МИНОБРНАУКИ РОССИИ
АСТРАХАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. В.Н. ТАТИЩЕВА

СОГЛАСОВАНО Руководитель ОПОП	УТВЕРЖДАЮ Заведующий кафедрой ПМИ
Гордеев И.И. Гордеев	М.В. Коломина
29 июня 2022 г.	29 июня 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
СТРУКТУРИРОВАНИЕ, РАЗМЕТКА И ОБОГАЩЕНИЕ ДАННЫХ

Составитель(-и)	**Муромцев Д.И., к.т.н., доцент, ИТМО** **Выборнова О.Н., к.т.н., доцент каф. ИБ, АГУ** **Гордеев И.И., к.ф.м.н., доцент каф. ПМИ, АГУ**
Направление подготовки / специальность	**09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ**
Направленность (профиль) ОПОП	**ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**
Квалификация (степень)	**магистр**
Форма обучения	**очная**
Год приема	**2022**
Курс	**2**

Астрахань – 2022

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целью освоения дисциплины «Структурирование, разметка и обогащение данных» является формирование у студентов компетенций в области анализа изображений и видео, а также анализа естественного языка с помощью методов искусственного интеллекта.

1.2. Задачи освоения дисциплины (модуля):

- изучение подходов и приобретение практических навыков в выборе и применении методов структурирования знаний для предметных областей;
- изучение подходов и приобретение практических навыков в выборе и применении методов представления знаний с помощью логических и продукционных методов, семантических сетей и фреймов, объектно-ориентированных методов;
- приобретение практических навыков выбора и применения методов обработки и распространения знаний для разработки программных компонентов систем, основанных на знаниях, и приложений.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина Б1.В.Д.02.01 «Структурирование, разметка и обогащение данных» относится к части, формируемой участниками образовательных отношений учебного плана направления подготовки 09.04.02 Информационные системы и технологии, 2022 года набора.

2.2. Для изучения данной учебной дисциплины (модуля) необходимы следующие знания, умения и навыки, формируемые предшествующими дисциплинами:

- Обработка и анализ данных.
- Методы машинного обучения.
- Прикладной искусственный интеллект.

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной:

- Интеллектуальный анализ данных.

Также дисциплина «Структурирование, разметка и обогащение данных» поможет студентам при реализации задач преддипломной практики и написанию магистерской диссертации.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

Процесс изучения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки:

а) профессиональных (ПК):

ПК-3 – Способен создавать и применять методы распределенного искусственного интеллекта для создания интеллектуальных сред и семантического веба.

ПК-5 – Способен выбирать и применять методы инженерии знаний для создания систем, основанных на знаниях.

Таблица 1. Декомпозиция результатов обучения

Код компетенции	Планируемые результаты освоения дисциплины (модуля)		
	Знать (1)	Уметь (2)	Владеть (3)
ПК-3 ПК-3.2. Применяет методы распределенного искусственного интеллекта для построения семантического веба.	ПК-3.2.1 методы построения онтологических систем, онтологические языки, логические исчисления для их описания.	ПК-3.2.2 применять и разрабатывать технологии онтологического поиска, вывода на онтологиях и онтологической разметки для создания систем интернета, интранета и систем онтологического поиска и распределенного вывода на семантическом Вебе.	ПК-3.2.3 навыками использования инструментов для построения семантического веба.

ПК-5 ПК-5.1. Выбирает и применяет методы сбора и извлечения знаний.	ПК-5.1.1 методологические подходы к выбору и разработке методов получения знаний инженером по знаниям от экспертов; извлечения знаний из данных и текстов и применения соответствующих инструментальных средств.	ПК-5.1.2 методологические подходы к выбору и разработке методов получения знаний инженером по знаниям от экспертов; извлечения знаний из данных и текстов	ПК-5.1.3 навыками использования инструментов извлечения знаний из данных и текстов.
ПК-5 ПК-5.2. Выбирает и применяет методы структурирования знаний.	ПК-5.2.1 методологические подходы к выбору и применению методов структурирования знаний для предметных областей в виде ментальных карт, таксономий, деревьев целей и решений.	ПК-5.2.2 выбирать и применять методы структурирования знаний для построения концептуальных моделей знаний (онтологий знаний).	ПК-5.2.3 навыками использования инструментов структурирования знаний в виде ментальных карт, таксономий, деревьев целей и решений.
ПК-5 ПК-5.3. Выбирает и применяет методы представления знаний.	ПК-5.3.1 методологические подходы к выбору и применению методов представления знаний с помощью логических и продукционных методов, семантических сетей и фреймов, объектно-ориентированных методов.	ПК-5.3.2 выбирать и применять методы представления знаний для проектирования базы знаний для предметных областей.	ПК-5.3.3 навыками построения баз знаний на основе формальных языков представления знаний.
ПК-5 ПК-5.4. Выбирает и применяет методы обработки и распространения знаний.	ПК-5.4.1 методологические подходы к выбору и применению методов обработки и распространения знаний с помощью дедукции, индукции и абдукции, согласования экспертных оценок и нечеткого вывода.	ПК-5.4.2 выбирать и применять методы обработки и распространения знаний для разработки программных компонентов систем, основанных на знаниях, и приложений.	ПК-5.4.3 навыками разработки поисковых запросов к базам знаний, использование средств структурированного поиска.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) в зачетных единицах **4 з. е.** Всего 144 часа: 24 часа выделено на контактную работу обучающихся с преподавателем (лекции – 12, лабораторные работы – 12), 120 часов – на самостоятельную работу обучающихся.

Таблица 2.

Структура и содержание дисциплины (модуля)

№ п/п	Наименование радела (темы)	Семестр	Неделя семестра	Контактная работа (в часах)			Самостоят. работа		Формы текущего контроля успеваемости *(по неделям семестра)* Форма промежуточной аттестации *(по семестрам)*
				Л	ПЗ	ЛР	КР	СР	
1	Качество данных, подходы и инструменты	4		2		2		24	Лабораторная работа 1
2	Структурирование данных для систем ИИ			2		2		24	Лабораторная работа 2-4
3	Разметка наборов данных			4		4		36	Лабораторная работа 5-7
4	Обогащение данных			4		4		36	Лабораторные работы 8-10
ИТОГО				**12**		**12**		**120**	**ЭКЗАМЕН**

Условные обозначения:

Л – занятия лекционного типа; ПЗ – практические занятия, ЛР – лабораторные работы; КР – курсовая работа; СР – самостоятельная работа по отдельным темам.

Таблица 3.
Матрица соотнесения тем/разделов
учебной дисциплины/модуля и формируемых в них компетенций

Темы, Разделы дисциплины	Кол-во часов	Компетенции		
		ПК 3	ПК 5	общее количество компетенций
Качество данных, подходы и инструменты	28	+	+	2
Структурирование данных для систем ИИ	28	+	+	2
Разметка наборов данных	44	+	+	2
Обогащение данных	44	+	+	2
Итого	**144**			

Краткое содержание дисциплины

Тема 1. Качество данных, подходы и инструменты

Основные понятия качества данных. Мастер-данные. Инструменты управления, качеством данных, интеграцией и очисткой данных, управление метаданнымии.

Тема 2. Структурирование данных для систем ИИ

Основные подходы к структурированию данных. Модели данных в системах ИИ (таблицы, иерархические структуры, графы). Системы извлечения фактов. Базы знаний. Методы автоматического пополнения баз знаний. Классификация и кластеризация данных. Сегментация данных. «Зашумленные» данные на примере текстовых постов (чаты, социальные сети). Модели семантической структуры данных и онтологии. Онтологические модели и языки. Методы онтологического поиска и распределенного вывода. Семантический Веб. Нейронные сети в задачах парсинга данных. Вероятностный парсинг. Подготовка выборочных данных.

Тема 3. Разметка наборов данных

Основные понятия и методы разметки цифровых изображений. Распознавание объектов на изображениях. Обнаружение движения на изображениях. Разметка текстовых данных. Применение скрытых Марковских моделей для частеречной разметки. Порождение морфологических гипотез (обработка несловарных словоформ). Синтаксическая разметка. Разметка речевых аудиоданных для диалоговых систем.

Тема 4. Обогащение данных

Понятие и задача обогащения данных. Проблемы малых выборок и низкочастотных данных. Аугментация наборов данных. Методы обогащения текстовых данных. Методы обогащения наборов изображений. Синтез речевых данных. Обогащение наборов временных рядов.

5. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

5.1. Указания по организации и проведению лекционных, практических (семинарских) и лабораторных занятий с перечнем учебно-методического обеспечения.

Лекционные занятия

Основной формой реализации теоретического обучения является лекция, которая представляет собой систематическое, последовательное изложение преподавателем-лектором учебного материала теоретического характера. Цель лекции – организация целенаправленной познавательной деятельности студентов по овладению программным материалом учебной дисциплины.

Порядок подготовки лекционного занятия включает в себя выполнение следующих этапов:

- изучение требований программы дисциплины;
- определение целей и задач лекции;
- разработка плана проведения лекции;
- подбор литературы (ознакомление с методической литературой, публикациями периодической печати по теме лекционного занятия);

- отбор необходимого и достаточного по содержанию учебного материала;
- определение методов, приемов и средств поддержания интереса, внимания, стимулирования творческого мышления студентов;
- написание конспекта лекции.

Лекция должна включать следующие разделы:

- формулировку темы лекции;
- указание основных изучаемых разделов или вопросов и предполагаемых затрат времени на их изложение;
- изложение вводной части;
- изложение основной части лекции;
- краткие выводы по каждому из вопросов;
- заключение;
- рекомендации литературных источников по излагаемым вопросам.

Лабораторные занятия

Лабораторное занятие − целенаправленная форма организации педагогического процесса, направленная на углубление научно-теоретических знаний и овладение определенными методами работы, в процессе которых вырабатываются умения и навыки выполнения тех или иных учебных действий в данной сфере науки. Они развивают научное мышление и речь, позволяют проверить знания студентов и выступают как средства оперативной обратной связи.

Правильно организованные лабораторные занятия ориентированы на решение следующих задач:

- обобщение, систематизация, углубление, закрепление полученных на лекциях и в процессе самостоятельной работы теоретических знаний по дисциплине (предмету);
- формирование практических умений и навыков, необходимых в будущей профессиональной деятельности, реализация единства интеллектуальной и практической деятельности;
- выработка при решении поставленных задач таких профессионально значимых качеств, как самостоятельность, ответственность, точность, творческая инициатива.

Состав заданий для лабораторного занятия должен быть спланирован с расчетом, чтобы за отведенное время они могли быть качественно выполнены большинством учащихся.

Лабораторные занятия должны так быть организованы, чтобы студенты ощущали нарастание сложности выполнения заданий, испытывали бы положительные эмоции от переживания собственного успеха в учении, поисками правильных и точных решений.

Самостоятельная работа

Самостоятельная работа – это вид учебной деятельности, которую студент совершает в установленное время и в установленном объеме индивидуально или в группе, без непосредственной помощи преподавателя (но при его контроле), руководствуясь сформированными ранее представлениями о порядке и правильности выполнения действий.

В учебном процессе образовательного учреждения выделяются два вида самостоятельной работы:

- аудиторная – выполняется на учебных занятиях, под непосредственным руководством преподавателя и по его заданию (выполнение самостоятельных работ; выполнение контрольных и практических работ; решение задач);
- внеаудиторная – выполняется по заданию преподавателя, но без его непосредственного участия (подготовка к аудиторным занятиям; изучение учебного материала, вынесенного на самостоятельную проработку; выполнение домашних заданий разнообразного характера; выполнение индивидуальных заданий, направленных на развитие у студентов самостоятельности и инициативы; подготовка к контрольной работе). Внеаудиторные самостоятельные работы представляют собой логическое продолжение аудиторных занятий, проводятся по заданию преподавателя, который инструктирует студентов и устанавливает сроки выполнения задания.

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Лекция

Лекция – основной вид обучения в вузе.

В лекции излагаются основные положения теории, ее понятия и законы, приводятся факты, показывающие связь теории с практикой.

Накануне лекции необходимо повторить содержание предыдущей лекции (а также теорию по изучаемой теме в школьных учебниках геометрии, если эта тема была представлена в них), а затем посмотреть тему очередной лекции по программе (по плану лекций).

Полезно вести записи (конспекты) лекций: для непонятных вопросов оставлять место при работе над темой лекции с учебными пособиями.

Записи лекций следует вести в отдельной тетради, оставляя место для дополнений во время самостоятельной работы.

При конспектировании лекций выделяйте главы и разделы, параграфы, подчеркивайте основное.

Лабораторное занятие

Лабораторное занятие – наиболее активный вид учебных занятий в вузе. Он предполагает самостоятельную работу над лекциями и учебными пособиями.

К каждому лабораторному занятию нужно готовиться. Подготовку следует начинать с повторения теории (по записям лекций или по учебному пособию). После этого нужно решать задачи из предложенного домашнего задания.

Организация самостоятельной работы

Самостоятельность в учебной работе способствует развитию заинтересованности студента в изучаемом материале, вырабатывает у него умение и потребность самостоятельно получать знания, что весьма важно для специалиста с высшим образованием. Самостоятельная работа студентов представлена в следующих формах:

- работа с учебной литературой и конспектом лекций с целью подготовки к лабораторным занятиям, составление конспектов тем, выносимых на самостоятельную проработку;
- систематическое выполнение домашних работ.

Таблица 4.

Содержание самостоятельной работы обучающихся

Номер радела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Формы работы
1	Качество данных, подходы и инструменты	24	Изучение теоретического материала. Подготовка к лабораторным работам.
2	Структурирование данных для систем ИИ	24	Изучение теоретического материала. Подготовка к лабораторным работам.
3	Разметка наборов данных	36	Изучение теоретического материала. Подготовка к лабораторным работам.
4	Обогащение данных	36	Изучение теоретического материала. Подготовка к лабораторным работам.

5.3. Виды и формы письменных работ, предусмотренных при освоении дисциплины, выполняемые обучающимися самостоятельно.

Отчет по лабораторной работе – оформляется и отчитывается в электронном виде: формат листа А4, книжная ориентация страницы. Отчеты по всем лабораторным работам имеют единый титульный лист, на котором указывается наименование дисциплины, ФИО и группа исполнителя, ФИО преподавателя, принимающего отчеты. В отчете по каждой лабораторной работе должно быть представлено наименование работы, цель, ход выполнения работы (скриншоты, краткое текстовое описание), выводы по результатам работы.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

При реализации различных видов учебной работы по дисциплине могут использоваться электронное обучение и дистанционные образовательные технологии.

6.1. Образовательные технологии

Учебные занятия по дисциплине могут проводиться с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line или off-line в формах.

№	Формы	Описание
1	Лекция-дискуссия	Лекция-дискуссия специально не назначается, а возникает достаточно спонтанно на большинстве лекций. Студенты устно высказываю своё мнение по ходу лекции, дискутируют как с лектором, так и между собой. Также дискуссии иногда возникают при защите лабораторных работ.
2	Исследовательские методы в обучении	Дает возможность учащимся самостоятельно пополнять свои знания, глубоко вникать в изучаемую проблему и предполагать пути ее решения, что важно при формировании мировоззрения.
3	Лабораторные работы	Формирование навыков использования современных компьютерных технологий по отладке программ.
4	Самостоятельная работа	Работа с ресурсами Internet, подготовка к лабораторным работам.

6.2. Информационные технологии

При реализации различных видов учебной и внеучебной работы используются следующие информационные технологии:

- система управления обучением LMS Moodle;
- использование возможностей Интернета в учебном процессе (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление обучающихся с оценками и т.д.);
- использование электронных учебников и различных сайтов (например, электронные библиотеки, журналы и т.д.) как источник информации;
- использование возможностей электронной почты;
- использование средств представления учебной информации (электронных учебных пособий, применение новых технологий для проведения занятий с использованием презентаций и т.д.);
- использование интерактивных средств взаимодействия участников образовательного процесса (технологии дистанционного или открытого обучения в глобальной сети);
- использование интегрированных образовательных сред, где главной составляющей являются не только применяемые технологии, но и содержательная часть, т.е. информационные ресурсы (доступ к мировым информационным ресурсам, на базе которых строится учебный процесс).

6.3. Перечень программного обеспечения и информационных справочных систем

а) Перечень лицензионного учебного программного обеспечения

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда
Google Chrome	Браузер
Microsoft Office 2013, Microsoft Office Project 2013 , Microsoft Office Visio 2013	Офисная программа
7-zip	Архиватор
Microsoft Windows 7 Professional	Операционная система
Kaspersky Endpoint Security	Средство антивирусной защиты
Notepad++	Текстовый редактор
PyCharm EDU	Среда разработки
R	Программная среда вычислений
Scilab	Пакет прикладных математических программ
Sofa Stats	Программное обеспечение для статистики, анализа и отчетности
VirtualBox	Программный продукт виртуализации операционных систем
VMware (Player)	Программный продукт виртуализации операционных систем
Microsoft Visual Studio	Среда разработки

Oracle SQL Developer	Среда разработки
IBM SPSS Statistics 21	Программа для статистической обработки данных

б) Современные профессиональные базы данных и информационные справочные системы:

1. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем»: https://library.asu.edu.ru.
2. Электронный каталог «Научные журналы АГУ»: http://journal.asu.edu.ru/.
3. Универсальная справочно-информационная полнотекстовая база данных периодических изданий ООО «ИВИС»: http://dlib.eastview.com/
4. Электронно-библиотечная система elibrary. http://elibrary.ru
5. Справочная правовая система КонсультантПлюс: http://www.consultant.ru
6. Информационно-правовое обеспечение «Система ГАРАНТ»: http://garant-astrakhan.ru

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине «Структурирование, разметка и обогащение данных» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 5

Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

№ п/п	Контролируемые разделы, темы дисциплины (модуля)	Код контролируемой компетенции (компетенций)	Наименование оценочного средства
1.	Качество данных, подходы и инструменты	ПК-3, ПК-5	Лабораторная работа 1
2.	Структурирование данных для систем ИИ	ПК-3, ПК-5	Лабораторная работа 2-4
3.	Разметка наборов данных	ПК-3, ПК-5	Лабораторная работа 5-7
4.	Обогащение данных	ПК-3, ПК-5	Лабораторные работы 8-10

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Для оценки результатов обучения применяются следующие критерии.

Таблица 6

Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

Таблица 7

Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4 «хорошо»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, не способен применить знание теоретического материала при выполнении заданий, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание при подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задание

7.3. Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

Тема 1. Качество данных, подходы и инструменты

Лабораторная работа № 1. Анализ качества и очистка данных в системе Open Refine.

Тема 2. Структурирование данных для систем ИИ

Лабораторные работы № 2. Преобразование и сегментация изображений в системе OpenCV.
Лабораторные работы № 3. Реализация алгоритма определение движения на видео на платформе Raspberry Pi
Лабораторные работы № 4. Ускорение работы алгоритмов обработки изображений с помощью платформы NVIDIA Jetson

Тема 3. Разметка наборов данных

Лабораторные работы №5. Создание аннотатора, реализующего алгоритмы токенизации и сегментации на предложения в библиотеке Stanford CoreNLP
Лабораторные работы № 6. Морфологический и синтаксический анализ
Лабораторные работы № 7. Векторные представления для моделирования семантики

Тема 4. Обогащение данных

Лабораторные работы №8. Кластеризация текстовых коллекций
Лабораторные работы № 9. Обработка и индексирование текстовых коллекций
Лабораторные работы № 10. Увеличение выборок данных для машинного обучения.

Пример лабораторной работы

Лабораторная работа № 5 «Разметка данных с помощью морфологического анализа»

а) Применить морфоанализатор для русского языка, используя модель (https://drive.google.com/drive/folders/0B4TmAgcGLMriMG96cFZSSWhWcEU?usp=sharing), текст можно использовать произвольный (но не менее 20 предложений).

б) Проверить автоматическую частеречную разметку и лемматизацию в .conll файле (https://drive.google.com/open?id=0B4TmAgcGLMriV0hCc2VNbHNadzQ) на корректность тэгов и лексем. (описание conll формата http://universaldependencies.org/format.html)

Можно пользоваться НКРЯ http://www.ruscorpora.ru/

Требования к результату:

1) список исправлений в формате:

sent_id \n
text \n
строка с ошибкой в частеречной разметке \n
строка с исправлениями \n

2) Список частей речи с расшифровками (http://universaldependencies.org/u/pos/index.html)

Содержание отчета

1. Цель и задачи лабораторной работы.
2. Исходные данные.
3. Полученная разметка.
4. Результаты проверки разметки.
5. Анализ и интерпретация полученных результатов.

Перечень вопросов к экзамену

1. Уровни лингвистического анализа в интеллектуальных системах.
2. Сегментация зашумленного текста.
3. Морфологический анализ текста.
4. Модели синтаксической структуры высказываний на естественном языке. Формальные грамматики и деревья зависимостей.
5. Алгоритмы синтаксического анализа (chart parsing, transition-based parsing)
6. Обучение синтаксических анализаторов с учителем.
7. Скрытые марковские модели и их применение в задачах частеречной разметки

Типовые практические задания для подготовки к экзамену

1. Произвести сегментацию (токенизацию) входного текста.
2. Произвести морфологический анализ текста.
3. Произвести синтаксический анализ входного текста.
4. Разработать шаблоны для поиска именованных сущностей в тексте.
5. Произвести поиск именованных сущностей в базе знаний.
6. Разработать онтологию для заданного фрагмента предметной области.
7. Произвести поиск по семантическим метаданным в базе знаний.

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

Фонды оценочных средств по дисциплине

Фонд оценочных средств позволяет оценить знания, умения и уровень приобретенных компетенций.

Фонд оценочных средств по дисциплине включает:

- вопросы к экзамену;
- комплект лабораторных работ.

Оценка качества освоения программы дисциплины включает текущий контроль успеваемости, промежуточную аттестацию, итоговую аттестацию.

Отчет по лабораторной работе

Отчет по лабораторной работе представляется в электронном виде. Защита отчета проходит в форме доклада студента по выполненной работе и ответов на вопросы преподавателя. В случае, если оформление отчета и поведение студента во время защиты соответствуют указанным требованиям, студент получает максимальное количество баллов.

Основаниями для снижения количества баллов в диапазоне от max до min являются:

- небрежное выполнение,
- отсутствие выводов,
- нарушение сроков предоставления отчета.

Отчет не может быть принят и подлежит доработке в случае:

- отсутствия необходимых разделов,
- неверных результатов расчета.

Экзамен

Основаниями для снижения оценки являются:

- ошибки в объяснениях и комментариях при верно выполненном задании;
- неполный ответ.

В соответствии с балльно-рейтинговой системой БАРС по дисциплине отводится 100 баллов (50 баллов на семестровую часть: 40 баллов – текущие формы контроля и до 10 баллов – на бонусы; 50 баллов – на экзаменационную часть).

Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения соответствующих работ. Он предусматривает проверку готовности студентов к плановым занятиям, оценку качества и самостоятельности выполнения заданий на практических занятиях, проверку правильности решения задач, выданных на самостоятельную проработку.

На экзамене осуществляется комплексная проверка знаний, навыков и умений студентов по материалу дисциплины на основании ответов на теоретические вопросы и решения практических задач.

Преподаватель, реализующий дисциплину, в зависимости от уровня подготовленности, обучающихся может использовать иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

а) Основная литература

1. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский.; Пер. с польского И. Д. Рудинского. - 2-е изд. , стереотип. - Москва: Горячая линия - Телеком, 2012. - 384 с. - ISBN 978-5-9912-0320-3. - Текст : электронный// ЭБС «Консультант студента»: [сайт]. - URL: https://www.studentlibrary.ru/book/ISBN9785991203203.html
2. Шапиро Л. , Стокман Дж. Компьютерное зрение / Л. Шапиро, Дж. Стокман; пер. с англ. - 4-е изд. - Москва : Лаборатория знаний, 2020. - 763 с. Систем. требования: Adobe Reader XI ; экран 10". (Лучший зарубежный учебник) - ISBN 978-5-00101-696-0. - Текст: электронный // ЭБС «Консультант студента»: [сайт]. - URL: https://www.studentlibrary.ru/book/ISBN9785001016960.html

б) Дополнительная литература

1. Переводческая семантография. Запись при устном переводе : Учебное пособие / Аликина Е.В. – М.: Издательство Юрайт, 2017 .– 145 .– (Бакалавр и магистр. Академический курс) .– ISBN 978-5-534-04601-4 : 56.25, 4 .– URL:http://www.biblio-online.ru/book/833E687A-36DC-478A-B7B0-263DF89F25AC
2. Грас Д. (2020) Data Science. Наука о данных с нуля. – СПб. : BHV-СПб, 2020 – 416 С. – ISBN 978-5-9775-6731-2
3. Селянкин Владимир Васильевич (2021) Компьютерное зрение. Анализ и обработка изображений. Учебное пособие для вузов. – М. : Издательство Лань, 2021 .– 152 С.
4. Р. Клетте. Компьютерное зрение. Теория и алгоритмы. – М. : Издательство ДМК Пресс, 2019 .– 506 С.
5. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М. : Издательство Техносфера, 2019 .– 1104 С.

в) Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимый для освоения дисциплины (модуля)

1. Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». Многопрофильный образовательный ресурс «Консультант студента» является электронной библиотечной системой, предоставляющей туп через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Каталог содержит более 15 000 наименований изданий. www.studentlibrary.ru.
2. Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». www.biblio-online.ru

3. Data Quality Fundamentals. – Образовательный портал udemy. – Режим доступа: https://www.udemy.com/course/data-quality-fundamentals/ .
4. Обработка изображений. – Образовательный портал stepik. – Режим доступа: https://stepik.org/course/1280/promo .
5. Getting Started with AI on Jetson Nano. – Образовательный портал nvidia – Режим доступа: https://courses.nvidia.com/courses/course-v1:DLI+S-RX-02+V2/about .
6. Введение в обработку естественного языка. – Образовательный портал Stepik. – Режим доступа: https://stepik.org/course/1233/promo

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Учебные аудитории, библиотеки АГУ, компьютерные классы, мультимедийные аудитории.

При необходимости рабочая программа дисциплины (модуля) может быть адаптирована для обеспечения образовательного процесса инвалидов и лиц с ограниченными возможностями здоровья, в том числе для обучения с применением дистанционных образовательных технологий. Для этого требуется заявление студента (его законного представителя) и заключение психолого-медико-педагогической комиссии (ПМПК).