

УДК 007:519.816

© А. А. Юминов  
yulof@mail.ru

## ЭКСПЕРТНАЯ СИСТЕМА СИНТЕЗА МОРФОЛОГИЧЕСКОЙ, СИНТАКСИЧЕСКОЙ И СЕМАНТИЧЕСКОЙ МОДЕЛЕЙ ТЕКСТА НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

**Ключевые слова:** морфологический анализ, синтаксический анализ, семантика текста, правила, логический вывод.

**Abstract.** In article take up methods of analysis natural language texts and logical synthesis of morphological, syntactical and semantical subject models conformably to interactive learning systems and natural language computer- aided system interfaces. Builded models are showing in tabular and graphical form. Expert can edit logical deduction rules and algorithm manage meta-rules. User can verify system understanding of text in visual form and can revise text and text models.

Одним из направлений совершенствования современных программных продуктов является приближение языка взаимодействия программ с пользователями к естественному. Идеальный вариант организации взаимодействия – проводить его на естественном языке. Это положение дел приводит к актуальности «понимания» компьютерами текста, особенно в автоматизированных обучающих системах.

**Модель знаний.** Иерархия элементов БЗ: *Дерево* -> *Концепт* -> *Экземпляр*

Дерево задает иерархию концептов. Концепт базы знаний (понятие) определяет подмножество экземпляров, у которых значение параметров удовлетворяют данному понятию, то есть экземпляры являются листьями деревьев [1].

**Семантический анализ** проводится с использованием машины логического вывода на основе базы знаний, содержащей правила, словари и справочники, а также структуру рабочей модели. Модели текста и модели модель сюжета хранятся в рабочей модели (в базе данных фактов).

**Описание правил.** Правила анализа можно разбить на два вида: метаправила, управляющие алгоритмом анализа, и, собственно, правила, описывающие закономерности.

Предусловия правил представляют собой дизъюнкцию конъюнкций элементарных условий. Каждое такое элементарное условие может быть представлено как функция логического типа. Следствия правил в общем случае хранят список процедур с параметрами.

**Описание рабочей модели.** Рабочая модель предназначена для хранения исходных и текущих фактов в структурированном виде. Ее структура описывается концептами.

Подобная организация рабочей модели удобна для ведения истории и просмотра состояния модели в пошаговом режиме. Рабочая модель данных отображается в табличном виде. Может быть разработан специализированный графический интерфейс с возможностью редактирования.

**Укрупненный алгоритм логического вывода.** Назовем *задачей* последовательность групп правил. Выполнить задачу — значит выполнить правила в порядке расположения их в группах, а групп в задаче с учетом счетчиков и с соответствующими возвратами и прерываниями, декларированными в процедурах управления алгоритмом.

**Семантический анализ** выполняется поэтапно. Данный текст разбивается на слова, выполняется синтез и уточнение морфологической модели текста. Далее синтезируется начально заполненная синтаксическая модель данного текста. Модель уточняется на основе правил синтаксического анализа. После этого каждому слову ставится в соответствие понятие из базы знаний предметной области, синтезируется и уточняется модель сюжета

на основе правил семантического анализа [2, 3].

Целью **морфологического анализа** является установление морфемного состава слова, а также морфологических признаков, используемых в задачах синтаксического и семантического анализа.

**База знаний морфологического анализа** состоит из списка перечислимых типов (описывающих морфологические признаки), словарей (описывающих значение морфем), справочников (описывающих взаимосвязь морфем и признаков), структуры рабочей модели данных (для хранения текстов, слов и вариантов разбора слов) и метаправил (описывающих последовательность разбора слов разных частей речи).

Цель **синтаксического анализа** — определение границ простых предложений, описывающих процесс, участников процесса (которые могут быть уточнены).

**База знаний синтаксического анализа** содержит перечислимые типы; правила анализа; метаправила управления алгоритмом анализа, а также структуру рабочей модели (содержащую элементы: текст, предложение, вариант разбора предложения, простое предложение, слово и словосочетание, отношения).

**Описание правил.** Синтаксический анализ проводится на основе правил. Правила группируются по назначению в группы (этапы анализа): поиск словосочетаний, поиск ролей слов, поиск простых предложений, восстановление слов и связей из контекста.

Целью **семантического анализа** является построение модели сюжета, описанного текстом.

База знаний семантического анализа состоит из тезауруса, правил, метаправил, дерева классификации понятий.

Семантический анализ строится на основе этапов: поиск понятий, соответствующих словам; поиск объектов сюжета; поиск процессов; построение временной диаграммы.

Подсистема семантического анализа текста может быть использована в интерактивных обучающих обучающих системах и в естественных языковых интерфейсах различных автоматизирован-

ных систем: модель сюжета может быть интерпретирована как запрос к базе данных (знаний) на основе интерпретируемых шаблонов.

На пути создания коммерчески успешного приложения, использующего семантический анализ стоит ряд сложностей к которым относятся, прежде всего, сложность грамматики языка, большое количество понятий предметных областей, динамичность языка, а также многозначность, присущая как на уровне грамматики, так и на смысловом уровне, что приводит к большим объемам и динамичности баз знаний. В этом свете особую роль играет наглядность, интуитивная понятность и удобство работы с пополнением/модификацией баз знаний. Автору видится положительным в этом контексте применение описанного подхода к организации семантического анализа.

Система построена на основе СУБЗ КГ[4].

### Список литературы

1. Кучуганов В.Н., Семантика графической информации // материалы международной научно-технической конференции «Интеллектуальные САПР». Таганрог: Издательство ТРТУ, 2002, №3(26) с. 157-166
2. Попов Э.В., Общение с ЭВМ на естественном языке //М., Наука, 1982
3. Искусственный интеллект. – В 3-х кн. Кн. 1. Системы общения и экспертные системы: Справочник. /Под ред. Д.А. Поспелова. - М.: Наука 1990
4. Кучуганов В.Н., Габдрахманов И.Н. Планирование задач в сложноструктурированных ситуациях // Труды международных научно-технических конференций «Интеллектуальные системы (IEEE AIS'04)» и «Интеллектуальные САПР» (CAD 2004). Научное издание в 3-х томах. М.: Изд-во Физико-математической литературы, 2004, Т.1. - С. 214-223.