Russian Scholarship Project
open doors
Association "Global Universities"

Computer & Data Science
Program outline

This document outlines the scope of themes that may be included in the Olympiad test. They are grouped by areas and followed by recommended references in English.

The skill set of the winner of the Olympiad in Computer & Data Science

To win the Open Doors Olympiad in Computer & Data Science, you must have analytical, technological and research competencies. Those listed below are in line with the focal points of master's programs offered at Russian universities.

Analytical skills:

- knowledge, skills and abilities to analyze and use various sources of information, quantitative and qualitative methods for designing software, nodes of information transmission networks and network protocols; to analyze their characteristics, model and assess information security threats

Practice-oriented competencies:

- knowledge, skills and abilities to organize and plan research into the modeling of information and communication systems, communication channels, information threats; to prepare and act on research and technical documentation, scientific reports, reviews, reports and articles

Research competencies:

- knowledge, skills and abilities to apply and develop analytical information systems, to develop empirical, quantitative (semi-empirical) or qualitative models of processes and objects, to assess and interpret research findings

Competence	**The winner of the Open Doors Olympiad in Computer and Data Sciences must:**		
	know	**be able to**	**possess the skill in**
analytical	the theoretical foundations of computer science; the basic principles	analyze trends in the software market;	applying the mathematical foundations of computer science

	of algorithms and data structures; criteria for evaluating the effectiveness of algorithms and data structures; programming languages system programming languages, basic standards and programming interfaces; code formatting standards documents regulating information security (international standards and recommendations)		in software development and research using the theoretical foundations of information security in information security threat prevention and research
Practice-oriented	the principles of creating software for analysis, recognition and processing of information, as well as for digital signal processing systems;	design distributed information systems, their components and interaction protocols; design systems with parallel data processing, high-	using methods and algorithms for solving control problems and designing automation objects;

	state of the art tools for protecting information from unauthorized access basic principles of algorithms and data structures; criteria for evaluating the effectiveness of algorithms and data structures; programming language(s) system programming languages, basic standards and programming interfaces; code formatting standards; the principles of organization and operation of the computer; the functional structure of the computer; systems and formats of commands; the composition, purposes and	performance systems and their components; make a reasonable choice of standard information security tools when solving a specific problem; use information security tools; make a reasonable choice of tools to protect information from unauthorized access solve problems by creating program codes; use existing algorithms and data structures; evaluate the effectiveness of existing algorithms and data structures; create and read the program code; use tools and programming	setting up and using standard information security tools; performing the initial configuration and operation of information security tools; performing system security checks; employing existing methods and algorithms for solving recognition and data processing problems; using existing methods, algorithms and solutions for digital signal processing problems; reading program codes, using version control systems and tools; developing

	algorithms of arithmetic and logical operations; special machine operations; composition of micro-operations; ways to speed up the execution of operations; the organization and functioning principles of operating and control devices, methods and means of generating functional control signals; temporal characteristics of a hierarchical memory system, principles of organizing cache memory	environments, IDEs, SDKs and one or several related version control systems; formalize the program code according to standards deal with the forms and formats of presentation of numerical and non-numerical information, analyze microprograms of operations in a computer, determine the compatibility of microoperations;	system software using version control systems, instrumental environments and command line tools; using system utilities and developing multitasking (multiprocessing) and multithreaded system applications; program design; designing technical tools for information exchange in systems with a bus architecture, as well as for data exchange with remote peripheral devices; designing effective command systems (formats, addressing methods, lists of operations);
Research	basic programming	apply optimization	describing

	paradigms; existing approaches to the verification of software models; promising research methods and ways to solve professional problems based on global trends in computer and information technology	methods when solving professional problems	information processes and objects using basic information-theoretic models; interpreting the results of model analysis, taking into account the capabilities and limitations of the methods used; creating new algorithms and data structures; evaluating the effectiveness of new algorithms and data structures

Content

Section 1. Applied mathematics

1. Comparing the order of growth of typical mathematical functions. Big-O asymptotic notation. [Rosen2018: section 3.2] [Cormen2009: section 3]
2. The Taylor expansion of a function. [Apostol1991, chapter 7]
3. Linear space: definition, examples. Dimension of a linear space. Matrices: rank, determinant, inverse matrix. Eigenvalues and eigenvectors. [Cormen2009:appendix D] [Strang2016]
4. Modular arithmetic. Fermat's little theorem. Finite fields of residues: definition, construction, performing computations in finite fields. The Euclidean

Algorithm. GCDs as Linear Combinations. Bezout's Theorem. [Rosen2018: sections 4.1—4.3]

5. Number systems (decimal, binary, hexadecimal, octal). Converting between representations in different bases. [Rosen2018, section 4.2]
6. Boolean algebra. Standard boolean operations: conjunction, disjunction, negation, implication, equivalence, exclusive OR, Sheffer stroke. Standard representations of boolean functions: DNF, CNF, simplifying expressions using distributive property, De Morgan rules, absorption laws. [Wegener1987: sections 1.1, 1.2] [Rosen2018: sections 12.1, 12.2]
7. Boolean circuits. Complexity measures: circuit size and depth. Construction of Boolean circuits for Boolean functions. [Wegener1987: sections 1.2, 1.3] [Rosen2018: sections 12.3, 12.4]
8. Predicates. Logical inference. Representing predicates with quantified formulas. [Rosen2018: chapter 1]
9. Rules of inference. Resolution. [Rosen2018: section 1.6.5]
10. Finite automata. Non-deterministic finite automata. Regular languages. [Hopcroft2013: sections 2.2, 2.3, 2.5] [Rosen2018: sections 13.2, 13.3]
11. Regular expressions. Converting between a regular expression and an automaton. POSIX Extended regular expressions. [Hopcroft2013: sections 3.1, 3.2] [Rosen2018: section 13.4] [POSIXERE]
12. Recursive definitions and structural induction. [Hopcroft2013: section 1.4] [Rosen2018: section 5.3]
13. Divisibility properties. GCD and LCM. Euclid's algorithm. [Rosen2018: section 4.3]
14. Probability: basic definitions and properties. [Stirzaker2003, chapter 1]
15. Distributions. Probability density functions. [Stirzaker2003, section 7.1]
16. Law of total probability. [Stirzaker2003, sections 2.1, 2.2]
17. Bayes' theorem. [Rosen2018: section 7.3]
18. Basic counting. Permutations and combinations with and without replacement. [Rosen2018: sections 6.1—6.3] [Cormen2009: Appendix C]
19. Asymptotic growth of combinatorial numbers. [Cormen2009: Appendix C]
20. Generating combinations and permutations. [Rosen2018: section 6.6]
21. Inclusion-exclusion formula. [Rosen2018: sections 8.5, 8.6]
22. Graphs: undirected, directed, bipartite, complete. Subgraphs: induced subgraphs, spanning trees. Distances in graphs. Depth-first and breadth-first graph traversal. [Rosen2018: chapter 10]
23. Trees. Minimal spanning trees problem. [Rosen2018: chapter 11]
24. Planar graphs. Euler's formula [Rosen2018: section 10.7.2]

Section 2. Software engineering

1. The SQL language, basic concepts
2. The purpose and main functions of the assembler
3. The architectural principles of von Neumann computers. The main types of computers with non-von Neumann architectures
4. The central processing unit: its functions and composition
5. Computer architecture. CISC and RISC architectures
6. The Intel x86 processor programming model
7. Intel x86 processor addressing modes
8. The concept of pipelined and superscalar processing of an instruction stream
9. The concept and classification of hardware interfaces
10. The purpose and main functions of the interrupt system
11. Software and hardware interrupts and differences between them
12. Function calls, call stacks, calling conventions
13. The hierarchical organization of computer memory
14. The notion and concepts of virtual memory
15. Paging and segment memory organization
16. Algorithms for replacing virtual memory pages
17. Process management in operating systems
18. Memory management in operating systems
19. Memory organization. RAM. Stack memory. Types of stacks. Memory segmentation
20. Purpose and functions of operating systems. Classification of operating systems. Operating system architecture. Windows and UNIX
21. Basic principles of data management and file systems
22. The purpose and principles of video adapter operation
23. The arithmetic foundations of computers. Positional number systems and operations in them
24. Data storage structure on external storage media
25. Hardware and software for information processes
26. Principles of modular, component, object-oriented design
27. Design patterns
28. Software system modeling, UML. Modern approaches to automatic program synthesis
29. Basic principles of object-oriented programming
30. C ++ programming language. Classes. Constructors and destructors
31. Inline functions. Inline functions in a class definition
32. The THIS pointer
33. The NEW and DELETE operators

34. References to objects
35. Function overloading. Default arguments. Operator overloading
36. Inheritance. Protected members of the class. Multiple inheritance. Virtual base classes
37. Formatted I / O. I / O manipulators. Custom I / O functions. File I / O. Unformatted I / O
38. Virtual functions. Derived class pointers
39. Exception handling
40. Dynamic identification and type casting
41. Namespaces
42. Static members of a class. Constant and modifiable members of a class
43. Library of standard templates. Container classes
44. Library of standard templates. Iterators
45. Library of standard templates. Algorithms
46. ORM (Object-Relational Mapping)
47. Automatic garbage collection mechanisms
48. Interfaces and Collections
49. Structured exception handling

Section 3. Information systems and computing

1. Information: collecting, transmitting, processing, accumulation, measurement; information units. Information and entropy
2. Data structures
3. Signal coding and quantization
4. The concept of the information system and information technology
5. Hardware and software information technology tools
6. Analog and digital information processing. Data processing devices and their characteristics
7. Multitasking
8. Parallel processing. The main classes of parallel systems
9. The concept of a system. Systems in engineering, economics and nature
10. Types of systems. The control object and the control system. Information Feedback. The subject of engineering cybernetics and information theory
11. The methodological framework for modeling. Axioms of modeling theory Characteristics of systems models. The objectives and problems of systems modeling

12. Types of control systems. Automatic and automated systems. Control actions. The concept of homeostasis. Adaptation problems. Information processes in systems
13. Modeling as a research method in cybernetics. Types of models. Models of technical, biological and socio-economic systems. The "black box" concept. The problem of identification. Model validity
14. Pragmatic, semantic and syntactic aspects of information
15. Programming languages. Procedure-oriented and object-oriented programming. Ways of describing algorithms. The Unified system of program documentation
16. Software life cycle
17. Programming styles. Procedural, structural style and object-oriented programming
18. The processor and the instruction system. The structure diagram of a microprocessor. Interaction between the functional blocks of a processor in instruction execution
19. Types of machine commands. Machine instruction formats. Addressing methods. Structure of machine instructions. Logical elements of instructions. Types of assembler instructions
20. Classification of personal computer memory types. Hard drive structure. Logical organization of information
21. The principles of the structural and functional organization of computer networks. Packet-switched networks. TCP/IP protocol stack. Addressing in IP networks
22. Multithreading: basic concepts. Methods of synchronization in multithreaded applications
23. Process interaction. Shared memory, synchronization tools. Message queues and other means of data exchange
24. The distribution, use and management of computer system resources Basic scheduling approaches and algorithms. Real-time and split-time systems
25. The reference model of open systems interaction (ISO OSI model) and its purpose. Data encapsulation. Layers of the OSI reference model
26. Collision. Collision domain. Narrowcast, multicast and broadcast
27. Broadcast domain. Hubs, switches and bridges and their operation in collision and broadcast domains

Section 4. Information security

1. Basic concepts of information security. Key factors affecting information security
2. Major international information security standards, their purpose and scope
3. Information security policy, its place and role in organization management
4. Personal data protection in organization management; legal regulation and protection requirements
5. Information security threats. Approaches to modeling information threats and attackers
6. Software vulnerabilities. Criteria and scoring system for vulnerability severity assessment. Open source vulnerability databases
7. Approaches to information risk management. Risk assessment. Risk scales and measurement criteria
8. Security information and event management (SIEM) systems. Design principles, tasks and goals of SIEM systems
9. Intrusion, prevention and detection systems in computer systems and networks; their purpose, goals and design principles
10. Network firewall systems: principles and tasks
11. Virtual Private Network (VPN): purpose, design principles and selection criteria
12. Access control models and systems
13. Identification, authentication and authorization. Authentication and authorization techniques
14. Steganography: the purpose and tasks. Techniques for embedding hidden information in executable files
15. Symmetric encryption systems. Block ciphers, their operation modes
16. Modern standards and protocols of symmetric encryption
17. Public key cryptosystems. The principles of construction. Modern standards and public-key encryption protocols
18. Digital signature protocols and schemes. One-way function, design principles and implementation variants of digital signature schemes
19. Public-key infrastructure, certification centers and digital certificates

Section 5. Data preprocessing and analysis

1. Decision support systems (DSS)
2. OLTP systems. Transaction concept
3. The concept of data warehouse (DW). Their purpose. Architecture
4. Physical DW. Virtual HD. Data marts (DM)

5. Data transfer. The ETL process. Data cleansing. DW and analysis
6. Types of data warehouses. Their purposes and architectures.
7. Multidimensional data models. Representation of data as a multidimensional cube
8. The OLAP system. OLAP system architecture. The star and snowflake schemas
9. Types of OLAP systems. Multidimensional OLAP (MOLAP). Relational OLAP (ROLAP). Hybrid OLAP (HOLAP)
10. Machine learning. Definition. Purposes
11. Machine learning functions: classification, regression, frequent set search, clustering
12. Machine learning models. Predictive and descriptive machine learning models. Machine learning methods
13. Knowledge discovery process. Stages of data mining
14. Classification models: classification rules, decision trees, regression functions, neural networks
15. Classification algorithms: Naive Base, C 4.5, BackProp, Support Vector Machine, etc.
16. Clustering models: centroid, hierarchical, density, graph, etc. Distance function
17. Clustering algorithms: hierarchical, k-Means, DBScan, SOM
18. Frequent sets. Association rules. Support. Credibility
19. Algorithms for frequent set search: Apriori, FPG, etc.
20. Deep learning. Definition. Deep learning and machine learning
21. The concept of the neuron and the neural network
22. Types of neural networks
23. BackPropagation algorithm
24. SGD algorithm. Loss function
25. Definition and properties of Big Data
26.

Recommended literature

Section1. Applied mathematics

1. [Hopcroft2013] Hopcroft, John E.; Motwani, Rajeev; Ullman, Jeffrey D. (2013). Introduction to Automata Theory, Languages, and Computation (3rd ed.). Pearson. ISBN 978-1292039053.

2. [Wegener1987] Wegener, Ingo. (1987) The Complexity of Boolean Functions. John Wiley and Sons. ISBN 3-519-02107-2. Free electronic edition: https://eccc.weizmann.ac.il/static/books/The_Complexity_of_Boolean_Functions/
3. [Rosen2018] Rosen, Kenneth. (2018) Discrete Mathematics and Its Applications. McGraw-Hill Education; 8 edition. ISBN: 978-1259676512
4. [Cormen2009] Thomas H. Cormen. Charles E. Leiserson. Ronald L. Rivest. Clifford Stein. (2009) Introduction to Algorithms, 3rd Edition (The MIT Press) ISBN 978-0262033848
5. [Strang2016] Strang, Gilbert.(2016) Introduction to Linear Algebra, Fifth Edition. Wellesley-Cambridge Press. ISBN: 978-0980232776

Section 2. Software engineering

1. Structured computer organization/ Andrew S. Tanenbaum, Todd Austin — 6th edition, sections:
   - 2.1, 2.2, 2.4
   - 3.3, 3.5, 3.7
   - 5.1, 5.4, 5.5, 5.6
   - 6.1
   - 7.1
2. Modern operating systems/ Andrew S. Tanenbaum, Herbert Bos — 4th edition, sections:
   - 1.1, 1.4, 1.5
   - 2.1 2.2
   - 3.1, 3.2, 3.3, 3.4
   - 4.1, 4.2, 4.3, 4.4, 4.5
   - 5.6
   - 10.1, 10.2
   - 11.1, 11.3
3. Learning SQL/ Alan Beaulieu — 2nd edition, chapters: 1-5
4. The C programming language/ Brian Kernighan, Dennis Ritchie — 2nd edition, chapters: 1-7
5. Structured Computer Organization/Andrew Tanenbaum, sections:
   - 2.1, 2.2, 2.4
   - 3.3, 3.5, 3.7
   - 5.1, 5.4, 5.5, 5.6
   - 6.1
   - 7.1
6. Modern Operating Systems / Andrew Tanenbaum, Herbert Bos, sections:
   - 1.1, 1.4, 1.5
   - 2.1 2.2
   - 3.1, 3.2, 3.3, 3.4
   - 4.1, 4.2, 4.3, 4.4, 4.5
   - 5.6

- 10.1, 10.2
- 11.1, 11.3

7. Learning SQL/ Alan Beaulieu, chapters: 1-5
8. The C Programming Language/ Brian Kernighan and Dennis Ritchie, chapters: 1-7
9. Programming: principles and practice using C++/ Bjarne Stroustrup, chapters: 4, 5, 8, 9, 14, 17, 18, 19, 20, 21, Appendix A
10. C++ Primer Plus Sixth Edition / Stephen Prata, chapters: 4, 7, 9, 10, 11, 12, 13, 16,
11. Design Patterns: Elements of Reusable Object-Oriented Software / [Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, Grady Booch, chapters: 1 - 5
12. Learning UML 2.0. / Kim Hamilton, Russell Miles, chapters: 1-3

Section 3. Information systems and computing

1. Rainer R. K., Prince B., Cegielski C. G. Introduction to Information Systems: R. Kelly Rainer, Brad Prince, Casey Cegielski. – John Wiley & Sons Singapore Pte. Limited, 2015.
2. Tanenbaum A. S. et al. Computer networks /Prentice-Hall international editions. – 1996.-813 p.
3. Hopcroft J. E., Motwani R., Ullman J. D. Introduction to automata theory, languages, and computation. 2006

Section 4. Information security

1. M. Rhodes-Ousley. Information Security: The Complete Reference, Second Edition 2nd Edition Publisher : McGraw-Hill Education; 2nd edition. 2013. 896 pages
2. C Eagle, K. Nance. The Ghidra Book: The Definitive Guide No Starch Press (September 1, 2020), 608 p.
3. W. Easttom. Computer Security Fundamentals. 4th Edition. Pearson IT Certification; 4th edition (October 22, 2019). 512 pages
4. C. P. Schultz, B. Perciaccante. Kali Linux Cookbook - Second Edition: Effective penetration testing solutions. Packt Publishing; 2nd Revised edition (September 12, 2017). 438 pages
5. N. Ferguson, B. Schneier, T. Kohno. Cryptography Engineering: Design Principles and Practical Applications 1st Edition, Wiley, 2011.P. 386.

Section 5. Data analysis and machine learning

1. Ian W., Elbe F. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Department of computer science University of Waikato. 3th ed, 2011
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. Springer. 2001
3. Alpaydin, Ethem (2010). Introduction to Machine Learning. London: The MIT Press. ISBN 978-0-262-01243-0. Retrieved 4 February 2017.

4. Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2
5. Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). Foundations of Machine Learning. USA, Massachusetts: MIT Press. ISBN 9780262018258.

Recommended online courses

Section 1. Applied mathematics

1. Calculus and Optimization for Machine Learning
2. https://www.coursera.org/learn/calculus-and-optimization-for-machine-learning
3. Mathematical Thinking in Computer Science
4. https://www.coursera.org/learn/what-is-a-proof
5. Combinatorics and Probability
6. https://www.coursera.org/learn/combinatorics
7. Introduction to Graph Theory
8. https://www.coursera.org/learn/graphs
9. Number Theory and Cryptography
10. https://www.coursera.org/learn/number-theory-cryptography

Section 2. Software engineering

1. Computer Architecture https://www.coursera.org/learn/comparch
11. Build a Modern Computer from First Principles: From N and to Tetris https://www.coursera.org/learn/build-a-computer
12. Coding for Everyone: C and C++ Specialization https://www.coursera.org/specializations/coding-for-everyone
13. Design Patterns https://www.coursera.org/learn/design-patterns
14. C++ Data Structures in the STL https://www.coursera.org/projects/cpp-data-structures-in-the-stl
15. Object-Oriented Data Structures in C++ https://www.coursera.org/learn/cs-fundamentals-1
16. Software Design and Architecture https://www.coursera.org/specializations/software-design-architecture
17. Introduction to Structured Query Language (SQL) https://www.coursera.org/learn/intro-sql

Section 3. Information systems and computing

1. Information Systems, URL: https://www.coursera.org/specializations/information-systems

2. Fundamentals of Network Communication, https://www.coursera.org/learn/fundamentals-network-communications
3. Introduction to Computer Science and Programming, https://www.coursera.org/specializations/introduction-computer-science-programming
4. TCP/IP and Advanced Topics URL: https://www.coursera.org/learn/tcp-ip-advanced?specialization=computer-communications
5. Packet Switching Networks and Algorithms URL: https://www.coursera.org/learn/packet-switching-networks-algorithms?specialization=computer-communications

Section 4. Information security

1. Information Security: Context and Introduction. University of London. Skill level - Beginner, URL: https://www.coursera.org/learn/information-security-data
2. The Foundations of Cybersecurity. University System of Georgia. Skill level - Beginner. URL https://www.coursera.org/learn/foundations-cybersecurity
3. Real-Time Cyber Threat Detection and Mitigation New York University. Skill level - Intermediate. URL: https://www.coursera.org/learn/real-time-cyber-threat-detection
4. Cryptography. University of Maryland, College Park. URL: https://www.coursera.org/learn/cryptography
5. Software security. University of Maryland https://www.coursera.org/learn/software-security

Section 5. Data analysis and machine learning

1. Free Machine Learning Course (fast.ai) https://www.fast.ai/
2. Machine Learning Course by Stanford University (Coursera)
3. https://www.coursera.org/learn/machine-learning
4. Deep Learning Course (deeplearning.ai) https://www.coursera.org/specializations/deep-learning
5. Free Machine Learning Data Science Course (Harvard University) https://www.edx.org/professional-certificate/harvardx-data-science
6. Free Machine Learning Introduction Course (Udacity) https://www.udacity.com/course/intro-to-machine-learning-nanodegree--nd229
7. Machine Learning Course (Stanford School of Engineering) https://online.stanford.edu/courses/cs229-machine-learning
8. Free Machine Learning Courses (edX) https://www.edx.org/learn/machine-learning