

Министерство образования Республики Беларусь

Учреждение образования

“Белорусский государственный университет информатики и  
радиоэлектроники”

Кафедра информационных интеллектуальных технологий

**Лабораторная работа 6**

**“Диалоговая система с поддержкой  
естественного языка”**

Выполнил

гр.121701

Мулярчик Д.С.

Лемантович Д.К.

Проверил

Крапивин Ю. Б

Минск 2024

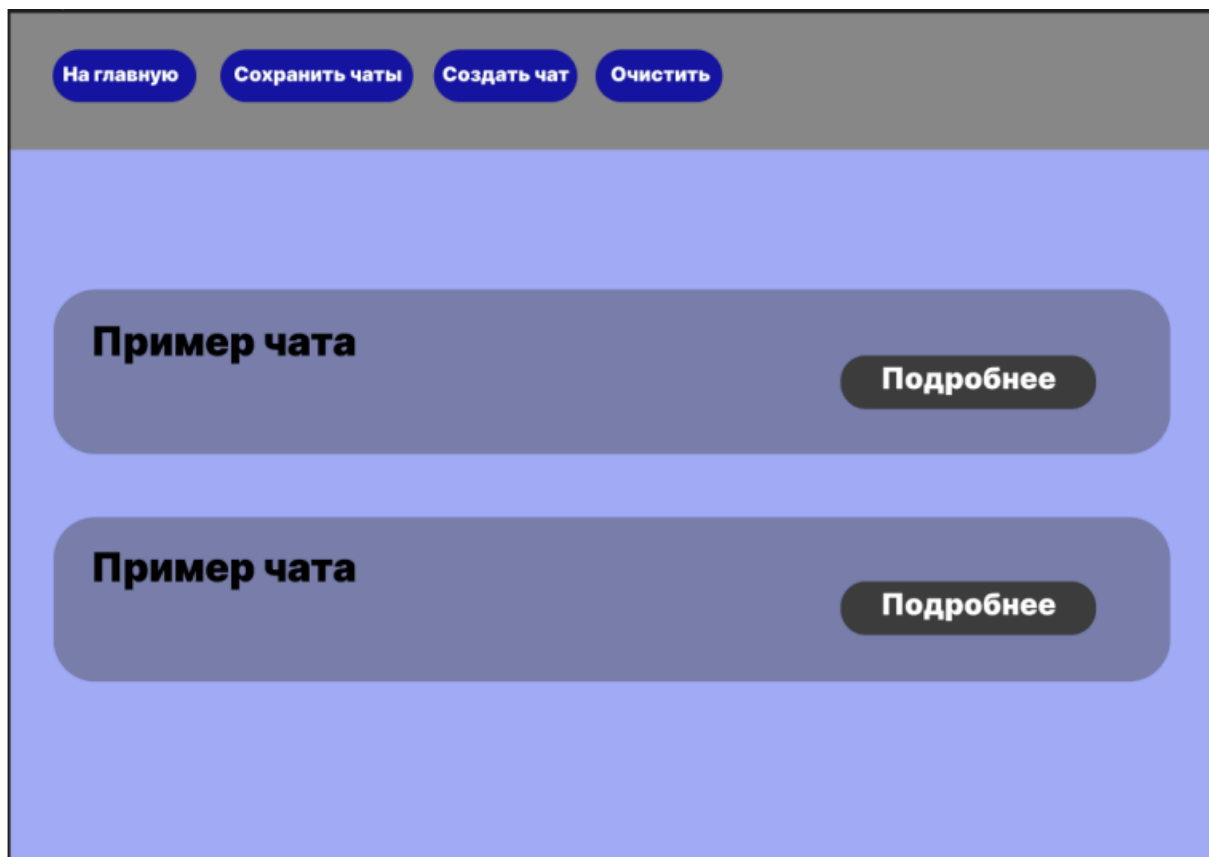
**Цель работы:** Освоить принципы разработки диалоговых систем с поддержкой естественного языка

**Задание:** Изучить основы создания диалоговых систем с поддержкой естественного языка. Закрепить навыки программирования при решении задач организации диалогового взаимодействия с поддержкой естественного языка

**Задача:** Предметная область: “Транспорт”. Язык: русский.

## Макет интерфейса

Главная страница



Страница чата

На главную

Сохранить чаты

Создать чат

Очистить

X

Ответ

Вопрос

Введите вопрос

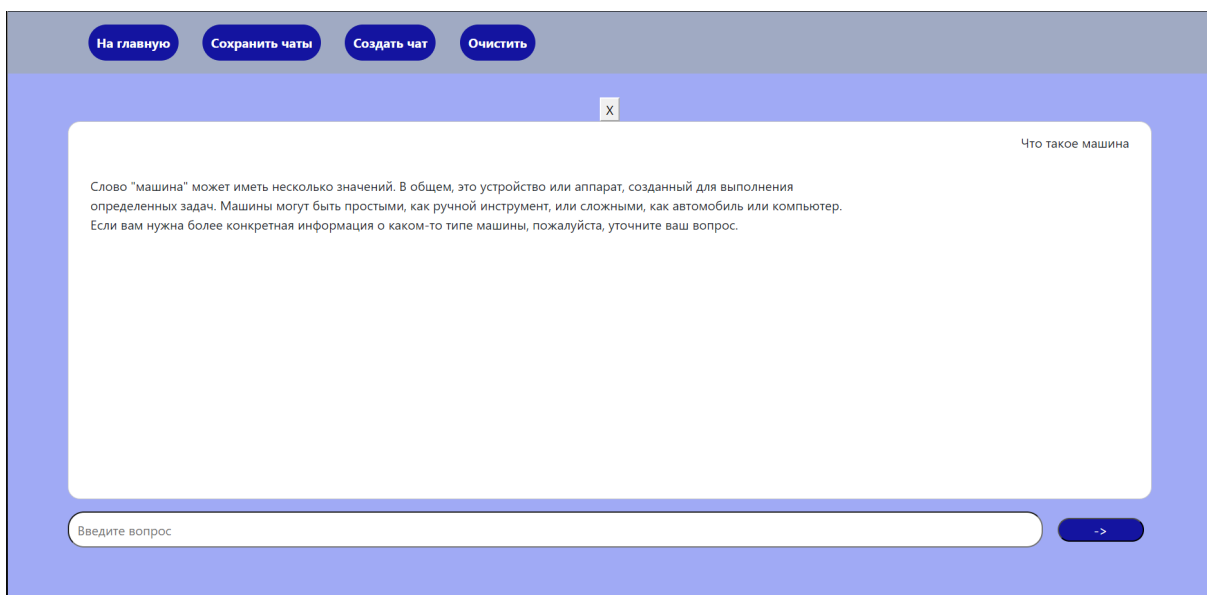
→

**Реализованный интерфейс**

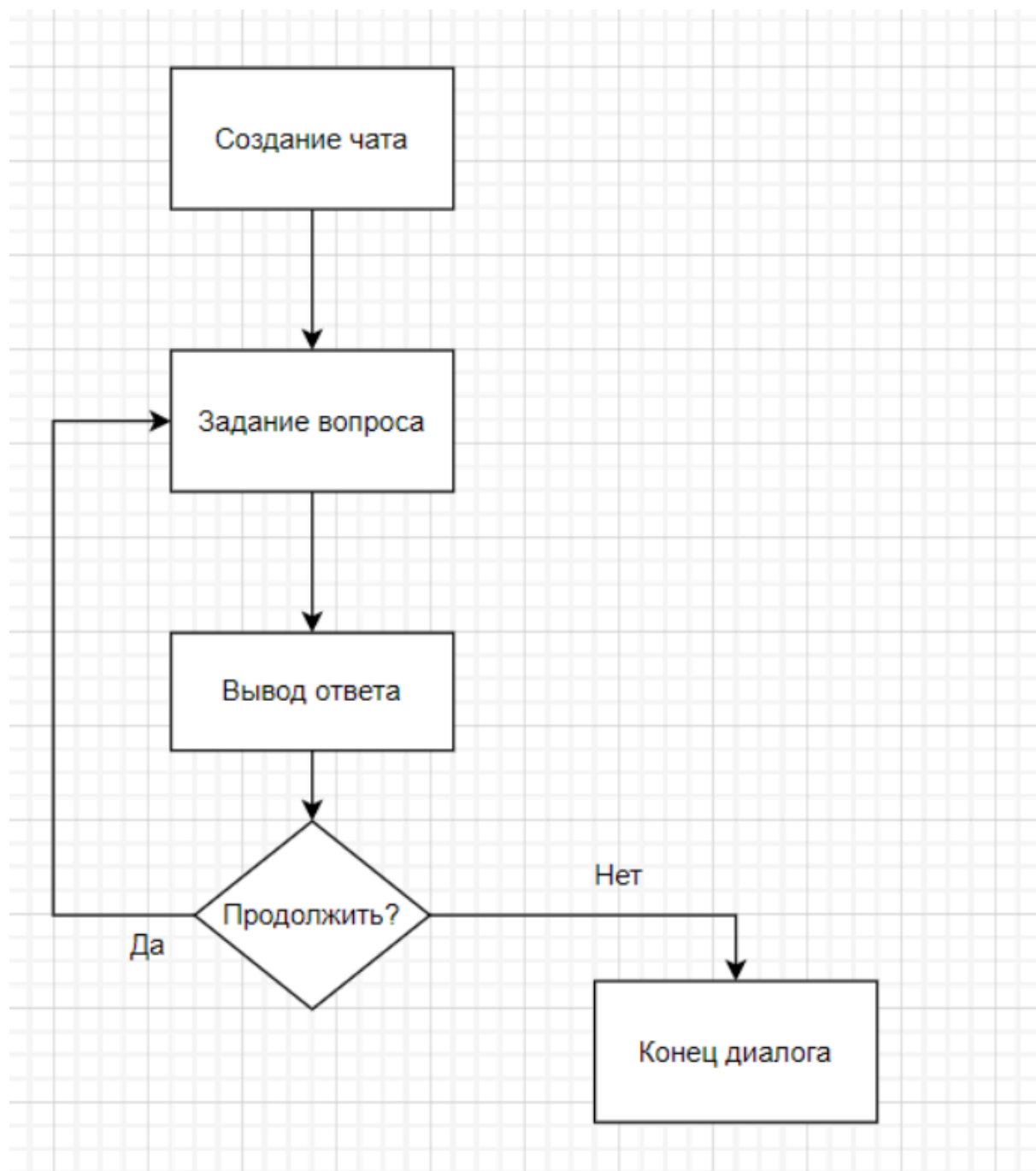
## Главная страница



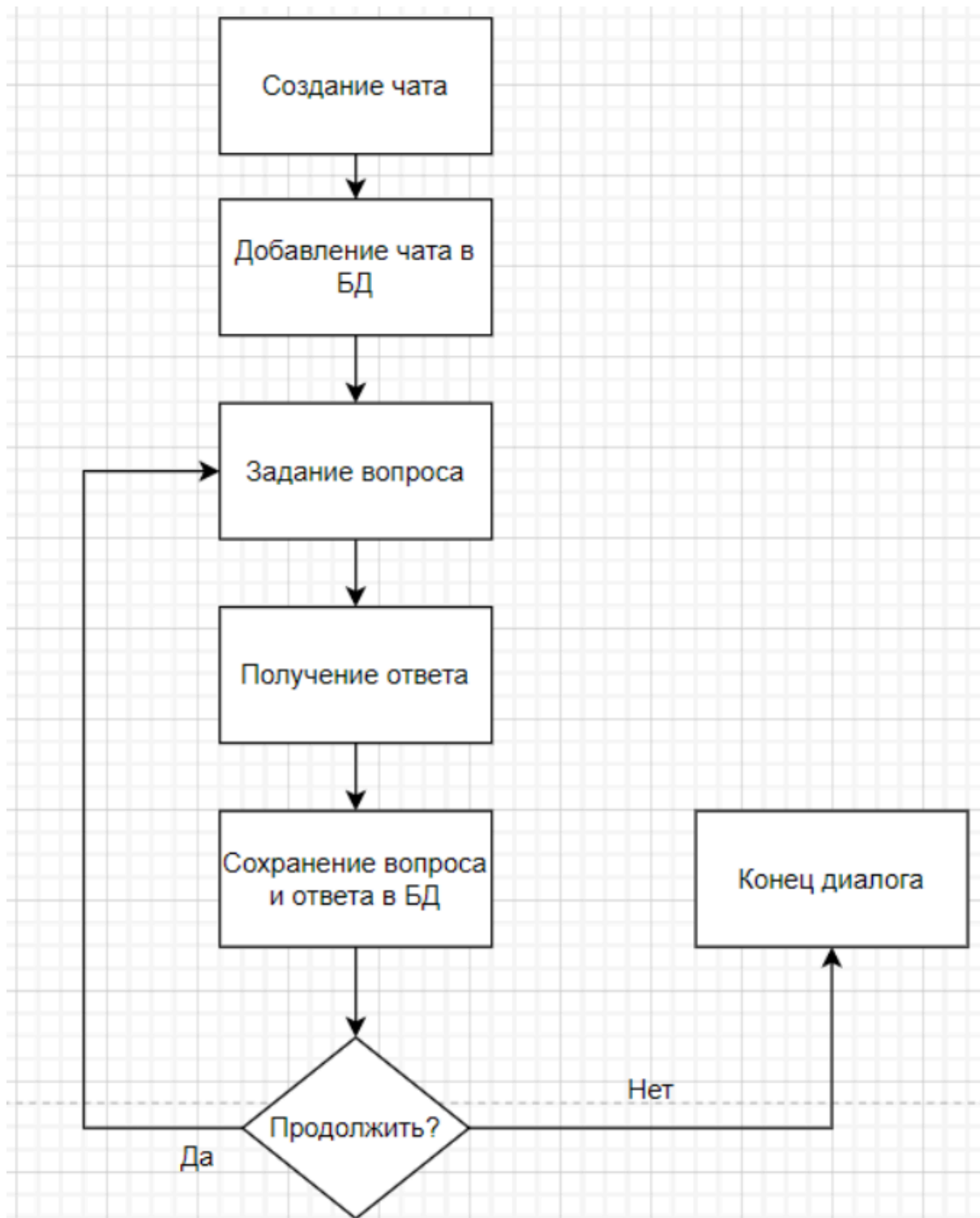
## Страница чата



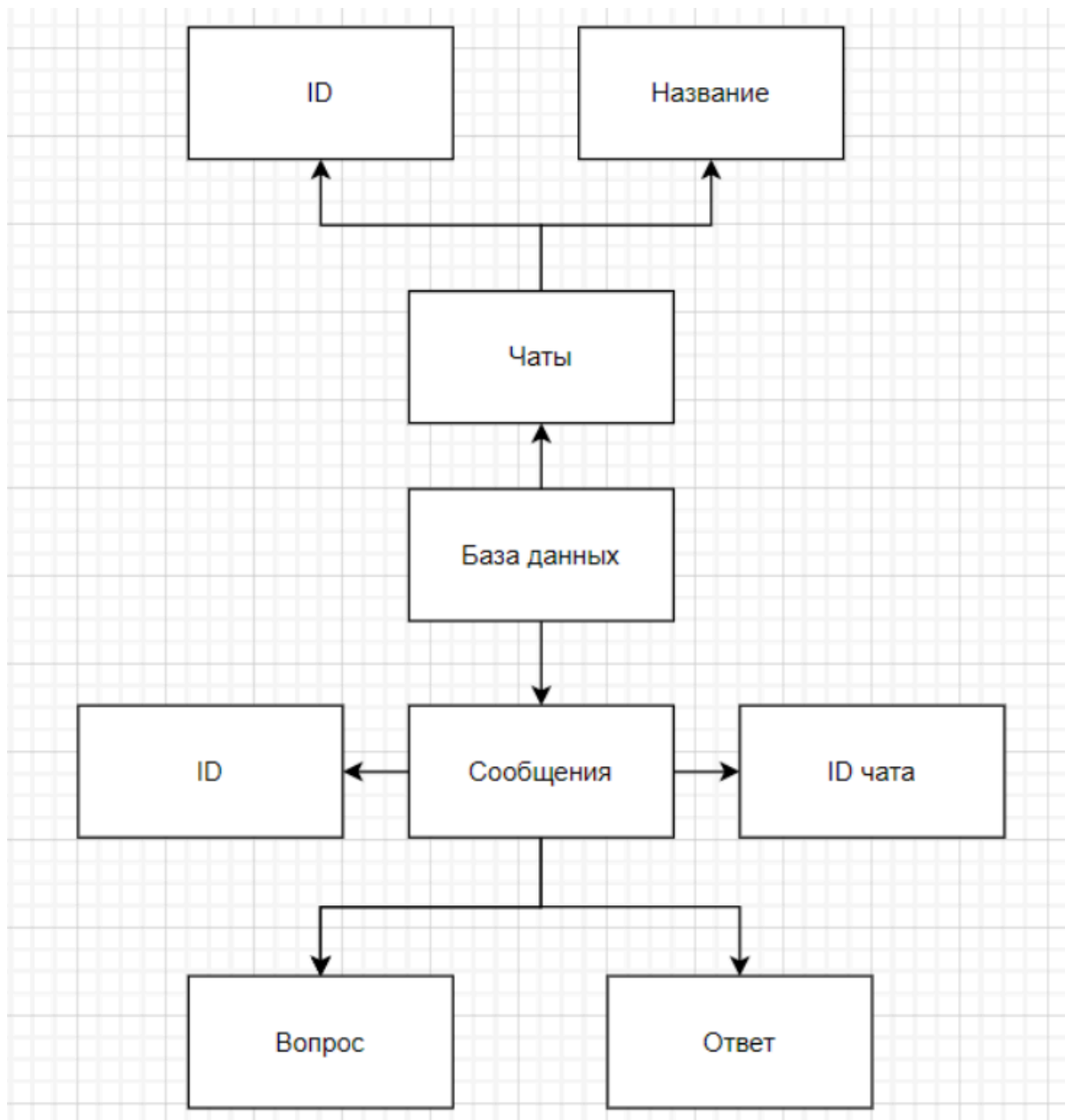
## Логическая структура диалога



Структура данных для хранения входной и выходной информации:



## Структура хранения данных



## Структурная схема приложения



**Вопросы:**



Типовая структура базы знаний для решения задач автоматической обработки текста естественного языка.

Типовая структура базы знаний для решения задач автоматической обработки текста естественного языка может включать следующие компоненты:

Лексическая база данных: В этом компоненте содержится информация о словах, их морфологических формах, синонимах, антонимах и других лексических связях.

Грамматические правила: Здесь содержатся правила грамматики, которые определяют синтаксическую структуру предложений.

Семантические правила: В этом компоненте содержатся правила, определяющие семантические связи и отношения между словами и фразами.

Правила вывода и логические правила: Здесь содержатся правила для логического вывода и рассуждений. Логические правила, которые позволяют делать выводы на основе имеющихся знаний.

Статистические модели: Это может включать модели машинного обучения, используемые для классификации текстов, извлечения информации или определения тональности текста.

База фактов и знаний: Здесь содержится фактическая информация, которая используется для решения конкретных задач.

Уровни изучения текста, связь с разделами лингвистики.

Фонетический и фонологический уровень: Этот уровень изучения текста связан с звуковыми аспектами языка. Он включает анализ звуковой структуры слов и произношения, а также исследование фонологических правил и закономерностей, определяющих звуковую организацию языка.

Морфологический уровень: На этом уровне изучается структура слов и их формы.

Синтаксический уровень: Синтаксический уровень изучает структуру предложений и отношения между словами в рамках предложения.

Семантический уровень: На семантическом уровне изучается смысловое содержание слов, фраз и предложений..

Прагматический уровень: Прагматический уровень изучает использование языка в контексте коммуникации.

=====

Системы составляющих.

Системы составляющих (constituency parsing systems) являются методами анализа синтаксической структуры предложений, которые строят деревья составляющих для предложений. Дерево составляющих представляет структуру предложения в виде иерархической композиции фразовых групп (constituents), таких как подлежащее, сказуемое, определения и т. д.

Системы составляющих включают в себя набор правил, которые определяют, какие фразовые группы могут сочетаться друг с другом и как они могут быть комбинированы для образования более крупных фразовых групп. Алгоритмы составляющего анализа применяют эти правила для структурирования предложения и построения его дерева составляющих.

Вот некоторые известные системы составляющих:

**СҮК-алгоритм (Cocke-Younger-Kasami):** Это один из наиболее известных алгоритмов для составляющего анализа. Он основан на динамическом программировании и может работать с контекстно-свободной грамматикой.

**Earley-алгоритм:** Этот алгоритм использует метод динамического предсказания и распространения для построения дерева составляющих. Он может работать с произвольными контекстно-свободными грамматиками.

**Статистические модели:** Некоторые системы составляющих используют статистические модели, основанные на машинном обучении. Эти модели могут обучаться на размеченных корпусах предложений и использовать статистические признаки для прогнозирования структуры предложения.

**Lexicalized Parsing:** Этот подход к составляющему анализу учитывает лексическую информацию о словах, такую как их части речи и лексические категории. Он позволяет более точно моделировать синтаксические зависимости и улучшает производительность систем составляющих.

Автоматизация обработки текста: этап синтаксического анализа.

Этап синтаксического анализа в автоматизации обработки текста является одним из ключевых этапов и заключается в анализе синтаксической структуры предложений. Главная задача этого этапа состоит в определении синтаксических отношений между словами и фразами в тексте.

Вот некоторые шаги, которые обычно выполняются на этапе синтаксического анализа:

Токенизация: Исходный текст разбивается на отдельные токены, такие как слова или пунктуационные знаки. Токенизация помогает выделить основные элементы текста, с которыми будет работать алгоритм.

Частеречная разметка: Каждому токenu присваивается частеречная метка, которая указывает на его часть речи (существительное, глагол, прилагательное и т. д.). Частеречная разметка позволяет определить грамматическую роль каждого слова в предложении.

Синтаксический анализ: На этом этапе проводится анализ синтаксических связей между словами в предложении. Он включает в себя построение синтаксического дерева или графа зависимостей, который отображает синтаксические отношения между словами (например, подлежащее-сказуемое, определение-определяемое и т. д.). Алгоритмы, используемые для синтаксического анализа, могут быть основаны на правилах грамматики, статистических моделях или комбинации обоих подходов.

Извлечение структурированной информации: После проведения синтаксического анализа можно извлечь структурированную информацию

из текста. Например, можно извлечь подлежащее и сказуемое, определить связи между сущностями или выделить важные фразы или конструкции.