

ПРОЕКТИРОВАНИЕ БАЗ ЗНАНИЙ ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМ ПО РУССКОМУ ЯЗЫКУ

К.В. Русецкий, Д.Д. Якупова, А.Ф. Хусаинов, Д.И. Ильюшениа
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Кафедра романской филологии, Казанский федеральный университет
НИИ "Прикладная семиотика" Академии Наук Республики Татарстан
Минск, Республика Беларусь
Казань, Российская федерация
E-mail: rusetski.k@gmail.com

В данной работе рассматривается универсальный подход к формализации лингвистических знаний различного рода на примере предметной области текстов русского языка. Далее рассматривается, как описанный подход и реализованные с его помощью фрагменты баз знаний могут быть использованы в качестве основы для построения интеллектуальных обучающих систем по русскому языку.

ВВЕДЕНИЕ

Электронные средства обучения (ЭСО) языку – программные средства учебного назначения, в которых отражена определенная предметная область, в той или иной степени реализована технология ее изучения, обеспечены условия для реализации различных видов учебной деятельности. Такие системы получили достаточно большое распространение в ВУЗах при дистанционной форме обучения. Однако современные системы такого класса не владеют материалом, которому учат, поскольку не могут отвечать на вопросы учащегося, а лишь выдают учебный материал. Наиболее приемлемым решением данной проблемы видится создание интеллектуальных обучающих систем (ИОС[1]). Для создания таких систем необходимо прежде всего разработать способ представления естественно-языковых конструкций и знаний о языке в памяти интеллектуальных систем. Желательно, чтобы, при формальной строгости, такой способ также позволял наглядно представлять конструкции выбранного языка. Наиболее хорошо данным требованиям отвечает представление в виде семантической сети. Представить любые знания, в том числе и естественно-языковые, позволяет формальный графовый язык, такой как Semantic Code (SC) и его графический эквивалент Semantic Code Graphical (SCg)[2].

Лингвистическая база знаний является информационно-содержательным ядром ИОС, включающим знания о языке. При проектировании базы знаний иностранного языка следует учитывать как лингвистические универсалии, так и частные особенности строения конкретного языка. Вне зависимости от изучаемого языка, в составе базы знаний ИОС выделяются компоненты «Словарь», «Грамматический справочник», «Корпус текстов». С учетом особенностей, например, английского языка можно выделить компонент «Неправильные глаголы».

I. СЛОВАРЬ

Накопление достаточного словарного запаса является неотъемлемой частью изучения любого иностранного языка. Поэтому одним из важнейших компонентов ИОС является словарь. В отличие от современных словарей, словарь в рамках ИОС является семантически структурированным. Это означает, что в нем:

- каждому слову приписывается грамматическая информация (часть речи, ее признаки);
- каждое слово связано с другими словами различными семантическими отношениями (например, быть синонимом, антонимом, гипонимом, и пр.);
- словам ставятся в соответствие примеры их употребления в текстах, которые хранятся и систематизируются в компоненте «Корпус текстов».

II. ГРАММАТИЧЕСКИЙ СПРАВОЧНИК

Особенностью представления грамматического справочника ИОС является то, что информация в нем рассчитана как на использование (изучение) учащимися, так и на реализацию операций автоматического создания заданий тестов. Основу справочника составляет набор грамматических правил соответствующего языка, которые представлены в двух видах: 1) формализованных правил и закономерностей образования словоформ, словосочетаний, предложений, описанных на логическом sc-языке; 2) учебных естественно-языковых текстов объяснений для учащихся указанных правил.

III. НЕПРАВИЛЬНЫЕ ГЛАГОЛЫ

В рамках данного компонента предусматривается хранение и формальное описание временных форм неправильных глаголов. В английском языке, к примеру, существует более 200 широко используемых глаголов [3], не следующих

общим правилам образования форм, что свидетельствует о важности такого компонента базы знаний. Все три формы неправильных глаголов представляются в базе знаний в явном виде.

IV. КОРПУС ТЕКСТОВ

Для создания разнообразных заданий и тестов необходимы источники иноязычных текстов [4]. Практически все тесты создаются на основе оригинальных текстов. Именно поэтому в рассматриваемой ИОС специально выделен компонент «Корпус текстов», который преследует две цели: а) автоматизированное создания заданий тестов; б) обеспечение связей слов из словаря с их словоупотреблениями в текстах.

Поскольку ранее упомянутый язык SC кодирует семантические сети с базовой теоретико-множественной интерпретацией, то для последующей формализации удобнее всего будет рассмотреть корпус текстов с точки зрения теории множеств. Корпус представляет собой неориентированное множество текстов. Текст есть ориентированное множество предложений. Предложение является ориентированным множеством вхождений словоформ. Абзац – подмножество предложений текста, выделяемое на письме красной строкой.

V. ФОРМАЛИЗАЦИЯ РАЗДЕЛИТЕЛЕЙ

Разделители (пробелы, знаки препинания) обычно отделяют некоторые подмножества текста (слова, причастные, деепричастные обороты, простые части сложных предложений и собственно предложения). Кроме того, знак препинания, ставящийся в конце предложения, определяется типом предложения – вопросительное, восклицательное либо повествовательное [5]. Вышесказанное означает, что вместо самих знаков препинания в базе знаний имеет смысл формализовывать те классы фрагментов текста, которые они обособляют. Это позволяет, кроме прочего, абстрагироваться от особенностей постановки знаков препинания в разных языках (ярким примером являются вопросительные и восклицательные предложения в испанском языке).

VI. МАШИНА ОБРАБОТКИ ЗНАНИЙ

Помимо базы знаний, неотъемлемой частью любой ИОС является также машина обработки знаний (МОЗ), обеспечивающая возможность решать различного рода задачи. МОЗ представлена информационно-поисковыми операциями и операциями генерации новых знаний. Информационно-поисковые операции используются для навигации пользователя по семантической сети, поиска ответов на заданные вопросы и пр. Например: поиск определения понятия, поиск примеров для понятия, вывод семантической окрестности узла. Операции генерации новых знаний используются для получения новых

знаний на основе существующих. К таким операциям могут относиться: генерация задания пользователю по определенной тематике, анализ выполненного пользователем задания.

Система операций является агентно-ориентированной и представляет собой набор sc-операций, условием иницирования которых является появление в памяти системы некоторой определенной конструкции. При этом операции взаимодействуют между собой через память системы посредством генерации конструкций, являющихся условиями иницирования для другой операции. При таком подходе становится возможным обеспечить гибкость и расширяемость решателя путем добавления или удаления из его состава некоторого набора операций.

Отличительной особенностью решателя задач как многоагентной системы в рамках данного подхода является принцип взаимодействия операций-агентов. Агенты обмениваются сообщениями исключительно через общую память путем использования соответствующего языка взаимодействия (языка вопросов-ответов), в отличие от большинства классических многоагентных систем, в которых агенты обмениваются сообщениями непосредственно друг с другом. В рассматриваемом подходе каждый агент, формулируя вопросную конструкцию в памяти, априори не знает, какой из агентов будет обрабатывать указанную конструкцию, а лишь дожидается появления в памяти факта окончания обработки вопроса. В случае наличия такого соответствия, агент начнет обработку указанного вопроса (решение поставленной задачи), и в результате работы сгенерирует некоторый ответ на поставленный вопрос.

Более подробно процесс проектирования операций и предъявляемые к ним требования рассмотрены в соответствующих работах [6].

1. Интеллектуальные обучающие системы и виртуальные учебные организации: Монография / В.В.Голенков, В.Б. Тарасов, О.Е. Елисеева и др.; Под ред. В.В. Голенкова, В.Б. Тарасова — Мн.: БГУИР, 2001. — 488 с.
2. Голенков, В.В. Представление и обработка знаний в графодинамических ассоциативных машинах / В. В. Голенков [и др.] — Мн. : БГУИР, 2001.
3. English Irregular Verbs [Электронный ресурс]. — 2013. — Режим доступа: <http://www.usingenglish.com/reference/irregular-verbs/>. — Дата доступа: 4 ноября 2013 г.
4. Централизованное тестирование. Английский язык : сборник тестов / Респ. ин-т контроля знания М-ва образования Респ. Беларусь. — Минск: Аверсэв, 2012.
5. Розенталь Д. Э. Современный русский язык / Д. Э. Розенталь, И. Б. Голуб, М. А. Теленкова. — 11-е изд. — М.: Айрис-пресс, 2009.
6. Шункевич, Д. В. Семантические модели и средства проектирования машин обработки знаний / Д. В. Шункевич // Материалы международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» — Минск, 2013.