

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

Кафедра «Программное обеспечение автоматизированных систем»



УТВЕРЖДАЮ:

Первый проректор

Т.Р. Змызгова

« 31 »августа 2022 г.

Рабочая программа учебной дисциплины

ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

образовательной программы высшего образования –
программы бакалавриата

09.03.03 Прикладная информатика

Направленность

Интеллектуальные информационные системы и технологии


Форма обучения: очная, заочная

Рабочая программа дисциплины «Интеллектуальные информационные системы и технологии» составлена в соответствии с учебным планом программы бакалавриата: «Прикладная информатика», (Интеллектуальные информационные системы и технологии), утвержденным

- для очной формы обучения 30 августа 2022 г.
- для заочной формы обучения 30 августа 2022 г.

Рабочая программа дисциплины одобрена на заседании кафедры «Программное обеспечение автоматизированных систем» 30.08.2022 года, протокол № 1.

Рабочую программу составил:
доцент кафедры ПОАС


_____/О.С. Черепанов/

Согласовано:

Заведующий кафедрой
«Программное обеспечение
автоматизированных систем»


_____/В. К. Волк/

Специалист
по учебно-методической работе
Учебно-методического отдела


_____/Г. В Казанкова/

Начальник
Управления образовательной
деятельности


_____/И. В. Григоренко/

1. ОБЪЕМ ДИСЦИПЛИНЫ

Всего: 3 зачетные единицы трудоемкости (108 академических часов)

Очная форма обучения

Вид учебной работы	Семестр	
	Очная форма	Заочная форма
	7	7
Аудиторные занятия (контактная работа с преподавателем), всего часов в том числе:	32	12
Лекции	16	6
Лабораторные работы	16	6
Аудиторные занятия в интерактивной форме, часов	-	
Самостоятельная работа, всего часов в том числе:	76	96
Подготовка к экзамену	27	27
Другие виды самостоятельной работы	13	33
Выполнение курсовой работы	36	36
Вид промежуточной аттестации	экзамен	экзамен
Общая трудоемкость дисциплины и трудоемкость по семестрам, часов	108	108

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Интеллектуальные информационные системы и технологии» относится к вариативной части блока 1 модуль «Методы интеллектуальной обработки данных».

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

- Дискретная математика.
- Теория вероятностей и математическая статистика.
- Основы программирования.
- Базы данных.
- Интеллектуальный анализ данных.
- Технологии параллельного программирования.
- Объектно-ориентированное программирование.

Результаты обучения по дисциплине необходимы для изучения дисциплин: «Методы распознавания и идентификации объектов», «Теория систем и системный анализ», и выполнения выпускной квалификационной работы.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Целью освоения дисциплины «Интеллектуальные информационные системы и технологии» 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 осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (УК-1);
- способность определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений (УК-2);
- способность применять системный подход, современные математические методы и технологии для формализации решения прикладных задач (ПК-1);
- способность организовывать и проводить работы по исследованию объектов профессиональной деятельности, выявлять информационные по-

требности пользователей, формировать требования к информационной системе на основе анализа бизнес-процессов предметной области (ПК-3);

- способность моделировать прикладные (бизнес) процессы и предметную область, использовать методы и инструментальные средства разработки программных проектов на стадиях технического задания, технологии концептуального, функционального и логического проектирования (ПК-4);

- способность проектировать информационные системы по видам обеспечения (ПК-6);

- способность осуществлять презентацию информационной системы и начальное обучение пользователей (ПК-14).

В результате изучения дисциплины обучающийся должен

Знать

- способы представления и хранения неструктурированных и слабоструктурированных данных (УК-1, ПК-1, ПК-3).

- методы поиска и обработки неструктурированных и слабоструктурированных данных (ПК-1, УК-2).

Уметь

- реализовывать программное обеспечение по интеллектуальной обработке данных (ПК-4, ПК-6, ПК-14).

Владеть:

- современными программными средствами, используемыми при разработке интеллектуальных информационных систем (ПК-4, ПК-6, ПК-11).

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Учебно-тематический план

Очная форма обучения

№	Наименование раздела	Количество часов контактной работы с преподавателем	
		Лекции	Лабораторные работы
1	Представление, хранение, обработка и передача слабоструктурированных и неструктурированных данных	6	0
	Рубежный контроль 1		2
2	Программные средства интеллектуальной обработки данных	10	12
	Рубежный контроль 2		2
Всего:		16	16

Заочная форма обучения

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

4.2. Содержание лекционных занятий

Наименование и содержание лекции	Часов контактной работы с преподавателем	
	Очная форма	Заочная форма
Раздел 1. Представление, хранение, обработка и передача слабоструктурированных и неструктурированных данных		
<p>Лекция 1. Введение в интеллектуальные информационные системы.</p> <p>Понятие «интеллектуальные информационные системы». История развития интеллектуальных информационных систем. Классификация программных средства интеллектуальной обработки данных (ИОД). Перечень и назначение современных программных средств ИОД.</p>	2	2
<p>Лекция 2. Уровни структурированности данных.</p> <p>Понятие «структура данных». Уровни структурированности данных. Способы 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>	2	0
<p>Лекция 3. 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>	2	0
<p>Раздел 2. Программные средства интеллектуальной обработки данных</p>		
<p>Лекция 4. 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	2
<p>Лекция 5. 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>	2	
<p>Лекция 6. 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	2
<p>Лекция 7. 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>	2	0
<p>Лекция 8. Spark GraphX</p> <p>Работа с графами в Spark. Операции над графами. Pregel API. Строители графов. Оптимизированное представление графов. Графовые алгоритмы. Разбор примеров кода.</p>	2	0
<p>Итого:</p>	16	6

4.3. Лабораторные работы

Номер раздела, темы	Наименование раздела	Наименование практической работы	Норматив времени, час.	
			очная форма	заочная форма
1	Представление, хранение, обработка и передача слабоструктурированных и неструктурированных данных	Рубежный контроль №1	2	-
2	Программные средства интеллектуальной обработки данных	Работа с Dataframes в Spark. Простая OLAP-аналитика.	2	3
		Обработка слабоструктурированных данных, используя алгоритмы машинного обучения MLlib	4	3
		Построение графа знаний, используя алгоритмы обработки естественного языка и онтологии	6	0
		Рубежный контроль № 2	2	-
Всего:			16	6

4.4 Курсовая работа

Основная цель курсовой работы — формирование практических навыков разработки программного обеспечения с использованием Apache Spark.

Объектом разработки является компьютерная программа. Назначение, область применения, функциональные требования формулируются студентом самостоятельно и согласуются с руководителем в рамках составления технического задания на курсовую работу.

К защите курсовой работы представляется программный код, набор исполняемых файлов, достаточный для демонстрации работы программы, а также комплект проектной, программной и эксплуатационной документации.

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.

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

В процессе выполнения курсовой работы студент реализует программное обеспечение в соответствии с индивидуальным заданием, согласованным с руководителем. Требования по содержанию, оформлению курсовой работы представлены в соответствующем методическом пособии.

Самостоятельная работа студентов включает в себя проработку материала лекционного курса дисциплины, подготовку и выполнение лабораторных работ и курсовой работы, а также подготовку к рубежным контролям (для очной формы обучения) и к экзамену.

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

Рекомендуемый режим самостоятельной работы

Наименование вида самостоятельной работы	Рекомендуемая трудоемкость, акад. час.	
	Очная форма	Заочная форма
Самостоятельное изучение тем дисциплины:		
Представление, хранение и обработка слабоструктурированных и неструктурированных данных	3	30
Программные средства интеллектуальной обработки данных	1	15
Программные средства интеллектуальной обработки данных	2	15
Подготовка к лабораторным работам (по 1 ч. на каждое занятие)	8	3
Подготовка к рубежным контролям (по 1 часу на каждый рубеж)		-
Подготовка к курсовой работе	36	36
Подготовка к экзамену	27	27

	Всего:	76	96
--	---------------	-----------	-----------

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

6.1. Перечень оценочных средств

1. Балльно-рейтинговая система контроля и оценки академической активности студентов в КГУ (для очной формы обучения).
2. Отчеты студентов по лабораторным занятиям.
3. Банк заданий к рубежным контролям № 1, № 2 (для очной формы обучения).
4. Курсовая работа.
5. Вопросы к экзамену.

6.2. Система балльно-рейтинговой оценки работы студентов по дисциплине

№	Наименование	Содержание					
		Распределение баллов за 7 семестр					
1	Распределение баллов за семестры по видам учебной работы, сроки сдачи учебной работы (доводится до сведения студентов на первом учебном занятии)	Вид учебной работы:	Посещение лекций	Выполнение и защита результатов лабораторных работ	Рубежный контроль №1	Рубежный контроль №2	Экзамен
		Балльная оценка:	16 x 8 = 86	156 + 156 + 156 = 456	86	96	306
2	Критерий пересчета баллов в традиционную оценку по итогам работы в семестре и экзамена			60 и менее баллов – неудовлетворительно; 61...73 – удовлетворительно; 74...90 – хорошо; 91...100 – отлично.			
3	Критерии допуска к промежуточной аттестации, возможности получения, возможность получения бонусных баллов			Для допуска к промежуточной аттестации (экзамену) студент должен набрать не менее 50 баллов и выполнить все лабораторные работы и курсовую работу. Для получения «автоматически» оценки «удовлетворительно» студенту необходимо набрать 68 баллов. По согласованию с преподавателем студенту, набравшему 68 баллов, студенту могут быть добавлены дополнительные (бонусные) баллы за активность на консультациях, активное участие в научной и методической работе, оригинальность принятых решений в ходе выполнения лабораторных работ, за участие в значимых учебных и внеучебных мероприятиях кафедры, и выставлена автоматическая оценка «хорошо» или «отлично».			

4	<p>Формы и виды учебной работы для неуспевающих (восстановившихся на курсе обучения) студентов для получения недостающих баллов в конце семестра</p> <p>В случае если к промежуточной аттестации (экзамену) набрана сумма менее 50 баллов, студенту необходимо набрать недостающее количество баллов за счет выполнения дополнительных заданий, до конца последней (зачетной) недели семестра. При этом необходимо проработать материал всех пропущенных лабораторных работ.</p> <p>Формы дополнительных заданий (назначаются преподавателем):</p> <ul style="list-style-type: none"> - выполнение и защита пропущенной лабораторной работы (при невозможности дополнительного проведения лабораторной работы преподаватель устанавливает форму дополнительного задания по тематике пропущенной лабораторной работы самостоятельно) – до 8 баллов. <p>Ликвидация академических задолженностей, возникших из-за разности в учебных планах при переводе или восстановлении, проводится путем выполнения дополнительных заданий, форма и объем которых определяется преподавателем.</p>																				
5	<p>Распределение баллов за семестры по видам учебной работы, сроки сдачи учебной работы (доводятся до сведения студентов на первом учебном занятии)</p> <table border="1" data-bbox="446 1041 917 2083"> <tr> <td colspan="5" data-bbox="877 1041 917 2083" style="text-align: center;">Курсовая работа</td> </tr> <tr> <td data-bbox="662 1041 877 1366">Качество представленной документации</td> <td data-bbox="662 1366 877 1691">20 б</td> <td data-bbox="662 1691 877 1915">Качество выполненной работы</td> <td data-bbox="662 1915 877 2083">50 б</td> <td data-bbox="662 2083 877 2083"></td> </tr> <tr> <td data-bbox="446 1041 662 1366"></td> <td data-bbox="446 1366 662 1691">20 б</td> <td data-bbox="446 1691 662 1915">Качество защиты</td> <td data-bbox="446 1915 662 2083">30 б</td> <td data-bbox="446 2083 662 2083"></td> </tr> <tr> <td data-bbox="446 2083 662 2083"></td> <td data-bbox="446 2083 662 2083"></td> <td data-bbox="446 2083 662 2083" style="text-align: center;">Всего</td> <td data-bbox="446 2083 662 2083"></td> <td data-bbox="446 2083 662 2083" style="text-align: center;">100 б</td> </tr> </table>	Курсовая работа					Качество представленной документации	20 б	Качество выполненной работы	50 б			20 б	Качество защиты	30 б				Всего		100 б
Курсовая работа																					
Качество представленной документации	20 б	Качество выполненной работы	50 б																		
	20 б	Качество защиты	30 б																		
		Всего		100 б																	

6.3. Процедура оценивания результатов освоения дисциплины

Рубежные контроли проводятся в форме письменного тестирования, экзамен в виде ответа на вопросы.

Перед проведением рубежного контроля преподаватель прорабатывает со студентами основную материал соответствующих разделов дисциплины в форме краткой лекции-дискуссии.

Варианты заданий для рубежных контролей № 1, № 2 состоят из 10 вопросов.

На каждую подготовку к рубежному контролю студенту отводится 1 академический час.

Преподаватель оценивает в баллах результаты рубежных контролей каждого студента по количеству правильных ответов и заносит в ведомость учета текущей успеваемости.

На экзамене студенту предлагается ответить на 2 вопроса. Вопросы к экзамену доводятся до студентов на последней лекции в семестре. Каждый вопрос оценивается в 15 баллов. На подготовку ответа студенту отводится 1 астрономический час.

Результаты текущего контроля успеваемости и экзамена заносятся преподавателем в экзаменационную ведомость, которая сдается в организационный отдел института в день проведения экзамена, а также выставляются в зачетную книжку студента.

6.4. Примеры оценочных средств для рубежных контролей и экзамена

6.4.1 Примеры заданий для рубежного контроля №1

1. Какие форматы не подходят для хранения и передачи слабоструктурированных данных?

- а) XML;
- б) JSON;
- в) EDU;
- г) JPEG;
- д) RDF.

2. Какие утверждения относятся к XPath?

- а) язык, который используется для выборки узлов из XML-документов;
- б) функциональный язык программирования и язык запросов к XML-документам;
- в) представлен в виде модели дерева с семью узлами, а именно инструкциями обработки, элементами, узлами документа, атрибутами, пространствами имен, текстовыми узлами и комментариями;
- г) помогает создавать синтаксис XML-документов.

3. Что включается в общую структуру онтологии:

- а) понятия;
- б) отношения;
- в) аксиомы;
- г) теоремы.

6.4.2 Примеры заданий для рубежного контроля №2

1. Какая структура данных в Apache Spark обрабатывается медленнее?
 - а) Dataset;
 - б) RDD;
 - в) DataFrame;
 - г) все одинаково.
2. Spark-приложения выполняют вычисления MapReduce...
 - а) в оперативной памяти;
 - б) в кэше;
 - в) с использованием внешнего хранилища состояний;
 - г) на жестком диске.
3. Какой подход Spark Streaming реализует к обработке данных?
 - а) микро-пакетный;
 - б) распределенный;
 - в) потоковый;
 - г) процедурный.

6.4.3 Примерный перечень вопросов для экзамену

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.

6.5. 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

ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ
образовательной программы высшего образования –
программы бакалавриата

09.03.03 Прикладная информатика

направленность

Интеллектуальные информационные системы и технологии

формы обучения – очная

Трудоемкость освоения дисциплины – 3 зач. ед. (108 акад. часов)

Семестры: 7-й (очная, заочная форма обучения)

Промежуточная аттестация: экзамен в 7-м семестре

Содержание дисциплины

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

Раздел 2. Программные средства интеллектуальной обработки данных