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

УТВЕРЖДАЮ
Руководитель ООП
С.М. Дудаков
2023 г.

Рабочая программа дисциплины (с аннотацией)

Основы компьютерной лингвистики

Направление подготовки
03.02 — Прикладная математика и информатика

Профиль подготовки
Искусственый интеллект и анализ данных

для студентов 4 курса
Форма обучения — очная

Составитель(и):
- к.ф.-м.н. доц. Карлов Б.Н.

Тверь — 2023

I. Аннотация

1. Цель и задачи дисциплины:

Цель курса — ознакомить студентов с основными понятиями компьютерной лингвистики, с различными способами задания языков, с возможностью применения ЭВМ для обработки естественных языков.

2. Место дисциплины в структуре ООП

Дисциплина входит в раздел «Элективные дисциплины» части, формируемой участниками образовательных отношений, блока 1.

Предварительные знания и навыки. Знание курсов «Математическая логика и теория алгоритмов», «Теория автоматов и формальных языков», «Теория вероятностей и математическая статистика».

Дальнейшее использование. Полученные знания используются для итоговой государственной аттестации, прохождении практики, а также в дальнейшей трудовой деятельности выпускников.

3. Объем дисциплины: 3 зачетных единицы, 108 академических часа, **в том числе:**

контактная аудиторная работа: лекции 30 часов, практические занятия 30 часов;
контактная внеаудиторная работа: контроль самостоятельной работы 0 часов, в том числе курсовая работа 0 часов;
самостоятельная работа: 48 часов, в том числе контроль 0 часов.

4. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
ПК-3, Способен осуществлять концептуальное моделирование проблемной области и проводить формализацию представления знаний в системах искусственного интеллекта	ПК-3.1, Разрабатывает концептуальную модель проблемной области системы искусственного интеллекта
ПК-5, Способен использовать инструментальные средства для решения задач машинного обучения	ПК-5.1, Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи

ПК-7 Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК-7.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях

5. Форма промежуточной аттестации и семестр прохождения: зачёт в 7 семестре.

6. Язык преподавания: русский.

II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)					Самостоятельная работа, в том числе Контроль (час.)
		Лекции		Практические занятия		Контроль сам. раб., в т.ч. курсовая работа	
		Всего	в т.ч. практическая подготовка	Всего	в т.ч. практическая		
1	2	3	4	5	6	7	8
Лексический анализ	24	6		6/0			12
Статистические методы обработки языка	28	8		8/0			12
Синтаксический анализ	28	8		8/0			12
Применение нейронных сетей для обработки языка	28	8		8/0			12
Итого	108	30	0	30/0	0	0	48

Учебная программа дисциплины

1. Лексический анализ.
   1) Конечные автоматы и конечные преобразователи.
   2) Алгоритм моделирования недетерминированного конечного автомата.
   3) Префиксные деревья. Представление множества слов в виде префиксного дерева.
   4) Применение конечных преобразователей для морфологического анализа слов.

5) Алгоритм Портера.
6) Исправление орфографических ошибок. Редакционное расстояние. Алгоритм Вагнера-Фишера.

2. Статистические методы обработки языка.
   1) Условные вероятности. Формула Байеса.
   2) N-граммы. Сглаживание Лапласа, Уиттена-Белла, Гуда-Тьюринга.
   3) Применение N-грамм для исправления орфографических ошибок с учётом контекста.
   4) Цепи Маркова, скрытые марковские модели. Алгоритм Витерби.
   5) Определение частей речи на основе правил и на основе скрытых марковских моделей.
   6) Модели word2vec и GloVe.
   7) Оценка качества моделей. Энтропия, перекрёстная энтропия.
3. Синтаксический анализ.
   1) Порождающие грамматики. Иерархия Хомского.
   2) Контекстно-зависимые и контекстно-свободные грамматики. Деревья вывода.
   3) Эквивалентность контекстно-зависимых грамматик и линейно-ограниченных автоматов.
   4) Классические категориальные грамматики. Эквивалентность КС-грамматик и классических категориальных грамматик.
   5) Системы составляющих и деревья зависимостей. Связь деревьев зависимостей и систем составляющих.
   6) Синтаксический анализ на основе КС-грамматик. Алгоритмы Кока-Янгера-Касами и Эрли.
   7) Слабо-контекстные грамматики. Множественные контекстно-свободные грамматики.
   8) Головные грамматики, линейные индексные грамматики, комбинаторные категориальные грамматики, ТАГ-грамматики.
   9) Категориальные грамматики зависимостей (КГЗ). Алгоритм анализа КГЗ.
   10) Основы λ-исчисления. λ-термы, β-редукция, нормальная форма. Формулировка теоремы Чёрча-Россера.
   11) Применения комбинаторных категориальных грамматик для представления семантики в виде λ-термов.
4. Применение нейронных сетей для обработки языка.
   1) Искусственные нейроны. Искусственные нейронные сети. Функции активации.
   2) Метод обратного распространения ошибки.
   3) Рекуррентные нейронные сети.
   4) Архитектура LSTM.
   5) Использование нейронных сетей в компьютерной лингвистике: модели языка, машинный перевод.

III. Образовательные технологии

Учебная программа – наименование разделов и тем	Вид занятия	Образовательные технологии
Лексический анализ	лекции, практические занятия	изложение теоретического материала, решение задач
Статистические методы обработки языка	лекции, практические занятия	изложение теоретического материала, решение задач
Синтаксический анализ	лекции, практические занятия	изложение теоретического материала, решение задач
Применение нейронных сетей для обработки языка	лекции, практические занятия	изложение теоретического материала, решение задач

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

Типовые контрольные задания и/или критерии для проверки индикатора ПК-3.1

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Знать понятия порождающей грамматики, системы составляющих, дерева зависимостей, связь систем составляющих и деревьев зависимостей с грамматиками	Примеры вопросов к зачёту: • Порождающие грамматики. Иерархия Хомского. • Контекстно-зависимые и контекстно-свободные грамматики. Деревья вывода. • Эквивалентность контекстно-зависимых грамматик и линейно-ограниченных автоматов. • Классические категориальные грамматики. Эквивалентность КС-грамматик и классических категориальных грамматик. • Системы составляющих и деревья зависимостей. Связь деревьев зависимостей и систем составляющих. • Синтаксический анализ на основе КС-грамматик. Алгоритмы Кока-Янгера-Касами и Эрли. • Слабо-контекстные грамматики.	оценка 3 — знает определения основных понятий, оценка 4 — кроме того знает основные свойства порождающих грамматик, систем составляющих и деревьев зависимостей, оценка 5 — кроме того знает алгоритмы анализа

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	Множественные контекстно-свободные грамматики. • Головные грамматики, линейные индексные грамматики, комбинаторные категориальные грамматики, ТАГ-грамматики. • Категориальные грамматики зависимостей (КГЗ). Алгоритм анализа КГЗ. • Основы λ-исчисления. λ-термы, β-редукция, нормальная форма. Формулировка теоремы Чёрча-Россера. • Применения комбинаторных категориальных грамматик для представления семантики в виде λ-термов.	
Уметь строить различные грамматики и автоматы по описанию языка	Примеры задач для контрольных работ: • Постройте детерминированный конечный преобразователь для выполнения следующей операции. На вход подаётся слово $w\$$ в алфавите $\{a,b,c,\$\}$, символ $\$$ помечает конец входного слова. Требуется вставить символ c после каждого блока символов a нечётной длины. • Постройте ТАГ-грамматику для языка $L = \{ a^i b^j c^k : 0 < i < j < k \}$. • Постройте комбинаторную категориальную грамматику для языка $L = \{ w * w^{-1} * w : w \in \{ 0, 1 \}^* \}$. • Постройте контекстно-свободную грамматику для языка $L = \{ a^i b^j c^k d^l : i < k$ или $j \geqslant l \}$. • Постройте классическую категориальную грамматику для языка $L = \{ a^i b^j c^k : i, j, k \geqslant 0, i = j$ или $j = k \}$. • Постройте категориальную грамматику зависимостей для языка	оценка 3 — умеет строить конечные автоматы и конечные преобразователи по описанию языка или отношения, оценка 4 — кроме того умеет строить контекстно-свободные и классические категориальные грамматики по описанию языка, оценка 5 — кроме того умеет слабо контекстные грамматики по описанию языка

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	$L = \{ w \in \{ a, b, c, d \}^+ : \lvert w\rvert_a + \lvert w\rvert_c \geqslant \lvert w\rvert_b + \lvert w\rvert_d$ и все символы a стоят правее c $\}$. Примеры тем для самостоятельной работы: • Напишите программу, реализующую алгоритм Портера для английского и русского языков. • Напишите программу, которая по контекстно-свободной грамматике и входному слову определяет, выводимо ли слово в грамматике. Реализуйте алгоритмы Кока-Янгера-Касами и Эрли. • Напишите программу, которая по классической категориальной грамматике и предложению на русском языке строит все его размеченные деревья зависимостей.	
Уметь строить формальное представление синтаксиса и семантики предложений на естественных языках	Примеры задач для контрольных работ: • Постройте размеченную иерархизованную систему составляющих для предложения «К востоку от боровых озёр лежат громадные мещёрские болота — мшары». • Постройте размеченное дерево зависимостей для предложения «Для точной диагностики заболеваний внутренних органов человека рентген незаменим». • Дано предложение «Мастер внимательно осматривал станок». Его словам сопоставлены следующие категории и λ-термы: – мастер — NP : M – внимательно — C : B – осматривал — (((S \ NP) \ C) / NP) :λxyz.Ozxy – станок — NP : C	оценка 3 — умеет строить системы составляющих и деревья зависимостей для предложений, оценка 4 — кроме того умеет строить размеченные иерархизованные системы составляющих и размеченные деревья зависимостей, оценка 5 — кроме того умеет представлять семантику в виде λ-термов

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	Сократите категории до S и упростите получающийся λ-терм. На первом шаге примените к первой категории правило (> T), а на втором шаге примените к третьей и четвертой категориям правило (>).	

Типовые контрольные задания и/или критерии для проверки индикатора ПК-5.1

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Знать основные методы лексического анализа текста	Примеры вопросов к зачёту: • Конечные автоматы и конечные преобразователи. • Алгоритм моделирования недетерминированного конечного автомата. • Префиксные деревья. Представление множества слов в виде префиксного дерева. • Применение конечных преобразователей для морфологического анализа слов. • Алгоритм Портера. • Исправление орфографических ошибок. Редакционное расстояние. Алгоритм Вагнера-Фишера.	оценка 3 — знает понятия конечного автомата и конечного преобразователя, оценка 4 — кроме того знает основные методы морфологического анализа слов, оценка 5 — кроме того знает алгоритмы исправления орфографических ошибок
Уметь использовать статистические методы обработки текста	Примеры задач для контрольных работ: • Найдите частоты биграмм для предложения «Ворон к ворону летит, ворон ворону кричит». Примените к получившимся частотам сглаживание Уиттена-Белла. Найдите вероятность предложения «Летит к ворону ворон». • Предположим, что в предложении написано слово «стор» с опечаткой.	оценка 3 — умеет использовать формулу Байеса, оценка 4 — кроме того умеет использовать N-граммы, оценка 5 — кроме того умеет выполнять сглаживание различными спосо-

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	Словарь содержит правильные слова «стар», «стог», «стон», «сток», «стоп», «тор», «сто», «сор», «створ». Пусть частоты этих слов в корпусе равны 0,2%, 0,1%, 0,3%, 0,5%, 0,1%, 1%, 0,2%, 0,3%, 0,3%. Пусть вероятность замены символа равна 1%, удаления – 2%, вставки – 3%. Найдите наиболее вероятное правильное написание слова. Примеры тем для самостоятельной работы: • Напишите программу, которая по корпусу русских текстов находит частоты N-грамм с использованием сглаживания Гуда-Тьюринга и после этого по заданному предложению вычисляет его вероятность.	бами

Типовые контрольные задания и/или критерии для проверки индикатора ПК-7.1

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
Знать основные статистические методы анализа текста	Примеры вопросов к зачёту: • Условные вероятности. Формула Байеса. • N-граммы. Сглаживание Лапласа, Уиттена-Белла, Гуда-Тьюринга. • Применение N-грамм для исправления орфографических ошибок с учётом контекста. • Цепи Маркова, скрытые марковские модели. Алгоритм Витерби. • Определение частей речи на основе правил и на основе скрытых марковских моделей. • Модели word2vec и GloVe.	оценка 3 — знает понятие N-граммы, сглаживания, скрытой марковской модели, моделей word2vec и GloVe, оценка 4 — кроме того знает алгоритмы, использующие эти модели, оценка 5 — кроме того знает методы оценки качества

Требования к обучающемуся	Типовые контрольные задания для оценки знаний, умений, навыков	Показатели и критерии оценивания, шкала оценивания
	• Оценка качества моделей. Энтропия, перекрёстная энтропия.	моделей
Знать основные применения нейронных сетей для обработки текстов	Примеры вопросов к зачёту: • Искусственные нейроны. Искусственные нейронные сети. Функции активации. • Метод обратного распространения ошибки. • Рекуррентные нейронные сети. • Архитектура LSTM. • Использование нейронных сетей в компьютерной лингвистике: модели языка, машинный перевод.	оценка 3 — знает простейшие архитектуры нейронных сетей и их применения для обработки текстов, оценка 4 — кроме того знает основные методы обучения, оценка 5 — кроме того знает более сложные архитектуры

V. Учебно-методическое и информационное обеспечение дисциплины

1. Рекомендованная литература

а) Основная литература

[1] Волосатова, Т.М. Информатика и лингвистика [Электронный ресурс]: Учебное пособие/Волосатова Т.М., Чичварин Н.В. — Электрон. дан. — М.: НИЦ ИНФРА-М, 2016. — 196 с.: 60х90 1/16. — (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977-0 — Режим доступа: https://znanium.com/catalog/document?id=422587

[2] Марченков, С.С. Конечные автоматы [Электронный ресурс]: учеб. пособие — Электрон. дан. — Москва: Физматлит, 2008. — 56 с. — Режим доступа: https://e.lanbook.com/book/59510. — Загл. с экрана.

[3] Короткова, М.А. Задачник по курсу "Математическая лингвистика и теория автоматов": учебное пособие для вузов [Электронный ресурс]: учеб. пособие / М.А. Короткова, Е.Е. Трифонова. — Электрон. дан. — Москва: НИЯУ МИФИ, 2012. — 92 с. — Режим доступа: https://e.lanbook.com/book/75843. — Загл. с экрана.

б) Дополнительная литература

[4] Федосеева, Л.И. Основы теории конечных автоматов и формальных языков [Электронный ресурс]: учеб. пособие / Л.И. Федосеева, Р.М. Адилов, М.Н. Шмокин. — Электрон. дан. — Пенза: ПензГТУ, 2013. — 136 с. — Ре-

жим доступа: https://e.lanbook.com/book/62703. — Загл. с экрана.

2. Программное обеспечение

Наименование помещений	Программное обеспечение
Ауд. 201а (компьютерная лаборатория ПМиК) (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Перечень программного обеспечения (со свободными лицензиями): Linux Kubuntu, KDE, TeXLive, TeXStudio, LibreOffice, GIMP, Gwenview, ImageMagick, Okular, Skanlite, Google Chrome, KDE Connect, Konversation, KRDC, KTorrent, Thunderbird, Elisa, VLC media player, PulseAudio, KAppTemplate, KDevelop, pgAdmin4, PostgreSQL, Qt, QtCreator, R, RStudio, Visual Studio Code, Perl, Python, Ruby, clang, clang++, gcc, g++, nasm, flex, bison, Maxima, Octave, Dolphin, HTop, Konsole, KSystemLog, Xterm, Ark, Kate, Kcalc, Krusader, Spectacle, Vim.

3. Современные профессиональные базы данных и информационные справочные системы

[1] ЭБС «ZNANIUM.COM» http://www.znanium.com
[2] ЭБС «Университетская библиотека онлайн» https://biblioclub.ru
[3] ЭБС IPRbooks http://www.iprbookshop.ru
[4] ЭБС «Лань» http://e.lanbook.com
[5] ЭБС «Юрайт» https://urait.ru
[6] ЭБС ТвГУ http://megapro.tversu.ru/megapro/Web
[7] Научная электронная библиотека eLIBRARY.RU (подписка на журналы) https://elibrary.ru/projects/subscription/rus_titles_open.asp
[8] Репозитарий ТвГУ http://eprints.tversu.ru

4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

[1] Natural Language Processing, http://www.learnerstv.org/Free-Computer-Science-Video-lectures-ltv676-Page1.htm
[2] Natural Language Processing with Deep Learning, http://web.stanford.edu/class/cs224n/
[3] Московский центр непрерывного математического образования, http://www.mccme.ru/

VI. Методические материалы для обучающихся по освоению дисциплины

Важной составляющей данного раздела РПД являются требования к

рейтинг-контролю с указанием баллов, распределенных между модулями и видами работы обучающихся.

Максимальная сумма баллов по учебной дисциплине, заканчивающейся зачетом, по итогам семестра составляет 100 баллов (50 баллов - 1-й модуль и 50 баллов - 2-й модуль).

Студенту, набравшему 40 баллов и выше по итогам работы в семестре, в экзаменационной ведомости и зачетной книжке выставляется оценка «зачтено». Студент, набравший до 39 баллов включительно, сдает зачет.

Распределение баллов по модулям устанавливается преподавателем и может корректироваться.

Примеры задач для подготовки к контрольным работам

1. Постройте недетерминированный конечный автомат с 11 состояниями для языка
   $L = \{w \in \{a,b\}^* : |w| \geq 4$ и первые две буквы слова не равны последним двум$\}$
   и обоснуйте его правильность. С помощью алгоритма моделирования проверьте, какие из слов *abb*, *baaba*, *bbab*, *abbabba* распознаются этим автоматом.
2. С помощью алгоритма Вагнера-Фишера найдите редакционное расстояние между словами «алгоритм» и «логарифм».
3. Дана КС-грамматика с начальным нетерминалом E и следующими правилами:

   $$E \to TE'$$
   $$E' \to +TE' \mid \varepsilon$$
   $$T \to FT'$$
   $$T' \to *FT' \mid \varepsilon$$
   $$F \to c \mid (E)$$

   С помощью алгоритма Эрли определите, какие из следующих слов выводимы в этой грамматике, и постройте для них деревья вывода: (c+c)*c, c+c+c+c, ((c)), (c+c, c*c**c, c*c).
4. Постройте комбинаторную категориальную грамматику для языка
   $$L = \{w\#w^{-1} : w - \text{правильное скобочное слово}\}$$
   и обоснуйте ее правильность.
5. Постройте КЗ-грамматику для языка $L = \{a^n : n$ — составное число$\}$.
6. Постройте размеченную иерархизованную систему составляющих и размеченное дерево зависимостей для предложения «Тощая торговка вяленой воблой торчала среди ящиков».
7. Обобщенной КЗ-грамматикой (ОКЗ-грамматикой) называется порождающая грамматика $G = (N, \Sigma, P, S)$, правила которой имеют вид $\xi A \eta \to \xi \alpha \eta$, где ξ, η, $\alpha \in (\Sigma \cup N)^*$, $A \in N$. Докажите, что любой рекурсивно перечислимый язык порождается некоторой ОКЗ-грамматикой.
8. Докажите, что класс языков типа 0 замкнут относительно тасовки:
   $TAC(L_1, L_2) = \{x_1 y_1 \ldots x_n y_n : x_i, y_j \in \Sigma^*, x_1 \ldots x_n = x, y_1 \ldots y_n = y, x \in L_1, y \in L_2\}$.

Требования к рейтинг контролю (7 семестр)

Контрольная работа 1. Темы: конечные преобразователи, статистические методы обработки текста, КЗ-грамматики. Пример задания:

1. Постройте детерминированный конечный преобразователь для выполнения следующей операции. На вход подаётся слово w\$ в алфавите $\{a,b,c,\$\}$, символ \$ помечает конец входного слова. Требуется удалить те символы a, которые стоят между двумя символами b. Например, из слова *acbabaabcab*\$ должно получиться *acbbaabcab*\$.
2. Найдите частоты биграмм для предложения «Без устали, без устали смотрю, смотрю в окно». Примените к получившимся частотам сглаживание Уиттена-Белла. Найдите вероятность предложения «Смотрю в окно без устали».
3. Постройте неукорачивающую грамматику, порождающую следующий язык:

$$L = \{a^n b^m a^n b^m : m, n > 0\}.$$

За решение каждой задачи выставляется максимум 10 баллов.

Самостоятельная работа 1. Темы: конечные преобразователи, статистические методы обработки текста, КЗ-грамматики. Пример задания:
Напишите программу, которая исправляет орфографические ошибки в отдельных словах с использованием формулы Байеса.
За решение задачи выставляется максимум 10 баллов.

Контрольная работа 2. Темы: слабо контекстные грамматики, категориальные грамматики, системы составляющих, деревья зависимостей. Пример задания:

1. Постройте ТАГ-грамматику для языка

$$L = \{ a^i b^j c^{i+2} : j > i > 0 \}$$

2. Постройте категориальную грамматику зависимостей для языка

$$L = \{ a^i b^{2i} w : i > 0, w \in \{ c, d \}^*, |w|_c < |w|_a \}$$

3. Постройте 3-множественную КС-грамматику для языка

$$L = \{www^{-1}w^{-1} : w \in \{ a, b \}^* \}$$

4. Постройте размеченную иерархизованную систему составляющих и размеченное дерево зависимостей для предложения «К востоку от боровых озёр лежат громадные мещёрские болота — мшары».
5. Для следующего предложения сократите в указанном порядке категории до S и приведите получающийся λ-терм к нормальной форме.

Мастер	внимательно	осматривал	станок
NP : M	C : B	(((S \ NP) \ C) / NP) : $\lambda xyz.Ozxy$	NP : C
(>**T**)		(>)	

За решение каждой задачи выставляется максимум 10 баллов.

Самостоятельная работа 2. Темы: слабо контекстные грамматики, системы

составляющих, деревья зависимостей. Пример задания:
Напишите программу, которая по системе составляющих строит согласованное с ней дерево зависимостей.
За решение задачи выставляется максимум 10 баллов.

Общая сумма. В сумме за все задачи выставляет не более 100 баллов.

Вопросы к зачёту

1. Лексический анализ.
   1) Конечные автоматы и конечные преобразователи.
   2) Алгоритм моделирования недетерминированного конечного автомата.
   3) Префиксные деревья. Представление множества слов в виде префиксного дерева.
   4) Применение конечных преобразователей для морфологического анализа слов.
   5) Алгоритм Портера.
   6) Исправление орфографических ошибок. Редакционное расстояние. Алгоритм Вагнера-Фишера.
2. Статистические методы обработки языка.
   1) Условные вероятности. Формула Байеса.
   2) N-граммы. Сглаживание Лапласа, Уиттена-Белла, Гуда-Тьюринга.
   3) Применение N-грамм для исправления орфографических ошибок с учётом контекста.
   4) Цепи Маркова, скрытые марковские модели. Алгоритм Витерби.
   5) Определение частей речи на основе правил и на основе скрытых марковских моделей.
   6) Модели word2vec и GloVe.
   7) Оценка качества моделей. Энтропия, перекрёстная энтропия.
3. Синтаксический анализ.
   1) Порождающие грамматики. Иерархия Хомского.
   2) Контекстно-зависимые и контекстно-свободные грамматики. Деревья вывода.
   3) Эквивалентность контекстно-зависимых грамматик и линейно-ограниченных автоматов.
   4) Классические категориальные грамматики. Эквивалентность КС-грамматик и классических категориальных грамматик.
   5) Системы составляющих и деревья зависимостей. Связь деревьев зависимостей и систем составляющих.
   6) Синтаксический анализ на основе КС-грамматик. Алгоритмы Кока-Янгера-Касами и Эрли.
   7) Слабо-контекстные грамматики. Множественные контекстно-свободные грамматики.
   8) Головные грамматики, линейные индексные грамматики, комбинаторные категориальные грамматики, ТАГ-грамматики.
   9) Категориальные грамматики зависимостей (КГЗ). Алгоритм анализа

КГЗ.

10) Основы λ-исчисления. λ-термы, β-редукция, нормальная форма. Формулировка теоремы Чёрча-Россера.
11) Применения комбинаторных категориальных грамматик для представления семантики в виде λ-термов.

4. Применение нейронных сетей для обработки языка.
   1) Искусственные нейроны. Искусственные нейронные сети. Функции активации.
   2) Метод обратного распространения ошибки.
   3) Рекуррентные нейронные сети.
   4) Архитектура LSTM.
   5) Использование нейронных сетей в компьютерной лингвистике: модели языка, машинный перевод.

VII. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Для аудиторной работы

Наименование помещений	Материально-техническое оснащение помещений
Ауд. 308 (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Набор учебной мебели, экран проектор.

Для самостоятельной работы

Наименование помещений	Материально-техническое оснащение помещений
Ауд. 201а (компьютерная лаборатория ПМиК) (170002, Тверская обл., г. Тверь, пер. Садовый, д. 35)	Набор учебной мебели, доска маркерная, компьютер, сервер (системный блок), концентратор сетевой.

VIII. Сведения об обновлении рабочей программы дисциплины

№ п/п	Обновленный раздел рабочей программы дисциплины	Описание внесённых изменений	Дата и протокол заседания кафедры, утвердившего изменения
1	11. 2) Программное обеспечение	Внесены изменения в список ПО	От 24.08.2023 года, протокол

			№ 1 ученого совета факультета
2	V. 1) Рекомендуемая литература	Обновление ссылок на литературу	От 24.08.2023 года, протокол № 1 ученого совета факультета
3			
4			