

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

На правах рукописи

Симановский Андрей Александрович

СОГЛАСОВАННЫЕ ЭВОЛЮЦИОННЫЕ ТРАНСФОРМАЦИИ
ВЗАИМОЗАВИСИМЫХ СЛАБОСТРУКТУРИРОВАННЫХ И
РЕЛЯЦИОННЫХ СХЕМ

05.13.11 — Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Санкт-Петербург
2007

Работа выполнена на кафедре системного программирования
математико-механического факультета Санкт-Петербургского
государственного университета.

Научный руководитель: доктор физико-математических наук,
профессор Новиков Борис Асенович

Официальные оппоненты: доктор технических наук,
Кузнецов Сергей Дмитриевич

кандидат физико-математических наук,
Новиков Федор Александрович

Ведущая организация: ГОУ Южно-Уральский государственный
университет

Защита диссертации состоится “___” _____ 200_ года в __ часов на
заседании диссертационного совета Д212.232.51 по защите диссертаций
на соискание ученой степени доктора наук при Санкт-Петербургском
государственном университете по адресу: 198504, Санкт-Петербург,
Старый Петергоф, Университетский пр., д. 28, математико-механический
факультет Санкт-Петербургского государственного университета.

С диссертацией можно ознакомиться в Научной библиотеке Санкт-
Петербургского государственного университета по адресу: 199034, Санкт-
Петербург, Университетская наб., д. 7/9.

Автореферат разослан “___” _____ 200_ года.

Ученый секретарь
диссертационного совета
доктор физико-математических наук,
профессор

Б. К. Мартыненко

Общая характеристика работы

Актуальность темы

Ряд современных приложений, работающих с базами данных, таких, например, как электронные системы учета медицинской информации, требуют от уровня представления данных возможности модификации (изменения) со временем хранимых данных и их схем. Процесс изменения схемы данных в подобных системах является неотъемлемой частью жизненного цикла приложения. Таким образом, поддержка модификаций схемы хранимых данных со временем является актуальной задачей для производителей СУБД.

Существует несколько подходов к решению этой задачи в рамках СУБД. Они имеют различные области применимости, предоставляют неодинаковый набор возможностей. Эволюция схем [18] обеспечивает высокую производительность системы за счет меньшего объема предоставляемых возможностей. Системы поддержки эволюции схем упорядочивают действия при изменении схем, позволяют описывать семантику изменений в предметной области и отражать изменения предметной области, происходящие со временем, в базе данных, распространяя эти изменения на схему хранимых данных и сами данные. Они избавляют от необходимости создания и реализации *ad-hoc* алгоритмов по изменению схем и данных при каждой модификации. В настоящее время существуют системы поддержки эволюции схем данных для реляционной [4, 5, 11, 9, 17], объектно-ориентированной [20, 19, 13, 6, 12, 10, 3, 15, 14, 8, 16], XML [21, 1, 7] моделей данных.

Использование XML данных требует от приложений организации эффективного хранения и доступа к XML данным. Одним из решений является сохранение данных в реляционных СУБД. Это позволяет не только организовывать хранение и доступ к данным, но и использовать такие отсутствующие в XML, но полезные для приложения механизмы, как индексы, транзакции, многопользовательский доступ и т.д. Также немаловажным фактором, влияющим на принятие такой технологии,

является отсутствие на сегодняшний день реализации XML базы данных, удовлетворяющей требованиям широкого спектра приложений. Рассматриваемые приложения часто используют XML и такие языки запросов как XQuery в качестве интерфейса к уровню представления данных, который в свою очередь использует мощные промышленные реляционные СУБД в качестве хранилища данных, реализуя логику преобразования работы с моделью данных XML в работу с реляционной моделью [22]. Как следствие, возникает вопрос эволюции связанных отображением пары XML и реляционной схем.

Ранее предлагаемые решения в рамках эволюции схем не рассматривали необходимость совместной эволюции схем данных, использующих различные модели. Недавно возник интерес к подобным совместным изменениям схем и отображений между ними, называемым также адаптацией схем, как к проблеме управления моделями данных [2], но вопрос исследуется только в рамках интеграции схем, подхода, хотя и наиболее общего для изменяющихся схем, но в то же время, и наиболее “тяжеловесного”, приводящего к значительным накладным расходам при работе приложения. Предлагаемое решение использует эволюцию схем для решения поставленной задачи изменения пары связанных отображением XML- и реляционной схем, решая задачу адаптации в частном случае, но более эффективным методом.

Цели работы

Данная работа исследует применимость эволюции схем для описания эволюционных трансформаций взаимозависимых слабоструктурированных и реляционных схем. Цели, преследуемые работой, включают:

- Построение модели эволюции, описывающей изменение взаимозависимых слабоструктурированных и реляционных схем данных.

- Разработка множества элементарных операций, позволяющих описывать трансформации взаимозависимых схем и и связывающих их отображений.
- Описание интеграционных трансформаций схем в рамках эволюционных преобразований взаимозависимых схем.

Основные результаты

В работе получены следующие основные результаты:

- Разработана аксиоматическая модель эволюции XML- и реляционной схем, описывающих XML-документ и его представление, хранимое в реляционной базе. Предложенная модель позволяет выделить общую и независимые части обеих схем. Модель позволяет описывать семантические инварианты схем данных и сохранять их при эволюционных трансформациях схем.
- Разработана классификация элементарных преобразований, описывающих трансформации связанных схем. Набор преобразований включает как преобразования характерные исключительно для схем XML-документов или схем реляционных представлений документов, так и преобразования, совместно меняющие сразу обе схемы. Подобная классификация позволяет не только производить произвольные эволюционные трансформации связанных схем, но и вносить ограниченные изменения только в одну из схем, например, с целью оптимизации работы системы.
- Предложен метод декларативного описания трансформаций схем. Метод позволяет строить возможные цепочки элементарных преобразований и ранжировать различные пути трансформации по предложенным в работе критериям, позволяя, например, при внесении изменений в одну из схем, эффективно определить необходимость изменений в связанной схеме.

- Реализован прототип системы, позволяющий применять изложенные в работе методы построения эволюционных преобразований схем на практике.

Научная новизна

В данной работе впервые предложена модель эволюции взаимозависимых слабоструктурированных и реляционных схем. Остальные полученные результаты также являются новыми и дополняют результаты предшествующих работ.

Практическая и теоретическая ценность

С теоретической точки зрения, в работе предложена простая аксиоматическая модель, позволяющая описывать эволюционные трансформации взаимозависимых слабоструктурированных и реляционных схем и связывающих их отображений. Кроме того, в работе исследован вопрос реализации автоматизированного поиска интерграционных трансформаций схем.

Практическая ценность работы состоит в том, что предложенная модель может быть использована для автоматизации эволюционных преобразований существующих XML-реляционных систем, например, с помощью предлагаемого прототипа.

Апробация работы

Результаты работы докладывались

- на Второй Всероссийской конференции “Методы и Средства Обработки информации” (Москва, октябрь 2005)
- на Девятой Восточно-Европейской конференции “Advances in Databases and Information Systems” (Таллин, Эстония, сентябрь 2005)
- на Шестой конференции “Baltic Conference on Databases and Information Systems” (Рига, Латвия, июнь 2004)

- на Первом коллоквиуме “Spring Colloquium for Young Researchers in Databases and Information Systems (SYRCoDIS)” (Санкт-Петербург, май 2004)
- на семинарах группы теории баз данных при лаборатории исследования операций НИИММ

Публикации

Основные результаты представлены в шести работах автора, перечисленных в прилагающемся списке работ автора.

Структура и объем диссертации

Диссертация состоит из пяти глав, заключения и списка литературы. Основной текст диссертации занимает 78 страниц машинописного текста. Библиография содержит 79 наименований. Общий объем диссертации 84 страницы. Рисунки и таблицы нумеруются по главам.

Содержание работы

Первая глава содержит обзор исследуемого вопроса, в том числе обсуждает понятия связанные со схемами данных, темпоральными базами данных, эволюцией схем, XML-реляционными отображениями и адаптацией схем. В главе также дан обзор существующих систем поддержки эволюции схем для реляционных, объектных, XML моделей данных и отображений между XML и реляционными данными. В главе вводится понятие семантической интерпретации XML документа с точки зрения запросов.

Во **второй главе** предложена аксиоматическая система поддержки эволюции пары взаимосвязанных XML- и реляционной схем. Основной идеей системы является выделение промежуточной схемы-медиатора реляционных и XML-данных, являющейся общей частью схем.

Система описывает отображение исходных схем в схему-медиатор и распространение изменений схемы-медиатора в исходные схемы.

Первый раздел второй главы посвящен представлению данных в системе. Второй раздел второй главы содержит определение схемы медиатора. В разделе вводятся понятия графа XML-схемы и факторизованного графа XML-схемы и схемы-медиатора. Отображение XML-схемы в схему-медиатор является композицией отображения, получающего граф XML-схемы по описанию XML-схемы и факторизации графа XML-схемы.

Третий раздел второй главы обсуждает отображение реляционной схемы в схему-медиатор.

Четвертый раздел второй главы вводит аксиомы системы и обсуждает связь сущностей схемы-медиатора и объектов предметной области приложения. Данная связь выражается через ассоциирование с каждой вершиной схемы-медиатора набора множеств. Аксиомы схемы-медиатора являются соотношениями между этими множествами. Выполнение аксиом гарантирует согласованность схемы-медиатора с предметной областью приложения.

Утверждение 1. Пусть абстрактная интерпретация — это заданное на множестве интерпретаций XML-схемы отображение во множество отображений $\Theta : I^* \times DTD^* \rightarrow (V \rightarrow V^*)$, $\Theta(DTD, I)(v) = \{u \in G_{DTD.V} \mid I(v) \equiv h(u) \wedge h \neq const\}$. Тогда на множестве канонических XML-схем система согласованных с синтаксической интерпретацией приложения множеств P_e и N_e полна относительно абстрактной интерпретации приложения.

В **третьей главе** приведена таксономия элементарных операций предлагаемой системы.

Преобразования схемы-медиатора представляют собой набор операций над графом с помеченными атрибутами вершинами, сохраняющих аксиомы. Операции над схемой-медиатором включают в себя добавление и удаление атрибута, добавление и удаление дуги, добавление и удаление вершины, а также слияние двух вершин и расщепление вершины.

Операции над схемой-медиатором позволяют из некоторой исходной схемы-медиатора получать любую наперед заданную схему-медиатор, имеющую данные множества значимых предшественников, последователей и атрибутов.

Утверждение 2. *Введенные операции позволяют из некоторой исходной схемы-медиатора получать любую наперед заданную схему-медиатор, имеющую данные множества значимых предшественников, последователей и атрибутов.*

Второй раздел третьей главы рассматривает преобразования XML-схемы. Ряд операций, изменяющих граф XML-схемы, удобнее ввести, оперируя графом XML-схемы. Остальные операции не меняют граф XML-схемы и выражаются исключительно в терминах DTD. Набор операций включает перемещение вершины вниз/вверх по иерархии, добавление и удаление “листовой” вершины, добавление и удаление верхней дуги, перемещение дуги вверх/вниз, маркирование/демаркирование дуги.

Утверждение 3. *Введенные операции, изменяющие граф XML-схемы, позволяют получать произвольные графы XML-схемы, соответствующие данной схеме-медиатору.*

Утверждение 4. *Совокупность введенных над XML-схемой операций, затрагивающих только XML-схему, позволяет описать, т.е. получить, имея некоторого представителя, класс всех XML-схем, соответствующих данной схеме-медиатору.*

Четвертый раздел третьей главы посвящен операциям, преобразующим реляционную схему. Набор операций включает добавление и удаление дублирующего столбца из таблицы.

Утверждение 5. *Операции добавления и удаления дублирующего атрибута из реляционной переменной позволяют преобразовать любую исходную реляционную схему, таблицы которой имеют столбцы с именами, образующими фиксированное множество, и которой соответствует фиксированная схема-медиатор, в данную целевую*

реляционную схему, таблицы которой имеют столбцы с именами, образующими то же множество, и которой соответствует та же схема-медиатор, при условии, что в обоих преобразованиях реляционных схем в схему-медиатор каждый столбец использован в определении хотя бы одного представления.

Четвертая глава исследует вопрос трансформации схем в рамках предложенной модели эволюции.

В первом разделе главы вводятся функция эволюции и граф эволюции. Трансформация схемы может быть представлена как множество путей в графе эволюции, исходящих из данной начальной схемы S_0 и заканчивающихся в схемах, удовлетворяющих условиям R .

Второй раздел четвертой главы посвящен эвристике, позволяющей сделать задачу поиска на графе эволюции конечной и ранжированию путей эволюции.

В **пятой главе** приводится описание прототипа, реализующего предложенную модель эволюции. Существующие программные средства, осуществляющие поддержку эволюции схем, не позволяют работать с парой взаимосвязанных схем. Прототип состоит из двух частей — SQL-скрипта, порождающего в реляционной базе данных, осуществляющей хранение данных, необходимые для системы поддержки эволюции таблицы и хранимые процедуры, и веб-приложения, реализующего интерфейс, обеспечивающий визуализацию и возможность изменять схемы.

Заключение содержит список основных результатов, полученных в работе.

Список литературы

- [1] *Кукс С. В.* Аксиоматизация эволюции схемы xml-баз данных // *Программирование.* — 2003. — Т. 29, № 3. — С. 140–146.
- [2] *Bernstein P. A.* Applying model management to classical meta data problems. // *CIDR.* — 2003.
- [3] *Chen J.-L., McLeod D.* Schema evolution for object-based accounting database systems. // *ISOOMS.* — 1994. — Pp. 40–52.
- [4] *Clifford J., Croker A.* The historical relational data model (hrdm) and algebra based on lifespans // *Proceedings of the Third International Conference on Data Engineering.* — Washington, DC, USA: IEEE Computer Society, 1987. — Pp. 528–537.
- [5] *Dadam P., Teuhola J.* Managing schema versions in a time-versioned non-first-normal-form relational database. // *BTW.* — 1987. — Pp. 161–179.
- [6] *Kim W., Chou H.-T.* Versions of schema for object-oriented databases. // *VLDB.* — 1988. — Pp. 148–159.
- [7] *Leonardi E., Bhowmick S. S.* Detecting changes on unordered xml documents using relational databases: a schema-conscious approach. // *CIKM.* — 2005. — Pp. 509–516.
- [8] *Lerner B. S.* A model for compound type changes encountered in schema evolution // *ACM Trans. Database Syst.* — 2000. — Vol. 25, no. 1. — Pp. 83–127.
- [9] *McKenzie L. E., Snodgrass R. T.* Evaluation of relational algebras incorporating the time dimension in databases // *ACM Comput. Surv.* — 1991. — Vol. 23, no. 4. — Pp. 501–543.
- [10] *Monk S., Sommerville I.* Schema evolution in oodbs using class versioning // *SIGMOD Rec.* — 1993. — Vol. 22, no. 3. — Pp. 16–22.

- [11] *Narayanaswamy K., Rao K. V. B.* An incremental mechanism for schema evolution in engineering domains // Proceedings of the Fourth International Conference on Data Engineering. — Washington, DC, USA: IEEE Computer Society, 1988. — Pp. 294–301.
- [12] *Osborn S. L.* The role of polymorphism in schema evolution in an object-oriented database // *IEEE Transactions on Knowledge and Data Engineering.* — 1989. — Vol. 1, no. 3. — Pp. 310–317.
- [13] *Penney D. J., Stein J.* Class modification in the gemstone object-oriented dbms // OOPSLA '87: Conference proceedings on Object-oriented programming systems, languages and applications. — New York, NY, USA: ACM Press, 1987. — Pp. 111–117.
- [14] *Peters R. J., Özsu M. T.* An axiomatic model of dynamic schema evolution in objectbase systems // *ACM Trans. Database Syst.* — 1997. — Vol. 22, no. 1. — Pp. 75–114.
- [15] *Ra Y.-G., Rundensteiner E. A.* A transparent object-oriented schema change approach using view evolution. // ICDE. — 1995. — Pp. 165–172.
- [16] *Rashid A., Sawyer P.* Object database evolution using separation of concerns. // *SIGMOD Record.* — 2000. — Vol. 29, no. 4. — Pp. 26–33.
- [17] *Roddick J. F.* Sql/se: a query language extension for databases supporting schema evolution // *SIGMOD Rec.* — 1992. — Vol. 21, no. 3. — Pp. 10–16.
- [18] *Roddick J. F.* A survey of schema versioning issues for database systems // *Information and Software Technology.* — 1995. — Vol. 37, no. 7. — Pp. 383–393. citeseer.ist.psu.edu/roddick95survey.html.
- [19] Semantics and implementation of schema evolution in object-oriented databases. / J. Banerjee, W. Kim, H.-J. Kim, H. F. Korth // SIGMOD Conference. — 1987. — Pp. 311–322.
- [20] *Skarra A. H., Zdonik S. B.* The management of changing types in an object-oriented database. // OOPSLA. — 1986. — Pp. 483–495.

- [21] Xem: Managing the evolution of xml documents. / H. Su, D. Kramer, L. Chen et al. // RIDE-DM. — 2001. — Pp. 103–110.
- [22] Xquery implementation in a relational database system / S. Pal, I. Cseri, O. Seeliger et al. // VLDB '05: Proceedings of the 31st international conference on Very large data bases. — VLDB Endowment, 2005. — Pp. 1175–1186.

Работы автора по теме диссертации

- S. V. Соох, A. A. Simanovsky. Regular Expressions in XML Schema Evolution. *Вісник Національного технічного університету “Харківський політехнічний інститут”*: Збірник наукових праць. Тематичний випуск “Системний аналіз, управління та інформаційні технології” 1(1). — Харьков, Украина, 2004 — с. 24–38.
- A. Simanovsky. Evolution of Schema of XML-documents Stored in a Relational Database. *Proceedings of 6th Baltic DBIS Conference*. — Riga, Latvia, 2004. — с. 192–204.
- A. Simanovsky. Applying the Reconfiguration-Design Formalism to XML Stored in a Relational Database. *Proceedings of the Spring Young Researcher’s Colloquium on Database and Information Systems*. — Санкт-Петербург, 2004. — с. 75–77.
- A. Simanovsky. Three Layer Evolution Model for XML Stored in Relational Databases. *Proceedings of the Ninth East-European Conference on Advances in Databases and Information Systems*. — Tallinn, Estonia, 2005. — с. 66–79.
- А. А. Симановский. Автоматизация разработки схем баз данных для XML-реляционных систем. *Методы и средства обработки информации. Труды Второй Всероссийской научно-технической конференции*. — Москва, 2005. — с. 113–118.

- А. А. Симановский. Совместные эволюционные трансформации взаимозависимых XML и реляционных схем. Вестник Санкт-Петербургского университета. Серия 10. 2007. Выпуск 1 — с. 102–111