

Документ подписан простой электронной подписью

Информация о владельце:

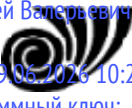
ФИО: Таскаев Сергей Валерьевич

Должность: Ректор

Дата подписания: 29.06.2026 10:25:50

Уникальный программный ключ:

04c19ed8bfb98f356cb77a486b9a8788b8522523

	МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)
Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю) "Современные технологии поиска и обработки информации" по направлению подготовки (специальности) 03.03.03 "Радиофизика" направленности (профилю) "Телекоммуникационные системы и информационные технологии" ФГБОУ ВО «ЧелГУ»	

стр. 1

**Фонд оценочных средств для промежуточной аттестации
по дисциплине (модулю)
Современные технологии поиска и обработки информации**

Направление подготовки (специальность)
03.03.03 Радиофизика

Направленность (профиль)
Телекоммуникационные системы и информационные технологии

Присваиваемая квалификация (степень)
Бакалавр

Форма обучения
Очная

Год набора 2026

Челябинск, 2026 г.



МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю) "Современные технологии поиска и обработки информации" по направлению подготовки (специальности) 03.03.03 "Радиофизика" направленности (профилю) "Телекоммуникационные системы и информационные технологии" ФГБОУ ВО «ЧелГУ»

стр. 2

Содержание

1. Паспорт фонда оценочных средств
2. Перечень формируемых компетенций
 - 2.1. Компетенции, закреплённые за дисциплиной
3. Содержание оценочных средств по дисциплине
 - 3.1. Виды оценочных средств
 - 3.2. Содержание оценочных средств
4. Порядок проведения и критерии оценивания промежуточной аттестации
 - 4.1. Порядок проведения промежуточной аттестации
 - 4.2. Критерии оценивания промежуточной аттестации по видам оценочных средств
 - 4.3. Результаты промежуточной аттестации и уровни сформированности компетенций



1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

Направление подготовки: *03.03.03 «Радиофизика»*

Направленность: Телекоммуникационные системы и информационные технологии

Дисциплина: Современные технологии поиска и обработки информации

Семестр изучения: *1 семестр*

Форма промежуточной аттестации: *зачет*

2. ПЕРЕЧЕНЬ ФОРМИРУЕМЫХ КОМПЕТЕНЦИЙ

2.1. Компетенции, закреплённые за дисциплиной

Изучение дисциплины «К.М.01.01 Современные технологии поиска и обработки информации» направлено на формирование следующих компетенций:

Коды компетенции согласно ФГОС (ОПОП ВО)	Содержание компетенций согласно ФГОС (ОПОП ВО)	Индикаторы достижения компетенции согласно ОПОП	Перечень планируемых результатов обучения по дисциплине
1	2	3	4
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Выполняет поиск информации, определяет критерии системного анализа поставленных задач УК-1.2. Использует критический анализ, систематизацию и обобщение информации для решения поставленных задач.	Знать: Основы выполнения эффективного поиска информации Уметь: Определять критерии системного анализа для поставленных задач Владеть: Навыками системного анализа и поиска информации
ОПК-3	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-3.1. Имеет представление об основных существующих информационных технологиях, используемых при решении профессиональных задач. ОПК-3.2. Демонстрирует умения использовать существующие информационные технологии при решении задач профессиональной деятельности. ОПК-3.3. Имеет практический опыт использования существующих информационных технологий при решении задач профессиональной деятельности.	Знать: Основные существующие информационные технологии, которые используются при решении задач профессиональной деятельности. Уметь: Использовать существующие информационные технологии для решения задач профессиональной деятельности. Владеть: Основными существующими информационными технологиями



3. СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

3.1 Виды оценочных средств

Код, наименование компетенции согласно ФГОС	Перечень планируемых результатов обучения по дисциплине	Контролируемые темы/разделы	Се-местр	Номер задания	Наименование оценочного средства
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	Знать: Основы выполнения эффективного поиска информации. Уметь: Определять критерии системного анализа для поставленных задач. Владеть: Навыками системного анализа и поиска информации.	Информационный поиск, Электронные библиотечные системы, Поисковые системы, основы поиска в вебе, Лексикон и списки словопозиций, Словари и нечеткий поиск, Сжатие индекса.	1	Самостоятельные работы № 1-3. Теоретические вопросы № 1-14.	Самостоятельные работы. Теоретические вопросы к зачету.
ОПК-3 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	Знать: Основные существующие информационные технологии, которые используются при решении задач профессиональной деятельности. Уметь: Использовать существующие информационные технологии для решения задач профессиональной деятельности. Владеть: Основными существующими информационными технологиями при решении задач профессиональной деятельности.	Булев поиск, Инвертированный индекс, Поисковые системы, Лексикон и списки словопозиций, Словари и нечеткий поиск, Сжатие индекса.	1	Самостоятельные работы № 4-6. Теоретические вопросы № 15-29.	Самостоятельные работы. Теоретические вопросы к зачету.



3.2 Содержание оценочных средств

Вопросы для подготовки к зачёту:

1. Понятие информационного поиска, его цель и задачи;
2. Классификация информационного поиска;
3. Модель булева поиска;
4. Определение эффективности системы информационного поиска: точность и полнота;
5. Инвертированный индекс. Этапы его построения;
6. Оптимизация запроса;
7. Структурная единица документа. Проблема детализации индексирования;
8. Лексема и термин: определение и различия;
9. Стоп-слова: определение, способ создания списка стоп-слов, использование стоп-слов в системах информационного поиска;
10. Нормализация лексем. Классы эквивалентности;
11. Индексация ударений и диакритических знаков;
12. Индексирование заглавных букв;
13. Стемминг. Пример;
14. Лемматизация. Пример;
15. Указатели пропуска;
16. Фразовые запросы. Двухсловные индексы;
17. Фразовые запросы. Координатный индекс;
18. Комбинированная схема обработки фразовых запросов;
19. Реализация словаря. Хеширование;
20. Реализация словаря. Деревья поиска;
21. Запросы с джокером и их обработка;
22. К-граммный индекс для обработки запросов с джокером;
23. Реализация исправления опечаток;
24. Фонетические исправления. Soundex-индекс;
25. Характеристики аппаратного обеспечения, влияющие на обработку запросов;
26. Архитектура MapReduce;
27. Сжатие словаря: цель и методы;
28. Сжатие инвертированного файла;
29. Классификация текстов.



Самостоятельная работа №1 по теме «Булев поиск».

Вариант 1

1. Прочитайте следующие документы:

Документ 1.

В Интернете с каждым днём скапливается всё больше информации, когда-либо созданной и вновь создаваемой людьми. Равнодоступность большей части информации в Интернете уравнивает возможности доступа к этой информации как обычных пользователей Интернета и журналистов локальных СМИ, так и сотрудников мировых информационных агентств. Благодаря Интернету перед каждым человеком ежедневно и даже ежесекундно открывается доступ к многомиллионной аудитории, которой он может передать свой информационный материал, полученный, например, с помощью обычного мобильного телефона с диктофоном и встроенной фотокамерой. Следовательно, уровень монополизации деятельности по распространению информации также снижается благодаря Интернету.

Документ 2.

До недавнего времени ограничения в прямой коммуникации между людьми, порождаемые пространством и временем, во многом определяли потребность людей в услугах журналистов. По мере роста общего количества пользователей Интернета, а среди них – числа владеющих английским языком, эти ограничения всё в большей степени снимаются, что закономерно ведёт к уменьшению спроса на услуги журналистов. Одновременно с этим растёт объем «сырой» информации, доступной каждому отдельному пользователю Интернета, что актуализирует проблему её отбора и редактирования. Последнее всегда входило в перечень функций журналистики, но с ростом числа пользователей Интернета эффективный информационный поиск начинает приобретать всё большую значимость не только в журналистской деятельности, но и в других разнообразных сферах общественной деятельности. Таким образом, информационный поиск – это процесс поиска неструктурированной документальной информации.



Документ 3.

Поиск информации представляет собой процесс выявления в некотором множестве документов (текстов), которые посвящены заданной теме (предмету) и удовлетворяют заранее определенному условию поиска (запросу), а также содержат необходимые (соответствующие информационной потребности) факты, сведения и данные. Процесс поиска включает последовательность операций, направленных на сбор, обработку и предоставление необходимой информации заинтересованным лицам.

Документ 4.

Комплекс программ, предназначенных для информационного поиска, называется поисковой машиной. Обычно является частью поисковой системы – автоматизированного программно-аппаратного комплекса с веб-интерфейсом, предоставляющего возможность поиска информации в Интернете. Самая известная поисковая система в мире – это Google, самая популярная в России – Яндекс, а одной из самых старых поисковых систем является Yahoo. Как уже было отмечено ранее, в архитектуре поисковой системы можно выделить поисковую машину – ядро системы, представленное набором программных модулей; базу данных или индекс, хранящую информацию обо всех известных поисковой системе Интернет ресурсах; и набор сайтов, являющих собой точки входа пользователей в систему. Все это соответствует классической трехуровневой архитектуре информационных систем: есть пользовательский интерфейс, бизнес логика, которая в данном случае представлена реализацией алгоритмов поиска и база данных.

Документ 5.

Для того, чтобы найти в Интернете требуемую информацию, необходимо знать либо адрес её местоположения (например, адрес html-страницы или файла), либо пользователя Интернета, который может предоставить информацию. Если мы не знаем ни адреса, ни человека, который мог бы нам помочь, то следует перейти к вопросам «Как можно узнать адрес размещения информации?» или «Как найти человека, который мог бы нам помочь с поиском информации?». При этом не следует переоценивать возможности Интернета. Лучшие результаты может дать совмещение онлайн-овых и оффлайн-овых методов поиска информации.



Документ 6.

Сегодня существует достаточно большое количество методов информационного поиска в Интернете и через Интернет. В каждом конкретном случае успешность поиска определяется знаниями возможных методов и навыками владения ими, знанием этнических языков, на которых эта информация может быть представлена, либо нашими социальными связями.

2. Постройте матрицу инцидентности «термин — документ» для терминов: Интернет, поиск, запрос, индекс, информация, система, машина. Словоформа и регистр значения не имеют. Термины расположите в таблице в алфавитном порядке.

3. Обработайте запрос поиск AND информация AND NOT интернет, взяв векторы для терминов из матрицы инцидентности и выполнив поразрядные операции AND и NOT. В ответе укажите список документов.

4. Составьте инвертированный список для коллекции документов Вашего варианта используя в качестве словаря термины из пункта 2.

Самостоятельная работа №2 по теме «Лексикон и списки словопозиций»

Вариант 1

1. Разбейте текст каждого документа на лексемы:

Документ 1.

Состав и интерпретация поддерживаемых метасимволов. Часто называется «диалектом» регулярного выражения.

Документ 2.

Особенности взаимодействия регулярных выражений с языком или программой.

Документ 3.

Специфика применения регулярных выражений к тексту.



2. Выполните предварительную лингвистическую обработку лексем с помощью нормализации лексем и игнорирования стоп-слов (союзы, предлоги).

3. Составьте инвертированный индекс для данной коллекции документов с указанием частоты и списка словопозиций для каждого термина. Термины расположите в алфавитном порядке.

4. Нормализуйте лексемы в запросе: *регулярные AND выражения* и обработайте его.

Самостоятельная работа №3
по теме «Словопозиции с координатами и фразовые запросы»

Вариант 1

1. Разбейте текст каждого документа на лексемы:

Документ 1.

Как правило, поисковые машины поддерживают два режима: режим простого поиска и режим расширенного поиска.

Документ 2.

Можно просто вводить через пробел одно или несколько слов; поиск слов всевозможными окончаниями моделируется символом * в конце слова.

Документ 3.

Многие системы позволяют искать словосочетания или фразу, для этого необходимо ее заключить в кавычки.

Документ 4.

Возможно обязательное включение или исключение определенных слов.

Документ 5.

Основная проблема поиска по примитивно составленному запросу заключается в том, что поисковая машина найдет все страницы, на которых указанные слова встречаются в любой части документа.



Документ 6.

Как правило, количество найденных страниц будет слишком велико.

2. Составьте координатный индекс для данной коллекции документов в формате термин док1: <позиция1, позиция2, ...>; док2: <позиция1, позиция2, ...>; и т.д., используя следующий словарь терминов: поиск, машина, запрос, режим, слово, документ, страница. Термины расположите в алфавитном порядке.

3. Какие документы удовлетворяют фразовому запросу поисковая машина?

4. Укажите набор документов, удовлетворяющих запросу: режим /2 поиск.

Самостоятельная работа №4 по теме «Поиск по сходству»

Вариант 1

1. С помощью перестановочного индекса определите термины лексикона **поисковик, иск, поиск, постфикс, пост, пуск**, соответствующие запросу **по*ск**.

2. Используя алгоритм динамического программирования, вычислите расстояние редактирования между строками **сильный** и **стильный**.

3. Составьте биграммный индекс терминов лексикона **стена, струна, сторона, оборона, рана, охрана** и найдите термины, содержащие по крайней мере две из всех биграмм в запросе **страна**.

4. Найдите коэффициент Жаккара для строк запроса и каждого термина, полученного в предыдущем пункте. Для каких терминов этот коэффициент наибольший?

Самостоятельная работа №5 по теме «Фонетические исправления»

Вариант 1

1. Используя алгоритм фонетического хеширования и таблицу кодирования русских букв составьте soundex-индекс для следующего словаря терминов:

собор, луг, забор, сбор, информатика, запор, инфракрасный, инвентарь, зубр, лук



МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю) "Современные технологии поиска и обработки информации" по направлению подготовки (специальности) 03.03.03 "Радиофизика" направленности (профилю) "Телекоммуникационные системы и информационные технологии" ФГБОУ ВО «ЧелГУ»

стр. 11

Таблица кодирования русских букв

0. у, е, ё, ы, а, о, э, я, и, ю, ь, ъ
1. б, п
2. ф, в
3. ж, з, с, х
4. к, г
5. ц, ч, ш, щ
6. д, т
7. л, й
8. м, н
9. р

2. Какие термины имеют одинаковые Soundex-коды?

3. Преобразуйте термин запроса **СОБР** в Soundex-код и выполните поиск по soundex-индексу.

Самостоятельная работа №6 по теме «Сжатие индекса»

Вариант 1

1. Создайте блочное хранение ($k = 4$) для следующего словаря терминов: поиска, информации, поиску, сбора, информационный, поисковой, сборный, информация, сбору, поисковик, сборного, информацию.

2. Выполните дальнейшее сжатие с помощью фронтальной упаковки. Для обозначения общего префикса используйте символ \diamond (Alt-код символа Alt+9674).

3. Вычислите коды, полученные с помощью схемы байтового кодирования для следующего инвертированного списка: $\langle 4, 10, 11, 12, 15, 62, 63, 265, 268, 270, 400 \rangle$ Запишите бинарные коды блоками по 8 бит.



4. ПОРЯДОК ПРОВЕДЕНИЯ И КРИТЕРИИ ОЦЕНИВАНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

4.1. Порядок проведения промежуточной аттестации

Зачет проводится в 2 этапа. На первом этапе студенту необходимо предоставить все самостоятельные работы, которые были заданы в течении семестра. На втором этапе студент отвечает на два теоретических вопроса. Продолжительность 1 час.

4.2. Критерии оценивания промежуточной аттестации по видам оценочных средств.

4.2.1 Критерии оценивания теоретического вопроса

Максимальный балл за ответ на теоретический вопрос — 10 баллов.

Зачтено/ 9-10 баллов	Зачтено/ 7-8 баллов	Зачтено/ 5-6 баллов	Незачтено/ 0-4 балла
Высокий уровень освоения проверяемых компетенций	Средний уровень освоения проверяемых компетенций	Базовый уровень освоения проверяемых компетенций	Недостаточный уровень освоения проверяемых компетенций
Обучающийся отлично знает материал, умеет анализировать проблему и решает её, применяя знания и навыки, полученные на занятиях и в ходе самостоятельной работы. Способен аргументировано изложить свою точку зрения на поставленную проблему. Обучающийся практически не допускает ошибок.	Обучающийся хорошо знает материал, умеет анализировать проблему, решает её, применяя знания и навыки, полученные на занятиях, способен аргументировано изложить свою точку зрения, Обучающийся допускает незначительные ошибки.	Обучающийся знаком с материалом, владеет основными навыками, полученными в ходе практических занятий. Обучающийся допускает фактические ошибки.	Обучающийся не знает основных положений вопроса, не ориентируется в основных понятиях, излагает материал с трудом, с грубыми фактическими ошибками, либо отказывается от ответов на вопросы.



4.3. Результаты промежуточной аттестации и уровни сформированности компетенций

При подведении итогов учитываются результаты текущей аттестации. Полученные за текущую аттестацию баллы суммируются с баллами, полученными за каждый этап при прохождении промежуточной аттестации:

0-40 баллов - незачтено;

41-50 баллов – зачтено.

Особенности проведения процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья обозначены в рабочей программе дисциплины (модуля).

Уровни сформированности компетенций определяется следующим образом:

1. Высокий уровень сформированности компетенций соответствует оценке зачтено:
 - предполагает формирование компетенций на высоком уровне, готовность к самостоятельной профессиональной деятельности;
 - студент способен аргументировать собственную точку зрения по дискуссионным вопросам дисциплины.
2. Средний уровень соответствует оценке зачтено:
 - предполагает формирование компетенций на более высоком уровне;
 - студент способен давать развернутые ответы на теоретические вопросы дисциплины на уровне не ниже оценки «удовлетворительно».
3. Базовый уровень соответствует оценке зачтено:
 - предполагает формирование компетенций на начальном уровне: знание основных этапов построения инвертированного индекса;
 - студент способен отвечать на вопросы в форме закрытого теста. Количество правильных ответов – не менее 50%.
4. Низкий уровень соответствует оценке незачтено.

