Федеральное государственное бюджетное образовательное учреждение
высшего образования

Московский государственный университет имени М. В. Ломоносова

Механико-математический факультет

|  |  |
| --- | --- |
|  | УТВЕРЖДАЮДекан механико-математического факультета МГУ\_\_\_\_\_\_\_\_\_\_\_\_\_\_/А. И. Шафаревич /«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г. |
|  | М.П. |

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

|  |  |
| --- | --- |
| Наименование дисциплины (модуля): | Введение в компьютерную лингвистику |
| Уровень высшего образования: | Бакалавриат, магистратура, специалитет |
| Направление подготовки / специальность: | Межфакультетский, по выбору студента |
| Направленность профиль)/специализация | Междисциплинарный общеобразовательный |
| Форма обучения: | Очная |
| Язык преподавания: | Русский |
| Автор программы: | Сорокин Алексей Андреевич |

Рабочая программа рассмотрена и одобрена

На заседании кафедры математической логики и теории алгоритмов

(протокол № 9 от 07.09.2022)

Москва 2022

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки/ специальности для студентов всех факультетов МГУ в соответствии с приказом № 43 от 13 февраля 2013 г.

**Содержимое**

[1. Место дисциплины (модуля) в структуре ОПОП ВО 3](#_Toc76748145)

[2. Объем дисциплины (модуля) 3](#_Toc76748146)

[3. Формат обучения 3](#_Toc76748147)

[4. Входные требования для освоения дисциплины (модуля) 3](#_Toc76748148)

[5. Результаты обучения по дисциплине (модулю) 3](#_Toc76748149)

[6. Содержание дисциплины (модуля) 4](#_Toc76748150)

[6.1. Структура дисциплины (модуля) 4](#_Toc76748151)

[7. Ресурсное обеспечение 4](#_Toc76748152)

[7.1. Список основной литературы 4](#_Toc76748153)

[7.2. Список дополнительной литературы (при наличии) 4](#_Toc76748154)

[7.3. Список программного обеспечения 4](#_Toc76748155)

[7.4. Список баз данных и информационных справочных систем 5](#_Toc76748156)

[7.5. Список ресурсов сети «Интернет» 5](#_Toc76748157)

[7.6. Материально-техническое обеспечение 5](#_Toc76748158)

[8. Фонд оценочных средств 5](#_Toc76748159)

[8.1. Текущий контроль успеваемости 5](#_Toc76748160)

[8.2. Промежуточная аттестация 5](#_Toc76748161)

1. Место дисциплины (модуля) в структуре ОПОП ВО

Является дисциплиной по выбору, избираемой в обязательном порядке.

1. Объем дисциплины (модуля)

Объем дисциплины (модуля) составляет 1 з.е., 36 академических часов, в том числе 24 академических часа, отведенных на контактную работу обучающихся с преподавателем и 12 академических часов на самостоятельную работу обучающихся.

1. Формат обучения

Дисциплина реализуется в очной форме с использованием электронного обучения

и дистанционных образовательных технологий (для выполнения домашних заданий).

1. Преподаватели

Дисциплину ведут ассистент кафедры математической логики и теории алгоритмов А. А. Сорокин и профессор кафедры математической логики и теории алгоритмов М. Р. Пентус.

1. Входные требования для освоения дисциплины (модуля)

Предварительные условия отсутствуют

1. Результаты обучения по дисциплине (модулю)

|  |  |  |
| --- | --- | --- |
| **Компетенции выпускников** | **Индикаторы достижения компетенций, реализуемые в настоящей дисциплине (модуле)[[1]](#footnote-1)** | **Планируемые результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций** |
| **Компетенция СПК-1**Знание идей и теоретических основ компьютерной лингвистики, умение применять её методы при решении различных задач. | **Индикатор СПК-1.1**Знает идеи и теоретические основы компьютерной лингвистики. | Знать основные понятия и простейшие алгоритмы компьютерной лингвистики. |
| **Индикатор СПК-1.2**Умеет применять простейшие методы компьютерной лингвистики. | Знать важнейшие модели компьютерной лингвистики.Уметь решать базовые задачи компьютерной лингвистики. |

1. Содержание дисциплины (модуля)

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**  |
|  | **Контактная работа (работа во взаимодействии с преподавателем)** **Виды контактной работы, ак. ч.** | Самостоятельная работа,Самостоятельная работа ак. ч. |
| Всего, ак. ч.  | Ауд., ак. ч. | Лекции\*, ак. ч. | Семинары\*, прак., ак. ч. |
| Введение.  | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Обзор методов и приложений. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Основы теории формальных языков, регулярные выражения. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Конечные автоматы и преобразователи. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Корпусная лингвистика. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Исправление опечаток, расстояние Левенштейна. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Введение в машинное обучение, линейные модели. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Введение в нейронные сети, их приложения в компьютерной лингвистике. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Автоматический перевод. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Большие языковые модели, их предобучение и дообучение. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Основные библиотеки для решения задач компьютерной лингвистики. | 3,0 | 2,0 | 2,0 |  | 1,0 |
| Промежуточная аттестация: зачет | 3,0 | 2,0 | 2,0 |  | 1,0 |
| **Итого** | **36** | **24** | **24** |  | **12** |

1. Ресурсное обеспечение
	1. Список основной литературы:

1. D. Jurafsky and J. Martin Speech and Language Processing (3 ed.) Доступно по ссылке https://web.stanford.edu/~jurafsky/slp3/

2. Прикладная и компьютерная лингвистика. Под ред. И. С. Николаева. URSS, 2017.

3. С. Хайкин. Нейронные сети: полный курс. 2-е изд. М.: Вильямс, 2019.

4. I. Goodfellow, J. Bengio, A. Courville Deep Learning https://www.deeplearningbook.org (электронный ресурс)

* 1. Список программного обеспечения

Использования не предполагается (при использовании ДОТ – Zoom или эквивалентная по функциям система). Возможно использование Excel или аналога и MathCad или аналога.

* 1. Список баз данных и информационных справочных систем

Общие библиотечные ресурсы

* 1. Список ресурсов сети «Интернет»

1. D. Jurafsky and J. Martin Speech and Language Processing <https://web.stanford.edu/~jurafsky/slp3/>

2. I. Goodfellow, J. Bengio, A. Courville Deep Learning [https://www.deeplearningbook.org](https://www.deeplearningbook.org/)

3. E. Voita Natural Language Processing course <https://lena-voita.github.io/nlp_course.html>

4. курс А.А. Сорокина «Нейросетевой анализ естественного языка» <https://github.com/deeppavlov/raai_summer_school_nlp_2021>

* 1. Материально-техническое обеспечение

Ноутбук или ПК с доступом в Интернет, проектор.

1. Фонд оценочных средств

**Вопросы к зачёту**

1. Регулярные выражения, их применения в компьютерной лингвистике.
2. Конечные автоматы, их применения в компьютерной лингвистике.
3. Конечные преобразователи, их применения в компьютерной лингвистике.
4. Вероятность текста, энграммные модели.
5. Лингвистические корпуса текстов и их количественный анализ.
6. Автоматическое исправление опечаток. Расстояние Левенштейна.
7. Основные типы задач машинного обучения. Их приложения в компьютерной лингвистике.
8. Линейные модели классификации.
9. Нейронные сети в задачах классификации, их приложения.
10. Нейросетевые методы в задаче машинного перевода.
11. Предобученные языковые модели: BERT.
12. Предобученные языковые модели: GPT.
13. Библиотеки Spacy и UDPipe для автоматического морфо­син­таксического анализа.
14. Библиотека transformers для задач компьютерной лингвистики.
	1. Текущий контроль успеваемости

Выполнение самостоятельных заданий. Опросы и дискуссии в рабочем порядке.

* 1. Промежуточная аттестация

Зачёт

1. В настоящем столбце должны быть указаны только те индикаторы достижения компетенций, которые связаны с данной дисциплиной (модулем) согласно таблице 4.1. Общей характеристики ОПОП. [↑](#footnote-ref-1)