

МИНОБРНАУКИ РОССИИ



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

ИНСТИТУТ ЛИНГВИСТИКИ
Учебно-научный центр компьютерной лингвистики

Лингвистические основы машинного перевода

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.04.03 Фундаментальная и прикладная лингвистика

Код и наименование направления подготовки/специальности

Магистерская программа: Фундаментальная и компьютерная лингвистика

Наименование направленности (профиля)/ специализации

Уровень высшего образования: **магистратура**

Форма обучения: **очная**

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2024

Лингвистические основы машинного перевода
Рабочая программа дисциплины

Составитель(и):

к. филол. н., профессор Л.Л.Иомдин

Ответственный редактор:

д. филол. н., профессор В.И.Подлеская

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной лингвистики
№ 5 от 26 марта 2024 г.

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 оценки по дисциплине.....	
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».	
6.3	Профессиональные базы данных и информационно-справочные системы.....	
7.	Материально-техническое обеспечение дисциплины	
8.	Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов.....	
9.	Методические материалы.....	
9.1	Планы семинарских/ практических/ лабораторных занятий	
9.2	Методические рекомендации по подготовке письменных работ	
9.3	Иные материалы.....	

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Предметом дисциплины является система описания языка с особенностями, присущими такому описанию, когда оно осуществляется в целях автоматической обработки текстов, в первую очередь машинного перевода, основанного на правилах. Основной внимание уделяется 1) тому, как текст на входном языке анализируется на разных уровнях (морфологическом, синтаксическом, семантическом), как соотносятся грамматическое описание языка для целей автоматической обработки текста и словарь языка, 2) тому, как осуществляется преобразование глубинной структуры входного языка в структуру выходного языка и 3) тому, как синтезируется текст на выходном языке. Курс предполагает интенсивную практическую работу магистрантов, которые должны освоить основные методы и подходы, применяемые лингвистами, разрабатывающими прикладные системы автоматической обработки текстов, и научиться самостоятельно создавать адекватные правила такой обработки.

Курс направлен на решение следующих задач:

- дать представление о месте теоретической лингвистики в задачах, решаемых компьютерной лингвистикой;
- познакомить магистрантов с важнейшими областями междисциплинарных исследований на стыке лингвистики со смежными дисциплинами, в первую очередь с компьютерной наукой;
- научить магистрантов обнаруживать типологические сходства и различия естественных языков, между которыми осуществляется машинный перевод, и творчески использовать полученные знания;
- научить магистрантов пользоваться методами обратной связи, т.е. применять полученные при разработке автоматических систем результаты для извлечения новых знаний о естественном языке;
- познакомить магистрантов с современными подходами к решению задачи машинного перевода, в том числе с гибридными и статистическими подходами и приемами машинного обучения.

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

Компетенция	Индикаторы компетенций	Результаты обучения
ПК-2 Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	ПК-2.1 Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных	Знать: <ul style="list-style-type: none"> – основные принципы обработки информации; – базовые принципы корпусной лингвистики, лексикографии, математической статистики; – базовые представления о языковом разнообразии; – наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных;

		<p>Уметь:</p> <ul style="list-style-type: none"> – пользоваться лингвистическими корпусами, электронными словарями и базами данных; – применять основные принципы корпусной лингвистики, лексикографии, математической статистики; <p>Владеть:</p> <ul style="list-style-type: none"> – принципами создания лингвистических корпусов, электронных словарей и баз данных.
<p>ПК-3 Способен использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем</p>	<p>ПК-3.1 Знает основные системы автоматической обработки звучащей речи и текстов на естественном языке; базовые принципы автоматической обработки языковых данных; основные интеллектуальные и информационные электронные системы и принципы работы с ними</p>	<p>Знать:</p> <ul style="list-style-type: none"> – основные системы автоматической обработки звучащей речи и текстов на естественном языке; – базовые принципы автоматической обработки языковых данных; – основные интеллектуальные и информационные электронные системы и принципы работы с ними; <p>Уметь:</p> <ul style="list-style-type: none"> – использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем; <p>Владеть:</p> <ul style="list-style-type: none"> – навыками автоматической обработки языковых данных.

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Лингвистические основы машинного перевода» относится к части, формируемой участниками образовательных отношений.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: Лингвистическая и онтологическая семантика.

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических часов.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
1	Лекции	4
1	Практические занятия	16
2	Лекции	4
2	Практические занятия	16
Всего:		40

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 68 академических часов.

3. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание
1	Лингвистика как наука о языке.	<ul style="list-style-type: none"> • Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика; • Лингвистика и прагматика.
2	Лингвистическое моделирование.	<ul style="list-style-type: none"> • Действующие модели языка; • Теория «Смысл – Текст» и возможности ее использования для построения систем автоматической обработки текста.
3	Грамматика и словарь естественного языка.	<ul style="list-style-type: none"> • Представление об интегральном описании языка.
4.	Краткий обзор формальных грамматик.	<ul style="list-style-type: none"> • Порождающие грамматики; • Грамматики составляющих и грамматики зависимостей; • Гибридные грамматики.
5.	Автоматический анализ и синтез текста.	<ul style="list-style-type: none"> • Морфологический анализ текста.
6.	Синтаксический анализ текста (парсинг).	<ul style="list-style-type: none"> • Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх»; • Правилые и статистические подходы к автоматической обработке текста.
7.	Языковая неоднозначность	<ul style="list-style-type: none"> • Типы неоднозначности;

	как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста.	<ul style="list-style-type: none"> • Интерактивное разрешение лексической и синтаксической неоднозначности; • Применение онтологии для разрешения неоднозначности; • Методы машинного обучения в применении к разрешению однозначности.
8.	Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке.	<ul style="list-style-type: none"> • Система машинного перевода как механизм обратной связи и источник новых лингвистических знаний.
9.	Типы систем машинного перевода.	<ul style="list-style-type: none"> • Автоматический и автоматизированный перевод; • Память переводов; • Интерлингва; • UNL (универсальный сетевой язык) как тип интерлингвы и его использование в задаче машинного перевода.
10.	Морфологический компонент системы автоматической обработки текстов.	<ul style="list-style-type: none"> • Морфологическая структура слова и предложения.
11	Синтаксический компонент системы автоматической обработки текстов.	<ul style="list-style-type: none"> • Синтаксическая структура предложения; • Алгоритм синтаксического анализа; • Синтаксические отношения. Синтагмы
12	Словарь системы автоматической обработки текстов.	<ul style="list-style-type: none"> • Словарь системы машинного перевода; • Структура словарной статьи; • Синтаксические признаки; • Семантические признаки (дескрипторы); • Теория валентностей; • Модель управления.
13	Правила межъязыкового перевода	<ul style="list-style-type: none"> • Межъязыкового перевод в узком смысле (трансфер).
14	Лексические функции в машинном переводе.	<ul style="list-style-type: none"> • Лексические функции в машинном переводе.
15	Краткий обзор действующих систем машинного перевода.	<ul style="list-style-type: none"> • Статистические системы машинного перевода.
16	Краткий обзор смежных задач прикладной лингвистики.	<ul style="list-style-type: none"> • Информационный поиск; • Интеллектуальный анализа данных; Синонимическое перифразирование высказываний и его прикладное значение.
17	Некоторые современные цифровые лингвистические ресурсы и их роль в задачах автоматической обработки текстов.	<ul style="list-style-type: none"> • Аннотированные корпуса текстов. Word Net, Frame Net, Treebanks (включая SynTagRus), Semantic Web).

4. Образовательные технологии

Для проведения учебных занятий по дисциплине используются различные образовательные технологии. Для организации учебного процесса может быть использовано электронное обучение и (или) дистанционные образовательные технологии.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- домашние задания	5 баллов	30 баллов
- выполнение заданий на семинаре	5 баллов	10 баллов
- участие в соревновании	20 баллов	20 баллов
Промежуточная аттестация – зачет		40 баллов
Итого за семестр		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	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>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.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ C	хорошо/ зачтено	<p>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.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p>

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	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

В качестве домашних заданий предлагаются задания следующих типов

Реферирование научного сочинения из рекомендованного списка

Выполнение практического задания: работа с конкретной системой машинного перевода ЭТАП-3. Наблюдение за алгоритмом синтаксического анализа.

Выполнение практического задания: работа с конкретной системой машинного перевода ЭТАП-3. Наблюдение за решением задачи разрешения неоднозначностей разных типов.

Выполнение практического задания: работа с системой UNL.

Выполнение практического задания: построение морфологических структур предложений разных типов на примере различных естественных языков.

Выполнение практических заданий: 1) ручное построение синтаксических структур предложений разных типов на примере различных естественных языков, 2) самостоятельное построение синтаксических правил на специальном формальном языке.

Выполнение практического задания: создание словарных статей словарей разных языков для машинного перевода.

Выполнение практического задания: написание пробных правил трансфера для перевода с русского языка на английский и обратно.

Выполнение практического задания: работа с базой данных лексических функций.

Выполнение практического задания: работа с доступными в Интернете статистическими системами перевода и критический анализ систем лексических функций.

Выполнение практического задания: выработка навыков пользования конкретными цифровыми лингвистическими ресурсами.

Зачет и экзамен ориентированы на следующие контрольные вопросы

1. Действующая модель языка «Смысл – Текст»: основные положения.
2. Определение предмета лингвистических дисциплин - фонетики, морфологии, синтаксиса, семантики – в свете подхода «Смысл – Текст»
3. Интегральное описание языка в смысле Ю.Д.Апресяна – основные положения.
4. Грамматики составляющих и грамматики зависимостей в свете задачи автоматической обработки текстов.
5. Гибридные грамматики и возможности их компьютерной реализации.
6. Морфологический анализ текста. Морфологическая структура слова и предложения.
7. Синтаксический анализ текста. Правильные и статистические подходы к автоматической обработке текста.
8. Синтаксическая структура предложения. Синтаксические отношения. Синтагмы.
9. Языковая неоднозначность и методы ее разрешения при автоматической обработке текста. Типы неоднозначности. Интерактивное разрешение лексической и синтаксической неоднозначности.
10. Типы систем машинного перевода. Автоматический и автоматизированный перевод. Память переводов. Интерлингва.
11. UNL (универсальный сетевой язык) как тип интерлингвы и его использование в задаче машинного перевода.
12. Словарь системы машинного перевода. Структура словарной статьи.
13. Синтаксические признаки.
14. Семантические признаки (дескрипторы).
15. Теория валентностей. Модель управления.
16. Правила межъязыкового перевода в узком смысле (трансфер).
17. Лексические функции в машинном переводе.
18. Статистические системы машинного перевода: общий подход.
19. Некоторые современные цифровые лингвистические ресурсы и их роль в задачах автоматической обработки текстов. Аннотированные корпуса текстов. Word Net, Frame Net, Treebanks (включая SynTagRus), Semantic Web).

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Основная литература

1. Bahdanau D., Cho K., Bengio Y., Neural machine translation by Jointly learning to align and translate. 2015 <https://arxiv.org/pdf/1409.0473.pdf>
2. Birch A. et al. HUME: Human UCCA-Based Evaluation of Machine Translation 2016 <https://arxiv.org/pdf/1607.00030.pdf>
3. Denkowski M., Lavie A., Meteor Universal: Language Specific Translation Evaluation for Any Target Language – Workshop on statistical machine translation, 2014 <http://www.aclweb.org/anthology/W14-3348>
4. Koehn et al. Moses: open source toolkit for statistical machine translation. 2007 <https://aclanthology.info/pdf/P/P07/P07-2045.pdf>
5. Papineni K. et al. BLEU: a method for automatic Evaluation of machine translation. 2002 <https://aclanthology.info/pdf/P/P02/P02-1040.pdf>

6. Papineni K. et al. BLEU: a method for automatic Evaluation of machine translation. 2002 <https://aclanthology.info/pdf/P/P02/P02-1040.pdf>
7. The Routledge Encyclopedia of Translation Technology. Chan Sin-Wai (Ed.) Routledge, 2015.
8. Wu Y. Et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation 2016 [https://arxiv.org/pdf/1609.08144.pdf%20\(7.pdf](https://arxiv.org/pdf/1609.08144.pdf%20(7.pdf)

Рекомендованная литература

1. Баранов А.Н. Введение в прикладную лингвистику. Серия "Новый лингвистический учебник". М.: Эдиториал УРСС. 2001. Глава 2, раздел 1.3.1. Моделирование общения (с. 20-31); Глава 4, разделы 1.3.1. – 1.3.4. «Естественный» перевод: лингвистические проблемы (с. 143-163); 1.4. Машинный перевод (с. 168-178).
 2. Carl Pollard, Ivan A. Sag: Head-Driven Phrase Structure Grammar. Chicago: University of Chicago Press. 1994.
 3. Nirenburg, Sergei, Harold L. Somers, and Yorick Wilks (eds.) Readings in Machine Translation. , Cambridge, Massachusetts: MIT Press. 2003
 4. Boguslavsky I, Iomdin L. Nivre J. Parsing the Russian Dependency Treebank. Proceedings of COLING-2008. Manchester, 2008.
 5. Philipp Koehn. Statistical Machine Translation. Cambridge University Press. 2009.
 6. Jury D. Apresjan, Igor M. Boguslavsky, Leonid L. Iomdin, Leonid L. Tsinman. Lexical Functions in Actual NLP-Applications // Selected Lexical and Grammatical Issues in the Meaning–Text Theory. In honour of Igor Mel'čuk. (Ed. by Leo Wanner). John Benjamins, Studies in Language Companion. Series 84. ISBN 978 90 272 3094 2. 2007. P. 199-230.
 7. The Routledge Encyclopedia of Translation Technology. Chan Sin-Wai (Ed.) Routledge, 2015.
 8. Мельчук И.А. Опыт теории лингвистических моделей класса «Смысл – Текст».– М.: Наука, 1995.
 9. Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистический процессор для сложных информационных систем. М.: Наука, 1992.
- Apresjan Ju, Boguslavsky I., Iomdin L et al. ЕТАР-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003, First International Conference on Meaning – Text Theory (June 16-18 2003). Paris: École Normale Supérieure, 2003. P. 279-288.

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

7. Материально-техническое обеспечение дисциплины

№п /п	Наименование ПО	Производитель	Способ распространения
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное

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 SE; 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

9.1 Планы семинарских/ практических/ лабораторных занятий

1. Лингвистика как наука о языке. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика. Лингвистика и прагматика.
2. Лингвистическое моделирование. Действующие модели языка. Теория «Смысл – Текст» и возможности ее использования для построения систем автоматической обработки текста.
3. Грамматика и словарь естественного языка. Представление об интегральном описании языка.
4. Краткий обзор формальных грамматик. Порождающие грамматики. Грамматики составляющих и грамматики зависимостей. Гибридные грамматики.
5. Автоматический анализ и синтез текста. Морфологический анализ текста. Синтаксический анализ текста (парсинг). Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх». Правильные и статистические подходы к автоматической обработке текста.
6. Языковая неоднозначность как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста. Типы неоднозначности. Интерактивное разрешение лексической и синтаксической неоднозначности. Применение онтологии для разрешения неоднозначности. Методы машинного обучения в применении к разрешению неоднозначности.
7. Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке. Система машинного перевода как механизм обратной связи и источник новых лингвистических знаний.
8. Типы систем машинного перевода. Автоматический и автоматизированный перевод. Память переводов. Интерлингва. UNL (универсальный сетевой язык) как тип интерлингвы и его использование в задаче машинного перевода.
9. Морфологический компонент системы автоматической обработки текстов. Морфологическая структура слова и предложения.

10. Синтаксический компонент системы автоматической обработки текстов. Синтаксическая структура предложения. Алгоритм синтаксического анализа. Синтаксические отношения. Синтагмы.

11. Словарь системы автоматической обработки текстов. Словарь системы машинного перевода. Структура словарной статьи. Синтаксические признаки. Семантические признаки (дескрипторы). Теория валентностей. Модель управления.

12. Правила межъязыкового перевода в узком смысле (трансфер).

13. Лексические функции в машинном переводе.

14. Краткий обзор действующих систем машинного перевода. Статистические системы машинного перевода.

15. Краткий обзор смежных задач прикладной лингвистики. Информационный поиск. Интеллектуальный анализ данных. Синонимическое перифразирование высказываний и его прикладное значение.

16. Некоторые современные цифровые лингвистические ресурсы и их роль в задачах автоматической обработки текстов. Аннотированные корпуса текстов. Word Net, Frame Net, Treebanks (включая SynTagRus), Semantic Web).

9.2 Другие материалы

Все необходимые для обучения материалы даются на лекциях и практических занятиях.