

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт филологии и межкультурной коммуникации
Высшая школа зарубежной филологии и межкультурной коммуникации им. И.А. Бодуэна де Куртенэ



УТВЕРЖДАЮ:

Проректор по образовательной деятельности КФУ

 Е.А. Турилова

28 февраля 2025 г.

подписано электронно-цифровой подписью

Программа дисциплины

Автоматический анализ текста (на английском языке)

Направление подготовки: 45.03.01 - Филология

Профиль подготовки: Прикладная филология: иностранный (английский) язык в международной коммуникации

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): доцент, к.н. Гурьянов И.О. (кафедра романо-германской филологии, Высшая школа зарубежной филологии и межкультурной коммуникации им И А Бодуэна де Куртенэ), IOGuryanov@kpfu.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

| Шифр компетенции | Расшифровка приобретаемой компетенции |
|-------------------------|---|
| ПК-2 | Способен проводить научные исследования в конкретной области филологического знания с формулировкой аргументированных умозаключений и выводов |
| ПК-7 | Способность эффективно применять базовые математические знания и информационные технологии при решении прикладных задач |
| УК-1 | Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач |
| УК-4 | Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах) |

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

1. Основные концепции и определения автоматического анализа текста.
2. Методы обработки естественного языка (NLP).
3. Алгоритмы машинного обучения, применяемые в анализе текста.
4. Инструменты и библиотеки для автоматического анализа текста.

5. Этические аспекты использования технологий анализа текста.

Должен уметь:

1. Применять методы NLP для решения конкретных задач.
2. Реализовывать алгоритмы машинного обучения на практике.
3. Анализировать и интерпретировать результаты обработки текстовых данных.
4. Работать с библиотеками Python для автоматического анализа текста.
5. Проводить исследовательскую работу в области автоматического анализа текста.

Должен владеть:

1. Умение структурировать текстовые данные для анализа.
2. Навыки работы с большими объемами текстовой информации.
3. Способность визуализировать данные и результаты анализа.
4. Умение критически оценивать источники информации.

5. Навыки командной работы при выполнении проектов и задач в области анализа текста.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.18 Дисциплины (модули)" основной профессиональной образовательной программы 45.03.01 "Филология (Прикладная филология: иностранный (английский) язык в международной коммуникации)" и относится к части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 4 курсе в 7 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) на 108 часа(ов).

Контактная работа - 49 часа(ов), в том числе лекции - 16 часа(ов), практические занятия - 32 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 1 часа(ов).

Самостоятельная работа - 32 часа(ов).

Контроль (зачёт / экзамен) - 27 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 7 семестре.

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

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

| N | Разделы дисциплины / модуля | Се-местр | Виды и часы контактной работы, их трудоемкость (в часах) | | | | | | Само-стоя-тель-ная ра-бота |
|----|--|----------|--|--------------------|------------------------------|---------------------------|-----------------------------|---------------------------|----------------------------|
| | | | Лекции, всего | Лекции в эл. форме | Практи-ческие занятия, всего | Практи-ческие в эл. форме | Лабора-торные работы, всего | Лабора-торные в эл. форме | |
| 1. | Тема 1. 1. Introduction to Natural Language Processing (NLP) | 7 | 4 | 0 | 8 | 0 | 0 | 0 | 8 |
| 2. | Тема 2. 2. Text Classification Techniques | 7 | 4 | 0 | 8 | 0 | 0 | 0 | 8 |
| 3. | Тема 3. 3. Sentiment Analysis | 7 | 4 | 0 | 8 | 0 | 0 | 0 | 8 |
| 4. | Тема 4. 4. Information Extraction and Named Entity Recognition (NER) | 7 | 4 | 0 | 8 | 0 | 0 | 0 | 8 |
| | Итого | | 16 | 0 | 32 | 0 | 0 | 0 | 32 |

4.2 Содержание дисциплины (модуля)

Тема 1. 1. Introduction to Natural Language Processing (NLP)

Overview: Define NLP and its importance in automated text analysis.

- Key Concepts: Tokenization, stemming, lemmatization, and part-of-speech tagging.
- Tools: Introduction to popular NLP libraries (e.g., NLTK, spaCy).

These topics provide a comprehensive foundation for students to understand and apply automated text analysis techniques effectively.

Тема 2. 2. Text Classification Techniques

- Overview: Explore various text classification methods and their applications.
- Key Concepts: Supervised vs. unsupervised learning, feature extraction, and model evaluation metrics.
- Practical Exercise: Implement a simple text classification model using machine learning algorithms.

These topics provide a comprehensive foundation for students to understand and apply automated text analysis techniques effectively.

Тема 3. 3. Sentiment Analysis

- Overview: Understand sentiment analysis and its significance in understanding public opinion.
- Key Concepts: Lexicon-based approaches vs. machine learning approaches, sentiment polarity, and emotion detection.
- Case Study: Analyze social media data to assess public sentiment on a specific topic.

These topics provide a comprehensive foundation for students to understand and apply automated text analysis techniques effectively.

Тема 4. Information Extraction and Named Entity Recognition (NER)

- Overview: Examine techniques for extracting structured information from unstructured text.
- Key Concepts: Named entity recognition, relation extraction, and event extraction.
- Hands-on Activity: Use NER tools to identify entities in a given dataset and discuss the results.

These topics provide a comprehensive foundation for students to understand and apply automated text analysis techniques effectively.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99б/ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Справочные материалы - <https://medium.com/>

Справочные материалы - <https://towardsdatascience.com/>

Справочные материалы - <https://www.kaggle.com/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

| Вид работ | Методические рекомендации |
|-----------|--|
| лекции | <p>Лекции</p> <p>1. Структура лекций:</p> <ul style="list-style-type: none"> - Введение в тему: определение автоматического анализа текста, его цели и задачи. - Обзор методов и технологий: машинное обучение, обработка естественного языка (NLP), статистические методы. - Примеры применения: анализ настроений, классификация текстов, извлечение информации. <p>2. Использование мультимедиа:</p> <ul style="list-style-type: none"> - Включение видео и интерактивных материалов для иллюстрации сложных концепций. - Презентации с графиками и схемами для наглядности. <p>3. Обсуждение кейсов:</p> <ul style="list-style-type: none"> - Разбор реальных примеров из бизнеса и науки, чтобы студенты увидели практическую ценность. <p>4. Вопросы и ответы:</p> <ul style="list-style-type: none"> - Выделение времени для обсуждения вопросов студентов, чтобы углубить понимание материала. <p>5. Рекомендованная литература:</p> <ul style="list-style-type: none"> - Предоставление списка научных статей и книг для дополнительного чтения. |

| Вид работ | Методические рекомендации |
|------------------------|---|
| практические занятия | <p>Практические занятия</p> <ol style="list-style-type: none"> 1. Работа с инструментами: <ul style="list-style-type: none"> - Ознакомление с популярными библиотеками для анализа текста (например, NLTK, spaCy, Hugging Face Transformers). - Проведение практических лабораторных работ по обработке текстов. 2. Проектные работы: <ul style="list-style-type: none"> - Создание небольших проектов по анализу текстов на основе реальных данных (например, анализ отзывов). 3. Групповая работа: <ul style="list-style-type: none"> - Формирование рабочих групп для совместного выполнения заданий, что способствует обмену идеями. 4. Применение алгоритмов: <ul style="list-style-type: none"> - Практика в реализации различных алгоритмов машинного обучения для решения задач анализа текста. 5. Обратная связь: <ul style="list-style-type: none"> - Регулярное предоставление обратной связи по выполненным заданиям для улучшения навыков. |
| самостоятельная работа | <p>Самостоятельная работа</p> <ol style="list-style-type: none"> 1. Исследовательские проекты: <ul style="list-style-type: none"> - Студенты выбирают тему для самостоятельного исследования и представляют результаты в виде отчета или презентации. 2. Чтение литературы: <ul style="list-style-type: none"> - Чтение научных статей и книг по актуальным темам в области автоматического анализа текста. 3. Онлайн-курсы: <ul style="list-style-type: none"> - Рекомендация прохождения онлайн-курсов по обработке естественного языка и машинному обучению. 4. Участие в хакатонах: <ul style="list-style-type: none"> - Поощрение участия в конкурсах и хакатонах по анализу данных и NLP. 5. Ведение дневника: <ul style="list-style-type: none"> - Ведение дневника изучения, где студенты фиксируют свои достижения и трудности. |

| Вид работ | Методические рекомендации |
|-----------|---|
| экзамен | <p>Экзамен</p> <p>1. Формат экзамена:</p> <ul style="list-style-type: none"> - Комбинированный экзамен: теоретическая часть (вопросы) и практическая часть (решение задач). <p>2. Темы экзамена:</p> <ul style="list-style-type: none"> - Основные концепции автоматического анализа текста, методы обработки данных, примеры применения. <p>3. Подготовка к экзамену:</p> <ul style="list-style-type: none"> - Проведение ревиционных занятий перед экзаменом для повторения ключевых тем. <p>4. Критерии оценки:</p> <ul style="list-style-type: none"> - Четкие критерии оценки как теоретической, так и практической части экзамена. <p>5. Обратная связь после экзамена:</p> <ul style="list-style-type: none"> - Обсуждение результатов экзамена с целью выявления сильных и слабых сторон у студентов. |

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;

- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 45.03.01 "Филология" и профилю подготовки "Прикладная филология: иностранный (английский) язык в международной коммуникации".

*Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.18 Автоматический анализ текста (на английском языке)*

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 45.03.01 - Филология

Профиль подготовки: Прикладная филология: иностранный (английский) язык в международной коммуникации

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Основная литература:

1. Куликов, В. А. Автоматический анализ текстов: теория и практика / В. А. Куликов. - М.: Наука, 2018. - 320 с.
2. Соловьев, А. И. Основы обработки естественного языка / А. И. Соловьев, Н. В. Петров. - СПб.: Питер, 2020. - 256 с.
3. Михайлов, Д. В. Введение в машинное обучение для анализа текста / Д. В. Михайлов. - Екатеринбург: Урал. ун-т, 2019. - 180 с.
4. Баранов, И. А. Методы автоматического извлечения информации из текстов / И. А. Баранов. - Казань: Казанский университет, 2021. - 210 с.
5. Петрова, Е. С. Текстовая аналитика: от данных к решениям / Е. С. Петрова, А. Н. Смирнов. - Новосибирск: Сибирское университетское издательство, 2022. - 300 с.

Дополнительная литература:

1. Лебедев, С. Ю. Анализ тональности текстов: подходы и инструменты / С. Ю. Лебедев. - Москва: МГТУ им. Баумана, 2020. - 150 с.
2. Иванова, Т. Н. Обработка естественного языка в современных системах / Т. Н. Иванова, В. П. Сергеева. - Ростов н/Д: Ростовский университет, 2019. - 220 с.
3. Громова, А. В. Прикладные аспекты анализа текстов / А. В. Громова, И. В. Устинов. - Владивосток: Дальневосточное издательство, 2021. - 275 с.
4. Кузнецов, Р. С. Алгоритмы и методы обработки текстов / Р. С. Кузнецов, О. В. Сидорова. - Краснодар: Кубанский государственный университет, 2022. - 190 с.
5. Федоров, П. А. Современные подходы к автоматическому анализу текстов / П. А. Федоров, Н. А. Васильева. - Челябинск: Челябинский государственный университет, 2018. - 240 с.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.18 Автоматический анализ текста (на английском языке)

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 45.03.01 - Филология

Профиль подготовки: Прикладная филология: иностранный (английский) язык в международной коммуникации

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows