

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Филологический факультет

УТВЕРЖДЕНО:
Декан
И. В.Тубалова

Рабочая программа дисциплины

Информационные технологии в филологии

по направлению подготовки

45.04.01 Филология

Направленность (профиль) подготовки:
Русский язык как иностранный

Форма обучения
Очная

Квалификация
Магистр

Год приема
2025

СОГЛАСОВАНО:
Руководитель ОП
М.М. Угрюмова
Председатель УМК
Ю.А. Тихомирова

Томск – 2025

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 Способен использовать в профессиональной деятельности, в том числе педагогической, знания современной научной парадигмы в области филологии и динамики ее развития, системы методологических принципов и методических приемов филологического исследования.

ОПК-3 Способен владеть широким спектром методов и приемов филологической работы с различными типами текстов.

ПК-1 Выполнение отдельных заданий в рамках решения исследовательских задач в сфере русского языка под руководством более квалифицированного работника.

УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном языке, для академического и профессионального взаимодействия.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.3 Способен к самостоятельному поиску информации в области новых, в том числе междисциплинарных, отраслей филологии, их методологических принципов и методических приемов для расширения сферы своей научной деятельности.

ИОПК-3.3 Способен к самостоятельному поиску информации о новом эмпирическом материале, концепциях и теориях в избранной конкретной области филологии, новациях в ее методологических принципах и методических приемах для расширения сферы своей научной деятельности.

ИПК-1.2 Реализует под руководством более квалифицированного работника план решения исследовательской задачи, используя необходимые информационные ресурсы, методы получения данных, при необходимости производя его корректировку в свете полученных промежуточных результатов.

ИУК-4.1 Обосновывает выбор актуальных коммуникативных технологий (информационные технологии, модерирование, медиация и др.) для обеспечения академического и профессионального взаимодействия.

2. Задачи освоения дисциплины

Целями освоения учебной дисциплины «Основные направления лингвистического обеспечения новых информационных технологий» – систематизация знаний по проблемам алгоритмизации и моделирования для решения лингвистических и исследовательских задач, с помощью современных информационных технологий, организации ведения результатов исследовательской деятельности с применением специализированных программ.

- Освоить инструменты ведения отчетности и написания статей.
- Научиться применять основные программные средства с целью увеличения эффективности обработки естественного языка.
- Научиться владеть формальными грамматиками, автоматизировать работу извлечения сущностей из текстового массива данных.
- Уметь интегрировать средства автоматизации обработки естественного языка в программный код.
- Знать и уметь пользоваться операционными системами (UNIX)

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплины (модули)».

Дисциплина относится к обязательной части образовательной программы.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Первый семестр, зачет

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования. Пререквизитами дисциплины являются следующие дисциплины: курс семантики и фонетики, основы программирования. Постреквизиты: Язык программирования R, учебная и производственная практики, Text mining с R.

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 2 з.е., 72 часов, из которых:

-лекции: 10 ч.

-практические занятия: 24 ч.

в том числе практическая подготовка: 24 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в информационные системы

1.1. Информационные технологии в современном мире

1.2. История информационных технологий

1.3. Архитектура и устройство персональных компьютеров

Тема 2. Типы операционных систем

2.1. Операционная система Windows

2.2. Операционная система Linux (Deb)

2.3. Работа в среде Linux

2.4. Создание и работа в системе контроля репозиторий (git)

2.5. Основные команды git

Тема 3. Морфологический анализ данных

3.1. Морфологические анализаторы

3.2. Работа в морфологическом анализаторе mystem, pymorphy

3.3. Использование сценариев и формальных грамматик. Работа ПО Tomita-parser

Тема 4. Организация научных исследований

4.1. Формирование структуры (отчета, диссертации, исследования) и содержания с помощью автоматизации в MS Office

4.2. Автоматическое формирование литературы (Mendeley, Zotero)

Тема 5. Организация научных исследований в Latex

5.1 Введение в Latex

5.2 Организация структуры документа

5.3 Работа с изображениями

5.4 Работа с формулами

5.5 Создание презентаций

Тема 6. Интеллектуальные алгоритмы и лингвистика

6.1. Тенденции развития современных технологий искусственного интеллекта

6.2. Использование интеллектуальных алгоритмов для решения лингвистических

6.3. Типы предварительной обработки текста для применения интеллектуальных алгоритмов

6.4. Проблема получения обучающих данных

Тема 7. Примеры интеллектуальных алгоритмов

7.1. Простейшая нейросеть на примере перцептрона

- 7.2. Обучение модели и оценка качества
- 7.3. Синонимическая близость и алгебраические операции над векторами
- 7.4. Рекуррентная модель и кодирование в вектор предложения
- 7.5. Модели глубокого обучения
- Тема 8. Инструменты для разработки интеллектуальных алгоритмов
- 8.1. Язык Python и технология CUDA
- 8.2. TensorFlow и репозиторий моделей Google
- 8.3. Библиотека глубокого обучения PyTorch
- 8.4. HuggingFace и библиотека Transformers
- 8.5. Библиотеки NLTK и Rumorphy
- Тема 9. Трансдукционные модели и трансформер
- 9.4. BERT как кодирующая половина трансформера
- 9.5. Идея Fine tuning
- 9.6. Генеративные модели и GPT как декодирующая половина трансформера

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, проведения контрольных работ, тестов по лекционному материалу, деловых игр по темам, выполнения домашних заданий, и фиксируется в форме контрольной точки не менее одного раза в семестр.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет в первом семестре проводится в письменной и устной форме по выбранному проекту. Проект предполагает логическое изложение теоретического блока с привязкой к практической деятельности. Продолжительность зачета 1 час.

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

11. Учебно-методическое обеспечение

- а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://lms.tsu.ru/course/view.php?id=26935>
- б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.
- в) План семинарских / практических занятий по дисциплине.
- г) Методические указания по проведению лабораторных работ.

12. Перечень учебной литературы и ресурсов сети Интернет

- а) основная литература:
 - Степанов А.Н. Информатика: учебник для вузов / А.Н. Степанов. – СПб.: Питер, 2015 – 720 с.
 - Jurafsky Daniel, James H. Martin. Speech and Language Processing. / An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Second Edition. Upper Saddle River, NJ, 2019. <https://www.cs.colorado.edu/~martin/slp2.html>
 - Николаев И.С. / Прикладная и компьютерная лингвистика. Изд. 2 URSS. 2017. 320 с. ISBN 978-5-9710-4633-2

б) дополнительная литература:

– Щипицина Л. Информационные технологии в лингвистике: учеб. пособие / Л. Щипицина. – М.: Флинта, 2015. – 128 с.

– Кодзасов С.В. Алгоритмы преобразования русских орфографических текстов в фонетическую запись / С.В. Кодзасов М.: МГУ, 1970. 130 с/

– Коваль С. А. Лингвистические проблемы компьютерной морфологии. СПб., 2005. Леонтьева Н. Н. Автоматическое понимание текстов. Системы, модели, ресурсы. М., 2006. Ляшевская О. Н. и др. Оценка методов автоматического анализа текста: морфологические парсеры русского языка. Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог–2010». Вып. 9(16). М., 2010.

– Шаров С. А., Беликов В. И., Копылов Н. Ю., Сорокин А. А., Шаврина Т. О. Корпус с автоматически снятой морфологической неоднозначностью: К методике лингвистических исследований. Компьютерная лингвистика и интеллектуальные технологии. // Диалог. М., 2015.

<http://www.dialog-21.ru/digests/dialog2015/materials/pdf/SharoffSAetal.pdf>

в) ресурсы сети Интернет:

– открытые онлайн-курсы

– Журнал «Эксперт» - <http://www.expert.ru>

– Официальный сайт Федеральной службы государственной статистики РФ - www.gsk.ru

– Официальный сайт Всемирного банка - www.worldbank.org

– Система контроля версий ПО <https://github.com/>

– Создание документов Latex <https://www.overleaf.com/>

– Морфологический анализатор Mystem <https://yandex.ru/dev/mystem/>

– Извлечение фактов (формальные грамматики) Tomita-parser <https://yandex.ru/dev/tomita/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://clib.tgu.ru/>

– Электронная библиотека (репозиторий) ТГУ – <http://vlib.tgu.ru/>

– ЭБС Лань – <http://e.lanbook.com/>

– ЭБС Консультант студента – <http://www.studentlibrary.ru/>

– Образовательная платформа Юпайт – <https://urait.ru/>

– ЭБС ZNANIUM.com – <https://znanium.com/>

– ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных (при наличии):

– Университетская информационная система РОССИЯ – <https://uisrussia.msu.ru/>

– Единая межведомственная информационно-статистическая система (ЕМИСС) – <https://www.fedstat.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Лаборатории, оборудованные ...

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешенном формате («Актру»).

15. Информация о разработчиках

Степаненко Андрей Александрович, НИ Томский государственный университет, ассистент кафедры общего славяно-русского языкознания и классической филологии