

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)

ИНСТИТУТ ЛИНГВИСТИКИ

Кафедра теоретической и прикладной лингвистики

**Технологии искусственного интеллекта в гуманитарных
исследованиях**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Направление 45.03.02 «Лингвистика»

Направленность «Язык и коммуникация»

Уровень высшего образования: бакалавриат

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2023

Технологии искусственного интеллекта в гуманитарных исследованиях

Рабочая программа дисциплины

Составители:

Составители:

д.ф.н., доцент, заведующий кафедрой теоретической и прикладной лингвистики
Воронежского государственного университета Шилихина Ксения Михаловна,
к.ф.н., доцент кафедры теоретической и прикладной лингвистики Воронежского
государственного университета Донина Ольга Валерьевна

УТВЕРЖДЕНО

Протокол заседания кафедры ТиПЛ

№ 4 от 31.03.2023

ОГЛАВЛЕНИЕ

1. Пояснительная записка

1.1 Цель и задачи дисциплины

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

1.3. Место дисциплины в структуре образовательной программы

2. Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

5.1. Система оценивания

5.2. Критерии выставления оценок

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

7. Материально-техническое обеспечение дисциплины

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

9. Методические материалы

9.1. Планы практических (семинарских, лабораторных) занятий

9.2. Методические рекомендации по подготовке письменных работ

9.3. Иные материалы

Приложения

Приложение 1. Аннотация дисциплины

1. Пояснительная записка

1.1. Цели и задачи дисциплины

Целью данного курса является знакомство с технологиями искусственного интеллекта, которые применяются в различных сферах филологической и лингвистической деятельности. В рамках курса предусматривается ознакомление учащихся с основными типами лингвистических информационных ресурсов и с инструментами, которые могут быть полезны при решении базовых филологических и лингвистических прикладных задач.

Задачи курса – приобретение студентами навыков работы с лингвистическими базами данных, корпусами текстов и электронными словарями, а также получение представлений о возможном применении данных средств и ресурсов при проведении исследований.

1.2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ПК-3 – Способен к научно-исследовательской деятельности	<i>ПК-3.1</i> – Владеет основами методов научного исследования, информационной и библиографической культурой	По окончании курса студент должен знать различные типы электронных лингвистических ресурсов, уметь работать с технологиями искусственного интеллекта в области обработки естественного языка и грамотно применять их в соответствии с типом задачи,
	<i>ПК-3.2</i> – Владеет стандартными методиками поиска, анализа и обработки материала исследования	владеть навыками работы с электронными словарями, лингвистическими базами данных, приложениями для автоматического анализа морфологии и синтаксиса, системами автоматизированного перевода.

1.3. Место дисциплины в структуре образовательной программы

Курс «Технологии искусственного интеллекта в гуманитарных исследованиях» является частью, формируемой участниками образовательного процесса, Блока 1 дисциплин учебного плана ОП ВО бакалавриата «Лингвистика» 45.03.02 –Лингвистика. Дисциплина читается в 5 семестре 3 курса академического бакалавриата, реализуется УНЦ компьютерной лингвистики.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения курсов «Прикладная лингвистика», «Точные методы в гуманитарных науках».

В результате освоения дисциплины формируются знания, умения и владения, необходимые для подготовки к производственной практике, к государственной итоговой аттестации.

2. Структура дисциплины в образовательной программе

Структура дисциплины для очной формы обучения

Общая трудоёмкость дисциплины составляет 3 з.е., 108 ч., в том числе контактная работа обучающихся с преподавателем 28 ч. (8 ч. лекций, 20 ч. практических занятий), самостоятельная работа обучающихся 80 ч., в том числе промежуточная аттестация 18 ч.

№ п/п	Раздел дисциплины/темы	Семестр	Виды учебной работы (в часах)						Формы текущего контроля успеваемости, форма промежуточной аттестации
			контактная					Самостоятель- ная работа	
			Лекции	Семинар	Практические занятия	Лабораторные занятия	Промежуточ- ная аттестация		
1.	Раздел 1. <i>Искусственный интеллект в гуманитарной сфере</i>	5	1		2			6	1. Дискуссия.
2.	Раздел 2. <i>Филологические и лингвистические ресурсы и программы</i>	5	1		4			10	Практический анализ ресурсов и программ
3.	Раздел 3. <i>Национальные и иные корпусы</i>	5	1		4			10	Дискуссия Практический анализ корпусных данных
4.	Раздел 4. «Google books Ngram Viewer и поиск словосочетаний»	5	2		4			12	Дискуссия Практический анализ данных
5.	Раздел 5. <i>Компьютерная лексикография и системы представления лексических знаний</i>	5	1		2			12	Дискуссия Практический анализ данных
	Раздел 6. <i>Инструменты искусственного интеллекта</i>	5	2		4			12	Дискуссия
	Экзамен	7						18	
	Итого:		8		20			80	

3. Содержание дисциплины

Тема 1. «Искусственный интеллект в гуманитарной сфере»

Цели и задачи цифровой филологии. Направления цифровой филологии.
Известные проекты.

Тема 2. «Филологические и лингвистические ресурсы и программы»

Лингвистические базы данных.

Типы языковых (лингвистических) ресурсов.

Лингвистическое обеспечение информационного поиска.

Обзор ресурсов:

1. Ethnologue: Languages of the World. About the 23rd Edition. URL: <https://www.ethnologue.com/about/this-edition>
2. Glottolog. URL: <https://glottolog.org>
3. PHOIBLE 2.0. URL: <http://phoible.org>
4. The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: <http://wals.info>
5. База данных «Языки мира». URL: https://iling-ran.ru/web/ru/news/201008_langworld

Тема 3. «Национальные и иные корпуса»

История корпусной лингвистики в мире и в России. Понятие корпуса. Модели классификации корпусов. Национальный корпус русского языка. Британский национальный корпус. Другие корпуса русского языка. Примеры применения корпусов текстов для лингвистических исследований.

Тема 4. «Google books Ngram Viewer и поиск словосочетаний»

Сервисы исследования биграмм и n-грамм. Программы для статистической обработки данных.

Ресурс Google Books. Google Books Ngram Viewer.

Расширенный поиск в системах «Яндекс» и Google

Тема 5. «Компьютерная лексикография и системы представления лексических знаний»

Способы использования корпусов в лексикографии. Лингвостатистика. Частотные словари.

Компьютерная лексикография. Системы создания электронных словарей.

WordNet. Русский WordNet.

Средства построения частотных словарей.

Система Multitran

Тема 6. «Инструменты искусственного интеллекта»

Автоматическая разметка текстов (морфологический и синтаксический уровни).

Экспертные системы. Обучение с учителем. Самообучающиеся системы: нейронные сети.

Диалоговые системы. Многомерный анализ данных.

Статистический машинный перевод.

4. Образовательные технологии

Дисциплина «Технологии искусственного интеллекта в гуманитарных исследованиях» реализуется преимущественно интерактивно – в форме интерактивных лекций, семинарских занятий и в различных видах коллективной и самостоятельной работы студента.

В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	2	3	4
1.	Название раздела	Лекция. Практическое занятие. Самостоятельная работа	Лекция с использованием видеоматериалов Развернутая дискуссия. Выполнение практических заданий. Консультирование и проверка домашних заданий посредством электронной почты и в обсуждениях на практических занятиях

5. Оценка планируемых результатов обучения

5.1. Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль: - опрос - участие в дискуссии на семинаре	5 баллов 10 баллов	30 баллов 30 баллов
Промежуточная аттестация (защита проекта)		40 баллов
Итого за семестр зачёт		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

5.2. Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	«отлично»/ «зачтено (отлично)»/«зачте но»	Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации. Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Свободно ориентируется в учебной и профессиональной литературе. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».
82-68/ C	«хорошо»/ «зачтено (хорошо)»/«зачте но»	Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей. Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами. Достаточно хорошо ориентируется в учебной и профессиональной литературе. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».
67-50/ D,E	«удовлетворительно»/ «зачтено (удовлетворительно)»/ «зачтено»	Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	«неудовлетворительно»/ не зачтено	Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Вопросы для подготовки к текущим аттестациям

1. Перечислите корпуса, которые входят в состав Национального корпуса русского языка.

2. По каким характеристикам можно проводить поиск в Национальном корпусе русского языка?
3. Что такое формат KWIC и для чего он используется?
4. Какие Вам известны учебные корпуса текстов? В чем заключаются особенности их разметки?
5. Какая разметка доступна в Хельсинкском аннотированном корпусе русского языка? Есть ли отличия от разметки в НКРЯ?
6. Назовите режимы поиска в корпусах Лидского университета. В чем их отличия?
7. Какой объем Британского национального корпуса? На какие части делится корпус?
8. Какие метапризнаки описаны в Британском национальном корпусе?
9. В чем заключаются отличия корпусов Minus, Maius и Maximum из семейства Aranea?
10. Для чего используется параметр «сглаживание» в Google books Ngram Viewer?
11. Как устроен частотный словарь?
12. Поиск по документам на каких языках реализован в ИПС «Яндекс»?
13. В каких частях документа можно проводить поиск в ИПС Google?
14. Документы каких типов можно искать в ИПС Google?
15. Какие категории признаков описаны во «Всемирном атласе языковых структур»?
16. Языки каких статусов выделяются в базе данных «Этнолог»?

Темы рефератов / презентаций

1. Информационно-поисковые языки.
2. Лингвистические электронные библиотеки.
3. Автоматический графематический анализ текстов.
4. Автоматическая морфологическая разметка текстов.
5. Методы снятия морфологической омонимии.
6. Способы использования корпусов в лексикографии.
7. WordNet. Русский WordNet.
8. Средства построения частотных словарей.
9. Система Multitran.
10. Использование корпусов в лингвистических исследованиях.
11. Многомерный анализ данных.
12. Экспертные системы.

Критерии оценивания рефератов / презентаций

Оценка	Критерии оценки
«Отлично»	1. Доклад четко выстроен, сопровождается презентацией, отражающей проблематику доклада, имеет четкие выводы, полностью характеризующие работу. Автор отлично ориентируется в материале и отвечает на все вопросы, владеет специальной терминологией.

	2. Студент активно участвует в обсуждении докладов других студентов из группы.
«Хорошо»	1. Доклад структурирован, презентация не отражает в полной мере содержание доклада, допущены неточности, выводы нечеткие. Автор затрудняется ответить на вопросы, слабо использует специальную терминологию. 2. Студент участвует в обсуждении докладов других студентов из группы.
«Удовлетворительно»	1. Доклад структурирован, содержание не полностью раскрывает цель доклада, отсутствует презентация. 2. Студент пассивен при обсуждении докладов других студентов из группы.
«Неудовлетворительно»	1. Доклад не структурирован/отсутствует, презентация отсутствует. Автор не владеет материалом. 2. Студент не участвует в обсуждении докладов.

Вопросы для подготовки к зачету

1. Искусственный интеллект в гуманитарной сфере: цели и задачи.
2. Филологические и лингвистические программы и ресурсы, в том числе работающие в режиме онлайн.
3. Поисковые системы Интернета с точки зрения филолога.
4. Лингвостатистика. Частотные словари.
5. Компьютерная лексикография. Системы создания электронных словарей.
6. Корпуса текстов и их роль в современной лингвистике.
7. Корпусная лингвистика в России. Корпусы русского языка.
8. Классификация (типология) корпусов.
9. Понятие разметки. Типы разметки.
10. Корпус как поисковая система (корпусные менеджеры).
11. Коллокации. Сервисы исследования биграмм и n-грамм.
12. Лингвистические исследования, базирующиеся на корпусах.
13. Корпус как инструмент обучения языку.
14. Лингвистические базы данных и экспертные системы.
15. Подходы к классификации: обучение с учителем и обучение без учителя.
16. Самообучающиеся системы: нейронные сети.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

Рекомендуемые интернет-ресурсы

6. ABBYY Россия. URL: <http://www.abbyy.ru/products/>
7. Czech National Corpus. URL: <http://ucnk.ff.cuni.cz/english/index.php>
8. Ethnologue. Languages of the World. URL: <http://www.ethnologue.com/>
9. ICAME Corpus Collection. URL: <http://icame.uib.no/corpora.html>
10. The ELRA catalogue. URL: <http://catalogue.elra.info/en-us/>

11. The LDC Corpus Catalog. URL: <https://catalog.ldc.upenn.edu/byyear>
12. Russian Corpora in Tübingen <https://www.lingexp.uni-tuebingen.de/sfb441/b1/en/korpora.html>
13. Атлас языков мира. URL: <http://wals.info/>
14. База данных «Языки мира». URL: https://iling-ran.ru/web/ru/news/201008_langworld
15. Древнерусские берестяные грамоты. URL: <http://gramoty.ru/>
16. Корпус русского литературного языка. URL: <http://narusco.ru/>
17. Корпуса и коллекции Марка Дэвиса. URL: 1) <https://www.mark-davies.info/> / 2) <https://www.english-corpora.org/>
18. British National Corpus. URL: <http://www.natcorp.ox.ac.uk/>
19. Машинный фонд русского языка. URL: <http://cfrl.ruslang.ru/index0.php>
20. Мультитран. URL: <http://multitrans.ru>
21. Национальный корпус русского языка. URL: <http://ruscorpora.ru/>
22. Новый частотный словарь русской лексики. URL: <http://dict.ruslang.ru/freq.php>
23. Программное обеспечение в области автоматической обработки текста. URL: http://www.aot.ru/online_demo.html
24. Сайт «Вавилонская башня». URL: <http://starling.rinet.ru/>
25. Сайт компании «Стэл — компьютерные системы». URL: <http://www.stel.ru/>
26. Сайт компании «Центр речевых технологий». URL: <http://www.speechpro.ru/>
27. Словари, созданные на основе Национального корпуса русского языка. URL: <http://dict.ruslang.ru/>
28. Упсальский корпус русского языка. URL: <https://snd.gu.se/en/catalogue/study/ext0071>
29. Хельсинский аннотированный корпус. URL: <http://h248.it.helsinki.fi/hanco/>

7. Материально-техническое обеспечение дисциплины

Занятия по курсу можно проводить с максимальной эффективностью в компьютерном классе или аудитории с доступом в Интернет, проектором и экраном для презентаций. Необходимо также наличие доски или флипчарта, чтобы преподаватель мог разбирать примеры по ходу объяснения и записывать задания. Для самостоятельной работы студентам необходимо рабочее место, оборудованное персональным компьютером с доступом в Интернет, аудио- и видеоплеером (Windows Media Player, MPC, WinAmp, VLC и т.п.) а также офисными программами (Microsoft Office, OpenOffice, LibreOffice, Zoho Office и т.п.).

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
 - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
 - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
 - письменные задания оформляются увеличенным шрифтом;
 - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:

- лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
- письменные задания выполняются на компьютере в письменной форме;
- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением;
 - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
 - в печатной форме увеличенным шрифтом;
 - в форме электронного документа;
 - в форме аудиофайла.
- для глухих и слабослышащих:
 - в печатной форме;
 - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - в печатной форме;
 - в форме электронного документа;
 - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
 - устройством для сканирования и чтения с камерой SARA CE;
 - дисплеем Брайля PAC Mate 20;
 - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
 - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
 - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - передвижными, регулируемые эргономическими партами СИ-1;
 - компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

Примерные задания для практических / лабораторных занятий ПО ТЕМЕ 2.

1. Поиск на ресурсе Glottolog. Найдите индоевропейскую семью языков. Какие группы в нее входят?
2. Поиск на ресурсе Ethnologue. Перечислите наиболее многочисленные языковые семьи. В каких странах наблюдается наибольшее разнообразие языков? Входят ли в топ-10 европейские государства?
3. Поиск на ресурсе Ethnologue. К какой языковой семье относится баскский язык? Приведите примеры похожих на него в этом отношении языков. В каких странах они распространены?
4. Поиск на ресурсе Wals. Какой порядок следования подлежащего и сказуемого преобладает в языках? Приведите примеры языков, в которых реализованы разные значения данного признака.
5. Поиск на ресурсе Wals. Найдите славянские языки, на которых говорят в Германии. Назовите их.

Примерные задания для практических / лабораторных занятий ПО ТЕМЕ 3

1. На сайте НКРЯ. Найдите в рекламных текстах упоминание еды и напитков. Проанализируйте 20 первых результатов. Есть ли среди них ошибочные? Чем это можно объяснить?
2. На сайте НКРЯ. Найдите примеры для следующих вариативных форм: *дверьми* и *дверями*, *лошадьми* и *лошадьми*, *дочерьми* и *дочерями*; *мета*ет и *меч*ет, *брызж*ет и *брызг*ает, *двига*ет и *движ*ет. Проанализируйте, какая из форм в соответствующей паре используется чаще? Есть ли между ними разница?
3. На сайте НКРЯ. Проанализируйте примеры согласования глагола с количественной группой. Например, *пришло несколько человек*, *большинство студентов ответили*, *несколько деревьев пострадало* и под. Можно ли сформулировать правило, когда выбирается форма множественного и единственного числа?
4. На сайте НКРЯ. Найдите значения, которые могут передаваться при помощи жеста «всплеснуть руками».
5. На сайте корпуса русских учебных текстов. Приведите примеры ошибок в курсовых работах, связанных с использованием разговорного стиля.
6. На сайте BNC. Найдите сочетания существительных в форме множественного числа, объединенных сочинительным союзом *and*. Например, *ladies and gentlemen*.

Рекомендуемые интернет-ресурсы

1. British National Corpus. URL: <http://bncweb.lancs.ac.uk>
2. Czech National Corpus. URL: <http://ucnk.ff.cuni.cz/english/index.php>
3. Корпус русского литературного языка. URL: <http://narusco.ru/>
4. Лингвистические корпуса и сервисы. URL: <http://web-corpora.net/>
5. Национальный корпус русского языка. URL: <http://ruscorpora.ru/>
6. Система NoSketchEngine и корпуса семейства Aranea. URL: <http://unesco.uniba.sk/aranea>
7. Хельсинский аннотированный корпус. URL: <http://h248.it.helsinki.fi/hanco/>

Примерные задания для практических / лабораторных занятий по теме 4:

1. Найдите сочетания глагола *заказать* с существительным в русскоязычном корпусе. Внимание: задание можно интерпретировать по-разному. Каким образом в таком случае будут отличаться результаты? Прокомментируйте полученные данные.
1. Проверьте, насколько частотны результаты для словосочетаний *small question* и *little question* в англоязычных корпусах (общем, американском и британском).
2. Проследите, как менялась частотность следующих слов начиная с 1940-х: *омбудсмен*, *хайп*, *франчайзинг*, *коуч*, *айпад*, *джойстик*, *пейджер*, *ваучер*, *секвестр*, *волонтер*.

Примерные задания для практических / лабораторных занятий по теме 5:

1. На сайте ФЭБ. В ЭНИ «История всемирной литературы» найдите статью, посвященную творчеству Ч. Диккенса.
2. На сайте ФЭБ. Найдите письма С. А. Толстой к своему супругу.
3. На сайте словаря глагольной сочетаемости. Найдите словосочетания, в которых один из компонентов имеет положительную оценку (например, *жизнь удалась*).
4. На сайте частотного словаря. Найдите в словаре частотности в художественных и публицистических текстах для следующих слов: *перестройка, гласность, трешка, хлыст, абвер, налоговый, зорко, парить*. Проинтерпретируйте полученные результаты.

Рекомендуемые интернет-ресурсы:

1. Грамматический словарь новых слов русского языка. URL: <http://dict.ruslang.ru/gram.php>
2. Новый частотный словарь русской лексики. URL: <http://dict.ruslang.ru/freq.php>
3. Словарь русской идиоматики. URL: <http://dict.ruslang.ru/magn.php>
4. Словарь глагольной сочетаемости непредметных имен русского языка. URL: http://dict.ruslang.ru/abstr_noun.php
5. Фундаментальная электронная библиотека «Русская литература и фольклор». URL: <http://feb-web.ru/>
1. Мультитран. URL: <http://multitran.ru>

Примерные задания для практических / лабораторных занятий по теме 6.

1. Выполнить морфологическую разметку предложенного текста.
2. Выполнить автоматический перевод при помощи систем «Яндекс», Google и Promt.

Рекомендуемые интернет-ресурсы:

1. Лингвистический процессор «ЭТАП-3». URL: <http://proling.iitp.ru/ru/etap>
2. Программное обеспечение в области автоматической обработки текста. URL: www.aot.ru
3. Платформа «Текстометрия». URL: <http://textometrie.org>

АННОТАЦИЯ

Целью курса

- знакомство с технологиями искусственного интеллекта, которые применяются в различных сферах филологической и лингвистической деятельности.
- ознакомление учащихся с основными типами лингвистических информационных ресурсов и с инструментами, которые могут быть полезны при решении базовых филологических и лингвистических прикладных задач.

Задачи курса

- приобретение студентами навыков работы с лингвистическими базами данных, корпусами текстов и электронными словарями,
- получение представлений о возможном применении данных средств и ресурсов при проведении исследований.

В результате освоения дисциплины обучающийся должен:

знать

различные типы электронных лингвистических ресурсов,

уметь

работать с технологиями искусственного интеллекта в области обработки естественного языка и грамотно применять их в соответствии с типом задачи,

владеть

навыками работы с электронными словарями, лингвистическими базами данных, приложениями для автоматического анализа морфологии и синтаксиса, системами автоматизированного перевода.