

# Правительство Российской Федерации

Федеральное государственное автономное образовательное учреждение  
высшего профессионального образования

## «Национальный исследовательский университет «Высшая школа экономики»

Факультет БИЗНЕС-ИНФОРМАТИКИ  
Отделение ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ

### Программа дисциплины

### Компьютерная лингвистика

для направления 010400.68 «Прикладная математика и информатика»  
подготовки магистров

Авторы: Большакова Е.И. (eibolshakova@hse.ru),  
Лукашевич Н.В. (louk\_nat@mail.ru)

Рекомендована секцией УМС  
«Прикладная математика  
и информатика»

Председатель  
\_\_\_\_\_ Кузнецов С.О.  
« \_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Одобрена на заседании кафедры  
Анализа данных  
и искусственного интеллекта

Зав. кафедрой  
\_\_\_\_\_ Кузнецов С.О.  
« \_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Утверждена УС факультета  
бизнес-информатики

Ученый секретарь  
\_\_\_\_\_  
« \_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Москва

# **I. Пояснительная записка**

## **Авторы программы**

кандидат физико-математических наук Е.И. Большакова,  
кандидат физико-математических наук Н.В. Лукашевич

## **Требования к студентам**

Предполагаются базовые знания по дисциплинам «Дискретная математика», «Информатика и программирование», «Алгоритмы и структуры данных».

## **Аннотация**

Курс «Компьютерная лингвистика» продолжает изучение методов и средств автоматической обработки текстов (АОТ) на естественном языке, начатое в рамках магистерского курса «Введение в интеллектуальные информационные системы». Задачей курса является рассмотрение различных моделей компьютерной лингвистики, применяемых в современных предметно-ориентированных информационных системах. Изучаются виды используемых лингвистических ресурсов и методы их создания. Рассматриваются прикладные задачи АОТ, требующие многоуровневого анализа и синтеза текста (такие как машинный перевод, генерация текста, задача извлечения информации и знаний из текста).

## **Учебные задачи курса**

Основная задача курса – изучение основных принципов и методов автоматической обработки текстов на естественном языке (ЕЯ). В результате изучения дисциплины студенты должны:

- Знать основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;
- Знать основные принципы представления знаний о предметной области в виде рубрикаторов, тезаурусов, онтологий;
- Понимать существенные отличия естественных языков от искусственных и особенности компьютерных моделей естественного языка;
- Знать типичные программные системы (приложения) в области АОТ и их архитектурные особенности;
- Уметь работать с одной из систем АОТ и/или инструментальным средством для разработки систем АОТ.

## II. Тематический план курса «Компьютерная лингвистика»

№	Название темы	Всего часов по дисциплине	Аудиторные часы		Самосто- ятельная работа
			Лекции	Сем. и практика	
1	Введение	10	4	0	6
2	Статистические модели	20	4	4	12
3	Морфологические модели	22	4	4	14
4	Методы синтаксического анализа	28	6	4	18
5	Модели дискурса и семантики	20	4	4	12
6	Разработка и применение лингвистических ресурсов	38	6	10	22
7	Построение приложений КЛ	24	4	6	14
	Итого	162	32	32	98

## III. Источники информации

**Базовый учебник** – ридер «Автоматическая обработка текста», составленный по следующим источникам:

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И. и др. – М.: МИЭМ, 2011.
2. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011.
3. The Oxford Handbook on Computational Linguistics. R. Mitkov (Ed.). Oxford University Press, 2005.

### Дополнительная литература

1. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.
2. Болховитянов А.В., Гусев А.В., Чеповский А.М. Морфологические модели компьютерной лингвистики: учеб. пособие – М. МГУП, 2010.
3. Ван Дейк Т.А., Кинч В. Стратегия понимания связного текста.// Новое в зарубежной лингвистике. Вып. XXIII — М., Прогресс, 1988, с. 153-211.
4. Васильев В. Г., Кривенко М. П. Методы автоматизированной обработки текстов. – М.: ИПИ РАН, 2008.
5. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения. — М., Наука, 1985.
6. Зубов А.В., Зубова И.И. Основы искусственного интеллекта для лингвистов. – М., Логос, 2007.
7. Касевич В.Б. Элементы общей лингвистики. — М., Наука, 1977.
8. Кобозева И.М. Лингвистическая семантика. – М., 2009.
9. Кронгауз М.А. Семантика. - М.: Издательский центр «Академия», 2005.

10. Маккьюин К. Дискурсивные стратегии для синтеза текста на естественном языке // Новое в зарубежной лингвистике. Вып. XXIV. М.: Прогресс, 1989, с.311-356.
11. Леонтьева Н. Н. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие – М.: Академия, 2006.
12. Пиотровский Р.Г. , Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. – М.: Высшая школа, 1977.
13. Чатуев М.Б., Чеповский А.М. Частотные методы в компьютерной лингвистике: учебное пособие – М. МГУП, 2011.
14. Biber, D., Conrad S., and Reppen D. Corpus Linguistics. Investigating Language Structure and Use. Cambridge University Press, Cambridge, 1998.
15. Bolshakov, I.A., Gelbukh A. Computational Linguistics. Models, Resources, Applications. Mexico, IPN, 2004.
16. Jurafsky D., Martin J. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Prentice Hall, 2000.
17. Manning, Ch. D., H. Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999.
18. Word Net: an Electronic Lexical Database. /Edit. by Christiane Fellbaum. Cambridge, MIT Press, 1998.

#### IV. Формы контроля и структура итоговой оценки

Текущий контроль – контрольная работа (80 мин.), домашнее задание в третьем модуле.

Промежуточный контроль – устный **зачёт** в конце третьего модуля;

Итоговый контроль – письменный экзамен в конце четвертого модуля.

В третьем модуле **оценка за текущий контроль** рассчитывается следующим образом:

$$O_{\text{текущий}} = 0,5 \cdot O_{\text{кр}} + 0,5 \cdot O_{\text{д/з}}$$

**Оценка промежуточного контроля** в форме зачета выставляется по формуле:

$$O_{\text{промежуточный}} = 0,5 \cdot O_{\text{зачет}} + 0,5 \cdot O_{\text{текущий}}$$

где  $O_{\text{зачет}}$  – оценка за работу непосредственно на устном зачете.

**Итоговая оценка по курсу** выставляется по следующей формуле:

$$O_{\text{итоговый}} = 0,5 \cdot O_{\text{экзамен}} + 0,5 \cdot O_{\text{сам. работа}}$$

где  $O_{\text{экзамен}}$  – оценка за работу непосредственно на экзамене, а  $O_{\text{сам. работа}}$  – оценка самостоятельной работы студентов при выполнении домашних заданий по текущим темам курса в четвертом модуле.

В диплом выставляется **результатирующая оценка** по данной учебной дисциплине по формуле:

$$O_{\text{дисциплина}} = 0,5 \cdot O_{\text{промежуточный}} + 0,5 \cdot O_{\text{итоговый}}$$

Таблица соответствия оценок по десятибалльной системе и системе зачет/незачет

Оценка по 10-балльной шкале	Оценка по 5-балльной шкале
1	незачет
2	
3	
4	зачет
5	
6	
7	
8	
9	
10	

**Таблица соответствия оценок по десятибалльной и пятибалльной системе**

<b>По десятибалльной шкале</b>	<b>По пятибалльной системе</b>
1 – неудовлетворительно 2 – очень плохо 3 – плохо	неудовлетворительно – 2
4 – удовлетворительно 5 – весьма удовлетворительно	удовлетворительно – 3
6 – хорошо 7 – очень хорошо	хорошо – 4
8 – почти отлично 9 – отлично 10 – блестяще	отлично – 5

## **У. Содержание курса «Компьютерная лингвистика»**

### **Тема 1. Введение**

1. Компьютерная лингвистика и автоматическая обработка текстов на естественном языке (ЕЯ): основные задачи и история развития. Междисциплинарный характер направления, связь со смежными научными дисциплинами.

2. Особенности ЕЯ, понятия языкового знака и языковой системы. Принципиальные отличия естественных и искусственных (формальных) языков: открытость, избыточность, нестандартная сочетаемость, асимметрия знаков и смыслов.

3. Уровни языковой системы (от фонетики до дискурса), их взаимосвязь. Основные единицы текста. Уровень фонем и символов. Синтаксический и морфологический уровни. Лексическая система. Словоформы и лексемы.

4. Понятие модели в компьютерной лингвистике. Основные уровни обработки текста и виды моделей. Модель «Смысл-Текст». Лингвистический процессор и лингвистические ресурсы (компьютерные словари и тезаурусы, грамматики, корпуса текстов).

### **Основная литература**

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И. и др. – М.: МИЭМ, 2011.
2. Леонтьева Н. Н. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие – М.: Академия, 2006.

### **Дополнительная литература**

1. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.
2. Касевич В.Б. Элементы общей лингвистики. — М., Наука, 1977.
3. Bolshakov, I.A., Gelbukh A. Computational Linguistics. Models, Resources, Applications. Mexico, IPN, 2004.
4. The Oxford Handbook on Computational Linguistics. R. Mitkov (Ed.). Oxford University Press, 2005.

## **Тема 2. Статистические модели**

1. Статистика словоупотреблений в текстах на ЕЯ. Закон Ципфа-Мандельброта и его интерпретация. Статистические характеристики слов ЕЯ. Средняя длина словоформы, зависимость длины слова от его частоты. Частотные словари ЕЯ.

2. Статистика встречаемости букв и буквосочетаний: биграмм, триграмм, N-грамм. Метод N-грамм для букв и для слов текста. Статистические языковые модели. Цепи Маркова и их применение. Алгоритм Витерби. Приложения статистических моделей.

4. Статистика встречаемости сочетаний слов, понятие коллокации. Устойчивые словосочетания, методы их автоматического извлечения на базе статистики. Меры устойчивости.

### **Основная литература**

1. Manning, Ch. D., H. Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999.
2. The Oxford Handbook on Computational Linguistics. R. Mitkov (Ed.). Oxford University Press, 2005.

### **Дополнительная литература**

1. Jurafsky D., Martin J. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, 2000.
2. Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. – М.: Высшая школа, 1977.
3. Чатуев М.Б., Чеповский А.М. Частотные методы в компьютерной лингвистике: учебное пособие – М. МГУП, 2011.

## **Тема 3. Морфологические модели**

1. Основные понятия морфологических моделей: морфема, аффикс, основа, флексия, словоизменяющая парадигма, флективный класс. Принципы построения морфологических моделей на базе словаря. Словари основ и словари словоформ.

2. Состав морфологического словаря. Особенности русской морфологии. Морфологическая зона толково-комбинаторного словаря в модели «Смысл-Текст». Морфологический синтез. Сравнение морфологических моделей.

3. Виды морфологического анализа: выделение основы, лемматизация, полный морфоанализ. Основные методы морфологического анализа: анализ на базе словаря, бессловарный анализ, анализ без опоры на грамматику. Программные модули автоматического морфологического анализа для русского языка.

4. Морфемный состав слова. Виды морфем. Морфотактики. Вычислительные модели морфологии на базе конечных автоматов. Приложения морфологических моделей.

### **Основная литература**

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И. и др. – М.: МИЭМ, 2011.
2. Васильев В. Г., Кривенко М. П. Методы автоматизированной обработки текстов. – М.: ИПИ РАН, 2008.
3. Леонтьева Н. Н. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие – М.: Академия, 2006.

### **Дополнительная литература**

1. Апресян Ю.Д. и др. Лингвистическое обеспечение системы ЭТАП-2. М.: Наука, 1989.

2. Болховитянов А.В., Гусев А.В., Чеповский А.М. Морфологические модели компьютерной лингвистики: учеб. пособие – М. МГУП, 2010.
3. The Oxford Handbook on Computational Linguistics. R. Mitkov (Ed.). Oxford University Press, 2005.

#### **Тема 4. Методы синтаксического анализа**

1. Различные подходы к анализу синтаксиса предложений ЕЯ. Основная задача синтаксического анализа. Синтаксические деревья: деревья непосредственных составляющих и деревья зависимостей. Понятия синтаксического предиката, валентности и актанта, модели управления. Синтаксические связи. Синтаксическая зона толково-комбинаторного словаря в модели «Смысл-Текст».

2. Трансформационный подход. Формальные языки и грамматики, классификация по Хомскому. Контекстно-зависимые, контекстно-свободные и регулярные (автоматные) грамматики. Задача синтаксического разбора для контекстно-свободных (КС) грамматик. Нисходящий и восходящий разбор.

3. Методы и алгоритмы синтаксического анализа для контекстно-свободных языков. Синтаксический разбор на базе автоматов и преобразователей с конечным числом состояний (Finite State Transducers). Расширенные сети переходов Вудса. Анализ на основе унификационных грамматик.

4. Частичный синтаксический анализ. Словосочетания и их основные типы. Выделение словосочетаний и синтаксических групп. Задача синтаксической сегментации текста. Словосочетания и коллокации.

#### **Основная литература**

1. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения. — М., Наука, 1985.
2. Леонтьева Н. Н. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие – М.: Академия, 2006.
3. The Oxford Handbook on Computational Linguistics. R. Mitkov (Ed.). Oxford University Press, 2005.

#### **Дополнительная литература**

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И. и др. – М.: МИЭМ, 2011.
2. Апресян Ю.Д. и др. Лингвистическое обеспечение системы ЭТАП-2. М.: Наука, 1989.
3. Васильев В. Г., Кривенко М. П. Методы автоматизированной обработки текстов. – М.: ИПИ РАН, 2008.
4. Касевич В.Б. Элементы общей лингвистики. — М., Наука, 1977.
5. Bolshakov, I.A., Gelbukh A. Computational Linguistics. Models, Resources, Applications. Mexico, IPN, 2004.
6. Jurafsky D., Martin J. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, 2000.

#### **Тема 5. Модели дискурса и семантики**

1. Характеристики связного текста (дискурса): тематическая связность, риторическая связность, лексическая связность, референциальная связность. Целостность и связность. Анафорические ссылки, кореференция, лексические повторы, дискурсивные слова.

2. Тематическая и композиционная структура текста. Сверхфразовые единства. Композиционные и дискурсивные особенности текстов разных жанров и стилей, их учет при обработке текстов.

3. Моделирование свойств связного текста. Построение лексических цепочек, автоматическое разрешение референции, автоматическое выявление риторической, композиционной, тематической структуры текста.

4. Модели представления семантики. Язык предикатов первого и высших порядков. Типизированные структуры.

### **Основная литература**

1. Леонтьева Н. Н. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие – М.: Академия, 2006.
2. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011.

### **Дополнительная литература**

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И. и др. – М.: МИЭМ, 2011.
2. Ван Дейк Т.А., Кинч В. Стратегия понимания связного текста.// Новое в зарубежной лингвистике. Вып. XXIII — М., Прогресс, 1988, с. 153-211.
3. Зубов А.В., Зубова И.И. Основы искусственного интеллекта для лингвистов. – М., Логос, 2007.
4. Кобозева И.М. Лингвистическая семантика. – М., 2009.
5. Кронгауз М.А. Семантика. - М.: Издательский центр «Академия», 2005.

## **Тема 7. Разработка и применение лингвистических ресурсов**

1. Система понятий и терминов как основа описания предметной области. Синонимия и лексическая многозначность. Смысловые (парадигматические) отношения лексических единиц. Лексические ресурсы WordNet и EuroNet.

2. Рубрикаторы и таксономии: функции и принципы создания. Информационно-поисковые тезаурусы (ИПТ). Стандарты по разработке ИПТ. Онтологии в информационно-поисковых системах.

3. Методы извлечения терминологических слов и словосочетаний из текстов. Способы оценки качества извлечения. Извлечение терминов как многофакторный процесс. Извлечение терминологических сочетаний в задаче индексирования текстов. Методы выявления терминологических связей: извлечение синонимов терминов, установление родовидовых отношений. Автоматизация построения таксономий.

4. Коллекции и корпуса текстов. Корпусная лингвистика. Виды разметки текстов. Параллельные и псевдопараллельные корпуса текстов, их применение.

### **Основная литература**

1. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011.
2. Biber, D., Conrad S., and Reppen D. Corpus Linguistics. Investigating Language Structure and Use. Cambridge University Press, Cambridge, 1998.
3. Hirst, G. Ontology and the Lexicon. In.: Handbook on Ontologies in Information Systems. Berlin, Springer, 2003.
4. Word Net: an Electronic Lexical Database. /Edit. by Christiane Fellbaum. Cambridge, MIT Press, 1998.



## Дополнительная литература

1. Кобозева И.М. Лингвистическая семантика. – М., 2009.
2. Национальный Корпус Русского Языка. <http://ruscorpora.ru>
3. Manning, Ch. D., H. Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999.
4. The Oxford Handbook on Computational Linguistics. R. Mitkov (Ed.). Oxford University Press, 2005.

## Тема 7. Построение приложений КЛ

1. Обзор приложений компьютерной лингвистики: машинный перевод, распознавание речи, автоматизация редактирования текстов, обучение языку, генерация текстов, диалог на ЕЯ, индексирование, реферирование и аннотирование текстов, извлечение информации и знаний из текстов. Основные подходы к разработке приложений: подход, основанный на знаниях (правилах) и подход, основанный статистике и машинном обучении.

2. Машинный перевод (МП). Лингвистические стратегии машинного перевода и поколения систем МП. Автоматический перевод, основанный на правилах. Интерлингва. Оценки качества машинного перевода. Статистический машинный перевод.

3. Генерация текстов документов. Стратегии синтеза текста. Генерация многоязыковых руководств пользователя (инструкций) по формальному описанию проблемной области.

4. Извлечение информации из текстов (Information Extraction) как одно из основных направлений КЛ. Технологии извлечения при подходе. Лингвистические шаблоны и их использование. Проблемы распознавания кореференции. Качество извлечения информации.

5. Автоматическое индексирование текстов. Извлечение ключевых слов.

## Основная литература

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И. и др. – М.: МИЭМ, 2011.
2. Леонтьева Н. Н. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие – М.: Академия, 2006.
3. The Oxford Handbook on Computational Linguistics. R. Mitkov (Ed.). Oxford University Press, 2005.

## Дополнительная литература

1. Васильев В. Г., Кривенко М. П. Методы автоматизированной обработки текстов. – М.: ИПИ РАН, 2008.
2. Маккьюин К. Дискурсивные стратегии для синтеза текста на естественном языке // Новое в зарубежной лингвистике. Вып. XXIV. М.: Прогресс, 1989, с.311-356.
3. Brown P., Pietra S., Mercer R., Pietra V. The Mathematics of Statistical Machine Translation. // Computational Linguistics, Vol. 19(2): 263-311. 1993.

## VI. Тематика заданий по формам текущего контроля

### Примеры заданий на контрольной работе

1. Дать определения терминов: *N-грамма; лексема; валентность; кореференция*. Проиллюстрировать эти понятия на примере слова *прослушивание*.
2. Что такое статистическая языковая модель? В каких приложениях компьютерной лингвистики применяются эти модели?
3. Что является результатом полного морфологического анализа заданной словоформы? Поясните на примере конкретной словоформы.

4. В чем отличие синтаксических деревьев непосредственно составляющих от синтаксических деревьев зависимостей?
5. Для заданной формальной грамматики и описать порождаемый ею язык. Привести примеры предложений (фраз) языка, и для одного из них указать соответствующее синтаксическое дерево разбора.
6. В заданном связном тексте указать и кратко пояснить межфразовые связи.
7. Какие основные виды смысловых связей терминов представлены тезаурусах?

### Примеры домашних заданий

1. Построение для заданной коллекции текста статистической языковой модели.
2. Разработка программы на базе конечного автомата для сегментации и начального лексического анализа неформатированного текста на естественном языке.
3. Разработка программы для синтаксического анализа текста методом рекурсивного спуска на основе контекстно-свободной грамматики.
4. Для заданного слова русского языка найти и сравнить его толкования в различных толковых словарях, а также в Национальном корпусе русского языка.
5. Для заданного набора словосочетаний отобрать явные термины и явные нетермины, объяснить принятые решения.

## VII. Вопросы для оценки качества освоения дисциплины

### Тема 1.

1. С какими научными дисциплинами связана область автоматической обработки текстов?
2. В чем особенности естественного языка как знаковой системы?
3. Перечислите основные отличия естественных языков от искусственных.
4. В чем суть явления полисемии? омонимии? Приведите примеры.
5. Перечислите основные уровни (подсистемы) языковой системы.
6. В чем особенности компьютерных моделей естественного языка?
7. Назовите основные уровни (модули) автоматического анализа текста.

### Тема 2.

8. Охарактеризуйте количественную меру информации и неопределенности.
9. В чем заключается закон Ципфа-Мальдельброта?
10. Какие основные статистические характеристики применяются для изучения естественного языка?
11. Объясните понятие цепи Маркова.
12. Что такое коллокация?
13. Объясните понятие устойчивого словосочетания.
14. Что такое мера взаимной информации  $MI$ ?
15. Какие статистические меры применяются для извлечения коллокаций?

### Тема 3.

16. Охарактеризуйте понятие лексемы.
17. Что такое морфема? аффикс? Какие виды аффиксов вы знаете?
18. Чем основа слова отличается от корня? Приведите примеры.
19. Что такое словоизменяемая парадигма?
20. Какая информация представляется в морфологических словарях?
21. Что такое лемматизация?
22. Назовите основные стратегии морфологического анализа.
23. Приведите пример морфологической омонимии.

### Тема 4.

24. Назовите основные подходы в анализу синтаксиса предложений ЕЯ.
25. Что такое синтаксическое дерево?

26. В чем особенность деревьев составляющих? Приведите пример.
27. В чем особенность деревьев зависимостей? Приведите пример.
28. Что такое валентность? Актант? Приведите примеры.
29. Опишите основные особенности трансформационного метода по Хомскому.
30. Что такое формальный язык? Формальная грамматика?
31. Опишите классификацию формальных грамматик по Хомскому.
32. Какие методы и алгоритмы анализа контекстно-свободных языков вы знаете?
33. Что такое преобразователь с конечным числом состояний и для чего он используется при обработке текстов?
34. В чем состоит синтаксическая сегментация текста?
35. Какие типы словосочетаний вы знаете?

Тема 5.

36. Назовите отличительные характеристики связного текста.
37. Что такое анафорическая ссылка?
38. Поясните понятие сверхфразового единства.
39. Приведите пример дискурсивной особенности текста.
40. Что такое автоматическое разрешение референции?
41. Объясните понятие лексической цепочки. Приведите примеры.
42. Что такое тематическая структура текстов?
43. Укажите принципы автоматического разрешения референции.
44. Что такое риторическая структура текста?
45. Какие модели семантики текста вы знаете?

Тема 6.

46. Что такое термин? Приведите примеры.
47. Назовите основные свойства терминов.
48. Что такое информационно-поисковый тезаурус?
49. Что такое дескриптор тезауруса?
50. Охарактеризуйте понятие лингвистической онтологии.
51. Какие свойства родовидовых (таксономических отношений) вы знаете?
52. Укажите принципы установления родовидовых (таксономических) отношений.
53. Назовите свойства отношения часть-целое.
54. Какие подвиды отношения часть-целое вы можете назвать?
55. Что такое параллельный и псевдопараллельный корпус?

Тема 7.

56. Назовите типичные приложения автоматической обработки текстов.
57. Укажите приложения АОТ, в которых нужен морфологический анализ.
58. В каких приложениях АОТ применяется синтаксический анализ?
59. Укажите особенности задачи извлечения информации из текстов.
60. Укажите основные стратегии машинного перевода.
61. Что такое интерлингва?
62. Назовите известные системы машинного перевода.
63. Как измеряется качество машинного перевода?
64. Что такое мера *BLEU*?
65. В каких прикладных задачах применяется генерация текста?

Авторы программы: \_\_\_\_\_ / Большакова Е.И. /  
 \_\_\_\_\_ / Лукашевич Н.В. /