

А.А. Єгошина

Государственный университет информатики и искусственного интеллекта

Средства анализа естественно-языковых on-line ответов обучаемых для систем web-обучения

В последние годы активизировались теоретические и прикладные работы по развитию комплексной и многоаспектной лингвистической стороны проблемы человеко-машинного диалога в процессе Web-обучения [1-5].

Знания об индивидуальной грамматической модели языка общения пользователя системы Web-обучения являются необходимой составляющей общей модели обучаемого. Возможности компьютерной системы по приобретению грамматических знаний, формированию комплекса грамматических словарей позволяют решать проблему адаптации ЕЯ-интерфейса при смене языка общения.

Таким образом, для повышения интеллектуального уровня средств общения пользователей с системами Web-обучения актуальной является разработка программных средств, расширяющих лингвистическое обеспечение существующих систем.

Задачей лингвистического процессора (ЛП) является преобразование естественно-языкового предложения (или даже целого текста), являющегося online-ответом обучаемого, в некоторый набор семантических структур, являющихся формальным представлением “смысла” ответа. Цель такого преобразования - обеспечить исходные данные для работы подсистемы тестирования средств управления Web-обучением.

ЛП выполнен в виде библиотеки, доступной различным приложениям системы управления Web-обучением и в первую очередь подсистемы тестирования. Общая структура лингвистического процессора и связи с компонентами системы Web-обучения, приведены на рисунке 1.

Рассмотрим основные программные блоки ЛП, их словарная поддержка, взаимодействие с пользователем, а также структуры данных, передаваемые между программными блоками.

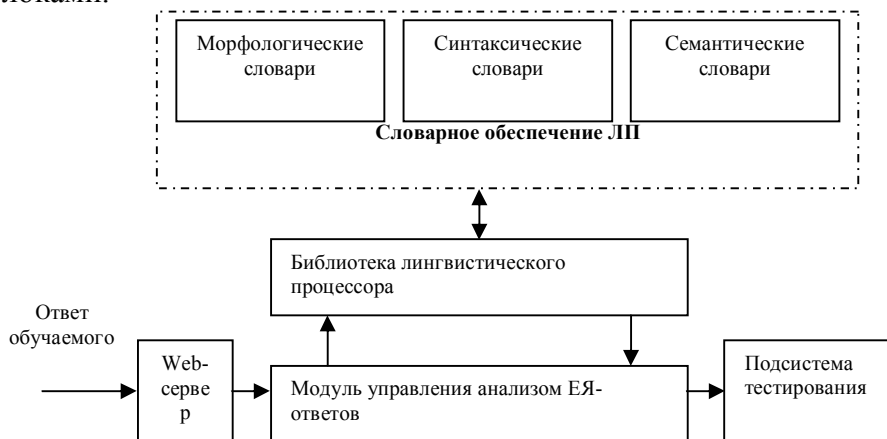


Рисунок 1 - Структура лингвистического процессора

Блок лексического анализа. В ЛП используется простейший лексический анализатор, который выполняет вспомогательные функции, не имеет возможности настройки или взаимодействия с пользователем и поэтому реализован как часть блока морфологического анализа.

Блок лексического анализа принимает исходный текст непосредственно от элементов пользовательского интерфейса — а именно, от текстового редактора. Анализируемое предложение попадает на вход лексического анализатора в виде массива ASCII-символов, содержащего прописные и строчные буквы русского алфавита, цифры, знаки пунктуации [5]. Вместе с указателем на массив символов в процедуру лексического разбора передается переменная - счетчик байт в этом массиве.

Полученный массив анализатор должен преобразовать в массив лексических единиц. Здесь под этим термином подразумевается слово, число или знак препинания. Для каждой лексической единицы формируется отдельная строка, в которую копируются все символы, принадлежащие данной лексической

единице [10]. При этом удаляются пробелы, символы переноса, конца строки и незнакомые символы.

Указатели на все сформированные таким образом строки анализатор заносит в выходной динамический массив, который является результатом его работы.

Блок морфологического анализа. В этом модуле присутствуют не только функции морфологического анализа, но и средства поддержки и редактирования словарных файлов, необходимых анализатору. Исходными данными для работы морфологического анализатора служат результаты предварительного лексического анализа.

Для каждой строки, полученной из массива лексических единиц, анализатор формирует одну или несколько записей. Несколько записей составляется для тех слов, для которых поиск в словаре дал неоднозначный результат, т. е. было найдено несколько омонимов.

Лингвистическая информация, необходимая для выполнения морфологического анализа, представлена в словарных файлах: файл словаря основ; файл аффиксов для различных частей речи; файл исключений.

Основной критерий при разбиении слова на основу и аффикс — основа должна оставаться неизменной во всех возможных словоформах данного слова. Из словаря основ поочередно выбираются все основы, совпадающие с начальными буквами анализируемого слова, и для каждой такой основы перебираются все возможные для нее аффиксы. В случае точного совпадения очередного варианта “основа+аффикс” с анализируемым словом вариант анализа считается успешным, и в программу передается морфологическая информация, соответствующая данной основе и данному аффиксу. При этом, как правило, постоянные морфологические параметры определяются основой слова, а переменные — аффиксом.

Блок синтаксического анализа реализует следующие возможности [4]:

- многовариантный анализ с последовательной выдачей вариантов;

- введение ограничений на время выполнения анализа;
- выдача частичных вариантов разбора при невозможности сформировать ни одного полного варианта;
- отсутствие ограничений на тип условий, проверяемых правилами синтагматики, а также на действия, выполняемые правилами;
- возможность автоматически генерировать исходные тексты для вновь создаваемых правил анализа.

Блок семантического анализа. Наиболее простым и универсальным средством представления знаний в системах искусственного интеллекта является семантическая сеть. По роду хранимой информации выделяют два типа семантических сетей: А-сети и К-сети [6]. Первые содержат множество объектов и отношений, допустимых в данной предметной области; вторые — множество объектов и отношений, присутствующих в описании конкретной ситуации. Задачей семантического анализатора ЛП является преобразование синтаксического дерева зависимостей в соответствующий фрагмент К-сети. При этом может выполняться проверка допустимости каждого семантического отношения по опорной А-сети. Семантическая сеть может представляться двумя способами: либо в виде списка вершин, для каждой из которых показываются все инцидентные ребра и соответствующие смежные вершины, либо в виде списка ребер, с указанием начальной и конечной вершины для каждого ребра [1].

Предложенный лингвистический процессор позволяет повысить адаптивность системы Web – обучения. Извлеченный из ЕЯ- ответа «смысл», представленный в виде К-сети поступает в подсистему тестирования. На основе анализа семантики ответа обучаемого в дальнейшем подсистемой представления знаний формируется самая подходящая индивидуально спланированная последовательность модулей знаний для обучения и работы с определенным порядком следования обучающих заданий (примеров, вопросов, задач и т.п.). Тем самым для обучаемого строится “оптимальный путь” сквозь обучающий материал.

Применение разработанных средств ЕЯ – общения в системах Web- обучения позволяет расширить возможности и интеллектуальность таких систем за счет реализации обратной связи с обучаемыми на качественно новом уровне. Предлагаемые средства интерфейса позволяют проводить интеллектуальный анализ решений обучаемого. В отличие от не интеллектуальных проверяющих программ, которые не могут сказать ничего более чем ответ правильный или нет, интеллектуальные анализаторы позволяют сказать, что именно неправильно или неполно и какие отсутствующие или неверные знания ответственны за ошибку. Интеллектуальные анализаторы могут предоставлять обучаемому интенсивную обратную связь об ошибках и корректировать модель обучаемого.

Кроме интеллектуального анализа ответов, предложенные средства интерфейса обеспечивают адаптивную поддержку навигации обучаемого в гиперпространстве, изменяя появление видимых ссылок. На основе ответов система может адаптивно сортировать, аннотировать или частично прятать ссылки на текущей странице для облегчения выбора пользователем следующей ссылки.

1. Бухараев Р.Г., Сулейманов Д.Ш., 1990 Бухараев Р.Г., Сулейманов Д.Ш. Семантический анализ в вопросно-ответных системах. - Казань: Изд-во Казан. ун-та. - 1990. -124 с.

2. Иванова В.Г., Галеева И.Х., 1996 Труды Международного семинара «Искусственный интеллект в образовании» в двух томах (Казань, 1-4 октября 1996 г.)// Под ред. Иванова В.Г., Галеева И.Х. -Казань. -1996.

3. Искусственный интеллект. — В 3-х кн. Кн.2. Модели и методы: Справочник. /Под ред. Д.А.Поспелова. —М.: Наука, 1990.

4. Компьютерный синтаксический анализ: описание моделей и направлений разработок.— Г. Д. Карпова, Ю. К. Пирогова, Т. Ю. Кобзарева, Е. В. Микаэлян. // Итоги науки и техники (серия “Вычислительные науки”). Т.6. —М.: ВИНТИ, 1991.

5. Нариньяни А.С., Лингвистические процессоры ЗАПСИБ (1-я и 2-я части). Препринт ВЦ СО АН СССР, N 199, 1979

6. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах. —М.: Наука, 1989.