

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ярошенко Николай Николаевич
Должность: проректор по учебно-методической деятельности
Дата подписания: 04.06.2026 11:02
Уникальный программный ключ:
25cc77c6d2a242799b1569189212ec549db4bb3f

Федеральное государственное бюджетное образовательное учреждение

высшего образования

Московский государственный институт культуры

УТВЕРЖДЕНО
Председатель УМС
Библиотечно-информационного
факультета
Боронина Н. В.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДИСЦИПЛИНЫ (МОДУЛЯ)
Б1.О.20 Лингвистическое обеспечение информационных систем

Направление подготовки/специальности (код, наименование): 09.03.02
Информационные системы и технологии

Профиль подготовки/специализация: Информационные системы и цифровые технологии в культуре

Квалификация (степень) выпускника: Бакалавр

Форма обучения: очная

*(РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов)*

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели:

Целью освоения дисциплины является: обучение студентов базовым знаниям в области разработки лингвистического обеспечения как одной из ключевых подсистем, обеспечивающей поддержку эффективного пользовательского интерфейса в работе с информационно-поисковыми системами, базами данных и знаний.

Задачи:

- 1) формирование теоретических основ в области разработки лингвистического обеспечения.
- 2) знакомство студентов с инструментальными средствами и стандартами, поддерживающими разработку лингвистического обеспечения.
- 3) изучение методики решения задач в области проектной деятельности

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Дисциплина «Лингвистическое обеспечение информационных систем» входит в состав Блока 1 «Дисциплины (модули)» и относится к обязательной части ОПОП по направлению подготовки 09.03.02 Информационные системы и технологии, профиль - Информационные системы и цифровые технологии в культуре.

Дисциплина «Лингвистическое обеспечение информационных систем» изучается в седьмом семестре. Входные знания, умения и компетенции, необходимые для изучения данного курса, формируются в процессе изучения таких дисциплин, как «Проектирование ИС», «Лингвистические средства информационных технологий» и «Теоретические основы информатики». В результате освоения дисциплины формируются знания, умения и навыки, необходимые для изучения следующих дисциплин и прохождения практик.

Взаимосвязь курса с другими дисциплинами ООП способствует планомерному формированию необходимых компетенций и углубленной подготовке студентов к решению специальных практических профессиональных задач.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование компетенций (элементов следующих компетенций) в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки (специальности) 09.03.02 Информационные системы и технологии:

Перечень планируемых результатов обучения по дисциплине (модулю).

Компетенция (код и наименование)	Индикаторы компетенций	Результаты обучения

<p>ПК-1 Способен понимать и применять в практической деятельности теоретические основы технических, социально-гуманитарных и междисциплинарных знаний, историю и прогнозы развития информационной сферы</p>	<p>ПК-1.3. Понимает и применяет в информационной деятельности основы лингвистики</p>	<p>Знает:</p> <ul style="list-style-type: none"> - структуру и назначение ИПЯ; правила создания лексико-семантической основы ИПЯ; теоретические основы типологии ИПЯ; отличительные особенности классификационных, вербальных и объектно-признаковых ИПЯ, а также области их применения; - состав, структуру и функции лингвистических средств информационной технологии; - современное состояние и тенденции развития лингвистического обеспечения ИС; состав, структуру, виды и функции лингвистического обеспечения ИС; - требования и критерии выбора компонентов при проектировании лингвистического обеспечения ИС; состав и содержание работ по проектированию и ведению лингвистического обеспечения ИС; нормативно-справочную базу проектирования лингвистического обеспечения ИС; - области применения лингвистических средств в сфере ИС и ИТ <p>Умеет:</p> <ul style="list-style-type: none"> - осуществлять основные процессы создания лексико-семантической основы ИПЯ: отбор, нормализацию, систематизацию лексических единиц, организационное оформление ИПЯ; - индексировать документы и запросы с использованием различных типов ИПЯ; - создавать и поддерживать в рабочем состоянии (актуализировать) лингвистические средства информационной технологии - осуществлять основные этапы проектирования лингвистического обеспечения ИС; обосновать и аргументировать состав и структуру проектируемого лингвистического обеспечения ИС, соответствующего целям и задачам конкретной ИС; разрабатывать и поддерживать в рабочем состоянии компоненты лингвистического обеспечения ИС;
---	--	---

		<p>выбирать и рационально использовать лингвистические средства для оптимизации использования информационных ресурсов и производства информационных продуктов</p>
		<p>Владеет:</p> <ul style="list-style-type: none"> - пониманием инструментов лингвистики в машинном переводе, автоматическом распознавании символов и речи, автоматическом извлечении данных, реферировании текстов, построении систем управления знаниями, создании электронных словарей, тезаурусов, онтологий, корпусной лингвистики, SEO; - способностью обосновать и аргументировать состав и структуру лингвистических средств, адекватных целям и задачам используемой информационной технологии; технологией предкоординатного и посткоординатного индексирования документов и запросов; методами и способами информационной диагностики предметной области и информационного моделирования с помощью лингвистических средств; - понятийным аппаратом, характеризующим лингвистическое обеспечение ИС; технологиями создания локальных лингвистических средств ИС; методами и способами повышения эффективности использования информационных ресурсов и производства информационных продуктов и услуг ИС с помощью лингвистических средств

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (модуля)

4.1 Объем дисциплины (модуля)

Объем (общая трудоемкость) дисциплины «Лингвистическое обеспечение информационных систем» составляет 3 з.е., 108 акад. часов, из них:

Контактных: 19 акад.ч.

СРС: 89 акад.ч.

Форма контроля: зачет.

4.2. Структура дисциплины для очной формы обучения.

	Раздел дисциплины	Семестр	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах) в т.ч. в интерактивной форме					Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
			Лекции	Семинары/ практические	Консультации	ИКР	СРС	
1	Язык, как средство представления информации.	7	6				18	Экспресс-опрос
2	Языковые интерфейсы.	7	4				14	Экспресс-опрос
3	Представление информации и языки обработки данных в ИС.	7	6	4			14	Экспресс-опрос Проверка отчётов по практическим работам
4	Разработка элементов информационно-поисковых систем.	7	4				14	Экспресс-опрос
5	Моделирование лингвистического обеспечения ИС.	7	6	4			14	Тестирование Проверка отчётов по практическим работам
	Форма итогового контроля							Зачет
	Всего 108 час		26	8			74	

ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Текущий контроль выполнения заданий (контроль формирования компетенций) осуществляется регулярно, начиная с первой недели семестра (входящий контроль). Текущий контроль освоения отдельных разделов дисциплины осуществляется при помощи опроса и тестового материала в завершении изучения каждого раздела. Система текущего контроля успеваемости служит не только оценке уровня компетентностной подготовки обучающегося и способствует в дальнейшем наиболее качественному и объективному оцениванию его в ходе промежуточной аттестации, но и самооценке обучающегося, стимулируя его усилия.

Промежуточная аттестация проводится в форме зачета.

Система оценивания

Форма контроля	Компетенция	Оценка
Текущий контроль: - опрос - участие в дискуссии на семинаре Выполнение практических работ	ПК-1.3.	зачтено/не зачтено зачтено/не зачтено отлично/хорошо/удовлетворительно/неудовлетворительно
Промежуточная аттестация зачет	ПК - 1.3	зачтено /не зачтено

Критерии оценки результатов по дисциплине

Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
«зачтено»	<p>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.</p> <p>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.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне «достаточный».</p>
«не зачтено»	<p>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.</p> <p>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.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p>

Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
	Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

Тестовые задания №1

№ п/п	Вопрос	Варианты ответов
1.	Что оценивает прагматический подход к измерению информации?	<ul style="list-style-type: none"> 1) Объем памяти, занимаемый сообщением 2) Синтаксическую правильность построения сообщения 3) Полезность, ценность информации для потребителя
2.	Какой метод морфологического анализа основан на использовании словаря основ и правил словоизменения?	<ul style="list-style-type: none"> 1) Вероятностный 2) Алгоритмический 3) Нейросетевой
3.	Какой подход к автоматическому анализу текста использует размеченные корпуса примеров для обучения моделей?	<ul style="list-style-type: none"> 1) Правилковой (экспертный) 2) Статистический 3) Дедуктивный
4.	Что представляет собой иерархический классификатор?	<ul style="list-style-type: none"> 1) Линейный список терминов в алфавитном порядке 2) Матрицу ассоциаций терминов 3) Древовидную структуру, где каждая рубрика может содержать подрубрики
5.	Какова основная функция конструкторов запросов в информационных системах?	<ul style="list-style-type: none"> 1) Автоматический перевод запросов на иностранные языки 2) Визуальное или формализованное создание поисковых запросов 3) Оценка релевантности найденных документов

6.	Какая модель коммуникации описывает процесс передачи информации как движение сообщения от источника через канал связи к получателю?	<ol style="list-style-type: none"> 1) Линейная 2) Трансакционная 3) Интеракционная
7.	Что семиотическом треугольнике связывает понятие (смысл, концепт)?	<ol style="list-style-type: none"> 1) Знак и обозначаемый предмет 2) Знак и адресат сообщения 3) Другие знаки в системе
8.	Что в концептуальном моделировании предметной области показывает "связь" (relationship) между сущностями?	<ol style="list-style-type: none"> 1) Только временную последовательность событий 2) Исключительно наследование свойств 3) Ассоциацию, существующую между объектами разных классов
9.	Какой метод абстрагирования позволяет выделить общие характеристики объектов, отвлекаясь от их частных различий?	<ol style="list-style-type: none"> 1) Анализ 2) Обобщение 3) Декомпозиция
10.	Какой принцип лежит в основе поиска информации в гипертекстовых системах?	<ol style="list-style-type: none"> 1) Переход между связанными документами по ссылкам 2) Последовательный перебор всех релевантных документов 3) Формирование поискового образа документа путем ключевых слов

Тестовые задания №2

№ п/п	Вопрос	Варианты ответов
1.	Что является основной задачей морфологического анализа в системах обработки естественного языка?	<ol style="list-style-type: none"> 1) Определение синтаксической структуры предложения 2) Приведение словоформы к ее

		<p>нормальной форме и определение грамматических признаков</p> <p>3) Семантическая разметка текста</p>
2.	Для решения какой задачи в обработке текстов наиболее эффективно применяются конечные автоматы?	<p>1) Семантический анализ текста</p> <p>2) Лексический анализ</p> <p>3) Машинный перевод сложных предложений</p>
3.	Через что в атрибутивном представлении информации описывается сущность?	<p>1) Ее уникальное имя</p> <p>2) Ее свойства и их значения</p> <p>3) Ее иерархии классов</p>
4.	Что такое тезаурус в контексте информационно-поисковых систем?	<p>1) Алфавитный указатель терминов</p> <p>2) Программа для проверки орфографии</p> <p>3) Словарь, в котором отражены отношения между терминами</p>
5.	Какая логическая форма выражается предложением: "Солнце является звездой"?	<p>1) Умозаключение</p> <p>2) Понятие</p> <p>3) Суждение</p>
6.	Что является результатом синтаксического анализа предложения?	<p>1) Дерево зависимостей, отражающее структуру предложения</p> <p>2) Набор ключевых слов для поиска</p> <p>3) Список всех слов с их морфологическими характеристиками</p>
7.	Что представляет собой конечный автомат, построенный по правилам регулярной грамматики?	<p>1) Иерархическое дерево классов</p> <p>2) Релятивную таблицу логических формул</p> <p>3) Граф, вершинами которого являются состояния, а ребрами - переходы по терминальным символам</p>
8.	Чем характеризуются синтагматические отношения между терминами в тексте?	<p>1) Отношениями синонимии и гипонимии (родо-видовыми связями) в словаре</p>

		<p>2) Линейной сочетаемостью, совместной встречаемостью в одном контексте или предложении</p> <p>3) Фонетическим сходством слов</p>
9.	Какую функцию выполняет язык при представлении информации, обеспечивая возможность ее хранения и передачи во времени?	<p>1) Кумулятивную</p> <p>2) Коммуникативную</p> <p>3) Мемориальную</p>
10.	Как в семиотике называется отношение между знаком и обозначаемым им объектом реальности?	<p>1) Прагматика</p> <p>2) Синтактика</p> <p>3) Семантика</p>

Примерный перечень вопросов к зачету

1. Место и назначение лингвистического обеспечения, как одной из обеспечивающих подсистем информационной системы (ИС), в контексте этапов жизненного цикла ИС.
2. Моделирующая функция языка. Понятие, суждение, умозаключение.
3. Коммуникативная функция языка. Знаковое представление информации.
4. Семантический и прагматический подходы к измерению информации. Соотношение понятий “информация” и “данные”.
5. Семиотический треугольник. Понятие и знак. Текст как знак и как сообщение.
6. Естественно-языковые интерфейсы. Морфологический анализ и синтез форм слов.
7. Синтаксис, словосочетания, предложения. Подходы к автоматизации процесса анализа слов и предложений.
8. Основные компоненты ИС, применяющие данные подходы. Инструментальные средства, выполняющие данные работы.
9. Формальные методы описания искусственных языков. Грамматический разбор.
10. Отображение правил регулярной грамматики на состоянии конечного автомата.
11. Выделение лексических единиц языка с использованием конечных автоматов. Действующие стандарты в данной области.
12. Формализованные представления информации в процессе разработки ИС.
13. Языковые средства, создаваемые на основании анализа предметных областей, описываемых точными переменными
14. Соотношение основных понятий: знак, понятие, сущность, свойство, характеристические признаки, связи.
15. Ролевой и понятийный подходы к выделению отношений.
16. Теоретико-множественные модели идентификации.
17. Типология и способы задания признаков. Методы абстрагирования.
18. Теоретико-множественная модель представления объектов. Использование классификаторов, кодификаторов, нормативных списков, тезаурусов.
19. Методы и средства структурирования информационных запросов.
20. Иерархический классификатор предметной области.
21. Конструкторы запросов. Поиск по ссылкам (гипертекст).
22. Теоретико-множественное и линейные представления сообщений, запросов, массивов документов, тезауруса.

23. Матрицы ассоциации документов, терминов и их свойства.
24. Модели и меры оценки силы парадигматических и синтагматических отношений.
25. Меры, основанные на структурно-частотных характеристиках терминов.

СЕМИНАРЫ/ ПРАКТИЧЕСКИЕ ЗАДАНИЯ

Практическое 1. Естественно-языковые интерфейсы. Подходы к автоматизации процесса анализа слов и предложений. Основные компоненты информационных систем, применяющие данные подходы. Инструментальные средства, выполняющие данные работы. Синтаксические деревья. Задачи грамматического разбора и вывода. Отношения над множествами. Алгоритмы эквивалентного преобразования грамматик. Понятие автоматной грамматики. Преобразование конечного автомата. Выделение лексических единиц языка с использованием конечных автоматов.

Методические рекомендации по выполнению:

- Изучите основные подходы к автоматическому анализу текста (морфологический, синтаксический, семантический).
- Разберите архитектуру компонентов информационных систем, работающих с естественным языком (морфологический анализатор, синтаксический парсер).
- Освойте инструментальные средства для построения синтаксических деревьев (например, NLTK в Python или онлайн-парсеры).
- Проанализируйте алгоритмы преобразования грамматик и приведите пример приведения грамматики к автоматному виду.
- Реализуйте алгоритм выделения лексических единиц (токенов) с использованием конечного автомата.

Критерии оценки:

- Правильность построения синтаксического дерева для заданного предложения.
- Корректность выполнения эквивалентного преобразования грамматики.
- Работоспособность реализованного автомата для выделения лексем.
- Глубина анализа компонентов информационных систем.

Практическое 2. Применение таблицы решений в процессе разработки интерпретаторов языков предметно-ориентированных информационных систем.

Методические рекомендации по выполнению:

- Изучите структуру таблицы решений (условия, действия, правила).
- Выберите предметную область для разработки интерпретатора (например, расчет налогов, проверка кредитной заявки, логика работы чат-бота).
- Постройте таблицу решений, описывающую логику работы выбранной системы.
- Разработайте алгоритм интерпретации таблицы решений (обход правил, проверка условий, выполнение действий).
- Реализуйте интерпретатор на языке программирования высокого уровня.

Критерии оценки:

- Полнота и непротиворечивость построенной таблицы решений.
- Корректность работы алгоритма интерпретации.
- Соответствие реализованного интерпретатора исходной таблице.
- Возможность модификации правил без изменения кода программы.

ТЕСТОВЫЕ ЗАДАНИЯ ПО ДИСЦИПЛИНЕ

Вариант 1

Открытые вопросы:

1. Что такое лингвистическое обеспечение ИС? Опишите его место и назначение на этапах жизненного цикла информационной системы.
2. Раскройте суть моделирующей функции языка. Приведите примеры понятий, суждений и умозаключений в контексте информационных систем.
3. Что представляет собой семиотический треугольник? Объясните, как текст может выступать одновременно знаком и сообщением.
4. Опишите подходы к автоматизации синтаксического анализа слов и предложений. Приведите примеры инструментальных средств для выполнения этой задачи.
5. Что такое формально-языковые модели в разработке ИС? Приведите 2–3 примера их применения.

Закрытые вопросы:

1. Какая функция языка отвечает за передачу информации между субъектами?
 - моделирующая;
 - коммуникативная;
 - когнитивная;
 - экспрессивная.
2. Что из перечисленного НЕ относится к компонентам лингвистического обеспечения ИС?
 - словари и тезаурусы;
 - алгоритмы морфологического анализа;
 - базы данных пользователей;
 - грамматические правила формальных языков.
3. Какой подход к измерению информации учитывает её смысловую нагрузку?
 - синтаксический;
 - семантический;
 - статистический;
 - алгоритмический.
4. Что отображает конечный автомат в контексте обработки естественного языка?
 - правила регулярной грамматики;
 - структуру базы данных;
 - архитектуру ИС;
 - алгоритм шифрования.
5. Какое средство используется для структурирования информационных запросов?
 - иерархический классификатор;
 - операционная система;
 - сетевой протокол;
 - система резервного копирования.
6. Что такое тезаурус в контексте лингвистического обеспечения?
 - словарь синонимов;
 - база данных пользователей;
 - набор программных модулей;
 - протокол передачи данных.
7. Какой метод абстрагирования основан на выделении общих свойств объектов?
 - классификация;
 - агрегация;
 - композиция;

- наследование.
- 8. Что представляют собой матрицы ассоциации документов и терминов?
 - таблицы частот встречаемости терминов в документах;
 - схемы баз данных;
 - сетевые топологии;
 - алгоритмы шифрования.
- 9. Какой тип отношений отражает связь «часть-целое»?
 - парадигматические;
 - синтагматические;
 - иерархические;
 - ассоциативные.
- 10. Что такое гипертекст с точки зрения лингвистического обеспечения?
 - текст с гиперссылками для навигации;
 - зашифрованный текст;
 - сжатый текст;
 - многоязычный текст.

Вариант 2

Открытые вопросы:

1. Объясните коммуникативную функцию языка и её роль в знаковом представлении информации в ИС. Приведите пример знаковой системы в информационных технологиях.
2. В чём разница между понятиями «информация» и «данные» с семантической и прагматической точек зрения? Приведите примеры.
3. Опишите процесс морфологического анализа и синтеза форм слов. Какие задачи он решает в естественно-языковых интерфейсах?
4. Что такое теоретико-множественная модель представления объектов? Как используются классификаторы и кодификаторы в ИС?
5. Объясните ролевой и понятийный подходы к выделению отношений. Приведите пример их применения в проектировании ИС.

Закрытые вопросы:

1. Какой компонент лингвистического обеспечения отвечает за анализ структуры предложений?
 - морфологический анализатор;
 - синтаксический анализатор;
 - семантический интерпретатор;
 - прагматический модуль.
2. Что такое регулярная грамматика в контексте формальных языков?
 - набор правил для генерации строк определённого вида;
 - метод шифрования данных;
 - структура базы данных;
 - протокол сетевого взаимодействия.
3. Какой стандарт регулирует представление лингвистических данных в ИС?
 - ГОСТ Р ИСО/МЭК 9126;
 - Unicode;
 - SQL;
 - HTTP.
4. Что такое конструкторы запросов в ИС?

- инструменты для формирования поисковых запросов;
 - модули шифрования;
 - компоненты интерфейса пользователя;
 - системы резервного копирования.
5. Какой метод используется для оценки силы парадигматических отношений?
 - анализ частоты совместной встречаемости терминов;
 - расчёт длины текста;
 - измерение времени обработки запроса;
 - подсчёт количества гиперссылок.
 6. Что такое кодификатор в ИС?
 - справочник с кодами для объектов предметной области;
 - программа для шифрования;
 - устройство ввода данных;
 - сетевой адаптер.
 7. Какой подход к структурированию информации использует иерархию классов?
 - теоретико-множественный;
 - объектно-ориентированный;
 - реляционный;
 - сетевой.
 8. Что такое синтагматические отношения?
 - связи между элементами в линейной последовательности (например, в предложении);
 - отношения синонимии;
 - иерархия классов;
 - ассоциации по смыслу.
 9. Какой инструмент используется для поиска по ссылкам в гипертекстовых системах?
 - браузер;
 - компилятор;
 - отладчик;
 - архиватор.
 10. Что такое характеристический признак в моделировании предметной области?
 - свойство объекта, используемое для его идентификации;
 - тип данных;
 - идентификатор пользователя;
 - пароль доступа.

Ключи к тестам

Вариант 1:

Ответы на открытые вопросы:

1. Лингвистическое обеспечение — совокупность языковых средств для взаимодействия пользователей с ИС. Оно задействовано на всех этапах ЖЦ: анализ требований (формулировка сценариев), проектирование (разработка тезаурусов), реализация (создание интерфейсов), эксплуатация (поддержка пользователей).
2. Моделирующая функция — отображение реальности через язык. Понятие: «пользователь ИС»; суждение: «Пользователь вводит запрос»; умозаключение: «Если запрос корректен, система вернёт результат».

3. Семиотический треугольник: знак (слово) — понятие (значение) — объект (реальный предмет). Текст как знак передаёт смысл, как сообщение — несёт информацию для получателя.
4. Подходы: правила на основе грамматик, статистические модели (n-граммы), нейронные сети. Инструменты: NLTK, SpaCy, Stanford NLP.
5. Формально-языковые модели — строгие описания синтаксиса и семантики (BNF, EBNF). Примеры: SQL для запросов, XML для разметки данных, UML для моделирования.

Ответы на закрытые вопросы:

1. коммуникативная
2. базы данных пользователей
3. семантический
4. правила регулярной грамматики
5. иерархический классификатор
6. словарь синонимов
7. классификация
8. таблицы частот встречаемости терминов в документах
9. иерархические
10. текст с гиперссылками для навигации

Вариант 2:

Ответы на открытые вопросы:

1. Коммуникативная функция — передача информации через знаки. Пример знаковой системы: URL (адрес ресурса) как знак, указывающий на веб-страницу.
2. Данные — сырые факты (числа, текст), информация — данные с контекстом и смыслом. Семантика: «37 °C» — температура; прагматика: «37 °C» означает лихорадку в медицинском контексте.
3. Морфологический анализ — определение части речи и формы слова; синтез — генерация форм. Задачи: нормализация запросов, исправление опечаток, поддержка многоязычных интерфейсов.
4. Теоретико-множественная модель — представление объектов как элементов множеств. Классификаторы (ОКПД) и кодификаторы (ISBN) структурируют данные для поиска и анализа.
5. Ролевой подход: выделение отношений по функциям («автор пишет статью»). Понятийный: группировка по семантике («статья» относится к «публикациям»). Пример: проектирование базы данных научных статей.

Ответы на закрытые вопросы:

1. синтаксический анализатор
2. набор правил для генерации строк определённого вида
3. Unicode
4. инструменты для формирования поисковых запросов
5. анализ частоты совместной встречаемости терминов
6. справочник с кодами для объектов предметной области
7. объектно-ориентированный
8. связи между элементами в линейной последовательности (например, в предложении)

9. браузер

10. свойство объекта, используемое для его идентификации