

SOFTWARE OF INFORMATION AND SEARCH SYSTEMS. PART 1. NOSQL DATABASES Working program of the academic discipline (Syllabus)

Details of the academic discipline The level is higher First (undergraduate) education Branch of knowledge 12 Information technologies Specialty 121 Software engineering **Educational program** Software engineering of multimedia and information retrieval systems **Status disciplines** Normative Form of education daytime Year of training, semester 4th year of study, 7 semester Scope of the discipline Lectures: 36 hours, laboratory work: 18 hours, independent work: 66 hours. Semester control/ control Exam, modular control work, calendar control measures Language of teaching English Lessons schedule According to the schedule for the autumn semester of the current academic year http://roz.kpi.ua/ Information about the Lecturer: assistant Department of system programming and specialized computer course leader / teachers systems (SPiSKS) Radchenko Kostiantyn Oleksandrovych,

Laboratory works: assistant Department of SPiSKS Radchenko Kostiantyn Oleksandrovych, radchenko.kostiantyn@III.kpi.ua Placement of the course Google classroom: KP-94. https://classroom.google.com/c/NTQ1NjUwMTA2Nzgw?cjc=swwv4ei

radchenko.kostiantyn@III.kpi.ua

Program of educational discipline

1. Description of the educational discipline, its purpose, subject of study and learning outcomes

The study of the discipline "Software for information and search systems. Part 1. NoSQL databases" allows students to develop the competencies necessary for solving practical problems of professional activity related to the development of software for information and search systems in the context of non-relational databases.

The purpose study of the discipline "Software for information and search systems. Part 1. NoSQL databases" is the formation of students of education systems of theoretical and practical knowledge in the field of non-relational databases, as well as data storage and processing mechanisms different from relational databases; acquiring knowledge on the installation and operation of document-oriented database management systems, which does not require a description of the table schema.

Subject disciplines "Software for information and search systems. Part 1. NoSQL databases" are methods, algorithms and models used to develop and study non-relational databases.

The study of the discipline "Software for information and search systems. Part 1. NoSQL databases" contributes to the formation of **professional competencies (FC)** in students, necessary for solving practical tasks of professional activity, related to the development and use of information retrieval methods in software.

PC07 Knowledge of information data models, the ability to create software for data storage, retrieval and processing.

PC13 Ability to reasonably select and master software development and maintenance tools.

PC14 Ability to algorithmic and logical thinking.

PC15 Ability to apply fundamental and interdisciplinary knowledge to build advanced retrieval algorithms.

PC17 Ability to develop software for information retrieval systems.

PC20 Ability to apply the acquired fundamental mathematical knowledge to develop calculation methods in the multimedia and information retrieval systems creation.

Study of the discipline "Information Retrieval Systems Software 2. Methods of Organization of Information Retrieval" contributes to the formation in students of the following **program learning outcomes (PLO)** according to the educational program:

PLO03 To know the software life cycle basic processes, phases and iterations.

PLO05 To know and apply relevant mathematical concepts, domain methods, system and object-oriented analysis and mathematical modeling for software development.

PLO06 Ability to select and use the appropriate task of software development methodology.

PLO07 To know and to apply in practice the fundamental concepts, paradigms and basic principles of the functioning of language, instrumental and computational tools of software engineering.

PLO08 To know and to be able to develop a human-machine interface.

PLO10 To conduct a pre-project survey of the subject area, system analysis of the design object.

PLO11 To select initial data for design, guided by formal methods of describing requirements and modeling.

PLO12 To apply effective approaches to software design in practice.

PLO13 To know and apply methods of developing algorithms, designing software and data and knowledge structures.

PLO15 To choose programming languages and development technologies to solve the problems of creating and maintaining software.

PLO16 To have the software development, design approval and all types of software documentation release skills.

PLO17 To be able to apply methods of component software development.

PLO18 To know and be able to apply information technology of processing, storage and transmission of data.

PLO19 To know and be able to apply software verification and validation methods.

PLO20 To know approaches to evaluation and quality assurance of software.

PLO25 To know and to be able to use fundamental mathematical tools in the algorithms construction and modern software development.

PLO31 To be able to identify, analyze and document software requirements for multimedia and information retrieval systems.

PLO38 To be able to apply programming technologies for multimedia and information retrieval systems software development.

PLO39 To know the types of search engines, the principles of their construction, the methods and algorithms for performing different kinds of information retrieval in them.

PLO40 To know and be able to apply in practice the methods and criteria for estimating the effectiveness of information retrieval.

PLO42 To know the basic presentation models of textual and multimedia information and methods of its pre-processing for use in the design of information retrieval systems.

PLO43 To know and be able to use in practice the existing software resources and libraries for processing of textual information and multimedia data in information retrieval systems.

PLO44 To know the most common query languages used in the development of information retrieval systems.

The study of the discipline "Software for information and search systems. Part 1. NoSQL databases" contributes to the formation in students of the following **program learning outcomes (PLO)** according to the educational program:

PLO03 To know the software life cycle basic processes, phases and iterations.

PLO05 To know and apply relevant mathematical concepts, domain methods, system and object-oriented analysis and mathematical modeling for software development.

PLO06 Ability to select and use the appropriate task of software development methodology.

PLO07 To know and to apply in practice the fundamental concepts, paradigms and basic principles of the functioning of language, instrumental and computational tools of software engineering.

PLO08 To know and to be able to develop a human-machine interface.

PLO10 To conduct a pre-project survey of the subject area, system analysis of the design object.

PLO11 To select initial data for design, guided by formal methods of describing requirements and modeling.

PLO12 To apply effective approaches to software design in practice.

PLO13 To know and apply methods of developing algorithms, designing software and data and knowledge structures.

PLO15 To choose programming languages and development technologies to solve the problems of creating and maintaining software.

PLO16 To have the software development, design approval and all types of software documentation release skills.

PLO17 To be able to apply methods of component software development.

PLO18 To know and be able to apply information technology of processing, storage and transmission of data.

PLO19 To know and be able to apply software verification and validation methods.

PLO20 To know approaches to evaluation and quality assurance of software.

PLO25 To know and to be able to use fundamental mathematical tools in the algorithms construction and modern software development.

PLO31 To be able to identify, analyze and document software requirements for multimedia and information retrieval systems.

PLO38 To be able to apply programming technologies for multimedia and information retrieval systems software development.

PLO39 To know the types of search engines, the principles of their construction, the methods and algorithms for performing different kinds of information retrieval in them.

PLO40 To know and be able to apply in practice the methods and criteria for estimating the effectiveness of information retrieval.

PLO42 To know the basic presentation models of textual and multimedia information and methods of its pre-processing for use in the design of information retrieval systems.

PLO43 To know and be able to use in practice the existing software resources and libraries for processing of textual information and multimedia data in information retrieval systems.

PLO44 To know the most common query languages used in the development of information retrieval systems.

2. Pre-requisites and post-requisites of the discipline (place in the structural and logical scheme of training according to the relevant educational program)

The successful study of the discipline "Software for information and search systems. Part 1. NoSQL databases" is preceded by the study of the disciplines "Databases", "Databases. Course work", "Programming", "Programming. Course work" of the curriculum of bachelor's training in the specialty 121 Software engineering.

Theoretical knowledge and practical skills obtained during the study of the discipline "Software for information and search systems. Part 1. NoSQL databases" ensure successful study of the discipline "Software of information and search systems. Part 2. Methods of organizing information search» and execution of course works and bachelor's theses in the specialty 121 Software Engineering.

3. Content of the academic discipline

The discipline "Software for information and search systems. Part 1. NoSQL databases" involves the study of the following topics:

Topic 1. Introduction to search engines and services

Topic 2. Unstructured databases

Topic 3. NoSQL databases as information and search systems

Modular control work

Examination

4. Educational materials and resources

Basic literature:

1. Luc Perkins with Eric Redmond and Jim R. Wilson. Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement, Second Edition. \cdot 2018. – 354 P.

2. S. Amer-Yahia, MD Choudhury, M. Feldman, N. Golbandi, R. Lempel, and C. Yu. Automatic construction of travel itineraries using social breadcrumbs. In Proc. 21st ACM Conference on Hypertext and Hypermedia (Hypertext'2010), pages 35–44, June 2010.

3. 9. C. Anderson. The Long Tail - Why the Future of Business is Selling Less of More. Hyperion Books, New York NY, 2006.

Additional literature:

- 4. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence 1st Edition, Kindle Edition. by Pramod J. Sadalage, Martin Fowler.
- 5. SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management 1st ed. 2019 Edition by Andreas Meier, Michael Kaufmann.
- 6. Data Modeling With NoSQL Database null Edition by Ajit Singh , Sultan Ahmad.
- 7. Database Internals: A Deep Dive into How Distributed Data Systems Work 1st Edition by Alex Petrov.
- 8. MongoDB: The Definitive Guide: Powerful and Scalable Data Storage 3rd Edition by Shannon Bradshaw , Eoin Brazil , Kristina Chodorow.
- 9. NoSQL for Mere Mortals 1st Edition. by Dan Sullivan Sullivan.
- 10. Next Generation Databases: NoSQLand Big Data 1st ed. Edition. by Guy Harrison.
- 11. Mastering MongoDB 4.x: Expert techniques to run high-volume and fault-tolerant database solutions using MongoDB 4.x, 2nd Edition. by Alex Giamas.
- 12. NoSQL Databases A Complete Guide 2020 Edition Kindle Edition by Gerardus Blokdyk.
- 13. Getting Started with NoSQL Paperback by Gaurav Vaish
- 14. Making Sense of NoSQL: A guide for managers and the rest of us 1st Edition by Dan McCreary, Ann Kelly
- 15. Next Generation Databases: NoSQLand Big Data 1st ed. Edition by Guy Harrison
- 16. NoSQL and SQL Data Modeling: Bringing Together Data, Semantics, and Software First Edition by Ted Hills
- 17. Nosql the Ruby Way Paperback by Durran Jordan
- 18. Professional NoSQL 1st Edition by Shashank Tiwari
- 19. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence by Pramod J. Sadalage
- 20. MongoDB: The Definitive Guide by Kristina Chodorow
- 21. Redis in Action by Josiah L. Carlson
- 22. MongoDB in Action by Kyle Banker
- 23. Redis Cookbook by Tiago Macedo
- 24. Making Sense of NoSQL: A guide for managers and the rest of us by Dan McCreary
- 25. Hadoop: The Definitive Guide by Tom White (Goodreads Author)
- 26. Scaling MongoDB by Kristina Chodorow
- 27. Cassandra: The Definitive Guide by Eben Hewitt
- 28. Mongodb and Python: Patterns and Processes for the Popular Document-Oriented Database by Niall O'Higgins
- 29. HBase: The Definitive Guide by Lars George
- 30. CouchDB: The Definitive Guide: Time to Relax by J. Chris Anderson
- 31. Elasticsearch in Action (ebook) by Radu Gheorghe
- 32. The Little Redis Book (ebook) by Karl Seguin
- 33. Php And Mongo Db Web Development Beginner's Guide by Rubayeet Islam (Goodreads Author)
- 34. The Definitive Guide To Mongo Db: The No Sql Database For Cloud And Desktop Computing by Peter; Thielen, Wouter Membrey
- 35. The DynamoDB Book (ebook) by Alex DeBrie
- 36. Must Learn KQL: Essential Learning for the Cloud-focused Data Scientist (Kindle Edition) by Rod Trent

- 37. Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine by Clinton Gormley
- 38. Getting Started with CouchDB by M.C. Brown
- 39. Designing Data-Intensive Applications (ebook) by Martin Kleppmann (Goodreads Author)
- 40. Learning Redis (Kindle Edition) by Vinoo Das
- 41. Big Data Analytics with R and Hadoop by Vignesh Prajapati
- 42. Amazon Web Services in Action by Andreas Wittig
- 43. Neo4j in Action by Jonas Partner
- 44. Big Data: Principles and best practices of scalable realtime data systems by Nathan Marz
- 45. Redis: The Definitive Guide: Data modeling, caching, and messaging by Jay A. Kreibich
- 46. Professional NoSQL by Shashank Tiwari
- 47. Dynamodb Applied Design Patterns (ebook) by Uchit Vyas
- 48. MongoDB with Python and Ming (Kindle Edition) by Rick Copeland
- 49. MongoDB Aggregation Framework Principles and Examples (Kindle Edition) by John Lynn
- 50. Amazon DynamoDB Developer Guide (Kindle Edition) by Amazon Web Services
- 51. MongoDB Applied Design Patterns by Rick Copeland
- 52. Building Node Applications with MongoDB and Backbone by Mike Wilson
- 53. Ruby and MongoDB Web Development Beginner's Guide by Gautam Rege

Use to master the practical skills of the discipline. The materials are freely available on the Internet.

Educational content

5. Methods of mastering an educational discipline (educational component)

No. z/p	Type of study class	Description of the training session		
Topic 1. Introduction to non-relational databases				
1	Lecture 1. Introduction to non- relational databases.	Overview of course content, introduction to non- relational databases, SAR theorem		
		Tasks on IWS: item 6 No. 1		
2	Lecture 2. Disadvantages of the relational data model. Overview of alternative database models.	A brief overview of the advantages and disadvantages of the relational data model. Features of modern applications for which the relational data model is ineffective. Characteristics of aggregate databases as alternative models - advantages over the relational model.		
		Task on IWS: item 6 No. 2		
3	Lecture 3. Object-relational linking.	Peculiarities of applying the principles of object- oriented programming in the design of databases. Object-relational linking as a compromise way of applying object and relational concepts. Analysis of examples of creating information systems based on the application of object-relational linking.		
		Tasks on IWS:. item 6 No. 3		

4	Lecture 4. Basics of object databases.	Object-oriented model of databases. Storage of objects in the database. Technologies for accessing object databases.
		Task on IWS: item 6 No. 4
5	Lecture 5. Concepts of non- relational databases (part 1).	Aggregate approach to building database models. Non-relational data models: key-value data models, column family models.
		Tasks on IWS:. item 6 No. 5
6	Lecture 6. Concepts of non- relational databases (Part 2).	Non-relational data models: document data model, graph data model. Construction and comparison of non-relational data models on the example of creating a database for some application area.
		Tasks on IWS:. item 6 No. 6
7	Laboratory work 1. Introduction to DBMS Neo4j	Task: Using Neo4j DBMS software to implement a module for data processing of the subject area.
		Tasks on IWS: item 6 No. 7
	Topic 2. Uns	structured databases
8	Lecture 7. Unstructured databases.	A materialized representation. Data access modeling.
		Tasks on IWS:. item 6 No. 8
9	Lecture 8. Distribution models.	Single-server replication, fragmentation, master-slave replication, peer-to-peer replication.
		Task on IWS: p. 6 No. 9
10	Lecture 9. Consistency.	Consistency of updates, reading. Weakening of coherence. Weakening Durability. Quorums.
		Assignment on IWS: item 6 No. 10
11	Laboratory work 2. Getting to know the Redis DBMS	Task: Using Redis DBMS software, implement a module for data processing of the subject area.
		Tasks on IWS: item 6 No. 11.
12	Lecture 10. Stamps of versions.	Commercial and system transactions. Version stamps on several nodes.
		Assignment on IWS: item 6 No. 12
13	Lecture 11. Mapping-convolution.	Basics of the MapReduce pattern. Distribution and unification. Constituent calculations in the mapping- convolution scheme.
		Assignment on IWS: item 6 No. 13
14	Lecture 12. Non-relational	Big Data. Horizontal, vertical scaling
	databases in Big Data.	Task on IWS: item 6 No. 14
15	Lecture 13. Practical use of non- relational databases in applications.	Overview of several DBMS built on the basis of non- relational database models (Redis, MongoDB, Neo4J, Cassandra).
		Assignment on IWS: item 6 No. 15

Topic 3. NoSQL databases as information and search systems		
16	Lecture 14. NoSQL databases in the development of information and search systems	General rules for using and referring to NoSQL DBMS when developing information and search systems. Analysis of the application of software tools for the implementation of non-relational database models on an example.
		Assignment on IWS: item 6 No. 16
17	Lecture 15. Basic and additional features of MongoDB.	MongoDB capabilities: search, insert, update, cursors, JOIN analogues, profiling and web interface.
		Assignment on IWS: item 6 No. 17
18	Lecture 16. Basic and additional features of Riak	Riak Capabilities: CRUD, Links and MIME Types, REST, MIME Types, Mapreduce and Server Clusters
		Assignment on IWS: item 6 No. 18
19	Lab 3. Working with a NoSQL graph database in the Microsoft SQL Server 2017 DBMS environment	Task: to build a graph database for the selected subject area. Task on IWS: item 6 No. 19
20	Lecture 17. Basic and additional features of HBase	HBase Capabilities: CRUD and Table Administration, HBase Configuration and Shell, Loading Wikipedia, Bloom Compression and Filters, On-Disk Horizontal Scaling Engine, Building a Scanner.
21	Lecture 18. Basic and additional features of Neo4J	Features of Neo4J: Graphs, Groovy and CRUD operations, Gremlin , Pipelines and Vertices , Schemaless Social Network , REST Interface , Transactions , Backup Assignment on IWS: item 6 No. 21
Modular control work		

6. Independent work of students (IWS)

The discipline "Software for information and search systems 1. NoSQL databases" is based on independent preparations for classroom classes on theoretical and practical topics.

No.	The name of the topic submitted for independent processing	Number of hours	literature
1	Preparation for the lecture 1	1	1; 2; 3;
2	Preparation for lecture 2	1	1; 2; 3;
3	Preparation for the lecture 3	1	1; 2; 3;
4	Preparation for the lecture 4	1	1; 2; 3;
5	Preparation for the lecture 5	1	1; 2; 3;
6	Preparation for the lecture 6	1	1; 2; 3;

7	Preparation for laboratory work 1	5	1-5
8	Preparation for the lecture 7	1	1; 2; 3;
9	Preparation for the lecture 8	1	1; 2; 3;
10	Preparation for the lecture 9	1	1; 2; 3;
11	Preparation for laboratory work 2	5	1-5
12	Preparation for lecture 10	1	1; 2; 3;
13	Preparation for lecture 11	1	1; 2; 3;
14	Preparation for lecture 12	1	1; 2; 3;
15	Preparation for lecture 13	1	1; 2; 3;
16	Preparation for lecture 14	1	1; 2; 3;
17	Preparation for lecture 15	1	1; 2; 3;
18	Preparation for lecture 16	1	1; 2; 3;
19	Preparation for laboratory work 3	5	1-5
20	Preparation for lecture 17	1	1; 2; 3;
21	Preparation for lecture 18	1	1; 2; 3;
22	Preparation for modular control work	3	1-53
23	Preparation for the exam	30	1-53

Policy and control

7. Policy of academic discipline (educational component)

Attending lectures is mandatory.

Attending classes on laboratory work may be occasional and as needed for consultation/protection of laboratory work.

Rules of behavior in classes: activity, respect for those present, turning off phones.

Adherence to the policy of academic integrity.

Rules for the protection of laboratory work: work must be done in accordance with the tasks and according to the option.

8. Types of control and rating system for evaluating learning outcomes (RSO)

During the semester, students perform 3 laboratory works. The maximum number of points for each laboratory work: 13 points.

Points are awarded for:

- performance quality: 0-8 points;

- answer during job defense: 0-5 points.

Performance evaluation criteria:

8 points – the work is done qualitatively, in full; 5-7 points - the work is done qualitatively, in full, but has shortcomings; 1-4 points – the work is completed in full, but contains significant errors; 0 points - the work is not completed in full.

Answer evaluation criteria:

4-5 points – the answer is complete, well-argued;
1-3 points – there are significant errors in the answer;
0 points - there is no answer or the answer is incorrect.

The maximum number of points for performing and defending laboratory work: 15 points × 3 lab. do = 36 points.

The assignment for the modular test consists of 3 theoretical and 2 practical questions. The answer to each theoretical question is evaluated by 2 points. The answer to each practical question is evaluated by 4 points.

Evaluation criteria for each theoretical question: 2 points – the answer is correct, complete, well-argued; 1 points – there are significant errors in the answer; 0 points - there is no answer or the answer is incorrect.

Evaluation criteria for each practical question: 4 points – the answer is correct, complete, well-argued; 2-3 points - in general, the answer is correct, but has flaws; 1 points – there are significant errors in the answer; 0 points - there is no answer or the answer is incorrect.

The maximum number of points for a modular control work: 2 points × 3 questions + 4 points × 2 questions = 14 points.

The rating scale for the discipline is equal to:

R = RS = Rlab work + RMKR + Exam = 36 points + 14 points + 50 points = 100 points.

Calendar control: is conducted twice a semester as a monitoring of the current state of fulfillment of the syllabus requirements.

At the first certification (8th week), the student receives "credited" if his current rating is at least 7 points (50% of the maximum number of points a student can receive before the first certification).

At the second certification (14th week), the student receives "passed" if his current rating is at least 13 points (50% of the maximum number of points a student can receive before the second certification).

Semester control: exam

Conditions for admission to semester control:

With a semester rating (RC) of not less than 30 points and the enrollment of all the work of the computer workshop, the student is admitted to the exam. After passing the exam, a grade is assigned according to the table (Table of correspondence of rating points to grades on the university scale).

A prerequisite for admission to the exam is the completion and defense of all laboratory work.

Table of correspondence of rating points to grades on the university scale:

Scores	Rating
100-95	Perfectly
94-85	Very good
84-75	Fine
74-65	Satisfactorily
64-60	Enough
Less than 60	Unsatisfactorily
Admission conditions not met	Not allowed

9. Additional information on the discipline (educational component)

The list of questions submitted for semester control is given in Appendix 1.

Working program of the academic discipline (syllabus):

Folded assistant Radchenko K.O.

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)

- 1. Purpose and scope of XML. Node types. Syntax requirements. (well-formed).
- 2. Rules for forming a valid document. XML validation.
- 3. Types of analyzers (parsers) of XML documents.
- 4. XML DOM: graphical presentation of the document, program interface.
- 5. The XPath language. Syntax. Application of XPath in Web technologies.
- 6. The general syntax of the XSLT language. Language constructions. Scopes and libraries for converting XML documents.
- 7. The XQuery language. Syntax, language constructions (FLOWER, functions).
- 8. Scaling problems of distributed systems. Vertical and horizontal scaling. CAP theorem.
- 9. Types and areas of application of non-relational databases.
- 10. Replication and sharding in non-relational databases.
- 11. Fields of application of the redis DBMS. Basic data structures of redis. Examples of commands.
- 12. Approaches to organizing data scaling facilities in redis.
- 13. Commands for organizing the publish-subscribe algorithm and the message queue. Queue differences from publish-subscribe.
- 14. Big Data characteristics, Volume, Variety, Velocity, Veracity
- 15. Characteristics of document databases
- 16. Data model in MongoDB, basic data types, attached documents
- 17. Replication in MongoDB: tasks of master and secondary nodes. Types of secondary nodes.
- 18. Sharding in MongoDB: sharding key, sharding methods.
- 19. Cluster structure during sharding, roles of servers.
- 20. MongoDB insert, update, delete, find operations. Search in attached documents. Examples.
- 21. The aggregation pipeline in MongoDB: the main stages of projection, match, limit, sort, group, unwind. Examples.
- 22. Fields of application of graph databases, advantages and disadvantages of their use.
- 23. Data model in Neo4J graph DBMS. The role of model components.
- 24. Syntax of the Neo4J Cypher data access language.
- 25. Examples of using the MATCH, CREATE, RETURN sections in the Cypher language.
- 26. Databases of the Wide Column type. Differences from other non-relational DBMS, scope of application. Advantages and disadvantages.
- 27. Cassandra cluster architecture. Cluster components. Features of the CQL language.
- 28. Non-relational databases of the "search server" type (search engine). Features of application, differences from other DBMS, scope of application. Advantages and disadvantages.
- 29. DBMS elasicsearch: characteristics, advantages and disadvantages, data access language.
- *30. Key components of the elasicsearch search server (node, cluster, document, index). elasicsearch API feature.*