

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**  
федеральное государственное автономное образовательное учреждение  
высшего образования «Балтийский федеральный университет имени Иммануила  
Канта»  
Высшая школа компьютерных наук и прикладной математики

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**«Методы и средства информационного поиска»**

**Шифр: 01.03.02**

**Направление подготовки: «Прикладная математика и информатика»**

**Профиль: «Искусственный интеллект и анализ данных»**

**Квалификация (степень) выпускника: бакалавр**

Калининград  
2023

## Лист согласования

### Составители:

1. Верещагин Сергей Дмитриевич, к. ф.-м.н., доцент
2. Верещагин Михаил Дмитриевич, к. ф.-м.н., доцент
3. Мищук Богдан Ростиславович, к. ф.-м.н., доцент

Рабочая программа утверждена на заседании  
Ученого совета ОНК «Институт высоких технологий»

Протокол № 4 от «24» января 2023 г.

Председатель Ученого совета ОНК  
«Институт высоких технологий»

Профессор, д.ф.-м.н.

А.В. Юров

Руководитель ОПОП ВО

Е.П. Ставицкая

## Содержание

1. Наименование дисциплины «Методы и средства информационного поиска».
2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы.
3. Место дисциплины в структуре образовательной программы.
4. Виды учебной работы по дисциплине.
5. Содержание дисциплины, в том числе практической подготовки в рамках дисциплины, структурированное по темам.
6. Рекомендуемая тематика учебных занятий в форме контактной работы.
7. Методические рекомендации по видам занятий
8. Фонд оценочных средств
  - 8.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы в рамках учебной дисциплины
  - 8.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности в процессе текущего контроля
  - 8.3. Перечень вопросов и заданий для промежуточной аттестации по дисциплине
  - 8.4. Планируемые уровни сформированности компетенций обучающихся и критерии оценивания
9. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

### 1. Наименование дисциплины: «Методы и средства информационного поиска».

**Целью** курса «Методы и средства информационного поиска» - изучить основные математические методы поиска и обработки информации.

### 2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Код компетенции	Результаты освоения образовательной программы (ИДК)	Результаты обучения по дисциплине
ПК-7. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК-7.1. Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях ПК-7.2. Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения	<ul style="list-style-type: none"><li>• знать<ol style="list-style-type: none"><li>1. основные особенности естественного языка, уровней языковой системы и моделей обработки текстов;</li><li>2. современные модели информационного поиска; методы автоматической классификации и кластеризации текстов;</li></ol></li><li>• уметь<ol style="list-style-type: none"><li>1. применять на практике модели информационного поиска для решения задач в рамках информационных систем,</li><li>2. применять методы классификации, кластеризации для извлечения знаний и информации из текстов;</li></ol></li><li>• владеть<ol style="list-style-type: none"><li>1. навыками выбора методов решения конкретной задачи автоматической обработки текстов (статистический, инженерно-лингвистический, комбинированный);</li><li>2. навыками анализа результатов обработки текстов для коррекции используемых алгоритмов обработки текстов.</li></ol></li></ul>

### 3. Место дисциплины в структуре образовательной программы

«Методы и средства информационного поиска» представляет собой дисциплину Части, формируемой участниками образовательных отношений (Б1.В.ДВ.01.02), дисциплина по выбору, направления подготовки бакалавриата 01.03.02 «Прикладная математика и информатика», профиль «Искусственный интеллект и анализ данных».

### 4. Виды учебной работы по дисциплине.

Виды учебной работы по дисциплине зафиксированы учебным планом основной профессиональной образовательной программы по указанному направлению и профилю,

выражаются в академических часах. Часы контактной работы и самостоятельной работы обучающегося и часы, отводимые на процедуры контроля, могут различаться в учебных планах ОПОП по формам обучения. Объем контактной работы включает часы контактной аудиторной работы (лекции/практические занятия/ лабораторные работы), контактной внеаудиторной работы (контроль самостоятельной работы), часы контактной работы в период аттестации. Контактная работа, в том числе может проводиться посредством электронной информационно-образовательной среды университета с использованием ресурсов сети Интернет и дистанционных технологий

## 5. Содержание дисциплины, структурированное по темам (разделам)

Исходя из рамок, установленных учебным планом по трудоемкости и видам учебной работы по дисциплине, преподаватель самостоятельно выбирает тематику занятий по формам и количеству часов проведения контактной работы: лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации преподавателем и (или) занятия семинарского типа (семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия), и (или) групповые консультации, и (или) индивидуальную работу обучающихся с преподавателем, в том числе индивидуальные консультации (по курсовым работам/проектам – при наличии курсовой работы/проекта по данной дисциплине в учебном плане). Рекомендуемая тематика занятий максимально полно реализуется в контактной работе с обучающимися очной формы обучения. В случае реализации образовательной программы в заочной / очно-заочной форме трудоемкость дисциплины сохраняется, однако объем учебного материала в значительной части осваивается обучающимися в форме самостоятельной работы. При этом требования к ожидаемым образовательным результатам обучающихся по данной дисциплине не зависят от формы реализации образовательной программы.

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
1.	Тема 1. Введение:	Определение сферы информационного поиска, задачи информационного поиска, Информационно-поисковые системы различной направленности. Архитектура информационно-поисковых систем
2.	Тема 2. Модели информационного поиска.	Модели информационного поиска. Оценка качества информационного поиска .
3.	Тема 3. Методы расширения информационно-поисковых запросов.	Методы расширения информационно-поисковых запросов. Вопросно-ответные системы. Диалоговые системы
4.	Тема 4. Учет различных факторов.	Анализ ссылок, логи запросов, анализ кликов, персонализация выдачи, поисковые сессии, словосочетания и близость расположения, тематические вероятностные модели. Комбинирование факторов. Модели Learning to-rank..
5.	Тема 5. Автоматическая классификация и кластеризация текстов.	Типы классификации. Тематическая классификация и анализ тональности. Методы классификации. Особенности методов классификации. Методы автоматической кластеризации. Автоматическое аннотирование.

## 6. Рекомендуемая тематика учебных занятий в форме контактной работы

Рекомендуемая тематика учебных занятий лекционного типа (предусматривающих преимущественную передачу учебной информации преподавателями):

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
6.	Тема 1. Введение:	Лекция 1. Задачи информационного поиска Лекция 2. Архитектура информационно-поисковых систем
7.	Тема 2. Модели информационного поиска.	Лекция 3-4. Модели информационного поиска. Лекция 5. Оценка качества информационного поиска .
8.	Тема 3. Методы расширения информационно-поисковых запросов.	Лекция 6-7. Методы расширения информационно-поисковых запросов. Лекция 8. Вопросно-ответные системы. Лекция 9. Диалоговые системы
9.	Тема 4. Учет различных факторов.	Лекция 10-11. Анализ ссылок, логи запросов, анализ кликов, персонализация выдачи, поисковые сессии, словосочетания и близость расположения, тематические вероятностные модели. Лекция 12. Комбинирование факторов. Лекция 13. Модели Learning to-rank..
10.	Тема 5. Автоматическая классификация и кластеризация текстов.	Лекция 14. Типы классификации. Тематическая классификация и анализ тональности. Лекция 15. Методы классификации. Лекция 16. Особенности методов классификации. Лекция 17. Методы автоматической кластеризации. Лекция 18. Автоматическое аннотирование.

### Рекомендуемая тематика практических занятий:

1. Задачи информационного поиска
2. Архитектура информационно-поисковых систем
3. Модели информационного поиска.
4. Оценка качества информационного поиска .
5. Методы расширения информационно-поисковых запросов.
6. Вопросно-ответные системы.
7. Диалоговые системы
8. Анализ ссылок, логи запросов, анализ кликов, персонализация выдачи, поисковые сессии, словосочетания и близость расположения, тематические вероятностные модели.

9. Комбинирование факторов.
10. Модели Learning to-rank..
11. Типы классификации. Тематическая классификация и анализ тональности.
12. Методы классификации.
13. Особенности методов классификации.
14. Методы автоматической кластеризации.
15. Автоматическое аннотирование.

### **Требования к самостоятельной работе обучающихся**

1. Работа с лекционным материалом, предусматривающая проработку конспекта лекций и учебной литературы, по всем темам из п. 6 настоящей рабочей программы.
2. Выполнение домашнего задания, предусматривающего решение задач, выполнение упражнений, выдаваемых на практических занятиях, по всем темам из п. 6 настоящей рабочей программы.

Руководствуясь положениями статьи 47 и статьи 48 Федерального закона от 29 декабря 2012 г. N 273-ФЗ «Об образовании в Российской Федерации» научно-педагогические работники и иные лица, привлекаемые университетом к реализации данной образовательной программы, пользуются предоставленными академическими правами и свободами в части свободы преподавания, свободы от вмешательства в профессиональную деятельность; свободы выбора и использования педагогически обоснованных форм, средств, методов обучения и воспитания; права на творческую инициативу, разработку и применение авторских программ и методов обучения и воспитания в пределах реализуемой образовательной программы и отдельной дисциплины.

Исходя из рамок, установленных учебным планом по трудоемкости и видам учебной работы по дисциплине, преподаватель самостоятельно выбирает тематику занятий по формам и количеству часов проведения контактной работы: лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации преподавателем и (или) занятия семинарского типа (семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия), и (или) групповые консультации, и (или) индивидуальную работу обучающихся с преподавателем, в том числе индивидуальные консультации (по курсовым работам/проектам – при наличии курсовой работы/проекта по данной дисциплине в учебном плане).

Рекомендуемая тематика занятий максимально полно реализуется в контактной работе с обучающимися очной формы обучения. В случае реализации образовательной программы в заочной / очно-заочной форме трудоемкость дисциплины сохраняется, однако объем учебного материала в значительной части осваивается обучающимися в форме самостоятельной работы. При этом требования к ожидаемым образовательным результатам обучающихся по данной дисциплине не зависят от формы реализации образовательной программы.

### **7. Методические рекомендации по видам занятий**

Лекционные занятия.

В ходе лекционных занятий обучающимся рекомендуется выполнять следующие действия. Вести конспектирование учебного материала. Обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации по их применению. Задавать преподавателю

уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Желательно оставить в рабочих конспектах поля, на которых во внеаудиторное время можно сделать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений.

Практические и семинарские занятия.

На практических и семинарских занятиях в зависимости от темы занятия выполняется поиск информации по решению проблем, практические упражнения, контрольные работы, выработка индивидуальных или групповых решений, итоговое обсуждение с обменом знаниями, участие в круглых столах, разбор конкретных ситуаций, командная работа, представление портфолио и т.п.

Самостоятельная работа.

Самостоятельная работа осуществляется в виде изучения литературы, эмпирических данных по публикациям и конкретных ситуаций из практики, подготовке индивидуальных работ, работа с лекционным материалом, самостоятельное изучение отдельных тем дисциплины; поиск и обзор литературы и электронных источников; чтение и изучение учебника и учебных пособий.

## 8. Фонд оценочных средств

### 8.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы в рамках учебной дисциплины

Основными этапами формирования указанных компетенций при изучении обучающимися дисциплины являются последовательное изучение содержательно связанных между собой тем учебных занятий. Изучение каждой темы предполагает овладение обучающимися необходимыми компетенциями. Результат аттестации обучающихся на различных этапах формирования компетенций показывает уровень освоения компетенций.

Контролируемые разделы (темы) дисциплины	Индекс контролируемой компетенции (или её части)	Оценочные средства по этапам формирования компетенций
		текущий контроль по дисциплине
Тема 1. Введение:	ПК-7	Решение задач
Тема 2. Модели информационного поиска.	ПК-7	Решение задач
Тема 3. Методы расширения информационно-поисковых запросов.	ПК-7	Решение задач
Тема 4. Учет различных факторов.	ПК-7	Решение задач
Тема 5. Автоматическая классификация и кластеризация текстов.	ПК-7	Решение задач

### 8.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности в процессе текущего контроля



Текущий контроль успеваемости осуществляется путем оценки результатов выполнения заданий практических (семинарских) занятий, самостоятельной работы, предусмотренных учебным планом и посещения занятий/активность на занятиях.

В качестве оценочных средств текущего контроля успеваемости предусмотрены:

### **контрольная работа**

Контрольная работа № 1: состоит из 4 теоретических вопросов и 2 задач на методы оценки качества информационного поиска и модели информационного поиска.

*Типовые задачи контрольной:*

1. По запросу две системы выдали документы в такой последовательности:

Система 1. RNNNNR, больше релевантных документов найдено не было

Система 2. RNNNNNNNR - больше релевантных документов найдено не было.

Какая система ищет лучше по мере средней точности. Показать расчет по соотв. формуле.

2. В системе булевского поиска по заданному запросу эксперт нашел 100 релевантных документов. Система нашла 200 релевантных документов, из них 50 документов найдено правильно. Какова полнота, точность и F мера качества поиска по данному запросу

3. Запрос к поисковой системе состоит из двух слов: a b. На этот запрос выданы следующие документы:

1. a b c d
2. a a a
3. b b c
4. a b b c

При выдаче на запрос пользователь отметил релевантными документ 1 и документ 4. Записать вектора представления запроса и документов: только частоты слов, без нормализации, без idf. Как изменится вектор запроса при применении подхода relevance feedback, если учет исходного запроса идет с коэффициентом 0.7, а релевантных документов с коэффициентом 0.3.

4. Составить матрицу переходов для вычисления PageRank для данного набора страниц (коэффициент телепортации - 0.1)

Контрольная работа № 2: состоит из 4 теоретических вопросов и 2 задач, связанных с приложениями информационного поиска

*Типовые задачи контрольной:*

1. Какое предложение, выданное автоматическими системами, будет лучше оценено по мерам Rouge1 и Rouge2.
2. Дано расположение точек на координатной плоскости. Показать, как будет работать алгоритм кластеризации k-means.

3. Даны вектора документов. Какие документы первые соединятся в кластер при методе аггломеративной кластеризации. Представление документов - вектор частот, мера близости - косинусная.

### 8.3. Перечень вопросов и заданий для промежуточной аттестации по дисциплине

#### Вопросы к зачету

1. Определение сферы информационного поиска, задачи информационного поиска,
2. Информационно-поисковые системы различной направленности.
3. Архитектура информационно-поисковых систем
4. Модели информационного поиска.
5. Оценка качества информационного поиска .
6. Методы расширения информационно-поисковых запросов.
7. Вопросно-ответные системы.
8. Диалоговые системы
9. Анализ ссылок, логи запросов, анализ кликов, персонализация выдачи, поисковые сессии, словосочетания и близость расположения, тематические вероятностные модели.
10. Комбинирование факторов. Модели Learning to-rank..
11. Типы классификации.
12. Тематическая классификация и анализ тональности.
13. Методы классификации. Особенности методов классификации.
14. Методы автоматической кластеризации.
15. Автоматическое аннотирование.

#### Типовые задания зачета:

1. Имеется два класса документов. В каждом классе содержатся документы, представленные своими векторами:

Класс 1

D1: (2, 2, 1, 1)

D2: (4, 2, 1,3)

Класс 2

D3: (1, 1, 2, 2)

D4: (3, 1, 2, 4)

К какому классу будет относиться документ: D5 (2, 1, 2, 2) по методу Россіо

2. Определите сходство запроса с документами по языковой модели информационного поиска. Текстовая коллекция состоит из двух документов. Лямбда равна 0.5.

Запрос: королева красоты

Документы:

1. В Италии победительницу конкурса красоты облили серной кислотой

2. Бывший парень облил итальянскую модель кислотой

3. Задача: определить класс документа D4 методом Байеса

- Даны документы и их классы C1 и C2
- D1=(X1,X2, X3) C1
- D2=(X1, X2, X4, X7) C2
- D3=(X4, X5, X6) C2
- Определить класс документа:
- D4 (X2, X4, X6, X7)

#### 8.4. Планируемые уровни сформированности компетенций обучающихся и критерии оценивания

Уровни	Содержательное описание уровня	Основные признаки выделения уровня (этапы формирования компетенции, критерии оценки сформированности)	Пятибалльная шкала (академическая) оценка	Двухбалльная шкала, зачет	БРС, % освоения (рейтинговая оценка)
Повышенный	Творческая деятельность	<i>Включает нижестоящий уровень.</i> Умение самостоятельно принимать решение, решать проблему/задачу теоретического и прикладного характера на основе изученных методов, приемов, технологий	отлично	зачтено	86-100
Базовый	Применение знаний и умений в более широких контекстах учебной и профессиональной деятельности, нежели по образцу с большей степени самостоятельности и инициативы	<i>Включает нижестоящий уровень.</i> Способность собирать, систематизировать, анализировать и грамотно использовать информацию из самостоятельно найденных теоретических источников и иллюстрировать ими теоретические положения или обосновывать практику применения	хорошо		71-85
Удовлетворительный (достаточный)	Репродуктивная деятельность	Изложение в пределах задач курса теоретически и практически контролируемого материала	удовлетворительно		55-70
Недостаточный	Отсутствие признаков удовлетворительного уровня		неудовлетворительно	не зачтено	Менее 55

## **9. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.**

### **Основная литература**

1. Григорьев, А. А. Методы и алгоритмы обработки данных : учебное пособие / А.А. Григорьев, Е.А. Исаев. — 2-е изд., перераб. и доп. — Москва : ИНФРА-М, 2022. — 383 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/1032305. - ISBN 978-5-16-015581-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1862852> (дата обращения: 04.04.2023). – Режим доступа: по подписке.
2. **Дополнительная литература**
  1. Гребенюк, М. П. Методы обработки статистических данных : учебно-методическое пособие к выполнению лабораторной работы для студентов специальности ТБЖ / М. П. Гребенюк, Н. А. Корниенко. - Москва : РУТ (МИИТ), 2018. - 23 с. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1896882> (дата обращения: 04.04.2023). – Режим доступа: по подписке.

## **10. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля).**

- НЭБ Национальная электронная библиотека, диссертации и прочие издания
- ЭБС Консультант студента
- ПРОСПЕКТ ЭБС
- ЭБС ZNANIUM.COM
- ЭБС IBOOKS.RU
- Электронно-библиотечная система (ЭБС) Кантитана (<https://elib.kantiana.ru/>)

## **11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.**

Программное обеспечение обучения включает в себя:

- система электронного образовательного контента БФУ им. И. Канта – [www.lms3.kantiana.ru](http://www.lms3.kantiana.ru), обеспечивающую разработку и комплексное использование электронных образовательных ресурсов;
- серверное программное обеспечение, необходимое для функционирования сервера и связи с системой электронного обучения через Интернет;
- корпоративная платформа [webinar.ru](http://webinar.ru);
- установленное на рабочих местах обучающихся ПО: Microsoft Windows 7, Microsoft Office Standart 2010, антивирусное программное обеспечение Kaspersky Endpoint Security.
- СУБД PostgreSQL (Свободное ПО, лицензия - Freeware).
- MongoDB (Свободное ПО, лицензия - Freeware).
- Python 2.7.15 (Anaconda2 5.2.0 64-bit)
- Python 3.6.5 (Anaconda3 5.2.0 64-bit)

## **12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.**

Для проведения занятий лекционного типа, практических и семинарских занятий используются специальные помещения (учебные аудитории), оборудованные техническими

средствами обучения – мультимедийной проекционной техникой. Для проведения занятий лекционного типа используются наборы демонстрационного оборудования.

Для проведения лабораторных работ, (практических занятий – при необходимости) используются специальные помещения (учебные аудитории), оснащенные специализированным лабораторным оборудованием: персональными компьютерами с возможностью выхода в интернет и с установленным программным обеспечением, заявленным в п.11.

Для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации используются специальные помещения (учебные аудитории), оборудованные специализированной мебелью (для обучающихся), меловой / маркерной доской.

Для организации самостоятельной работы обучающимся предоставляются помещения, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

Для обучения инвалидов и лиц с ограниченными возможностями здоровья университетом могут быть представлены специализированные средства обучения, в том числе технические средства коллективного и индивидуального пользования.