

В.В. КАРАСЮК, канд. техн. наук, НЮАУ (г. Харьков)

ОПИСАНИЕ КОНЦЕПЦИЙ ПРЕДМЕТНОЙ ОБЛАСТИ НА БАЗЕ ОНТОЛОГИЙ ДЛЯ СИСТЕМ ОБУЧЕНИЯ

У статті розглянута модель схеми знань для систем електронного навчання. Для опису концептуальної моделі пропонується використати ациклічний орієнтований граф. Також пропонується процедура автоматичного поповнення бази знань за рахунок вибору відповідної інформації із Internet. Ця процедура включає апарат генотипів, що підвищує швидкість та зменшує складність алгоритмів формування онтологій. Розглянуті зв'язки між концепціями у моделі онтологій проблемної області. Вказані обмеження, які поставлені у поточній реалізації даної процедури та перспективні напрямки досліджень.

In article the model of organization of knowledge base for systems of electronic training is considered. For the description of conceptual model it is offered to use acyclic oriented graph. Procedure of automatic updating of the knowledge base also is offered at the expense of a choice of the appropriate information from Internet. This procedure is described by means of genotypes, that raises speed and reduces complexity of algorithms of formation ontologies. The communications between the concepts in model of ontologies of a subject domain are considered. The restrictions are specified which are put in the current realization of this procedure and the perspective directions of researches are specified.

Введение. Электронное обучение (ЭО) в настоящее время является реальной альтернативой традиционному обучению. ЭО можно уверенно рассматривать как удобный инструмент, который снижает затраты на образование и повышает его оперативность. В рамках ЭО можно эффективно организовать самостоятельную работу студентов по изучению учебных материалов и контроль знаний. Однако построение систем ЭО наталкивается на необходимость «ручной» подготовки электронных образовательных материалов как обучающих, так и предназначенных для контроля. Если решить, хотя бы частично, проблему автоматизации подготовки учебных материалов для ЭО, то это позволило бы резко повысить эффективность и массовость использования специализированных обучающих систем.

Анализ исследований по проблемам электронного образования и подготовки обучающих материалов. Необходимость использования понятийного тезауруса как средства интеллектуализации процессов обработки информации была выявлена при разработке первых информационно-поисковых систем [1]. По данному научному направлению имеется большое количество публикаций, что свидетельствует о становлении теоретических и методологических основ данного направления и выработке единой терминологии. Значительный вклад в это направление сделан учеными НТУ «ХПИ», ХНУРЭ, Киевского НУ им. Т.Г. Шевченко, Московского государственного университета экономики, статистики и информатики, Современного гуманитарного университета, и другими. Однако заметно, что

отсутствует единый подход к построению и реализации систем ЭО. Для логического анализа формальных понятий предметной области систем ЭО применяются: аппарат соответствия Галуа, теория решеток, операции над графами, генетические алгоритмы [2], нейронные сети, онтологии и другие математические модели. В то же время, как отмечено в [3], построение логически четкой иерархии научных терминов – практически невозможно.

Проблемы автоматизации подготовки учебного контента для ЭО рассмотрена в работах Арзамасцева А.А. [4], Парфенова И.И. [5], Шуневича Б.И. [6], Кухаренко В.Н. [7], Роберта И.В. [8], Каука В.И., Образцова П.И., Кривули Г.Ф., Преждо Л.Н., Якубайтиса Э.Я. и других. Перспективными, с учетом возможностей автоматизации поддержки, являются системы на основе семантических сетей. В то же время мало внимания уделено в указанных работах проблеме содержания предметной базы знаний таких систем в актуальном состоянии, особенно в предметных областях с формализованным и динамически изменяющимся содержанием (например, в области информатики). Актуальность данной проблематики настолько высока, что в рамках консорциума W3C разработан язык OWL, который призван упростить процесс поиска знаний в документах Интернета, возложив необходимость знания предметной области и описание контекста поиска полностью на авторов документа и систему поиска. С другой стороны, содержательное наполнение учебных материалов для ЭО достаточно формализовано, что снижает уровень требований к процедурам их обработки.

Цель статьи. Рассмотреть принципы построения средств автоматического отбора учебного материала по заданной тематике. Предложить модель представления данных в базе знаний обучающей системы, которая будет поддерживаться соответствующей технологией актуализации данных в ней. Разработать на основе предложенной модели процедуру автоматического пополнения базы знаний обучающей системы.

Модель представления схемы знаний для электронного обучения. Современные модели представления и интеграции информационных ресурсов активно развиваются и внедряются. Онтологии позволяют производить автоматизированную обработку семантики информации, предоставленной в Интернет, с целью её эффективного использования (представления, преобразования, поиска). Соответствующий принцип обработки данных базируется на представлении Интернета как глобальной базы знаний и ориентирован не на осмысление информации, а на обеспечение семантической интероперабельности информационных ресурсов, т.е. автоматизированную интерпретацию и обработку информации. Поскольку онтология определяет термины конкретной предметной области знаний, она должна описываться формальным языком, основанным на принципах математической логики.

Онтология представляется в виде гетерархии (ослабленной или запутанной иерархии), в которой вершины представляют концепции

предметной области. Свойства каждой концепции описываются с помощью механизма «атрибут – значение». Формально такое описание представляется в виде ациклического ориентированного графа. Связи между концепциями описываются с помощью дополнительных логических утверждений. В общем случае вершины могут иметь более одного предшественника.

Язык XML предназначен для разметки документов произвольной структуры. Корректные XML документы содержат дерево вложенных открывающих и закрывающих тегов, каждый из которых может содержать несколько пар «атрибут – значение». Таким образом, если выполнить предварительную разметку документа, то с помощью XML можно создать онтологию отдельного документа и целой предметной области. Консорциум W3C активно работает над внедрением XML в структуру SemanticWeb – Web следующего поколения. Но документы в сети Интернет, как правило, в настоящее время такой разметки не имеют. Поэтому использование XML для этих целей отодвигается на дальнюю перспективу.

Можно сделать вывод, что схему знаний учебных материалов для электронного обучения целесообразно представить как онтологию в виде ациклического ориентированного графа.

Сущность процедуры автоматического формирования базы знаний.

Распространенные в Интернет мультиагентные механизмы поиска тем не менее оставляют пользователям «словарную проблему» (vocabulary problem). Она состоит в том, что, с одной стороны, слова многозначны, а с другой – одни и те же концепции могут быть выражены по-разному. Очень трудно построить онтологию, которая соответствовала бы тематическому разнообразию информации в источниках базы знаний. Пример частного решения этой проблемы – реализация в 90-х годах прошлого века в машине поиска AltaVista сервиса AltaVista Refine, который позволял устранять неоднозначность терминов запроса с помощью словаря совместной встречаемости слов [9].

В онтологии используются ограниченные виды связей между концепциями. Например, в [10] предлагается использовать 11 видов связей: ввести, передать функции обладания (дать, взять, ...), неявное действие, передать информацию, изменить местоположение, создавать мысли, двигать, прикладывать силу, вытащить наружу, физически захватить, произвести звук, направить орган чувств к определенному стимулу. Для реальной предметной области, с которой в настоящее время проводится моделирование, достаточно использовать 7. «Концентраторами» сети онтологии на первом этапе являются концепции *система*, *файл*, *папка*, *ярлык*, и другие базовые понятия. Наполнение базы знаний системы ЭО начинается с терминов, которые впоследствии могут стать «концентраторами». Далее она будет дополняться смысловыми типами, по которым поисковые агенты в источниках знаний будут собирать информацию. Далее включаются механизмы, характерные для генетических алгоритмов. Работа генетического алгоритма начинается с

кодирования базы, то есть с формирования хромосом. Каждый ген в хромосоме характеризуется своим значением (концепцией). Закодированная хромосома является генотипом. Если рассмотренным образом представить содержательную структуру собранного материала, то формальными методами можно определить тождественность хромосом существующей базы знаний обучающей системы и исследуемого источника. То есть возможно идентифицировать структуру предложений потенциального учебного текста. Далее необходимо определить его качество и включить его в виде новых неколайдерных генов (концепций) в существующую структуру онтологии. Таким образом идет расширение онтологии базы знаний в заданной проблемной области.

В настоящее время программно реализуется вариант системы, в котором моделируется онтология (база знаний), включающая понятия из проблемной области курса информатики. В структуре базы знаний поддерживаются отношения типа «IS-A» [11].

Выводы. В процедуру автоматического формирования онтологии системы ЭО можно ввести аппарат генотипов, что повышает скорость работы и сокращает сложность алгоритмов перебора. В дальнейшем планируется изучить возможность повышения качества формирования онтологии предлагаемого вида и разработать процедуру получения учебного текста из имеющейся онтологии.

Список литературы: 1. Шемакин Ю.И. Тезаурус в автоматизированных системах управления и обработки информации. – М.: Воениздат, 1974. – 192 с. 2. Карасюк В.В. Модель системы управления знаниями в обучении // Вісник Національного технічного університету "Харківський політехнічний інститут". Збірник наукових праць. Тематичний випуск: Системний аналіз, управління та інформаційні технології. – Харків: НТУ "ХПИ". – 2005. – № 18. – С. 97 – 100. 3. Налимов В.В. Вероятностная модель языка. О соотношении естественных и искусственных языков. – М.: Наука, 1979. – 304 с. 4. Китаевская Т.Ю., Арзамасцев А.А., Слетков Д.В. Автоматизированная система проектирования содержания обучения информатики в вузе // Информатика и образование. – 2004. – № 12. – С. 100 – 105. 5. Парфенова М.Я., Парфенов И.И. Методологический базис интеллектуальности открытых информационных систем дистанционного обучения // Открытое образование. – 2002. – № 6. – С. 32 – 42. 6. Шуневич Б.И. Дистанционное обучение на Украине: учебные программы, курсы, программное обеспечение // Открытое образование. – 2003. – № 2. – С. 41 – 48. 7. Виртуальная учебная среда "Web-класс ХПИ" / Под ред. Кухаренко В.Н. – Харьков: НТУ "ХПИ", 2003. – 59 с. 8. Роберт И.В. О понятийном аппарате информатизации образования // Информатика и образование. – 2002. – № 12. – С. 2–6; 2004. – № 5. – С. 22 – 29; 2004. – № 6. – С. 63 – 70. 9. Альшанский Г.А., Браславский П.И., Титов П.В. Формирование информационных запросов к машинам поиска интернета на основе тезауруса: семантико-ориентированный подход // Труды VIII Международной конференции по электронным публикациям "EL-Pub2003". 8 – 10 октября 2003 года, Новосибирск, Академгородок, <http://www.ict.nsc.ru/ws/elpub2003/5964/>. 10. Шенк Р. Обработка концептуальной информации. – М.: Энергия, 1980. – 360 с. 11. Цаленко М.Ш. Моделирование семантики в базах данных. – М.: Гл. ред. физ.-мат. лит., 1989. – 288 с.

Поступила в редакцию 07.04.2006