, algorithms. Types of machine learning. Machine learning models. Machine learning algorithms. Gradient descent method. Learning NN by the method of backpropagation of the error. Advantages of machine learning.* *The place of deep learning among machine learning methods. Concepts of "deep network" and "deep learning". Models (algorithms) of deep learning. The main types of deep neural networks of deep learning: direct propagation networks; recurrent neural networks; convolutional neural networks; convolutional inverse deep networks; generative competitive networks. Deep learning software.* *Materials for self-studying: p.6 №6, 32.*

7	*Lecture 5. Tools for creating artificial intelligence systems*	*Means of building an expert system. Classification of tools for the development of an expert system. Artificial intelligence programming languages. Lisp is a functional programming language. Lisp functions and suggestions. Shells of expert systems. CLIPS is a software environment for developing expert systems. Modern frameworks of machine and deep learning: Tensor Flow, PyTorch, Keras, MXNet, CNTK (Microsoft Cognitive Toolkit), Caffe, Deeplearning4j, Chainer.* *Materials for self-studying: п.6 №7.*
8	*Computer class 2. Multimedia data recognition software tools (part 2)*	*Task: Using selected software tools (frameworks, APIs and libraries) to implement speech/text/image recognition in the information retrieval system* *Materials for self-studying: p.6 №8.*
Topic 2. Text recognition in information retrieval systems		
9	*Lecture 6. Information retrieval systems*	*Retrieval systems. Types of retrieval. Full text retrieval. Architecture of the information retrieval system. Indexing. Retrieval request. Indexing in retrieval engines. Retrieval performance parameters.* *Materials for self-studying: p.6 №9.*
10	*Lecture 7. Models of full-text retrieval in information retrieval systems*	*Ranking algorithms. Basic retrieval models: classical Boolean model, extended Boolean search, retrieval using fuzzy sets, vector-space model, probabilistic retrieval model and decision-making algorithms, retrieval using linguistic models.* *Materials for self-studying: p.6 №10.*
11	*Computer class 3. Implementation of the graphic interface of the information retrieval system*	*Task: With the help of selected software tools, implement a graphical interface for the developed retrieval information system.* *Materials for self-studying: p.6 №11.*
12	*Lecture 8. Retrieval tasks, methods and technologies of information retrieval*	*Phonetic word retrieval algorithms. Soundex algorithm. NYSIIS algorithm. Metaphone algorithm. Application of phonetic coding. Algorithms for determining phonetic distance. Hamming distance. Levenshtein distance. Jaro distance. Distance based on N-grams. Finding the longest common subsequence (LCS distance).* *Materials for self-studying: p.6 №12.*
13	*Lecture 9. Text classification and clustering*	*Classification task. Approaches to classification. The task of classification with a teacher, without a teacher. Naive Bayesian classifier. Logistic regression. Method of support vectors. The method of k-nearest neighbors. Decision tree. Artificial neural networks.* *Materials for self-studying: p.6 №13*

14	*Computer class 4. Authorization and authentication module in the information retrieval system (part 1)*	*Task: Analyze software tools and design a database in the developed information retrieval system.* *Materials for self-studying: p.6 №14.*
Topic 3. Image recognition in information retrieval systems		
15	*Lecture 10. Computer vision*	*Computer vision systems. Characteristics of input data. Receiving and processing registered data. Components of the digital image recognition system. Typical problems with image recognition. Levels of digital image processing. Representation of digital graphic data.* *Materials for self-studying: p.6 №15.*
16	*Lecture 11. Technologies for extracting features of digital images*	*Image filtering. Gradation transformations. Statistical and textural methods of image analysis.* *Materials for self-studying: p.6 №16.*
17	*Computer class 4. Authorization and authentication module in the information retrieval system (part 2)*	*Task: Using software, to implement the authorization and authentication module for the developed retrieval information system.* *Materials for self-studying: p.6 №17.*
18	*Lecture 12. Methods of selection of objects on digital images. Segmentation*	*Threshold segmentation: segmentation with global threshold and adaptive threshold. Segmentation based on growing areas. Morphological segmentation. Segmentation based on clustering. Segmentation based on watershed.* *Materials for self-studying: p.6 №18.*
19	*Lecture 13. Contour analysis of digital images*	*Operators for defining image edges. Active contour method. Contour tracing algorithms. Algorithms "Moore-Neighbor Tracing", "Redial Sweep", "Theo Pavlidi's Algorithm".* *Materials for self-studying: p.6 №19.*
20	*Computer class 5. Testing the information retrieval system (part 1)*	*Task: Perform functional testing of the developed information retrieval system.* *Materials for self-studying: p.6 №20.*
21	*Lecture 14. Digital image recognition technologies*	*The method of segmentation of homogeneous images. Motion tracking methods for compression and recognition of video data of the MPEG family.* *Materials for self-studying: p.6 №21.*
22	*Lecture 15. Neural networks for data retrieval*	*Neural network methods of image recognition. Multi-layer neural networks without feedback and with feedback. The concept of key and control points on a face image. Definition of the type of neural network model. Principles of adaptation of architectural parameters. Definition of convolutional neural network parameters.*

		Materials for self-studying: p.6 №22.
23	*Computer class 5. Testing the information retrieval system (part 2)*	*Task: Determine performance and security indicators for the developed information search system.* *Materials for self-studying: p.6 №23.*
Topic 4. Sound recognition in information retrieval systems		
24	*Lecture 16. Speech recognition*	*Digital sound parameters used to analyze audio signals (descriptors, features). Spectral features. The average value of the spectrum of the analyzed signal. Relative spectrum power in bands.* *Time signatures of sound signals. Segment duration, phonemes. Segment height. Segment form factor.* *Amplitude-frequency characteristics of sound signals. Intensity, amplitude. Energy. Frequency of the main tone. Modulation of the main tone.* *Materials for self-studying: p.6 №24.*
25	*Lecture 17. Speech signal recognition technologies*	*Forward propagation neural networks as a means of sound information analysis. Multi-layered perceptron. Networks with error backpropagation. Convolutional neural networks as an effective means of sound signal analysis.* *Materials for self-studying: p.6 №25.*
26	*Computer class 5. Testing the information retrieval system (part 3)*	*Task: Perform user interface and usability testing of the developed information retrieval system.* *Materials for self-studying: p.6 №26.*
Modular control work		

6. Self-study

The discipline "Artificial intelligence technologies for information retrieval systems" is based on self-study for a classes on theoretical and practical topics.

No	*Topic for self-studying*	*Hours*	*Literature*
1	*Preparation to the lecture 1*	*1*	*1; 2; 3; 10*
2	*Preparation to the computer class 1*	*1,5*	*11-13*
3	*Preparation to the lecture 2*	*1*	*1; 3*
4	*Preparation to the lecture 3*	*1*	*1; 3; 9; 10*
5	*Preparation to the computer class 2 (part 1)*	*1,5*	*11-13*
6	*Preparation to the lecture 4*	*1*	*3; 9; 10*
7	*Preparation to the lecture 5*	*1*	*3; 9; 10*
8	*Preparation to the computer class 2 (part 2)*	*1,5*	*11-13*
9	*Preparation to the lecture 6*	*1*	*2;*

10	*Preparation to the lecture 7*	*1*	*2;*
11	*Preparation to the computer class 3*	*1,5*	*11-13*
12	*Preparation to the lecture 8*	*1*	*2; 6*
13	*Preparation to the lecture 9*	*1*	*1; 2; 3*
14	*Preparation to the computer class 4 (part 1)*	*1,5*	*11-13*
15	*Preparation to the lecture 10*	*1*	*4; 5*
16	*Preparation to the lecture 11*	*1*	*4; 5*
17	*Preparation to the computer class 4 (part 2)*	*1,5*	*11-13*
18	*Preparation to the lecture 12*	*1*	*4*
19	*Preparation to the lecture 13*	*1*	*4*
20	*Preparation to the computer class 5 (part 1)*	*1,5*	*11-13*
21	*Preparation to the lecture 14*	*1*	*4*
22	*Preparation to the lecture 15*	*1*	*4*
23	*Preparation to the computer class 5 (part 2)*	*1,5*	*11-13*
24	*Preparation to the lecture 16*	*1*	*10*
25	*Preparation to the lecture 17*	*1*	*10*
26	*Preparation to the computer class 5 (part 3)*	*1,5*	*11-13*
27	*Preparation to the modular control work*	*4*	*1-10*
28	*Preparation to the final test*	*6*	*1-10*
29	*Internet of Things (IoT). IoT Challenges: Security and Privacy. Some features of IoT development.*	*6*	*2; 3; 4*
30	*Knowledge presentation models. Problems of presenting knowledge. Choosing a form of knowledge presentation. Approaches to presenting knowledge. Classification of knowledge representation models.*	*4*	*7; 8; 10*
31	*Concept of semantics. Semantic grid objects and relationships between them. Types of connections between objects of semantic networks. Types of semantic networks. Classification of semantic networks.*	*2,5*	*1;3; 7; 8; 9*
32	*Genetic algorithms and their application. Basic concepts. Classical genetic algorithm. Evolutionary and genetic algorithms. Genetic algorithms and neural networks. Advantages and disadvantages of genetic algorithms. Approaches to improving genetic algorithms and overcoming their shortcomings.*	*13*	*7*

Політика та контроль

7. Course policy

Attending lectures is mandatory.

Attending computer classes can be occasional.

Rules of conduct: activity, taking part in discussions, respect to teacher and groupmates, muting cellphones.

Compliance with the policy of academic integrity.

8. Monitoring and grading policy

During the first lecture, students become acquainted with the rating system of evaluation, which is based on the Regulations on the system of evaluation of learning outcomes https://document.kpi.ua/files/2020_1-273.pdf.

The semester credit module rating of a student is calculated based on 100-point scale. Points are given for 5 laboratory works (R_L) and modular control work. The maximum value of points for one work is equal to 10.

For laboratory work, points are awarded for:
– correct functioning of the developed programs: 0-5 points;
– defense of the results obtained in the work, answers to theoretical questions of the teacher, completeness of the work protocol: 0-3 points.
– timeliness of preparation of the protocol for laboratory class, completeness of performance of the theoretical task: 0-2 points.

The maximum value of laboratory works (R_L) makes 10 points × 5 laboratory works = 50 points.

The modular control work consists of 3 theoretical questions and 2 practical questions. The maximum value of points for one question is equal to 10.
Criteria for evaluating each question:
– correct and meaningful answer: 9-10 points;
– correct answer, incomplete explanations: 5-8 points;
– the answer contains errors: 1-4 points;
– no answer or the answer is incorrect: 0 points.

The maximum value** of modular control work (R_{MCW}) makes **10 points × 5 questions = 50 points.

The maximum value of a rating for a semester (R_S) makes $R_C = R_L + R_{MCW}$ = 100 points.

Calendar attestation of students (8 and 14 weeks of semesters) for the discipline is carried out according to the value of the current rating of the student at the time of attestation. If the value of this rating is not less than 50% of the maximum possible at the time of certification, the student is considered satisfactorily certified. Otherwise, in the attestation statement is set "unsatisfactory".

A necessary condition for obtaining a test by a student is the performance modular control work and defense of all laboratory works with a sum of at least 60%. Students who do not have academic debt can also increase their grades by taking a final test.

The final performance score or the results of the Final/ Pass Test are adopted by university grading system as follows:

Score	*Grade*
100-95	*Excellent*
94-85	*Very good*
84-75	*Good*
74-65	*Satisfactory*
64-60	*Sufficient*
Below 60	*Fail*
Course requirements are not met	*Not Graded*

9. Additional information about the course

Topics for the calendar control in the Attachment 1.

Syllabus of the course

Is designed by teacher PhD, Senior lecturer, Shkurat Oksana;

Adopted by Computer Systems Software Department (protocol № 12 from 26.04.23)

Approved by the Faculty Board of Methodology (protocol № 10 from 26.05.23)

Attachment 1. Topics for the calendar control

1. Artificial neural networks. Architecture of artificial neural network.
2. Single-layer and multi-layer artificial neural networks.
3. Classification of artificial neural networks.
4. Training of artificial neural networks.
5. Machine learning. Problems solved by machine learning methods.
6. Types of machine learning. Machine learning models. Machine learning algorithms.
7. The method of gradient descent.
8. Training of neural networks by the method of error back propagation.
9. The place of deep learning among machine learning methods.
10. Concepts of "deep network" and "deep learning". Models (algorithms) of deep learning.
11. The main types of deep neural networks of deep learning: direct propagation networks; recurrent neural networks; convolutional neural networks; convolutional inverse deep networks; generative competitive networks.
12. Means of building an expert system. Classification of tools for the development of an expert system.
13. Artificial intelligence programming languages.
14. Lisp is a functional programming language.
15. Lisp functions and suggestions.
16. Shells of expert systems. CLIPS is a software environment for developing expert systems.
17. Modern machine and deep learning frameworks: Tensor Flow, PyTorch, Keras, MXNet, CNTK (Microsoft Cognitive Toolkit), Caffe, Deeplearning4j, Chainer.
18. Composition and structure of automated information and retrieval systems.
19. Classification of information and retrieval systems.
20. Retrieval tasks, means and technologies of information retrieval.
21. Retrieval organization, technology and basic methods.
22. Basic retrieval models: probabilistic retrieval, retrieval using language models, extended Boolean retrieval, retrieval with hidden semantic indexing.
23. Representation of a digital image as a retrieval object.
24. Components of image recognition systems.
25. Technologies of statistical image analysis.
26. Technologies of texture analysis of images.
27. Technologies of structural image analysis. Image segmentation methods. Methods of contour analysis of images.
28. Stages of the classification task: feature generation, feature selection, classifier construction, classification quality assessment.
29. Mathematical formulation of the classification problem. The decisive rule.
30. Classification task with the teacher.
31. The problem of classification without a teacher.
32. Feature, feature vector, feature space.
33. Neural network methods of image recognition.
34. Multi-layer neural networks without feedback and with feedback.
35. The concept of key and control points on a face image.
36. Image retrieval technologies by metadata.
37. Image retrieval technologies by content.
38. Speech signal as a retrieval object.
39. Sound descriptors used to analyze audio signals.
40. Spectral features. The average value of the spectrum of the analyzed signal. Relative spectrum power in bands.
41. Time signatures of sound signals. Segment duration, phonemes. Segment height. Segment form factor.
42. Amplitude-frequency characteristics of sound signals. Intensity, amplitude. Energy. Frequency of the main tone. Modulation of the main tone.

43. Neural networks of forward propagation as a means of sound information analysis. Multi-layered perceptron.
44. Neural networks with error backpropagation.
45. Convolutional neural networks as an effective means of sound signal analysis.