

004:378

АНАЛИЗ ПОДХОДОВ К ХРАНЕНИЮ КОНТЕНТА ОБРАЗОВАТЕЛЬНОГО ПОРТАЛА ВУЗА

© 2008 г. А.Н. Иванченко, О.М. Колокольцева

Роль образовательных порталов в повышении качества обучения

В связи с предстоящим вступлением России во Всемирную торговую организацию актуальным становится повышение качества всех сфер деятельности, включая образование.

Особую значимость в эффективном управлении качеством образования приобретают принципы его открытости: предоставление разнообразных наборов образовательных услуг, возможность непрерывного обучения и получения высшего и дополнительного профессионального образования широкими слоями населения [1]. Развитие этих принципов возможно на основе соблюдения требований образовательных стандартов и создания интегрированной и распределенной информационно-образовательной среды отдаленного образования, ведущая роль в формировании которой принадлежит вузам, поскольку именно они открывают доступ к получению высшего образования тем, кто успешно оканчивает среднюю школу или средние специальные учебные заведения.

Создание единой информационно-образовательной среды вуза, представление, интеграция информации (учебные материалы, тестовые задания, расписания занятий, новости, объявления, справочники, административные базы данных), организацию дистанционного обучения и взаимодействие учебного заведения с другими организациями сферы образования обеспечивает образовательный портал вуза.

Цель любого портала – это прежде всего предоставление конкретному пользователю необходимой информации в течение минимального времени и без дополнительных временных затрат на просмотр информационных материалов, переключение между различными интерфейсами и т.д. Для достижения этой цели используются механизмы персонализации и администрирования контента [2].

Персонализация – это система регистрации пользователя. Она позволяет применять механизмы регистрации информации и анализа работы пользователя, с помощью которых удается определить ту область, которая может заинтересовать посетителей портала.

Под образовательным контентом понимаются Интернет-ресурсы (образовательные и учебно-методические материалы), созданные специально для использования в процессе обучения на определенной ступени образования. Контент представляет собой эффективное средство мотивации, «привязывания» посе-

тителя к portalу, поэтому его правильный подбор и администрирование определяют успешное функционирование портала в целом [2].

При создании и накоплении информационных ресурсов в вузах необходимо обеспечить интеграцию образовательного контента, созданного различными кафедрами и другими подразделениями (например, сотрудниками Центра новых информационных технологий или филиалов) вуза. Поскольку кафедры и подразделения вуза, как правило, являются территориально распределенными, учебный материал может существовать в различных версиях и периодически корректироваться. При таких условиях необходим удобный, быстрый доступ к отдельным информационным ресурсам и транспортировка получателям нужной информации.

Для повышения качества обучения необходимо наполнять курсы мультимедийной информацией, которая значительно увеличивает воспринимаемость теоретического материала. Изучение многих дисциплин невозможно без проведения лабораторных работ. Со временем оборудование устаревает, и в большинстве случаев учебное заведение не способно создать лабораторные установки для их проведения по финансовым или иным причинам. Для решения этой проблемы можно использовать программы-тренажеры, средства имитационного моделирования, а также включать в электронные учебники видеоматериалы, содержащие описание и демонстрацию проведения лабораторной работы. При наличии большого количества таких учебных материалов возникает проблема их хранения и доставки пользователю. Качество и скорость передачи учебных материалов по сети зависит от пропускной способности каналов. Предоставление материалов на CD-ROM позволяет передавать большие объемы мультимедийной информации эффективным способом, не связанным с проблемой пропускной способности каналов. Однако такой способ передачи является трудоемким с точки зрения времени и финансовых затрат на формирование и пересылку учебных материалов обучающимся [3].

Проблемы управления аудио, видео и мультимедийным контентом, его хранением, интеграцией, доступом к нему и доставкой пользователям образовательного портала являются актуальными, так как в последнее время отмечается тенденция увеличения объема образовательной мультимедиа-информации [4]. Необходима разработка стратегий для управления хранением и использования такого контента. При

этом следует стремиться к достижению высокой реактивности портала и как можно более полному удовлетворению запросов пользователей.

Подходы к хранению контента образовательных порталов

Образовательный контент большинства порталов размещается в *централизованном хранилище*, которое содержит множество образовательных информационных ресурсов (ОИР) $R = \{r_i, i = \overline{1, n}\}$ и их метаданные, а также данные для функционирования интерактивных сервисов.

Организация централизованного хранения образовательного контента реализуется на основе двухзвенной и трехзвенной архитектуры клиент–сервер. *Двухзвенная архитектура клиент–сервер* позволяет создать портал в условиях ограниченного бюджета проекта при относительно небольшом количестве пользователей портала, но не подходит для порталов с большим объемом мультимедийных образовательных ресурсов [5, 6]. *Трехзвенную архитектуру клиент–сервер* целесообразно применять при большом количестве одновременно работающих клиентов, интенсивно генерирующих запросы к базе данных, и больших объемах образовательных ресурсов, предоставляемых пользователям.

При построении порталов на основе двухзвенной и трехзвенной архитектур клиент–сервер объединение образовательных информационных ресурсов в единое хранилище данных может быть затруднительно по следующим причинам:

- на процесс интеграции влияет территориальная распределенность подразделений, поскольку многие вузы являются многофилиальными структурами с множеством кафедр, ведущих подготовку по широкому набору специальностей;

- ограниченность ресурсов жесткого диска сервера баз данных. Так как мультимедийные курсы и видеофайлы имеют большой объем, то не всегда возможно представить все образовательные ресурсы в рамках единого хранилища портала.

Решением вышеуказанных проблем является построение образовательного портала на основе концепции распределенного хранения информации или, в случае использования в учебном процессе мультимедийных учебных курсов, *применение специальных систем хранения*, например RAID-массивов, SAN и NAS.

Использование систем хранения RAID, SAN и NAS решает проблему ограниченности ресурсов жесткого диска сервера баз данных, но не избавляет от сложностей интеграции информационных ресурсов в единое хранилище портала. Кроме того, системы хранения данных имеют достаточно высокую стоимость.

При *распределенном хранении образовательного контента* возможны два подхода к хранению информационных ресурсов: либо существует множество баз данных $D = \{d_j, j = \overline{1, m}\}$ ОИР, либо множество информационных ресурсов $R = \{r_i, i = \overline{1, n}\}$ размещается

в распределенной базе данных. Концепция распределенного хранения информационных ресурсов реализуется при построении образовательного портала на основе распределенного хранилища информации на основе концепции GRID [3, 7–9].

Распределенный подход к хранению контента позволяет решить и проблему обновления ОИР, и проблемы их интеграции. Но при таком подходе предъявляются требования к качеству, загруженности канала передачи данных и доступности серверов, на которых размещены ОИР. Для устранения недостатков централизованного и распределенного подходов необходимо исследовать *комбинированный подход к хранению образовательного контента*. В этом случае возможна репликация части информационных ресурсов на web-сервер или сервер баз данных с других серверов или рабочих станций, т.е. к определенным ОИР осуществляется централизованный доступ, а к другим – децентрализованный. Для реализации комбинированного подхода может быть предложена следующая концепция: используется двухуровневая архитектура клиент–сервер, в базе данных на web-сервере размещаются метаданные ОИР и данные, необходимые для работы служб портала. Образовательный контент хранится в распределенной базе данных на файл-серверах кафедр и подразделений образовательного учреждения, ответственных за их создание. Доставка ОИР, хранящихся на файл-серверах, осуществляется по локальной сети вуза.

Управление образовательным контентом при комбинированном подходе к его хранению

При использовании комбинированного подхода актуальность приобретают следующие задачи: определение критерия, на основе которого будет производиться репликация; оценка использования образовательного контента в учебном процессе; выбор информационных ресурсов, которые необходимо реплицировать; перемещение и удаление ОИР из базы данных web-сервера; возможность прогнозирования спроса пользователей портала на определенные ОИР.

Выбор критерия. Репликация информационных ресурсов на web-сервер может осуществляться в зависимости от следующих критериев: минимизация объема передаваемых данных по сети ($\min V$), минимизация времени реакции системы на запросы пользователей ($\min T$), минимизация стоимости хранения ОИР ($\min C$); минимизация загруженности web-сервера ($\min L$). Наиболее важными критериями из них являются минимизация времени реакции системы на запросы пользователей ($\min T$) и загруженности сервера баз данных или web-сервера ($\min L$).

Проблемы определения загрузки сервера баз данных и web-сервера были рассмотрены в [10–12]. В работе [10] проводится анализ системы Интернет-порталов, которая представлена крупными Интернет-порталами (федеральными, региональными, специализированными), и получены аналитические выражения для расчета основных параметров систем: пропускная способность, загрузка web-серверов и серверов

ЕД параметры очередей, среднее время обслуживания заявки заданного класса. В статье [11] рассматривается модель обработки запросов системы управления web-сайтами и исследуется задача минимизации отказов в обработке запросов. В диссертационной работе [12] предложен подход к прогнозированию функциональной загруженности портала с использованием искусственных нейронных сетей.

Дальнейших исследований требуют вопросы загруженности web-сервера и сервера баз данных, времени реакции портала на запросы пользователя в зависимости от размещения образовательных ресурсов. Также необходимо рассмотреть возможность перемещения (удаления) ОИР с web-сервера при достижении высокой загруженности, которая является одним из факторов, влияющих на количество отказов в обработке запросов.

Формирование информационного наполнения портала и оценка эффективности информационных ресурсов. Для решения задачи управления хранением ОИР необходимо проводить мониторинг их использования в учебном процессе. Самый распространенный подход к оценке учебного воздействия ОИР – экспертная оценка, которая позволяет выявить их качество и предсказать, как тот или иной ресурс может быть использован в процессе обучения. Однако данная оценка является субъективной, поскольку не учитывается востребованность образовательного контента в учебном процессе.

В работе [13] для оценки эффективности использования Интернет-ресурсов предложен подход, основанный на значении параметра «величина конверсии», который показывает отношение количества пользователей, достигших некоторой цели (U_s), к общему числу пользователей (U) и определяется по формуле: $CR_s = U_s / U$. «Цель» – это действие или группа действий, которые являются значимыми для данного информационного ресурса, например, прохождение он-лайн-тестирования, участие в интерактивных уроках, загрузка ресурса.

Для оценки использования ресурса в учебном процессе необходима адаптация способа оценки, предложенной в [13], на основе анализа статистики и журналов web-сервера портала, запросов пользователей к образовательным ресурсам. При этом следует учесть, что пользователей можно разделить на две сгруппированные категории: зарегистрированные пользователи, пользующиеся порталом в целях дистанционного обучения, незарегистрированные пользователи, которые также могут использовать образовательный ресурс в учебных целях. Кроме того, при определении индикатора использования ОИР в учебном процессе могут быть приняты во внимание опросы, проведенные среди пользователей портала, а также рейтинги ОИР.

Репликация ОИР. Решение задачи распределения и динамического перераспределения ОИР было предложено в работе [14]. Объектом исследования является отраслевая система дистанционного обучения, в которой ОИР, насыщенные мультимедийной инфор-

мацией, распределены по территориальным округам. Задача перераспределения и динамического управления перемещением ОИР сводится к анализу связности требований на типы ресурсов, для описания этого процесса используются марковские цепи. Модели, предложенные в работе [14], позволяют проводить перераспределение ОИР на основе прогноза использования ресурсов, но не учитывают реальное использование образовательных ресурсов в учебном процессе и качество выполнения запросов пользователей, которое зависит, в первую очередь, от загруженности web-сервера и реактивности образовательного портала. Поэтому разработка методов и алгоритмов для решения задач, возникающих при комбинированном подходе к хранению образовательного контента, является актуальной.

Прогнозирование использования образовательного контента. При выборе информационных ресурсов и их оценке нужно принять решение, какие ресурсы будут храниться централизованно, а к каким будет осуществляться динамический доступ. При этом необходимо учитывать прогноз использования ресурсов для изменения структуры образовательного контента портала: удалять из базы данных web-сервера наименее запрашиваемые ресурсы и включать в информационное наполнение более «перспективные» ресурсы, которые будут использоваться в образовательных целях. При составлении прогнозов следует учесть данные востребованности ресурсов, соответствие ОИР дисциплинам учебного плана, расписание занятий на следующий семестр студентов очной формы обучения, график сессий и установочных занятий студентов заочной формы обучения.

С учетом выявленных требований задача управления контентом образовательного портала может быть сформулирована следующим образом. Множество пользователей образовательного портала $A = \{A_s, s = \overline{1, s_0}\}$ формируют запросы в соответствии с интенсивностями $\lambda = (\lambda_1, \dots, \lambda_s, \dots, \lambda_{s_0})$. Время формирования запроса s -м пользователем является случайной величиной, распределенной по экспоненциальному закону с плотностью $f_s(t) = \lambda_s e^{-\lambda_s t}$, $s = \overline{1, s_0}$. Образовательные информационные ресурсы представлены множеством $R = \{r_i, i = \overline{1, n}\}$. ОИР $r_i \in R$ характеризуется подмножеством $X = \{K, V, TP\}$, где $K = \{k_d, d = \overline{1, K_0}\}$ – множество разделяемых единиц контента, связанных в рамках определенной модели, K_0 – количество разделяемых единиц контента, составляющих ОИР, $V = \{v_d, d = \overline{1, K_0}\}$, где v_d – объем в байтах каждой разделяемой единицы контента, $TP = \{tp_d, d = \overline{1, K_0}\}$, где tp_d – тип разделяемой единицы контента (текст, изображение, аудиофрагмент, видеофрагмент).

Полное множество запросов состоит из двух непересекающихся подмножеств Q^F – множество запросов на чтение, Q^W – множество запросов на обновление ОИР. Связь между множеством запросов

$Q = Q^F \cup Q^W$ и множеством ОИР задается матрицами $B^F = \|b_{vd}^F\|$ и $B^W = \|b_{kd}^W\|$. Элементы b_{vd}^F матрицы B^F , идентифицирующие взаимосвязи между запросами на чтение и образовательными информационными ресурсами R , определяются следующим образом:

$$b_{vd}^F = \begin{cases} 1, & \text{если ОИР используется при выполнении} \\ & \text{запроса } q_v \in Q^F; \\ 0 & \text{– в противном случае.} \end{cases}$$

Элементы b_{kd}^W матрицы B^W формируются так:

$$b_{kd}^W = \begin{cases} 1, & \text{если } q_p \in Q^W \text{ модифицирует ОИР } r_i; \\ 0 & \text{– в противном случае.} \end{cases}$$

Все ОИР размещены на множестве узлов ЛВС $U = \{u_m, m = \overline{1, M}\}$. Параметры, определяющие характеристики сети, заданы тройкой вида: $P = \langle WR, DA, SP \rangle$: $WR = (WR_1, \dots, WR_m, \dots, WR_M)$, где WR_m – среднее время поиска и считывания данных в узле сети $u_m \in U$; $DA = (DA_1, \dots, DA_m, \dots, DA_M)$, где DA_m – среднее время доступа к данным узла сети; $SP = \|SP_{mh}\|$, $m, h = \overline{1, M}$, где SP_{mh} – среднее время передачи данных по каналам сети между узлами $u_m \in U$ и $u_h \in U$.

Распределение ОИР по узлам ЛВС задается в виде бинарной матрицы $G = \|g_{mi}\|$, $m = \overline{1, M}$, $i = \overline{1, n}$, элементы которой принимают значения $g_{mi} = 1$, если ОИР r_i хранится на узле u_m , в противном случае $g_{mi} = 0$.

Предполагается, что выбор пользователем запроса носит вероятностный характер, значения вероятностей определяются элементами матрицы $F = \|f_{sy}\|$, $s = \overline{1, S_0}$, $y = \overline{1, Q_0}$, f_{sy} – вероятность формирования пользователем A_s запроса $q_y \in Q^F \cup Q^W$.

Необходимо построить математическую модель, которая позволяла бы найти такое подмножество ОИР $R^{res} = \{r_j^{res}, j = \overline{1, n}\}$ множества R , которые необходимо реплицировать на web-сервер. При этом среднее время реакции системы на запросы пользователей T и количество невыполненных запросов пользователя Q^{ost} из-за загруженности web-сервера должно быть минимальным.

На параметры модели накладываются следующие ограничения:

1. ОИР могут быть размещены либо на одной из рабочих станций ЛВС, либо одновременно и на узле ЛВС, и на web-сервере, т.е.

$$r_i \in R: \sum_{i=1}^n g_{mi} \leq 2.$$

2. Если ОИР копируется на web-сервер, то происходит репликация всех разделяемых единиц контента k_d , из которых формируется этот ОИР.

3. Суммарный объем разделяемых единиц контента всех ОИР, реплицируемых на web-сервер, не должен превышать объема, выделенного для их хранения на жестком диске web-сервера:

$$\sum_{i=1}^n \sum_{d=1}^{K_0} v_{id} \leq VD,$$

где VD – количество байт на жестком диске web-сервера, выделенных для хранения ОИР.

4. При оценке времени реакции системы на запросы пользователей будут учитываться только запросы на чтение ОИР $q_{vd} \in Q^F$, поскольку редактировать ОИР, как правило, могут пользователи, имеющие соответствующие права: администраторы портала, авторы курсов, и основная нагрузка на портал состоит в выполнении запросов на предоставление и чтение образовательной информации.

Таким образом, задача нахождения множества R^{res} будет иметь вид: найти R^{res} такое, что $T \rightarrow \min$,

$Q^{ost} \rightarrow \min$, при условии, что $r_i \in R: \sum_{i=1}^n g_{mi} \leq 2$,

$$\sum_{i=1}^n \sum_{d=1}^{K_0} v_{id} \leq VD.$$

Заключение

При рассмотрении подходов к хранению контента образовательных порталов можно сделать следующие выводы:

1. При хранении в единой базе данных образовательных информационных ресурсов обеспечивается высокая реактивность запросов пользователей портала, а также возможность централизованного резервного копирования и восстановления базы данных. Недостаток такой концепции – трудность управления обновлением образовательных информационных ресурсов.

2. Распределенный подход позволяет реализовать портал, в котором необходимо работать с большим объемом мультимедийной информации в условиях ограниченности ресурсов жесткого диска сервера, а также портал распределенного университета, когда сложно следить за обновлением информации и сложно интегрировать все информационные ресурсы в единое хранилище данных. Однако при таком подходе предъявляются высокие требования к пропускной способности и загрузке каналов передачи данных.

3. Для построения портала вуза с большим объемом мультимедийной информации целесообразным является использование комбинированного подхода к хранению контента, при котором часть образовательного контента реплицируется на web-сервер с другими серверами или порталами. Реализация такого подхода к хранению образовательного контента позволяет решить проблемы интеграции ОИР, низкой реактивности и пропускной способности каналов передачи данных.

4. При комбинированном подходе к хранению образовательного контента необходимо определить количество информационных ресурсов R^{res} , к которым будет осуществляться централизованный доступ. В качестве критериев эффективности предполагается использовать время реакции системы на запросы пользователей T и число невыполненных запросов пользователей Q^{lost} .

5. Необходима разработка методик оценки использования ОИР в учебном процессе и методики прогнозирования спроса пользователей на ОИР с целью разработки стратегии развития образовательного портала.

Литература

1. Основы открытого образования / А.А. Андреев, С.Л. Каплан, Г.А. Краснова, С.Л. Лобачев, К.Ю. Лупанов, А.А. Поляков, А.А. Скамницкий, В.И. Солдаткин; Отв. ред. В.И. Солдаткин. Т. 1. М., 2002.
2. Комарцев Е.М. Образовательные порталы как средство систематизации и структурирования информации: Дис. ... канд. пед. наук. Ставрополь, 2004. 207 с.
3. Васильев П.М. Методы и технологии доступа к видеoinформации системы дистанционного обучения с использованием распределенной базы данных: дис. ... канд. техн. наук. Дубна, 2006.
4. Герасимов В.В., Гугель Ю.В., Курмышев Н.В., Сигалов А.В. Система образовательных порталов России: анализ телекоммуникационной инфраструктуры, общие требования к аппаратным платформам, технические аспекты размещения // Образовательные порталы России. 2004. Вып. 1. С. 25 – 129.
5. Позднеев Б.М., Буханов А.Н. Анализ систем управления информационным содержанием образовательных порталов // http://magazine.stankin.ru/arch/n_20/3/index.html, 2002.
6. Лясин А.С. Как создать портал в Internet: Основы использования web-технологий. М., 2003.
7. Анищенко Н.Г., Васильев П.М., Кореньков В.В., Крюков Ю.А. Хранение и доставка учебных видеоматериалов при дистанционной форме обучения // <http://www.physicsnet.ru/ru/problems/journal/61>, 2007.
8. Жижимов О.Л., Мазов Н.А., Фролов А.С. Доступ к базам данных ISIS из Internet и построение распределенной информационной системы // Вычислительные технологии. 1997. Т. 2. № 3. С. 45 – 50.
9. Шокин Ю.И., Ламин В.А., Федотов А.М., Барахнин В.Б., Жижимов О.Л., Мазов Н.А., Пищик Б.Н., Покровский Н.Н. Распределенная информационная система «Виртуальный музей науки и техники СО РАН» // Тр. 5-й Всерос. науч. конф. RCDL'2003. Новосибирск, 2003. С. 112 – 116.
10. Райнова О.Д. Разработка моделей и методов повышения эффективности функционирования системы образовательных Интернет-порталов: Автореф. дис. ... канд. техн. наук. М., 2006.
11. Быков М.Ю. Модель процесса обработки запросов системой управления web-сайтами // Вычислительные методы и программирование. 2005. Т. 6. С. 53 – 56.
12. Хапланов Н.Л. Разработка методов и средств построения распределенного информационно-справочного портала дистанционного обучения: дис. ... канд. техн. наук. М., 2004.
13. Соколов С.Н. Экспертиза образовательного Интернет-ресурса на основе анализа его использования // Новые информационные технологии в образовании: Материалы междунар. науч.-практ. конф. Екатеринбург, 26 – 28 февраля 2007 г. / Рос. гос. проф.-пед. ун-т. Екатеринбург, 2007. Ч. 2. С. 48 – 49.
14. Мельник В.Г. Перераспределение информационных ресурсов в телекоммуникационной системе дистанционного обучения: Дис.... канд. техн. наук. М., 2004.

Южно-Российский государственный технический университет
(Новочеркасский политехнический институт)

11 декабря 2007 г.