

Применение концепции адаптивного контента для электронного образовательного ресурса

Д.А. Алдунин

аспирант кафедры инноваций и бизнеса в сфере информационных технологий
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: daldunin@hse.ru

Аннотация

В настоящее время рынок электронного обучения быстро растет как в финансовом, так и в географическом плане. Электронные образовательные ресурсы все чаще привлекают мультикультурную аудиторию, становятся доступны для людей с различным уровнем образования. В то же время в различных культурах существует своя специфика восприятия информации и различные подходы к образовательному процессу. Данная работа посвящена иллюстрации возможного решения для адаптации контента электронного образовательного ресурса для мультикультурной аудитории. Описанное решение основано на применении концепции адаптивного контента для построения индивидуальной образовательной траектории и подготовки контента в соответствии с индивидуальными культурными особенностями обучающихся и его компетенциями (имеющимися и желаемыми).

В ходе исследования была разработана модель образовательного процесса, ориентированная на конкретного обучающегося. В статье представлены как модель верхнего уровня, так и детализированная модель. В статье также описаны основные принципы формирования индивидуальной образовательной траектории на основании имеющихся и желаемых компетенций пользователя и статистических данных о его стиле обучения. В качестве демонстрации возможности применять исторические данные о том, как образовательный стиль влияет на успешное завершение курса, приведен статистический анализ данных об изучении студентами семи курсов на платформах HarvardX и MITx в течении 2013/14 учебного года, показавший статистическую значимость целого ряда параметров. Также в работе представлено сравнение алгоритмов, применимых для оценки вероятности успешного завершения курса обучающимся в зависимости от его стиля обучения.

Ключевые слова: электронное обучение, кросс-культурный контекст, математическое моделирование, веб-разработка, индивидуальная образовательная траектория, образовательная аналитика.

Цитирование: Aldunin D.A. Application of the adaptive content concept for an e-learning resource // Business Informatics. 2016. No. 4 (38). P. 27–34. DOI: 10.17323/1998-0663.2016.4.27.34.

Введение

В настоящее время кросс-культурная проблема является одним из основных аспектов, которые следует учитывать при создании электронного ресурса, ориентированного на представителей различных культур. С началом использования информационных технологий для доставки информации появились новые инструменты и методы ее представления. Некоторые из них могут как частично уменьшить, так и увеличить психологическую напряженность процесса приобретения

знаний. Согласно Г. Узилевскому [1], эргономическая семиотика является научной практикой, которая изучает проблемы, общие для семиотики, лингвистики и эргономики. Она способна решать целый ряд проблем, связанных с необходимостью сделать информацию понятной для целевой аудитории, в соответствии с когнитивной спецификой, вызванной культурными различиями. В эпоху всемирного проникновения Интернета в большинство областей жизни применение кросс-культурных аспектов эргономической семиотики очень важно.

С одной стороны, контент обычных электронных ресурсов, безусловно, воплощает национальную культуру своих создателей [2]. С другой стороны, контент должен соответствовать культурным и прагматическим ожиданиям обучающегося для максимальной эффективности представления информации. Например, можно заметить, что западный образовательный контент характеризуется открытостью, четкой логикой и предсказуемостью, дозированной информацией и отсутствием скрытого содержания, в то время как восточный контент, как правило, имеет сложную структуру и очень зависит от контекста [2].

Прагматические факторы обязательно должны быть приняты во внимание в процессе создания и мониторинга ресурсов, ориентированных на мультикультурную аудиторию. Существуют различные когнитивные модели восприятия и обработки информации в культурах Востока (коллективизм и склонность к диалектике) и Запада (индивидуализм и склонность к формальной логике). На Западе информация в основном воспринимается непосредственно через призму индивидуального восприятия человека и накладывается на существующий информационный фон. На Востоке пользователи просто собирают информацию без личного критического анализа при потреблении информационного контента [2].

1. Базовые принципы классификации культур

Центральным элементом предлагаемой концепции адаптивного контента является профиль обучаемого, состоящий из двух частей: культурно-когнитивного и компетентностного профилей. Культурно-когнитивный профиль содержит информацию о культурных особенностях — когнитивной специфике, специфике хода мыслей и эмоциональных (аффективных) различиях. Согласно исследованиям М. Холодной [3], когнитивный стиль отражает способ восприятия, анализа, структурирования и категоризации мира, стиль обучения. Компетентностный профиль содержит перечень знаний и умений учащегося и вектор его или ее образовательного и профессионального развития (то, чем он или она интересуется для дальнейшего обучения).

Имеет смысл основывать культурно-когнитивный профиль на шести культурных измерениях, предложенных Г. Хофстеде [4, 5], поскольку существует тест, который позволяет получить точные значе-

ния для каждого измерения, что делает этот подход очень удобным для программирования. К измерениям относятся: дистанцированность от власти (PDI), коллективизм против индивидуализма (IND), мужественность против женственности (MAS), избегание неопределенности (UAV), долгосрочная ориентация против краткосрочной (LTO), потворство своим желаниям против сдержанности (IVR). Также возможно провести сопоставление с классификациями культур Ф. Тромпенаарса [6], Г. Триандиса [7] и Р. Льюиса [8]. Классификация Хофстеде обусловлена окружением индивида, в то время как остальные классификации носят более индивидуальный характер и, таким образом, дополняют первую. Профиль также может быть расширен с помощью бинарной классификации стилей обучения, введенной Р. Фелдером и Л. Сильверман [9]. В этой модели стили обучения описываются как значения из пар крайностей: активный или рефлексивный, сенсорный или интуитивный, вербальный или визуальный, последовательный или глобальный. Для этой модели также существует тест для получения значений для конкретного обучающегося [10].

Компетентностный профиль должен быть основан на классификаторе компетентности. Компетенции следует разделить на систематические (ключевые компетенции, такие как тайм-менеджмент, умение анализировать статистические данные, работать в группе и т.д.) и профессиональные (связанные с конкретной профессиональной областью). Профессиональные компетенции также должны быть разделены на две группы: инструментальные и социально-личностные. Последние связаны с культурно-когнитивным профилем, таким образом осуществляется связь между профилями.

2. Модель обучающегося

Различными исследователями было представлено множество моделей обучающегося и процесса обучения. В качестве примеров можно привести онтологию области знания П. Данга и А. Флореа, предназначенную для применения в многоагентной системе электронного обучения [11], или модель А. Гарсиа-Габота, принимающую во внимание устройство, используемое обучающимся для доступа к электронному ресурсу [12]. Однако эти модели не несут комплексного видения обучающегося, специфики электронного обучения процесса и моделирования предметной области. Предлагаемая модель онтологии (*рисунок 1*) позволяет решить эту проблему.

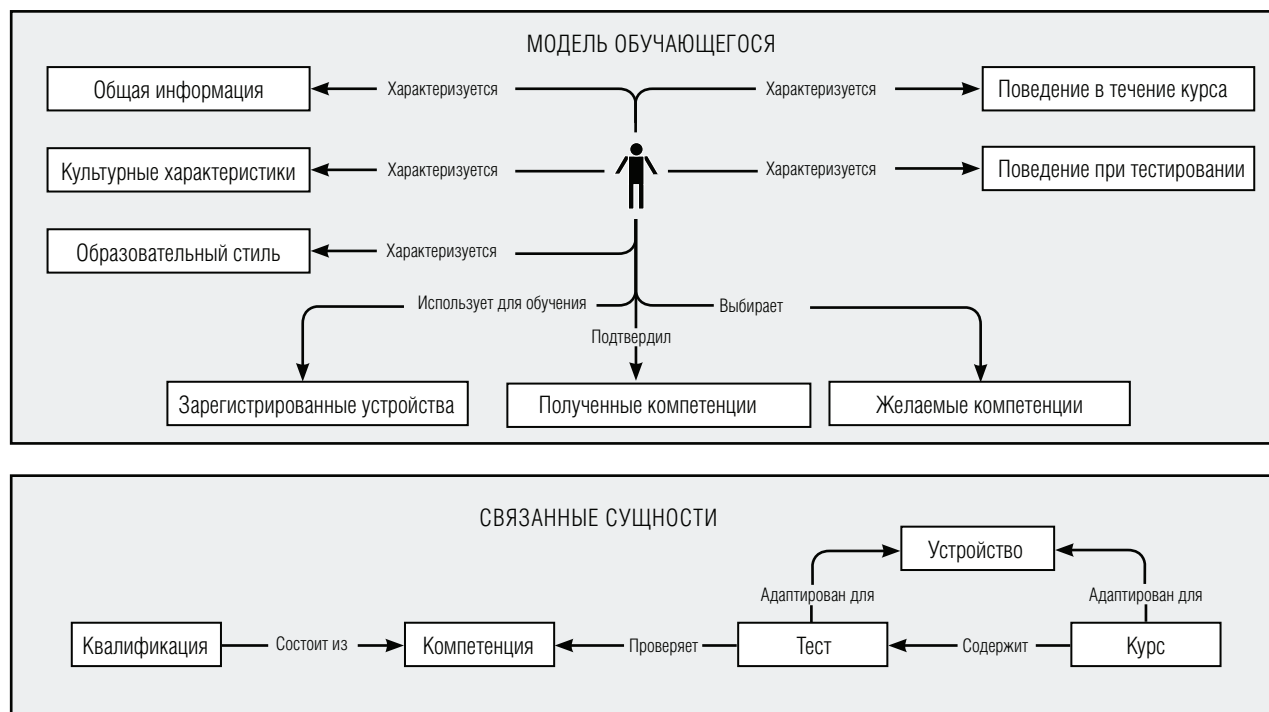


Рис. 1. Онтология, ориентированная на обучающегося

Полная модель, описывающая отношения и атрибуты сущностей, представлена на *рисунке 2*.

Ориентация представленной модели на обучающегося тесно связана с идеей клиентоориентированности в современной торговле, которая является ключом к бизнес-зрелости¹. Поскольку электронное обучение, по сути, является бизнес-процессом с определенными выгодами для обеих сторон, очевидно, что ориентированность на обучающегося является важным шагом для развития электронного обучения во всем мире.

3. Индивидуальная образовательная траектория

Индивидуальная образовательная траектория (ИОТ) — план обучения, учитывающий предпочтения обучающегося в окончательном наборе приобретенных компетенций, а также его склонности к конкретным формам представления познавательных объектов и задач, различным методам исследования и формам выражения идей, связанные с культурными и личностными характеристиками. Обучающийся должен выбрать образовательные курсы или желаемые компетенции для формирования ИОТ. Каждый курс может быть охарактеризован набором компетенций в соответствии с упо-

мянутым ранее классификатором компетенций, продолжительностью и стоимостью. Кроме того, длительность некоторых курсов может зависеть от конкретного обучающегося. Время, затрачиваемое на приобретение одной компетенции, также может быть различно для разных курсов. Кроме того, могут быть сформированы квалификации (наборы компетенций). Поэтому, когда обучающийся выбирает квалификацию (например, бизнес-аналитика в ИТ), это означает, что он должен получить ряд компетенций (анализ бизнес-процессов, системный анализ, моделирование процессов и т.д.), на основе чего может быть сформирован перечень необходимых курсов. Существуют различные способы формирования этого перечня: свести к минимуму количество курсов, свести к минимуму продолжительность, минимизировать цену и т.д. Обучающийся также должен иметь возможность выбирать курсы самостоятельно.

Таким образом, становится возможным создать математическую модель с целью получения набора из n компетенций в m курсах. Для каждой компетенции $comp$ существует множество курсов, которые обеспечивают получение этой компетенции $\langle c_1, \dots, c_i \rangle, 1 \leq i \leq m$; и наоборот, для каждого курса есть множество компетенций, получение которых

¹ "The journey toward greater customer centricity", Ernst & Young LLP, 2013

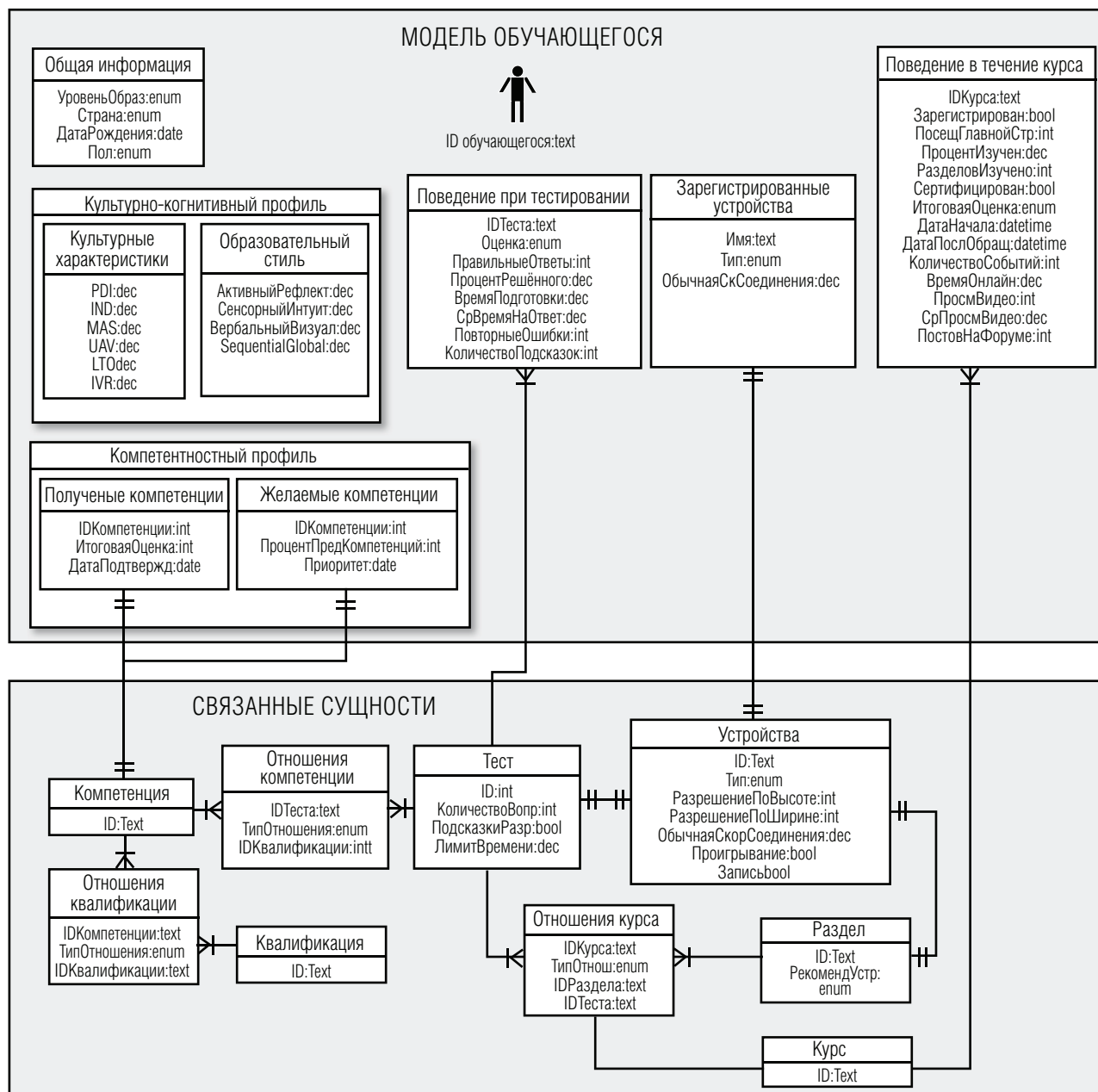


Рис. 2. Детализированная модель с атрибутами и отношениями сущностей

он обеспечивает $\langle postcomp_1, \dots, postcomp_i \rangle, 1 \leq i \leq n$. Кроме того, для каждого курса есть списки предварительно необходимых компетенций $\langle precomp_1, \dots, precomp_i \rangle, 1 \leq i \leq n$.

Для каждого курса существует значение $0 \leq h \leq 1$, которое характеризует сложность прохождения этого курса в зависимости от доли предварительно необходимых для этого курса компетенций, которыми уже обладает обучающийся.

Значения времени, необходимого для приобретения компетенции, могут быть определены как

$$T = CT \cdot (1 - H), \quad (1)$$

$$\text{где } T = \begin{bmatrix} T_1 \\ \dots \\ T_m \end{bmatrix}$$

время, затрачиваемое на прохождение курсов;

$$CT = \begin{bmatrix} CT_1 \\ \dots \\ CT_m \end{bmatrix}$$

общее время прохождения курсов;

$$H = \begin{bmatrix} H_1 \\ \dots \\ H_m \end{bmatrix} \text{ — сложность курсов.}$$

Итоговое время прохождения всех необходимых курсов зависит от предельного количества курсов для одновременного изучения и последовательности изучения. Поскольку общее время зависит от того, насколько подготовлен обучающийся для каждого курса, максимизация перекрестного использования компетенций в процессе изучения курсов является идеальным способом определить последовательность их изучения и минимизировать общее время. Это означает, что компетенции, получаемые в ходе изучения одного курса, должны быть использованы при изучении другого и так далее, пока это возможно.

Общая стоимость изучения курсов зависит от общего времени из-за таких постоянных затрат на единицу времени (FC), как плата за подключение к Интернету, амортизация устройства и т.д.:

$$PR_{Total} = \sum_{i=1}^m PR_i + FC \cdot T_{Total}, \quad (2)$$

где $PR = \begin{bmatrix} PR_1 \\ \dots \\ PR_m \end{bmatrix}$ — цена курсов.

Знание культурной специфики обучающегося позволяет прогнозировать возможные трудности в ходе изучения конкретного курса и даже вероятность успешной сдачи экзамена. Прогнозирование вероятных трудностей возможно благодаря следующим видам анализа:

- ♦ анализ того, владеет ли обучающийся необходимыми компетенциями;
- ♦ анализ корреляции между культурной спецификой обучаемого и необходимыми компетенциями;
- ♦ анализ того, как учащиеся со схожими культурными характеристиками прошли этот курс;
- ♦ анализ того, какое поведение обучающихся коррелирует с успешным завершением курса.

В качестве примера последнего типа анализа с помощью линейной регрессии были проанализированы несколько курсов из обезличенного набора данных о прохождении обучающимися курсов HarvardX и MITx в 2013 академическом году². В Таблице 1, представлены коэффициенты регрессии, определяющие, какие модели поведения являются более или менее важными для успешного прохождения курса. Все они имеют p-value ниже 0,05, что означает статистическую значимость всех коэффициентов. Коэффициент детерминации (R^2) определяет, какой процент результатов можно объяснить с помощью модели.

Такой анализ позволяет непосредственно поддерживать учащихся, вероятно имеющих неправильный стиль обучения, или, по крайней мере, информировать их о возможных проблемах и их причинах.

4. Адаптация образовательного контента

Учебный курс может быть представлен в виде ряда учебных материалов (мультимедийные или текстовые документы). Каждый обучающий материал должен быть разделен на блоки. Блок является атомарной частью учебного материала и должен быть помечен наиболее подходящими для него диапазонами культурных измерений Хофстеде (одно, два или даже все шесть измерений для очень специфичного контента) и индексом внутри материала. Такая концепция позволяет построить наиболее ценный вариант учебного материала для конкретного обучающегося. Например, если описывается случай успеха компании, индивидуалистам может быть представлен контент о личном вкладе великого лидера, в то время как тем, кто избегает неопределенности, могут быть представлены статистические данные и т.д. Когда обучающийся открывает учебный материал, этот материал должен быть подготовлен путем выбора наиболее подходящего блока для каждого значения индекса или блока по умолчанию, если подходящий блок не найден.

Кроме того, представление учебных материалов также может быть адаптировано. Есть пять ключевых моментов для такой адаптации [9]:

- ♦ **метафоры** — типичные образы, используемые в дизайне;
- ♦ **ментальные модели** — модели хода мысли из реальной жизни, использование которых стимулируется при использовании сайта;
- ♦ **навигация** — особенности перемещения внутри веб-страниц и между ними;
- ♦ **взаимодействие** — специфика человеко-компьютерного взаимодействия;
- ♦ **внешнее оформление** — особенности дизайна: цвета, формы, разметка, звуки, использование мультимедиа и т.д.

Адаптация по большинству из этих пунктов может быть автоматизирована. Например, алгоритмы компьютерного зрения (например, CaptionBot, основанный на Microsoft Cognitive Services) позволяют автоматически описывать изображения.

² <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26147>

Таблица 1.

Регрессионный анализ поведения, приводящего к успешной сдаче курсов

Название курса	R ²	Число взаимодействий с курсом	Число дней активности	Число просмотров видео	Число изученных глав	Число постов на форуме
Введение в химию твердого тела	0,81	0,479	0,182	–0,237	0,462	–0,007
Введение в компьютерные науки и программирование	0,78	0,370	0,406	–0,162	0,240	–0,005
Введение в биологию	0,75	0,814	0,173	–0,328	0,164	–0,021
Электричество и магнетизм	0,82	0,676	0,341	–0,243	0,104	–0,006
Механика	0,84	0,467	0,050	–0,043	0,454	0,017
Здоровье в цифрах	0,74	0,624	0,070	–0,295	0,424	–0,017
Здоровье человека и глобальные климатические изменения	0,67	0,793	0,031	–0,286	0,236	0,015

5. Математические модели и алгоритмы

Применение концепции адаптивного контента требует использования некоторых математических моделей и алгоритмов. Первая задача, которая должна быть решена, — как прогнозировать вероятность успешного завершения курса.

Для решения этой задачи можно использовать исторические данные о том, как обучающиеся с различными значениями культурных характеристик справлялись с курсом. Для подобного анализа панельных данных может быть использована логистическая регрессия. Бинарная логистическая модель используется для оценки вероятности бинарного ответа, основываясь на независимых переменных — предикторах:

$$\mathbb{P}\{y = 1|x\} = F(x_1, x_2, \dots, x_n), y \in \{0, 1\}. \quad (3)$$

В данном случае бинарный ответ — это успешное или неудачное завершение курса. Анализируя статистику предыдущих обучающихся, можно обнаружить, насколько важны различные параметры для достижения успеха в том или ином курсе и использовать меры этой значимости (коэффициенты регрессии θ), умноженные на значения для текущего обучающегося x для того, чтобы предсказать его результат. Логистическая функция (или сигмоид) записывается следующим образом:

$$F = \frac{1}{1 + e^{-\theta^T x}} = \frac{1}{1 + e^{-(\theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n)}}. \quad (4)$$

Коэффициенты регрессии должны быть определены для каждого курса с использованием исторических данных об окончании курса в качестве

обучающей выборки. С точки зрения минимизации затрат машинного времени не имеет никакого смысла производить пересчет коэффициентов регрессии после получения результата для каждого нового обучающегося, поскольку его влияние будет очень низким. Например, в обезличенном наборе данных о курсах HarvardX и MITx за 2013 академический год³ содержится более 640 000 записей (каждая представляет собой описание деятельности одного обучающегося в одном курсе EdX) всего о 13 учебных курсах.

Следующей задачей является формирование последовательности курсов. Как уже отмечалось, этот алгоритм должен быть основан на максимизации перекрестного использования компетенций в процессе изучения курсов. Обход в глубину (depth-first traversal) является наиболее подходящим алгоритмом для решения этой задачи. Он позволяет построить оптимальную последовательность из списка курсов и компетенций, которые они требуют и обеспечивают. Преимуществами этого алгоритма являются простота и преодоление неоптимальности обычного перебора.

Последней задачей верхнего уровня является формирование групп студентов со схожими характеристиками. Этот процесс может иметь разные цели: от формирования точного количества групп в рамках одного курса до формирования групп по одному или нескольким признакам. Для различных целей имеет смысл использовать различные алгоритмы кластеризации, такие как метод k-средних, DBSCAN или алгоритм сдвига среднего. Полное описание алгоритмов можно найти, например, в работе [13].

³ <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26147>

Кластеризация методом k-средних

Этот метод векторного квантования стремится разделить n наблюдений $(x_1, x_2, \dots, x_n)$ на $k \leq n$ кластеров $S = \{S_1, S_2, \dots, S_n\}$ на основе минимизации суммарного квадратичного отклонения точек кластеров от центров этих кластеров:

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2, \quad (5)$$

где μ_i — центр масс кластера S_i .

Преимущества:

1. Кластеризация методом k-средних является NP-трудной задачей, однако, существуют эффективные эвристические алгоритмы. Это делает метод k-средних более быстрым, чем большинство других алгоритмов кластеризации.

Недостатки:

1. Кластеризация методом k-средних требует предварительно определенного количества кластеров.

2. Кластеризация методом k-средних имеет высокую чувствительность к выбросам и шуму.

3. Кластеризация методом k-средних не очень хорошо работает с кластерами некруглой формы.

DBSCAN (Density-based spatial clustering of applications with noise — Плотностный алгоритм кластеризации пространственных данных с присутствием шума)

Этот алгоритм группирует точки, которые упакованы в пространстве тесно (имеют много соседей) и помечает точки, лежащие в одиночку, как выбросы (низкая плотность, ближайшие соседи слишком далеко). DBSCAN является одним из наиболее цитируемых в научной литературе⁴.

Преимущества:

1. DBSCAN не требует заранее определенного количества кластеров, в отличие от кластеризации методом k-средних.

2. DBSCAN способен находить кластеры любой формы.

3. DBSCAN является устойчивым к выбросам.

4. DBSCAN требует только два параметра (ε и минимальное количество точек, необходимое для формирования плотной области) и почти игнорирует порядок, в котором анализируются записи.

Недостатки:

1. DBSCAN не является полностью детерминированным.

2. Масштаб и данные должны быть хорошо изучены для правильного выбора параметров.

Алгоритм сдвига среднего

Этот алгоритм основан на непараметрическом анализе пространства признаков и использовании дискретных данных для определения местоположения максимумов функции плотности, породившей данные. Алгоритм сдвига среднего является итеративным и начинается с начальной оценки x :

$$m(x) = \frac{\sum_{x_i \in N(x)} x_i K(x_i - x)}{\sum_{x_i \in N(x)} K(x_i - x)}, \quad (6)$$

где K — взвешенное среднее значение плотности в окне;

$N(x)$ — окрестность x (множество точек, для которых $K(x) \neq 0$).

Разность $m(x) - x$, собственно, и называется сдвигом среднего.

Преимущества:

1. Алгоритм сдвига среднего не требует заранее определенного количества кластеров, в отличие от метода k-средних.

2. Алгоритм сдвига среднего способен находить кластеры любой формы.

3. Алгоритм сдвига среднего требует только один параметр (размер окна).

4. Алгоритм сдвига среднего хорошо приспособлен для использования гауссового ядра.

Недостатки:

1. Выбор размера окна для алгоритма сдвига среднего нетривиален.

Обращаясь к технической части, важно отметить, что для реализации подобного сервиса следует использовать базу, а не хранилище данных, так как основным типом запросов будет выбор, а не вставка записей. Интеграция с существующими информационными системами (такими как системы управления обучением в университетах и т.д.) может быть реализована с помощью промежуточного репликационного слоя и табличных триггеров. Таким образом, любое изменение в одной системе может

⁴ DBSCAN is on rank 41 within most cited data mining articles according to Microsoft academic search, when accessed on: 21.08.16

быть воспроизведено сразу в ряде таблиц другой системы. Это позволит не адаптировать структуру базы данных сервиса для каждой интеграции.

Заключение

Применение концепции адаптивного контента описанным способом может значительно улучшить процесс электронного обучения за счет следующих аспектов:

- ♦ повышение доступности для понимания учебных материалов. Процесс электронного обучения не обеспечивает такой поддержки тьютора, как при очном обучении; это в значительной степени автодидактический процесс. Таким образом, очень важно уменьшить психологическое напряжение, вызванное чувством одиночества, недружелюбным интерфейсом и непонятными учебными материалами;

- ♦ возможность построить индивидуальную образовательную траекторию;

- ♦ возможность сформировать компетентностный профиль обучающегося. По сути, такой профиль является готовой частью резюме и портфолио.

В качестве направлений дальнейших исследова-

ний можно отметить следующие проблемы:

- ♦ необходимость приоритизации культурных характеристик в процессе формирования учебных материалов в случае смешанных культур (таких как китайские студенты, обучающиеся в США в течение длительного времени) и, следовательно, развитие системы взвешенных коэффициентов;

- ♦ проблемы, связанные с дифференциацией культурных групп в целевой аудитории (взаимодействие представителей разных групп и т.д.);

- ♦ необходимость мониторинга удовлетворенности пользователей и прогнозирование изменений в предпочтениях пользователей на основе данных о пользователях со схожими параметрами.

Данная публикация является частью серии работ, посвященных анализу соответствия культурной специфики целевой аудитории и методов адаптации контента и дизайна, необходимой для обеспечения максимального комфорта и легкости процесса электронного обучения для мультикультурной аудитории. Конечным результатом будет разработка программной среды, которая облегчит подготовку и отображение контента, чувствительного к культурной специфике. ■

Литература

1. Узилевский Г.Я. Начала эргономической семиотики. Орел: ОРАГС, 2000.
2. Huang K.-H., Deng Y.-Sh. (2008) Social interaction design in cultural context: A case study of a traditional social activity // International Journal of Design. 2008. No. 2. P. 81–96.
3. Холодная М.А. Когнитивные стили. О природе индивидуального ума. СПб: Питер, 2004.
4. Hofstede G. Culture's consequences: International differences in work related values. Beverly Hills, CA: Sage Publications, 1980.
5. Hofstede G., Hofstede J.G., Minkov M. Cultures and organizations: Software of the mind. N.Y.: McGraw-Hill, 2010.
6. Trompenaars F., Hampden-Turner C. Managing people across cultures. London: Capstone Publishing, 2004.
7. Triandis H. Culture and self-deception: A theoretical perspective // Social Behavior and Personality. 2011. No. 1. P. 3–14
8. Lewis R. When cultures collide. Managing successfully across cultures. Boston, London: Nicholas Brealey International, 2006.
9. Felder R.M., Silverman L.K. Learning and teaching styles in engineering education // Engineering Education. 1998. No. 7 (78). P. 674–681.
10. Soloman B.A., Felder R.M. Index of learning styles questionnaire. North Carolina State University. [Электронный ресурс]: <http://www.engr.ncsu.edu/learningstyles/ilsweb.html> (дата обращения 01.09.2016).
11. Dung P.Q., Florea A.M. Adaptation to learners' learning styles in a multi-agent e-learning system // Internet Learning. 2013. No. 1. P. 11–20.
12. Garcia-Gabot A. A proposal of a multi-agent system for adapting learning contents to user competencies, context and mobile device. Research paper. Bratislava: Slovak University of Technology, 2013.
13. Gan G., Ma C., Wu J. Data clustering: Theory, algorithms, and applications. Philadelphia: SIAM, 2007.