

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Воронежский государственный технический университет»

УТВЕРЖДАЮ

Декан ФЭМИТ

Баркалов С.А.



2023 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

«Технологии информационного поиска и анализа web-контента»

Направление подготовки 27.03.03 Системный анализ и управление

Профиль Бизнес-аналитика и системы больших данных

Квалификация выпускника бакалавр

Нормативный период обучения 4 года

Форма обучения очная

Год начала подготовки 2023

Автор программы

Мерозов В.Д.

Заведующий кафедрой

управления

Баркалов С.А.

Руководитель ОПОП

Терешченко Д.С.

Воронеж 2023

## 1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

**1.1. Цели дисциплины:** знакомство студентов с технологиями информационного поиска и анализа web-контента, а также приобретение навыков исследования данных и разработки web-приложений.

### 1.2. Задачи освоения дисциплины

- знакомство студентов с базовыми технологиями построения информационно-поисковых веб-сайтов;
- знакомство с основными видами информационно-поисковых веб-сайтов и принципами их проектирования;
- получение навыков проведения анализа web-контента и представления его результатов заказчику;
- развитие самостоятельности при создании информационно-поисковых веб-сайтов с использованием изученных технологий.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Технологии информационного поиска и анализа web-контента» относится к дисциплинам обязательной части блока Б1.

## 3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Технологии информационного поиска и анализа web-контента» направлен на формирование следующих компетенций:

ПК-1 - способен подготавливать данные и формировать требования к результатам аналитических работ в области анализа данных;

ПК-2 - способен проводить аналитические исследования с помощью методов системного анализа в соответствии с требованиями заказчика;

ПК-3 – способен осуществлять проектирование систем управления среднего и крупного масштаба и сложности

| Компетенция | Результаты обучения, характеризующие сформированность компетенции   |
|-------------|---|
| ПК-1        | Знать основные принципы подготовки данных и формирования требований к результатам аналитических работ в области анализа web-контента; |
|             | Уметь применять современные отечественные и зарубежные технологии для создания информационно-поисковых веб-сайтов                     |
|             | Владеть навыками программирования и проектирования информационно-поисковых веб-сайтов   |
| ПК-2        | Знать основы проведения анализа web-контента и представления его результатов заказчику  |

|      |  |  |
|------|--|--|
|      |  | Уметь проводить и представлять результаты анализа web-контента                             |
|      |  | Владеть- навыками и приёмами проведения и оформления результатов анализа web-контента      |
| ПК-3 |  | Знать – основы проектирования информационно-поисковых веб-сайтов                           |
|      |  | Уметь – применять технологию проектирования информационно-поисковых веб-сайтов на практике |
|      |  | Владеть – приёмами и методами проектирования информационно-поисковых веб-сайтов            |

#### 4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Технологии информационного поиска и анализа web-контента» составляет 4 з.е.

Распределение трудоемкости дисциплины по видам занятий  
**очная форма обучения**

| Виды учебной работы                             | Всего часов | Семестры |
|---|-------------|----------|
|   |             | 1        |
| <b>Аудиторные занятия (всего)</b>               | 54          | 54       |
| В том числе:                                    |             |          |
| Лекции  | 18          | 18       |
| Лабораторные работы (ЛР)                        | 36          | 36       |
| Практические работы (ПР)                        | 16          | 16       |
| <b>Самостоятельная работа</b>                   | 74          | 74       |
| <b>Курсовая работа</b>                          | +           | +        |
| Виды промежуточной аттестации - зачет с оценкой | +           | +        |
| Общая трудоемкость:<br>академические часы       | 144         | 144      |
| зач.ед.   | 4           | 4        |

#### 5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

**5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий**

**очная форма обучения**

| № п/п | Наименование темы   | Содержание раздела   | Лек ц | Лаб. зан. | Пр | СРС | Всего, час |
|-------|---|--|-------|-----------|----|-----|------------|
| 1     | Введение в информационный поиск и обработку естественного языка     | Основные понятия информационного поиска и обработки естественного языка. Модели информационного поиска. Булевская модель.  | 4     | 8         | 4  | 12  | 28         |
| 2     | Словари и нечёткий поиск  | Поисковые структуры для словарей. Построение индекса. Сжатие индекса.  | 4     | 8         | 4  | 12  | 28         |
| 3     | Ранжирование, взвешивание терминов и модель векторного пространства | Модель векторного пространства:<br>– параметрические и зонные индексы;<br>– частота термина и взвешивание;<br>– модель векторного пространства для ранжирования; | 4     | 8         | 2  | 12  | 26         |

|              |  |   |           |           |           |           |            |
|--------------|--|---|-----------|-----------|-----------|-----------|------------|
|              |  | – варианты функций tf-idf.<br>Ранжирование в полнофункциональной поисковой системе:<br>– эффективное ранжирование<br>– компоненты информационно-поисковой системы<br>– влияние операторов языка запросов на ранжирование в векторном пространстве..   |           |           |           |           |            |
| 4            | Оценка информационного поиска  | XML-поиск. Факторы, влияющие на результат информационного поиска. Основные параметры оценки поисковой системы. Стандартные текстовые коллекции. Оценка неранжированных результатов поиска. Оценка ранжированных результатов поиска. Обратная связь по релевантности и расширение запроса. Качество системы и её полезность для пользователя. Снippets.  | 2         | 4         | 2         | 12        | 20         |
| 5            | Введение в обработку естественно-языковых текстов  | Общее понятие процесса обработки естественно-языковых текстов. Методика обработки естественно-языковых текстов и машинное обучение. Плоская кластеризация: Иерархическая кластеризация. Разложение матриц и латентно-семантическое индексирование. Основы поиска в вебе. Обход и индексирование веба.   | 2         | 4         | 2         | 12        | 20         |
| 6            | Программирование и проектирование систем обработки естественных языков: задачи морфологического анализа, морфологический разбор, стемминг, лемматизация. | 1. Обобщенный алгоритм обработки текста на естественном языке. 2. Морфологический и синтаксический разбор текстов. 3. Фазы обобщенного алгоритма обработки текста на естественном языке: синтаксический анализ, семантический анализ, прагматический анализ 4. Анализ и индексирование текста. 5. Стемминг. 6. Лемматизация. 7. Сравнение функций стемминга и лемматизации для поиска текста с учетом морфологии. | 2         | 4         | 2         | 14        | 22         |
| <b>Итого</b> |  |   | <b>18</b> | <b>36</b> | <b>16</b> | <b>74</b> | <b>144</b> |

## 5.2 Перечень лабораторных работ

|   |
|---|
| 1. Работа с поисковыми сервисами.                     |
| 2. Поиск информации в сети Интернет.                  |
| 3. Русскоязычные вербальные ИПС: сравнительный анализ |
| 4. Глобальные вербальные ИПС: изучение и поиск.       |
| 5. Глобальные вербальные ИПС: ранжирование выдачи     |
| 6. Метапоисковые системы.                             |

## 5.3 Перечень практических работ

Укажите перечень практических работ

|   |
|---|
| 1. Создание и оптимизация сайта.                    |
| 2. Регистрация сайта в поисковых системах.          |
| 3. Добавление в поисковые системы файла sitemap.xml |
| 4. Добавление в поисковые системы файла robots.txt. |
| 5. Добавление на сайт инструментов аналитики.       |
| 6. Анализ (seo-анализ) сайтов-конкурентов           |

## **6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ**

В соответствии с учебным планом освоение дисциплины предусматривает выполнение курсовой работы для очной формы обучения.

Примерная тематика курсовой работы:

1. Разработка информационно-поискового Web-сайта в области политических новостей
2. Разработка информационно-поискового Web-сайта в области экономических новостей
3. Разработка информационно-поискового Web-сайта в области культурных новостей
4. Разработка информационно-поискового Web-сайта в области спортивных новостей
5. Разработка информационно-поискового Web-сайта научных статей
6. Разработка информационно-поискового Web-сайта справочных данных
7. Разработка информационно-поискового Web-сайта учебных материалов
8. Разработка информационно-поискового Web-сайта научно-технической информации
9. Разработка информационно-поискового Web-сайта в области искусственного интеллекта
10. Разработка информационно-поискового Web-сайта в области медицины
11. Разработка информационно-поискового Web-сайта в области компьютерного зрения
12. Разработка информационно-поискового Web-сайта в области программной инженерии
13. Разработка информационно-поискового Web-сайта в области глубокого машинного обучения
14. Разработка информационно-поискового Web-сайта в области систем поддержки принятия решений
15. Разработка информационно-поискового Web-сайта в области экспертных систем

Задачи, решаемые при выполнении курсовой работы:

- Закрепить на практике основные теоретические положения, изученные на лекциях;
- Разработка основных видов веб-приложений и освоение принципов их проектирования;
- Получение навыков создания веб-приложений с помощью современных веб-технологий.

## 7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

### 7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

#### 7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

| Компетенция | Результаты обучения, характеризующие сформированность компетенции   | Критерии оценивания  | Аттестован  | Не аттестован   |
|-------------|---|--|---|---|
| ПК-1        | Знать основные принципы подготовки данных и формирования требований к результатам аналитических работ в области анализа web-контента; | Знает основные принципы подготовки данных и формирования требований к результатам аналитических работ в области анализа web-контента | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
|             | Уметь применять современные отечественные и зарубежные технологии для создания информационно-поисковых веб-сайтов                     | Умеет программировать web-приложение на основе современных языков и сред разработки  | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
|             | Владеть навыками программирования и проектирования информационно-поисковых веб-сайтов   | Владеет языками процедурного и объектно-ориентированного программирования  | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
| ПК-2        | Знать основы проведения анализа web-контента и представления его результатов заказчику  | Знает как установить web-приложение в сети Интернет  | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
|             | Уметь проводить и представлять результаты анализа web-контента  | Умеет выбрать компоненты языковых средств для работы web-приложения  | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
|             | Владеть навыками и приемами проведения и оформления результатов анализа web-контента  | Владеет методами установки web-приложения  | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |
| ПК-3        | Знать – основы проектирования информационно-поисковых веб-сайтов  | знает – основы проектирования и обеспечения целостности базы данных  | Выполнение работ в срок, предусмотренный в рабочих программах | Невыполнение работ в срок, предусмотренный в рабочих программах |

|  |  |  |   |  |   |
|--|--|--|---|--|---|
|  |  | web-приложений   |   |  |   |
|  | Уметь – применять технологию проектирования информационно-поисковых веб-сайтов на практике | умеет – проектировать базы данных web-приложений и обеспечивать их целостность             | Выполнение работ в срок, предусмотренный в рабочих программах |  | Невыполнение работ в срок, предусмотренный в рабочих программах |
|  | Владеть – приёмами и методами проектирования информационно-поисковых веб-сайтов            | владеет – технологиями проектирования и обеспечения целостности базы данных web-приложений | Выполнение работ в срок, предусмотренный в рабочих программах |  | Невыполнение работ в срок, предусмотренный в рабочих программах |

### 7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 1 семестре для очной формы обучения по четырехбалльной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно».

| Компетенция | Результаты обучения, характеризующие сформированность компетенции   | Критерии оценивания                                      | Отлично  | Хорошо  | Удовл.   | Неудовл.                             |
|-------------|---|--|--|---|--|--------------------------------------|
| ПК-1        | Знать основные принципы подготовки данных и формирования требований к результатам аналитических работ в области анализа web-контента; | Тест   | Выполнение теста на 90-100%                            | Выполнение теста на 80-90%  | Выполнение теста на 70-80%                               | В тесте менее 70% правильных ответов |
|             | Уметь применять современные отечественные и зарубежные технологии для создания информационно-поисковых веб-сайтов                     | Решение стандартных практических задач                   | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены                     |
|             | Владеть навыками программирования и проектирования информационно-поисковых веб-сайтов   | Решение прикладных задач в конкретной предметной области | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены                     |

|      |   |  |  |   |  |                                      |
|------|---|--|--|---|--|--------------------------------------|
| ПК-2 | Знать основы проведения проведения анализа web-контента и представления его результатов заказчику | Тест   | Выполнение теста на 90-100%                            | Выполнение теста на 80-90%  | Выполнение теста на 70-80%                               | В тесте менее 70% правильных ответов |
|      | Уметь проводить и представлять результаты анализа web-контента                                    | Решение стандартных практических задач                   | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены                     |
|      | Владеть- навыками и приемами проведения и оформления результатов анализа web-контента             | Решение прикладных задач в конкретной предметной области | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены                     |
| ПК-3 | Знать – основы проектирования информационно-поисковых веб-сайтов                                  | Тест   | Выполнение теста на 90-100%                            | Выполнение теста на 80-90%  | Выполнение теста на 70-80%                               | В тесте менее 70% правильных ответов |
|      | Уметь – применять технологию проектирования информационно-поисковых веб-сайтов на практике        | Решение стандартных практических задач                   | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены                     |
|      | Владеть – приемами и методами проектирования информационно-поисковых веб-сайтов                   | Решение прикладных задач в конкретной предметной области | Задачи решены в полном объеме и получены верные ответы | Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах | Продемонстрирован верный ход решения в большинстве задач | Задачи не решены                     |

## **7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)**

### **7.2.1 Примерный перечень заданий для подготовки к тестированию**

#### *Задание 1*

*1. Установите соответствие терминов и определений:*

*А) Данные 1) совокупность сведений, познаний в какой-либо области*

*Б) Информация 2) сведения, необходимые для какого-либо вывода, решения, процедуры*

*В) Знания 3) совокупность собранных и аналитически обработанных сведений, требующихся для принятия оптимального решения при устранении некоторой проблемы, а также сам процесс передачи или получения этих сведений.*

#### *Задание 2*

*Тег – это:*

*+ фрагмент текста на Web-странице*

*- адрес Web-документа*

*- адрес компьютера в сети*

*- команда языка программирования HTML*

#### *Задание 3*

*Выделите лишние элементы.*

*Ценность информации зависит от следующих характеристик:*

*1) полезность;*

*2) достоверность;*

*+3) сохранность;*

*4) своевременность;*

*5) полнота.*

#### *Задание 4*

*Назначение классификации источников информации:*

*1) индексация источников;*

*2) навигация в информационном потоке;*

*3) облегчение поиска;*

*4) идентификация источника*

*5) всё вместе.*

#### *Задание 5*

*Для того, чтобы установить на Web-странице отображение текста определенным шрифтом, используется атрибут тега <font>:*

*- kbd*

*- code*

*+ type*

*- cite*

#### *Задание 6*

*Чтобы найти наиболее достоверную документальную информацию целесообразно воспользоваться:*

*1) книгой;*

*2) журналом;*

*3) продолжающимся изданием;*

*4) трудами конференций;*

*5) непубликуемыми документами;*

*+6) описаниями патентов.*

#### *Задание 7*

*Добавление бегущей строки в Web-документ осуществляется тегом*

*- runstring*

- *stroka*
- *marquee*
- *target*

#### *Задание 8*

*Вставка рисунка в Web-документ осуществляется тегом*

- *photo*
- *img*
- *picture*
- *videoimg*

#### *Задание 9*

*К редакторам, в рамках которых может формироваться текст Web-документа, относятся:*

- + *WordPad, FrontPage Express, Notepad, MS Word*
- *FrontPage Express, HomeSite, PowerPoint*
- *MS Word, MS Excel, CorelDraw*
- *MS Windows, MS Word, HomeSite*

#### *Задание 10*

*Для выделения активного участка карты используется тег*

- *domain*
- *shape*
- *square*
- + *area*

### **7.2.2 Примерный перечень заданий для решения стандартных задач**

#### *Задание 1*

*Чтобы найти наиболее свежую документальную информацию целесообразно воспользоваться:*

- 1) *книгой;*
- 2) *журналом;*
- 3) *продолжающимся изданием;*
- 4) *трудами конференций;*
- 5) *непубликуемыми документами;*
- 6) *описаниями патентов.*

#### *Задание 2*

*Укажите рациональную последовательность поиска с использованием поисковых машин:*

- 1) *Отбор поисковых машин 1*
- 2) *Составление тезауруса 2*
- 3) *Определение географических регионов поиска 3*
- 4) *Формирование и выполнение запросов к поисковым машинам 4*
- 5) *Обработка результата запроса 5.*

#### *Задание 3*

*URL – это:*

- а) *протокол сетей ETHERNET;*
- б) *название сетей внутри WEB-пространства;*
- в) *адрес WEB-страницы, который отображается в поле АДРЕС браузера.*

#### *Задание 4*

*В какой части работы гипертекстовой информационной системы предусматривается непосредственное участие человека:*

- 1) *в работе поисковых машин;*
- 2) *при индексации информационных источников;*
- 3) *при классификации каталогов ресурсов..*

### Задание 5

Ключевые слова – это слова:

- 1) способные в совокупности представлять смысл текста;
- +2) формирующие существенные признаки текста;
- 3) имеющие максимальную частоту в тексте.

### Задание 6

Расширенный запрос...границы поиска:

- 1) сужает;
- 2) в зависимости от вида источника информации может расширять или сужать;
- 3) расширяет..

### Задание 7

Элементом HTML-документа называется:

- а) начальный дескриптор;
- б) заголовок WEB-страницы;
- в) начальный и конечный дескриптор вместе с заключенным между ними содержимым;
- г) символы, служебные слова, дескрипторы, фреймы, списки, типы полей формы, формы.

### Задание 8

Контейнер – это:

В чём инновационность поисковой системы Google:

- 1) большая скорость поискового робота Googlebot;
- 2) оригинальность интерфейса;
- 3) применение алгоритма ссылочного ранжирования PageRank.

### Задание 9

Гипермедиа:

- а) так называются средства представления информации в WEB-пространстве;
- б) это фильмы и анимация в INTERNET;
- в) это применение принципов гипертекста к средствам мультимедиа.

### Задание 10

Гипертекст – это:

- а) текст, который используется для передачи информации в WEB-пространстве;
- б) особый текст, в котором есть ссылка на другую WEB-страницу или документ;
- в) текст с линейной структурой.

## 7.2.3 Примерный перечень заданий для решения прикладных задач

### Задание 1

.... краткая характеристика содержания произведений печати или рукописи:

- 1) Проспект;
- 2) Бюллетень;
- 3) Реферативный сборник;
- 4) Аннотация.

### Задание 2

...аннотированный сборник публикаций, классифицированный по системе универсального десятичного классификатора:

- 1) Проспект;
- 2) Бюллетень;
- 3) Реферативный сборник;

#### 4) Аннотация.

##### Задание 3

CGI-сценарии – это программы, предназначенные для:

- а) обеспечения доступа к системам программирования;
- б) обработки информации, полученной от форм;
- в) взаимодействия с браузером.

##### Задание 4

... краткое информационное издание, рекламного характера с описанием товаров и условий их приобретения:

- 1) Проспект;
- 2) Бюллетень;
- 3) Реферативный сборник;
- 4) Аннотация.

##### Задание 5

HTML – формы:

- а) это специальные программы;
- б) предназначены для отображения и сбора информации;
- в) выполняют функцию обработки информации.

##### Задание 6

Фреймы используются для оформления следующих элементов:

- а) оглавления, форм, служебной информации о браузере;
- б) оглавления, неподвижных элементов интерфейса, форм и результатов;
- в) скрытой информации, которая используется браузером.

##### Задание 7

... краткое периодическое или продолжающееся информационное издание, посвященное какому-либо кругу вопросов, с включением графических изображений:

- 1) Проспект;
- 2) Бюллетень;
- 3) Реферативный сборник;
- 4) Аннотация.

##### Задание 8

Алгоритм анализа документальных источников информации включает:

- 1) анализ структуры документа по оглавлению;
- 2) оценка содержания по реферату;
- 3) библиографическая оценка документа;
- 4) проверка наличия в анализируемом источнике информации авторской гипотезы решения проблемы;
- 5) проверка наличия доказательства реальности выдвинутой автором гипотезы;
- 6) оценка степени практической реализации гипотезы автора и стадии её внедрения;
- 7) проведение критического анализа теоретических положений, предпосылок и выводов автора;
- 8) сопоставление условий проведения эксперимента, представленных в анализируемом источнике информации, и условий решения проблемы, стоящей перед исследователем;
- 9) выявление нерешенных вопросов, перспектив дальнейших усовершенствований;

- 10) выделение прототипа;
- 11) разработку гипотезу предстоящих исследований.

#### Задание 9

Какие технологии анализа электронной информации позволяют проводить тематический анализ текста:

- 1) OLAP-технологии;
- 2) Knowledge Discovery in Databases (KDD);
- 3) TextAnalyst;
- 4) Oracle InterMedia Text;
- 5) Russian Context Optimizer (RCO);
- 6) Система PolyAnalyst;
- 7) Контент-анализ.

#### Задание 10

Исследовательские инструменты контент-анализа:

- 1) пакет офисных программ;
- 2) классификатор контент-анализа;
- 3) протокол итогов анализа;
- 4) регистрационная карточка;
- 5) инструкция исследователю;
- 6) список проанализированных документов.

### 7.2.4 Примерный перечень вопросов для подготовки к зачету

Основные понятия информационного поиска и обработки естественного языка.

Модели информационного поиска.

Булевская модель.

Поисковые структуры для словарей.

Построение индекса.

Сжатие индекса.

Модель векторного пространства:

- параметрические и зонные индексы;
- частота термина и взвешивание;
- модель векторного пространства для ранжирования;
- варианты функций tf-idf.

Ранжирование в полнофункциональной поисковой системе:

- эффективное ранжирование
- компоненты информационно-поисковой системы
- влияние операторов языка запросов на ранжирование в векторном пространстве.

XML-поиск.

Факторы, влияющие на результат информационного поиска.

Основные параметры оценки поисковой системы.

Стандартные текстовые коллекции.

Оценка неранжированных результатов поиска.

Оценка ранжированных результатов поиска.

Обратная связь по релевантности и расширение запроса.

Качество системы и её полезность для пользователя.

Снипеты.

Общее понятие процесса обработки естественно-языковых текстов.

Методика обработки естественно-языковых текстов и машинное обучение.

Плоская кластеризация:

Иерархическая кластеризация.

Разложение матриц и латентно-семантическое индексирование.

Основы поиска в вебе.  
 Обход и индексирование веба.  
 Обобщенный алгоритм обработки текста на естественном языке.  
 Морфологический и синтаксический разбор текстов.  
 Фазы обобщенного алгоритма обработки текста на естественном языке: синтаксический анализ, семантический анализ, прагматический анализ.  
 Анализ и индексирование текста.  
 Стемминг.  
 Лемматизация.  
 Сравнение функций стемминга и лемматизации для поиска текста с учетом морфологии.

### **7.2.5 Примерный перечень заданий для решения прикладных задач**

Не предусмотрено учебным планом

### **7.2.6. Методика выставления оценки при проведении промежуточной аттестации**

*Зачёт проводится по билетам, каждый из которых содержит 2 вопроса. Каждый ответ на вопрос в билете оценивается по пяти бальной шкале*

*1. Оценка «Неудовлетворительно» ставится в случае, если студент не смог ответить на один из вопросов билета.*

*2. Оценка «Удовлетворительно» ставится в случае, если студент получил удовлетворительные оценки за ответы на вопросы билета.*

*3. Оценка «Хорошо» ставится в случае, если студент получил хорошие оценки за ответы на вопросы билета.*

*4. Оценка «Отлично» ставится, если студент получил отличные оценки за ответы на вопросы билета.*

### **7.2.7 Паспорт оценочных материалов**

| № п/п | Контролируемые разделы (темы) дисциплины                            | Код контролируемой компетенции | Наименование оценочного средства                                |
|-------|---|--------------------------------|---|
| 1     | Введение в информационный поиск и обработку естественного языка     | ПК-1, ПК-2, ПК-3               | Тест, защита лабораторных работ, требования к курсовому проекту |
| 2     | Словари и нечёткий поиск  | ПК-1, ПК-2, ПК-3               | Тест, защита лабораторных работ, требования к курсовому проекту |
| 3     | Ранжирование, взвешивание терминов и модель векторного пространства | ПК-1, ПК-2, ПК-3               | Тест, защита лабораторных работ, требования к курсовому проекту |
| 4     | Оценка информационного поиска                                       | ПК-1, ПК-2, ПК-3               | Тест, защита лабораторных работ, требования к курсовому проекту |
| 5     | Введение в обработку естественно-языковых текстов                   | ПК-1, ПК-2, ПК-3               | Тест, защита лабораторных работ, требования к курсовому проекту |
| 6     | Программирование и проек-   | ПК-1, ПК-2,                    | Тест, защита лабора-  |

|  |  |      |  |
|--|--|------|--|
|  | тирование систем обработки естественных языков: задачи морфологического анализа, морфологический разбор, стемминг, лемматизация. | ПК-3 | торных работ, требования к курсовому проекту |
|--|--|------|--|

### **7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методике выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методике выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методике выставления оценки при проведении промежуточной аттестации.

Защита курсовой работы, курсового проекта или отчета по всем видам практик осуществляется согласно требованиям, предъявляемым к работе, описанным в методических материалах. Примерное время защиты на одного студента составляет 20 мин.

## **8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **8.1 Перечень учебной литературы, необходимой для освоения дисциплины**

1. Хахаев И. А., Кучинский В. Ф. *Технологии обработки текстовой информации в LibreOffice* / СПб: НИУ ИТМО, 2016. – 340 с. (<https://e.lanbook.com/book/91441>)

2. Кузнецов М., Симдянов И., Гольшиев С.. *PHP5. Практика создания web-сайтов.*/ СПб: БХВ-Петербург, 2009 – 298 с.

3. Кандаурова Н.В., Чеканов В.С. *Технологии обработки информации: учебное пособие* / Ставрополь:СКФУ, 2014 – 175 с. (<https://biblioclub.ru/index.php?page=book&id=457753>)

4. Ли Н.И., Ахметшина А.И., Резванова Э. А. *Технология обработки текстовой информации: учебное пособие* / Казань: КНИТУ, 2016 – 84 с. (<https://biblioclub.ru/index.php?page=book&id=560807>)

5. Тузовский А.Ф. *Проектирование Интернет приложений: учебно-методическое пособие* / А.Ф. Тузовский; Томский политехнический университет. -Томск:Изд-Томского политехнического университета, 2010. - 200 с.

6. Сычев А.В. *Web-технологии. Учебный курс.* - Воронежский государственный университет, 2009 – 306 с.

7. Щербаков А.Ю. Интернет-аналитика: поиск и оценка информации в web-ресурсах: практическое пособие /М: Книжный мир, 2012 – 78 с. (<https://biblioclub.ru/index.php?page=book&id=89693>)

**8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:**

1. Справочник HTML и CSS. [Электронный ресурс]. Режим доступа: [htmlbook.ru](http://htmlbook.ru)

2. Справочник JavaScript [Электронный ресурс]. Режим доступа: [javascript.ru](http://javascript.ru)

3. PHP. Scripting language that is especially suited to web development. [Электронный ресурс]. Режим доступа: [php.net](http://php.net)

4. HTML5 WebSocket. [Электронный ресурс]. Режим доступа: [www.websocket.org](http://www.websocket.org)

5. Хабр – индустрия высоких технологий. [Электронный ресурс]. Режим доступа: [habr.com](http://habr.com)

6. Метаязык на основе CSS, предназначенный для увеличения уровня абстракции CSS кода и упрощения файлов каскадных таблиц стилей. [Электронный ресурс]. Режим доступа: [sass-scss.ru](http://sass-scss.ru)

7. PostCSS. Tool for transforming CSS with JavaScript. [Электронный ресурс]. Режим доступа: [postcss.org](http://postcss.org)

8. Less. Динамический язык стилей. [Электронный ресурс]. Режим доступа: [lesscss.org](http://lesscss.org)

9. Синтаксис регулярных выражений. [Электронный ресурс]. Режим доступа: [regexstudio.com](http://regexstudio.com)

## **9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА**

- 1.
2. Компьютерный класс, который позволяет реализовать образовательные возможности с доступом в сеть Интернет на скорости 6 мегабит в секунду. С возможностью проводить групповые занятия с обучаемыми, а также онлайн (оффлайн) тестирование.
3. Библиотечный электронный читальный зал с доступом к электронным ресурсам библиотек страны и мира. В количестве 3-х мест.
4. Персональный компьютер с предустановленным лицензионным программным обеспечением не ниже Windows XP, Office 2007, которое позволяет работать с видео-аудио материалами, создавать и демонстрировать презентации, с выходом в сеть Интернет
5. Ноутбук с предустановленным лицензионным программным обеспечением не ниже Windows XP, Office 2007, которое позволяет работать с видео-аудио материалами с выходом в сеть Интернет.

## 10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Технологии информационного поиска и анализа web-контента» читаются лекции, проводятся лабораторные работы, практические работы и выполняется курсовая работа.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Практические работы могут проводиться в виде семинаров, круглых столов и др.

Методика выполнения курсовой работы изложена в учебно-методическом пособии. Выполнять этапы курсовой работы должны своевременно и в установленные сроки.

Контроль усвоения материала дисциплины производится проверкой курсовой работы, защитой курсовой работы.

| Вид учебных занятий    | Деятельность студента  |
|------------------------|--|
| Лекция                 | Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометить важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии. |
| Лабораторная работа    | Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.   |
| Практическая работа    | Практические работы позволяют более детально разобраться с лекционным материалом и научиться применять теоретические знания, полученные на лекции для выполнения лабораторных работ. В рамках практических работ необходимо: разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и ответить на контрольные вопросы.   |
| Самостоятельная работа | Самостоятельная работа студентов способствует глубоко-   |

|  |  |
|--|--|
|  | <p>му усвоения учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие:</p> <ul style="list-style-type: none"> <li>- работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций;</li> <li>- выполнение домашних заданий и расчетов;</li> <li>- работа над темами для самостоятельного изучения;</li> <li>- участие в работе студенческих научных конференций, олимпиад;</li> <li>- подготовка к промежуточной аттестации.</li> </ul> |
| <p>Подготовка к промежуточной аттестации</p> | <p>Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед зачетом с оценкой три дня эффективнее всего использовать для повторения и систематизации материала.</p>   |