

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В.Ломоносова»

УТВЕРЖДАЮ

Декан факультета ВМК МГУ
имени М.В. Ломоносова

академик

Е.И. Моисеев



« ____ » _____ 2018 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Обработка текстов»

Уровень высшего образования – подготовка научно-педагогических кадров в аспирантуре

Направление подготовки – 09.06.01 «Информатика и вычислительная техника»

Направленность (профиль) – 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

1. НАИМЕНОВАНИЕ ДИСЦИПЛИНЫ

Обработка текстов

2. УРОВЕНЬ ВЫСШЕГО ОБРАЗОВАНИЯ

аспирантура

3. НАПРАВЛЕНИЕ ПОДГОТОВКИ, НАПРАВЛЕННОСТЬ (ПРОФИЛЬ) ПОДГОТОВКИ

Направление подготовки: 09.06.01 «Информатика и вычислительная техника»

Направленность (профиль) подготовки: 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

4. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к специальным курсам по выбору вариативной части образовательной программы.

5. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

Дисциплина участвует в формировании следующих компетенций образовательных программ:

Формируемые компетенции	Планируемые результаты обучения
Владением методологией теоретических и экспериментальных исследований в области профессиональной деятельности (ОПК-1)	З1 (ОПК-1) ЗНАТЬ: современные математические методы, применяющиеся для решения задач в области естественных наук, экономики, социологии и информационно-коммуникационных технологий У1 (ОПК-1) УМЕТЬ: применять современные методы постановки и анализа задач в области математики и информатики В1 (ОПК-1) ВЛАДЕТЬ: навыками оптимального выбора современных методов и средств

	постановки и анализа задач в области математики и информатики
<p>Владение современными методами построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также методами разработки и реализации алгоритмов их решения на основе фундаментальных знаний в области математики и информатики</p> <p>(ПК-1)</p>	<p>ЗНАТЬ: классические методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также базовые методы разработки и реализации алгоритмов их решения;</p> <p>УМЕТЬ: применять классические методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также базовые методы разработки и реализации алгоритмов их решения;</p> <p>ВЛАДЕТЬ: базовыми навыками выбора методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также методов разработки и реализации алгоритмов их решения.</p>
<p>Способность разрабатывать и реализовывать алгоритмы организации работы современных вычислительных комплексов и компьютерных сетей</p> <p>(ПК-2)</p>	<p>ЗНАТЬ: классические методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей;</p> <p>УМЕТЬ: применять классические методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей;</p> <p>ВЛАДЕТЬ: базовыми навыками выбора методов разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей</p>

6. ОБЪЕМ ДИСЦИПЛИНЫ

Объем дисциплины составляет 3 зачетные единицы, всего 108 часов.

36 часов составляет контактная работа с преподавателем – 32 часа занятий лекционного типа; 4 часа мероприятий текущего контроля успеваемости.

72 часа составляет самостоятельная работа учащегося.

7. ВХОДНЫЕ ТРЕБОВАНИЯ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Учащиеся должны владеть знаниями по математическим курсам, а также по курсам, связанным с основами программирования и системами программирования, в объеме, соответствующем основным образовательным программам магистратуры по укрупненным группам направлений и специальностей 01.00.00 «Математика и механика», 02.00.00 «Компьютерные и информационные науки».

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Процесс изложения учебного материала сопровождается презентациями с использованием мультимедийного проектора.

9. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Курс знакомит слушателей с основными проблемами компьютерной обработки текстов. Рассматриваются как фундаментальные понятия и идеи, так и современные исследования в данной области. Особое внимание уделяется применению методов машинного обучения, которые активно развиваются в настоящее время и показывают лучшие результаты. В курсе даётся обзор базовых задач в области обработки текстов на естественном языке, описываются проблемы, возникающих при разработке систем текстового анализа, и подходы к их решению.

The course is devoted to natural language processing. Fundamental concepts and ideas, as well as modern research are described. Particular attention is paid to the use of machine learning methods which are being actively developed at the present time and show the best results. The course provides an overview of basic natural language processing tasks, describes the problems arising in the development of text analysis systems, and approaches to solving them.

Наименование и краткое содержание разделов и тем	Всего (часы)	В том числе	
		Контактная работа (работа во взаимодействии с преподавателем), часы из них	Самостоятельная работа учащегося, часы из них

дисциплины, форма промежуточной аттестации по дисциплине		Занятия лекционно го типа	Занятия семинарско го типа	Групповые консультац ии	Индивидуальн ые консультации	Учебные занятия, направленн ые на проведение текущего контроля успеваемост и: коллоквиум ы, практическ ие контрольны е занятия и др.	Всего	Выполнен ие домашних заданий	Подготов ка реферато в и т.п..	Всего
Тема 1. Распознавание именованных сущностей, машинное обучение с учителем. Основы машинного обучения с учителем, линейные классификаторы: метод опорных векторов, логистическая	3	2	-	-	-	-	2	1	-	1

регрессия. Задача распознавания именованных сущностей: постановка задачи, области применения, простейшие алгоритмы решения задачи, методы оценки качества.										
Тема 2. Разметка последовательности, нейронные сети. Понятие разметки последовательности и на примере задачи распознавания именованных сущностей. Условные случайные поля, алгоритм Витерби. Нейронные сети. Нелинейность и дифференцируемость функций в нейронных сетях. Алгоритм	3	2	-	-	-	-	2	1	-	1

<p>обратного распространения ошибок. Типовые практики обучения нейронных сетей.</p>										
<p>Тема 3.Синонимия: дистрибутивные векторные представления слов.</p> <p>Векторные представления слов. Дистрибутивная гипотеза. Локальные модели векторов слов: continous skip-gram, continous bag of words. Иерархический softmax и negative</p>	<p>3</p>	<p>2</p>	<p>-</p>	<p>-</p>	<p>-</p>	<p>-</p>	<p>2</p>	<p>1</p>	<p>-</p>	<p>1</p>

<p>sampling. Глобальные модели векторов слов: GloVe. Способы оценки качества векторов слов.</p>										
---	--	--	--	--	--	--	--	--	--	--

<p>Тема 4.Символьные представления слов.</p> <p>Использование символьного состава слов в методах обработки текстов на примере задачи распознавания именованных сущностей. Представления слов на основе мешка символьных последовательностей, попарного кодирования байтов, рекуррентных и сверточных нейронных сетей.</p>	5	4	-	-	-	-	4	1	-	1
--	---	---	---	---	---	---	---	---	---	---

<p>Тема 5. Базовые задачи обработки текстов.</p> <p>Сегментация текста: задачи токенизации и определения границ предложений. Задача определения языка текста. Профили языка. Naïve Bayes классификатор. Оценка качества классификации. Задача определения частей речи и морфологического анализа. Multi- label классификация. Лемматизация: Ripple Down Rules и LemmaGen. Грамматическая омонимия. Лемматизация как задача классификации.</p>	5	4	-	-	-	-	4	1	-	1
--	---	---	---	---	---	---	---	---	---	---

<p>Тема 6. Синтаксический анализ.</p> <p>Понятие формальной грамматики, иерархия Хомского. Генерация текста по формальной грамматике. Дерево составляющих. Разбор предложения по грамматике составляющих: Алгоритм Кока-Янгера-Касами. Дерево зависимостей. Разбор предложения на основе переходов: Arc-standard, Arc-eager. Методы разбора для непроективных деревьев: Attardi's system, Online reordering. Синтаксический</p>	5	4	-	-	-	-	4	1	-	1
--	---	---	---	---	---	---	---	---	---	---

<p>разбор на основе Stack LSTM. Оценка качества построенных деревьев разбора.</p>										
<p>Тема 7. Машинный перевод.</p> <p>Модель зашумленного канала. Статистический машинный перевод. Языковая модель. Модель перевода. Выравнивания. IBM Model 1. Декодирование. Нейросетевой машинный перевод. Архитектура кодировщик-декодировщик.</p>	5	4	-	-	-	-	4	1	-	1

<p>Методы кодирования. Методы декодирования. Оценка качества.</p>										
<p>Тема 8. Разрешение кореферентности.</p> <p>Кореферентность и анафоричность. Модель "пара упоминаний". Классификатор пар упоминаний. Отбор пар упоминаний. Объединение пар упоминаний в кореферентные цепочки. Модель "сущность - упоминание". Модели на основе ранжирования. Поиск</p>	3	2	-	-	-	-	2	1	-	1

упоминаний в тексте. Оценка качества.										
Тема 9. Другие задачи обработки текстов. Варианты постановки задачи извлечения мнений. Анализ тональности текста. Анализ тональности на уровне объектов (аспектов). Задача автоматического реферирования. Варианты постановок задачи. Статистические методы. Графовые	3	2	-	-	-	-	2	1	-	1

<p>методы. Методы оценки качества рефератов. Диалоговые системы. Диалоги с определенной целью. Диалоги без цели.</p>										
<p>Тема 10. Извлечение отношений.</p> <p>Постановки задачи извлечения отношений. Методы извлечения отношений, ограниченные рамками одного предложения: с применением синтаксической структуры и без нее. Методы извлечения сложных отношений: механизм</p>	3	2	-	-	-	-	2	1	-	1

внимания, обобщение рекуррентных нейронных сетей на случай графа.										
Тема 11. Привязка к базам знаний. Базы знаний. Виды контекста: локальный, глобальный. Поиск упоминаний в тексте. Оценка качества.	3	2	-	-	-	-	2	1	-	1
Тема 12. Перенос знаний, совместное обучение. Методы переноса знаний на основе построения векторных представлений слов, учитывающих контекст словоупотребления. Понятие	3	2	-	-	-	-	2	1	-	1

совместного обучения. Методы совместного обучения на базе нейронных сетей.										
Выполнение практического задания № 1	12	-	-	-	-	-	0	12		12
Выполнение практического задания № 2	12	-	-	-	-	-	0	12		12
Промежуточная аттестация - устный экзамен	40	-	-	-	-	4	4	36	-	36
Итого	108	36					72			

10. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ УЧАЩИХСЯ

Самостоятельная работа учащихся состоит в изучении лекционного материала, учебно-методической литературы, выполнении практических заданий и промежуточной аттестации.

11. РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ

Основная учебно-методическая литература

- 1) Daniel Jurafsky and James H. Martin. 2008. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Second Edition. Prentice Hall.
- 2) Christopher D. Manning and Hinrich Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press.
- 3) Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python. O'Reilly Media, 2009 (<http://www.nltk.org/book>)
- 4) Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep learning. MIT Press. 2016

Ресурсы информационно-телекоммуникационной сети «Интернет»

- 1) Морфологический анализатор pymorphy2: <https://pymorphy2.readthedocs.io/>
- 2) Natural Language Toolkit: <https://www.nltk.org/>
- 3) TensorFlow: <https://www.tensorflow.org/>
- 4) Keras: The Python Deep Learning library: <https://keras.io/>
- 5) Scikit-learn: <https://scikit-learn.org/>

Информационные технологии, используемые в процессе обучения

В процессе обучения используются технологии, доступные в пакетах pymorphy2, nltk, tensorflow, keras, scikit-learn: морфологический анализ текста, классификаторы SVM, случайный лес, нейронные сети прямого распространения, рекуррентные и сверточные нейронные сети.

Материально-техническая база

Медиапроектор, ноутбук и экран для проведения лекций-презентаций.

12. ЯЗЫК ПРЕПОДАВАНИЯ

Русский

13. РАЗРАБОТЧИК ПРОГРАММЫ

Разработчики – И.А. Андрианов, В.Д. Майоров, А.А. Сысоев, Д.Ю. Турдаков

14. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

РЕЗУЛЬТАТ ОБУЧЕНИЯ	КРИТЕРИИ и ПОКАЗАТЕЛИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ					ОЦЕНОЧНЫЕ СРЕДСТВА
	1	2	3	4	5	
	Неудовлетворительн о	Неудовлетворительн о	Удовлетворительн о	Хорошо	Отлично	
ЗНАТЬ: современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных	Отсутствие знаний	Фрагментарные представления о современных методах построения и анализа математических моделей, возникающих при решении естественнонаучных	В целом сформированные, но неполные знания о современных методах построения и анализа математических моделей, возникающих при	Сформированные, но содержащие отдельные пробелы знания о современных методах построения и анализа математических	Сформированные систематические знания о современных методах построения и анализа математических моделей,	Дисциплины вариативной части, факультативные дисциплины

задач, а также современные методы разработки и реализации алгоритмов их решения Код 31 (ПК-1)		задач, а также современных методах разработки и реализации алгоритмов их решения	решении естественнонаучных задач, а также современных методах разработки и реализации алгоритмов их решения	моделей, возникающих при решении естественнонаучных задач, а также современных методах разработки и реализации алгоритмов их решения	возникающих при решении естественнонаучных задач, а также современных методах разработки и реализации алгоритмов их решения	
УМЕТЬ: применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения Код У1 (ПК-1)	Отсутствие умений	Фрагментарные умения применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения	В целом успешное, но не систематическое умение применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения	Успешное, но содержащее отдельные пробелы умение применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения	Сформированное умение применять современные методы построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современные методы разработки и реализации алгоритмов их решения	Исследовательская практика
ВЛАДЕТЬ: навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при	Отсутствие навыков	Фрагментарное владение навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при решении естественнонаучных задач, а также	В целом успешное, но не полное владение навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при	Успешное, но содержащее отдельные пробелы владение навыками оптимального выбора современных методов построения	Сформированное владение навыками оптимального выбора современных методов построения и анализа математических моделей, возникающих при	Научные исследования

решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения Код В1 (ПК-1)		современных методов разработки и реализации алгоритмов их решения	решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения	и анализа математических моделей, возникающих при решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения	решении естественнонаучных задач, а также современных методов разработки и реализации алгоритмов их решения	
ЗНАТЬ: современные методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения Код 31 (ПК-2)	Отсутствие знаний	Фрагментарные представления о современных методах разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	В целом сформированные, но неполные знания о современных методах разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Сформированные, но содержащие отдельные пробелы знания о современных методах разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Сформированные систематические знания о современных методах разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Дисциплины вариативной части, факультативные дисциплины
УМЕТЬ: применять современные методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Отсутствие умений	Фрагментарные умения применять современные методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	В целом успешное, но не систематическое умение применять современные методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей	Успешное, но содержащее отдельные пробелы умение применять современные методы разработки и реализации алгоритмов организации работы вычислительных комплексов и	Сформированное умение применять современные методы разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Исследовательская практика

поколения Код У1 (ПК-2)			последнего поколения	компьютерных сетей последнего поколения	поколения	
ВЛАДЕТЬ: навыками оптимального выбора современных методов разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения Код В1 (ПК-2)	Отсутствие навыков	Фрагментарное владение навыками оптимального выбора современных методов разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	В целом успешное, но не полное владение навыками оптимального выбора современных методов разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Успешное, но содержащее отдельные пробелы владение навыками оптимального выбора современных методов разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Сформированное владение навыками оптимального выбора современных методов разработки и реализации алгоритмов организации работы вычислительных комплексов и компьютерных сетей последнего поколения	Научные исследования

<p>ЗНАТЬ: современные математические методы, применяющиеся для решения задач в области естественных наук, экономики, социологии и информационно-коммуникационных технологий Код 31 (ОПК-1)</p>	<p>Отсутствие знаний</p>	<p>Фрагментарные представления о современных математических методах, применяющихся для решения задач в области естественных наук, экономики, социологии и информационно-коммуникационных технологий</p>	<p>В целом сформированные, но неполные знания о современных математических методах, применяющихся для решения задач в области естественных наук, экономики, социологии и информационно-коммуникационных технологий</p>	<p>Сформированные, но отдельные пробелы в знаниях о современных математических методах, применяющихся для решения задач в области естественных наук, экономики, социологии и информационно-коммуникационных технологий</p>	<p>Сформированные систематические знания о современных математических методах, применяющихся для решения задач в области естественных наук, экономики, социологии и информационно-коммуникационных технологий</p>	<p>Устный экзамен</p>
<p>УМЕТЬ: применять современные методы постановки анализа задач в области математики и информатики Код У1 (ОПК-1)</p>	<p>Отсутствие умений</p>	<p>Фрагментарные умения применять современные методы постановки анализа задач в области математики и информатики</p>	<p>В целом успешное, но не систематическое умение применять современные методы постановки анализа задач в области математики и информатики</p>	<p>Успешное, но содержащее отдельные пробелы умение применять современные методы постановки анализа задач в области математики и информатики</p>	<p>Сформированное умение применять современные методы постановки анализа задач в области математики и информатики</p>	<p>Устный экзамен</p>
<p>ВЛАДЕТЬ: навыками оптимального выбора современных методов и средств постановки анализа задач в области математики и информатики</p>	<p>Отсутствие навыков</p>	<p>Фрагментарное владение навыками оптимального выбора современных методов и средств постановки анализа задач в области математики и информатики</p>	<p>В целом успешное, но не полное владение навыками оптимального выбора современных методов и средств постановки анализа задач в области математики и информатики</p>	<p>Успешное, но содержащее отдельные пробелы владение навыками оптимального выбора современных методов и средств постановки анализа задач в области математики и информатики</p>	<p>Сформированное владение навыками оптимального выбора современных методов и средств постановки анализа задач в области математики и информатики</p>	<p>Устный экзамен</p>

Код В1 (ОПК-1)				постановки анализа задач в области математики и информатики		
-----------------------	--	--	--	--	--	--

Фонды оценочных средств

Примеры практических заданий

- 1) Выявление метонимичных упоминаний геолокаций.
- 2) Определение языка короткого текста.
- 3) Определение авторства текста.
- 4) Определение ключевых слов для научно-технической литературы.
- 5) Определение эмоциональной окраски текста.

Примерные вопросы для промежуточного контроля успеваемости

- 1) Постановка задачи распознавания именованных сущностей.
- 2) Области применения распознавания именованных сущностей.
- 3) Машинное обучение с учителем.
- 4) Метод опорных векторов.
- 5) Логистическая регрессия.
- 6) Оценка качества распознавания именованных сущностей.
- 7) Разметка последовательности.
- 8) Условные случайные поля.
- 9) Алгоритм Витерби.
- 10) Нейронные сети прямого распространения.
- 11) Алгоритм обратного распространения ошибки.
- 12) Нейронные сети прямого распространения. Скорость обучения.
- 13) Регуляризация нейронных сетей.
- 14) Рекуррентные нейронные сети.
- 15) Векторные представления слов. Дистрибутивная гипотеза. Локальные модели векторов слов: модели continuous skip-gram и continuous bag of words.
- 16) Векторные представления слов. Дистрибутивная гипотеза. Вычислительная сложность softmax. Иерархический softmax и negative sampling.
- 17) Векторные представления слов. Дистрибутивная гипотеза. Матрица совместной встречаемости слов. Глобальные модели векторов слов: модель GloVe.

- 18) Векторные представления слов. Проблема редких слов в моделях векторных представлений слов. Мера близости слов. Способы оценки качества векторов слов.
- 19) Необходимость применения символьного состава слов в методах обработки текстов.
- 20) Мешок символьных последовательностей.
- 21) Парное кодирование байтов.
- 22) Символьные представления слов на базе рекуррентных нейронных сетей.
- 23) Символьные представления слов на базе классических сверточных нейронных сетей.
- 24) Символьные представления слов на базе сверточных нейронных сетей с расширениями.
- 25) Сегментация текста: задачи токенизации и определения границ предложений.
- 26) Задача определения языка текста. Профили языка. Оценка качества классификации. Проблема коротких и мультиязычных текстов.
- 27) Задача определения языка текста. Naïve Bayes классификатор. Оценка качества классификации. Проблема коротких и мультиязычных текстов.
- 28) Задача определения частей речи и морфологического анализа. Разметка последовательности. Multi-label классификация.
- 29) Задача нормализации слов. Ripple Down Rules и LemmaGen. Грамматическая омонимия и способы борьбы с ней.
- 30) Задача нормализации слов. Нормализация слов как задача классификации. Грамматическая омонимия и способы борьбы с ней.
- 31) Синтаксический анализ. Дерево составляющих. Понятие формальной грамматики, иерархия Хомского. Генерация текста по формальной грамматике.
- 32) Синтаксический анализ. Дерево составляющих. Разбор предложения по грамматике составляющих. Алгоритм Кока-Янгера-Касами. Многозначность разбора, выбор лучшего разбора.
- 33) Синтаксический анализ. Дерево зависимостей. Разбор предложения на основе переходов: Arc-standard, Arc-eager.
- 34) Синтаксический анализ. Дерево зависимостей. Проективность деревьев разбора. Методы разбора для непроективных деревьев: Attardi's system, Online reordering.
- 35) Синтаксический анализ. Дерево зависимостей. Stack LSTM. Синтаксический разбор на основе Stack LSTM.
- 36) Синтаксический анализ. Дерево составляющих и дерево зависимостей. Оценка качества построенных деревьев разбора.
- 37) Машинный перевод. Модель зашумленного канала.
- 38) Статистический машинный перевод. Языковая модель.
- 39) Статистический машинный перевод. Модель перевода.
- 40) Статистический машинный перевод. Выравнивания. IBM Model 1.
- 41) Статистический машинный перевод. Декодирование.
- 42) Нейросетевой машинный перевод. Архитектура кодировщик-декодировщик.
- 43) Нейросетевой машинный перевод. Методы кодирования.
- 44) Нейросетевой машинный перевод. Методы декодирования.
- 45) Нейросетевой машинный перевод. Оценка качества.
- 46) Корреляция и анафоричность.

- 47) Разрешение кореферентности. Модель "пара упоминаний".
- 48) Разрешение кореферентности. Классификатор пар упоминаний.
- 49) Разрешение кореферентности. Отбор пар упоминаний.
- 50) Разрешение кореферентности. Объединение пар упоминаний в кореферентные цепочки.
- 51) Разрешение кореферентности. Модель "сущность - упоминание".
- 52) Разрешение кореферентности. Модели на основе ранжирования.
- 53) Разрешение кореферентности. Поиск упоминаний в тексте.
- 54) Разрешение кореферентности. Оценка качества.
- 55) Варианты постановки задачи извлечения мнений: анализ тональности текста, извлечение мнений. Методы оценки качества.
- 56) Задача автоматического реферирования. Варианты постановок задач.
- 57) Задача автоматического реферирования. Экстрактивный и абстрактивный реферат. Статистические методы решения задачи.
- 58) Задача автоматического реферирования. Экстрактивный и абстрактивный реферат. Графовые методы решения задачи.
- 59) Задача автоматического реферирования. Оценка качества построенных рефератов. Ручные, автоматизированные и автоматические методы оценки качества.
- 60) Диалоговые системы. Диалоги с определенной целью. Общая схема решения.
- 61) Диалоговые системы. Диалоги без цели. Идеи подходов к решению задачи.
- 62) Постановки задачи извлечения отношений.
- 63) Методы извлечения отношений на основе синтаксической структуры.
- 64) Методы извлечения отношений без синтаксической структуры.
- 65) Методы извлечения сложных отношений. Механизм внимания
- 66) Методы извлечения сложных отношений. Обобщение рекуррентных нейронных сетей на случай графа.
- 67) Базы знаний.
- 68) Привязка к базам знаний. Использование локального контекста.
- 69) Привязка к базам знаний. Использование глобального контекста.
- 70) Привязка к базам знаний. Поиск упоминаний в тексте.
- 71) Привязка к базам знаний. Оценка качества.
- 72) Методы переноса знаний на основе построения векторных представлений слов, учитывающих контекст словоупотребления.
- 73) Совместное обучение. Методы совместного обучения на основе нейронных сетей.

Методические материалы для проведения процедур оценивания результатов обучения

Для промежуточной аттестации предусмотрено выполнение двух практических заданий (примеры заданий приведены выше). Каждое задание оценивается от 1 до 2.5 баллов (задание, получающее оценку менее 1 балла, отправляется на доработку). Учащийся, успешно сдавший оба практических задания получает по их итогам "предварительную" экзаменационную оценку. Учащиеся, получившие оценку "отлично", освобождаются от устного экзамена (за который им выставляется "отлично"); учащиеся, получившие оценку "хорошо" имеют

право либо засчитать эту оценку как экзаменационную, либо сдавать устный экзамен. Остальные учащиеся получают оценку на устном экзамене.