

ОПРЕДЕЛЕНИЕ СТРУКТУРЫ ПРЕДМЕТНОЙ МОДЕЛИ ОБУЧАЕМОГО НА ОСНОВЕ УЧЕБНОГО ТЕКСТА

Янковская Т. В., студентка,
Щеголькова В.А., преподаватель ШИСум ГУ

На сегодняшний день наиболее известными способами представления знаний об обучаемом являются стереотипная и оверлейная модели. Первая используется для адаптации к группам, вторая позволяет полностью индивидуализировать обучение, являясь «функцией усвояемости» на концептном графе предметной области. Зависимость от структуры учебного материала не позволяет сделать ее универсальной для всех систем. Кроме того, если предметная область большая, то ручной способ построения является трудоемким. Т.о. возникает задача автоматизированного способа построения структуры предметной модели обучаемого (оверлейной модели), которую удобнее представить в виде графа.

Так как она основана на учебном материале, можно воспользоваться результатами исследований в области автоматического структурирования массивов текста. Однако обзор литературы [1,2,3] и др. показал, что в общем случае эта проблема пока не имеет однозначного и эффективного решения.

Решение нашей задачи по сравнению с общей облегчено тем, что для обучения обычно предоставляется размеченный текст. Он разделен на разделы и подразделы, выделены основные понятия и упражнения, определено функциональное назначение отдельных фрагментов. Т.о. автоматизированное создание графа предметной области на размеченном текстовом материале выглядит более оптимистично.

Пусть задана нормативная модель обучаемого в виде множества $tr_j, (j = \overline{1..n})$ ключевых слов, представляющих требования к знаниям и умениям.

Концептом K назовем неделимый элемент предметной области, предназначенный для изучения обучаемым и оцениваемый системой. Формально концепт можно представить в виде кортежа $K = (id, tr, keys)$, где id - уникальный идентификатор, tr - ключевое слово, определяемое концептом (может отсутствовать), $keys$ - список ключевых слов, которые встречаются в тексте концепта.

Пусть учебный текст состоит из множества $K_i, (i = \overline{1..m})$ концептов. Воспользуемся логичной упорядоченностью текста и пронумеруем концепты в порядке следования. Построим ориентированный граф $G(K, E)$ концептов предметной области на основе связи по ключевым словам. Для этого определим множество связей E .

Концепты K_p и K_q находятся в отношении $K_p \Rightarrow K_q$ («для изучения K_q необходимо знать K_p »), если ключевое слово tr_p концепта K_p находится в списке ключевых слов $keys_q$ концепта K_q . В матрице инцидентности таким парам поставим в соответствие значение 1, обратному отношению поставим в соответствие значение -1.

На основании построенного графа можно решить задачу предоставления фрагментов материала для обучения. При этом сформированный фрагмент должен быть тематически связан с некоторым понятием или группой понятий.

В работе использована идея «островной кластеризации», предложенная в исследовании [3] для тематической группировки больших корпусов текстов и сведения их в иерархическую структуру. В результате работы алгоритма было получено множество подграфов-островов, каждый из которых является тематическим фрагментом.

1. Виноградова Н.В., Митрофанова О.А., Паничева П.В. Автоматическая классификация терминов в русскоязычном корпусе текстов по корпусной лингвистике // Труды 9ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», - RCDL'2007. <http://www.dialog-21.ru/dialog2007/>
2. Ермаков А.Е. Автоматизация онтологического инжиниринга в системах извлечения знаний из текста // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008. - Москва, Наука, -2008.
3. Киселев М. В. Метод кластеризации текстов, учитывающий совместную встречаемость ключевых терминов, и его применение к анализу тематической структуры новостного потока, а также ее динамики / М. В. Киселев, В. С. Пивоваров, М. М. Шмулевич // Интернет-математика 2005. Автоматическая обработка веб-данных, - М. -2005. - с. 412-435.