

Диалог на естественном языке в компьютерной системе обучения иностранному языку

Dialogue in a natural language in computer system of training for foreign language

И.Н.Дайман

<mailto:pildery@gmail.com>

В настоящей работе описывается реализация подхода к тестированию знаний обучаемого в форме диалога на ЕЯ с обучающей системой. Показано, как организуется диалог, исходя из структуры урока и знаний обучаемого, как разрешаются неоднозначности и как по итогам диалога производится оценка знаний обучаемого. В процессе организации и управления диалогом выделены и разработаны для частной задачи проблема недоопределенности, связанной с ЕЯ, и проблема согласования систем знаний обучающей системы и обучаемого.

Введение

Организация процесса тестирования и анализа знаний обучаемого является одной из важнейших проблем компьютерных обучающих систем, обеспечивающих, с одной стороны, адекватность системы пользователю, с другой – создающих у пользователя эффект учителя и повышающих эффективность обучения.

В данной работе предлагается пример организации тестирования знаний обучаемого в форме диалога на языке обучения. В качестве пользовательского интерфейса для реализации такого диалога используется имитация электронного чата.

Тест направлен на анализ знания обучаемым материала текущего урока. На основании этого анализа формируется управляющее воздействие в виде множества переходов, каждый из которых указывает обучаемому необходимость повторения того или иного материала, а также предлагает изучение нового.

Постановка задачи

Пусть задана методика обучения, которая предполагает обучение правильному употреблению словоформ в реальных речевых ситуациях.

Пусть знания системы представлены в виде набора X – множество корней, Y – множество аффиксов, F – множество правил, определяющих правильную словоформу.

$\forall \{x_i\} \subseteq X \exists f_l \in F : \exists \{y_k\} \subseteq Y \wedge f_l(\{x_i\} \cup \{y_k\})$ – частично упорядоченное множество.

Иначе говоря, для любого подмножества корней можно найти правило, которое определяет соответствующее ему подмножество аффиксов и задает отношение частичного порядка на множестве-объединении корней и аффиксов.

Такое частично упорядоченное множество будем называть правильной словоформой.

(1) Пример

grossen Buecher

Здесь курсивом показан корень, жирным шрифтом с подчеркиванием выделен аффикс (в данном случае окончание), правило – склонение имен прилагательных и существительных по падежам, окончание в мн. ч, им.п.

Знанием системы назовем совокупность правильных словоформ.

Сложность знания определяем как мощность множества $f_l(\{x_i\} \cup \{y_k\})$.

В рассматриваемом примере сложность знания равна 4. Понятно, что чем больше исходно набор корней и правило, определяющее словоформу, тем более сложным будет знание. Примером более сложного знания могут служить глагольные рамки в конструкциях временных форм, расширенные определения, проч.

Предметная область представлена в виде фиксированного набора уроков, каждый из которых содержит заданный материал для изучения.

Пусть задан некоторый произвольный урок с номером ρ .

Множество знаний обучаемого на заданном уроке ρ представлено недоопределенным множеством $O = \langle O^+, O^-, M_o \rangle$, где O^+ – знания, которые обучаемый усвоил, O^- – знания, которые обучаемый не усвоил, M_o – недоопределенное число (н-число), мощность множества O .

Множество O задано на универсуме всех знаний системы U .

Задача: необходимо, в процессе теста обучаемого на заданном уроке доопределить множество его знаний.

Это значит, что в результате теста обучаемого будет точно известно, какие из знаний он усвоил на данном уроке, а какие не усвоил.

Изменение знаний обучаемого, процесс обучения.

Цель обучения на фрагменте обучения с номером ρ есть достижение получения обучаемым того количества знаний, которое предполагается к изучению на заданном уроке.

В общем случае эта цель как правило не достигается в результате обучения и знания обучаемого представляются в виде недоопределенного множества, что отражает состояние обучаемого в результате изучения материала конкретного урока.

Процесс обучения проходит по следующим этапам:

1. реализация обучающего воздействия;
2. тестирование знаний обучаемого;
3. формирование управляющего воздействия, переход на следующий фрагмент обучения.

Реализация обучающего воздействия определяется методикой обучения и достигается за счет применения различных видов деятельности на текущем этапе обучения, вариативности примеров и способе подачи материала.

По окончании этапа реализации обучающего воздействия модель знаний обучаемого есть недоопределенное множество. На этапе тестирования происходит доопределение модели обучаемого и анализ качества его знаний.

На основании доопределения модели обучаемого и оценки качества его знаний формируется управляющее воздействие, выбирается номер фрагмента обучения, на который система рекомендует перейти обучаемому.

Далее процесс обучения, изменения знаний обучаемого, доопределения и перехода продолжается до тех пор, пока обучаемый не прекратит работу с обучающей системой или пока не пройдет обучение на всех фрагментах.

В данной работе не рассматривается алгоритм формирования управляющего воздействия на основании оценки знаний обучаемого.

По окончании обучения модель обучаемого представлена доопределенным множеством O со следующими ограничениями:

1. набор знаний, которые обучаемый усвоил (O^+), известны и определены;
2. набор знаний, которые обучаемый не усвоил (O^-), известны и определены.

Тест обучаемого в форме диалога с обучающей системой

Тестирование знаний обучаемого преследует достижение следующих целей:

1. доопределение модели обучаемого с целью уточнения информации о знаниях, которые он усвоил;
2. выявление качества знаний обучаемого на каждом шаге тестирования.

Сложность знаний, изучаемых на заданном фрагменте обучения определяет сложность теста.

Тестирование производится в форме диалога на языке обучения, в качестве пользовательского интерфейса используется режим имитации электронного чата.

Лингвистическим материалом для построения вопросов системы в диалоге служит материал данного фрагмента обучения. Лингвистическим материалом для анализа ответа обучаемого служит материал обучающей системы в целом.

Каждый шаг диалога предполагает проверку наличия у пользователя единицы знания из набора, который предполагалось освоить на заданном фрагменте обучения. Шаг диалога состоит из вопроса системы и ответа обучаемого.

Вопрос системы формируется методистом. Ответ пользователя имеет несколько степеней свободы в зависимости от сложности знаний, которые подлежат проверке. Степень свободы определяет произвольность ответов пользователя и их вариативность. Основное ограничение, которое используется при тестировании обучаемого заключается в том, что знания обучаемого могут быть идентифицированы только на основании знаний системы.

Порядок выбора вопросов в диалоге со стороны системы определяется спецификой данного фрагмента обучения. Он может быть как произвольным, так и заданным, исходя из сложности знаний.

Условия проведения теста задают два основных типа ошибок при ответе обучаемого на вопрос системы:

1. ошибки в знаниях, которые предполагаются к изучению;
2. ошибки неправильного написания слов.

Рассмотрим более подробно ошибки первого вида и укажем точную разницу между ошибками первого и второго вида.

Исходя из определения знания, имеем, что ошибки в знании обучаемого могут быть следующими:

1. неправильный выбор правила для построения словоформы;
2. отсутствие необходимого аффикса;
3. отсутствие необходимого корня.

Варианты 2 и 3 означают невозможность идентифицировать в ответе обучаемого соответствующий элемент. Это может происходить в силу либо отсутствия такого элемента, либо его неправильной записи.

Запись элемента, которая не позволяет его распознать (найти среди ожидаемых), будем считать ошибкой в знаниях обучаемого – ошибка первого вида.

В случае, если элемент можно распознать как ожидаемый, будем считать что имеет место ошибка второго вида.

В случае ошибки второго вида ответ обучаемого может быть скорректирован в диалоге функцией корректировки ошибок.

Если корректировка выполнена удачно, ошибка не учитывается, в противном случае она идентифицируется как ошибка в знаниях соответствующего вида.

Для простоты будем считать, что на каждом шаге диалога от обучаемого ожидается использование только одной единицы знаний (правильное употребление словоформы).

Пусть E – множество ошибок, в соответствии с вышесказанным, это множество содержит элементы трех видов:

$$\begin{pmatrix} \text{Ошибка_сочетания} \\ \text{Незнание_корней} \\ \text{Незнание_аффиксов} \end{pmatrix}$$

Алгоритм доопределения множества O_p , диалог системы и обучаемого.

ШАГ1: $i = 1$

ШАГ2: ПОКА($i \leq P$) ДЕЛАТЬ

ШАГ2.1: Выбор_Вопроса_Системы

ШАГ2.2: Выбор_Шаблона_Ответа

$$\left(f_{l_{\text{сис}}}^i \left(\{x_{l_{\text{сис}}}^i\} \cup \{y_{k_{\text{сис}}}^i\} \right) \right)$$

ШАГ2.3: Ожидание_Ввода($f_l^i(\{x_l^i\} \cup \{y_k^i\})$)

$$\text{ШАГ2.4: ЕСЛИ} \left(\begin{array}{l} \text{Оценка_разницы} \\ (f_{l_{\text{сис}}}^i, f_l^i) \leq \Delta \end{array} \right) \text{ТО}$$

$$\text{ШАГ2.4.1: } O_{\rho}^+ = O_{\rho}^+ \cup f_{l_{\text{сис}}}^i$$

ШАГ2.4.2: ВСЕЕСЛИ

$$\text{ШАГ2.5: ЕСЛИ} \left(\begin{array}{l} \text{Оценка_разницы} \\ (f_{l_{\text{сис}}}^i, f_l^i) > \Delta \end{array} \right) \text{ТО}$$

$$\text{ШАГ2.5.1: ЕСЛИ} \left(\begin{array}{l} \text{Ошибка}(f_l^i, f_{l_{\text{сис}}}^i) = \\ \text{ИСТИНА} \end{array} \right) \text{ТО}$$

$$\text{ШАГ2.5.2: } O_{\rho}^- = O_{\rho}^- \cup f_{l_{\text{сис}}}^i$$

$$\text{ШАГ2.5.3: ИНАЧЕ}(O_{\rho}^+ = O_{\rho}^+ \cup f_{l_{\text{сис}}}^i)$$

ШАГ2.5.4: ВСЕЕСЛИ

ШАГ2.5.5: ВСЕЕСЛИ

ШАГ2.6: $i = i + 1$

ШАГ2.7: ПЕРЕХОД_ШАГ2

ШАГ2.8: ВСЕПОКА

ШАГ3: КОНЕЦ

В приведенном алгоритме умеют место следующие функции:

1. *Выбор_Вопроса_Системы* – функция, действие которой для данного фрагмента обучения задается на этапе составления курса;
2. *Выбор_Шаблона_Ответа* – функция, действие которой для данного фрагмента обучения задается на этапе составления курса;
3. *Ожидание_Ввода* – функция, которая принимает информацию от обучаемого, специфика ее действия и реализации связана с выбором вида пользовательского интерфейса;
4. *Оценка_разницы* – функция, которая оценивает качество знания обучаемого, осуществляет распознавание введенной обучаемым информации с ожидавшейся;
5. *Ошибка* – функция, которая выполняет идентификацию ошибки и корректировку ошибки обучаемого. В случае, если корректировка произошла успешно, функция возвращает значение *ЛОЖЬ*, иначе – значение *ИСТИНА*.

Алгоритм идентификации и корректировки ошибок

Корректировка предполагается для одной выделенной словоформы. Идентификация ошибки происходит в соответствии со сказанным выше.

Функция *Ошибка* реализует следующие действия:

1. идентификация вида ошибки;
2. запрос на получение данных для корректировки ошибки;
3. анализ выбора обучаемого;
4. исполнение корректировки;
5. пополнение множества E;
6. возврат результата.

ШАГ1: ERR = Идентификация _ошибки

$(f_{l_{сис}}^i, f_i^i)$

ШАГ2: ЕСЛИ($ERR \in ERR_{cor}$)ТО

ШАГ2.1: SEL = Выбрать _похожие(f_i^i)

$f_{l_{сис}}^i \in SEL$

ШАГ2.2: Ожидание _ввода

ШАГ2.3: ЕСЛИ($f_i^{i'} \equiv f_{l_{сис}}^i$)ТО

ШАГ2.3.1: ВОЗВРАТ (ЛОЖЬ)

ШАГ2.3.2: ИНАЧЕ ($E = E \cup ERR$)

ШАГ2.3.3: ВОЗВРАТ (ИСТИНА)

ШАГ2.3.4: ВСЕЕСЛИ

ШАГ2.4: ИНАЧЕ ($E = E \cup ERR$)

ШАГ2.5: ВОЗВРАТ (ИСТИНА)

ШАГ2.6: ВСЕЕСЛИ

ШАГ3: КОНЕЦ

В приведенном алгоритме имеют место следующие функции:

1. *Выбрать _похожие* – функция библиотеки доступа к лингвистическому материалу системы, понятие «похожий» определяется методикой курса;
2. *Ожидание _ввода* – интерфейсная функция, в данном случае предполагается выбор из списка предложенных вариантов.

Пример теста и корректировки ошибок

Приводим пример шага диалога, который демонстрирует работу приведенных алгоритмов. Все указанные в алгоритмах функции есть функции библиотеки доступа к данным WordFun.

Исходные данные

Lektion 1, Seite 16, Uebung 6, Muster.

Der Tisch, teuer, wie, *mein*, *ist*

V.: Wie meinen Sie, ist der Tisch teuer?

O.: Ja, ich _____ Tisch _____ teuer.

Ожидаемый ответ

Ja, ich *meine*, **der** Tisch **ist** teuer

Полученный ответ

Ja, ich *meine*, **der** Tisch teuer

Результат

Зафиксирована ошибка отсутствия глагольной формы.

Зафиксировано знание глагольного спряжения (1 л., ед.ч., н.в.)

Зафиксировано знание артикля (м.р., ед. ч., им. п.).

Заключение

Изложенный подход к анализу знаний обучаемого используется в пилотном продукте системы «Обучение иностранному языку» на базе существующего источника лингвистической информации и библиотеки доступа к данным WordFun.

Эксплуатация этого продукта показывает эффективность такого метода с точки зрения адекватности оценки знаний обучаемого и обеспечения активного тренинга обучаемого в изучении и правильном употреблении словоформ иностранного языка.

Текущая реализация использует методику изучения немецкого языка русскоговорящими пользователями, методика разработана С.Кочетовой именно с акцентом на изучение и правильное употребление словоформ немецкого языка.

Список литературы

1. Нариньяни А.С. Недоопределенность в системе представления и обработки знаний // Техническая кибернетика. М.: Наука, 1986, №5.
2. Нариньяни А.С. Недоопределенные множества – новый тип данных в системе представления знаний //Новосибирск: ВЦ СО АН СССР, 1980.
3. Попов Э.В. Общение с ЭВМ на естественном языке // М.: Едиториал УРСС, 2004.
4. Кочетова С.О. Alles Klar. Учебник немецкого языка для начинающих // СПб.: Союз, 2005.