

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

Филологический факультет

УТВЕРЖДАЮ:
Декан филологического факультета
И. В. Тубалова
« 31 » августа 2023 г.



Рабочая программа дисциплины

Информационные технологии и основы информационной культуры в лингвистике

по направлению подготовки

45.03.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) подготовки:
Фундаментальная и прикладная лингвистика

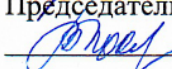
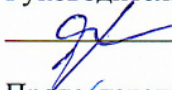
Форма обучения
Очная

Квалификация
Бакалавр

Год приема
2023

Код дисциплины в учебном плане: Б1.О.19

СОГЛАСОВАНО:
Руководитель ОПОП
А.В. Васильева
Председатель УМК
Ю.А. Тихомирова



Томск – 2023

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

- УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач;
- ОПК-7 – способность понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИУК-1.1. Осуществляет поиск информации, необходимой для решения задачи

ИОПК-7.3 Применяет современные технические средства и информационно-коммуникационные технологии для решения задач профессиональной деятельности.

ИОПК-7.2 Использует информационные ресурсы при решении стандартных задач профессиональной деятельности.

ИОПК-7.1 Осуществляет профессиональную деятельность с учетом требований информационной безопасности.

2. Задачи освоения дисциплины

Целями дисциплины «Информационные технологии и основы информационной культуры в лингвистике» являются формирование у студентов компетенций в области использования современных информационных и коммуникационных технологий для осуществления профессиональной лингвистической деятельности.

– освоить основные понятия прикладной лингвистики; направления использования ИКТ в лингвистике и возможности информационно-лингвистических технологий для решения профессиональных задач;

– основные понятия сферы информационной безопасности и основные методы защиты информации;

– грамотно использовать основные понятия прикладной лингвистики; решать профессиональные задачи с применением информационно-лингвистических технологий; выполнять основные мероприятия;

по защите информации при решении профессиональных задач;

– Освоить понятийный аппарат прикладной лингвистики; информационной культурой осуществления профессиональной деятельности с применением информационно-лингвистических технологий;

– Освоить возможности современных ИКТ для поддержки научного исследования;

– Уметь применять возможности современных ИКТ для поиска, систематизации, анализа и обработки информации в рамках проведения научного исследования.

3. Место дисциплины в структуре образовательной программы

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.

6. Язык реализации

Русский

7. Объем дисциплины

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

-лекции: 16 ч.

-практические занятия: 16 ч.

в том числе практическая подготовка: 16 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в курс

Понятие информации и информационных технологий. ИТ в лингвистике: направления использования

Аппаратное и программное обеспечение информационных технологий в лингвистике.

Операционные системы. Windows, Linux.

Прикладные разделы компьютерной лингвистики.

Тема 2. Фонетический уровень языка

Основы акустической теории речеобразования. Типологизация звуков АТР

Фонетическая разметка в программе PRAAT

Графическое представление звуков: спектрограмма, осциллограмма

Частота общего фона, форманты

Анализ и синтез звучащей речи

Тема 3. Компьютерная морфология

Анализ морфологии на основе правил.

Морфологический анализатор Rymorphy2, mystem, AOT, проект mystem+

Статистические методы анализа слов

Триграммная скрытая Марковская модель

Самостоятельная работа по теории

Тема 4. Извлечение информации из неструктурированного текстового массива данных

Распознавание сущностей: Natasha, Tomita-parser

Формальные грамматики

Распознавание отношений

Тема 5. Компьютерный синтаксис

Современные подходы к анализу синтаксических структур

Современные синтаксические анализаторы

Лингвистический процессор ЭТАП

Stanford NLP, RASP, OpenNLP, NLTK

9. Текущий контроль по дисциплине

Текущий контроль успеваемости обучающихся направлен на определение соответствия результатов обучения после освоения элемента по дисциплине, проводится путем контроля посещаемости, проведения контрольных работ, тестов по лекционному материалу, разработки кода, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр. Примерные задания текущего контроля:

Примерные задания по 4 модулю:

Извлеките факты из новостей темы «Футбол». Структурируйте информацию в виде таблицы html по виду: команда 1 – команда 2 – результат (победа, ничья, поражение) – счет

Примерный код:

National -> Noun<gram="geo"> | Noun<c-agr[1]> Noun<gram="geo", c-agr[1]>;

Club -> Noun<h-reg1, quoted>;

Club -> Word<h-reg1, l-quoted> Word* Word<r-quoted>;

Team -> Club | National;

Result -> Verb<kwtype="result_verb">;

Score -> AnyWord<wff=/[0-9]:[0-9]/>;

S -> Team AnyWord* Result AnyWord* Team AnyWord* Score;

S -> Team AnyWord* Team AnyWord* Result AnyWord* Score;

S -> Team AnyWord* Team AnyWord* Score;

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет проводится в устной форме по выбранному проекту. Проект предполагает логическое изложение теоретического блока с привязкой к практической деятельности и проверяет уровень овладения компетенций УК-1, ОПК-7, ИОПК-7.3, ИОПК-7.2, ИОПК-7.1, ИУК-1.1

Зачет по дисциплине принимается на основе достижения рубежных показателей в рейтинге (не ниже 55 баллов), при выполнении практических заданий, тестов, посещения занятий.

Рейтинг, баллы

1 – присутствие на лекции

1 – присутствие на занятии

1-3 – работа на занятии

1-36 – подготовка к занятию и работа на практическом занятии (в т.ч. д/з)

Результаты зачета определяются оценками «зачтено», «не зачтено».

Пример задачи:

Дан текст, при помощи контекстно-свободных формальных грамматики, извлеките из него сочетания по модели: «существительное» - «союз» - «существительное»

Примерный код:

```
encoding "utf8";
```

```
TTextMinerConfig {
```

```
  Dictionary = "colloc.gzt";
```

```
  PrettyOutput = "result.html";
```

```
  Input = {
```

```
    File = "articles.txt";
```

```
  }
```

```
  Articles = [
```

```
    { Name = " Colocation" }
```

```
  ]
```

```
  Facts = [
```

```
    { Name = " ColocationNcN" }
```

```
  ]
```

```
  Output = {
```

```
    Format = text;
```

```
  }
```

```
}
```

Критерии зачета обусловлены логической демонстрацией приобретенных компетенций в соответствии с текущей программой. Демонстрация предусматривает уверенное использование терминологии, понимание и корректное использование

математического аппарата, предусматривает корректность написания кода, его понимание и корректное использование в нем математических методов. Отметка «зачтено» выставляется за счет демонстрации полученных компетенций в практиках, домашних работах и итоговом задании: уверенное владение и понимание работы кода, знание и демонстрация в практике теоретических основ баз данных. Минимальный порог зачета составляет 55 баллов, ниже 55 – «не зачтено»

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=7571>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) План семинарских / практических занятий по дисциплине.

Тема 1. Введение в информатику. Понятие и представление информации. Операционные системы. Windows, Linux. Прикладные разделы компьютерной лингвистики.

Тема 2. Фонетический уровень языка. Основы акустической теории речеобразования. Типологизация звуков АТР. Фонетическая разметка в программе PRAAT. Графическое представление звуков: спектрограмма, осциллограмма. Частота общего фона, форманты. Анализ и синтез звучащей речи

Тема 3. Компьютерная морфология. Анализ морфологии на основе правил. Морфологический анализатор Rymorphy2, mystem, AOT, проект mystem+. Статистические методы анализа слов

Тема 4. Извлечение информации из неструктурированного текстового массива данных. Распознавание сущностей: Natasha, Tomita-parser. Формальные грамматики. Распознавание отношений

Тема 5. Компьютерный синтаксис. Современные подходы к анализу синтаксических структур. Современные синтаксические анализаторы. Лингвистический процессор ЭТАП. Stanford NLP, RASP, OpenNLP, NLTK

г) Методические указания по организации самостоятельной работы студентов.

Формы самостоятельной работы студентов разнообразны. Они включают в себя:

– изучение и систематизацию практических и теоретических примеров в рамках выполнения текущих заданий по предмету;

– изучение учебной, научной и методической литературы, материалов периодических изданий с привлечением электронных средств официальной, статистической, периодической и научной информации;

– подготовку докладов и презентаций, написание программного кода и его отладка;

– участие в работе студенческих конференций, комплексных научных исследованиях.

Самостоятельная работа приобщает студентов к научному творчеству, поиску и решению актуальных современных проблем.

Примеры самостоятельной работы студентов:

Скачайте текст, содержащий в себе неологизмы:

- проведите морфологический анализ лексем при помощи программы mystem

- создайте словарь неопределенных неологизмов в программе mystem

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, 2009. <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:

- Система ЭТАП-3: <http://proling.iitp.ru/ru/etap3>
- Синтаксический анализатор АОТ: <http://aot.ru/demo/synt.html>
- FrameNet <https://framenet.icsi.berkeley.edu>
- Stanford NLP <http://nlp.stanford.edu:8080/corenlp/process>
- Парсер RASP в составе системы Gate: последняя версия доступна по ссылке <http://ilexir.co.uk/applications/rasp/download/>
- OpenNLP: <https://opennlp.apache.org>
- Link Grammar Parser: <http://slashzone.ru/parser/parse.pl>
- Лингвистический пакет NLTK: <http://www.nltk.org/install.html>
- AIIRE <http://aiire.org>, <http://svn.aiire.org/repos/t>
- Томита-парсер: <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, Windows 7-10;
- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).
- язык программирования R (RStudio) и Python;
- Программа Mystem.

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

в) профессиональные базы данных:

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

– Единая межведомственная информационно-статистическая система (ЕМИСС) –
– Национальный корпус русского языка [Электронный ресурс]. URL: <https://ruscorpora.ru/>
Institute of Formal and Applied Linguistics [Электронный ресурс]. URL: <http://ufal.mff.cuni.cz/udpipe>

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 компьютерами (не ниже i3, RAM 8Gb), проектором

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, старший преподаватель кафедры общей, компьютерной и когнитивной лингвистики