

Федеральное государственное бюджетное образовательное
учреждение высшего образования
Московский государственный университет имени М.В. Ломоносова
Механико-математический факультет

**УТВЕРЖДАЮ**
декан механико-
математического факультета
_____/А.И.
Шафаревич /
« 14 » октября 2021г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

Теория баз данных

Уровень высшего образования:

магистратура

Направление подготовки / специальность:

02.04.01 "Математика и компьютерные науки" (3++)

Направленность (профиль)/специализация ОПОП:

Интеллектуальные системы. Теория и приложения

Форма обучения:

очная

Рабочая программа рассмотрена и одобрена
на заседании Ученого совета механико-математического факультета
(протокол № 7 от 14 октября 2021 года)

Москва 2021

Рабочая программа дисциплины (модуля) разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 02.04.01 "Математика и компьютерные науки" утвержденного Приказом Министерства образования и науки РФ от 10 января 2018 г. N 13.

1. Место дисциплины (модуля) в структуре ОПОП ВО:

дисциплина относится к блоку профессиональной подготовки вариативной части ОПОП ВО.

2. Входные требования для освоения дисциплины (модуля), предварительные условия (если есть):

Для того чтобы изучение дисциплины было возможно, обучающийся должен обладать следующими компетенциями:

Знать: основные понятия, концепции, результаты и методы дискретной математики, теории вероятностей, математического анализа и линейной алгебры.

Уметь: решать стандартные задачи дискретной математики и теории вероятностей.

3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников.

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ПК-8. Способен руководить проектами со стороны заказчика по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях	ПК-8.3. Решает прикладные задачи и реализует проекты в области сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений» со стороны заказчика	ПК-8.3. 3-1. Знает фундаментальные правила построения рекомендательных систем и систем поддержки принятия решений, основанных на интеллектуальных принципах, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений», основываясь на знании теории баз данных ПК-8.3. У-1. Умеет решать задачи по выполнению коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений» со стороны заказчика, основываясь на знании теории баз данных

4. Объем дисциплины (модуля) составляет 3 з.е., в том числе 36 академических часов, отведенных на контактную работу обучающихся с преподавателем, 72 академических часов на самостоятельную работу обучающихся.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий:

5.1. Структура дисциплины (модуля) по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий (в строгом соответствии с учебным планом)

Наименование разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Номинальные трудозатраты обучающегося			Всего академических часов	Форма текущего контроля успеваемости* (наименование)
	Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, академические часы		Самостоятельная работа обучающегося, академические часы		
	Занятия лекционного типа	Занятия семинарского типа			
Тема 1. Информационно-графовая модель данных.	8		19	27	опрос
Тема 2. Задача поиска идентичных объектов.	8		19	27	опрос
Тема 3. Задачи поиска с коротким ответом.	8		19	27	опрос
Тема 4. Задачи поиска с отношением частичного порядка.	8		19	27	опрос
Другие виды самостоятельной работы (отсутствуют)	—	—			—
Промежуточная аттестация (экзамен)					
Итого	32		76	108	—

5.2. Содержание разделов (тем) дисциплины

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
1.	Тема 1. Информационно-графовая модель данных.	Информационно-графовая модель данных. Критерий допустимости информационных графов. Критерий полноты базового множества.

		Сложность информационных графов. Теорема о существовании оптимальных информационных графов. Мощностная нижняя оценка. Случай оптимальности перебора.
2.	Тема 2. Задача поиска идентичных объектов.	Задача поиска идентичных объектов. Константный в среднем алгоритм поиска идентичных объектов. Оценки памяти константного в худшем случае алгоритма поиска идентичных объектов. Константный в худшем случае алгоритм поиска идентичных объектов с линейной памятью.
3.	Тема 3. Задачи поиска с коротким ответом.	Задачи поиска с коротким ответом. Древовидность оптимальных информационных графов для задач поиска с коротким ответом. Нижняя оценка сложности задач поиска с коротким ответом в случае равновероятных теней записей.
4.	Тема 4. Задачи поиска с отношением частичного порядка.	Задачи поиска с отношением частичного порядка. Теорема о существовании главных цепей для задач поиска с отношением частичного порядка. Задача включающего поиска. Верхняя оценка сложности включающего поиска. Теорема Краскала-Катоны. Нижняя оценка сложности включающего поиска. Нижняя оценка включающего поиска в классе древовидных схем.

6. Фонд оценочных средств (ФОС, оценочные и методические материалы) для оценивания результатов обучения по дисциплине (модулю).

6.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости, критерии и шкалы оценивания

Примеры заданий для самостоятельной работы

2.11. По аналогии с одномерной задачей интервального поиска приведите тип, описывающий n -мерные задачи интервального поиска.

2.13. Задача о метрической близости состоит в том, чтобы по произвольно взятой точке-запросу единичного n -мерного куба n -мерного евклидова пространства найти в конечном подмножестве этого куба (библиотеке) все точки, находящиеся на расстоянии не более чем R от точки запроса. Опишите тип, задающий задачи о метрической близости.

2.21. Пусть $X = \{1, 2, \dots, N\}$, $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, множество переключателей имеет вид

$$G = \{g_a(x) = \begin{cases} 1, & \text{если } x < a \\ 2, & \text{если } x = a \\ 3, & \text{если } x > a \end{cases} : a \in X\}, \quad (2.11)$$

базовое множество имеет вид $\mathcal{F} = \langle \emptyset, G \rangle$, $V = \{3, 5, 7, 11, 13, 17, 19\}$. Постройте информационный граф над базовым множеством $\mathcal{F}$, решающий ЗИП $I = \langle X, V, = \rangle$.

2.22. Пусть $X = \{1, 2, \dots, N\}$, $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, $V = \{y_1, y_2, \dots, y_k\} \subseteq X$. Предположим, что $y_1 < y_2 < \dots < y_k$. Метод блочного поиска с размером блока m , решающий задачу $I = \langle X, V, = \rangle$, состоит в следующем. Если на вход алгоритма поиска подается запрос $x \in X$, то, начиная с $i = 1$ до $i = k/m$, просматриваем записи $y_{i \cdot m}$. Если $x > y_{i \cdot m}$, то увеличиваем i на 1, иначе по очереди просматриваем записи $y_{(i-1)m+1}, y_{(i-1)m+2}, \dots, y_{i \cdot m}$ и сравниваем их с запросом x . При равенстве мы нашли нужную запись, если же ни для какой записи равенства не наблюдается, то ответ на запрос x пуст. Опишите базовое множество и постройте информационный граф над этим базовым множеством, который бы решал ЗИП $I = \langle X, V, = \rangle$ методом блочного поиска.

2.23. Пусть $X = \{1, 2, \dots, N\}$, $V \subseteq X$, ρ_c — отношение поиска, задаваемое на $X \times V$ и определяемое соотношением

$$x\rho_c y \iff (y \in V) \& (x \leq y) \& (\neg(\exists y')((y' \in V) \& (x \leq y') \& (y' < y))), \quad (2.12)$$

т. е. $x\rho_c y$, если $y \in V$ ближайшее справа к x . При выполнении этих условий ЗИП $I = \langle X, V, \rho_c \rangle$ называется задачей о близости. Пусть базовое множество имеет вид $\mathcal{F} = \langle \emptyset, G \rangle$, где множество переключателей G задается соотношением (2.3). Постройте информационный граф над базовым множеством $\mathcal{F}$, решающий ЗИП $I = \langle X, V, \rho_c \rangle$, если $V = \{3, 5, 7, 11, 13, 17, 19\}$.

2.24. Одномерная задача о доминировании задается типом

$$S_{dom1} = \langle [0, 1], [0, 1], \geq \rangle.$$

Пусть $V = \{y_1, y_2, \dots, y_k\} \subseteq [0, 1]$. Опишите некоторое базовое множество и постройте какой-либо информационный граф над этим базовым множеством, который бы решал ЗИП $I = \langle [0, 1], V, \geq \rangle$.

2.36. Пусть $X = \{1, 2, \dots, N\}$, $S = \langle X, X, =, \mathbf{P}, \sigma \rangle$ — тип поиска идентичных объектов, где $\sigma = 2^X$, $\mathbf{P}$ — равномерная вероятностная мера, то есть для любого $x \in X$ выполняется $\mathbf{P}(x) = 1/N$, $V = \{3, 5, 7, 11, 13, 17, 19\}$. Приведите мощностную нижнюю оценку для ЗИП $I = \langle X, V, = \rangle$.

2.37. Пусть $S_{dom1} = \langle [0, 1], [0, 1], \geq, \mathbf{P}, \sigma \rangle$ — тип одномерной задачи о доминировании, где $\mathbf{P}$ — равномерная вероятностная мера на $[0, 1]$, $V = \{y_1, y_2, \dots, y_k\} \subseteq [0, 1]$. Приведите мощностную нижнюю оценку для ЗИП $I = \langle [0, 1], V, \geq \rangle$.

2.38. Пусть $S_{int} = \langle X_{int}, Y_{int}, \rho_{int} \rangle$ — тип одномерного интервального поиска и на множестве запросов $X_{int} = \{(u, v) : 0 \leq u \leq v \leq 1\}$ задана равномерная вероятностная мера. $V = \{y_1, y_2, \dots, y_k\} \subseteq [0, 1]$. Приведите мощностную нижнюю оценку для ЗИП $I = \langle X_{int}, V, \rho_{int} \rangle$. Оцените сверху полученную величину.

2.41. Пусть $X = Y = [0, 1]^2$, $(x_1, x_2)\rho(y_1, y_2) \Leftrightarrow |x_1 - y_1| + |x_2 - y_2| \leq 1/10$. На множестве запросов задана равномерная вероятностная мера. Является ли $\langle [0, 1]^2, V, \rho \rangle$ задачей поиска с коротким ответом, если

- а) $V = \{(1/7, 1/7), (1/6, 1/4), (1/5, 5/12), (2/3, 0)\}$,
- б) $V = \{(0.1, 0.7), (0.2, 0.6), (0.3, 6/11), (0.5, 1/111)\}$?

2.42. Для каких баз данных задача поиска ближайшего соседа, описанная в примере 2.16, будет задачей с коротким ответом и равновероятными тенями записей.

2.43. Пусть $X = Y = [0, 10]^2$, $V = \{(1, 5), (3, 4), (6, 6), (8, 7)\}$,

$$(x_1, x_2)\rho(y_1, y_2) \Leftrightarrow |x_1 - y_1| \leq 1 \text{ \& } |x_2 - y_2| \leq 1, \quad (2.49)$$

$$f_{a,b}^{c,d}(x_1, x_2) = \begin{cases} 1, & \text{если } x_1 \in [a, b] \text{ \& } x_2 \in [c, d], \\ 0, & \text{иначе.} \end{cases}$$

Убедиться, что информационный граф, приведенный на рисунке 2.17, решает задачу $I = \langle [0, 10]^2, V, \rho \rangle$, перестроить этот граф в древовидный из класса $\mathcal{D}^I$ по алгоритму, описанному в доказательстве теоремы 27 и проверить, что сложность нового графа не больше, чем у старого, считая, что на множестве запросов задана равномерная вероятностная мера.

2.39. Пусть $I = \langle X, V, \rho_c \rangle$ — задача о близости, где $X = \{1, 2, \dots, N\}$, $V \subseteq X$, ρ_c — отношение поиска, задаваемое соотношением на $X \times V$ и определяемое соотношением (2.12). По аналогии с поиском идентичных объектов постройте ИГ, решающий задачу о близости I методом бинарного поиска. Получите соответствующие оценки сложности.

2.38. Пусть $S_{int} = \langle X_{int}, Y_{int}, \rho_{int} \rangle$ — тип одномерного интервального поиска и на множестве запросов $X_{int} = \{(u, v) : 0 \leq u \leq v \leq 1\}$ задана равномерная вероятностная мера. $V = \{y_1, y_2, \dots, y_k\} \subseteq [0, 1]$. Приведите мощностную нижнюю оценку для ЗИП $I = \langle X_{int}, V, \rho_{int} \rangle$. Оцените сверху полученную величину.

2.40. Пусть $I = \langle X, V, \rho_c \rangle$ — задача о близости, где $X = \{1, 2, \dots, N\}$, $V \subseteq X$, ρ_c — отношение поиска, задаваемое соотношением на $X \times V$ и определяемое соотношением (2.12). По аналогии с поиском идентичных объектов постройте ИГ, решающий задачу о близости I и имеющий константную сложность.

6.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации по дисциплине, критерии и шкалы оценивания

Контрольная работа N 2 — Вариант I

1. Пусть $S = \langle [0, 1]^2, [0, 1]^2, \rho \rangle$ — тип задач информационного поиска, где $(x_1, x_2)\rho(y_1, y_2) \Leftrightarrow x_1 \leq y_1 \& x_2 > y_2$. Будет ли полным для типа S базовое множество, состоящее из предикатов вида $f_{(a,b)}^i(x_1, x_2) = 1 \Leftrightarrow x_i \in (a, b]$, $i = 1, 2$, $a, b \in \mathbb{R}$? Ответ обосновать.
2. Пусть $X = \{1, 2, \dots, 100\}$, $S = \langle X, X, = \rangle$ — тип поиска идентичных объектов, базовое множество состоит из переключателей $g_a(x)$, принимающих значения 1 и 2 при $x < a$ и $x \geq a$ соответственно. Построить информационный граф, решающий ЗИП $I = \langle X, V, = \rangle$, где $V = \{5, 7, 12, 18\}$.
3. Задача поиска ближайшего соседа. Пусть $X = Y = [0, 10]$, отношение $x\rho y$ точно тогда, когда запись y из V является ближайшей к запросу x . Базовое множество состоит из предикатов вида $f_{a,b}(x) = 1 \Leftrightarrow x \in [a, b]$, $a, b \in \mathbb{R}$. Построить ИГ, решающий ЗИП с библиотекой $V = \{1, 2, 3, 5\}$, со средней сложностью меньшей чем 4, если запросы равномерно распределены по отрезку $[0, 10]$.
4. Для типа $S = \langle \{0, 1\}^4, \{0, 1\}^4, \rho \rangle$ построить полное базовое множество, если $(x_1, x_2, x_3, x_4)\rho(y_1, y_2, y_3, y_4) \Leftrightarrow (x_1 \leq y_1) \& (x_2 \geq y_2) \& (x_3 \sim y_3) \& (x_4 \vee \bar{y}_4)$. Построить ИГ, решающий ЗИП $I = \langle \{0, 1\}^4, V, \rho \rangle$, где $V = \{(0, 1, 1, 0), (1, 1, 1, 1), (1, 1, 0, 0), (0, 1, 1, 1), (1, 1, 0, 1)\}$, и вычислить его сложность при условии, что на $\{0, 1\}^4$ задана равномерная вероятностная мера.

Программа экзамена:

1. Информационно-графовая модель данных.
2. Критерий допустимости информационных графов.
3. Критерий полноты базового множества.
4. Сложность информационных графов.
5. Теорема о существовании оптимальных информационных графов.
6. Мощностная нижняя оценка.
7. Задача поиска идентичных объектов.
8. Константный в среднем алгоритм поиска идентичных объектов.
9. Оценки памяти константного в худшем случае алгоритма поиска идентичных объектов.
10. Константный в худшем случае алгоритм поиска идентичных объектов с линейной памятью.
11. Задачи поиска с коротким ответом.
12. Древовидность оптимальных информационных графов для задач поиска с коротким ответом.
13. Нижняя оценка сложности задач поиска с коротким ответом в случае равновероятных теней записей.
14. Теорема о существовании главных цепей записей для задачи поиска с отношением частичного порядка.
15. Задача включающего поиска.
16. Нижняя оценка сложности включающего поиска.
17. Асимптотика функции Шеннона сложности включающего поиска.
18. Нижняя оценка сложности включающего поиска в классе древовидных схем.

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
виды оценочных средств				
Знания (виды оценочных средств: опрос, тесты)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения (виды оценочных средств: практические задания)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности)	Успешное и систематическое умение

			непринципиального характера)	
Навыки (владения, опыт деятельности) (виды оценочных средств: выполнение и защита курсовой работы, отчет по практике, отчет по НИР и т.п.)	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

7. Ресурсное обеспечение:

7.1. Перечень основной и дополнительной литературы

Перечень основной и дополнительной учебной литературы:

- 1) Кудрявцев В.Б., Гасанов Э.Э., Подколзин А.С. Теория интеллектуальных систем: в 4 кн. Книга вторая. Базы данных. – М.: Издательские решения. 2018. – 212 с. – ISBN 978-5-4493-5158-6 (т. 2), ISBN 978-5-4493-5157-9.
- 2) Гасанов Э.Э., Кудрявцев В.Б., Теория хранения и поиска информации. Москва, «Физматлит», 2002..
- 3) Гасанов Э.Э. Теория сложности информационного поиска. Москва, изд-во механико-математического факультета МГУ, 2005.
- 8) Мейер Д. Теория реляционных баз данных. – М.: Мир, 1987.
- 9) Гасанов Э.Э., Кудрявцев В.Б., Теория хранения и поиска информации. Москва, «Физматлит», 2002..
- 10) Мендельсон Э. Введение в математическую логику. – М.: Наука, 1971.

7.2. Перечень лицензионного программного обеспечения, в том числе отечественного производства нет

7.3. Перечень профессиональных баз данных и информационных справочных систем нет

7.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <http://intsys.msu.ru/>
2. <http://intsys.msu.ru/science/books/>

7.5. Описание материально-технического обеспечения.

Аудитории для проведения лекционных занятий.

8. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.
9. Разработчик (разработчики) программы.

д.ф.-м.н., проф. Э.Э.Гасанов.