ДОКТОРСКА ПРОГРАМА
„ИНФОРМАЦИОННИ ТЕХНОЛОГИИ (ГОЛЕМИ ОБЕМИ ОТ ДАННИ)“

Професионално направление 4.6 Информатика и компютърни науки

КОНСПЕКТ

за конкурсен изпит за прием на докторанти

Пояснение: Конспектът е предназначен за конкурсен изпит за прием на докторанти. Той включва общ модул, насочен към фундаментална подготовка в областта на информатиката и информационните технологии, и специализиран модул, насочен към тематиката на големите обеми от данни. Изпитът има за цел да провери не само фактологични знания, но и способност за систематично мислене, анализ на проблем, избор на подходящ модел, аргументиране на техническо решение и ориентация към изследователска работа.

На писмения изпит се задават два въпроса: един от общия модул и един от специализирания модул. Устният изпит представлява събеседване по въпросите от конспекта, като може да бъде поставена кратка аналитична задача, казус или задача за сравнение на подходи. Не се изисква владеене на конкретна софтуерна платформа, език или индустриален инструмент, освен когато те се разглеждат като примери за по-общи принципи.

I. ОБЩ МОДУЛ: ОСНОВИ НА ИНФОРМАЦИОННИТЕ ТЕХНОЛОГИИ

1. Компютърни архитектури, операционни системи и управление на ресурси.

Анотация: Основни компоненти на компютърните системи; процесори, памет, входно-изходни устройства и йерархии на паметта;

Литература: [1], [2]

2. Операционни системи. Управление на ресурсите

Основни типове операционни системи. Файлови системи и тяхната логическа организация и физическо представяне. UNIX, MS Windows, MS DOS, Linux – обзор и сравнителна характеристика. Процеси, нишки, синхронизация и управление на ресурси; роля на операционната система при изпълнение на приложения; базови принципи на виртуализацията, паралелното изпълнение и използването на изчислителни ресурси при обработка на данни и обучение на модели.

Литература: [3]

3. Езици за програмиране (ЕП).

Анотация: Класификация на ЕП. Синтаксис и семантика на ЕП. Транслатори, компилатори и интерпретатори. Обща и сравнителна характеристика. Структурни, функционални, логически, обектно ориентирани ЕП.

Литература: [4]–[6]

4. Обектно-ориентиран подход за разработка на софтуер

Анотация: Същност и предимства. Основни понятия – обекти, класове, връзки, йерархии. Характеристики на обектно-ориентирания анализ и проектиране (абстракция, капсулация, наследяване, полиморфизъм, конкурентност, типизация).

Литература: [4], [5], [7]

5. Структури от данни

Анотация: Основни структури от данни - масиви, списъци, стекове, опашки, дървета, хеш-таблици и графи; алгоритми за търсене, сортиране, обхождане и оптимизация; Реализация. Основвни операции. Атрибути на данни. Примитивни и непримитивни данни. Абстрактни структтури от данни.

Литература: [4], [8]

6. Алгоритми и сложност

Анотация: Формално и неформално определение. Свойства на алгоритмите. Детерминирани и недетерминирани алгоритми. Сложност и оптималност на алгоритмите. Изчислимост, P и NP класове от задачи. Времева и пространствена сложност; рекурсия, динамично програмиране и жадни алгоритми; значение на алгоритмичната ефективност при анализ на данни, търсене, обучение и вземане на решения.

Литература: [8], [17]

7. Разпределени системи, мрежи и облачни среди.

Анотация: Основни принципи на компютърните мрежи и комуникационните протоколи; клиент-сървърни и многослойни архитектури; разпределение на изчисления и данни; латентност, пропускателност, отказо-устойчивост и мащабируемост; базови принципи на облачните изчисления и виртуализацията; значение на изчислителната инфраструктура при обучение и внедряване на системи за изкуствен интелект (ИИ).

Литература: [9]–[11]

8. Бази от данни и управление на информация.

Анотация: Релационен модел, схеми, ключове и нормализация; език за структурирани заявки (Structured Query Language, SQL) и основни операции върху данни; индекси, транзакции и консистентност; основни нерелационни модели на данни; съхранение, извличане и подготовка на данни за анализ; значение на качеството, структурата и проследимостта на данните при изграждане на интелигентни системи.

Литература: [11]–[13]

9. Интернет на нещата (Internet of Things).

Анотация: Основни технологии и стандарти. Основни приложения. Научни изследвания и политики в Европейския съюз.

Литература: [14]–[16]

10. Изкуствен интелект.

Анотация: историческо развитие, основни парадигми, методи, приложения и съвременни предизвикателства.

Литература: [17], [18]

II. СПЕЦИАЛИЗИРАН МОДУЛ: ГОЛЕМИ ОБЕМИ ОТ ДАННИ

9. Понятие, характеристики и жизнен цикъл на големите обеми от данни.

Анотация: Характеристики на големите обеми от данни - обем, скорост, разнообразие, достоверност и стойност; източници на структурирани, полуструктурирани и неструктурирани данни; жизнен цикъл на данните - събиране, съхранение, обработка, анализ, архивиране и унищожаване; връзка между изследователски въпрос, бизнес задача и архитектура за обработка на данни.

Литература: [11], [13]

10. Архитектури за съхранение и организация на големи данни.

Анотация: Файлови системи, обектни хранилища и разпределено съхранение; партициониране, репликация и отказоустойчивост; сравнение между data warehouse, data lake и lakehouse подходи; управление на сурови, обработени и доверени данни; значение на метаданните, версиите и проследимостта; критерии за избор на архитектура според обем, достъп, цена и надеждност.

Литература: [11], [13], [19]

11. Нерелационни и разпределени модели на данни.

Анотация: ключ-стойност модели, документни, колонно-фамилни и графови модели; денормализация и проектиране според заявките; шардване, репликация, консистентност и наличност; компромис между съгласуваност, наличност и устойчивост при разделяне на мрежата (Consistency, Availability, Partition Tolerance, CAP); ново поколение релационни системи за управление на бази от данни (NewSQL системи) като опит за съчетаване на хоризонтално мащабиране и транзакционни гаранции; критерии за избор на модел при различни типове натоварване.

Литература: [11], [20], [21]

12. Пакетна и поточна обработка на големи данни.

Анотация: Разлика между пакетна, микропакетна и поточна обработка; основни идеи на MapReduce, модели с насочени ациклични графи (Directed Acyclic Graphs) и разпределено изпълнение; обработка на събития в реално време; прозорци, време на събитие, закъснели данни и състояние; надеждност, повторяемост и възстановяване при грешки; типични приложения в мониторинг, Интернет на нещата (Internet of Things, IoT), анализ на потребителско поведение и оперативна аналитика.

Литература: [11], [22], [23]

13. Интеграция, качество и управление на данните.

Анотация: ETL и ELT процеси; интеграция на данни от различни източници; валидиране, профилиране и контрол на качеството; управление на метаданни, проследимост на произхода на данните и каталогизиране; роли и отговорности при управление на данни; принципи за откриваемост, достъпност, оперативна съвместимост и повторна използваемост на данните (Findable Accessible Interoperable Reusable, FAIR) в изследователски контекст; връзка между качество на данните, доверие в анализа и възпроизводимост на резултатите.

Литература: [13], [24]

14. Аналитични заявки, управленска аналитика и визуален анализ върху големи данни.

Анотация: Аналитични натоварвания и подходи многомерен анализ (Online Analytical Processing, OLAP); агрегиране, индексиране, партициониране и материализирани изгледи; основи на оптимизацията на аналитични заявки; приблизително изпълнение на аналитични заявки и извадкови методи като подходи при големи обеми от данни; изграждане на показатели, табла и визуализации; интерпретируемо представяне на резултати за научни, управленски и приложни цели.

Литература: [13], [21], [26]

15. Машинно обучение, сигурност и възпроизводимост при работа с големи данни.

Анотация: Подготовка на данни и признаци в мащаб; разделяне на данни при времеви, групирани и зависими наблюдения; особености на машинното обучение върху големи набори; управление на експерименти и проследимост между данни, модели и резултати; защита на лични данни, минимизация, псевдонимизация и риск от повторна идентификация; значение на възпроизводимостта и документацията при научна работа с големи данни.

Литература: [10], [24], [25]

Литература

[1] J. L. Hennessy, D. A. Patterson, and C. Kozyrakis, *Computer Architecture: A Quantitative Approach*, 7th ed. Cambridge, MA, USA: Morgan Kaufmann, 2025. [Online]. Available: https://www.amazon.com/Computer-Architecture-Quantitative-Approach-Kaufmann/dp/0443154066

[2] D. A. Patterson and J. L. Hennessy, *Computer Organization and Design RISC-V Edition: The Hardware/Software Interface*, 2nd ed. Cambridge, MA, USA: Morgan Kaufmann, 2020. [Online]. Available: https://www.amazon.com/Computer-Organization-Design-RISC-V-Architecture/dp/0128203315

[3] A. Silberschatz, P. B. Galvin, and G. Gagne, *Operating System Concepts*, 10th ed. Hoboken, NJ, USA: Wiley, 2018. [Online]. Available: https://www.amazon.com/Operating-System-Concepts/dp/1119439256

[4] M. L. Scott and J. Aldrich, *Programming Language Pragmatics*, 5th ed. Cambridge, MA, USA: Morgan Kaufmann, 2025. [Online]. Available: https://www.amazon.com/Programming-Language-Pragmatics-Michael-Scott/dp/0323999662

[5] M. Gabbrielli and S. Martini, *Programming Languages: Principles and Paradigms*, 2nd ed. Cham, Switzerland: Springer, 2023. [Online]. Available: https://www.amazon.com/Programming-Languages-Principles-Paradigms-Undergraduate/dp/3031341430

[6] K. D. Cooper and L. Torczon, *Engineering a Compiler*, 3rd ed. Cambridge, MA, USA: Morgan Kaufmann, 2022. [Online]. Available: https://www.amazon.com/Engineering-Compiler-Keith-D-Cooper/dp/0128154128

[7] I. Sommerville, *Software Engineering*, 10th ed. Boston, MA, USA: Pearson, 2015. [Online]. Available: https://www.amazon.com/Software-Engineering-10th-Ian-Sommerville/dp/0133943038

[8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 4th ed. Cambridge, MA, USA: MIT Press, 2022. [Online]. Available: https://www.amazon.com/Introduction-Algorithms-fourth-Thomas-Cormen/dp/026204630X

[9] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 8th ed. Boston, MA, USA: Pearson, 2021. [Online]. Available: https://www.amazon.com/Computer-Networking-James-Kurose/dp/0136681557

[10] C. Huyen, *Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2022. [Online]. Available: https://www.amazon.com/Designing-Machine-Learning-Systems-Production-Ready/dp/1098107969

[11] M. Kleppmann and C. Riccomini, *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2026. [Online]. Available: https://www.amazon.com/Designing-Data-Intensive-Applications-Reliable-Maintainable/dp/1098119061

[12] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database System Concepts*, 7th ed. New York, NY, USA: McGraw Hill, 2019. [Online]. Available: https://www.amazon.com/Database-Concepts-Abraham-Silberschatz-Professor/dp/0078022150

[13] J. Reis and M. Housley, *Fundamentals of Data Engineering: Plan and Build Robust Data Systems*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2022. [Online]. Available: https://www.amazon.com/Fundamentals-Data-Engineering-Robust-Systems/dp/1098108302

[14] S. Ziegler, R. Radócz, A. Quesada Rodriguez, and S. N. Matheu Garcia, Eds., *Springer Handbook of Internet of Things*, Cham, Switzerland: Springer, 2024. [Online]. Available: https://www.amazon.com/Springer-Handbook-Internet-Things-Handbooks/dp/3031396499

[15] R. C. Sofia and J. Soldatos, Eds., *Shaping the Future of IoT with Edge Intelligence: How Edge Computing Enables the Next Generation of IoT Applications*, River Publishers, 2023. [Online]. Available: https://a.co/d/015vP1QX

[16] P. G. Chiara, *The Internet of Things and EU Law: Cybersecurity, Privacy and Data Protection Challenges*, Cham, Switzerland: Springer, 2024. [Online]. Available: https://www.amazon.com/Internet-Things-Law-Cybersecurity-Protection/dp/3031676629

[17] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Boston, MA, USA: Pearson, 2020. [Online]. Available: https://www.amazon.com/Artificial-Intelligence-A-Modern-Approach/dp/0134610997

[18] V. Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, 1st ed. Cham, Switzerland: Springer, 2020. [Online]. Available: https://www.amazon.com/Responsible-Artificial-Intelligence-Foundations-Algorithms/dp/303030373X

[19] J. Serra, *Deciphering Data Architectures: Choosing Between a Modern Data Warehouse, Data Fabric, Data Lakehouse, and Data Mesh*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2024. [Online]. Available: https://www.amazon.com/Deciphering-Data-Architectures-Warehouse-Lakehouse/dp/1098150767

[20] P. J. Sadalage and M. Fowler, *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*, 1st ed. Boston, MA, USA: Addison-Wesley, 2012. [Online]. Available: https://www.amazon.com/NoSQL-Distilled-Emerging-Polyglot-Persistence/dp/0321826620

[21] A. Petrov, *Database Internals: A Deep Dive into How Distributed Data Systems Work*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2019. [Online]. Available: https://www.amazon.com/Database-Internals-Deep-Distributed-Systems/dp/1492040347

[22] J. S. Damji, B. Wenig, T. Das, and D. Lee, *Learning Spark: Lightning-Fast Data Analytics*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2020. [Online]. Available: https://www.amazon.com/Learning-Spark-Jules-Damji/dp/1492050040

[23] T. Akidau, S. Chernyak, and R. Lax, *Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2018. [Online]. Available: https://www.amazon.com/Streaming-Systems-Where-Large-Scale-Processing/dp/1491983876

[24] E. Eryurek, U. Gilad, V. Lakshmanan, and J. Kibunguchy-Grant, *Data Governance: The Definitive Guide: People, Processes, and Tools to Operationalize Data Trustworthiness*, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2021. [Online]. Available: https://www.amazon.com/Data-Governance-Definitive-Operationalize-Trustworthiness/dp/1492063495

[25] R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*, 3rd ed. Indianapolis, IN, USA: Wiley, 2020. [Online]. Available: https://www.amazon.com/Security-Engineering-Building-Dependable-Distributed/dp/1119642787

[26] P. Bruce, A. Bruce, and P. Gedeck, *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2020. [Online]. Available: https://www.amazon.com/Practical-Statistics-Data-Scientists-Essential/dp/149207294X