

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Максимов Алексей Борисович

Должность: директор департамента по образовательной политике

Дата подписания: 03.11.2023 14:46:44

Уникальный программный ключ:

8db180d1a3f02ac9e60521a5672742735c18b1d6

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»  
Факультет информационных технологий

УТВЕРЖДЕНО

Декан факультета

Информационных технологий

Д.Г. Демидов /



«16»

02

2023 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Информационный поиск**

Направление подготовки

**09.03.02 «Информационные системы и технологии»**

Профиль

**«Цифровая трансформация»**

Квалификация (степень) выпускника

**Бакалавр**

Формы обучения

**Очная**

**Москва 2023 г.**

**Разработчик(и):**

к.э.н., доцент, доцент кафедры  
«Информатика и информационные технологии»

/Н.А. Панова/

**Согласовано:**

Заведующий кафедрой «Информатика  
и информационные технологии»,  
к.т.н., доцент



/Е.В. Булатников/

## Содержание

1. Цели, задачи и планируемые результаты обучения по дисциплине.....	4
2. Место дисциплины в структуре ОП бакалавриата. ....	5
3. Структура и содержание дисциплины. ....	6
3.1 Виды учебной работы и трудоемкость (по формам обучения) .....	6
<b>3.2 Тематический план изучения дисциплины</b> .....	6
3.3 Содержание дисциплины.....	7
<b>3.4 Тематика семинарских/практических и лабораторных занятий</b> .....	9
<b>3.5 Тематика курсовых проектов (курсовых работ)</b> .....	10
4. Учебно-методическое и информационное обеспечение .....	10
4.1 Нормативные документы и ГОСТы.....	10
4.2 Основная литература.....	10
4.3 Дополнительная литература.....	11
4.4 Электронные образовательные ресурсы.....	11
4.5 Лицензионное и свободно распространяемое программное обеспечение .....	11
5. Материально-техническое обеспечение .....	11
6. Методические рекомендации .....	12
<b>6.1 Методические рекомендации для преподавателя по организации обучения</b> ...	12
7. Фонд оценочных средств .....	12
7.1 Методы контроля и оценивания результатов обучения .....	13
<b>7.2 Шкала и критерии оценивания результатов обучения</b> .....	13
<b>7.3 Оценочные средства</b> .....	15

## 1. Цели, задачи и планируемые результаты обучения по дисциплине

**Целью** дисциплины «Информационный поиск» является ознакомление обучающихся с базовыми понятиями теории информационного поиска, основными принципами организации информационно-поисковых систем, алгоритмами аналитико-синтетической переработки документов, а также методами обработки слабоструктурированных данных и извлечения информации, приобретение обучающимися навыков поиска материалов, формирования запросов, проверке достоверности полученной информации и работе с ней, изучение методов извлечения отношений, анализа тональности, аннотирования и кластеризации текстов с существующими программными реализациями этих методов.

**Задачами** изучения дисциплины являются:

- изучение информационного поиска как научной дисциплины;
- освоение методологических основ информационного поиска, его принципов, целей и задач;
- развитие умений и навыков поиска, проверки достоверности информации, формулирование запросов;
- изучение основ построения моделей информационно-поисковой системы.

В результате освоения дисциплины у обучающихся формируются следующие компетенции и должны быть достигнуты следующие результаты обучения как этап формирования соответствующих компетенций:

<b>Код компетенции</b>	<b>В результате освоения образовательной программы обучающийся должен обладать</b>	<b>Перечень планируемых результатов обучения по дисциплине</b>
ОПК-1	Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ИОПК-1.1. знает основы математики, физики, вычислительной техники и программирования ИОПК-1.2. умеет решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования ИОПК-1.3. имеет навыки теоретического и экспериментального исследования объектов профессиональной деятельности

ОПК-2	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ИОПК-2.1. знает современные информационные технологии и программные средства, основные виды и принципы работы информационных систем и информационных технологий; способы внедрения и интеграции современных информационных систем, способы оценки необходимости использования программных средств ИОПК-2.2. умеет использовать современные информационные технологии и программные средства, как в рамках отдельного предприятия, так и в рамках корпораций, государственных систем; внедрять и настраивать современные информационные системы, проводить интеграцию различных информационных систем и программных средств, оценивать необходимость использования программного средства для решения задач ИОПК-2.3. владеет навыками применения современных информационных технологий и программных средств, при решении задач в различных отраслях, внедрения и настройки современных информационных систем, оценки необходимости использования программных средств и информационных систем для решения задач
-------	--	--

## 2. Место дисциплины в структуре ОП бакалавриата.

Дисциплина «Информационный поиск» относится к обязательной части Блока 1 образовательной программы бакалавриата.

Дисциплина «Информационный поиск» взаимосвязана логически и содержательно-методически со следующими дисциплинами и практиками ОП:

- Цифровая культура компании;
- Основы цифровой трансформации;
- Информационная безопасность и защита информации;
- Распределенные компьютерные системы сбора и обработки данных;
- Цифровые экосистемы;
- Организация цифрового бизнеса;
- Корпоративные системы обработки данных;
- Нормативно-правовое регулирование цифровой среды;
- Документационное обеспечение цифровых процессов
- Проектно-технологическая практика;
- Научно-исследовательская работа;
- Преддипломная практика;

–Государственная итоговая аттестация (выполнение и защита ВКР).

### 3. Структура и содержание дисциплины.

Общая трудоемкость дисциплины составляет 5 зачетных единиц (180 часов).

#### 3.1 Виды учебной работы и трудоемкость (по формам обучения)

Форма обучения	курс	семестр	Трудоемкость дисциплины в часах						Форма итогового контроля
			Всего час./зач. ед.	Аудиторных часов	Лекции	Семинарские (практические) занятия	Лабораторные работы	Самостоятельная работа	
Очная	1	1	180/5	96	48	48	-	84	экзамен

#### Очная ф.о.

Вид учебной работы	Всего часов/зач. ед	Семестры
		1
<b>Аудиторные занятия (всего)</b>	<b>96</b>	<b>96</b>
<i>В том числе:</i>		
Лекции	48	48
Практические занятия (ПЗ)	48	48
Семинары (С)	-	-
Лабораторные работы (ЛР)	-	-
<b>Самостоятельная работа (всего)</b>	<b>84</b>	<b>84</b>
<i>В том числе:</i>		
Подготовка к практическим занятиям	52	52
Тестирование	30	30
<b>Вид промежуточной аттестации – зачет/экзамен</b>	<b>экзамен</b>	<b>экзамен</b>
<b>Общая трудоемкость час./зач. ед.</b>	<b>180/5</b>	<b>180/5</b>

#### 3.2 Тематический план изучения дисциплины (по формам обучения)

№ п/п	Раздел/тема дисциплины	Общая трудоемкость	Виды учебных занятий, включая самостоятельную работу обучающихся, час.	
			Контактная работа	Самостоятельная работа

		Всего	Лекции	Практические занятия	
1.	Тема 1. Информация – сущность, содержание и виды	15	4	4	7
2.	Тема 2. Введение в обработку естественного языка. Основные понятия информационного поиска	15	4	4	7
3.	Тема 3. Классификация и кластеризация текстов.	15	4	4	7
4.	Тема 4. Информационно-поисковые системы	15	4	4	7
5.	Тема 5. Булев поиск и фразовый поиск	15	4	4	7
6.	Тема 6. Оптимизация поиска	15	4	4	7
7.	Тема 7. Нечеткий поиск	15	4	4	7
8.	Тема 8. Ранжирование в информационном поиске	15	4	4	7
9.	Тема 9. Оценка результатов поиска	15	4	4	7
10.	Тема 10. Введение в машинный перевод	15	4	4	7
11.	Тема 11. Введение в извлечение информации.	15	4	4	7
12.	Тема 12. Методы машинного обучения в задаче извлечения информации.	15	4	4	7
13.	<b>ВСЕГО</b>	180	48	48	84
14.	<b>Экзамен</b>	–	–	–	–
15.	<b>ИТОГО</b>	180	48	48	84

### 3.3 Содержание дисциплины

#### **Тема 1. Информация – сущность, содержание и виды**

Сущность понятия «информация». Виды и критерии классификации информации. Информация как основа проведения информационного поиска.

#### **Тема 2. Введение в обработку естественного языка. Основные понятия информационного поиска**

История информационного поиска. Цели, задачи и проблемы поиска. Этапы поиска. Понятие поискового запроса. Точность, релевантность и полнота поиска. Виды информационного поиска.

Этапы анализа текста. Обзор основных приложений автоматического анализа текста (АОТ) (машинный перевод, информационный поиск, и т.д.). Слова, фразы, предложения, корпуса. Языковые модели. Автоматический морфологический анализ и синтез. Виды морфологического анализа: стемминг, лемматизация, полный морфоанализ. Принципы морфоанализа на базе словаря

основ или словаря словоформ. Морфологические процессоры для русского языка.

### **Тема 3. Классификация и кластеризация текстов.**

Классификация текстов как типичная задача обработки текстов в области TextMining. Обзор методов машинной классификации. Выбор признаков и метрик. Особенности кластеризации текстов. Рубрицирование текстовых документов. Обзор задач АОТ, решаемых на основе классификации текстов. Модели и методы автоматической классификации и кластеризации текстовой информации. Иерархические и вероятностные подходы. Интеллектуальный анализ данных.

### **Тема 4. Информационно-поисковые системы**

Сущность информационно-поисковой системы. Функции информационно-поисковых систем. Существующие современные информационно-поисковые системы, в том числе разработанные российскими производителями. Основные элементы информационно-поисковой системы: модуль индексирования, база данных, поисковый сервер.

### **Тема 5. Булев поиск и фразовый поиск**

Сущность понятия «булев поиск». Логические операторы «и», «или», «не». Инвертированный индекс и его построение. Матрица инцидентности. Проблемы кодирования и определения языка в информационном поиске. Сущность и содержание понятия «токен» («лексема»). Лемматизация и нормализация текста.

Сущность, содержание и применение фразового поиска. Двусловный индекс, его построение и использование в фразовом поиске. Координатный индекс, его построение и использование в фразовом поиске. Использование комбинированных схем при фразовом поиске.

### **Тема 6. Оптимизация поиска**

Необходимость и целесообразность оптимизации поиска. Направления оптимизации поиска. Сжатие индекса. Задачи сжатия индекса. Сжатие индекса «с потерями» и «без потерь». Стемминг и использование «стоп-слов». Сокращение словаря. Законы Ципра и Хигса. Ускорение поиска. Инкрементальное индексирование.

### **Тема 7. Нечеткий поиск**

Сущность и содержание понятия «нечеткий поиск». Необходимость и целесообразность использования алгоритмов нечеткого поиска. Использование дерева поиска. Перестановочный индекс, его построение и применение в нечетком поиске. Понятия «k-граммный индекс», «k-грамма». Недостатки нечеткого поиска. Применение нечеткого поиска при исправлении ошибок, опечаток.



### **Тема 8. Ранжирование в информационном поиске**

Сущность и содержание понятия «ранжирование», использование данного понятия в информационном поиске. Виды ранжирования: текстовое, вероятностное, ссылочное, поведенческое.

### **Тема 9. Оценка результатов поиска**

Оценка поиска и ее необходимость. Задачи оценки поиска. Оценка поиска с точки зрения разработчика. Оценка поиска с точки зрения пользователя. Количественные показатели оценки поиска. Качественные показатели оценки поиска. Подходы к оценке качественных показателей поиска: интерактивный, он-лайн метрики, офф-лайн метрики. Исходные данные для оценки. Точность, полнота и релевантность поиска как основа для оценки поиска.

### **Тема 10. Введение в машинный перевод**

Стратегии машинного перевода, основанного на лингвистических правилах. Статистический машинный перевод: особенности и виды. Принципы создания статистического переводчика.

### **Тема 11. Введение в извлечение информации.**

Основные способы представления смысла текста и модели представления знаний в искусственном интеллекте: семантические сети, язык предикатов. Семантический анализ текста на основе семантико-синтаксических моделей управления. Разметка частей речи. Выделение именованных сущностей. Извлечение информации и отношений из текста. Извлечение информации и знаний из текстов: особенности задачи и типы извлекаемых объектов. Понятие лингвистического шаблона для извлечения информации. Инструментальные программные средства для построения систем извлечения информации из текстов. Извлечение знаний под управлением онтологий в системах класса OntosMiner.

### **Тема 12. Методы машинного обучения в задаче извлечения информации.**

Формальные методы определения автора текста. Лингвостатистические параметры. Статистические методы атрибуции. Авторский инвариант и лингвистические спектры. Применение методов кластеризации и классификации для установления авторства текстов. Методы обнаружения спама: вероятностные и статистические, байесовский классификатор.

## **3.4 Тематика семинарских/практических и лабораторных занятий**

Практическое занятие 1. Информация – сущность, содержание и виды

Практическое занятие 2. Введение в обработку естественного языка. Основные понятия информационного поиска

Практическое занятие 3. Классификация и кластеризация текстов.

Практическое занятие 4. Информационно-поисковые системы

Практическое занятие 5. Булев поиск и фразовый поиск

Практическое занятие 6. Оптимизация поиска  
Практическое занятие 7. Нечеткий поиск  
Практическое занятие 8. Ранжирование в информационном поиске  
Практическое занятие 9. Оценка результатов поиска  
Практическое занятие 10. Введение в машинный перевод  
Практическое занятие 11. Введение в извлечение информации.  
Практическое занятие 12. Методы машинного обучения в задаче извлечения информации.

### **3.5 Тематика курсовых проектов (курсовых работ)**

Курсовой проект не предусмотрен

## **4. Учебно-методическое и информационное обеспечение**

### **4.1 Нормативные документы и ГОСТы**

1. Федеральный закон от 29 декабря 2012 года N 273-ФЗ «Об образовании в Российской Федерации» (с изменениями и дополнениями);
2. Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.02 Информационные системы и технологии, утвержденный Приказом Министерства образования и науки РФ от 19 сентября 2017 г. N 929 "Об утверждении федерального... Редакция с изменениями N 1456 от 26.11.2020;
3. Приказ Министерства образования и науки РФ от 05 апреля 2017 г. N 301 «Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры.
4. Федеральный закон от 27 июля 2006 г. N 149-ФЗ "Об информации, информационных технологиях и о защите информации" (с изменениями и дополнениями)

### **4.2 Основная литература**

1. Гасанов, Э. Э. Интеллектуальные системы. Теория хранения и поиска информации : учебник для вузов / Э. Э. Гасанов, В. Б. Кудрявцев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2021. — 271 с.  
<https://urait.ru/bcode/471008>
2. Сакова, О. Я. Аналитико-синтетическая переработка информации. Библиографическое описание информационных ресурсов : учебное пособие для вузов / О. Я. Сакова. — 2-е изд. — Москва : Издательство Юрайт, 2021 ; Кемерово : Кемеров. гос. ин-т культуры. — 123 с.  
<https://urait.ru/bcode/477587>

#### **4.3Дополнительная литература**

1. Федеральный проект «Цифровая экономика РФ»  
[https://digital.gov.ru/ru/activity/directions/858/?utm\\_referrer=https%3a%2f%2fwww.google.com%2f](https://digital.gov.ru/ru/activity/directions/858/?utm_referrer=https%3a%2f%2fwww.google.com%2f)
2. Цифровая организация. Тенденции и практики применения в России  
[https://filearchive.cnews.ru/files/reviews/forum2016/plenum/3\\_Shushkin.pdf](https://filearchive.cnews.ru/files/reviews/forum2016/plenum/3_Shushkin.pdf)
3. Что такое digital business <https://union-sp.ru/blog/chto-takoe-digital-biznes/>
4. Цифровые бизнес-модели <https://morethandigital.info/ru/11-tzifroviye-biznyes-modyeli-kotoriye-vi-dol%CA%90ni-znat-vkl-primyeri/>
5. Стратегический проект «Цифровой бизнес» <https://misis.ru/university/prioritet-2030/digital-business/>

#### **4.4Электронные образовательные ресурсы**

LMS-курс «Информационный поиск»

<https://online.mospolytech.ru/course/view.php?id=8664>

#### **4.5Лицензионное и свободно распространяемое программное обеспечение**

Программы пакета MicrosoftOffice (Word, Excel, PowerPoint)

Браузеры и поисковые системы «Яндекс», «Google»

Национальный корпус русского языка

#### **4.6 Современные профессиональные базы данных и информационные справочные системы:**

1. Справочно-правовая системы «КонсультантПлюс: Некоммерческая интернет-версия» <https://www.consultant.ru/online/>
2. Официальный интернет-портал правовой информации <http://pravo.gov.ru>
3. Российская национальная библиотека <http://www.nlr.ru>
4. ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/index.php>
5. Единое окно доступа к образовательным ресурсам Федеральный портал <http://window.edu.ru>
6. Научная электронная библиотека <http://www.elibrary.ru>
7. Российская государственная библиотека <http://www.rsl.ru>

#### **5.Материально-техническое обеспечение**

Лекционные аудитории общего фонда.

Компьютерный класс для проведения лабораторных работ.

Настенный/ переносной экран.

Переносной/ стационарный проектор для демонстрации слайдов.

Ноутбук для демонстрации слайдов.

Цифровая доска.

Библиотека, читальный зал.

## **6.Методические рекомендации**

### **6.1 Методические рекомендации для преподавателя по организации обучения**

Данный раздел настоящей рабочей программы предназначен для начинающих преподавателей и специалистов-практиков, не имеющих опыта преподавательской работы. Методика преподавания дисциплины «Информационный поиск» и реализация компетентностного подхода в изложении и восприятии материала предусматривает использование онлайн-курса в системе дистанционного обучения Университета, групповых и индивидуальных консультаций обучающихся с целью формирования и развития общепрофессиональных навыков.

Подробное содержание отдельных разделов дисциплины «Информационный поиск» рассматривается в п.5.3 рабочей программы. Примеры тестовых заданий для текущего и промежуточного контроля по дисциплине представлены в составе ФОС по дисциплине в п.8 настоящей рабочей программы.

Перечень основной и дополнительной литературы, баз данных и информационных справочных систем, необходимых в ходе преподавания дисциплины «Информационный поиск», приведен в п.6 настоящей рабочей программы.

### **6.2. Методические указания обучающимся**

Получение углубленных знаний по дисциплине достигается за счет активной самостоятельной работы обучающихся. Выделяемые часы целесообразно использовать для знакомства с учебной и научной литературой по проблемам дисциплины, анализа научных концепций. Промежуточная аттестация обучающихся в форме экзамена проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине проводится преподавателем, ведущим занятия по дисциплине. К промежуточной аттестации допускаются только обучающиеся, выполнившие все виды учебной работы, предусмотренные рабочей программой по дисциплине «Информационный поиск».

## **7.Фонд оценочных средств**

## **7.1 Методы контроля и оценивания результатов обучения**

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций: **практические задания, тестирование, экзамен.**

## **7.2 Шкала и критерии оценивания результатов обучения**

К промежуточной аттестации допускаются только студенты, выполнившие все виды учебной работы, предусмотренные рабочей программой по дисциплине «Организация цифрового бизнеса».

### **7.2.1 Критерии оценки ответа на экзамене**

#### **«Отлично»:**

Выполнены все виды учебной работы, предусмотренные учебным планом. Обучающийся выполнил и защитил все практические задания со средним баллом от 4,5 до 5. Итоговое тестирование выполнено на 85 — 100%. Обучающийся демонстрирует прочные теоретические знания, практические навыки, владеет терминами, делает аргументированные выводы и обобщения, приводит примеры, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, которые обучающийся может исправить самостоятельно.

#### **«Хорошо»:**

Выполнены все виды учебной работы, предусмотренные учебным планом. Обучающийся выполнил и защитил все практические задания со средним баллом от 4 до 4,5. Итоговое тестирование выполнено на 70 — 84%. Обучающийся демонстрирует достаточные теоретические знания, практические навыки, владеет терминами, делает аргументированные выводы и обобщения, приводит примеры, оперирует приобретенными знаниями, умениями, навыками. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации, которые обучающийся может исправить при незначительной коррекции преподавателем.

#### **«Удовлетворительно»:**

Выполнены все виды учебной работы, предусмотренные учебным планом. Обучающийся выполнил и защитил все практические задания со средним баллом ниже 4. Итоговое тестирование выполнено на 55 — 69%. Обучающийся демонстрирует неполное соответствие теоретических знаний, практических навыков, владеет терминами, делает аргументированные выводы и обобщения, приводит примеры, оперирует приобретенными знаниями, умениями, навыками. При этом могут быть допущены ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации, которые обучающийся может исправить при коррекции преподавателем.

#### **«Неудовлетворительно»:**

Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Обучающийся не выполнил одно или более заданий текущего и промежуточного контроля. Итоговое тестирование выполнено на 0 — 54%. Обучающийся демонстрирует незнание теоретических основ предмета, отсутствие практических навыков, не умеет делать аргументированные выводы и приводить примеры, не владеет терминами, проявляет отсутствие логичности и последовательности изложения, делает ошибки, которые не может исправить даже при коррекции преподавателем, отказывается отвечать на дополнительные вопросы, допускает значительные ошибки, испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

### **7.2.2. Критерии оценки работы обучающегося на практических занятиях**

**«5» (отлично):** выполнены все задания, предусмотренные на практическом занятии, обучающийся четко и без ошибок ответил на все контрольные вопросы, активно работал на занятиях, своевременно сдал и защитил работу.

**«4» (хорошо):** выполнены все задания, предусмотренные на практическом занятии, есть некоторые недочеты, обучающийся с корректирующими замечаниями преподавателя ответил на все контрольные вопросы, достаточно активно работал на занятиях, своевременно сдал и защитил работу.

**«3» (удовлетворительно):** выполнены все задания, предусмотренные на практическом занятии с замечаниями преподавателя; обучающийся ответил на все контрольные вопросы с замечаниями, сдал и защитил работу позже установленного срока.

**«2» (неудовлетворительно):** обучающийся не выполнил или выполнил неправильно задания, предусмотренные на практическом занятии, студент ответил на контрольные вопросы с ошибками или не ответил на контрольные вопросы, не сдал и не защитил практическую работу.

### **7.2.3. Критерии оценки тестирования**

Тестирование оценивается в соответствии с процентом правильных ответов, данных студентом на вопросы теста.

Стандартная шкала соответствия результатов тестирования выставяемой балльной оценке:

- «отлично» - свыше 85% правильных ответов;
- «хорошо» - от 70,1% до 85% правильных ответов;
- «удовлетворительно» - от 55,1% до 70% правильных ответов;
- от 0 до 55% правильных ответов – «неудовлетворительно»

**«5» (отлично):** тестируемый демонстрирует системные теоретические знания, владеет терминами и обладает способностью быстро реагировать на вопросы теста.

**«4» (хорошо):** тестируемый в целом демонстрирует системные теоретические знания, владеет большинством терминов и обладает способностью быстро реагировать на вопросы теста.

**«3» (удовлетворительно):** системные теоретические знания у тестируемого отсутствуют, он владеет некоторыми терминами и на вопросы теста реагирует достаточно медленно.

**«2» (неудовлетворительно):** системные теоретические знания у тестируемого отсутствуют, терминологией он не владеет и на вопросы теста реагирует медленно.

## 7.3 Оценочные средства

### 7.3.1. Текущий контроль (задания на практических занятиях)

Пример заданий текущего контроля:

1. Составить регулярное выражение, удовлетворяющее заданным требованиям.
2. Построить наиболее вероятную цепочку тегов (скрытых состояний) в заданной скрытой марковской модели по указанному предложению.
3. Вывести формулу для коэффициентов заданного алгоритма сглаживания  $n$ -граммной языковой модели.
4. Построить символьную триграммную языковую модель по заданному корпусу и с ее помощью построить распознаватель языка документа.
5. Вычислить перплексию  $n$ -граммной языковой модели с заданным сглаживанием.
6. На основе заданной обучающей выборки построить марковскую модель максимальной энтропии для выделения заданных именованных сущностей (имен собственных, географических названий и т. д.) из текста.

### 7.3.2. Промежуточный контроль (экзамен)

Экзамен может проводиться в форме компьютерного тестирования.

Пример тестового задания:

1. Информация - это:

- + сведения об окружающем мире и протекающих в нем процессах
- только то, что выдается в ответ на запрос
- только данные выраженные в форме слов или цифр
- опыт человека

2. Найдите соответствие:

виды информации по способу восприятия -> визуальная, звуковая, тактильная

виды информации по способу представления -> текстовая, графическая, видео, числовая

виды информации по общественному значению -> массовая, специальная, личная

3. Возможные проблемы, связанные с информацией в интернете:

- +Авторские права
- +Достоверность информации
- +Конфиденциальность информации
- +Платный или бесплатный доступ
- Доступность информации
- Полная недостоверность информации

4. Найдите соответствие между разными аспектами информации и определениями:

синтаксический -> с точки зрения техники передачи информации

семантический -> с точки зрения содержания и правильности толкования информации

прагматический -> с точки зрения полезности информации для потребителя

5. Найдите соответствие между терминами, имеющими отношение к информации:

потребитель -> субъект использующий информацию

код -> форма, в которой информация передается

источник информации -> элемент окружающего мира, сведения о котором являются объектами преобразования

6. Качество информации:

+это соответствие ее характеристик целям и задачам поиска

-это достоверность данных в выдаче информационно-поисковой системы

-это степень удовлетворенности поиском

-это объем информации в цифровом выражении

### **Примерные вопросы для подготовки к экзамену:**

1. Сущность и содержание понятия «Информация»
2. Виды информации в информационном поиске
3. Понятие информации
4. Виды и критерии классификации информации
5. История информационного поиска
6. Цели и задачи информационного поиска
7. Проблемы информационного поиска.
8. Поисковый запрос.
9. Основные виды информационного поиска.
10. Информационно-поисковые системы и их сущность.
11. Функции информационно-поисковых систем
12. Элементы информационно-поисковой системы
13. Булев поиск



14. Логические операторы, используемые в информационном поиске
15. Инвертированный индекс и его построение.
16. Матрица инцидентности и ее использование в информационном поиске.
17. Кодирование и его влияние на информационный поиск.
18. Определение языка в информационном поиске.
19. Сущность понятия «токен».
20. Лемматизация и нормализация текста.
21. Сущность и область применения фразового поиска.
22. Использование двусловного индекса для целей фразового поиска.
23. Построение, достоинства и недостатки двусловного индекса.
24. Координатный индекс.
25. Использование нескольких видов индексов в фразовом поиске.
26. Необходимость и целесообразность оптимизации поиска.
27. Основные направления оптимизации поиска.
28. Задачи оптимизации поиска.
29. Сжатие индекса при оптимизации поиска.
30. Сжатие индекса «с потерями» - область применения, достоинства и недостатки.
31. Сжатие индекса «без потерь» - область применения, достоинства и недостатки.
32. Сущность понятия «стемминг»
33. Стоп-слова и их использование для оптимизации поиска.
34. Сокращение словаря.
35. Законы Ципра и Хигса.
36. Ускорение поиска.
37. Инкрементальное индексирование.
38. Сущность и содержание понятия «нечеткий поиск».
39. Область применения нечеткого поиска.
40. Использование дерева поиска.
41. Перестановочный индекс, его построение и применение в нечетком поиске.
42. Понятия «k-граммный индекс», «k-грамма».
43. Недостатки нечеткого поиска.
44. Сущность и содержание понятия «ранжирование»
45. Виды ранжирования и области их применения, достоинства и недостатки
46. Текстовое ранжирование
47. Вероятностное ранжирование
48. Поведенческое ранжирование
49. Ссылочное ранжирование
50. Оценка поиска и ее необходимость.
51. Задачи оценки поиска.
52. Оценка поиска с точки зрения разработчика.
53. Оценка поиска с точки зрения пользователя.
54. Количественные показатели оценки поиска.
55. Качественные показатели оценки поиска.

56. Интерактивный подход к оценке качественных показателей поиска.
57. Он-лайн метрики при оценке качественных показателей поиска
58. Офф-лайн метрики при оценке качественных показателей поиска.
59. Исходные данные для оценки информационного поиска и его результатов
60. Расчет точности, полноты, релевантности поиска.