

МИНОБРНАУКИ РОССИИ

Ярославский государственный университет им. П.Г. Демидова

Кафедра вычислительных и программных систем

УТВЕРЖДАЮ

Декан факультета ИВТ

 Д.Ю. Чалый

«23» мая 2023 г.

**Рабочая программа дисциплины**  
«Автоматический анализ текстов»

**Направление подготовки**

02.04.02 Фундаментальная информатика и информационные технологии

**Направленность (профиль)**

«Искусственный интеллект и компьютерные науки»

**Квалификация выпускника**

Магистр

**Форма обучения**

очная

Программа рассмотрена  
на заседании кафедры  
от «21» апрель 2023 г.,  
протокол № 8

Программа одобрена НМК  
факультета ИВТ  
протокол №6  
от «28» апреля 2023 г.

Ярославль

### 1. Цели освоения дисциплины

Целями дисциплины «Автоматический анализ текстов» являются формирование у студентов системного представления о методах автоматической обработки естественно-языкового текста, ознакомлении студентов с теоретическими концепциями, практическими методами систем автоматической обработки естественно-языковых текстов.

### 2. Место дисциплины в структуре ОП магистратуры

Дисциплина «Автоматический анализ текстов» относится к вариативной части (дисциплина по выбору) ОП бакалавриата. В курсе рассматриваются вопросы получения специалистами компетенций в области автоматизации обработки естественных языков.

Содержание курса тесно связано фактически со всеми дисциплинами, которые изучались студентами. Освоению данной программы предшествуют учебные курсы по программированию и современным информационным технологиям.

Дисциплина «Автоматический анализ текстов» знакомит магистрантов с важнейшими областями междисциплинарных исследований на стыке информатики и лингвистики. Дисциплина способствует профессиональному росту студентов, повышению их общеметодологического уровня, а также дальнейшему развитию навыков научно-исследовательской деятельности.

### 3. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОП магистратуры

Процесс изучения дисциплины направлен на формирование следующих элементов компетенций в соответствии с ФГОС ВО, ОП ВО и приобретения следующих знаний, умений, навыков и (или) опыта деятельности:

Код компетенции	Формулировка компетенции	Перечень планируемых результатов обучения
ОПК-3	ОПК-3.1. Применяет принципы, методы и средства анализа и структурирования профессиональной информации для решения задач области создания и применения технологий и систем искусственного интеллекта ОПК-3.2. Проводит анализ современных методов и средств информатики и искусственного интеллекта для решения задач профессиональной деятельности ОПК-3.3. Анализирует профессиональную информацию, выделяет в ней главное, структурирует, оформляет и представляет в виде аналитических обзоров	Знать: общие принципы построения систем автоматической обработки текста; основные формальные модели для уровней графематического, морфологического, синтаксического и семантического анализа текстовых массивов; основные алгоритмы автоматической обработки текстовых массивов; практические возможности современных систем автоматической обработки естественного языка. Уметь: работать с системами морфологического и синтаксического анализа текста, осуществлять реферирование текста с

		<p>использованием компьютерных средств,</p> <ul style="list-style-type: none"> <li>• моделировать морфологические и синтаксические структуры текста.</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>• методами математической лингвистики для анализа текстовой информации,</li> <li>• навыками работы с автоматическими словарями</li> <li>• принципами построения систем автоматического индексирования и реферирования тестовых массивов</li> </ul>
--	--	---

#### 4. Объем, структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зач. ед., 108 акад. час.

№ п/п	Темы (разделы) дисциплины, их содержание	Семестр	Виды учебных занятий, включая самостоятельную работу студентов, и их трудоемкость (в академических часах)						Формы текущего контроля успеваемости  Форма промежуточной аттестации (по семестрам)
			Контактная работа						
			лекции	практические	лабораторные	консультации	аттестационные испытания	самостоятельная работа	
1.	Основные понятия математической лингвистики	1	2	2				12	
2.	Методы задания синтаксической структуры предложений	1	2	2				12	
3.	Принципы построения синтаксических анализаторов	1	2	2		1		12	Индивидуальное задание
4.	Теоретико-множественные модели языка	1	2	2				12	Лабораторная работа

5.	Тезаурусы и онтологии	1	1	2				12	Индивидуальное задание
6.	Модели и методы автоматической классификации и кластеризации текстовой информации	1	1	2				7,7	Лабораторная работа
7.	Автоматические системы извлечения информации	1	1	2					
8.	Лингвостатистические параметры	1	1	10		1			
<b>Всего за 1 семестр</b>			<b>12</b>	<b>24</b>		<b>2</b>	<b>0,3</b>	<b>67,7</b>	<b>Зачет</b>
<b>Всего</b>			<b>12</b>	<b>24</b>		<b>2</b>	<b>0,3</b>	<b>67,7</b>	

#### Содержание разделов дисциплины:

1. Основные понятия математической лингвистики. Устройство систем автоматизированной обработки текстов. Основные этапы построения. Графематический и морфологический анализ.
2. Методы задания синтаксической структуры предложений. Деревья подчинения. Проблемы синтаксического анализа.
3. Принципы построения синтаксических анализаторов. Применение морфологического и синтаксического анализа в поисковых системах.
4. Теоретико-множественные модели языка. Основные определения: отмеченные последовательности, контексты, дистрибутивные классы и др. Формализация понятий «часть речи» и «синтаксический тип». Формализация понятия «грамматический род». Формализация «категории падежа».
5. Тезаурусы и онтологии. WordNet. Общие принципы построения. Меры семантической близости. Частотные методы в компьютерной лингвистике.
6. Модели и методы автоматической классификации и кластеризации текстовой информации. Иерархические и вероятностные подходы. Интеллектуальный анализ данных.
7. Автоматические системы извлечения информации. Алгоритмические основы. Принципы обработки неструктурированной и плохо структурированной информации. Тематическая индексация текстов.
8. Лингвостатистические параметры. Статистические методы атрибуции. Авторский инвариант и лингвистические спектры. Применение методов кластеризации и классификации для установления авторства текстов.

#### 5. Образовательные технологии, используемые при осуществлении образовательного процесса по дисциплине

В процессе обучения используются следующие образовательные технологии:

**Лекция-беседа** или «диалог с аудиторией», является наиболее распространенной и сравнительно простой формой активного вовлечения студентов в учебный процесс. Эта лекция предполагает непосредственный контакт преподавателя с аудиторией. Преимущество лекции-беседы состоит в том, что она позволяет привлекать внимание студентов к наиболее важным вопросам темы, определять содержание и темп изложения учебного материала с учетом особенностей студентов.

**Мастер-класс** – это особая форма учебного занятия, когда преподаватель-мастер передает свой опыт путем прямого и комментированного показа последовательности действий, методов, приемов и форм педагогической деятельности. Целью проведения мастер-класса является профессиональное, интеллектуальное и эстетическое воспитание

студентов, и прежде всего, развитие в ходе мастер-класса способности студента самостоятельно и нестандартно мыслить.

**Лабораторная работа** – организация учебной работы с реальными материальными и информационными объектами, экспериментальная работа с аналоговыми моделями реальных объектов.

Практическое занятие – занятие, посвященное освоению конкретных умений и навыков и закреплению полученных на лекции знаний.

#### **6. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения и информационных справочных систем (при необходимости)**

В процессе осуществления образовательного процесса используются: для разработки документов, презентаций, для работы с электронными таблицами

OfficeStd 2013 RUS OLP NL Acdmc 021-10232

LibreOffice (свободное)

издательская система LaTeX;

– Среда разработки NetBeans 8.2 : [www.netbeans.org](http://www.netbeans.org). Доступ свободный

– OS Linux (свободная)

– для поиска учебной литературы библиотеки ЯрГУ – Автоматизированная библиотечная информационная система "БУКИ-NEXT" (АБИС "Буки-Next").

#### **7. Перечень основной и дополнительной учебной литературы, ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины**

а) основная:

1. Пентус, А. Е. Математическая теория формальных языков / Пентус А. Е. , Пентус М. Р. - Москва : Национальный Открытый Университет "ИНТУИТ", 2016. (Основы информатики и математики) - ISBN 5-9556-0062-0. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN5955600620.html>

б) дополнительная:

в) ресурсы сети «Интернет»

- Автоматическая обработка текстов на естественном языке. Учебное пособие. / Большакова, Е.И.; Клышинский, Э.С.; Ландэ, Д.В.; Носков, А.А.; Пескова, О.В.; Ягунова, Е.В. [Электронный ресурс] <https://nsu.ru/xmlui/handle/nsu/8982>

- Митренина О. В. ПРИКЛАДНАЯ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА. [Электронный ресурс] / Митренина О. В., Николаев И. С., Ландо Т. М., М: URSS, 2016. <https://elibrary.ru/item.asp?id=26365640>

Электронно-библиотечная система «Юрайт»( <https://urait.ru/> ).

Электронно-библиотечная система «Лань»(<https://e.lanbook.com/>).

#### **8. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине**

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине включает в свой состав специальные помещения:

-учебные аудитории для проведения занятий лекционного типа и практических занятий (семинаров);

- учебные аудитории для проведения групповых и индивидуальных консультаций,
- учебные аудитории для проведения текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы;
- помещения для хранения и профилактического обслуживания технических средств обучения.

Специальные помещения укомплектованы средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий, хранящиеся на электронных носителях и обеспечивающие тематические иллюстрации, соответствующие рабочим программам дисциплин.

Помещения для лабораторных занятий и самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Число посадочных мест в лекционной аудитории больше либо равно списочному составу потока, а в аудитории для практических занятий (семинаров) – списочному составу группы обучающихся.

**Автор(ы) :**

Старший преподаватель кафедры ВПС А.М.Васильев

**Приложение №1 к рабочей программе дисциплины  
«Автоматический анализ текстов»  
Фонд оценочных средств  
для проведения текущей и промежуточной аттестации студентов  
по дисциплине**

**1. Типовые контрольные задания или иные материалы, необходимые для  
оценки знаний, умений, навыков и (или) опыта деятельности,  
характеризующих этапы  
формирования компетенций**

**1.1. Контрольные задания и иные материалы, используемые в процессе  
текущей аттестации**

**Типовые индивидуальные задания**

№ п.п.	Тема
1	Проблемы автоматизации синтаксического анализа предложений.
2	Проблемы обнаружения кореференций и анафор в текстах на ЕЯ.
3	Применение алгоритмов и методов обработки текстовой информации в технике.
4	Применение алгоритмов и методов обработки текстовой информации в медицине.
5	Применение алгоритмов и методов обработки текстовой информации в системах безопасности.
6	Возможности программных приложений для анализа социальных сетей.
7	Проблемы автоматической идентификации авторов текстов.
8	Методы автоматического построения онтологий.
9	Сравнение алгоритмов обнаружения и исправления ошибок и опечаток.
10	Сравнение алгоритмов морфологического анализа.

Задание выполняется письменно в виде реферата.

Требования к выполнению индивидуального задания:

1. Наличие авторской позиции, самостоятельность суждений.

2. Степень раскрытия сущности проблемы

- соответствие плана теме реферата;

- соответствие содержания теме и плану реферата;

- владение терминологией и понятийным аппаратом проблемы;

- полнота и глубина раскрытия основных понятий проблемы;

- обоснованность способов и методов работы с материалом;

- умение работать с литературой, систематизировать и структурировать материал;

- умение обобщать, сопоставлять различные точки зрения по рассматриваемому

вопросу, аргументировать основные положения и выводы

3. Уровень выбора источников

- обоснованность,

- полнота использования,

- современность,

- научность;

4. Соблюдение требований к оформлению, грамотность:

- правильное оформление ссылок на используемую литературу;

- соблюдение требований к объему реферата (от 10 до 15 страниц);
- культура оформления: выделение абзацев
- грамотность и культура изложения;
- отсутствие орфографических и синтаксических ошибок,
- отсутствие стилистических погрешностей;
- отсутствие опечаток, сокращений слов, кроме общепринятых;
- научно-литературный стиль.

Работа, в которой дословно переписаны текст учебника, пособия или аналогичная работа, защищенная ранее другим студентом, не оценивается, а тема заменяется на новую.

#### Критерии оценивания

<b>Оценка</b>	<b>Критерии</b>
Отлично Уровень формирования компетенций: высокий	Выполнены все требования к заданию. Обозначена проблема и обоснована её актуальность, сделан анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция; рассмотрены примеры использования разбираемых методов и подходов к решению задачи, сформулированы выводы, тема раскрыта полностью Анализирует изученный материал, выделяет наиболее значимые для раскрытия темы факты, научные положения, соблюдает логическую последовательность в изложении материала
Хорошо Уровень формирования компетенций: продвинутый	Выполнены все требования к заданию. Обозначена проблема и обоснована её актуальность, сделан анализ хотя бы одной точки зрения на рассматриваемую проблему, изложена собственная позиция, но возможно нарушение логики изложения; рассмотрены примеры использования разбираемых методов и подходов к решению задачи, но неполно, сформулированы выводы, тема раскрыта, но не полностью Знает изученный материал, грамотно излагает факты, научные положения, соблюдает логическую последовательность в изложении материала с некоторыми неточностями
Удовлетворительно Уровень формирования компетенций: пороговый	Выполнена большая часть требований. Проблема обозначена без обоснования её актуальности, факты, методы, алгоритмы изложены частично и без анализа; рассмотрен пример использования разбираемых методов и подходов к решению задачи, но неполно или с ошибками Знает основной материал, но не усвоил его деталей, допускает неточности, нарушения логической последовательности в изложении
Неудовлетворительно	Не выполнены требования. Тема задания не раскрыта, обнаруживается существенное непонимание проблемы или задание не выполнено Не знает значительной части материала задания, допускает существенные ошибки



## Примеры лабораторных работ

### 1. Реализовать алгоритм выделения ключевых слов:

- TextRank — это алгоритм без обучения, основанный на графовых методах. Его основная идея - построить граф на основе входного текста, где вершинами будут кандидаты в ключевые слова, а ребрами — совместные появления слов в тексте, и ранжировать вершины с помощью специального графового алгоритма. Чем выше ранг у соседей данной вершины, тем выше будет ранг самой вершины.
- Topical PageRank — это алгоритм без обучения, похожий на TextRank. Также как и TextRank, данный метод строит граф кандидатов в ключевые слова, но использует улучшенный метод ранжирования. Основное различие между этими двумя алгоритмами заключается в том, что Topical PageRank перед стадией ранжирования вершин графа находит темы входного текста методом латентного размещения Дирихле. Данный шаг обеспечивает то, что выделенные ключевые слова будут соответствовать нужным темам. Далее алгоритм выполняет метод TextRank для каждой найденной темы.
- Кеа - это алгоритм с обучением для выделения ключевых слов из текстов. Сначала алгоритм находит набор ключевых слов-кандидатов с помощью лексических методов и вычисляет несколько статистических характеристик для каждого кандидата. Затем Кеа строит модель предсказаний для тренировочных текстов, у которых уже есть выделенные вручную ключевые слова. На последнем шаге алгоритм применяет метод наивного Байеса для определения ключевых слов тестовых документов. Также Кеа может использовать тезаурус для выбора ключевых слов из заданного словаря.
- Маui основывается на Кеа и выполняет те же шаги для выделения ключевых слов. Данные методы различаются тем, что Маui вычисляет больше статистических характеристик слов-кандидатов и применяет деревья решений вместо алгоритма наивного Байеса.

### 2. Определение характеристик качества выделения ключевых слов: точность, полнота и F-статистика.

#### 3. Построение и работа с тезаурусом. Виды связей в тезаурусе (Thesaurus construction and use: a practical manual, 2000):

1. эквивалентные: синонимы, лексические варианты, квазисинонимы
2. иерархические: гипонимы, гиперонимы, часть-целое
3. ассоциации

Методы, выделяющие конкретные связи: синонимы, гиперонимы

1. Синтаксические, использующие существующий тезаурус.  
Detection of synonymy links between terms: experiment and results (2001)  
Automatic Acquisition of Synonym Resources and Assessment of their Impact on the Enhanced Search in EHRs (2009)

- Projecting Corpus-Based Semantic Links on a Thesaurus (1999)  
 Learning syntactic patterns for automatic hypernym discovery (2004)
2. Синтаксические, не использующие существующий тезаурус.  
 Extracting Hyponymic Relations from Chinese Free Corpus (2006)  
 Building a hyponymy lexicon with hierarchical structure (2002)
  3. Методы, выделяющие несколько различных типов связей на основе страниц сайтов  
 Building a Web Thesaurus from Web Link Structure (2003)  
 Wikipedia Mining for an Association Web Thesaurus Construction (2007)
- Методы, выделяющие ассоциативные связи на основе меры схожести
    1. Статистические  
 Метод автоматического построения тезаурусов на основе статистической обработки текстов на естественном языке (2012)  
 Automatic construction of networks of concepts characterizing document databases (1992)  
 Construction of a dynamic Thesaurus and its use for associated information retrieval (1989)
    2. Методы кластеризации  
 An approach to the automatic construction of global thesauri (1990)
    3. Синтаксические  
 Explorations in Automatic Thesaurus Discovery (1994)
    4. Комбинированные  
 Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy (1997) - статистический + синтаксический

#### Критерии оценивания

Оценка	Критерии
Отлично Уровень формирования компетенций: высокий	Знает и применяет подходящие для решения алгоритмы, выбирает наиболее эффективный алгоритм. Создает полноценное приложение в среде разработки. Поясняет код и изменяет его при необходимости. Анализирует изученный материал, выделяет наиболее значимые для решения задачи факты, научные положения, соблюдает логическую последовательность в выполнении работы
Хорошо Уровень формирования компетенций: продвинутый	Знает и применяет подходящие для решения алгоритмы, выбирает наиболее эффективный алгоритм. Создает полноценное приложение в среде разработки. Поясняет код и изменяет его при необходимости с небольшими неточностями. Выделяет подходящие для решения задачи факты, научные положения, соблюдает логическую последовательность в выполнении работы с небольшими неточностями
Удовлетворительно Уровень формирования компетенций: пороговый	Знает и применяет подходящий для решения алгоритм, но с некоторыми ошибками. Создает полноценное приложение в среде разработки. С трудом поясняет код, не может изменить код при усложнении или существенном дополнении задачи.

	Выделяет подходящие для решения задачи факты, но нарушает логическую последовательность в выполнении работы
Неудовлетворительно	Не может подобрать и реализовать подходящий для решения алгоритм. Не может создать полноценное приложение в среде разработки. Не может пояснить и изменить код. Не знает материал темы задания, нарушает логическую последовательность в выполнении работы

### Список вопросов к зачету:

1. Перечислить направления компьютерной лингвистики.
2. Сформулировать общие принципы построения автоматизированных систем обработки текстов.
3. Разъяснить принципы работы графематического и морфологического анализаторов.
4. Перечислить методы задания синтаксической структуры предложений.
5. Разъяснить принципы работы фрагментационного и синтаксического анализаторов. Описать принцип их взаимодействия.
6. Изложить основные идеи семантического анализа.
7. Привести примеры мер семантической близости.
8. Дать определения отмеченных последовательностей, контекста, дистрибутивных классов.
9. Дать формальные определения частей речи, грамматического рода и категории падежа в терминах модели языка.
10. Суть теоретико-модельного подхода к исследованию семантики текстов.
11. Сформулировать принципы построения тезаурусов и онтологий. Сходства и отличия.
12. Применение частотных методов в компьютерной лингвистике. Перечислить,
13. описать, привести примеры.
14. В чем отличие между классификацией и кластеризацией текстов?
15. Перечислить методы классификации и кластеризации текстовой информации. Сформулировать основные принципы.
16. Разъяснить принципы работы автоматических систем извлечения информации.
17. Сформулировать принципы обработки неструктурированной и плохо структурированной информации. Индексация текстов.
18. Перечислить формальные методы атрибуции текстов.
19. Дать определения лингвостатистических параметров, авторского инварианта и лингвистических спектров.
20. Привести примеры использования методов кластеризации и классификации для определения авторства текстов.

### Методические указания по выставлению зачета

Каждый студент в течение семестра выполняет одно индивидуальное задание и одну лабораторную работу. Зачет выставляется по результатам выполнения индивидуального задания и лабораторной работы на оценку не ниже удовлетворительно. Задания могут быть сданы в течение семестра последовательно в процессе освоения материала или на зачете. В случае необходимости преподаватель в ходе сдачи задания может провести беседу по вопросам к зачету, связанным в первую очередь с тематикой индивидуального задания.

## **2. Перечень компетенций, этапы их формирования, описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкалы оценивания**

### **2.1. Шкала оценивания сформированности компетенций и ее описание**

Оценивание уровня сформированности компетенций в процессе освоения дисциплины осуществляется по следующей трехуровневой шкале:

*Пороговый уровень* - предполагает отражение тех ожидаемых результатов, которые определяют минимальный набор знаний и (или) умений и (или) навыков, полученных студентом в результате освоения дисциплины. Пороговый уровень является обязательным уровнем для студента к моменту завершения им освоения данной дисциплины.

*Продвинутый уровень* - предполагает способность студента использовать знания, умения, навыки и (или) опыт деятельности, полученные при освоении дисциплины, для решения профессиональных задач. Продвинутый уровень превосходит пороговый уровень по нескольким существенным признакам.

*Высокий уровень* - предполагает способность студента использовать потенциал интегрированных знаний, умений, навыков и (или) опыта деятельности, полученных при освоении дисциплины, для творческого решения профессиональных задач и самостоятельного поиска новых подходов в их решении путем комбинирования и использования известных способов решения применительно к конкретным условиям. Высокий уровень превосходит пороговый уровень по всем существенным признакам.

## 2.2. Перечень компетенций, этапы их формирования, описание показателей и критериев оценивания компетенций на различных этапах их формирования

Код компетенции	Форма контроля	Этапы формирования (№ темы (раздела))	Показатели оценивания	Шкала и критерии оценивания компетенций на различных этапах их формирования		
				Пороговый уровень	Продвинутый уровень	Высокий уровень
ОПК-3	Индивидуальные задания Лабораторная работа. Зачет.	1-8	<p>Знать:</p> <ul style="list-style-type: none"> <li>• общие принципы построения систем автоматической обработки текста;</li> <li>• основные формальные модели для уровней графематического, морфологического, синтаксического и семантического анализа текстовых массивов;</li> <li>• основные алгоритмы автоматической обработки текстовых массивов;</li> <li>• практические возможности современных систем автоматической</li> </ul>	<p>обучающийся должен:</p> <p>а) <i>знать</i>:</p> <ul style="list-style-type: none"> <li>• основные цели и задачи компьютерной лингвистики;</li> <li>• основные методы и подходы к автоматической обработке текстов (правилловые, статистические);</li> <li>• основные классы приложений, развиваемых на базе компьютерной лингвистики (информационный поиск, автоматический и автоматизированный перевод текстов);</li> <li>• основные классы цифровых</li> </ul>	<p>обучающийся должен:</p> <p>а) <i>знать</i>:</p> <ul style="list-style-type: none"> <li>• основные цели и задачи компьютерной лингвистики;</li> <li>• основные методы и подходы к автоматической обработке текстов (правилловые, статистические, в т.ч. машинное обучение, гибридные);</li> <li>• основные классы приложений, развиваемых на базе компьютерной лингвистики (информационный поиск, глубокий анализ данных, автоматический и автоматизированный перевод текстов с одного языка на другой, автоматическое аннотирование и реферирование</li> </ul>	<p>обучающийся должен:</p> <p>а) <i>знать</i>:</p> <ul style="list-style-type: none"> <li>• основные цели и задачи компьютерной лингвистики;</li> <li>• основные методы и подходы к автоматической обработке текстов (правилловые, статистические, в т.ч. машинное обучение, гибридные);</li> <li>• основные классы приложений, развиваемых на базе компьютерной лингвистики (информационный поиск, глубокий анализ данных, автоматический и автоматизированный перевод текстов с одного языка на другой, автоматическое аннотирование и реферирование</li> </ul>

		<p>обработки естественного языка.</p> <p>Уметь:</p> <ul style="list-style-type: none"> <li>• работать с системами морфологического и синтаксического анализа текста, осуществлять реферирование текста с использованием компьютерных средств, моделировать морфологические и синтаксические структуры текста.</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>• методами математической лингвистики для анализа текстовой информации, навыками работы с автоматическими словарями принципами построения систем автоматического индексирования и</li> </ul>	<p>лингвистических ресурсов, создаваемых методами компьютерной лингвистики (компьютерные словари);</p> <p>б) уметь:</p> <ul style="list-style-type: none"> <li>• строить базовые правила систем автоматической обработки текстов; строить базовые морфологические и синтаксические структуры предложения (на примере русского и английского языков);</li> </ul> <p>в) владеть:</p> <ul style="list-style-type: none"> <li>• принципами использования электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов и онтологий).</li> </ul>	<p>другой, автоматическое аннотирование и реферирование документов);</p> <ul style="list-style-type: none"> <li>• основные классы цифровых лингвистических ресурсов, создаваемых методами компьютерной лингвистики (компьютерные одноязычные и многоязычные словари);</li> </ul> <p>б) уметь:</p> <ul style="list-style-type: none"> <li>• строить базовые правила систем автоматической обработки текстов;</li> <li>• разбираться в правилах и алгоритмах автоматической обработки текстов;</li> <li>• строить базовые морфологические и синтаксические структуры предложения (на примере русского и английского языков);</li> </ul> <p>в) владеть:</p>	<p>документов, анализ тональности текста, человеко-машинное общение на естественном языке);</p> <ul style="list-style-type: none"> <li>• основные классы цифровых лингвистических ресурсов, создаваемых методами компьютерной лингвистики (компьютерные одноязычные и многоязычные словари, аннотированные корпуса текстов);</li> </ul> <p>б) уметь:</p> <ul style="list-style-type: none"> <li>• строить базовые правила систем автоматической обработки текстов; разбираться в правилах и алгоритмах автоматической обработки текстов; строить базовые морфологические и синтаксические структуры предложения (на примере русского и английского языков); использовать лингвистические технологии для</li> </ul>
--	--	--	--	--	--

			реферирования тестовых массивов.		<ul style="list-style-type: none"> <li>• навыком освоения большого объема информации;</li> <li>• принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов и онтологий) и умение пользоваться такими ресурсами.</li> </ul>	<p>проектирования систем автоматической обработки письменного текста на естественном языке</p> <p><i>в) владеть:</i></p> <ul style="list-style-type: none"> <li>• навыком освоения большого объема информации;</li> <li>• принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов и онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умение пользоваться такими ресурсами;</li> <li>• навыками постановки научно-исследовательских задач и навыками самостоятельной работы.</li> </ul>
ОПК-3	Индивидуальные задания Лабораторная работа. Зачет.	1-8	<p>Знать: методику проведения исследований.</p> <p>Уметь: выполнять исследовательские эксперименты</p> <p>Владеть:</p>	<p>обучающийся должен:</p> <p><i>а) знать:</i></p> <ul style="list-style-type: none"> <li>• основные методы и подходы к автоматической обработке текстов (правилые, статистические);</li> </ul>	<p>обучающийся должен:</p> <p><i>а) знать:</i></p> <ul style="list-style-type: none"> <li>• основные методы и подходы к автоматической обработке текстов (правилые, статистические);</li> </ul>	<p>обучающийся должен:</p> <p><i>а) знать:</i></p> <ul style="list-style-type: none"> <li>• различные методы и подходы к автоматической обработке текстов (правилые, статистические, в т.ч. машинное обучение, гибридные);</li> </ul>

			<ul style="list-style-type: none"> <li>• методами математической лингвистики для анализа текстовой информации</li> </ul>	<p><i>б) уметь:</i></p> <ul style="list-style-type: none"> <li>• использовать лингвистические технологии для проектирования систем автоматической обработки письменного текста на естественном языке</li> </ul> <p><i>в) владеть:</i></p> <ul style="list-style-type: none"> <li>• навыками самостоятельной работы.</li> </ul>	<p><i>б) уметь:</i></p> <ul style="list-style-type: none"> <li>• использовать лингвистические технологии для проектирования систем автоматической обработки письменного текста на естественном языке</li> <li>• проводить эксперименты по обработке текста</li> </ul> <p><i>в) владеть:</i></p> <ul style="list-style-type: none"> <li>• навыками постановки научно-исследовательских задач и навыками самостоятельной работы.</li> </ul>	<p><i>б) уметь:</i></p> <ul style="list-style-type: none"> <li>• использовать лингвистические технологии для проектирования систем автоматической обработки письменного текста на естественном языке</li> <li>• проводить эксперименты по обработке текста</li> <li>• рассчитывать статистические характеристики эксперимента</li> </ul> <p><i>в) владеть:</i></p> <ul style="list-style-type: none"> <li>• навыками постановки научно-исследовательских задач и навыками самостоятельной работы.</li> </ul>
--	--	--	--	--	---	--



### **3. Методические рекомендации преподавателю по процедуре оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций**

Целью процедуры оценивания является определение степени овладения студентом ожидаемыми результатами обучения (знаниями, умениями, навыками и (или) опытом деятельности).

Процедура оценивания степени овладения студентом ожидаемыми результатами обучения осуществляется с помощью методических материалов, представленных в разделе

«Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций»

#### **3.1 Критерии оценивания степени овладения знаниями, умениями, навыками и (или) опытом деятельности, определяющие уровни сформированности компетенций**

Пороговый уровень (общие характеристики):

- владение основным объемом знаний по программе дисциплины;
- знание основной терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы без существенных ошибок;
- владение инструментарием дисциплины, умение его использовать в решении стандартных (типовых) задач;
- способность самостоятельно применять типовые решения в рамках рабочей программы дисциплины;
- усвоение основной литературы, рекомендованной рабочей программой дисциплины;
- знание базовых теорий, концепций и направлений по изучаемой дисциплине;
- самостоятельная работа на практических и лабораторных занятиях, периодическое участие в групповых обсуждениях, достаточный уровень культуры исполнения заданий.

Продвинутый уровень (общие характеристики):

- достаточно полные и систематизированные знания в объеме программы дисциплины;
- использование основной терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать выводы;
- владение инструментарием дисциплины, умение его использовать в решении учебных и профессиональных задач;
- способность самостоятельно решать сложные задачи (проблемы) в рамках рабочей программы дисциплины;
- усвоение основной и дополнительной литературы, рекомендованной рабочей программой дисциплины;
- умение ориентироваться в базовых теориях, концепциях и направлениях по изучаемой дисциплине и давать им сравнительную оценку;
- самостоятельная работа на практических и лабораторных занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Высокий уровень (общие характеристики):

- систематизированные, глубокие и полные знания по всем разделам дисциплины;
- точное использование терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;

- безупречное владение инструментарием дисциплины, умение его использовать в постановке и решении научных и профессиональных задач;
- способность самостоятельно и творчески решать сложные задачи (проблемы) в рамках рабочей программы дисциплины;
- полное и глубокое усвоение основной и дополнительной литературы, рекомендованной рабочей программой дисциплины;
- умение ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине и давать им критическую оценку;
- активная самостоятельная работа на практических и лабораторных занятиях, творческое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

### **3.2 Описание процедуры выставления оценки**

В зависимости от уровня сформированности каждой компетенции по окончании освоения дисциплины студенту выставляется оценка «зачтено», «незачтено».

Показатели и критерии, используемые при выставлении оценки подробно описаны в разделе «Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций».

Высокий уровень формирования компетенций соответствует оценке «отлично» за индивидуальное задание и лабораторную работу.

Продвинутый уровень формирования компетенций соответствует оценке «хорошо» за индивидуальное задание и лабораторную работу.

Пороговый уровень формирования компетенций соответствует оценке «удовлетворительно» за индивидуальное задание и лабораторную работу.

Оценка «отлично» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована на высоком уровне.

Оценка «хорошо» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована не ниже, чем на продвинутом уровне.

Оценка «удовлетворительно» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована не ниже, чем на пороговом уровне.

Оценка «неудовлетворительно» выставляется студенту, у которого хотя бы одна компетенция (полностью или частично формируемая данной дисциплиной) сформирована ниже, чем на пороговом уровне.

Оценка «зачет» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована не ниже, чем на пороговом уровне.

Оценка «незачтено» выставляется студенту, у которого хотя бы одна компетенция (полностью или частично формируемая данной дисциплиной) сформирована ниже, чем на пороговом уровне.

## **Приложение №2 к рабочей программе дисциплины «Автоматический анализ текстов»**

### **Методические указания для студентов по освоению дисциплины**

Занятия проводятся в интерактивной форме с использованием мультимедиа-технологий. Занятия предполагают наличие дискуссий по поводу тех или иных вопросов разработки программных приложений осуществляемых в результате соответствующего предложения преподавателя.

Практическое применение полученных знаний отрабатывается и во время лабораторных занятий, ориентированных помимо закрепления лекционного материала на разбор различных модельных ситуаций, характерных для разработки современных программных продуктов. Для успешного освоения дисциплины очень важно решение достаточно большого количества задач, как в аудитории, так и самостоятельно в качестве домашних заданий. В основном такими задачами являются лабораторные работы различного объема. Примеры решения задач разбираются на занятиях. Для решения всех задач необходимо знать и понимать лекционный материал. Поэтому в процессе изучения дисциплины рекомендуется регулярное повторение пройденного лекционного материала. Материал, законспектированный на лекциях, необходимо дома еще раз проработать и при необходимости дополнять информацией, полученной на консультациях, практических занятиях или из учебной литературы.

Текущий контроль успеваемости осуществляется в форме опросов по основным понятиям и концепциям курса и защиты реферата по заданной теме, осуществляемый в ходе лабораторных занятий. Для самостоятельной работы студентам предлагается разрабатывать и реализовывать алгоритмы создания компьютерных систем с применением разобранных во время лекций и лабораторных занятий подходов и методик. Окончательная аттестация осуществляется в форме зачета, основную часть которого составляют результаты лабораторных работ, а также собеседование по тематике курса.

### **Учебно-методическое обеспечение самостоятельной работы студентов по дисциплине**

Для самостоятельной работы особенно рекомендуется использовать учебную литературу, указанную в разделе № 7 данной рабочей программы.

Также для подбора учебной литературы рекомендуется использовать широкий спектр интернет-ресурсов:

1. Электронно-библиотечная система «Университетская библиотека online» ([www.biblioclub.ru](http://www.biblioclub.ru)) - электронная библиотека, обеспечивающая доступ к наиболее востребованным материалам-первоисточникам, учебной, научной и художественной литературе ведущих издательств (\*регистрация в электронной библиотеке – только в сети университета. После регистрации работа с системой возможна с любой точки доступа в Internet.).

2. Для самостоятельного подбора литературы в библиотеке ЯрГУ рекомендуется использовать:

1. Личный кабинет ([http://lib.uniyar.ac.ru/opac/bk\\_login.php](http://lib.uniyar.ac.ru/opac/bk_login.php)) дает возможность получения on-line доступа к списку выданной в автоматизированном режиме литературы, просмотра и копирования электронных версий изданий сотрудников университета (учеб. и метод. пособия, тексты лекций и т.д.) Для работы в «Личном кабинете» необходимо зайти на сайт Научной библиотеки ЯрГУ с любой точки, имеющей доступ в Internet, в пункт меню

«Электронный каталог»; пройти процедуру авторизации, выбрав вкладку «Авторизация», и заполнить представленные поля информации.

2. Электронная библиотека учебных материалов ЯрГУ ([http://www.lib.uniyar.ac.ru/opac/bk\\_cat\\_find.php](http://www.lib.uniyar.ac.ru/opac/bk_cat_find.php)) содержит более 2500 полных текстов учебных и учебно-методических материалов по основным изучаемым

дисциплинам, изданных в университете. Доступ в сети университета, либо по логину/паролю.

3. Электронная картотека [«Книгообеспеченность»](#) ([http://www.lib.uniya.ac.ru/opac/bk\\_bookreq\\_find.php](http://www.lib.uniya.ac.ru/opac/bk_bookreq_find.php)) раскрывает учебный фонд научной библиотеки ЯрГУ, предоставляет оперативную информацию о состоянии книгообеспеченности дисциплин основной и дополнительной литературой, а также цикла дисциплин и специальностей. Электронная картотека [«Книгообеспеченность»](#) доступна в сети университета и через Личный кабинет.