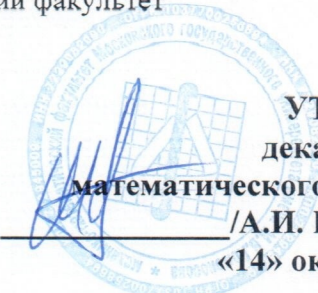


Федеральное государственное бюджетное образовательное учреждение высшего
профессионального образования
Московский государственный университет имени М.В. Ломоносова
Механико-математический факультет

 **УТВЕРЖДАЮ**
декан механико-
математического факультета
/А.И. Шафаревич /
«14» октября 2021г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

Машинное обучение и нейронные сети в компьютерной лингвистике

Уровень высшего образования:
магистратура

Направление подготовки / специальность:
02.04.01 "Математика и компьютерные науки" (3++)

Направленность (профиль) ОПОП:
Цифровые технологии и искусственный интеллект

Форма обучения:
очная

Рабочая программа рассмотрена и утверждена
на заседании Ученого совета Механико-математического факультета
(протокол №7, от 14 октября 2021 года)

Москва 2021

Рабочая программа дисциплины (модуля) разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 02.04.01 "Математика и компьютерные науки" утвержденного Приказом Министерства образования и науки РФ от 10 января 2018 г. N 13.

1. Место дисциплины (модуля) в структуре ОПОП ВО:

дисциплина относится к блоку профессиональной подготовки вариативной части ОПОП ВО.

2. Входные требования для освоения дисциплины (модуля), предварительные условия (если есть):

Для того чтобы изучение дисциплины было возможно, обучающийся должен обладать следующими компетенциями:

Знать: основные понятия, концепции, результаты и методы математической логики, машинного обучения.

Уметь: решать стандартные задачи математической логики.

3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников.

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ПК-7. Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях	ПК-7.3. Руководит проектами в области сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений»	ПК-7.3. 3-1. Знает фундаментальные правила построения рекомендательных систем и систем поддержки принятия решений, основанных на интеллектуальных принципах, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений» ПК-7.3. У-1. Умеет руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений»

4. Объем дисциплины (модуля) составляет 3 з.е., в том числе 32 академических часов, отведенных на контактную работу обучающихся с преподавателем, 76 академических часов на самостоятельную работу обучающихся.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий:

5.1. Структура дисциплины (модуля) по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий (в строгом соответствии с учебным планом)

Наименование разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Номинальные трудозатраты обучающегося		Самостоятельная работа обучающегося, академические часы	Всего академических часов	Форма текущего контроля успеваемости* (наименование)
	Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, академические часы				
	Занятия лекционного типа	Занятия семинарского типа			
Тема 1 Понятие семантического анализа текста	2		4	6	—
Тема 2 Основные методы семантического анализа	22		52	74	опрос
Тема 3 Семантический анализ нормативных актов, осуществляемый на основе правил	8		20	28	опрос
	—	—			—
Промежуточная аттестация (экзамен)					
Итого	32		76	108	—

5.2. Содержание разделов (тем) дисциплины

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
1.	Тема 1 Понятие семантического анализа текста.	Этапы семантического анализа. Лексический анализ текста. Морфологический анализ слов предложения. Синтаксический анализ предложения. Семантический анализ текста.

2.	Тема 2 Основные методы семантического анализа	Синтаксический анализ на основе машинного обучения. Синтаксический анализ с использованием промежуточных состояний. Индуктивный анализ синтаксических зависимостей. Пример модели нейронной сети для синтаксического анализа. Синтаксически размеченные корпуса (Universal Dependencies, Национальный корпус русского языка). Синтаксический анализ на основе правил. Теория лингвистических моделей «Смысл $\Leftrightarrow$ Текст». Синтаксический анализ в системе АОТ. Поверхностно-семантический анализ в системе АОТ. Синтаксический и семантический анализ в компьютерном решателе математических задач А.С. Подколзина. Синтаксический и семантический анализ, использующий как правила, так и машинное обучение.
3.	Тема 3 Семантический анализ нормативных актов, осуществляемый на основе правил	Общий принцип работы семантического анализа нормативных актов, осуществляемого на основе правил. Построение связного синтаксического графа предложения. Упрощение синтаксического графа. Отождествление сущностей, упомянутых в тексте нормативного акта. Построение логической формулы на основе упрощенного синтаксического графа предложения. Построение модели, вычисляющей значения сущностей, упомянутых в тексте нормативного акта.

6. Фонд оценочных средств (ФОС, оценочные и методические материалы) для оценивания результатов обучения по дисциплине (модулю).

6.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости, критерии и шкалы оценивания (в отсутствие утвержденных соответствующих локальных нормативных актов на факультете)

Контрольные вопросы для самостоятельной работы

1. Выполните морфологический анализ слов предложения «Буря мглою небо кроет, вихри снежные крутя».
2. Постройте дерево зависимостей для предложения «Ровно в девятнадцать ноль-ноль тридцать первого декабря прошлого года Андрей Т. лежал в постели и с покорной горечью размышлял о прошлом, настоящем и будущем».
3. Чем отличается синтаксический анализ от семантического?
4. Приведите пример схемы нейронной сети для синтаксического анализа предложения.
5. В чем преимущества и недостатки синтаксического анализа на основе правил по сравнению с синтаксическим анализом с помощью машинного обучения?
6. Осуществите синтаксическую разметку предложения «Яков Шенцер уже дожидался меня за столиком под четырехцветным полотняным тентом на каменной террасе старого дома» в формате Universal Dependencies.
7. Чем отличаются поверхностно-синтаксические и глубинно-синтаксические отношения для словосочетания «деревянный стол» в модели «Смысл $\Leftrightarrow$ Текст»?
8. Перечислите поверхностно-синтаксические отношения (согласно модели «Смысл $\Leftrightarrow$ Текст») в предложении «Через пятнадцать минут друзья вышли из дому».
9. Проведите синтаксический анализ предложения «На этом семинаре Сьюзен не выступала» согласно правилам системы AOT.
10. Перечислите семантические отношения (согласно системе AOT) в предложении «Он дал сыну два яблока.»
11. Получите результат семантического анализа предложения «Роботы терпеливо ждали в своих деревянных кабинах» в компьютерном решателе математических задач А.С. Подколзина.
12. Приведите пример правила для синтаксического анализа нормативных актов.
13. Постройте вручную синтаксический граф предложения «Настоящее Положение устанавливает правила формирования в бухгалтерском учете информации об основных средствах организации» так, как это предполагается делать по алгоритму семантического анализа нормативных актов, осуществляемого на основе правил.
14. Постройте вручную и затем упростите синтаксический граф предложения «Срок полезного использования объекта основных средств определяется организацией при принятии объекта к бухгалтерскому учету» так, как это предполагается делать по алгоритму семантического анализа нормативных актов, осуществляемого на основе правил.
15. Постройте логическую формулу для предложения «При способе уменьшаемого остатка годовая сумма амортизационных отчислений определяется исходя из остаточной стоимости объекта основных средств на начало отчетного года и нормы амортизации, исчисленной исходя из срока полезного использования этого объекта и коэффициента не выше 3, установленного организацией».

6.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации по дисциплине, критерии и шкалы оценивания

Программа экзамена:

1. Этапы семантического анализа текста.
2. Лексический анализ текста.
3. Морфологический анализ слов предложения.
4. Синтаксический анализ предложения.
5. Семантический анализ текста.
6. Синтаксический анализ на основе машинного обучения.
7. Синтаксический анализ с использованием промежуточных состояний.
8. Индуктивный анализ синтаксических зависимостей.
9. Пример модели нейронной сети для синтаксического анализа.
10. Синтаксический анализ на основе правил.
11. Теория лингвистических моделей «Смысл $\Leftrightarrow$ Текст».
12. Синтаксический анализ в системе АОТ.
13. Поверхностно-семантический анализ в системе АОТ.
14. Синтаксический и семантический анализ в компьютерном решателе математических задач А.С. Подколзина.
14. Синтаксический и семантический анализ, использующий как правила, так и машинное обучение.
16. Общий принцип работы семантического анализа нормативных актов, осуществляемого на основе правил.
17. Построение связного синтаксического графа предложения.
18. Упрощение синтаксического графа предложения.
19. Отождествление сущностей, упомянутых в тексте нормативного акта.
20. Построение логической формулы на основе упрощенного синтаксического графа предложения.
21. Построение модели, вычисляющей значения сущностей, упомянутых в тексте нормативного акта. .

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)

виды оценочных средств				
Знания (виды оценочных средств: опрос, тесты)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения (виды оценочных средств: практические задания)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) (виды оценочных средств: выполнение и защита курсовой работы, отчет по практике, отчет по НИР и т.п.)	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

7. Ресурсное обеспечение:

7.1. Перечень основной и дополнительной литературы

Перечень основной и дополнительной учебной литературы:

- 1) Подколзин А.С. Компьютерное моделирование логических процессов. – М.: Физматлит. 2008. – 1024 с. – ISBN 978-5-9221-1045-7.
- 2) Подколзин А.С. Компьютерное моделирование логических процессов. Том 6. Опыт обучения компьютерного решателя задач: понимание естественного языка и анализ рисунков. – депонировано в ВИНТИ РАН, 2019. - 758 с.
- 3) Мельчук И.А. Опыт теории лингвистических моделей «Смысл $\Leftrightarrow$ Текст». – М.: Школа «Языки русской культуры», 1999. - 346 с. – ISBN 5-7859-0078-5.
- 4) Кудрявцев В.Б., Гасанов Э.Э., Перпер Е.М. Автоматическая генерация компьютерной программы, моделирующей нормативно-правовой акт // Интеллектуальные системы. Теория и приложения (ранее: Интеллектуальные системы по 2014, № 2, ISSN 2075-9460).

– 2014. – Т. 18, вып. 2. – С. 133-156.

5) Перпер Е.М. О синтаксическом анализе юридических текстов // Интеллектуальные системы. Теория и приложения (ранее: Интеллектуальные системы по 2014, № 2, ISSN 2075-9460). – 2016. – Т. 20, вып. 2. – С. 31-49.

6) Chen D., Manning C.D. A Fast and Accurate Dependency Parser using Neural Networks // Proc. of EMNLP 2014. – 2014. – P. 740-750.

7) Hall J. MaltParser – An Architecture for Inductive Labeled Dependency Parsing. – 2006.

8) Anisimovich K.V., Druzhkin K.Ju., Minlos F.R., Petrova M.A., Selegey V.P., Zuev K.A. Syntactic and semantic parser based on ABBYY Compreno linguistic technologies // Международная конференция по компьютерной лингвистике «Диалог-2012». — 2012. — Vol. 2, P. 91—103.

7.2. Перечень лицензионного программного обеспечения, в том числе отечественного производства
нет

7.3. Перечень профессиональных баз данных и информационных справочных систем
нет

7.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <http://aot.ru/>

2. <https://universaldependencies.org/>

3. <https://ruscorpora.ru/>

7.5. Описание материально-технического обеспечения.

Аудитории для проведения лекционных занятий.

8. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.

9. Разработчик (разработчики) программы.

д.ф.-м.н., проф. Э.Э.Гасанов., к.ф.-м.н. Е.М.Перпер.