
УДК 044.03

М.А. Керносов

Харьковский национальный университет радиоэлектроники, Харьков

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ЗАПРОСА К КОНТЕНТУ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ

В работе проведен анализ существующих моделей запроса к различным элементам контента информационно-аналитической системы, которые отличаются формами представления и способами их обработки. Обоснована необходимость развития существующих моделей запроса с целью разработки модели, обеспечивающей унифицированное описание структур данных и электронных документов как основных элементов контента информационно-аналитической системы. Предложена математическая модель запроса, позволяющая обеспечить типизацию функций формирования и обработки информационных запросов к контенту системы.

Ключевые слова: запрос, контент, информационно-аналитическая система, модель, структура данных, электронный документ.

Введение и постановка проблемы

Повсеместное внедрение информационно-аналитических систем (ИАС) в бизнес-процессы (БП) современных предприятий обуславливает значительное влияние эффективности работы ИАС в быстро изменяющихся условиях на конкурентоспособность и рентабельность предприятия. Целью внедрения данных систем является максимально быстрое предоставление актуальных данных о состоянии БП предприятия по запросам пользователей.

Особенности ИАС предприятия в настоящее время заключаются в следующем:

- увеличение роли ИАС в БП предприятия как объекта автоматизации (ОА), широкое внедрение электронного документооборота и электронных форм отчетности;
- увеличение требований к гибкости ИАС, способности адаптации к изменениям БП ОА;

- необходимость управления контентом ИАС, вызванная изменениями запросов пользователей, потребностью во взаимодействии ИАС с другими системами;

- распространение интерфейсов интеграции систем различных классов с целью их функционального взаимодействия и обмена данными;

- потребность в данных, содержащихся в электронных документах, при решении функциональных задач (ФЗ) ИАС и, наоборот, потребность в фактографических данных при формировании выходных документов ФЗ системы;

- рост требований к надежности ИАС, актуальности данных в системе и оперативности их получения для принятия управленческих решений.

Вследствие данных тенденций популярными становятся системы с открытой модульной архитектурой, а контент ИАС образуется разнородными элементами, основными из которых являются струк-

туры данных и электронные документы, применяемые для решения отдельных ФЗ [1].

Рассмотренные тенденции и требования к современным ИАС предприятия и функциям управления контентом системы определяют актуальные направления научных исследований в данной области.

Результаты анализа современных тенденций развития ИАС позволяют сформулировать цель работы как разработку модели запроса к ИАС, обеспечивающей повышение эффективности управления контентом ИАС и адаптации таких систем к выполнению изменяющихся запросов пользователей.

Исследование существующих моделей запроса

Модели запроса к контенту ИАС можно разделить на две группы, в зависимости от типа элементов контента, к которым относятся как наборы данных, так и наборы электронных документов.

Первая группа моделей применяется в системах управления базами данных (СУБД) и характеризует запросы, формируемые и обрабатываемые в базах и хранилищах данных (БД и ХД). Запрос (S) к БД описывается следующим выражением:

$$S = \pi(\sigma_F(R_1 \times R_2 \times \dots \times R_n)),$$

где π – оператор селекции, который используется для получения любой требуемой перестановки, проекции или комбинации значений атрибутов результирующего набора данных (в современных СУБД задаётся в виде перечня наименований запрошенных столбцов таблиц БД); σ_F – условие выбора результирующего набора данных, которое соответствует секции «where» в структурированном языке запросов (Structured Query Language – SQL); F – логическое выражение, переданное в запросе (содержимое секции «where» в SQL); R_i – отношение (таблица БД или ХД ИАС), $i = \overline{1, n}$; n – количество таблиц, к которым сформирован запрос [2,3].

Такие запросы применяются в реляционной и объектно-реляционной моделях данных. Электронный документ, хранимый в БД или ХД ИАС, рассматривается как значение одного из атрибутов таблицы БД. Для геоинформационных систем существует расширение SQL функциями для обработки запросов к пространственным данным [4].

Вторая группа моделей запроса, рассчитанных на осуществление контекстного поиска в хранилище документов, применяется в поисковых системах (Google, Yandex и т.д.). Такие запросы используются в моделях поиска, базовыми из которых являются булевская, векторная и вероятностная [5,6].

В булевой модели применительно ко множеству документов осуществляется проверка наличия (или отсутствия) в каждом из них требуемых слов в соответствии с логическим условием, переданным в запросе. Пример запроса (S):

$$S = t_1 \vee (t_2 \wedge t_3),$$

где t_1, t_2, t_3 – слова, формирующие условие полнотекстового поиска.

Документ удовлетворяет запросу, если применительно к нему выполняется переданное в запросе логическое условие. При этом слова, формирующие условие запроса, должны принадлежать множеству словаря (T) поисковой системы, состоящего из n слов: $t \in T = \{t_1, t_2, \dots, t_n\}$ [6].

В векторной и вероятностной моделях поиска учитывается частота встречаемости слова в документе и «обратная встречаемость», т.е. «редкость» термина в хранилище документов. В таких моделях применяются сложные алгоритмы обработки запросов, но модель запроса практически неизменна.

В векторной модели поиска запрос (S) представлен в виде вектора словаря системы (T), содержащего единичные элементы в позициях, соответствующих словам, которые вошли в запрос, и нулевые в остальных случаях. Пример вектора запроса на поиск документа, содержащего слова t_1 и t_3 :

$$S = \{1, 0, 1, 0, 0, \dots, 0\}.$$

В вероятностной модели поиска запрос задаётся в виде множества. Предыдущий пример с применением вероятностной модели примет вид:

$$S = \{t_1, t_3\}.$$

Недостатком данных моделей запроса является то, что они позволяют осуществлять только контекстный поиск документов и не позволяют формировать запросы к наборам данных.

Существующие модели, технологии и методы моделирования запросов к контенту ИАС ориентированы на решение узкоспециализированных задач применительно только к одному виду контента системы. К таким задачам относится моделирование структуры запросов, планов их выполнения или бизнес-логики формирования элементов контента ИАС, взаимодействия ИАС с внешней средой, с пользователями т.д. Поэтому данные модели не в полной мере отражают современные требования к ИАС, что, в свою очередь, обуславливает необходимость их дальнейшего развития с целью проектирования и сопровождения ИАС, обладающих способностью адаптации к изменениям БП ОА и запросов пользователей вне зависимости от способов представления информации в системе. Это позволит обеспечить поддержку жизненного цикла информации и документа в ИАС, её наполнение актуальными данными, а также сократить время разработки ИАС и трудоемкость их сопровождения.

Разработка модели запроса к контенту информационно-аналитической системы

Для реализации функций управления разнородным контентом ИАС необходима унифициро-

ванная модель запроса к контенту системы. Взаимодействие функциональных модулей ИАС между собой и с внешней средой осуществляется посредством обмена структурами данных, включая издание и обработку документов. Во многих случаях официальные служебные документы являются источником новых данных, вносимых в ИАС. При этом выходные данные системы зависят не только от входных, новых данных, но и от набора данных, внесённых в ИАС к моменту выполнения запроса, т.е. от текущего состояния контента системы.

Запросы к контенту ИАС могут формироваться как пользователями системы с использованием комплексов средств автоматизации (КСА) при решении функциональных задач ИАС, так и внешними системами, например, с использованием предоставляемых web-сервисов системы с сервис-ориентированной архитектурой (Service-Oriented Architecture – SOA).

Запрос к контенту ИАС представляет собой совокупность двух составляющих: входных данных, которые должны быть переданы в систему для выполнения запроса, и набора данных, ожидаемых в качестве результата его выполнения. В SOA-системах информация о запросах к сервисам системы представляется в виде описания наборов входных и выходных параметров и публикуется в реестре сервисов системы, который используется модулями распределённой ИАС или внешними системами для формирования запросов к контенту ИАС. Для пользователей функции формирования запросов и обработки полученных результатов выполняет КСА функциональных задач системы.

В обоих случаях параметры запроса к контенту ИАС формализованы и предполагают однозначное преобразование входных параметров в ожидаемый результат. Специфика обработки запроса, алгоритмы обработки входных данных и формирования набора выходных данных при создании запроса пользователя к контенту ИАС не учитываются. Поэтому математическую модель запроса к контенту ИАС целесообразно представить в виде конечного автомата Милли, для которого функция выходов определяется на паре <состояние, входной сигнал>.

Представим модель запроса к контенту ИАС (Z) в обобщенном виде:

$$Z = \langle S, X, Y, s_0, \delta, \lambda \rangle,$$

где S – множество состояний контента ИАС, которые характеризуются набором фактографических данных и документов в момент выполнения запроса; $X = \langle d_{xf}, D_x, \xi \rangle$ – входные документы (D_x) и структуры данных (d_{xf}), переданные вместе с запросом к контенту ИАС; ξ – возмущающее воздействие на автоматизированные бизнес-процессы ОА и контент ИАС, которое приводит к появлению новых форм выходных документов и структур данных или к возникновению ошибок в документах и структурах данных,

формируемых на основе существующих форм; $Y = \langle d_y(R_x, X), D_y(d_{xf}, R_x), D_y(\xi), D_y(d_{xf}, R_x, D_y(\xi)) \rangle$ – выходные документы (D_y) и структуры данных ($d_y(R_x, X)$), полученные в результате выполнения запроса к контенту ИАС; $D_y(d_{xf}, R_x)$ – регулярные документы, являющиеся результатом обработки входных данных (d_{xf}); $D_y(\xi)$ – выходные документы, модифицированные в результате влияния возмущающего воздействия ξ ; $D_y(d_{xf}, R_x, D_y(\xi))$ – документы, издание которых вызвано обнаруженными несоответствиями в ранее сформированных выходных документах; S_0 – начальное состояние контента ИАС в момент внедрения системы; $\delta: R_x \times \langle d_{xf}, D_x, \xi \rangle \rightarrow R_y$ – функции переходов контента ИАС из одного состояния в другое. Реализуются как последовательности команд в информационном и программном обеспечении (ИО и ПО) ИАС; R_x – элементы контента (фактографические данные и документы), формирующие начальное состояние контента ИАС на момент выполнения запроса; R_y – элементы контента (фактографические данные и документы), формирующие конечное состояние контента ИАС в результате выполнения функций переходов, полученные в процессе обработки запросов к контенту ИАС; $\lambda: R_x \times \langle d_{xf}, D_x, \xi \rangle \rightarrow \langle d_y(R_x, X), D_y(d_{xf}, R_x), D_y(\xi), D_y(d_{xf}, R_x, D_y(\xi)) \rangle$ – функции выходов (последовательности команд в ИО и ПО ИАС) [7].

Таким образом, обработка запросов к контенту ИАС – развивающийся во времени процесс. При этом состояния контента зависят от входных данных и от всей предыстории функционирования ИАС, а модель запроса к контенту ИАС имеет вид:

$$Z = \langle S, \langle d_{xf}, D_x, \xi \rangle, \langle d_y(R_x, X), D_y(d_{xf}, R_x), D_y(\xi), D_y(d_{xf}, R_x, D_y(\xi)) \rangle, s_0, R_x \times \langle d_{xf}, D_x, \xi \rangle \rightarrow R_y, R_x \times \langle d_{xf}, D_x, \xi \rangle \rightarrow \langle d_y(R_x, X), D_y(d_{xf}, R_x), D_y(\xi), D_y(d_{xf}, R_x, D_y(\xi)) \rangle \rangle$$

Во множествах функций переходов и выходов, выполняемых во время обработки запросов, можно выделить два подмножества функций работы с документами:

а) $CF(R_x, d_{xf}, \xi) \in \lambda$ – множество функций, обрабатывающих входные данные (d_{xf}) и формирующих на их основе выходные документы (D_y):

$$CF: R_x \times \langle d_{xf}, \xi \rangle \rightarrow D_y;$$

б) $RF(R_x, D_x, \xi) \in \delta$ – множество функций обработки входных документов, преобразующих документы в данные, вносимые в ИАС:

$$RF: R_x \times \langle D_x, \xi \rangle \rightarrow R_y.$$

Реализация функций обработки данных и документов, необходимых для выполнения запросов к контенту ИАС, может быть описана с помощью теории графов, теории алгоритмов или другими способами. Физически эти функции могут быть реализованы в виде методов, обрабатывающих данные, внесенные пользователем в интерфейс ПО ИАС, и наборов операторов языка манипулирования данными в составе ИО системы и т.д. [7].

Из приведенной модели запроса к контенту ИАС видно, что данные $(d_{x\beta}, R_x)$, вносимые в документы, впоследствии попадают в ИАС и изменяют её состояние. Такая модель выдвигает требования к достоверности данных, вносимых в документы, качеству формирования, заполнения и обработки документов, поскольку в результате воздействия случайных факторов ξ , выход системы может не совпадать с ожидаемым результатом. В идеальной ситуации выходные данные R_y и документы D_y не содержат некорректных сведений и ошибок. Но из-за наличия случайной составляющей каждое из множеств выходных данных и документов, полученных в результате обработки запросов к контенту ИАС, будет представлено в виде двух подмножеств:

$$R_y = R_a \cup R_\xi, \quad D_y = D_a \cup D_\xi,$$

где R_a – подмножество корректных (актуальных) данных, соответствующих изменениям в БП; R_ξ – подмножество некорректных данных, снижающих степень достоверности и актуальности данных в ИАС; D_a – подмножество корректных документов; D_ξ – подмножество документов, содержащих некорректные данные [7].

Выводы

При разработке ИАС документ «Описание постановки задачи» регламентирует только типовые запросы к системе, необходимые результаты их выполнения и требования к обработке. Предложенная модель запроса к контенту ИАС основана на представлении электронного документа как структуры данных, которая может как передаваться в ИАС в качестве пара-

метра запроса, так и возвращаться пользователю в качестве результата его выполнения. Она позволяет упростить адаптацию системы к изменениям автоматизируемых БП, сопровождаемых потоками данных и документов, за счёт управления функциями переходов и выходов контента системы. Кроме того, после сбора статистических данных, отражающих частоту и особенности типовых запросов к контенту ИАС, а также влияние случайного возмущающего воздействия ξ на работу системы, возможно формирование решения по управлению обработкой запросов к контенту ИАС.

Предложенная в работе модель запроса к контенту ИАС позволяет снизить временные затраты на выполнение запросов за счет типизации функций управления данными и документами как основными информационными единицами контента системы.

Список литературы

1. Кернос М.А. Система электронного документооборота как элемент информационной системы / М.А. Кернос, В.М. Левыкин // Информатика и компьютерные технологии 2006 : 2-я Межд. студ. науч.-техн. конф., 13 декабря 2006 : сб. тр. – Донецк: ДонНТУ, 2006. – С. 353-354.
2. Codd E.F. A Relationship Model of Data for Large Shared Data Banks / E.F. Codd // Communications of the ACM. – 1970. – Vol. 13. – №6. – P. 377-387.
3. Пашинин О.В. Оптимизация запросов к базам данных / О.В. Пашинин // Математические структуры и моделирование. – 2007. – №17. – С. 100-107.
4. Шаши Ш. Основы пространственных баз данных / Ш.Шаши, Ч. Санжэй. – М. КУДИЦ-ОБРАЗ, 2004. – 336 с.
5. Brin S. The anatomy of a large-scale hypertextual Web search engine / S. Brin, L. Page // Computer Networks and ISDN Systems. – 1998. – Vol. 30. – №1-7. – P. 107-117.
6. Сегалович И. Как работают поисковые системы / И. Сегалович // Мир Интернет. – 2008. – № 10 (73). – С. 24-32.
7. Левыкин В.М. Задача моделирования системы электронного документооборота как компонента информационной системы / В.М. Левыкин, М.В. Евланов, М.А. Кернос // Нові технології. – 2007. – № 1-2 (15-16). – С. 65-69.

Поступила в редколлегию 12.04.2013

Рецензент: д-р техн. наук, проф. В.В. Бескорвайный, Харьковский национальный университет радиоэлектроники, Харьков.

МАТЕМАТИЧНА МОДЕЛЬ ЗАПИТУ ДО КОНТЕНТУ ІНФОРМАЦІЙНО-АНАЛІТИЧНОЇ СИСТЕМИ

М.А. Кернос

В роботі проведено аналіз існуючих моделей запиту до різних елементів контенту інформаційно-аналітичної системи, які відрізняються формами подання та способами їх обробки. Обґрунтовано необхідність розвитку існуючих моделей запиту з метою розробки моделі, що забезпечує уніфікований опис структур даних і електронних документів як основних елементів контенту інформаційно-аналітичної системи. Запропоновано математичну модель запиту, що дозволяє забезпечити типізацію функцій формування і обробки інформаційних запитів до контенту системи.

Ключові слова: запит, контент, інформаційно-аналітична система, модель, структура даних, електронний документ.

MATHEMATICAL MODEL OF QUERY TO CONTENT OF INFORMATION-ANALYTIC SYSTEM

M.A. Kernosov

The paper analyzes the existing models of the various elements of the query to information-analytic systems content that has different forms of representation and methods of processing. The necessity of the existing models improvement in order to develop a model that provides a unified description of the data structures and electronic documents as the main content items of information-analytical system is proved. A mathematical model of query that provides typing functions of forming and processing of information requests to the content of the system is suggested.

Keywords: query, content, information-analytic system, model, data structure, electronic document.