

МЕТОД ОРГАНИЗАЦИИ ЗНАНИЙ С ПОМОЩЬЮ WEB-ОНТОЛОГИЙ В ОБРАЗОВАТЕЛЬНОМ ПРОЦЕССЕ

Бойко П.А.,

научный руководитель канд. техн. наук Даничев А.А..

***Сибирский Федеральный Университет Институт космических и
информационных технологий***

Введение

Передача, хранение и преумножение знаний от поколений к поколениям во все времена определяли развитие человеческой цивилизации. В последние десятилетия знания, интеллектуальные ресурсы приобрели особую значимость в социальном и экономическом развитии общества.

Управление знаниями включает ряд различных моделей представления знаний, такие как продукционные модели, семантические, фреймовые, логические и онтологические модели. Все они по-разному работают, используют разные алгоритмы и способы обработки знаний. Семантические модели, например, используют семантические сети в качестве метода данной модели. Семантическая сеть — это информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними. Объектами могут быть понятия, события, свойства, процессы. С помощью онтологических моделей можно отобразить иерархию объектов, а так же легко осуществляется перенос онтологической модели в глобальную сеть, что называется web-онтологией.

Задача организации знаний в образовательном процессе

Рассмотрим проблему качества высшего образования с точки зрения освоения студентами понятийного аппарата дисциплин учебного направления естественно-научного цикла. Сейчас в современном образовании происходят изменения в способе формирования учебного процесса, тогда как основная мировая проблема состоит в информационном перенасыщении. Большая часть мировой информации содержится в Интернете в статьях, книгах, монографиях. И с каждым днем ее становится все больше, разнообразнее, многограннее. В основном запасы информации пополняются на просторах интернет ресурсов, но и все заинтересованные (студенты, например) чаще пользуются глобальной сетью, нежели бумажными носителями, книгами или другими печатными изданиями. Это можно объяснить тем, что необходимую информацию в Интернете гораздо удобнее искать. Это наиболее доступный способ в современном инновационном мире, и практически всю литературу можно найти в глобальной сети, а также, любого рода информацию можно получить сразу, как только в этом появилась необходимость. Поиск в Интернете привлекателен тем, что все данные там систематизированы и подобраны определенным образом. Это иногда создает неудобства, поскольку информация может быть недостоверна, некачественна или отсутствовать по данному запросу. Знаний, а именно научных знаний, очень много. Множество понятий во многих областях науки пересекаются или повторяются, но при этом, каждое из понятий индивидуально (то есть, по сути, такие термины будут являться словами-омонимами – одинаковые по звучанию и написанию, но абсолютно разные по значению) и для своей области науки имеет точное определение, значение. Такие понятия необходимо различать. И неверно изложенная информация в такой ситуации может привести к ошибочному пониманию той или иной дисциплины, темы, научной области. Для упрощения усвоения всю имеющуюся информацию необходимо

структурировать так, чтобы она так же оставалась доступной, быстрой и в то же время достоверной, хотя бы в пределах одного образовательного ресурса.

Решение задачи структурирования информации

Сформулированную проблему предлагается решить с помощью разработанной системы и алгоритма, которые подразумевают, что у студентов не будет необходимости использовать сразу несколько информационных ресурсов для самоподготовки. Кроме того, результаты запроса подстраиваются под характеристики студента, то есть его курса обучения и перечню дисциплин, согласно направлению его обучения.

Для решения сформулированной задачи необходимо выделить элементы информации, которую нужно структурировать, ими являются дисциплина, термин и определение. Из выше перечисленных моделей представления знаний воспользуемся семантической моделью, так как с помощью семантической сети такие элементы информации представляются проще для дальнейших исследований. Представим, что элементарным объектом в семантической сети является термин. Тогда мы можем повысить эффективность усвоения знаний в этой сети с помощью алгоритма, который учитывает отношения терминов в разных дисциплинах.

Теоретически алгоритм сводится к тому, что поиск термина ведется среди выбранных пользователем дисциплин. Однако порядок найденных дисциплин может быть произвольным. Для создания какого-либо порядка среди этих дисциплин необходимо ввести критерий, согласно которому и будет выполняться упорядочивание. Роль критерия в алгоритме будет играть количество переходов внутри дисциплины S . Таким образом, система адаптируется под предпочтения пользователя, руководствуясь предположением, что более тщательное изучение дисциплины является порядком предпочтения самого пользователя. Для начального изучения, когда пользователь еще не выполнял вход в дисциплины (или когда критерий S имеет одинаковые значения для нескольких найденных дисциплин), оператором вводится дополнительный критерий упорядочивания дисциплин S' при формировании семантической сети.

Шаг 1. Пусть имеется множество дисциплин $D = \{d_i; i = 1, \dots, m\}$ и полное множество терминов $T = \{t_j; j = 1, \dots, n\}$. Каждый термин имеет имя и описание $t_j = \{name, discription\}$.

Введем критерии упорядочивания для каждой дисциплины $S = \{s_i; i = 1, \dots, m\}$, где $s_i = a_i + \delta/b_i$, a_i – количество переходов внутри дисциплины, b_i – отрезок времени, который прошел с момента последнего перехода по дисциплине, δ – поправочный коэффициент; $S' = \{s'_i; i = 1, \dots, m\}$, где s'_i – порядковый номер дисциплины в установленном операторе ряде предпочтений дисциплин (в порядке убывания). Чем больше значение s_i или s'_i , тем предпочтительнее является дисциплина для пользователя.

Каждой дисциплине поставим в соответствие булевый вектор отношений $R_i = \{r_{ij}; j = 1, \dots, n\}$. Если $r_{ij} = 1$, то это означает, что j -й термин содержится в i -ой дисциплине, а $r_{ij} = 0$ – наоборот, не содержится.

Пусть пользователь описывается множеством $U = \{D^U\}$, где $D^U \subseteq D$ – подмножество выбранных пользователем дисциплин (здесь алгоритм описывается с точки зрения теории множеств; при реализации проще всего будет использовать идентификаторы). Соответственно, каждая дисциплина $d^U \in D^U$ имеет свой вектор отношений $R_{i^U} = \{r_{i^U j}; j = 1, \dots, n\}$, $i^U = 1, \dots, m^U$. Здесь m^U – мощность множества D^U .

Шаг 2. Пользователь пожелал узнать описание слова w (это слово может быть термином с описанием в словаре). Определяется, в каких дисциплинах присутствует описание w :

$$\{d_i \in D^c; \text{if } w = t_j \text{ and } r_{ij} = 1; i = 1, \dots, m; j = 1, \dots, n\}$$

где $D^c \subseteq D$ – подмножество дисциплин, в которых содержится описание w . Аналогично каждая дисциплина $d^c \in D^c$ имеет свой вектор отношений $R_{i^c} = \{r_{i^c j}; j = 1, \dots, n\}$, $i^c = 1, \dots, m^c$. Здесь m^c – мощность множества D^c .

Шаг 3. Находим пересечение $D^x = (D^U \cap D^c) \subseteq D$. Аналогично каждая дисциплина $d^x \in D^x$ имеет свой вектор отношений $R_{i^x} = \{r_{i^x j}; j = 1, \dots, n\}$, $i^x = 1, \dots, m^x$. Здесь m^x – мощность множества D^x .

Шаг 4. Если $D^x = \emptyset$ – пустое множество, то доступное описание w отсутствует. Конец алгоритма. В этой вариации алгоритма поиск происходит только в области знаний пользователя U . Возможен вариант, когда предлагается альтернативный порядок дисциплин, не включенных в дисциплины пользователя, но содержащий искомый термин.

Шаг 5. Если $d^x \in D^x$ существует одно или несколько. В этом случае, необходимо упорядочить найденные описания. Каждой дисциплине $d_{i^x}^x \in D^x$ поставлено в соответствие значение s_{i^x} . Упорядочим множество D^x по убыванию значения s_{i^x} . Если для каких-либо дисциплин значения критериев s_{i^x} одинаковы, то множество эти дисциплины располагаются в порядке, согласно критерию s'_{i^x} , относительно друг друга. Получим упорядоченное множество $\overline{D^x}$ мощности m^x .

Шаг 6. Выходом алгоритма станет вектор ссылок на термины $t_j = w$, где $r_{i^x j} = 1$, который является визуальным представлением web-онтологии. Конец алгоритма.

Подобные алгоритмы просты для понимания, однако они делают структуру любого сайта адаптивной, т.е. подстраивающейся под индивидуальные требования пользователя. Эффективность процесса адаптации во многом зависят от качества словаря. Под качеством понимается состав терминов и их описания, а также, правильно ли они размещены в дисциплинах. Описанных алгоритмов вполне хватит, чтобы реализовать простейшую web-онтологию для семантической сети, описывающей дисциплины естественно-научного цикла. Такой подход даст более продуктивный результат, в отличие от обычных сайтов.

Для корректности информации словари по каждой конкретной дисциплине должны формироваться преподавателями и профессорами, специализирующимися на этих дисциплинах. То есть всю литературу по дисциплине, которую он рекомендует, где содержится информация по терминам, профессор предоставляет для размещения на выделенном интернет ресурсе.

Заключение

Главным преимуществом разработанной модели web-онтологии для семантической сети, описывающей естественно-научный цикл дисциплин, является подстраивающаяся система отношений терминов под пользователя и содержание сети. Другими словами, если пользователь захочет узнать значение какого-либо термина, то система предложит наиболее подходящий ряд описаний этого термина (причем, он будет упорядочен: сначала самый подходящий, затем менее и так далее).

Таким образом, предложенный способ использования семантической сети и web-онтологии повышает информативность сети, достоверность, доступность информации. В дальнейшем планируется развитие описанных технологий и алгоритмов.