ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«24» июня 2025 г.
Протокол №2

Рабочая программа дисциплины (модуля)
«Большие идеи в компьютерных науках»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика»*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2025*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2025

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 4
3. **Тематический план** 6
4. **Содержание дисциплины (модуля)** 6
5. **Учебно-методическое обеспечение** 7
6. **Материально-техническое обеспечение** 7
7. **Методические и оценочные материалы** 9

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Большие идеи в компьютерных науках» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) помогает развивать у студентов критическое мышление и навыки решения проблем, необходимые для эффективной работы в быстро меняющейся технологической среде. Кроме того, понимание больших идей в компьютерных науках способствует инновациям и улучшению процессов в различных областях профессиональной деятельности.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика» и входит в Блок Факультативные дисциплины.

Дисциплина (модуль) изучается на 1 курсе во 2 семестре.

Цель изучения дисциплины (модуля): заключается в том, чтобы познакомить студентов с основными концепциями и принципами, формирующими современное понимание компьютерных наук и их применения.

Задачи изучения дисциплины (модуля):

— освоение основных принципов работы и структуру интернета с упором на графовую структуру;

— освоение методов поисковых систем при классификации и ранжировании веб-страниц, включая деревья метрик и методы оценки релевантности;

— формирование понимания алгоритма pagerank и его применение для оценки важности веб-страниц;

— формирование понимания базовых принципов теории аукционов и их практическое применение;

— умение определять релевантность веб-страниц, используя алгоритм pagerank;

— умение рассчитывать аукционы разного типа: аукцион первой цены, второй цены и vcg (vickrey-clarke-groves);

— проектирование и анализ механизмов ранжирования и аукционных стратегий в цифровой среде.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности.
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области
		ОПК-1.2.	Умеет анализировать и формулировать математические

			задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов.
		ПК-1.2.	Умеет корректно ставить и формулировать математические задачи, применять строгие методы доказательства и анализировать полученные результаты.
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач
ПК-2.	Способен решать типовые задачи профессиональной деятельности в области искусственного интеллекта, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области искусственного интеллекта, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
		Семинары	Консультации			
1	Введение в графовую структуру. Интернета Поисковые системы и дерево метрик	3	3		16	Домашнее задание Подготовка к семинару
2	Алгоритм PageRank и его применение	3	3		8	Домашнее задание Подготовка к семинару
3	Основы теории аукционов	3	3		8	Домашнее задание
	Зачет			2		
	Итого:	***9***	***9***	***2***	***18***	
Объем дисциплины (модуля) (в ак. ч.)		***38***				
Объем дисциплины (модуля) (в зач. ед.)		***1***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Введение в графовую структуру Интернета. Поисковые системы и дерево метрик	Как работает Google. Большая картина. Путь поиска запроса. Индексация. Структура интернета. Карта интернета. Индексация страниц. Поиск
2	Алгоритм PageRank и его применение	Основная идея PageRank. Basic PageRank. PageRank и линейная алгебра. Свойства PageRank. Альтернативный взгляд на PageRank
3	Основы теории аукционов	Реклама в поиске. Теория аукционов. Аукцион VCG. Свойства аукциона VCG

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Маркин, А. В. Системы графовых баз данных. Neo4j : учебник для вузов / А. В. Маркин. — Москва : Издательство Юрайт, 2025. — 303 с. — (Высшее образование). — ISBN 978-5-534-13996-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/567732.

Дополнительная литература:

1. Новиков, Ф. А. Символический искусственный интеллект: математические основы представления знаний : учебник для вузов / Ф. А. Новиков. — Москва : Издательство Юрайт, 2025. — 278 с. — (Высшее образование). — ISBN 978-5-534-00734-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/561410.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том

числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование"	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое

CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Большие идеи в компьютерных науках» в рамках текущего контроля успеваемости используются такие виды учебной работы, как семинары, консультации и домашние задания, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Семинар — это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Консультации – структурированные встречи, на которых преподаватели предоставляют индивидуальную или групповую помощь в освоении учебного материала, обсуждении вопросов и решении проблем, возникающих в процессе обучения.

Консультации могут включать разъяснение сложных тем, подготовку к экзаменам и помощь в выполнении проектных работ, что способствует более глубокому пониманию предмета и улучшению академической успеваемости.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Большие идеи в компьютерных науках»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме ***зачета***, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	**Оценка за зачет**	**Пятибалльная оценка**	**Общая характеристика результата обучения по дисциплине (модулю)**
10	Зачтено	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Зачтено	Отлично	
8	Зачтено	Отлично	
7	Зачтено	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи.
6	Зачтено	Хорошо	

Десятибалльная оценка	Оценка за зачет	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
			Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Зачтено	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине, но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Зачтено	Удовлетворительно	
3	Не зачтено	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не зачтено	Не сдан	
1	Не зачтено	Не сдан	

Дисциплина (модуль) «Большие идеи в компьютерных науках» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	60%	3	Набор заданий по темам недели
Аудиторная работа	40%	2	Активное участие в семинарах: ответы на вопросы преподавателя и участие в дискуссии

Формула расчёта итоговой оценки по дисциплине (модулю) «Большие идеи в компьютерных науках»: «0,6 × среднее за домашние задания + 0,4× среднее за аудиторную работу».

Для получения зачета нужно получить 4 или более баллов

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные темы семинаров

Введение в графовую структуру Интернета. Поисковые системы и дерево метрик

1. Что такое графовая структура и как она применяется для описания Интернета?
2. Как узлы и ребра представляют веб-страницы и ссылки между ними?
3. Какие основные характеристики графов используются для анализа структуры Интернета?
4. Как поисковые системы используют графовые структуры для индексации веб-страниц?

5. Что такое дерево метрик и как оно помогает в организации данных?
6. Как метрики, такие как степень узла и центральность, влияют на понимание важности страниц?
7. Какова роль алгоритмов в анализе графовой структуры Интернета?
8. Какие проблемы могут возникнуть при использовании графовых моделей для поиска информации?
9. Как графовая структура помогает в улучшении качества поиска в поисковых системах?
10. Какие примеры реальных приложений графовых структур можно привести в контексте Интернета?

Алгоритм PageRank и его применение

1. Что такое алгоритм PageRank и кто его разработал?
2. Как PageRank определяет важность веб-страниц?
3. Какие математические концепции лежат в основе алгоритма PageRank?
4. Как PageRank справляется с проблемой "спама" в поисковых системах?
5. Каковы основные шаги в вычислении PageRank для веб-страниц?
6. Как PageRank влияет на ранжирование результатов поиска в поисковых системах?
7. Какие альтернативные алгоритмы существуют для оценки важности страниц?
8. Как PageRank можно применять вне контекста веб-страниц (например, в социальных сетях)?
9. Как PageRank учитывает изменения в структуре Интернета со временем?
10. Какие ограничения и недостатки есть у алгоритма PageRank?

Основы теории аукционов

1. Что такое теория аукционов и каковы ее основные компоненты?
2. Каковы различные типы аукционов и как они отличаются друг от друга?
3. Какова роль информации в аукционных процессах?
4. Как аукционы могут быть использованы для эффективного распределения ресурсов?
5. Как теория аукционов применяется в контексте онлайн-рекламы?
6. Что такое "аукцион с закрытыми ставками" и как он работает?
7. Каковы преимущества и недостатки различных аукционных механизмов?
8. Как поведение участников аукциона влияет на его результаты?
9. Как теория аукционов может быть использована для оптимизации бизнес-процессов?
10. Какие примеры успешных аукционных платформ можно привести?

Примерные домашние задания

Домашнее задание №1

Поисковые системы

Задача 1.

А. Приведи два конкретных и реальных примера того, чем современный поисковик (Яндекс или Google) отличается от базового варианта с лекции. Для каждого примера покажи, на какую метрику из дерева метрик он влияет. Если необходимо, можешь добавить дополнительную метрику в дерево метрик.

Хорошие примеры: сценарий «задать вопрос голосом»; наличие рекламных страниц в топе выдачи.

Б. Предложи одну новую идею, которая могла бы улучшить существующие поисковые системы.

Покажи, на какую метрику из дерева метрик она может повлиять.

Графовая структура интернета

Определение. Рассмотрим граф G, состоящий из множества вершин $V = \{v_1, v_2, \ldots, v_n\}$ и множества ребер $E = \{e_1, e_2, \ldots, e_m\}$.

Матрицей смежности графа G называется матрица M размера $n \times n$, каждый элемент M_{ij} которой определяется так:

$$M_{ij} = \begin{cases} 1, & \text{если в графе } G \text{ существует ребро, идущее из вершины } v_i \text{ в вершину } v_j; \\ 0, & \text{иначе.} \end{cases}$$

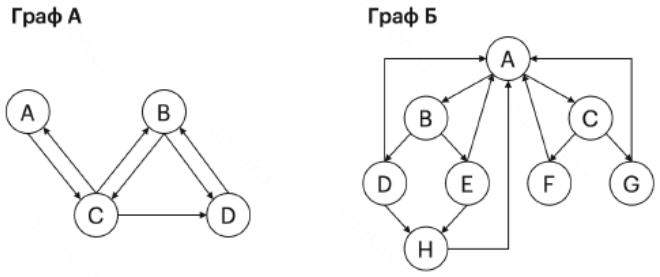


Рис. 1: Примеры графов

Построим матрицу смежности для графа А на рисунке 1. Будем считать, что вершины упорядочены по алфавиту: v_1 = A, v_2 = B, v_3 = C, v_4 = D. Тогда матрица смежности будет выглядеть так:

$$M = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Задание 2

А. Построй матрицу смежности для графа Б (рисунок 1). Считай, что вершины упорядочены по алфавиту.

Б. Построй граф по матрице смежности:

$$M = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

В. Выше мы работали с направленными графами — в них у каждого ребра есть направление (от вершины v_i к вершине v_j). Направленный граф подходит для моделирования интернета.

Действительно, если на странице v_i есть ссылка на страницу v_j, то на странице v_j может не быть обратной ссылки на страницу vi, поэтому нужны направленные рёбра.

Теперь рассмотрим ненаправленный граф — такой, в котором рёбра не имеют направлений (у рёбер нет стрелок). Ненаправленный граф, например, подходит для моделирования связей между городами: если из города А есть прямая дорога в город В, то и из города В в город А ведёт та же самая прямая дорога. Определение матрицы смежности подходит и ненаправленным графам: для каждого ребра $e \in E$ мы считаем, что оно идет в обе стороны (из v_i в v_j и обратно).

Построй матрицу смежности графа А (рисунок 1), считая его ненаправленным (игнорируй стрелки на ребрах). Какое уникальное свойство есть у матрицы смежности ненаправленного графа?

PageRank. Авторитетность страницы

В этом разделе мы переизобретём алгоритм PageRank, который определяет авторитетность страниц, основываясь на структуре ссылок. Его придумали основатели Google Сергей Брин и Ларри Пейдж. Все задачи ниже сформулированы для двух графов, изображённых на рисунке 1.

В основе метода PageRank лежит простая интуиция: ссылка со страницы A на страницу B добавляет странице B авторитетности. Например, страницу aviasales.ru цитируют сотни других страниц, а страницу nonametickets.ru вряд ли цитирует хоть кто-то. Поэтому даже если релевантность страницы nonametickets.ru высока, её авторитетность будет низкой.

Попробуем формализовать эту идею. Для этого посмотрим на ссылки в интернете как на голоса в го- лосовании.

Идея 1. *Авторитетность как результат голосования.* Пусть страницы голосуют друг за друга ссыл- ками. Если страница A ссылается на страницы $B_1, B_2, B_3, \ldots, B_k$, то она распределяет свой голос между этими страницами, то есть каждая страница $B_i, i \in \{1, 2, \ldots, k\}$ получает $1/k$-ю голоса от A.

Итоговая авторитетность каждой страницы после голосования равна сумме голосов, полученных от других страниц.

Посмотрим на пример графа A на рисунке 1. Таблица 1 показывает авторитетность каждой страницы по итогам голосования. Например, страница D получает $1/2$ голоса от B и $1/3$ голоса от C. Итоговая авторитетность страницы D равна $1/2 + 1/3 = 5/6$.

	A	B	C	D
Авторитетность	$1/3$	$4/3$	$3/2$	$5/6$

Таблица 1: Авторитетность страниц графа A по итогам голосования

Задание 3.

Подсчитай авторитетность каждой страницы графа Б после голосования. Чему равна сумма итоговых авторитетностей всех страниц?

Задание 4.

К сожалению, идея 1 сама по себе не даёт хорошего решения задачи определения авторитет- ности. Вернёмся к сайтам aviasales.ru и nonametickets.ru. Что должен сделать владелец nonametickets.ru, чтобы по итогам голосования получить авторитетность выше, чем у сайта aviasales.ru?

В этой задаче мы считаем, что каждый сайт состоит из одной страницы.

Как исправить проблему с идеей 1? Здесь мы подходим ко второй идее метода PageRank.

Идея 2. *Принцип итеративного улучшения.* После первого голосования у нас есть оценки авторитет- ности каждой страницы. Проведём еще один раунд голосования, но теперь каждая страница распре- деляет не 1 голос, а свою текущую авторитетность. Авторитетные страницы теперь имеют больший вес, а голоса страниц с низким авторитетом дисконтируются.

Применим вторую идею к графу A. Рассмотрим страницу C с текущей оценкой

авторитетности $3/2$. На странице C есть ссылки на три других страницы — A, B и D. Значит, каждая из этих страниц получает одну треть от текущей авторитетности C:

$$1/3 \times 3/2 = 1/2.$$

Аналогично рассмотрим страницу A, на которой есть всего одна ссылка — на страницу C. Текущая оценка авторитетности A равна $1/3$, значит, страница C получает всю текущую авторитетность от A: $1 \times 1/3 = 1/3$ голоса. Итоговые авторитетности страниц после второго раунда голосования даны в таблице 2.

	A	B	C	D
Авторитетность	$1/2$	$4/3$	1	$7/6$

Таблица 2: Авторитетность страниц по итогам второго раунда голосования на графе A

Задача 5.

Подсчитай авторитетность каждой страницы графа Б на рисунке 1 после второго раунда голосования. Чему равна сумма авторитетностей всех страниц?

Добавим финальную идею и, наконец, сформулируем базовую версию алгоритма PageRank.

Идея 3 Если второй раунд голосования улучшает оценку авторитетности, то нужно продолжать!

Basic PageRank Базовая версия алгоритма состоит из трех шагов:

Шаг 1. Устанавливаем изначальную авторитетность каждой страницы равной 1

Шаг 2. Фиксируем число итераций k

Шаг 3. Выполняем k раундов голосования по правилу:

Каждая страница равномерно делит свою текущую авторитетность между страницами, на которые она ссылается. Итоговая авторитетность каждой страницы после раунда голосования равна сумме голосов, полученных от других страниц.

О том, как выбрать число итераций k, мы поговорим на следующей лекции. А сейчас заметим, что после достаточно большого количества итераций алгоритма на графе A авторитетности страниц примут значения из таблицы 3.

	A	B	C	D
Авторитетность	$4/11$	$16/11$	$12/11$	$12/11$

Таблица 3: Значение авторитетностей страниц из графа A после большого количества итераций

Задача 6.

Проведи еще одну итерацию голосования на графе A, начиная со значений из таблицы 3. Прокомментируй результат.

Домашнее задание №2

Задача 1.

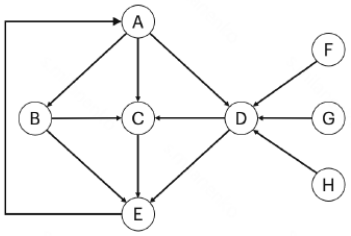
Рассмотрим матрицы A, B и вектора x_1, x_2:

$$A = \begin{pmatrix} 2 & 4 & 2 \\ 0 & -2 & 2 \\ 0 & 0 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1/3 & 5/6 & 0 \\ 1/3 & 0 & 0 \\ 1 & -1/6 & -1 \end{pmatrix}, \quad x_1 = \begin{pmatrix} 1 \\ 2 \\ 1/2 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -1 \\ 2 \\ 1/3 \end{pmatrix}$$

А. Найди транспонированные матрицы A^T *и* B^T.
Б. Вычисли произведения $A \cdot x_1$ *и* $B \cdot x_2$.
В. Вычисли значение $-1/2A \cdot x_1 + 3B \cdot x_2$.

Задача 2.

А. Построй матрицу смежности M для графа 1. Считай, что вершины упорядочены по алфавиту:
v_1 = A, v_2 = B и так далее.
Б. Транспонируй матрицу M и построй граф, соответствующий матрице смежности M^T.
В. Рассмотрим произвольные графы G_1 и G_2 с матрицами смежности M и M^T, соответственно.
Как эти графы связаны между собой?



Граф 1: Граф для задачи 2

Задача 3.

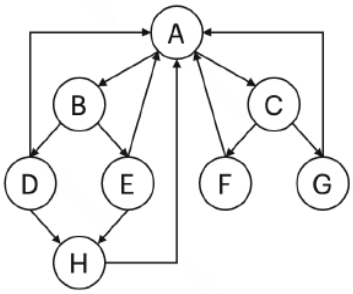
Рассмотрим матрицу A:

$$A = \begin{pmatrix} 4 & -1 \\ 3 & 0 \end{pmatrix}$$

Найди все собственные вектора матрицы A с собственным значением λ = 1.

Матричная форма PageRank

В этом разделе все задачи сформулированы на примере графа 2. Во всех задачах считай, что вершины упорядочены по алфавиту



Граф 2

Задача 4.

А. Составь матрицу голосования P и вектор авторитетностей $a^{(0)}$ для графа 2.
Б. Проведи вычисления двух итераций алгоритма Basic PageRank в матричной форме и найди вектора $a^{(1)}$ и $a^{(2)}$. В решении обязательно покажи свои вычисления.

Задача 5.

Может ли вектор x (определен ниже) быть предельным значением вектора авторитетностей в алгоритме Basic PageRank на графе 2? Если да, то можем ли мы заранее

утверждать, что алгоритм Basic PageRank при увеличении количества итераций k будет сходиться к вектору x?

$$x = \frac{1}{13}\begin{pmatrix} 32; & 16; & 16; & 8; & 8; & 8; & 8; & 8 \end{pmatrix}^T$$

Модификация Basic PageRank

Вы знаете, что базовая версия алгоритма PageRank не всегда сходится к осмысленным значениям авторитетностей. Такая проблема возникает для двух типов графов:

• Тип 1. Графы, в которых есть вершины без исходящих ребер. Например, Граф А на рисунке 3 — у вершины G нет исходящий ребер.

• Тип 2. Графы, в которых не все вершины соединены путями из ребер. Например, Граф Б на рисунке 3 — в нем из вершины G нельзя добраться до вершины A.

В этом разделе мы разработаем модификацию алгоритма, которая имеет хорошие свойства сходимости для любого графа. Модификация заключается в изменениях матрицы голосования P (задачи 6 и 7 – меняется только матрица голосования, всё остальное остается неизменным.)

Матрицу, которую ты получишь в результате, называют матрицей Google. Именно она лежит в основе алгоритма PageRank, который использовался для определения авторитетностей страниц на ранних этапах развития Google.

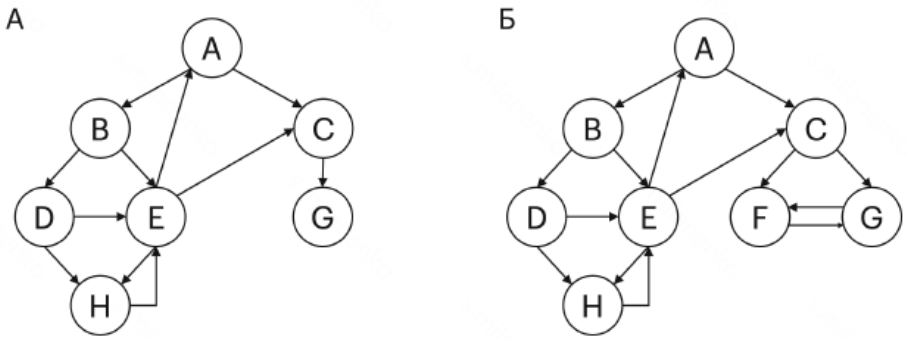


Рис. 3: Примеры проблемных графов для Basic PageRank

Задача 6.

Начнем с решения проблемы для первого типа графов.

Проблема: из-за отсутствия исходящих ребер у одной из вершин суммарная авторитетность страниц в графе уменьшается: уже на первой итерации суммарная авторитетность страниц уменьшается на 1, а в итоге может сойтись к нулю.

У этой проблемы есть простой фикс: если у вершины X нет исходящих ребер, то во время голосования X распределяет свою авторитетность равномерно на все вершины графа, включая себя.

А. Как наличие в графе вершины без исходящих ребер отражается на матрице голосования P?

Б. Как нужно изменить матрицу P, чтобы реализовать предложенный фикс?

В. Приведи интерпретацию предложенного фикса в терминах Random Serfer Model.

Задача 7.

Рассмотрим проблему второго типа графов.

Проблема: посмотрим на пример графа Б (рисунок 3). Вершины F и G в голосовании обмениваются авторитетностью и не отдают ее другим вершинам. Одновременно, в каждом раунде голосования вершины F и G получают авторитетность от других вершин (через C). В пределе вся авторитетность сконцентрируется на вершинах F и G, и Basic PageRank приведет к контринтуитивным результатам: высокоцитируемые вершины (E, H, . . .) останутся без авторитетности (таблица 1).

	A	B	C	D	E	F	G	H
Предельные значения	0	0	0	0	0	4	4	0

Таблица 1: Авторитетность страниц для графа Б (рисунок 3)

Эта проблема решается сложнее. Разберем технический фикс, который предложили создатели PageRank. Выберем параметр $s \in (0; 1)$ — scaling factor. Тогда в каждом раунде голосования каждая страница:

• Распределяет долю s авторитетности по правилам Basic PageRank.

• Распределяет долю $(1 - s)$ авторитетности равномерно по всем страницам, включая себя.

А. Как нужно изменить матрицу P, чтобы реализовать предложенный фикс?

Б. Приведи интерпретацию предложенного фикса в терминах Random Serfer Model.

В. Сформулируй полную версию алгоритма PageRank в терминах матричного умножения.

На практике оба фикса, предложенные в задачах 6 и 7, могут использоваться одновременно. Scaling factor из задачи 7 применяется по умолчанию, даже если в графе нет проблемы со связностью вершин. А фикс из задачи 6 добавляется если в графе находятся вершины без исходящих ребер.

Домашнее задание №3

Аукционы с одним товаром

Начнем с небольшой тренировки по простым аукционам с одним товаром (single-item auctions).

Задача 1.

А. В аукционе *первой* цены участвуют 8 кандидатов. Их ставки собраны в таблице 1. Кто победит в аукционе и какую стоимость заплатит?

Кандидат	A	B	C	D	E	F	G	H
Ставка	10	8	12	7	9	2	8	0

Таблица 1: Ставки кандидатов в пункте А

Б. В аукционе *второй* цены участвуют 8 кандидатов. Их ставки собраны в таблице 2. Кто победит в аукционе и какую стоимость заплатит?

Кандидат	A	B	C	D	E	F	G	H
Ставка	18	14	3	2	9	7	17	20

Таблица 2: Ставки кандидатов в пункте Б

Задача 2.

В аукционе ***второй*** цены участвуют 4 кандидата. Все участники ведут себя честно и делают ставки равные ценности, которую товар имеет для них: $b_i = v_i,\ i \in \{A, B, C, D\}$. Их ставки собраны в таблице 3.

Кандидат	A	B	C	D
Ставка	2	4	6	8

Таблица 3: Ставки кандидатов в задаче 2

А. Чему равна суммарная ***ценность***, полученная участниками по результатам аукциона? При подсчете ценности не учитывай стоимость, которую платят участники. Подсчитай только сумму ценностей (v), которую получили участники.

Б. Представь, что к аукциону присоединяется участник E и делает ставку $b_E = 10$. Чему будет равна суммарная ценность участников $\{A, B, C, D\}$ по результатам нового аукциона? Как изменение суммарной ценности связано с ценой, которую заплатит E?

Задача 3.

В аукционе ***второй*** цены участвуют 2 кандидата, которые действует независимо и не могут вступать в сговор. Организатор аукциона не готов продавать свой товар меньше, чем за s, поэтому он становится третьим участником и делает ставку s. Если его ставка победит, то он останется с товаром и заплатит сам себе итоговую цену p (фактически, организатор ничего не заплатит и оставит товар у себя).

Участники аукциона знают о таком поведении организатора и в точности знают значение s. Могут ли участники повысить свою ***выгоду***, отклонившись от честной стратегии, то есть сделав ставку $b \neq v$.

Задача 4.

По запросу «Авиабилеты в Москву» в поисковой выдаче выделены три слота под рекламу. Вероятности кликов на этих слотах равны 0.6 (слот #1), 0.5 (слот #2) и 0.1 (слот #3), соответственно. На показ по этому запросу претендуют четыре рекламодателя. Их ставки за клик собраны в таблице 4.

Кандидат	A	B	C	D
Ставка за клик	40	20	10	5

Таблица 4: Ставки кандидатов за клик в задаче 4

Распредели слоты с помощью аукциона VCG и посчитай какую цену заплатит каждый рекламодатель. В решении подробно распиши все расчеты.

Задача 5.

По запросу «Авиабилеты в Красноярск» в поисковой выдаче выделены два слота под рекламу. Вероятности кликов на этих слотах равны 0.4 (слот #1) и 0.3 (слот #2). На показ по этому запросу претендуют четыре рекламодателя. Их ставки собраны в таблице 5.

Кандидат	A	B	C	D
Ставка за клик	40	30	10	1

Таблица 5: Ставки кандидатов за клик в задаче 5

А. Распредели слоты с помощью аукциона VCG и посчитай какую цену заплатит каждый рекламодатель.

Б. Поисковая система решает выделить по этому запросу еще один рекламный слот. Вероятность клика по новому слоту равна 0.2, а его появление не влияет на два других слота. Распредели слоты с помощью аукциона VCG и посчитай какую цену заплатит каждый рекламодатель.

Б. Чему равен доход поисковой системы в случае А и Б? Было ли решение выделить третий слот верным?

Исследование аукциона второй цены

В этой задаче ты докажешь ключевое свойство аукциона второй цены.

Задача 6.

Докажи, что в аукционе второй цены оптимальная стратегия для каждого участника — быть честным. Независимо от того, как себя ведут другие участники, участнику i выгодно делать ставку, равную ценности, которую товар имеет для него: $b_i = v_i$.

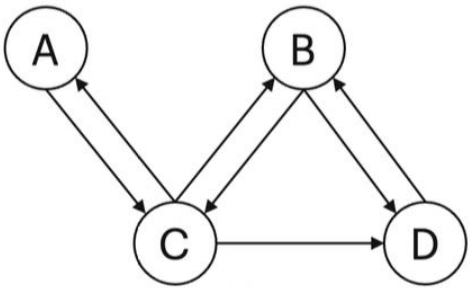
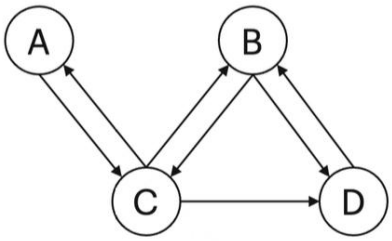
А. Докажи, что участник i не может увеличить свою выгоду за счет выбора ставки $b'_i > v_i$.

Б. Докажи, что участник i не может увеличить свою выгоду за счет выбора ставки $b'_i < v_i$.

Подсказка: Обрати внимание, что выбор ставки влияет на то, победил ли участник в аукционе или нет, но не влияет на размер оплаты в случае победы.

Задания для промежуточной аттестации по дисциплине (модулю)

<table>
<tr><th>№ п/п</th><th>Задание</th><th>Ответ</th><th>Компетенция</th></tr>
<tr><td>1.</td><td>Сопоставь методы (слева) и функции, которые эти методы выполняют в поисковой системе (справа).
Ответ запишите в последовательности букв, пример: ГВАБ (без пробелов и знаков препинания, большими буквами)
<table><tr><td>А. Алгоритмы обходы графов</td><td>1. Распределение рекламных слотов</td></tr><tr><td>Б. Алгоритмы парсинга</td><td>2. Построение карты интернета</td></tr><tr><td>В. Алгоритмы машинного обучения</td><td>3. Оценка релевантности страницы для запроса</td></tr><tr><td>Г. Аукционы
А.</td><td>4. Описание контента страницы</td></tr></table></td><td>ГАВБ</td><td>ОПК-1</td></tr>
<tr><td>2.</td><td>Пусть в аукционе первой цены разыгрывается товар, который представляет для участника Дональда ценность V. Какую ставку должен делать Дональд, если он ведет себя рационально?
А. Ставка равна V
Б. Ставка выше V
В. Ставка ниже V</td><td>В</td><td>ОПК-1</td></tr>
<tr><td>3.</td><td>Для каждого аукциона слева выбери эквивалентный аукцион справа.
Ответ запиши в последовательности букв, пример: АБ (без пробелов и знаков препинания, большими буквами)
<table><tr><td>А. Аукцион первой цены</td><td>1. Британский аукцион (цена начинается с 0 и повышается)</td></tr><tr><td>Б. Аукцион второй цены</td><td>2. Голландский аукцион</td></tr></table></td><td>БА/ба</td><td>УК-1</td></tr>
</table>

	A. (цена начинается с ∞ и понижается)		
4.	Выбери верные утверждения о методе PageRank А. Basic PageRank находит ненулевые авторитетности страниц для любого графа; Б. Если в графе есть вершина без исходящих ссылок, то итоговые авторитетности могут получиться нулевыми; В. Задача по поиску авторитетностей страниц сводится к задаче поиска собственного вектора определенной матрицы.	БВ/бв	УК-2
5.	Возьмем 4 интернет страницы, каждая из которых имеет по одной ссылке на каждую из трех других страниц. Сколько всего ребер будет в направленном графе, который отражает структуру интернета. В качестве ответа введи одно целое число. Пример ответа: 3	12	ПК-1
6.	Рассмотрим граф, изображенный на рисунке ниже. Чему равна сумма элементов матрицы смежности этого графа?  В качестве ответа введи одно целое число. Пример ответа: 3	7	ПК-1
7.	Рассмотрим граф, изображенный на рисунке ниже. Чему равна авторитетность страницы A после первой итерации голосования метода Basic PageRank?  Округли ответ до двух знаков после запятой. Пример ответа: 0,67 Комментарий: — бесконечная дробь 0,222222 округляется до 0,22 бесконечная дробь 0,555555 округляется до 0,56	0,33/0.33	ПК-1
8.	Рассмотрим граф, состоящий из n=10 вершин, в котором из каждой вершины есть ребро в хотя бы одну другую вершину. Чему равна сумма авторитетностей страниц после первого раунда голосования Basic PageRank? В качестве ответа введи одно целое число. Если в ответе получается не целое число, округли его до целого. Пример ответа: 5	10	ОПК-1
9.	В аукционе первой цены разыгрывается один товар и участвуют 8 кандидатов. Их ставки собраны в таблице ниже.	12	ОПК-1

	<table><tr><td>Кандидат</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td><td>H</td></tr><tr><td>Ставка</td><td>10</td><td>8</td><td>12</td><td>7</td><td>9</td><td>2</td><td>8</td><td>0</td></tr></table> Какую стоимость заплатит победитель аукциона? В качестве ответа введи одно целое число. Пример ответа: 5		
10.	По запросу «Отдых в Сочи» в поисковой выдаче выделены три слота под рекламу. Вероятности кликов на этих слотах равны 0,7 (слот #1), 0,4 (слот #2) и 0,1 (слот #3), соответственно. На показ по этому запросу претендуют четыре рекламодателя. <table><tr><td>Кандидат</td><td>A</td><td>B</td><td>C</td><td>D</td></tr><tr><td>Ставка за клик</td><td>40</td><td>20</td><td>10</td><td>5</td></tr></table> Чему равна ценность показа на втором слоте для рекламодателя B? В качестве ответа введи одно целое число. Пример ответа: 5	8	ПК-2