

Вопросы к государственному экзамену (обновлено 2026 г.)

Магистерская программа «Большие данные: инфраструктуры и методы решения задач»

Программа госэкзамена по направлению 01.04.02 ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА: общая часть

1. Модель цепной реакции в диффузионном приближении. Расчет критической массы реактора.
2. Модель Лотки-Вольтерра. Периодические колебания численности популяций.
3. Раскраски графов, хроматическое число графа. Критерий двухцветности графа. Теорема об оценке хроматического числа графа. Теорема Брукса (только формулировка). ([1] стр. 152-153, 32)
4. Наследственные свойства графов. Теорема об оценке наибольшего числа ребер в графе с наследственным свойством. Теорема о наибольшем числе ребер в графе без треугольников. Теорема Турана (только формулировка). [1] (Литература. Харари Ф. Теория графов. М.: Мир, 1973.)
5. Симплекс-метод для канонической задачи линейного программирования: идея метода и ее реализация, выбор стартовой угловой точки. См. материалы курса на сайте
6. Итерационные методы минимизации: скорейшего спуска, проекции градиента и Ньютона. См. материалы курса на сайте
7. Алгоритмы планирования процессов в современных ОС.
8. Методы синхронизации процессов и методы предотвращения тупиковых ситуаций в операционных системах.
9. Классификации архитектур вычислительных систем. Способы организации высокопроизводительных систем и основные принципы функционирования. Характеристики производительности, реальная и пиковая производительность, ускорение и эффективность.
10. Основные средства разработки для систем с общей и распределенной памятью. Основные характеристики пакетов OpenMP, Posix Threads, MPI, поддержка многопоточности в современном C++.
11. Эталонная модель Взаимодействия Открытых Систем (OSI). Функции уровней. Функции подуровня управления доступом к среде в многоабонентской сети. Назначение устройств объединения сетей: концентраторов, мостов, коммутаторов, маршрутизаторов.
12. Маршрутизация в глобальной компьютерной сети Интернет. Таблицы маршрутизации. Методы продвижения дейтаграмм (Forwarding). Задача выбора маршрутов (Routing): основные алгоритмы, их достоинства и недостатки.

Дополнительная часть.

1. Архитектура распределенной инфраструктуры Hadoop. Парадигма распределенного программирования MapReduce. Архитектура YARN.
 - *Chuck Lam. Hadoop in Action Tom White. Hadoop: The Definitive Guide, 3rd Edition*
 - *IBM Course: "Hadoop Fundamentals" (BD001)*
<https://bigdatauniversity.com/learn/hadoop/>
 - *Programming Pig: Dataflow Scripting with Hadoop, By Alan Gates, Publisher: O'Reilly Media*
 - *Programming Hive: Data Warehouse and Query Language for Hadoop By Edward Capriolo, Dean Wampler, Jason Rutherford, Publisher: O'Reilly Media*
 - *Jaql documentation. http://www-01.ibm.com/support/knowledgecenter/SSPT3X_3.0.0/com.ibm.swg.i.../*
 - *Материалы лекций: http://synth_serv/synthesis/student/BigData/seminar-hadoop/hadoop2015*
2. Общие характеристики NoSQL баз данных, и их преимущества. CAP теорема. Классификация NoSQL баз данных. Архитектура и модель данных HBase.
 - *Apache HBase web site: http://hbase.apache.org/*
 - *George, Lars. HBase The Definitive Guide (O'Reilly 2011)*
 - *"Brewer's CAP Theorem" posted by Julian Browne, January 11, 2009.*
<http://www.julianbrowne.com/article/viewer/brewers-cap-theorem>
 - *Материалы лекций: http://synth_serv/synthesis/student/BigData/seminar-hadoop/hadoop2015*
3. Разрешение сущностей. Выявление дубликатов, удаление дубликатов, установление связей между сущностями из разных исходных коллекций.
 - *А.Е. Вовченко, Л.А. Калиниченко, Д.Ю. Ковалев. Методы разрешения сущностей и слияния данных в ETL-процессе и их реализация в среде Hadoop. Информатика и ее применения, т. 8, вып. 4, 2014.*
 - *Peter Christen, Data Matching. Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Springer Data-Centric Systems and Applications, 2012*
 - *Материалы лекций: http://synth_serv/synthesis/student/BigData/seminar-information-extracti...*
4. Слияние данных. Стратегии разрешения конфликтов. Функции разрешения конфликтов. Слияние данных на основе операции Union. Слияние данных на основе операции Join.

- A.E. Вовченко, Л.А. Калиниченко, Д.Ю. Ковалев. *Методы разрешения сущностей и слияния данных в ETL-процессе и их реализация в среде Hadoop. Информатика и ее применения*, т. 8, вып. 4, 2014.
 - J. Bleiholder, F. Naumann. *Data Fusion. ACM Computing Survey 2009*.
 - Материалы лекций: http://synth_serv/synthesis/student/BigData/seminar-information-extracti...
5. Конъюнктивные запросы. Ответы, интерпретация запроса. Поглощение запросов. Ответы с учетом взглядов. Переписывание запросов. Гомоморфизм запросов, теорема о гомоморфизме запросов. Алгоритм проверки поглощения запросов, пример поглощения.
- Serge Abiteboul, Ioana Manolescu. *Web Data Management. Cambridge University Press, 2011.* - <http://webdam.inria.fr/Jorge/html/wdmch10.html> (раздел 9.2)
6. Подход Local-As-View к виртуальной интеграции реляционных ресурсов. Bucket-алгоритм переписывания запросов. Корректность переписывания. Примеры локальной и глобальной схем, взглядов (представлений), переписывания запроса.
- Serge Abiteboul, Ioana Manolescu. *Web Data Management. Cambridge University Press, 2011.* - <http://webdam.inria.fr/Jorge/html/wdmch10.html> (раздел 9.4)
7. Обмен данными – общая схема, формальная постановка задачи. Порождающие кортежи зависимости (tgd). Универсальные решения в обмене данными. Свойства универсальных решений. Процедура погони.
- Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, Lucian Popa: *Data exchange: semantics and query answering. Theor. Comput. Sci. 336(1): 89-124 (2005)* - <http://dblab.cs.toronto.edu/project/dataexchange/docs/tcs05.pdf>
 - Phokion G. Kolaitis. *A Tutorial on Schema Mappings & Data Exchange DEIS '10.* - <http://synthesis.ipi.ac.ru/synthesis/student/BigData/master-course-warehousing/DW16-2-DataExchange.pdf>
8. Многомерная модель данных. Факты, измерения, параметры. Иерархия измерений. Подходы к реализации многомерной модели. Схемы звезда и снежинка.
- Кристиан Йенсен, Торбен Бэч. *Технология многомерных баз данных.* <http://www.osp.ru/os/2002/01/180958/>
 - Christian S. Jensen, Torben Bach Pedersen, and Christian Thomsen. *Multidimensional Databases and Data Warehousing. Morgan & Claypool, 2010.* (раздел 2)
 - *Data Warehouse Design.* - <http://synthesis.ipi.ac.ru/synthesis/student/BigData/lectures/DW%204%20Data%20Warehouse%20Design.pdf> (разделы Logical Modeling, Physical Modeling)
9. Иерархические и спектральные алгоритмы обнаружения сетевых сообществ.
- M. Newman. *Networks: an Introduction. Oxford University Press, 2010. Глава 11.*

10. Модели эпидемий.

- M. Newman. *Networks: an Introduction*. Oxford University Press, 2010. Глава 17

11. Доверительные интервалы для среднего, для доли. Построение доверительного интервала на основе бутстрепа. Связь между проверкой гипотез и доверительными интервалами.

- Лагутин М. Б. *Наглядная математическая статистика*. 2007. (Глава 11)

12. Проверка гипотез. Ошибки I и II рода. Достижаемый уровень значимости. Критерий согласия Пирсона (хи-квадрат).

- Лагутин М. Б. *Наглядная математическая статистика*. 2007. (Главы 12, 13)

13. Поиск аномалий в данных с помощью методов кластеризации. Постановка задачи. Методы DBSCAN, K-means.

- K. В. Воронцов. *Машинное обучение, курс лекций*. Курс - <https://bit.ly/ML-Vorontsov>, слайды: <http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-s...>, видео: <https://rb.gy/u457yj>

14. Поиск аномалий во временных рядах. Постановка задачи. Авторегрессионная модель ARMA

C. Aggarwal. *Outlier analysis*. Springer, 2013. (Глава 8)

15. Модели прогнозирования на основе деревьев решений. Алгоритмы CHAID, CART, C4.5: критерии поиска разбиений, параметры ограничения роста и обрубания дерева. –

Hastie, Tibshirani and Friedman. *The Elements of Statistical Learning*. – Springer-Verlag, 2009. <https://link.springer.com/book/10.1007/978-0-387-84858-7>

16. Классическая линейная модель регрессии: описание модели, основные ограничения, метод наименьших квадратов.

Charu C. Aggarwal. *Data Mining: The Textbook*. Springer, 2015. Раздел 11.5.1