

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
"Казанский (Приволжский) федеральный университет"  
Высшая школа информационных технологий и интеллектуальных систем



**УТВЕРЖДАЮ**

Проректор по образовательной деятельности КФУ

\_\_\_\_\_ Д.А. Таюрский

"\_\_" \_\_\_\_\_ 20\_\_ г.

## **Программа дисциплины**

Основы информационного поиска

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Технологии разработки информационных систем

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

## **Содержание**

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
  - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
  - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и) доцент, к.н. (доцент) Невзорова О.А. (Кафедра информационных систем, отделение фундаментальной информатики и информационных технологий), Olga.Nevzorova@kpfu.ru

## 1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-4	Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения, способность формализовать предметную область программного проекта и разработать спецификации для компонентов программного продукта
УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- а) основы теории информационного поиска;
- б) основные принципы организации информационно-поисковых систем;
- в) основные этапы аналитической обработки документов.

Должен уметь:

- а) осуществлять обоснованный выбор технологии построения информационно-поисковой системы;
- б) практически реализовывать основные алгоритмы информационного поиска.

Должен владеть:

- а) основными технологиями построения информационно-поисковых систем;
- б) основными алгоритмами аналитической обработки документов.

Должен демонстрировать способность и готовность:

- понимать основные принципы информационного поиска;
- уметь реализовывать изученные алгоритмы и применять изученные технологии на практике;
- понимать принцип работы поисковых систем;
- пользоваться изученными методами поиска.

## 2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.11 Дисциплины (модули)" основной профессиональной образовательной программы 09.03.04 "Программная инженерия (Технологии разработки информационных систем)" и относится к вариативной части.

Осваивается на 4 курсе в 8 семестре.

## 3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 5 зачетных(ые) единиц(ы) на 180 часа(ов).

Контактная работа - 36 часа(ов), в том числе лекции - 18 часа(ов), практические занятия - 18 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 99 часа(ов).

Контроль (зачёт / экзамен) - 45 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 8 семестре.

## 4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

**4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)**

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)			Самостоятельная работа
			Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Многоуровневая модель информации	8	4	2	0	19
2.	Тема 2. Метаданные и обработка электронных ресурсов	8	4	4	0	20
3.	Тема 3. Наивный байесовский классификатор	8	2	4	0	20
4.	Тема 4. Векторная модель документа	8	4	4	0	20
5.	Тема 5. Анализ связей и ранжирование	8	4	4	0	20
	Итого		18	18	0	99

**4.2 Содержание дисциплины (модуля)****Тема 1. Многоуровневая модель информации**

Соотношение между понятиями "данные", "информация" и "знания". Информатика и семиотика. Проблема восприятия информации. Знак, концепт, денотат - треугольник Фреге. Связь между многоуровневой моделью информации и семиотическими моделями. Общие принципы организации информационно-поисковых систем. Знакомство с ГОСТ.73-96 "Поиск и распространение информации" и ГОСТ 7.74-96 "Информационно-поисковые языки". Проблема индексации документов. Проект Semantic Web: достоинства и недостатки.

**Тема 2. Метаданные и обработка электронных ресурсов**

Типы и иерархия метаданных. Особенности аналитико-синтетической переработки электронных документов. Модель информационно-поисковой системы. Основные принципы системного анализа. Формулировка требований к информационно-поисковой системе. Модели поиска: контекстный, атрибутивный, "по аналогии". Структура логических компонентов информационно-поисковой системы. Создание каталогов информационно-поисковых систем. Построение тезаурусов и онтологий информационно-поисковых систем. Типы тезаурусов и онтологий. Автоматизированная технология построения тезаурусов и онтологий. Извлечение метаданных из слабоструктурированных документов.

**Тема 3. Наивный байесовский классификатор**

Условная вероятность. Вероятность независимых событий. Теорема Байеса. Формула полной вероятности. Задача бинарной классификации. Логарифм отношения правдоподобия. Байесовская фильтрация нежелательных электронных писем ("спама"). Наивный байесовский классификатор для нескольких классов. Реализация наивного байесовского классификатора.

**Тема 4. Векторная модель документа**

Частота термина (tf). Обратная частота документа (idf). Вес по tf-idf. Представление документа в виде вектора. Интерпретация угла между векторами документов. Косинусная мера схожести документов. Локальные весовые функции: бинарная, tf, log-tf, augnorm. Глобальные весовые функции: бинарная, нормальная, ldf, gldf, энтропия. Ограничения и недостатки векторной модели документа. Латентно-семантический анализ. Терм-документная матрица. Свойства строк и столбцов терм-документной матрицы. Сингулярное разложение матрицы. Существование сингулярного разложения. Численные методы нахождения сингулярного разложения. Латентно-семантическое индексирование.

**Тема 5. Анализ связей и ранжирование**

Графовые структуры. Представление графовых структур в виде матриц. Разреженные матрицы. Представление разреженных матриц. Ранжирующий вектор PageRank. Поиск собственного вектора матрицы. Степенной метод поиска собственного значения. Степенной метод в вычислительной модели MapReduce. Цепи Маркова. Вектор начальных состояний. Матрица переходных вероятностей. Стационарное распределение цепи Маркова. Условие существования и единственности стационарного распределения. Связь PageRank и цепей Маркова.

**5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)**

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства образования и науки Российской Федерации от 5 апреля 2017 года №301)

Письмо Министерства образования Российской Федерации №14-55-996ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

## **6. Фонд оценочных средств по дисциплине (модулю)**

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

## **7. Перечень литературы, необходимой для освоения дисциплины (модуля)**

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

## **8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

Information Retrieval Wiki - <http://ir.dcs.gla.ac.uk/wiki/>

Introduction to Information Retrieval - <http://nlp.stanford.edu/IR-book/>

Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных - <http://www.machinelearning.ru/>

### 9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	Работа по лекциям включает в себя работу до лекции, работу во время лекции и работу после лекции. Студенты знакомы с учебным планом и преподаватель заранее сообщает тему следующей лекции. Студент должен ознакомиться с темой по материалам в сети Интернет, в виртуальной аудитории. Вопросы во время лекции поощряются по оценке преподавателя. После лекции материал прорабатывается и используется в лабораторных работах
практические занятия	Практические занятия выполняются по темам, определенным учебным планом. Легенды для конкретной работы предлагаются преподавателем. Каждая работа завершается отчетом. В отчете должны быть четко определены постановка задачи, используемый инструментарий, пути решения задачи, подробный ход решения задачи, выводы. Приветствуется обсуждение и возможные альтернативные варианты решения. Инструментарий зависит от имеющегося программного обеспечения.
самостоятельная работа	Самостоятельная работа включает в себя работу с лекционным материалом, подготовку к лабораторным работам и выполнение лабораторных работ вне аудитории, если это предлагается преподавателем, подготовку отчета, а также изучение нового материала по сети. Изучение нового материала по теме должно обязательно сопровождаться ознакомлением с новейшими достижениями, так как данная сфера относится к быстро развивающимся областям. Поэтому приветствуется включение в отчеты по лабораторным работам а также вопросы во время лекций по новейшим достижениям по изучаемой теме, это может поощряться преподавателем дополнительными баллами.
экзамен	Для текущего контроля успеваемости (по отдельным разделам дисциплины) и промежуточной аттестации (экзамен) рекомендуется использовать компьютерное тестирование, а для итогового контроля - письменный экзамен. Студент допускается к экзамену только после выполнения и защиты отчетов лабораторных работ. При подготовке к экзамену рекомендуется повторить материал лекций. При недостаточном понимании теоретических вопросов или затруднениях при решении задач следует посещать консультации преподавателя.

### 10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

### 11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Мультимедийная аудитория.

Компьютерный класс.

### 12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 09.03.04 "Программная инженерия" и профилю подготовки "Технологии разработки информационных систем".



Приложение 2  
к рабочей программе дисциплины (модуля)  
Б1.В.11 Основы информационного поиска

**Перечень литературы, необходимой для освоения дисциплины (модуля)**

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Технологии разработки информационных систем

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

**Основная литература:**

1. Голицына, О. Л. Информационные системы : учебное пособие / О. Л. Голицына, Н. В. Максимов, И. И. Попов. - 2-е изд. - Москва : ФОРУМ : ИНФРА-М, 2018. - 448 с. : ил. - (Высшее образование). - ISBN 978-5-16-100362-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/953245> (дата обращения: 04.03.2020). - Режим доступа : по подписке.
2. Барский А.Б., Искусственный интеллект и логические нейронные сети : учебное пособие / Барский А.Б. - Санкт-Петербург: ИЦ Интермедия, 2019. - 360 с. - ISBN 978-5-4383-0155-4 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785438301554.html> (дата обращения: 04.03.2020). - Режим доступа : по подписке.
3. Назаров С.В., Основы информационных технологий / Назаров С.В. и др. - Москва: Национальный Открытый Университет 'ИНТУИТ', 2016. - 530 с. - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : [http://www.studentlibrary.ru/book/intuit\\_182.html](http://www.studentlibrary.ru/book/intuit_182.html) (дата обращения: 04.03.2020). - Режим доступа : по подписке.

**Дополнительная литература:**

1. Кудинов, Ю. И. Основы современной информатики : учебное пособие / Ю. И. Кудинов, Ф. Ф. Пашенко. - 5-е изд., стер. - Санкт-Петербург : Лань, 2018. - 256 с. - ISBN 978-5-8114-0918-1. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/107061> (дата обращения: 04.03.2020). - Режим доступа: для авториз. пользователей.
2. Исаев, Г. Н. Моделирование информационных ресурсов: теория и решение задач: учебное пособие / Г.Н. Исаев. - Москва : Альфа-М: ИНФРА-М, 2010. - 224 с. - ISBN 978-5-98281-211-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/193771> (дата обращения: 04.03.2020). - Режим доступа : по подписке.



Приложение 3  
к рабочей программе дисциплины (модуля)  
Б1.В.11 Основы информационного поиска

**Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем**

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Технологии разработки информационных систем

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2020

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "Консультант студента", доступ к которой предоставлен обучающимся. Многопрофильный образовательный ресурс "Консультант студента" является электронной библиотечной системой (ЭБС), предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Полностью соответствует требованиям федеральных государственных образовательных стандартов высшего образования к комплектованию библиотек, в том числе электронных, в части формирования фондов основной и дополнительной литературы.