

Д.О. Прокофьев, аспирант 2-го года обучения, dprokofiev@hse.ru

Департамент компьютерной инженерии МИЭМ НИУ ВШЭ, Москва

В.А. Старых, кандидат технических наук, профессор, vstarykh@hse.ru

Руководитель департамента компьютерной инженерии МИЭМ НИУ ВШЭ, Москва

Моделирование образовательных процессов и их оптимизация на примере модели работы с электронными образовательными ресурсами

Аннотация

В ходе исследования рассмотрены основные проблемы автоматизации и оптимизации процессов в системе образования с помощью ВРМС и технологий класса больших данных, вопросы моделирования образовательных процессов, интеграции процессных и аналитических систем для построения образовательной инфраструктуры на примере модели работы с электронными образовательными ресурсами.

Ключевые слова: автоматизация образовательные процессы, ВРМС, большие данные, оптимизация процессов, моделирование процессов, электронные образовательные ресурсы.

Введение

Одним из приоритетных направлений государственной политики в области образования является информатизация. Были приняты и реализованы различные программы, направленные на внедрение информационно-коммуникационных технологий в образовательный процесс, результатом которых явилось: обеспечение всех школ компьютерами; повсеместный доступ к Интернет; создание информационных ресурсных центров. Таким образом формируется информационно-образовательная инфраструктура, для автоматизации процессов

которой используются различные типы информационных систем.

К основным процессам информационно-образовательной инфраструктуры можно отнести специализированные процедуры: собственно образовательные процессы, административные функции, формирование расписаний, планирование загрузки, ведение журнала, использования помещений, учет личных данных и многое другое [1]. Эти процессы требуют методологического сопровождения: полного, разнообразного, стандартизированного, обновляемого. Их можно и нужно описывать как модели, используя преимущества процессного подхода. Модели могут стать основой нового подхода к быстрой разработке сложных информационных систем, объединяющих инфоориентированный и процессный подходы, применительно к определенному классу задач автоматизации [2]. Достоинством такого подхода может служить и то, что модели процессов в совокупности с описаниями данных (справочников, классификаторов, , бизнес-правил (то есть комплекса регламентов, регулирующего законодательства и т.п.) могут быть расширены до комплексных моделей образовательной деятельности.

Комплексное моделирование образовательной деятельности

При моделировании образовательных процессов и включении их в комплексную модель необходимо помнить, что эти модели должны быть обобщены для возможного повторного использования и дальнейшей стандартизации — не имеет смысла тратить ресурсы на исследование предметной области и построение моделей каждый раз, когда требуется автоматизировать новые процессы, оптимизировать существующие, либо каким-либо другим образом вносить изменения в образовательные процессы. В таких случаях разумно предложить внедрение стандартизованных метаданных

процессов. Они могут храниться в открытых централизованных репозиториях моделей процессов, подобным уже функционирующим репозиториям моделей электронных учебных курсов [3].

Во-первых, метаданные процессов, или модели процессов, должны быть связаны с описаниями данных, которые конкретный процесс использует — метаданными, наборами классификаторов и справочников. Эта связь будет особенно важна в комплексной модели образовательной деятельности. Такая связь может быть реализована как расширение классических стандартов для описаний процессов, которые, как неоднократно отмечалось, по определению не поддерживают никаких прямых включений описания данных, либо ссылок на них. [4] Она не может привязываться к определенным стандартам и должна быть реализована как адаптер для подключения различных описаний метаданных.

Во-вторых, следует подчеркнуть, что такие модели должны иметь открытый формат для упрощения взаимодействия между различными субъектами системы образования, в том числе и не ограничивающихся образовательным пространством Российской Федерации, например, международных стандартизирующих организаций. Таким образом, эти модели должны соответствовать требованиям переносимости и интероперабельности.

Например, уже существует модель ресурсов для систем дистанционного обучения SCORM (Sharable Content Object Reference Model, «образцовая модель объекта содержимого для совместного использования»). Эта модель представляет собой сборник стандартов и спецификаций, содержащих требования к организации учебного материала и всей системе дистанционного обучения. Эта система является полностью удовлетворяет перечисленным выше требованиям, и может служить примером для построения моделей образовательных процессов.

Мониторинг образовательных процессов и построение моделей с обратной связью

При построении метаданных для образовательных процессов следует определить основные проблемы, которые необходимо будет решать с их помощью. Так, для современной образовательной инфраструктуры постоянно растет роль мониторинга, который чаще всего ведется с использованием систем построения отчетов, либо с применением систем бизнес-анализа — приложений, включающие графики, таблицы и отчеты для удобного операционного и стратегического анализа. Такие системы включают все инструменты, связанные с принятием решений на основе фактических данных. Таким образом, туда входят инструменты интеграции и очистки данных (ETL), аналитические хранилища данных и средства Data Mining. BI-технологии позволяют анализировать большие объемы информации, заостряя внимание пользователей на ключевых факторах эффективности, моделируя исход различных вариантов действий, отслеживая результаты принятия тех или иных решений. Поддержка решений фактами всегда являлась основной целью традиционных BI-технологий. Однако во многих случаях это может и не приносить ожидаемых результатов или широкого применения этих результатов в образовательных организациях. В том числе, это может быть вызвано трудностями, связанными с измерением ценности полученных результатов. Ценность выявляется исключительно в момент применения в контексте поддержки принятия решений, как автоматизированного, так и традиционного. Именно действия, выполняемые на основании информации, являются источником дополнительной ценности. В результате проведенного анализа отмечается несогласованность между современной бизнес-аналитикой как

проявлением подхода, ориентированного на данные и процессного подхода в целом. Традиционные BI-технологии, строго говоря, не предназначены для построения процессов и не отслеживают зависимости между процессами и данными, используемыми в этих процессах [4].

Другой характерной особенностью управленческих решений, построенных с использованием BI-компонентов является отсутствие обратной связи. Довольно часто отмечается необходимость своеобразного «повторного использования» информации в рамках организационных процессов принятия решений и определяются сложности построения систем закрытого цикла [5]. В том случае, если собираются необходимые факты, по заданному событию применяется только необходимая информация, а затем результаты принятых решений обогащают исходные факты, возможно достижение более значительной ценности информации по сравнению с использованием традиционных моделей. Задача построения обратной связи носит процессный характер. Именно отсутствие оперативных аспектов применения бизнес-анализа может сдерживать дальнейшую интеграцию информационных систем [5]. Для моделирования таких проблем следует учитывать некоторые особенности управления процессами. Как правило, бизнес-процесс выполняется в течении длительного периода времени, который фактически не ограничен – неделя, месяц, год и более. Особенно важным с точки зрения объединения подходов является то, что бизнес-процесс может оперировать большими объемами данных, потеря которых нежелательна. Эти аспекты необходимо учитывать при проектировании и описании моделей образовательных процессов как метаданных, используемых в системе управления процессами [6].

Исследование проблем организации данных в образовательных процессах

Современные системы предоставляют возможности управлять процессами в разнородных организациях и формировать разные по своему характеру управляющие регламенты, рекомендации и сервисы. Это усложняет использование процессного подхода для построения систем, моделирующих глобальные процессы. Управление информацией на протяжении всего жизненного цикла, включая ее создание, хранение, предоставление доступа, потерю актуальности и архивирование, в гетерогенных системах, которой и является информационно-образовательная среда, до сих пор является сложной задачей [2].

Многие исследователи в данном контексте обращают внимание на рост интереса к технологиям класса Big Data, связанный с постоянным увеличением объемов данных, которыми приходится оперировать в рамках образовательной инфраструктуры. Большие данные обеспечивают прозрачность, связанную с четким пониманием того, какие данные используются для получения результатов и какими алгоритмами, в том числе статистическими, эти данные обрабатываются. Современный опыт внедрения больших данных показывает, что организации, применяющие эти технологии, часто обнаруживают «неожиданные» полезные корреляции в данных различных назначений. Такие результаты невозможны в традиционной закрытой модели бизнес-анализа [6,7].

Наиболее важным с выбранной точки зрения преимуществом больших данных является выход за рамки традиционной бизнес-аналитики с приобретением возможности контролировать и автоматизировать процессы. С их помощью можно получать и анализировать не только результаты процесса, но и сведения о непосредственном ходе процесса. Накопленная таким образом информация для многих организаций является важным активом, содержащим в

том числе и сведения о рабочих операциях [6].

Построение модели образовательного процесса на примере работы с электронными образовательными ресурсами

В качестве примера образовательного процесса, позволяющего и требующего оптимизации с помощью перечисленных методов, можно привести процесс работы с электронными образовательными ресурсами (ЭОР).

Для работы с ними внедряются соответствующие методические рекомендации. Часто рекомендуется инициировать работу учителей по изучению открытых Интернет-ресурсов образовательных порталов, отбор, электронной поддержки к преподаваемым предметам из имеющегося контента, включению ссылок на электронные образовательные ресурсы в разрабатываемое календарно-тематическое планирование.

Кроме вышеназванных категорий преподаватель может подбирать электронные образовательные ресурсы к своему уроку на любых доступных сайтах сети Интернет. Для проведения уроков в дистанционном режиме рекомендуется использовать открытые Интернет-ресурсы, для подготовки к ЕГЭ и ГИА — специализированные обучающие сайты. При этом учитель может выступать в роли эксперта, самостоятельно оценивая найденные им материалы, и использовать на уроке только те из них, которые отвечают основным образовательным потребностям обучаемых, а также содержательно-методическим и дизайн-эргономическим требованиям.

Качество этой работы, а также качество самих рекомендаций оценивается только по внешним параметрам без какой-либо обратной связи, или сведений об этом процессе, что идеально подходит для применения предложенных методов анализа.

Исходные данными для анализа являются деперсонифицированные протоколы доступа к сети, собираемые частично при помощи специальных агентов, установленных на клиентских компьютерах, частично на общих концентраторах через систему мониторинга. Для агентов указывается базовая информация о размещении компьютера и/или проводимых с его помощью занятиях. Базовой единицей информации является соединение — комбинация сведений о HTTP-запросе и ответе на этот запрос. Для каждого соединения собираются следующие параметры:

1. идентификатор школы
2. деперсонифицированный идентификатор агента (при сборе информации с клиентских компьютеров)
3. точное время доступа для привязки соединения определенного агента к конкретной позиции в расписании занятий и факультативов, либо отнесение к рекреационному (неучебному) времени для фильтрации; данные о расписании для выбранных школ были взяты из системы электронных журналов
4. основная информация: адрес соединения, хост, параметры запроса
5. информация технологического характера: время ответа, скорость загрузки данных, использованные прокси и DNS-серверы, используемый провайдер и маршруты доступа, сведения о серверных ошибках и сбоях сети, прочие параметры качества обслуживания.

Эти параметры и формируют описание данных, которые используются в процессе работы с разнородными ЭОС — метаданные. Они описаны в формате XML. Это описание затем включается как ссылка при описании метаданных процесса.

С помощью фильтрующих адаптеров сбора размер журналов несколько уменьшается — отбрасывается малозначимая информация (технологические

соединения, вызванные, например, обновлением программного обеспечения, выполнением сервисных операций и т. п.)

Затем данные поступают в хранилище HDFS для анализа при помощи инструментов обработки больших данных. Для первичной обработки, например, для привязки к системе электронных журналов, анализа параметров запросов с выбором ключевых слов, дополнительной адаптивной фильтрации и объединения большого количества соединений в группы, а также для получения статистических контрольных параметров применяются Apache Spark и Hadoop, а для вывода моделей образовательного процесса использования электронных ресурсов — система ProM. ProM позволяет обобщить и визуализировать процессы в виде графов для удобного первичного анализа.

После этого полученные показатели привязываются к ключевым точкам процесса и сопоставляются с идеальной моделью. Разработан алгоритм сопоставления контрольных параметров и параметров процессов. Наиболее серьезные отклонения фиксируются и выводятся для анализа эксперту с описанием проблемных, либо дополняющих (оптимизирующих) точек (соединений) образовательного процесса.

Таким образом, возможность проведения комбинированного процессного анализа этих данных, собираемых для каждой школы, позволила поставить ряд вопросов, получение ответов на которые и стало целью применения описанной методологии. Совместно они дают представление о том, о качестве доступа к сети Интернет и его практической ценности в образовательных организациях.

1. Какова реальная загрузка оборудования, каналов, время отклика и скорость доступа к ресурсам?

Собираются основные сетевые показатели, которые позволяют выявить узкие места в архитектуре сети или текущей её реализации, а также сверить реальные параметры с заявленными провайдерами или другими

организациями, обслуживающими средства связи.

2. *Какими конкретно электронными образовательными ресурсами пользуются школы и в какой последовательности? Какова доля работы с ресурсами, перечисленными в методических рекомендациях и какова доля «паразитного доступа»?*

При построении процессных цепочек доступа к ресурсам для различных дисциплин, которые и составляют модель образовательного процесса могут быть выявлены новые ресурсы, отобранные преподавателями, другими сотрудниками образовательных организаций, либо самих учащихся. Все они обладают несистематизированным знанием об образовательном процессе и, таким образом, выступают как своего рода «эксперты на месте» — смысловой аналог термина «knowledge worker». Результаты, которые получаются из этого знания, можно сопоставить с идеальным или официальным представлением о процессе, и отражать расхождения в оптимизации методик и регламентов, регулирующих использование электронных образовательных ресурсов. Также можно понять, в каких условиях (в каких цепочках, при каких исходных запросах, на каких дисциплинах) велика вероятность ухода из пространства собственно образовательных ресурсов к сторонним ресурсам побочного характера — из этой информации можно сделать вывод, всегда ли нужно избегать таких ситуаций, и как можно изменить методику, чтобы уменьшить вероятность выхода из образовательного пространства.

3. *Какие ключевые слова используются при поиске информации?*

Например, метод позволяет выяснить, последовательности каких ключевых слов вносятся во время обучения по тем или иным дисциплинам, какие цепочки наиболее распространены, позволяет найти

информационные ресурсы, к которым чаще всего приводит поиск, а также проверить степень соответствия по тематике и качеству ресурсов для каждой из дисциплин. В каждой из них, либо для групп дисциплин строится кластеризованная карта предметной области, на которой размещаются поисковые термины и наиболее популярные ссылки для них, что способствует структурированию информации и развитию централизованных образовательных ресурсов.

4. *Необходимо ли изменять текущие «черные» и «белые» списки доступа там, где они применяются?*

Предоставляются возможности для анализа сведений о том, соединения с какими ресурсами запрещались фильтрами чаще всего и содержат ли цепочки доступа отфильтрованные ресурсы (не прерывались ли они таковыми). Кроме того, возможности по анализу ключевых слов позволяют выявить «нежелательные» кластеры запросов. На основании экспертного анализа перечисленных сведений представляются рекомендации по коррекции списков.

Для ответа на эти вопросы в результате анализа строится реальная модель образовательного процесса использования электронных ресурсов, а затем проводится её сопоставление с идеальной моделью, описанной в методических рекомендациях. Эти модели после дополнительной обработки также описываются в формате XML — таким образом выводятся метаданные процессов. Определяются основные контрольные показатели. Например, для решения сетевых проблем собираются простые контрольные показатели в разрезе по отдельным регионам сети, по образовательным организациям, по образовательным ресурсам. Эти показатели могут быть привязаны как к схеме сети, так и к обобщенным либо частным моделям образовательных процессов для выработки рекомендаций по оптимизации сетевых параметров. Такая

информация может быть сгенерирована специальными алгоритмами, вносящими адаптивные изменения в модели [8-10].

Оптимизированные модели процессов, содержащие ссылки на метаданные, могут быть импортированы в BPMS для проверки и, в случае работы с сетевыми параметрами — контролируемой автоматизации управления сетью, а в случае работы с процессными цепочками доступа — генерации расширенных методических рекомендаций доступа к образовательным ресурсам, уведомлений для образовательных учреждений и запросов на изменение «черных» и «белых» списков в том случае, если они применяются. Применённые рекомендации вводятся в практику, и анализ происходит вновь — таким образом осуществляется описанный ранее принцип обратной связи в управлении образовательным процессом. Таким образом, получен действующий элемент комплексной модели образовательной деятельности. После работ по стандартизации и обобщению эта модель может быть сохранена в централизованном репозитории для дальнейшего использования.

Для визуализации результаты записываются в аналитическую базу, а затем используется система Saiku, которая позволяет строить графики и отчеты в разрезе по организациям, периодам, часто используемым ресурсам и прочим ключевым изменениям.

Результаты применения методологии

В настоящий момент описанные методы обработки и анализа процессов реализуются в системе управления, которая в опытно-режиме применяется в ГУО Псковской области.

Ведется работа по созданию единого комплекса стандартов моделей — метамоделей для различных образовательных процессов.

В случае процесса работы с электронными образовательными ресурсами получены предварительные результаты, которые могут говорить о состоятельности методики — например, средняя скорость доступа к образовательным информационным ресурсам увеличилась на 5%, в региональный портал рекомендованных образовательных ресурсов и дистанционного обучения добавлено около 100 объединенных по стратегиям использования образовательных ресурсов, выявленных данным методом. Продолжается дальнейший сбор результатов для оценки качества оптимизации образовательных процессов.

Список литературы

1. Емельянова В. В., Драгунов А. В. Влияние трансформации управления на использование ИКТ в системе образования // Информатика и образование. 2013. № 8. С. 88–94.
2. Драгунов А. В. О некоторых подходах к быстрой разработке систем облачных приложений // Информатика и образование. 2013. № 7. С. 82–84.
3. Принципы и технологические основы создания открытых информационно-образовательных сред / А.И. Башмаков, В.А. Старых. – М.: БИНОМ. Лаборатория знаний, 2010. С. 8.
4. Barlow M. Real-Time Big Data Analytics: Emerging Architecture. 1st edition. O'Reilly Media, 2013. С. 27.
5. Belikov J., Petelenkov E., Vassiljeva K., Nomm S. Computational Intelligence Methods Based Design of Closed-Loop System // Neural Information Processing / ed. Lee M. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. Vol. 8226. С. 215–224.
6. Hull R., Su J., Vaculin R. Data management perspectives on business process management // Proceedings of SIGMOD '13: ACM SIGMOD International Conference on Management of Data. С. 943.

7. Lim E.-P., Chen H., Chen G. Business Intelligence and Analytics // ACM Transactions on Management Information Systems. ACM, 2013. Vol. 3. № 4. C. 1–10.
8. Reguieg H., Toumani F., Nezhad H.R.M. Using Mapreduce to scale events correlation discovery for business processes. HP Laboratories, 2012. C. 17.
9. Wang J., Crawl D., Altintas I. Kepler + Hadoop // Proc. 4th Work. Work. Support Large-Scale Sci. - Work. '09. New York, New York, USA: ACM Press, 2009. C. 1–8.
10. Nguyen P., Halem M. A MapReduce workflow system for architecting scientific data intensive applications // Proceeding 2nd Int. Work. Softw. Eng. cloud Comput. - SEACLOUD '11. New York, New York, USA: ACM Press, 2011. C. 57.