Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 09.10.2023 15:48:40
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Факультет информационных технологий

УТВЕРЖДАЮ
Декан факультета
«Информационные технологии»
_____/Д.Г.Демидов/
«10» мая 2022

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Алгоритмы и методы компьютерной лингвистики»

Направление подготовки/специальность
09.03.02 Информационные системы и технологии

Профиль/специализация
«Программное обеспечение игровой компьютерной индустрии»

Квалификация
Бакалавр

Формы обучения
Очная

Москва, 2022 г.

Разработчик(и):

к.т.н., доцент кафедры
«Информатика и информационные технологии»  / П. С. Новиков /

Согласовано:
Заведующий кафедрой
«Информатика и информационные технологии»,
к.т.н.  / Е.В. Булатников /

Содержание

1. Цели, задачи и планируемые результаты обучения по дисциплине 4
2. Место дисциплины в структуре образовательной программы 4
3. Структура и содержание дисциплины 5
3.1 Виды учебной работы и трудоемкость 5
3.2 Тематический план изучения дисциплины 5
3.3 Содержание дисциплины 6
3.4 Тематика семинарских/практических и лабораторных занятий 9
3.5 Тематика курсовых проектов (курсовых работ) 10
4. Учебно-методическое и информационное обеспечение 10
4.1 Нормативные документы и ГОСТ 10
4.2 Основная литература 11
4.3 Дополнительная литература 11
4.4 Электронные образовательные ресурсы 11
4.5 Лицензионное и свободно распространяемое программное обеспечение 11
4.6 Современные профессиональные базы данных и информационные справочные системы 11
5. Материально-техническое обеспечение 12
6. Методические рекомендации 12
6.1 Методические рекомендации для преподавателя по организации обучения 12
6.2 Методические указания для обучающихся по освоению дисциплины 12
7. Фонд оценочных средств 13
7.1 Методы контроля и оценивания результатов обучения 13
7.2 Шкала и критерии оценивания результатов обучения 13
7.3 Оценочные средства 14

1. Цели, задачи и планируемые результаты обучения по дисциплине

Целью освоения дисциплины «Алгоритмы и методы компьютерной лингвистики» является приобретение знаний и навыков в области компьютерной лингвистики, овладение навыками разработки и реализации, а также применение их на практике для решения задач и понимание проблем и ограничений, а также овладение основными алгоритмами и методами, используемыми в этой области.

Основные задачи освоения дисциплины «Алгоритмы и методы компьютерной лингвистики» включают:

Понимание основных принципов и роли алгоритмов и методов в обработке естественного языка.

Изучение различных алгоритмов и методов компьютерной лингвистики, таких как машинное обучение, статистический анализ и синтаксический анализ.

Овладение навыками разработки и реализации алгоритмов, используемых в компьютерной лингвистике.

Получение практического опыта работы с реальными задачами компьютерной лингвистики и применение изученных алгоритмов для их решения.

Обучение по дисциплине «Алгоритмы и методы компьютерной лингвистики» направлено на формирование у обучающихся следующих компетенций:

Код и наименование компетенций	Индикаторы достижения компетенции
ПК-1. Способен разрабатывать требования и проектировать программное обеспечение	ИПК-1.1. Знает способы разработки требований и проектирования программного обеспечения для игровой компьютерной индустрии и требований дизайна к готовым продуктам ИПК-1.2. Умеет проектировать программное обеспечение с применением современных инструментальных средств в игровой компьютерной индустрии ИПК-1.3. Имеет навыки разработки требований и проектирования информационных и автоматизированных сред для игровой компьютерной индустрии

2. Место дисциплины в структуре образовательной программы

Дисциплина относится к модулю «Элективные дисциплины» обязательной части Блока 1. учебного плана программы бакалавриата.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

- Прототипирование игровых интерфейсов;
- Рефакторинг;
- Анализ программного кода;
- Производственная практика (преддипломная);
- Выполнение и защита выпускной квалификационной работы.

3. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы (144 часа).

3.1 Виды учебной работы и трудоемкость

№ п/п	Вид учебной работы	Количество часов	Семестры
			6
1	**Аудиторные занятия**	**72**	**72**
	В том числе:		
1.1	Лекции	36	36
1.2	Семинарские/практические занятия		
1.3	Лабораторные занятия	36	36
2	**Самостоятельная работа**	**72**	**72**
	В том числе:		
2.1	Подготовка и выполнение лабораторных работ	72	72
3	**Промежуточная аттестация**		
	Зачет		зачет
	Итого:	**144**	**144**

3.2 Тематический план изучения дисциплины

№ п/п	Разделы/темы дисциплины	Трудоемкость, час					
		Всего	Аудиторная работа				Самостоятельная работа
			Лекции	Семинарские/практические занятия	Лабораторные занятия	Практическая подготовка	
1.1	Тема 1. Естественно-языковые знаковые системы	4	2				2
1.2.	Лабораторная работа № 1 «Частотный анализ текста»	4			2		2
2.1	Тема 2. Модели и методы представления и организации знаний	4	2				2
2.2	Лабораторная работа № 2 «Работа с текстом. Конспектирование материалов»	4			2		2
3.1	Тема 3. Количественная спецификация ЕЯ систем	4	2				2
3.2	Лабораторная работа № 3 «Построение моделей»	4			2		2
4.1	Тема 4. Логико-статистические методы извлечения знаний	4	2				2
4.2	Лабораторная работа № 4 «Построение моделей»	6			2		4

5.1	Тема 5. Технология автоматизированного построения словаря-тезауруса	4	2				2
5.2	Лабораторная работа № 5 «Создание алгоритмов»	6			2		4
6.1	Тема 6. «Пример исследования ЕЯ ресурса»	4	2				2
6.2	Лабораторная работа № 6 «Автоматическая обработка текста»	6			2		4
7.1	Тема 7. Теоретические основы семиотики информационных технологий	4	2				2
7.2	Лабораторная работа № 7 «Семиотический анализ информационных технологий»	6			2		4
8.1	Тема 8. Основы семиотики	4	2				2
8.2	Лабораторная работа № 8 «Семиотический анализ текстов»	6			2		4
9.1	Тема 9. Строение знаковых систем	6	4				2
9.2	Лабораторная работа № 9 «Анализ структуры знаковых систем»	8			4		4
10.1	Тема 10. Семиотическая концепция проектирования	6	4				2
10.2	Лабораторная работа № 10 «Применение семиотической концепции в проектировании информационных систем»	8			4		4
11.1	Тема 11. Концептуальное проектирование	6	4				2
11.2	Лабораторная работа № 11 «Разработка концептуальной модели информационной системы»	8			4		4
12.1	Тема 12. Метафорическое проектирование.	6	4				2
12.2	Лабораторная работа № 12 «Применение метафорического проектирования в разработке интерфейсов»	8			4		4
13.1	Тема 13. Парадигматическое проектирование.	6	4				2
13.2	Лабораторная работа № 13 «Применение парадигматического проектирования в разработке компьютерных алгоритмов»	8			4		4
Итого		**144**	**36**		**36**		**72**

3.3 Содержание дисциплины

Тема 1. Естественно-языковые знаковые системы:

- Определение и классификация естественно-языковых знаковых систем.
- Обзор основных компонентов естественно-языковых знаковых систем (фонология, морфология, синтаксис, семантика, прагматика).
- Изучение основных теорий и моделей, описывающих естественно-языковые знаковые

системы (например, структурная лингвистика, генеративная грамматика).
- Рассмотрение приложений и применения естественно-языковых знаковых систем в компьютерной лингвистике.

Тема 2. Модели и методы представления и организации знаний:
- Изучение основных моделей и представлений знаний в компьютерной лингвистике (например, онтологии, семантические сети, графовые модели).
- Рассмотрение различных методов и алгоритмов для организации и структурирования знаний.
- Изучение семантического анализа и семантического поиска в текстах с использованием моделей и методов представления знаний.
- Анализ использования моделей и методов представления знаний в приложениях компьютерной лингвистики.

Тема 3. Количественная спецификация ЕЯ систем:
- Рассмотрение методов количественного анализа естественно-языковых знаковых систем.
- Изучение статистических методов обработки текстов и анализа частотности использования слов и конструкций.
- Анализ корпусов текстов и использование методов для извлечения статистической информации о естественно-языковых знаковых системах.
- Рассмотрение приложений количественной спецификации естественно-языковых знаковых систем в компьютерной лингвистике.

Тема 4. Логико-статистические методы извлечения знаний:
- Изучение методов извлечения семантической информации из текстов с использованием логико-статистических моделей.
- Рассмотрение моделей и алгоритмов для автоматического извлечения и классификации информации (например, именованные сущности, отношения, события).
- Анализ использования логико-статистических методов извлечения знаний в приложениях компьютерной лингвистики, таких как авто-комплит, автоматическое аннотирование текстов и т.д.

Тема 5. Технология автоматизированного построения словаря-тезауруса:
- Ознакомление с методами и инструментами для автоматизированного построения словарей и тезаурусов.
- Рассмотрение алгоритмов для автоматической лемматизации, морфологического анализа и синтаксического разбора текстов.
- Изучение методов для автоматического извлечения семантических отношений и построения семантических сетей.
- Анализ использования технологии автоматизированного построения словаря-тезауруса в приложениях компьютерной лингвистики.

Тема 6. «Пример исследования ЕЯ ресурса»:
- Ознакомление со значимостью исследования естественно-языковых ресурсов в компьютерной лингвистике.
- Рассмотрение примеров исследования естественно-языковых текстов и корпусов для извлечения лингвистической информации и семантической структуры.

- Изучение методов аннотирования и разметки текстовых ресурсов для дальнейшего использования в компьютерной лингвистике.
- Анализ применения естественно-языковых ресурсов в различных задачах, таких как машинный перевод, распознавание речи, обработка естественного языка.

Тема 7. Теоретические основы семиотики информационных технологий:
- Ознакомление с основными понятиями и теоретическими принципами семиотики информационных технологий.
- Рассмотрение роли символов, знаков и сигналов для передачи и интерпретации информации в информационных системах.
- Изучение основных моделей и методов семиотического анализа в контексте информационных технологий.
- Анализ применения семиотических подходов в разработке и проектировании информационных систем и интерфейсов.

Тема 8. Основы семиотики:
- Рассмотрение основных понятий и теоретических подходов в семиотике.
- Изучение различных типов знаков и их классификаций в контексте семиотики.
- Ознакомление со структурой знаковых систем и взаимосвязью их компонентов.
- Анализ роли семиотики в различных областях, таких как лингвистика, психология, культурология.

Тема 9. Строение знаковых систем:
- Изучение основных компонентов знаковых систем (значение, образ, выражение).
- Рассмотрение различных подходов к анализу и классификации знаковых систем (например, троек zшинори, пирамида Пирса).
- Изучение структурных моделей знаковых систем и их приложений в различных областях, включая компьютерную лингвистику.
- Анализ взаимосвязи между знаковыми системами и естественными языками.

Тема 10. Семиотическая концепция проектирования:
- Ознакомление с основными принципами и методами семиотической концепции в проектировании информационных систем.
- Рассмотрение роли знаков, символов и семиотических моделей в процессе проектирования.
- Изучение принципов проектирования информационных систем с использованием семиотического подхода.
- Анализ применения семиотической концепции в различных областях, таких как веб-дизайн, разработка интерфейсов, создание информационных архитектур.

Тема 11. Концептуальное проектирование:
- Рассмотрение основных принципов и методов концептуального проектирования информационных систем.
- Изучение процесса создания и описания концептуальной модели информационной системы.
- Ознакомление с способами представления и структурирования знаний в концептуальных моделях.

- Анализ применения концептуального проектирования в различных областях, таких как базы данных, онтологии, системы управления знаниями.

Тема 12. Метафорическое проектирование.

Применение алгоритмов машинного обучения в компьютерной лингвистике, включая классификацию текстов, анализ тональности, перевод, извлечение информации и тематическое моделирование.

Тема 13. Парадигматическое проектирование.

Алгоритмы и методы обработки естественного языка в компьютерной лингвистике, включая синтаксический и семантический анализ, дисамбигуацию, генерацию текста и применение в реальных приложениях.

3.4 Тематика семинарских/практических и лабораторных занятий

Семинарские и практические занятия не предусмотрены.

Лабораторные занятия

Лабораторная работа № 1 «Частотный анализ текста»:
Задание: Напишите алгоритм для подсчета частоты встречаемости каждого слова в заданном тексте. Реализуйте программу, которая принимает текстовый файл в качестве входных данных и выводит количество и процентное соотношение каждого слова в тексте.

Лабораторная работа № 2 «Работа с текстом. Конспектирование материалов»:
Задание: Разработайте программу, которая позволит пользователю создавать и редактировать конспекты текстовых материалов. Реализуйте функции добавления, удаления и редактирования разделов и подразделов, а также сохранение и загрузку конспекта из файла.

Лабораторная работа № 3 «Построение моделей»:
Задание: Исследуйте и постройте модель машинного обучения для определения тональности текста. Разработайте программу, которая сможет классифицировать текст на положительный, отрицательный или нейтральный, основываясь на обученной модели.

Лабораторная работа № 4 «Построение моделей»:
Задание: Проанализируйте и сравните несколько алгоритмов кластеризации текстовых данных. Реализуйте программу, которая сможет проводить кластерный анализ текстовых документов и выводить результаты.

Лабораторная работа № 5 «Создание алгоритмов»:
Задание: Разработайте алгоритм для автоматического выделения ключевых слов из текста. Реализуйте программу, которая будет использовать этот алгоритм для создания списка ключевых слов на основе входного текстового файла.

Лабораторная работа № 6 «Автоматическая обработка текста»:
Задание: Разработайте программу для автоматической обработки текста: удаление стоп-слов, лемматизация, определение именованных сущностей и т.д. Реализуйте различные модули обработки текста, которые могут быть применены к разным текстовым данным.

Лабораторная работа № 7 «Семиотический анализ информационных технологий»:
Задание: Исследуйте и проанализируйте семиотические аспекты информационных техно-

логий. Разработайте программу, которая проведет семиотический анализ заданного информационного технологического процесса и выведет результаты анализа.

Лабораторная работа № 8 «Семиотический анализ текстов»:
Задание: Разработайте программу для семиотического анализа текстов. Пользователь должен иметь возможность вводить текст и получать результаты семиотического анализа, включающие анализ знаковых систем и их взаимодействия.

Лабораторная работа № 9 «Анализ структуры знаковых систем»:
Задание: Проведите анализ структуры заданной знаковой системы. Разработайте программу, которая позволит пользователю задавать иерархию знаков и определять их связи и взаимодействия.

Лабораторная работа № 10 «Применение семиотической концепции в проектировании информационных систем»:
Задание: Примените семиотическую концепцию при проектировании информационной системы. Разработайте программу, которая позволит создать прототип информационной системы с учетом семиотических аспектов, таких как знаковая интерпретация, коммуникация и взаимодействие.

Лабораторная работа № 11 «Разработка концептуальной модели информационной системы»:
Задание: Разработайте концептуальную модель информационной системы на основе заданных требований. Реализуйте программу, которая позволит создавать диаграммы классов, атрибутов и связей, представляющие концептуальную модель информационной системы.

Лабораторная работа № 12 «Применение метафорического проектирования в разработке интерфейсов»:
Задание: Примените метафорическое проектирование при разработке пользовательского интерфейса. Разработайте программу, которая позволит создать интерфейс, основанный на метафорах и представлениях, упрощающих взаимодействие пользователя с системой.

Лабораторная работа № 13 «Применение парадигматического проектирования в разработке компьютерных алгоритмов»:
Задание: Примените парадигматическое проектирование при разработке компьютерных алгоритмов. Разработайте программу, которая использует разные парадигмы программирования (например, процедурное, объектно-ориентированное или функциональное программирование) при реализации алгоритмов.

3.5 Тематика курсовых проектов (курсовых работ)

Курсовой проект не предусмотрен.

4. Учебно-методическое и информационное обеспечение

4.1 Нормативные документы и ГОСТ

1. Федеральный закон от 29 декабря 2012 года No 273-ФЗ «Об образовании в Российской Федерации» (с изменениями и дополнениями);

2. Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.02 Информационные системы и технологии, утвержденный Приказом Министерства образования и науки РФ от 19 сентября 2017 г. N 929 "Об утверждении федерального... Редакция с изменениями N 1456 от 26.11.2020;

3. Приказ Министерства образования и науки РФ от 05 апреля 2017 г. No 301 «Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры.

4.2 Основная литература

1. Павлов А. С., Добров Б. В. Метод обнаружения массово порожденных неестественных текстов на основе анализа тематической структуры // Вычислительные методы и программирование: новые вычислительные технологии. — 2011. — Т. 12.
2. Воронцов К. В., Потапенко А. А. Регуляризация, робастность и разреженность вероятностных тематических моделей // Компьютерные исследования и моделирование. — 2012. — Т. 4, № 4.
3. Воронцов К. В. Аддитивная регуляризация тематических моделей коллекций текстовых документов // Доклады РАН. — 2014. — Т. 456, № 3.
4. Айсина Р. М. Обзор средств визуализации тематических моделей коллекций текстовых документов // Машинное обучение и анализ данных (http://jmlda.org). — 2015. — Т. 1, № 11.
5. Воронцов К. В., Потапенко А. А. Модификации EM-алгоритма для вероятностного тематического моделирования // Машинное обучение и анализ данных. — 2013. — Т. 1, № 6.
6. Янина А. О., Воронцов К. В. Мультимодальные тематические модели для разведочного поиска в коллективном блоге // Машинное обучение и анализ данных. — 2016. — Т. 2, № 2.

4.3 Дополнительная литература

1. Моделирование языковой деятельности в интеллектуальных системах (Под ред. А.Е. Кибрика и А.С. Нариньяни). – М., 1987.
2. Искусственный интеллект. В 3-х кн. М., 1990. - Кн. 1. Системы общения и экспертные системы. Кн. 2. Модели и методы. Использование ЭВМ в лингвистических исследованиях. Киев, 1989.
3. Лакофф Дж., Джонсон М. Метафоры, которыми мы живем. // Язык и моделирование социального взаимодействия. - М., 1987.

4.4 Электронные образовательные ресурсы

ЭОР не предусмотрен.

4.5 Лицензионное и свободно распространяемое программное обеспечение

1. Python
2. Natural Language Toolkit (NLTK)
3. Stanford CoreNLP

4.6 Современные профессиональные базы данных и информационные справочные системы

1. https://urait.ru/
2. https://www.iprbookshop.ru/
3. https://e.lanbook.com/

5. Материально-техническое обеспечение

Компьютерные классы со следующей оснащенностью: столы, стулья, аудиторная доска, использование переносного мультимедийного комплекса (переносной проектор, персональный ноутбук). Персональные компьютеры, мониторы, мышки, клавиатуры. Рабочее место преподавателя: стол, стул.

Программное обеспечение: Microsoft Windows или Linux, Stanford CoreNLP, WordNet, WEKA.

6. Методические рекомендации

6.1 Методические рекомендации для преподавателя по организации обучения

Методика преподавания дисциплины «Алгоритмы и методы компьютерной лингвистики» предусматривает использование онлайн-курса в системе дистанционного обучения Университета, групповых и индивидуальных консультаций обучающихся, аудиторных занятий в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся.

Лабораторные работы по дисциплине «Алгоритмы и методы компьютерной лингвистики» осуществляются в форме самостоятельной проработки теоретического материала обучающимися; выполнения практического задания; защиты преподавателю лабораторной работы (знание теоретического материала и выполнение практического задания по теме лабораторной работы).

6.2 Методические указания для обучающихся по освоению дисциплины

Изучение дисциплины осуществляется в соответствии с учебным планом.

Методические указания для обучающихся по дисциплине "Алгоритмы и методы компьютерной лингвистики" включают в себя онлайн-курсы, групповые и индивидуальные консультации, аудиторные занятия и лабораторные работы. Студентам предоставляется возможность самостоятельно проработать теоретический материал и выполнить практические задания. Целью методики является формирование профессиональных навыков, необходимых для работы в области компьютерной лингвистики. Обучающиеся получат необходимые знания и навыки для работы с алгоритмами и методами компьютерной лингвистики, что поможет им успешно решать задачи в своей будущей профессиональной деятельности.

Текущий контроль осуществляется на аудиторных занятиях в виде защиты лабораторных работ. Критериями оценки результатов являются:

- уровень освоения теоретического материала;
- уровень владения практическими навыками (в виде вопросов по процессу выполнения лабораторных работ);
- умения обучающегося использовать теоретические знания при выполнении практических задач (в виде дополнительных заданий);
- сформированность компетенций;
- оформление материала в соответствии с требованиями.

Промежуточный контроль осуществляется на зачете в форме тестирования в системе дистанционного обучения Университета, включающего вопросы на знание тестирование программного обеспечения.

7. Фонд оценочных средств

7.1 Методы контроля и оценивания результатов обучения

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций: **лабораторные работы**, **зачёт**.

7.2 Шкала и критерии оценивания результатов обучения

К промежуточной аттестации допускаются только студенты, выполнившие все виды учебной работы, предусмотренные рабочей программой по дисциплине «Алгоритмы и методы компьютерной лингвистики».

3.2.1. Критерии оценки ответа на зачёте

(формирование компетенций — ПК-1.)

«Зачтено»:

Выполнены все виды учебной работы, предусмотренные учебным планом. Обучающийся демонстрирует прочные теоретические знания, практические навыки, владеет терминами, делает аргументированные выводы и обобщения, приводит примеры, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации, которые обучающийся может исправить при коррекции преподавателем.

«Не зачтено»:

Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Обучающийся демонстрирует незнание теоретических основ предмета, отсутствие практических навыков, не умеет делать аргументированные выводы и приводить примеры, не владеет терминами, проявляет отсутствие логичности и последовательности изложения, делает ошибки, которые не может исправить даже при коррекции преподавателем, отказывается отвечать на дополнительные вопросы, допускает значительные ошибки, испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

3.2.2. Критерии оценки работы обучающегося на лабораторных занятиях:

(формирование компетенций — ПК-1.)

«5» (отлично): выполнены все практические задания, предусмотренные лабораторными работами, обучающийся четко и без ошибок ответил на все контрольные вопросы, проявил творческий подход при выполнении заданий, смог выполнить дополнительные задания.

«4» (хорошо): выполнены все практические задания, предусмотренные лабораторными работами, обучающийся с корректирующими замечаниями преподавателя ответил на все контрольные вопросы, проявил творческий подход при выполнении заданий, смог частично выполнить дополнительные задания.

«3» (удовлетворительно): выполнены все практические задания, предусмотренные лабораторными работами, с замечаниями преподавателя; обучающийся ответил на все контрольные вопросы с замечаниями, дополнительные задания выполнены с замечаниями.

«2» (неудовлетворительно): обучающийся не выполнил или выполнил неправильно практические задания, предусмотренные лабораторными работами, обучающийся ответил на контрольные вопросы с ошибками или не ответил на контрольные вопросы, дополнительные задания выполнены неверно или не выполнены.

7.3 Оценочные средства

7.3.1 Промежуточная аттестация

Текущий контроль осуществляется на аудиторных занятиях в виде защиты лабораторных работ. Лабораторная работа – средство контроля усвоения учебного материала темы, раздела или разделов дисциплины, организованное как учебное занятие в виде демонстрации полученных навыков при решении поставленных практических задач.
Примеры вопросов к защите лабораторных работ (оцениваемые компетенции — ПК-1).

1. Какой язык программирования рекомендуется использовать для работы с алгоритмами и методами компьютерной лингвистики?
2. Что такое Natural Language Toolkit и какие инструменты для работы с текстами она содержит?
3. Какой программный комплекс позволяет проводить морфологический, синтаксический и семантический анализ текстов на различных языках?
4. Что такое библиотека Gensim и для чего она используется?
5. Какую информацию содержит лексическая база данных WordNet?
6. Какая библиотека для обработки естественного языка на языке Java содержит инструменты для токенизации, частеречной разметки, извлечения именованных сущностей и других задач?
7. Для чего используется библиотека Apache Lucene?
8. Что такое WEKA и какие методы анализа данных она содержит?
9. Какая среда разработки используется для статистического анализа данных?
10. Какие пакеты для работы с текстовыми данными и анализа естественного языка содержит язык программирования R?
11. Какие инструменты из перечисленных рекомендуется использовать для работы с тематическим моделированием и векторным представлением слов в текстах?
12. Какие инструменты из перечисленных рекомендуется использовать для поиска и индексации текстовых данных?
13. Какие инструменты из перечисленных рекомендуется использовать для классификации и кластеризации данных?