

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Курганский государственный университет»
(КГУ)

Кафедра «Программное обеспечение автоматизированных систем»



УТВЕРЖДАЮ:
Ректор
/ Дубив Н.В./
2019 г.

Рабочая программа учебной дисциплины
**ПРОГРАММНЫЕ СРЕДСТВА ИНТЕЛЛЕКТУАЛЬ-
НОЙ ОБРАБОТКИ ДАННЫХ**

образовательной программы высшего образования –
программы магистратуры

09.04.04 – Программная инженерия

Направленность:

**Методы и алгоритмы интеллектуальной обработки данных в информа-
ционно-вычислительных системах**

Форма обучения: заочная

Курган 2019

Рабочая программа дисциплины «Программные средства интеллектуальной обработки данных» составлена в соответствии с учебными планами по программе бакалавриата «Программная инженерия» (Методы и алгоритмы интеллектуальной обработки данных в информационно-вычислительных системах), утвержденными для заочной формы обучения «29» августа 2019 года.

Рабочая программа дисциплины одобрена на заседании кафедры «Программное обеспечение автоматизированных систем» «31» августа 2019 года, протокол № 1.

Рабочую программу составили:

Доцент кафедры
«Программное обеспечение
автоматизированных систем»

О.С. Черепанов

Согласовано:

Заведующий кафедрой
«Программное обеспечение
автоматизированных систем»
к.т.н., доцент

Т.Р. Змызгова

Специалист по учебно-методической
работе Учебно-методического отдела

Г.В. Казанкова

Начальник Управления образовательной
деятельности

С.Н. Синицын

1. ОБЪЕМ ДИСЦИПЛИНЫ

Всего: 4 зачетные единицы трудоемкости (144 академических часов)

Вид учебной работы	На всю дисциплину	Семестр
		3
Аудиторные занятия (контактная работа с преподавателем), всего часов	14	14
в том числе:		
Лекции	6	6
Лабораторные работы	8	8
Аудиторные занятия в интерактивной форме, часов	-	-
Самостоятельная работа, всего часов	130	130
в том числе:		
Подготовка к зачету	18	18
Другие виды самостоятельной работы	94	94
Контрольная работа	18	18
Вид промежуточной аттестации	зачет	зачет
Общая трудоемкость дисциплины и трудоемкость по семестрам, часов	144	14

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Программные средства интеллектуальной обработки данных» относится к базовой части блока 1 модуль «Методы и алгоритмы анализа данных».

Изучение дисциплины базируется на результатах обучения, сформированных при изучении следующих дисциплин:

- Вычислительная математика.
- Прикладная статистика.
- Методы оптимизации и принятия решений.
- Современные системы и методы высокопроизводительной обработки данных.

Результаты обучения по дисциплине необходимы для изучения дисциплин: «Технологии разработки интеллектуальных систем», «Междисциплинарная курсовая работа», и выполнения выпускной квалификационной работы.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Целью освоения дисциплины «Программные средства интеллектуальной обработки данных» является формирование практических навыков работы с современными программными средствами распределенной обработки неструктурированных и слабоструктурированных данных.

Задачами дисциплины являются:

- изучение способов представления, хранения и обработки слабоструктурированных и неструктурированных данных;
- знакомство с программным средством интеллектуальной обработки данных Apache Spark.

Компетенции, формируемые в результате освоения дисциплины:

- способность самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте (ОПК-1);
- способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности (ОПК-6).

В результате изучения дисциплины обучающийся должен

Знать

- способы представления, хранения и передачи неструктурированных и слабоструктурированных данных (ОПК-6).

- методы поиска и обработки неструктурированных и слабоструктурированных данных (ОПК-6).

Уметь

- реализовывать программное обеспечения по интеллектуальной обработки данных (ОПК-1).

Владеть:

- современными программными средствами, используемыми при разработке интеллектуальных информационных систем (ОПК-1).

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Учебно-тематический план

№	Наименование раздела	Количество часов контактной работы с преподавателем	
		Лекции	Практические работы
1	Представление, хранение, обработка и передача слабоструктурированных и неструктурированных данных	2	0
2	Программные средства интеллектуальной обработки данных	4	8
Всего:		6	8

4.2. Содержание лекционных занятий

Наименование и содержание лекции	Часов контактной работы с преподавателем
Раздел 1. Представление, хранение, обработка и передача слабоструктурированных и неструктурированных данных	
Введение в интеллектуальные информационные системы. Понятие «интеллектуальные информационные системы». История развития интеллектуальных информационных систем. Классификация программных средства интеллектуальной обработки данных (ИОД). Перечень и назначение современных программных средств ИОД.	1
Уровни структурированности данных. Понятие «структура данных». Уровни структурированности данных. Спо-	0,5

<p>собы представления и хранения слабоструктурированных и неструктурированных данных. Форматы представления слабоструктурированных данных. Языки XPath, XQuery, XPointer.</p> <p>Программные средства хранения слабоструктурированных и неструктурированных данных.</p>	
<p>Онтологии предметной области.</p> <p>Определение понятия «онтология». Структура онтологий. Классификация онтологий. Методы построения онтологий. Языки описания онтологий. Семантическая паутина (Semantic Web). Модель Бернерса-Ли. RDF. OWL. KIF. Библиотеки онтологий.</p>	0,5
<p>Раздел 2. Программные средства интеллектуальной обработки данных</p>	
<p>Введение в Spark. Spark Core</p> <p>Архитектура Spark. Компоненты Spark. Модель параллельных вычислений. Наборы RDD. Хранение данных в памяти и управление памятью. Типы наборов RDD. Функции, применяемые к RDD. Планирование и структура заданий Spark.</p>	2
<p>Spark SQL</p> <p>DataFrames. Схемы. API Dataframe. Преобразования DataFrames. SQL-выражения. Специализированные преобразования DataFrames. Наборы DataFrames и Datasets. Фреймворк Parquet. Таблицы Hive. Функции агрегирования. Оптимизаторы запросов. Соединения Spark Core и Spark SQL.</p>	
<p>MLLib</p> <p>Базовая статистика. ML-конвейеры. Преобразователи. Оценщики. Параметры. Извлечение, преобразование и выбор факторов. Классификация и регрессия. Кластерный анализ. Совместная фильтрация. Поиск частотных шаблонов. Примеры программ.</p>	2
<p>Spark Streaming</p> <p>Базовые концепции потоков. Инициализация потоков. Дискретизация потоков. Преобразования потоков данных. Использование Dataframes и SQL операций над потоками. MLLib операции. Интеграция с Apache Kafka.</p>	
<p>Spark GraphX</p> <p>Работа с графами в Spark. Операции над графами. Pregel API. Строители графов. Оптимизированное представление графов. Графовые алгоритмы.</p>	0

Разбор примеров кода.	
Итого:	6

4.3. Лабораторные работы

Номер раздела, темы	Наименование раздела	Наименование практической работы	Норматив времени, час.
1	Программные средства интеллектуальной обработки данных	Работа с Dataframes в Spark. Простая OLAP-аналитика.	4
		Обработка слобоструктурированных данных, используя алгоритмы машинного обучения MLlib	4
Всего:			8

4.4 Контрольная работа

Основная цель выполнения контрольной работы — ознакомление подходами реализации интеллектуальных информационных систем.

Контрольная работа предполагает написание подготовки доклада на заданную тему.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Во время лекций по дисциплине студентам рекомендуется конспектировать теоретический материал, отмечая важные моменты, на которые заострил внимание преподаватель, участвовать в опросах и дискуссиях. Перед лекций необходимо повторить выданный материал, зафиксировать непонятные моменты, чтобы обсудить их на занятии. Конспект лекций представлен в виде мультимедийных презентаций и включен в состав методического комплекса дисциплины.

Практический практикум включает выполнение задания по одному разделу дисциплины «Программные средства интеллектуальной обработки данных». Все работы выполняются в соответствии с заданием, выданным преподавателем.

В процессе выполнения контрольной работы студент реализует программное обеспечение в соответствии с индивидуальным заданием, согласо-

ванным с руководителем. Требования по содержанию, оформлению контрольной работы представлены в соответствующем методическом пособии.

Самостоятельная работа студентов включает в себя проработку материала лекционного курса дисциплины, подготовку и выполнение практических работы и контрольной работы, а также подготовку к зачету.

Для текущего контроля успеваемости обучения используется балльно-рейтинговая система контроля. Для получения высокой оценки настоятельно рекомендуется активно участвовать во время обсуждения материала дисциплины на лекционных и практических занятиях, а также тщательно его прорабатывать при самостоятельной работе.

Рекомендуемый режим самостоятельной работы

Наименование вида самостоятельной работы	Рекомендуемая трудоемкость, акад. час.
Самостоятельное изучение тем дисциплины:	78
Представление, хранение и обработка слабоструктурированных и неструктурированных данных	39
Программные средства интеллектуальной обработки данных	39
Выполнение контрольной работы	18
Подготовка к практическим работам	16
Подготовка к зачету	18
Всего:	130

6 ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

6.1 Перечень оценочных средств

Фонд оценочных средств по дисциплине включает следующие компоненты, включенные в состав учебно-методического комплекса дисциплины:

1. Балльно-рейтинговая система контроля и оценки академической активности обучающихся в КГУ
2. Вопросы для подготовки к зачету по дисциплине.
3. Образцы отчетов по практическим заданиям.
4. Контрольная работа.

6.1 Система балльно-рейтинговой оценки работы обучающихся по дисциплину

Таблица 6.1 – Рейтинговые балльные оценки по дисциплине

Вид учебной работы	Количество баллов
Посещение лекций	46 ч 3 = 126
Выполнение практических работ	4 занятия x 86 = 326
Контрольная работа	26 б
Зачет	30 б

Таблица 6.2 – Соответствие шкал оценивания

Рейтинговая оценка, баллов	Виды оценок промежуточной аттестации		Оценка ECTS
	Традиционная оценка		
91-100	Отлично (5)	Зачтено	A
84-90	Хорошо (4)		B
74-83			C
68-73	Удовлетворительно (3)		D
61-67			E
51-60	Неудовлетворительно (2)	Не зачтено	Fx
0-50			F

Для допуска к зачету обучающийся должен набрать по итогам текущего контроля не менее 50 баллов и при этом он должен выполнить и защитить все практические и контрольную работу.

Для получения оценки «зачтено» автоматически (без сдачи экзамена) достаточно набрать 61 балл по результатам текущего течение семестра.

Обучающемуся преподаватель вправе добавить до 30 дополнительных (бонусных) баллов за активность на учебных занятиях, оригинальность принимаемых решений при выполнении лабораторных работ и индивидуальных контрольных заданий.

В случае если по результатам текущего контроля набрано менее 50 баллов, он может набрать недостающее количество баллов, выполнив дополнительные индивидуальные задания до конца зачетной недели семестра.

Ликвидация академических задолженностей, возникших из-за разности в учебных планах при переводе или восстановлении, также проводится путем выполнения дополнительных индивидуальных заданий.

Состав дополнительных заданий, их количество, формы выполнения и максимальные балльные оценки определяются преподавателем и доводятся до обучающихся в момент выдачи заданий.

6.3 Процедура оценивания результатов освоения дисциплины

Зачет по дисциплине проводится в традиционной форме: обучающийся выполняет задания билета, включающего два теоретических вопроса, и отвечает экзаменатору. Оценивается полнота и правильность ответов на вопросы билета, а также его эрудиция в смежных вопросах.

6.4 Примеры оценочных средств

6.4.1 Примерный перечень вопросов для зачету

1. Понятие «Интеллектуальные информационные системы». История развития интеллектуальных информационных систем.
2. Классификация, перечень и назначение программных средства интеллектуальной обработки данных.
3. Понятие «структура данных». Уровню структурированности данных. Способы представления и хранения данных.
4. Форматы представления слабоструктурированных данных. Языки XPath, XQuery, XPointer.
5. Программные средства хранения слабоструктурированных и неструктурированных данных.
6. Определение понятия «онтология». Структура онтологий. Классификация онтологий. Методы построения онтологий.
7. Семантическая паутина. Модель Бернерса-Ли.
8. Архитектура Spark. Компоненты Spark.
9. Модель параллельных вычислений.
10. Наборы RDD. Хранение данных в памяти и управление памятью. Типы наборов RDD.
11. Функции, применяемые к RDD.
12. Планирование и структура заданий Spark.
13. DataFrames. Схемы. API Dataframe.
14. Преобразования DataFrames. SQL-выражения.

15. Специализированные преобразования DataFrames. Наборы DataFrames и Datasets.
16. Фреймворк Parquet.
17. Таблицы Hive.
18. Функции агрегирования.
19. Оптимизаторы запросов.
20. Соединения Spark Core и Spark SQL.
21. Базовая статистика MLlib.
22. ML-конвейеры. Преобразователи. Оценщики. Параметры.
23. Извлечение, преобразование и выбор факторов в MLlib.
24. Классификация и регрессия в MLlib.
25. Кластерный анализ в MLlib.
26. Совместная фильтрация в MLlib.
27. Поиск частотных шаблонов в MLlib.
28. Базовые концепции потоков. Инициализация потоков.
29. Дискретизация потоков. Преобразования потоков данных.
30. Использование Dataframes и SQL операций над потоками в Spark Streaming.
31. MLlib операции в Spark Streaming.
32. Интеграция Spark Streaming с Kafka.
33. Работа с графами в Spark. Операции над графами.
34. Pregel API. Строители графов в Spark GraphX.
35. Оптимизированное представление графов в Spark GraphX.
36. Графовые алгоритмы в Spark GraphX.

6.4.2 Примеры тем контрольной работы

1. Проектирование системы обработки потоковых данных в режиме реального времени.
2. Использование библиотеки PageRank для работы с графами.
3. Использование Spark для обработки Big Data.

6.5. Фонд оценочных средств

Полный банк заданий для текущего, рубежных контролей и промежуточной аттестации по дисциплине, показатели, критерии, шкалы оценивания компетенций, методические материалы, определяющие процедуры оценивания образовательных результатов, приведены в учебно-методическом комплексе дисциплины.

7. ОСНОВНАЯ УЧЕБНАЯ ЛИТЕРАТУРА

1. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. - СПб.: Питер, 2017. - 336 с.
2. Кац Г и др. W3C XML: XQuery от экспертов. Руководство по языку запросов. - М.: Издательство «КУДИЦ-Образ», 2005. - 480 с.
3. Карау Х., Венделл П., Конвински Э. Изучаем Spark. - М. ДМК-Пресс, 2015. - 304 с.

4. Перрен Ж. Spark в действии. - М. ДМК-Пресс, 2021. - 636 с.

8 МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

8.1 Техническое обеспечение

№	Наименование	Использование
1	Комплект: ноутбук, медиа-проектор, экран	Для демонстрации иллюстративного материала при чтении лекций.
2	Персональный компьютер стандартной комплектации	Используется в качестве инструмента и объекта исследования при выполнении лабораторных и контрольных работ.

8.2 Программное обеспечение

№	Наименование	Использование
1	IDE IntelliJ IDEA или PyCharm	Реализация программного обеспечения в рамках лабораторных работ и курсовой работы
2	Apache Spark	
3	Компилятор Scala	
4	Apache Kafka	

Аннотация
рабочей программы учебной дисциплины

**ПРОГРАММНЫЕ СРЕДСТВА ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ
ДАННЫХ**

образовательной программы высшего образования –
программы магистратуры

09.04.04 Программная инженерия

направленность

Методы и алгоритмы интеллектуальной обработки данных в информационно-вычислительных системах

формы обучения – заочная

Трудоемкость освоения дисциплины – 4 зач. ед. (144 акад. часов)

Семестры: 3-й

Промежуточная аттестация: зачет в 3-м семестре

Содержание дисциплины

Раздел. 1. Представление, хранение, обработка и передача слабоструктурированных и неструктурированных данных

Раздел 2. Программные средства интеллектуальной обработки данных