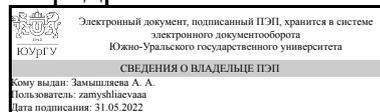


УТВЕРЖДАЮ:  
Заведующий выпускающей  
кафедрой



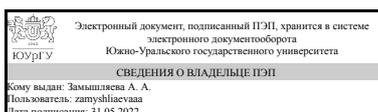
А. А. Замышляева

## РАБОЧАЯ ПРОГРАММА

**дисциплины 1.Ф.М1.02 Информационный поиск, анализ и предобработка данных для направления 09.04.04 Программная инженерия**  
**уровень** Магистратура  
**магистерская программа** Искусственный интеллект и машинное обучение в финтех индустрии  
**форма обучения** очная  
**кафедра-разработчик** Прикладная математика и программирование

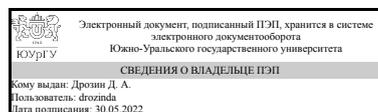
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.04.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 932

Зав.кафедрой разработчика,  
д.физ.-мат.н., проф.



А. А. Замышляева

Разработчик программы,  
к.экон.н., доцент



Д. А. Дрозин

## 1. Цели и задачи дисциплины

Целью дисциплины является изучение математических методов анализа данных, синтеза цифровых алгоритмов их обработки, развития навыков статистического моделирования и исследования различных процессов на ЭВМ. Основными задачами дисциплины являются: - овладение современными методами анализа данных; - закрепление навыков проведения анализа данных на ЭВМ, практическое применение статистических методов обработки информации при решении конкретных задач.

## Краткое содержание дисциплины

В настоящее время информационный поиск в сети интернет стал неотъемлемой частью жизни среднего человека цивилизованного мира. Поисковые машины (Google, Yandex и т.п.) работают в соответствии с заложенными в них алгоритмами. Знание принципов построения алгоритмов работы этих машин является обязательным, для управления контентом в сети интернет. В курсе рассматриваются основные методы анализа данных и методы информационного поиска. Кроме того, создание поисковых машин обусловлено рядом специфических задач. В курсе рассматриваются основные подходы решения этих задач и на протяжении курса формируется навык разработки локальной поисковой машины.

## 2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-5 Способен разрабатывать и применять алгоритмы анализа данных для решения прикладных задач	Знает: основные принципы сбора, хранения и предобработки данных Умеет: выбирать методы и средства для анализа данных, оценивать возможности и ограничения используемых методов, осуществлять дискретизацию непрерывных данных с учётом решаемой задачи Имеет практический опыт: сбора первичной информации, организации и хранения данных для конкретного исследования, применения методов предобработки данных

## 3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Нет	Технологическое предпринимательство, Интеллектуальный анализ текстов, Статистика для анализа данных, Математические модели принятия решений в условиях неопределенности

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Нет

#### 4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 38,25 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		1	
Общая трудоёмкость дисциплины	108	108	
<i>Аудиторные занятия:</i>	32	32	
Лекции (Л)	16	16	
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0	
Лабораторные работы (ЛР)	16	16	
<i>Самостоятельная работа (СРС)</i>	69,75	69,75	
Подготовка к экзамену	9,75	9,75	
Подготовка к лекциям	60	60	
Консультации и промежуточная аттестация	6,25	6,25	
Вид контроля (зачет, диф.зачет, экзамен)	-	зачет	

#### 5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Булев поиск	5	2	0	3
2	Лексикон и списки словопозиций	5	2	0	3
3	Словари и нечеткий поиск	5	2	0	3
4	Построение индекса	5	2	0	3
5	Сжатие индекса	6	2	0	4
6	Ранжирование, взвешивание терминов	2	2	0	0
7	Ранжирование в полнофункциональной поисковой системе	2	2	0	0
8	Оценка информационного поиска	2	2	0	0

##### 5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Основные понятия информационного поиска. Обработка булевых запросов. Сравнение расширенной булевой модели и ранжированного поиска.	2

2	2	Схематизация документа. Определение лексикона терминов. Быстрое пересечение инвертированных списков терминов.	2
3	3	Поисковые структуры для словарей. Запросы с джокером. Исправление опечаток	2
4	4	Основы аппаратного обеспечения. Блочное индексирование. Однопроходное индексирование.	2
5	5	Статистические характеристики терминов. Сжатие словаря. Сжатие инвертированного файла.	2
6	6	Параметрические и зонные индексы. Частота термина и взвешивание. Модель векторного пространства для ранжирования.	2
7	7	Эффективное ранжирование. Компоненты информационно-поисковой системы. Влияние операторов языка запросов на ранжирование в векторном пространстве.	2
8	8	Оценка информационно-поисковой системы. Стандартные тестовые коллекции. Оценка неранжированных результатов поиска.	2

## 5.2. Практические занятия, семинары

Не предусмотрены

## 5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1-2	1	Алгоритм построения инвертированного индекса	3
2-3	2	Алгоритм построения стемминга Ловинса	3
4-5	3	k-граммный индекс, двоичное дерево	3
5-6	4	Построение распределенного индекса	3
7-8	5	Сжатие словаря	4

## 5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Подготовка к экзамену	Симанков, В. С. Методы и алгоритмы поиска информации в Интернете : монография / В. С. Симанков, Д. М. Толкачев. — Москва : Креативная экономика, 2017. — 332 с. Кузьмин, А. В. Поиск в Интернете. Как искать, чтобы найти. Все, от поиска информации, файлов, видео и фотографий до поиска товаров и работы через Интернет : учебное пособие / А. В. Кузьмин, Н. Н. Золотарева ; под редакцией М. В. Финкова. — Санкт-Петербург : Наука и Техника, 2006. — 151 с.	1	9,75
Подготовка к лекциям	Симанков, В. С. Методы и алгоритмы поиска информации в Интернете : монография / В. С. Симанков, Д. М. Толкачев. — Москва : Креативная	1	60

	экономика, 2017. — 332 с. Кузьмин, А. В. Поиск в Интернете. Как искать, чтобы найти. Все, от поиска информации, файлов, видео и фотографий до поиска товаров и работы через Интернет : учебное пособие / А. В. Кузьмин, Н. Н. Золотарева ; под редакцией М. В. Финкова. — Санкт-Петербург : Наука и Техника, 2006. — 151 с.		
--	---	--	--

## 6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

### 6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	1	Текущий контроль	Алгоритм построения инвертированного индекса	20	6	Разработана работоспособная программа без ошибок – 1 балл; Студент может пояснить все строчки кода программы исходя из смысла задачи – 3 балл; В соответствии с требованиями сделан отчет – 2 балла.	зачет
2	1	Текущий контроль	Алгоритм построения стемминга Ловинса	20	6	Разработана работоспособная программа без ошибок – 1 балл; Студент может пояснить все строчки кода программы исходя из смысла задачи – 3 балл; В соответствии с требованиями сделан отчет – 2 балла.	зачет
3	1	Текущий контроль	к-граммный индекс, двоичное дерево	20	6	Разработана работоспособная программа без ошибок – 1 балл; Студент может пояснить все строчки кода программы исходя из смысла задачи – 3 балл; В соответствии с требованиями сделан отчет – 2 балла.	зачет
4	1	Текущий контроль	Построение распределенного индекса	20	6	Разработана работоспособная программа без ошибок – 1 балл; Студент может пояснить все строчки кода программы исходя из смысла задачи – 3 балл; В соответствии с требованиями сделан отчет – 2 балла.	зачет
5	1	Текущий контроль	Сжатие словаря	20	6	Разработана работоспособная программа без ошибок – 1 балл;	зачет

						Студент может пояснить все строчки кода программы исходя из смысла задачи – 3 балл; В соответствии с требованиями сделан отчет – 2 балла.	
6	1	Промежуточная аттестация	Экзамен	-	3	Если вопрос раскрыт полностью - 1 балл. Если вопрос раскрыт, но не полностью - 0.5 балла. Если вопрос не раскрыт - 0 баллов.	зачет

## 6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
зачет	На зачете происходит оценивание учебной деятельности обучающихся по дисциплине на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. Студент может улучшить свой рейтинг, пройдя контрольное мероприятие промежуточной аттестации, которое не является обязательным. Контрольное мероприятие промежуточной аттестации проводится в смешанной форме - письменно-устной. Студенту выдается билет, содержащий 3 вопроса. На подготовку выделяется 1 час, после чего студент сдает работу в письменном виде. Затем проводится собеседование. За каждый вопрос студент может максимум набрать один балл.	В соответствии с пп. 2.5, 2.6 Положения

## 6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ					
		1	2	3	4	5	6
ПК-5	Знает: основные принципы сбора, хранения и предобработки данных	+	+	+	+	+	+
ПК-5	Умеет: выбирать методы и средства для анализа данных, оценивать возможности и ограничения используемых методов, осуществлять дискретизацию непрерывных данных с учётом решаемой задачи	+	+	+	+	+	+
ПК-5	Имеет практический опыт: сбора первичной информации, организации и хранения данных для конкретного исследования, применения методов предобработки данных	+	+	+	+	+	+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

## 7. Учебно-методическое и информационное обеспечение дисциплины

### Печатная учебно-методическая документация

а) *основная литература:*

Не предусмотрена

б) *дополнительная литература:*

Не предусмотрена

в) *отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:*

Не предусмотрены

г) методические указания для студентов по освоению дисциплины:

1. ГОСТ Оформления отчетов

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. ГОСТ Оформления отчетов

### Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	Электронно-библиотечная система издательства Лань	Симанков, В. С. Методы и алгоритмы поиска информации в Интернете : монография / В. С. Симанков, Д. М. Толкачев. — Москва : Креативная экономика, 2017. — 332 с. — ISBN 978-5-9500501-8-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/116049">https://e.lanbook.com/book/116049</a> (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.
2	Дополнительная литература	Электронно-библиотечная система издательства Лань	Кузьмин, А. В. Поиск в Интернете. Как искать, чтобы найти. Все, от поиска информации, файлов, видео и фотографий до поиска товаров и работы через Интернет : учебное пособие / А. В. Кузьмин, Н. Н. Золотарева ; под редакцией М. В. Финкова. — Санкт-Петербург : Наука и Техника, 2006. — 151 с. — ISBN 5-94387-268-X. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/37349">https://e.lanbook.com/book/37349</a> (дата обращения: 15.09.2021). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

1. -Microsoft Visual Studio (бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

### 8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лекции	340 (3б)	Компьютеры
Лабораторные занятия	340 (3б)	Компьютеры
Лабораторные занятия	332 (3б)	Компьютеры