

МИНОБРНАУКИ РОССИИ

**Федеральное государственное автономное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Кафедра теоретического и прикладного языкознания

Авторы-составители: Худякова Екатерина Сергеевна

**Рабочая программа дисциплины
ВВЕДЕНИЕ В СПЕЦИАЛЬНОСТЬ
Код УМК 99515**

Утверждено
Протокол №9
от «30» мая 2023 г.

Пермь, 2023

1. Наименование дисциплины

Введение в специальность

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в обязательную часть Блока « Б.1 » образовательной программы по направлениям подготовки (специальностям):

Направление подготовки: **45.03.03** Фундаментальная и прикладная лингвистика
направленность Компьютерная лингвистика

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Введение в специальность** у обучающегося должны быть сформированы следующие компетенции:

45.03.03 Фундаментальная и прикладная лингвистика (направленность : Компьютерная лингвистика)

УК.9 Знает правовые и этические нормы, способен оценивать последствия нарушения этих норм

Индикаторы

УК.9.1 Ориентируется в правовых принципах и нормах в разных сферах жизнедеятельности и последствиях их нарушения

УК.9.2 Ориентируется в этических нормах поведения в разных видах профессиональной деятельности и последствиях их нарушения

ОПК.2 Способен понимать принципы работы современных информационно-коммуникационных технологий и использовать их для решения профессиональных задач с учетом требований информационной безопасности

Индикаторы

ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности

ОПК.3 Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики

Индикаторы

ОПК.3.1 демонстрирует знание основных понятий и категорий современной лингвистики

4. Объем и содержание дисциплины

Направление подготовки	45.03.03 Фундаментальная и прикладная лингвистика (направленность: Компьютерная лингвистика)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	1
Объем дисциплины (з.е.)	3
Объем дисциплины (ак.час.)	108
Контактная работа с преподавателем (ак.час.), в том числе:	42
Проведение лекционных занятий	14
Проведение практических занятий, семинаров	28
Самостоятельная работа (ак.час.)	66
Формы текущего контроля	Защищаемое контрольное мероприятие (2) Итоговое контрольное мероприятие (1)
Формы промежуточной аттестации	Зачет (1 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Компьютерная лингвистика

Введение в проблематику. Понятия предмета и объекта науки, понятие методологии, метода и методики. Необходимое обеспечение курса. Установка Python, ознакомление со средой разработки.

Понятие прикладной лингвистики. Основные проблемы, история возникновения. Методы ПЛ

Предпосылки развития ПЛ (научные – структурализм, математическая лингвистика, когнитивные науки, кибернетика, ИИ; технические – развитие вычислительной техники). Широкое и узкое понимание предмета ПЛ. Моделирование как базовый метод ПЛ. Методологические различия ПЛ и теоретической лингвистики.

Классификация предметных областей ПЛ по оптимизируемым функциям языка. ПЛ и смежные науки

Разделы ПЛ: оптимизация способов фиксации и хранения информации, способов передачи информации, интеллектуальных способностей человека, использования языка как средства коммуникации.

Понятие компьютерной лингвистики. Цели и задачи, методы и области применения

Причины развития КЛ. Понятие КЛ, предмет (лингвистическое обеспечение информатики). Понятие lingware. Автоматический анализ и синтез текстов на естественном языке: этапы и уровни. Перспективы развития КЛ.

Лингвистика и информатика. Понятие ИИ. Прикладная лингвистика и ИИ

Лингвистика и информатика: предмет информатики, задачи, области пересечения с лингвистикой.

Понятие ИИ: широкое и узкое понимание ИИ. Этапы развития ИИ (история развития дисциплины), области пересечения ПЛ и ИИ.

Свойства естественного языка. Возможности формализации ЕЯ (поуровневый анализ): ограничения и способы их преодоления. Проблема формализуемости языка.

Обработка текстовой информации на естественном языке

Понятие ситуации в лингвистике: семантическое и синтаксическое определения. Макроситуация текста. Методика ситуативного анализа

Понятие цельности и связности текста. Установление ядерно-периферийных и ядерно-ядерных отношений между высказываниями.

. «Информационный вес» фрагмента текста как база для обработки текста. Составление шаблона аннотации, выбор лексических заполнителей, составление словаря «стоп-слов». Способы представления знаний в КЛ. Онтологии как база для аннотирования: методы составления онтологий и тезаурусов. «Вес» фрагмента (статистические показатели веса) и «свертывание» текстовой структуры.

Преппроцессинг. Корпусный подход к подготовке лингвистических данных. Строковые методы в Python. Регулярные выражения для преппроцессинга.

Преппроцессинг: предварительная подготовка текстовых данных: причины и способы. Корпусный подход к подготовке лингвистических данных. Очистка данных. Строковые методы в Python: s.replace; s[:], s.find; s.split. Понятия массива, кортежа, списка. словаря. Составление словаря. Регулярные выражения для преппроцессинга.

Морфологический автоматический анализ

Принципы автоматического морфологического анализа, результаты: грамматические характеристики (стеммер), леммы (лемматизаторы). Типы МорфАн: со словарем флексий, лемм, бессловарные. Принципы работы, ограничения. Анализ работы лемматизатора и морфологического анализатора.

Фильтры на категориальные признаки.

Синтаксический автоматический анализ

Модели синтаксиса: непосредственно составляющие; дерево решений; лексикозависимые методы, лексико-синтаксические шаблоны; фреймовый синтаксис. NLTP-библиотека; treetagger. Ограничения и проблемы синтаксического анализа русского языка.

Понятие информационного поиска. Классификаторы текстов

Типы и виды информационного поиска (полнотекстовый, по метаданным), механизмы ИП: адресный, документальный, фактуальный. Методики осуществления ИП. Методы представления лингвистических данных: модель мешка слов, с морфологической предобработкой, с учетом структуры, модель Деб. Элементы машинного обучения для классификации текста (модальность, семантика, жанр).

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторные занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Коваленко Ю. В. Информационно-поисковые системы: Учебно-методическое пособие / Коваленко Ю. В. - Омск: Омская юридическая академия, 2017, ISBN 978-5-98065-148-0.-38.
<http://www.iprbookshop.ru/66817.html>
2. Сузи, Р. А. Язык программирования Python : учебное пособие / Р. А. Сузи. — 3-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 350 с. — ISBN 978-5-4497-0705-5. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/97589.html>
3. Титов, А. Н. Python. Обработка данных : учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. — Казань : Издательство КНИТУ, 2022. — 104 с. — ISBN 978-5-7882-3171-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <http://www.iprbookshop.ru/129220.html>
4. Connor P. Milliken. Python Projects for Beginners / Connor P. Milliken // Publisher Name: Apress, Berkeley, CA. — 2020. — 332 p. — ISBN 978-1-4842-5355-7. [Электронный ресурс].
<https://link.springer.com/book/10.1007/978-1-4842-5355-7>

Дополнительная:

1. Новое в зарубежной лингвистике. сборник статей : переводы. - Москва: Прогресс, 1989. Вып. 24. Компьютерная лингвистика/ сост., ред. и вступ. ст. д-ра филол. наук Б. Ю. Городецкого. - 1960.-432.- Библиогр.: с. 407-422. - Предм. указ.: с. 423-427. - Указ. компьютер. систем: с. 428-429
2. Ихсанов Н. Х. Формальное моделирование подмножеств естественного языка: автореферат дис. ... канд. техн. наук : 05.13.18/Н. Х. Ихсанов.-Ульяновск, 2000.-22.
3. Терминологический инжиниринг на иностранном языке. Осваиваем терминологию компьютерной безопасности: учебное пособие для студентов, обучающихся по направлению подготовки магистров «Фундаментальная и прикладная лингвистика»/М-во науки и высш. образования РФ, Перм. гос. нац. исслед. ун-т.-Пермь: ПГНИУ, 2020, ISBN 978-5-7944-3429-3.-145. <https://elis.psu.ru/node/595731>
4. Мокроусов М. Н. Разработка и исследование методов и системы семантического анализа естественно-языковых текстов: автореферат дис. ... канд. техн. наук : 05.13.01/М. Н. Мокроусов.-Ижевск, 2010.-24.

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<http://company.yandex.ru/technologies/search> технологии Яндекса для обработки текста

<https://colab.research.google.com/> среда разработки

<https://ruscorpora.ru/> Национальный корпус русского языка

transcriber.semograph.com/corpora корпус региолектной русской речи

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Введение в специальность** предполагает использование следующего программного обеспечения и информационных справочных систем:

Презентационные материалы (слайды по темам лекционных и практических занятий); доступ в режиме on-line в Электронную библиотечную систему (ЭБС); доступ в электронную информационно-образовательную среду университета.

Образовательный процесс по дисциплине «Информационные технологии в лингвистике» предполагает использование следующего программного обеспечения и информационных справочных систем:

1. офисный пакет
2. Оболочки для создания обучающих программ.
3. Онлайн приложения для создания мультимедиа презентаций.
4. Оболочки для создания экспертной системы.
5. Библиотеки для языка Python

Программы пп. 1–5 являются свободно распространяемыми (free software).

Для проведения практических занятий применяются ресурсы сети Интернет (<https://colab.research.google.com/>), приложения Mystem, Speech Analyzer, дистрибутив языков программирования Python и R Anaconda (все – свободно распространяемые), библиотеки языка программирования Python.

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

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

Презентационные материалы (слайды по темам лекционных и практических занятий); доступ в режиме on-line в Электронную библиотечную систему (ЭБС); доступ в электронную информационно-образовательную среду университета.

Для проведения практических занятий применяются ресурсы сети Интернет (<https://colab.research.google.com/>), приложения Mystem, Speech Analyzer, дистрибутив языков программирования Python и R Anaconda (все – свободно распространяемые), библиотеки языка программирования Python.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с

доступом к локальной и глобальной компьютерным сетям.

2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине
Введение в специальность**

**Планируемые результаты обучения по дисциплине для формирования компетенции.
Индикаторы и критерии их оценивания**

ОПК.3

Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
ОПК.3.1 демонстрирует знание основных понятий и категорий современной лингвистики	Знает базовые проблемы современной прикладной лингвистики; умеет использовать основные компьютерные лингвистические программы; умеет применять на практике методы автоматической обработки языкового материала; владеет методами и методиками автоматической обработки текста для решения конкретных задач автоматической обработки текста.	Неудовлетворител Не знает базовых проблем современной прикладной лингвистики; не умеет использовать компьютерные лингвистические программы; не умеет применять на практике никакие методы автоматической обработки языкового материала; не владеет никакими методиками автоматической обработки текста для решения конкретных задач автоматической обработки текста. Удовлетворительн Знает некоторые базовые проблемы современной прикладной лингвистики; умеет под руководством преподавателя использовать компьютерные лингвистические программы; умеет применять на практике некоторые методы автоматической обработки языкового материала; владеет некоторыми методиками автоматической обработки текста для решения конкретных задач автоматической обработки текста. Хорошо Знает базовые проблемы современной прикладной лингвистики; умеет использовать при консультативной помощи преподавателя основные компьютерные лингвистические программы; умеет применять на практике основные методы автоматической обработки языкового материала; в целом владеет методами и методиками автоматической обработки текста для решения конкретных задач автоматической обработки текста. Отлично

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		Отлично Знает базовые проблемы современной прикладной лингвистики; умеет самостоятельно использовать базовые компьютерные лингвистические программы; умеет применять на практике методы автоматической обработки языкового материала; владеет методами и методиками автоматической обработки текста для решения конкретных задач автоматической обработки текста.

ОПК.2

Способен понимать принципы работы современных информационно-коммуникационных технологий и использовать их для решения профессиональных задач с учетом требований информационной безопасности

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности	Владение методами применения лингвистического ПО; умение создавать и использовать лингвистическое ПО; знание принципов и технологий автоматической обработки речи для решения широкого круга задач.	Неудовлетворител Не умеет самостоятельно применять информационные технологии для решения профессиональных и исследовательских задач; не владеет лингвистическими технологиями для анализа ЕЯ; не знает широкий перечень прикладных систем по обработке естественного языка, подходящих для решения профессиональных и исследовательских задач. Удовлетворительн Не умеет самостоятельно применять информационные технологии для решения профессиональных и исследовательских задач; слабо владеет лингвистическими технологиями для анализа ЕЯ; знает некоторый перечень прикладных систем по обработке естественного языка, подходящих для решения профессиональных и исследовательских задач. Хорошо Умеет самостоятельно применять информационные технологии для решения профессиональных и исследовательских задач; частично владеет лингвистическими

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		Хорошо технологиями для анализа ЕЯ; знает некоторый перечень прикладных систем по обработке естественного языка, подходящих для решения профессиональных и исследовательских задач Отлично Умеет самостоятельно применять информационные технологии для решения профессиональных и исследовательских задач; владеет лингвистическими технологиями для анализа ЕЯ; знает широкий перечень прикладных систем по обработке естественного языка, подходящих для решения профессиональных и исследовательских задач.

УК.9

Знает правовые и этические нормы, способен оценивать последствия нарушения этих норм

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
УК.9.1 Ориентируется в правовых принципах и нормах в разных сферах жизнедеятельности и последствиях их нарушения	Знание правовых принципов и норм в сфере компьютерно-опосредованной коммуникации и последствиях их нарушения, умение применять положения медиаправа; владение базовыми положениями авторского права	Неудовлетворител Не знает основ кибербезопасности, авторского права, медиаправа; не умеет ориентироваться в нормативных документах; не владеет навыками применения правовых знаний в профессиональной деятельности. Удовлетворительн Знает некоторые теоретические основы кибербезопасности, авторского права, медиаправа; умеет ориентироваться в нормативных документах; в целом не владеет навыками применения правовых знаний в профессиональной деятельности. Хорошо Знает теоретические основы кибербезопасности, авторского права, медиаправа; умеет ориентироваться в правовых документах; в целом владеет навыками применения правовых знаний в профессиональной деятельности. Отлично

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		Отлично Знает теоретические основы кибербезопасности, авторского права, медиаправа; умеет ориентироваться в правовых документах; владеет навыками применения правовых знаний в профессиональной деятельности.
УК.9.2 Ориентируется в этических нормах поведения в разных видах профессиональной деятельности и последствиях их нарушения	Знать этические нормы поведения в профессии компьютерного лингвиста; уметь ориентироваться в правовых документах в сфере интернет-коммуникации, владеть нетикетом, владеть навыками соблюдения этических норм.	Неудовлетворител Не знает этических норм поведения в профессии; не умеет применять принципы нетикета; не владеет навыками соблюдения этических норм. Удовлетворительн Знает некоторые этические нормы поведения в профессии; умеет ориентироваться в нормативных документах в сфере интернет-коммуникации,; в целом не владеет навыками соблюдения этических норм. Хорошо Знает этические нормы поведения в профессии; в целом умеет применять принципы нетикета; в целом владеет навыками соблюдения этических норм Отлично Знает этические нормы поведения в профессии; умеет применять принципы нетикета; владеет навыками соблюдения этических норм.

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Зачет

Способ проведения мероприятия промежуточной аттестации : Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

Максимальное количество баллов : 100

Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 43 до 60

«неудовлетворительно» / «незачтено» менее 43 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.3.1 демонстрирует знание основных понятий и категорий современной лингвистики УК.9.2 Ориентируется в этических нормах поведения в разных видах профессиональной деятельности и последствиях их нарушения	Обработка текстовой информации на естественном языке Защищаемое контрольное мероприятие	Знание уровней представления лингвистической информации для компьютерной обработки текста. Владение принципами моделирование поуровневого анализа и синтеза речи. Умение применять приложения для автоматического поуровневого анализа и синтеза. Знание существующих подходов к компьютерным средствам моделирования речи

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности ОПК.3.1 демонстрирует знание основных понятий и категорий современной лингвистики УК.9.2 Ориентируется в этических нормах поведения в разных видах профессиональной деятельности и последствиях их нарушения УК.9.1 Ориентируется в правовых принципах и нормах в разных сферах жизнедеятельности и последствиях их нарушения	Морфологический автоматический анализ Защищаемое контрольное мероприятие	Знание принципов и моделей морфологического анализа, умение осуществлять автоматический морфологический анализ; владение технологиями МорфАн в Python

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.2.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности ОПК.3.1 демонстрирует знание основных понятий и категорий современной лингвистики УК.9.2 Ориентируется в этических нормах поведения в разных видах профессиональной деятельности и последствиях их нарушения УК.9.1 Ориентируется в правовых принципах и нормах в разных сферах жизнедеятельности и последствиях их нарушения	Понятие информационного поиска. Классификаторы текстов Итоговое контрольное мероприятие	Умение поставить учебную задачу по классификации текстов (на любом основании) и способность решить ее с применением прикладных лингвистических средств.

Спецификация мероприятий текущего контроля

Обработка текстовой информации на естественном языке

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

Показатели оценивания	Баллы
Способен написать собственное приложение для решения задачи. Предоставил результат решения 3 задач (за каждую можно получить 5 баллов). Ошибки в коде - минус 5 баллов.	15
Может выбрать адекватные методы решения предложенной проблемы и обосновать их. За неверный выбор студент получает 0 баллов.	5
Способен сделать обзор литературы и существующих программных средств по выбранному вопросу. За отсутствие обзора литературы или средств/приложений - минус 5 баллов. 5	5
Владеет методами препроцессинга коллекции документов в соответствии с проблемой (напр, свести до лексем, получить частеречные характеристики, получить синтаксическую	5

разметку).	

Морфологический автоматический анализ

Продолжительность проведения мероприятия промежуточной аттестации: **4 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставяемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

Показатели оценивания	Баллы
Может написать блок кода, позволяющий получить частотный словник всех словоформ (5), лексем (5), частей речи (5). Если код выполняется, результат работы получен - макс, за 1 ошибку в коде снимаются по 1 баллу.	15
Может подключить библиотеку rmorph2, осуществить с помощью нее автоматический морфологический анализ. Если код выполняется, результат работы получен - макс, за 1 ошибку в коде снимаются по 1 баллу.	10
Может экспортировать словники в /csv	3
Может провести препроцессинг текста, зашумляющих знаков нет.	2

Понятие информационного поиска. Классификаторы текстов

Продолжительность проведения мероприятия промежуточной аттестации: **4 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставяемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **17**

Показатели оценивания	Баллы
Способен написать собственное приложение для решения задачи. Ошибки в коде - минус 5 баллов.	20
Способен сделать обзор литературы и существующих программных средств по выбранному вопросу. За отсутствие обзора литературы или средств/приложений - минус 5 баллов.	10
Владеет методами препроцессинга коллекции документов в соответствии с проблемой (напр, свести до лексем, получить частеречные характеристики, получить синтаксическую разметку)	5
Может выбрать адекватные методы решения предложенной проблемы и обосновать их. За неверный выбор студент получает 0 баллов. За отсутствие обоснования со ссылками на публикации - минус 5 баллов.	5