

УДК 004.832.22

К.А. МЕТЕШКИН, Л.А. ФЕДОРЧЕНКО, Н.М. КОБЗИСТАЯ

Международный Славянский университет, Харьков, Украина

УЧЕБНАЯ ПРОГРАММА КАК ОБЪЕКТ ИССЛЕДОВАНИЯ КОРПУСНОЙ ЛИНГВИСТИКИ

Предлагается использовать методы корпусной лингвистики при создании лингвистического обеспечения обучающихся систем с интегрированным интеллектом.

искусственный интеллект, обучающие системы, лингвистическое обеспечение, корпусная лингвистика

Совершенствование информационных технологий и создание на их базе интеллектуальных систем различного назначения обусловило развитие прикладной лингвистики, методическая база которой направлена на создание лингвистического обеспечения этих систем.

Целью настоящей статьи является оценка возможности использования методов корпусной лингвистики при создании лингвистического обеспечения обучающихся систем с интегрированным интеллектом, в частности лингвистических баз знаний учебного назначения.

Одним из направлений прикладной лингвистики является корпусная лингвистика (КЛ), которая создана для исследования языкового материала и ориентирована на компьютерную обработку больших массивов текстовой информации. Одним из основополагающих понятий КЛ является понятие корпуса текстов (КТ). В работе [1] КТ определен как вид корпуса данных, единицами которого являются тексты или их достаточно значительные фрагменты, включающие, например, какие-то полные фрагменты макроструктуры текстов данной предметной области.

Из определения видно, что понятие «корпус текстов» достаточно размыто (нечетко), а его объем (фрагменты) задаются только лишь на макроуровне и не указываются его границы. Поэтому для целенаправленного использования методов КЛ и использо-

вания их при разработке компонент лингвистического обеспечения обучающихся систем с интегрированным интеллектом, сведения о которых приведены в работах [2 – 4], сформулируем определение, которое учитывает структуру текстовой информации.

Определение. Под корпусом текста будем понимать законченный смысловой массив текстовой информации, на структурных элементах которого могут быть заданы различного вида отношения.

Такая формулировка позволяет поставить в соответствие с понятием «корпус текста» любое печатное произведение или его составную часть. Например, КТ можно считать монографию или учебник, на структурных элементах которых заданы отношения строгого порядка, включения, принадлежности и т.д., как это показано в работе [6]. Другой пример – научная статья, элементами которой можно считать совокупность абзацев текста, определяющих начало и конец структурных ее частей (аналитической, смысловой и обобщающей), на которых задаются причинно-следственные отношения, отношения строгого порядка и т.д.

Исходя из прагматических целей использования в обучающих системах с интегрированным интеллектом текстовой информации, которая содержит в себе процедурные и декларативные знания их авторов, а также опираясь на сформулированное определение, предложим следующую классификацию КТ.

Будем различать КТ текстов по их целевому значению.

КТ учебного материала. К ним отнесем учебники, учебные пособия, конспекты лекций, другие учебные материалы и их структурные части, обеспечивающие обучающихся сведениями о явлениях, процессах, фактах некоторой предметной области.

КТ текстов методического материала. Здесь полагаем, что тексту и его структуре характерны признаки управления когнитивной деятельностью обучающихся. К ним отнесем методические разработки, учебные программы и другие методические материалы, которые также имеют определенную внутреннюю структуру.

КТ служебных документов. К ним отнесем тексты приказов, директив, инструкций и других служебных документов, которые регламентируют установление строгого соответствия между объектами (субъектами) образовательных процессов. Их составные части также могут быть связаны между собой определенными отношениями.

КТ словарно-справочных средств. К ним отнесем словари, справочники, энциклопедии и другие словарно-справочные средства, которые имеют специфическую внутреннюю структуру. Например, в толковом словаре между разъясняемыми словами и понятийной составляющей словаря определены отношения «Термин – понятия», а между терминами, составляющими все множество терминов словаря, устанавливается отношение строгого порядка, так как термины в словаре располагаются в алфавитном порядке.

Для продуктивного использования текстовой информации в обучающих системах с интегрированным интеллектом, конкретно при построении учебных баз знаний, необходимо задать минимальный и максимальный объем текстовых корпусов. Исходя из практики компьютерной верстки хотя бы рукописей учебного и методического материалов и работы в текстовых редакторах, будем считать, что

максимальный объем содержит Глобальный корпус, составляющий все лексическое произведение. Единичным КТ будем считать *предложение* (простое или сложное), которое определено в работе [6], как минимальная единица человеческой речи, представляющая собой грамматически организованное соединение слов, обладающее определенной смысловой и интонационной законченностью.

Учитывая вышесказанное, а также результаты исследований, приведенных в работе [5], где предлагаются методы формализации структур учебного и методического материала, покажем на примере представление учебной программы в виде КТ. Для этого на рис. 1 представим топологию КТ учебной программы, т.е. ее структурные элементы. Предварительно дадим определение учебной программе.

Определение. Учебная программа – это основной методический документ, определяющий цели, задачи и планируемые результаты обучения по конкретной учебной дисциплине во взаимосвязи с другими учебными дисциплинами и обеспечивающий планирование учебного процесса в вузе.

Определим отношения, которые задаются, исходя из смыслового содержания элементов КТ «Рабочая учебная программа». Для этого на теоретическом языке запишем основные соотношения, отражающие структуру рассматриваемого КТ.

Обозначим $Z = \{W, Q, S, E\}$ – КТ «Рабочая учебная программа», содержащий КТ: W – «Целевые установки»; Q – «Организационно-методические указания»; S – «Содержание»; E – «Информационно-методическое обеспечение».

Проведем анализ отношений между этими КТ. Из рис. 1 видно, что они связаны отношением строгого включения. Поэтому справедливы следующие соотношения:

$$\{W, Q, S, E\} \subset Z;$$

$$(w_c, w_y, w_b) \subset W; (q_j) \subset Q, j = \overline{1, n};$$

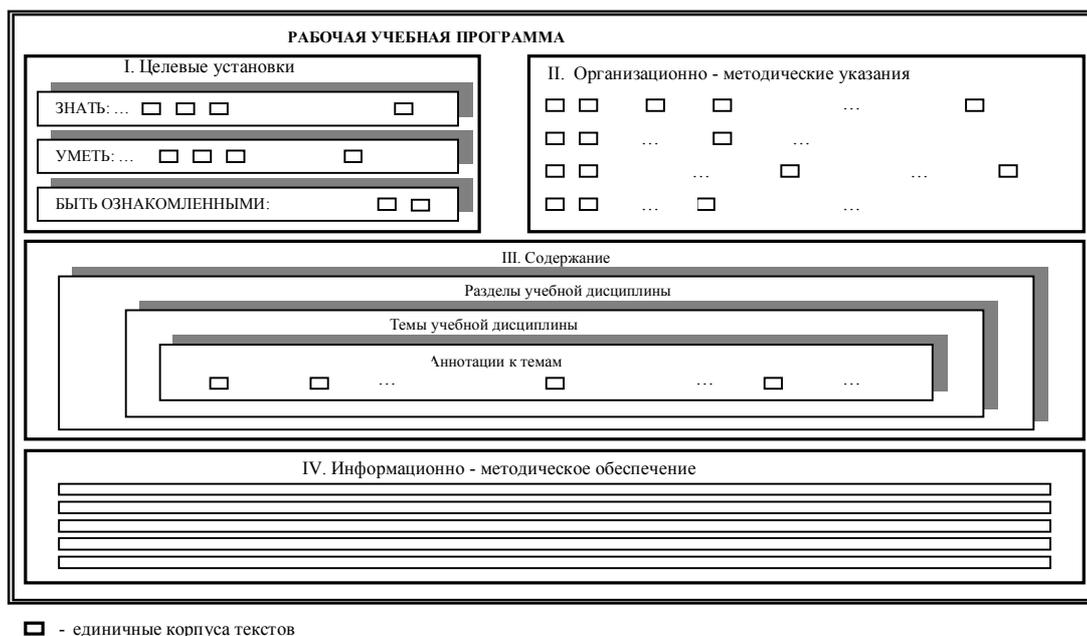


Рис. 1. Корпус текста «Рабочая учебная программа»

$S = \{ R, T, A \}; \left((A) \subset T \right) \subset R \subset S;$
 $(a_m) \subset A, m = \overline{1, k}; (e_v) \subset E, v = \overline{1, u},$
 где $w_c = \{ w_c^1, \dots, w_c^a \}, w_y = \{ w_y^1, \dots, w_y^b \},$
 $w_b = \{ w_b^1, \dots, w_b^z \}$ – КТ – «Знать», «Уметь»,
 «Быть ознакомленными» с соответствующими еди-
 ничными корпусами; q_j – единичные КТ, состав-
 ляющие корпус текста Q ; R – КТ «Раздел учебной
 дисциплины»; T – КТ «Тема учебной дисциплины»;
 a_m – единичные КТ, составляющих аннотации к m -й
 учебной теме; E – КТ «Информационно-методи-
 ческие материалы», которые состоят из e_v элемен-
 тарных КТ.

Кроме отношения строгого включения между выделенными КТ существует иерархия учебных целей. Видно, что целевые установки учебной дисциплины задаются смысловым содержанием корпуса текстов W , учебными целями разделов и тем. Начальные условия достижения этих целей заданы смысловым содержанием корпуса текстов Q .

Тогда, отождествляя целевые установки с корпусами текстов, получим граф целевых установок, изображенный на рис. 2, где W^g – глобальная цель учебной дисциплины; W^{p_n} – множество целей учеб-

ных разделов; $W_n^{t_1}$ – множество целей учебных тем; q_j – начальные условия достижения учебных целей.

Кроме иерархических (древовидных) отношений целевых установок, выделим отношение «Обеспечивать» между корпусами текстов R, T и E , которые можно записать как декартовы произведения:

$$Y \subset R \times E; U \subset T \times E.$$

Это обозначает, что каждому корпусу текстов R или T соответствует корпус текстов E . Эти отношения можно детализировать и поставить в соответствие каждому корпусу текстов T элементарный или несколько элементарных КТ из E .

Из краткого анализа отношений между КТ учебной программы видно, что она имеет сложную вложенную структуру с множеством разнотипных отношений между ее элементами, которая затрудняет автоматическую обработку такого рода текстовых документов. Однако сама идея декомпозиции сложных текстовых документов на КТ позволяет использовать эту возможность для представления текстовой информации различного объема в лингвистических базах знаний, в основу которых положены эвристические методы представления знаний, в частности, фреймовыми системами.

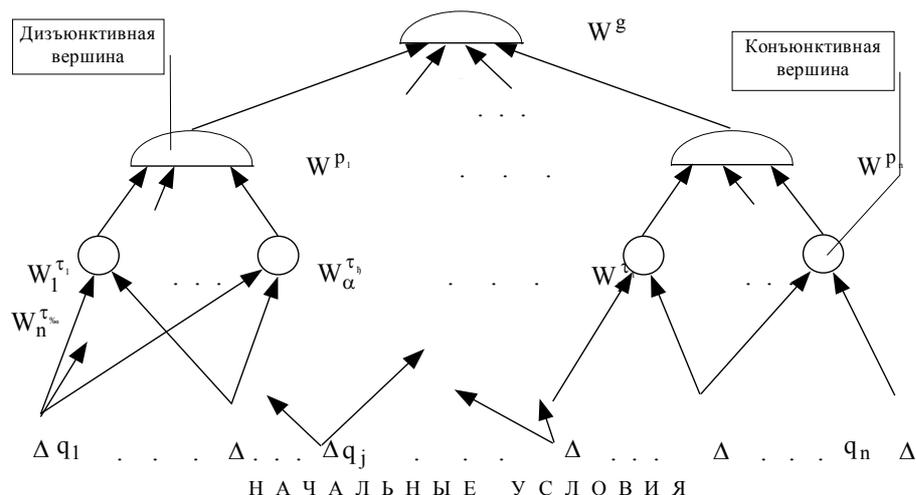


Рис. 2. Граф целевых установок, заданный с учетом топологии корпуса текстов «Учебная программа»

Топология КТ «Рабочая учебная программа» (рис. 1) показывает, что такая структура текстовых данных вполне приемлема для размещения их в соответствующих слотах фреймовой системы, а отношения между ними можно реализовать при помощи специальных присоединенных процедур или «демонов», которые также предназначены для организации отношений между данными, помещенными в соответствующие слоты. Подробно с построением фреймовых систем можно ознакомиться в работе [7].

Таким образом, применение методов КЛ и их использование при построении обучающих систем с интегрированным интеллектом позволяет осуществлять манипулирование различными по объему КТ, что открывает большие возможности для автоматического анализа текстовой информации, определения учебных программ, дублирующих учебный материал в рамках одной специальности. Кроме того, на этой основе можно строить модели профессиональных знаний преподавателей.

Литература

1. Баранов А.Н. Введение в прикладную лингвистику: учебное пособие. – М.: Эдиториал УРСС, 2001. – 360 с.

2. Метешкин К.А. Элементы технологии обучения на основе гибридного интеллекта // Новый коллегіум. – 2002. – № 3. – С. 33 – 37.

3. Метешкин К.А., Шаронова Н.В. Использование гибридного интеллекта в учебном процессе высших учебных заведений // Alma mater (Вестник высшей школы). – 2001. – № 11. – С. 10 – 15.

4. Метешкин К.А. Шаронова Н.В. Лингвистическое обеспечение обучающих систем с интегрированным интеллектом // Новый коллегіум. – 2002. – № 4/5. – С. 64 – 68.

5. Метешкин К.А. Теоретические основы построения интеллектуальных систем управления учебным процессом в вузе: Монография. – Х.: Экограф, 2000. – 278 с.

6. Розенталь Д.Э., Теленкова М.А. Словарь-справочник лингвистических терминов. – М.: Просвещение, 1985. – 399 с.

7. Осуга С. Обработка знаний. – М.: Мир, 1989. – 293 с.

Поступила в редакцию 02.06.2004

Рецензент: д-р техн. наук, проф. О.Е. Федорович, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.