


Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
Московский государственный университет имени М.В. Ломоносова  
Факультет вычислительной математики и кибернетики

**УТВЕРЖДАЮ**  
декан факультета вычислительной  
математики и кибернетики

**/И.А. Соколов /**  
**«27» сентября 2022г.**

## **ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

**по дисциплине**

**Методы и средства информационного поиска**

---

**Уровень высшего образования:**

**бакалавриат**

**Направление подготовки / специальность:**

**01.03.02 "Прикладная математика и информатика" (3++)**

**Направленность (профиль) ОПОП:**

**Искусственный интеллект и анализ данных**

**Форма обучения:**

**очная**

Рассмотрен и утвержден

*на заседании Ученого совета факультета ВМК*

*(протокол №7, от 27 сентября 2022 года)*

Москва 2022

## 1. ФОРМЫ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В процессе и по завершении изучения дисциплины оценивается формирование у студентов следующих компетенций:

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-7. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	<p>ПК-7.1. Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях</p> <p>ПК-7.2. Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения</p>	<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>

### 1.1. Текущий контроль успеваемости

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. RNNNNNNRR - 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.

## 1.2. Промежуточная аттестация

Промежуточная аттестация осуществляется в форме зачет

В качестве средств, используемых на промежуточной аттестации предусматривается:

Билеты

## 1.3. Типовые задания для проведения промежуточной аттестации

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 Россio

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:
- $D4 (X2, X4, X6, X7)$

## 2. КРИТЕРИИ ОЦЕНКИ ПО ДИСЦИПЛИНЕ

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
виды оценочных средств				
<b>Знания</b> (виды оценочных средств: приведены в п. 1.2.)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
<b>Умения</b> (виды оценочных средств: приведены в п. 1.2.)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности не принципиального характера)	Успешное и систематическое умение
<b>Навыки</b> (владения, опыт деятельности) (виды оценочных средств: приведены в п. 1.2..)	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