

УЧЕБНАЯ ПРОГРАММА КАК ОБЪЕКТ ИССЛЕДОВАНИЯ КОРПУСНОЙ ЛИНГВИСТИКИ

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 оценить возможность использования методов корпусной лингвистики при создании лингвистического обеспечения обучающих систем с интегрированным интеллектом, в частности лингвистических баз знаний учебного назначения.

Одним из направлений прикладной лингвистики является корпусная лингвистика, которая создана для исследования языкового материала и ориентирована на компьютерную обработку больших массивов текстовой информации. Одним из основополагающих понятий корпусной лингвистики является понятие корпуса текстов. В работе [1] корпус текстов 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 (не четко), а его объем (фрагменты) задаются только лишь на макро уровне и не указываются его границы. Поэтому для целенаправленного использования методов корпусной лингвистики и использования их при разработке компонент лингвистического обеспечения обучающих систем с интегрированным интеллектом, сведения о которых приведены в работах [2- 5], 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 понятию «корпус текста» любое печатное произведение или его составную часть. Например, корпусом текста 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, строгого порядка и т.д.

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

Будем различать корпуса текстов по их целевому назначению.

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

Корпуса текстов методического материала. Здесь будем полагать, 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 внутреннюю структуру.

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

Корпуса текстов словарно-справочных средств. К ним отнесем, словари, справочники, энциклопедии и другие словарно-справочные средства, которые имеют специфическую внутреннюю структуру. Например, в толковом словаре между словником и понятийной составляющей словаря определены отношения

«Термин – понятия», а между терминами 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 хотя бы минимальный и максимальный объем текстовых корпусов.

Исходя из практики компьютерной верстки рукописей учебного и методического материала и работы в текстовых редакторах, будем считать, что максимальный объем содержит Глобальный корпус составляющий все лексическое произведение. Единичным корпусом текста будем считать *предложение* (простое или сложное), которое определено в работе [7], как минимальная единица человеческой речи, представляющая собой грамматически организованное соединение слов, обладающее известной смысловой и интонационной законченностью.

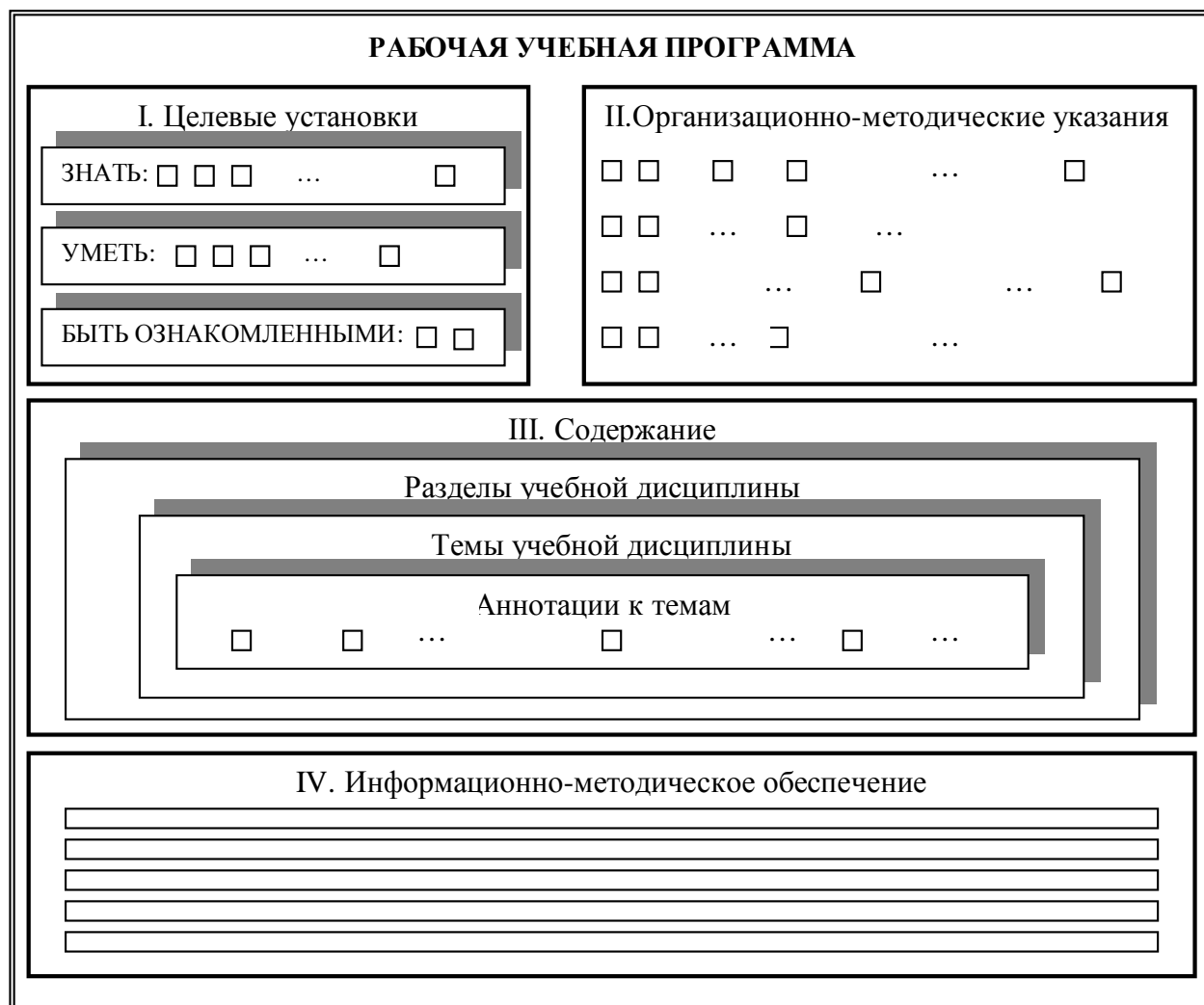
Учитывая вышесказанное, а также результаты исследований приведенных в работе [6], где предлагаются методы формализации структур учебного и методического материала, покажем на примере представление учебной программы в виде корпуса текста. Для этого на рис. 1 представим топологию корпусов текста учебной 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 корпуса текста «Рабочая учебная программа». Для этого на теоретико-множественном языке запишем основные соотношения, отражающие структуру рассматриваемого корпуса текста.

Обозначим $Z = \{W, Q, S, E\}$ - корпус текста «Рабочая учебная программа», содержащий корпус текстов W - «Целевых установок»; корпус текстов Q - «Орга-

низационно-методических указаний»; корпус текстов S – «Содержание»; корпус текстов E – «Информационно-методического обеспечения».



- единичные корпуса текстов

Рис. 1. Корпус текста «Рабочая учебная программа»

Проведем анализ отношений между этими корпусами текстов. Из рис.1 видно, что они связаны отношением строгого включения. Поэтому справедливы следующие соотношения $\{W, Q, S, E\} \subset Z$; $(w_c, w_y, w_b) \subset W$; $(q_j) \subset Q, j = \overline{1, n}$;

$S = \{R, T, A\}$; $((A) \subset T) \subset R) \subset S$; $(a_m) \subset A, m = \overline{1, k}$; $(e_v) \subset E, v = \overline{1, u}$, где

$w_c = \{w_c^1, \dots, w_c^\alpha\}$, $w_y = \{w_y^1, \dots, w_y^\beta\}$, $w_b = \{w_b^1, \dots, w_b^\chi\}$ - корпуса текстов

«Знать», «Уметь», «Быть ознакомленными» с соответствующими единичными корпусами; q_j - единичные корпуса текстов составляющие корпус текстов Q; R –

корпуса текстов «Раздел учебной дисциплины»; T – корпуса текстов «Тема учеб-

ной дисциплины»; a_m - единичные корпуса текстов, составляющих аннотации к m -й учебной теме; E – корпус текстов «Информационно-методические материалы», которые состоят из e_v элементарных корпусов текста.

Кроме отношения строгого включения между выделенными корпусами текстов существует иерархия учебных целей. Видно, что целевые установки учебной дисциплины задаются смысловым содержанием корпуса текстов W , учебные цели разделов и тем, а также начальные условия достижения этих целей смысловым содержанием корпуса текстов Q .

Тогда отождествляя целевые установки с корпусами текстов, получим граф целевых установок, изображенный на рис.2.

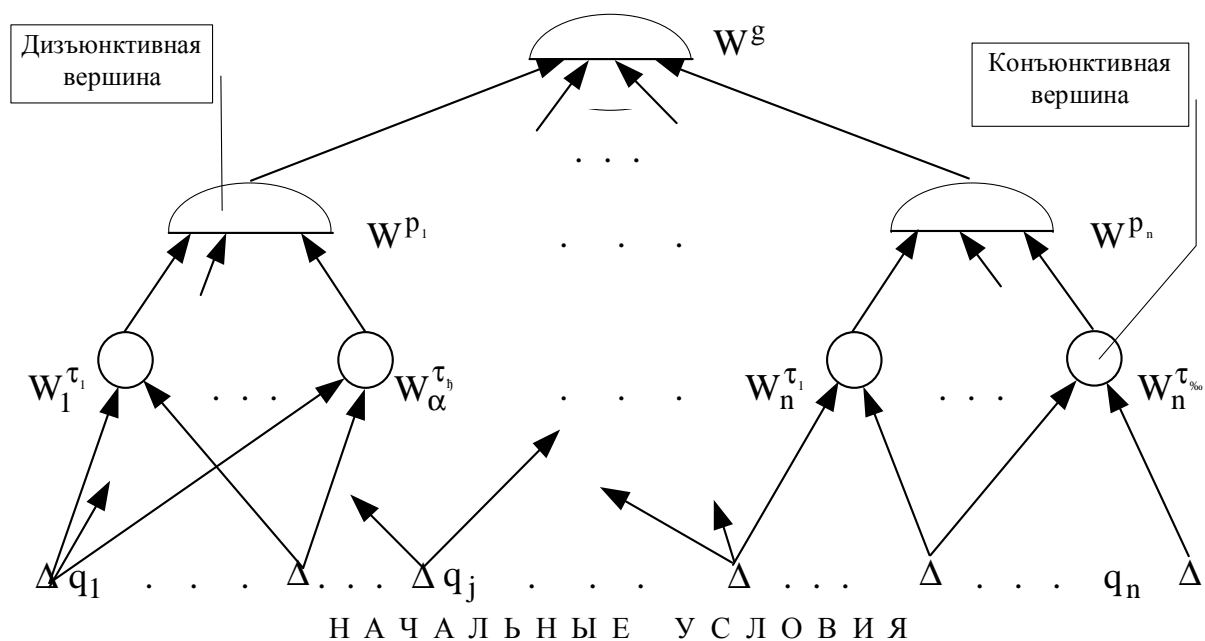


Рис. 2. Граф целевых установок, заданный с учетом топологии корпуса текстов «Учебная программа»

Здесь W^g - глобальная цель учебной дисциплины, W^{P_n} - множество целей учебных разделов, $W_n^{\tau_1}$ - 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 в виде декартовых произведений $Y \subset R \times E$, $U \subset T \times E$.

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

Из краткого анализа отношений между корпусами текстов 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 корпуса текстов 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.

Топология корпуса текстов «Рабочая учебная программа» (см. рис.1) показывает, что такая структура текстовых данных вполне приемлема для их размещения 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 о построении фреймовых систем можно ознакомиться в работе [8].

Таким образом, применение методов корпусной лингвистики и их использование при построении обучающих систем с интегрированным интеллектом позволяет осуществлять манипулирование различными по объему корпусами текстов, что открывает большие возможности для автоматического анализа текстовой информации, определения учебных программ, которые дублируют 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. Задача создания обучающих систем с гибридным интеллектом // Системи обробки інформації. Збірник наукових праць. Вип. 4(14).- Харків: НАНУ, ПАНМ, ХВУ, 2001.- С.13-18.
3. Метешкин К.А. Элементы технологии обучения на основе гибридного интеллекта // Новый коллегіум, 2002, №3, С. 33-37.
4. 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.
5. Метешкин К.А. Шаронова Н.В. Лингвистическое обеспечение обучающих систем с интегрированным интеллектом // Новый коллегіум, Харків, 2002, №4/5, С. 64 – 68.
6. Метешкин К.А. Теоретические основы построения интеллектуальных систем управления учебным процессом в вузе: Монография. - Харьков: Экограф, 2000. - 278 с.
7. Розенталь Д.Э., Теленкова М.А. Словарь-справочник лингвистических терминов: - М.: Просвещение, 1985. - 399 с.
8. Осуга С. Обработка знаний. - М.: Мир, 1989. - 293 с.