

КИБЕРНЕТИЧЕСКИЙ ПОДХОД К РЕШЕНИЮ ЗАДАЧИ РЕИНЖИНИРИНГА БАЗ ДАННЫХ

Введение

Результаты исследований состояния информатизации в различных компаниях, организациях, учреждениях свидетельствуют о том, что в настоящее время многие из них владеют определенными информационными системами организационного управления (ИСОУ). При этом для решения возникающих новых задач, связанных, как правило, с расширением деятельности, и соответственно, рассматриваемыми предметными областями (ПрО), они хотят иметь более функциональные, с улучшенными характеристиками качества информационные системы (ИС), требующие меньших затрат по сопровождению. В этих условиях востребованными становятся проекты: по разработке новых ИСОУ и их интеграции с существующими информационными системами; разработке новых ИСОУ с целью замены существующих ИС; модернизации существующих ИСОУ. Суть данных проектов заключается в проведении процедур реинжиниринга существующих ИС и их основного функционального компонента: базы данных (БД).

Одним из важных требований, предъявляемых к процессу реинжиниринга существующих ИСОУ и их БД, является своевременность завершения соответствующих проектов в рамках запланированного бюджета с заданными характеристиками качества. Критичность этого условия наглядно демонстрируют результаты анализа ИТ-проектов, который провели компетентные международные организации экспертов [1, 2]. Вывод из этого анализа не оптимистичный: более 60 % проектов были провалены или завершены с опозданием, причем с гораздо большими затратами, чем планировалось.

Налицо существование нерешенной проблемы, связанной с необходимостью своевременного создания, модернизации в рамках запланированного бюджета информационных систем, обладающих требуемыми качествами, и ограниченностью возможностей существующих методов проектирования. В отношении баз данных указанная ограниченность возможностей обусловлена ориентацией традиционной методологии их проектирования, используемой при реинжиниринге БД ИСОУ, на итерационную, достаточно сложную и трудоемкую процедуру создания уникальных концептуальной модели, логической и физической схем при разработке новой БД, либо на существенное их преобразование при модернизации.

В сложившейся ситуации возникает объективная потребность в пересмотре существующих подходов, методологий и технологий реинжиниринга баз данных и оценке целесообразности применения в проектах той или иной технологии.

Чтобы оценить возможность и целесообразность применения в проектах той или иной технологии, в первую очередь традиционной технологии проектирования реляционных баз данных (РБД), получивших наибольшее распространение в ИС рассматриваемого класса и удовлетворяющих требованиям к типу источника данных создаваемой системы, использовали кибернетический подход для таких исследований, в результате чего:

– определена модель как множество взаимосвязанных характеристик, образующих базис для спецификации требований к качеству и оценивания качества подвергающейся систематической трансформации БД ИСОУ, с соответствующими метриками качества (за основу в качестве прототипов были приняты характеристики и метрики, определенные в действующем в Украине стандарте ДСТУ ISO/IEC 9126, а также в серии международных стандартов SQuaRE – ISO/IEC 250xx), которые в совокупности можно представить в виде формализованного выражения:

$$Q_{DB} = \{H_i^{DB}, S_{ij}^{DB}, M_{jk}^{DB(i)}, At_{jl}^{DB(i)}\}, \quad (1)$$

где H_i^{DB} – i -я характеристика качества БД ($i=1, \dots, I$); S_{ij}^{DB} – j -я подхарактеристика ($j=1, \dots, J$) i -й характеристики качества; $M_{jk}^{DB(i)}$ – k -я метрика ($k=1, \dots, K$) j -й подхарактеристики i -й характеристики качества; $At_{jl}^{DB(i)}$ – l -й атрибут ($l=1, \dots, L$) j -й подхарактеристики i -й характеристики качества – переменная, которой присваивается значение в результате измерения (применения метрики), $At_{jl}^{DB(i)} \in Z$; $Z = (z_1, \dots, z_\Theta)$, например для соответствующих H_i^{DB} , S_{ij}^{DB} атрибутами качества являются: полнота, корректность реализации объектов схемы БД, способность к взаимодействию, оперативность устранения некорректных данных в БД, адаптация объектов схемы БД, непрерывность использования данных БД, среднее время отклика на запрос, среднее время, затрачиваемое на модификацию, и другие);

– формализована задача реинжиниринга БД ИСОУ и предложен способ ее решения.

Постановка задачи

Представим задачу реинжиниринга БД ИСОУ (под которой в дальнейшем будем понимать реляционную (объектно-реляционную) БД) в виде известной из общей теории адаптивных систем задачи структурной адаптации [3].

В условиях динамических изменений предметных областей в рамках запланированного бюджета необходимо осуществить своевременную реструктуризацию некоторой рассматриваемой реляционной БД ИСОУ с целью максимизации (минимизации) выбранных атрибутов (показателей) качества этой БД при обязательном обеспечении их требуемых значений.

Формализация задачи

Формализацию задачи структурной адаптации РБД ИСОУ начнем с представления схемы (рис. 1) системы управления объектом в условиях динамических изменений предметных областей (схемы адаптации РБД ИСОУ к изменениям условий функционирования).

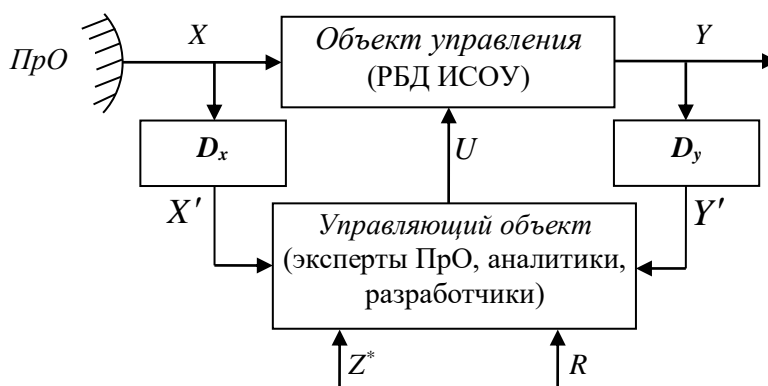


Рис. 1. Схема системы управления объектом

Каждый из приведенных на рис. 1 элементов представляет собой следующие компоненты системы управления и набор ограничительных требований:

– объект управления (ОУ) – РБД ИСОУ, которую необходимо адаптировать к изменениям условий функционирования;

– управляющий объект – задействованные в процессе реинжиниринга РБД ИСОУ различного профиля исполнители: эксперты ПрО, аналитики, разработчики, формирующие управляющие воздействия на основе анализа входной, выходной информации, используя для этого доступные модели, методы и средства;

– X – множество изменений в предметной области;

– Y – состояние сложного объекта управления – базы данных ИСОУ: $Y = F(X, U)$, где F – оператор связи входа X и выхода Y с учетом фактора управляющего воздействия U ;

– U – множество доступных моделей, методов и средств, используемых при реинжиниринге РБД ИСОУ;

– X' – информация об изменениях в ПрО, получаемая управляющим объектом с помощью датчика D_x ; Y' – информация о состоянии ОУ: схеме, оценках значений атрибутов качества РБД ИСОУ, получаемая управляющим объектом с помощью датчика D_y . Датчики измеряют (выделяют) только то, что используется (необходимо) в процессе управления. Поэтому получаемая из них информация не в полной мере отражает действительное состояние ОУ и реальные изменения в ПрО: $X' = D_x(X)$; $Y' = D_y(Y)$.

Датчиками D_x , D_y являются соответствующие инструментальные средства (CASE-средства), обеспечивающие процедуру оценки входных и выходных данных, необходимую экспертам ПрО, аналитикам, разработчикам для принятия решений в определении эффективных управляющих воздействий;

– Z^* (цель управления) – множество требований, предъявляемых к состоянию наблюдаемого объекта управления, которые связываются с Θ -мерным пространством значимых для адаптируемой РБД ИСОУ совокупности свойств (атрибутов), обеспечивающих ее способность удовлетворять установленные или предполагаемые потребности в соответствии с назначением: $Z^* = (z_1^*, \dots, z_\Theta^*)$, где z_i^* – требование к значению атрибута качества – $z_i \in Z$, $Z = (z_1, \dots, z_\Theta)$ (под z_i^* может пониматься также требование к одновременному выполнению совокупности значений атрибутов качеств). Цели-требования z_i^* ($i = 1, \dots, \Theta$) вектора Z^* могут быть представлены или при необходимости преобразованы к виду следующих форм (критериев):

а) равенств (для z_i^*): $z_i = a_i$ – это требование означает, что значение i -й целевой переменной $z_i = \varphi_i(Y')$ должно быть равно заданной величине a_i ;

б) неравенств (для z_j^*): $z_j \leq b_j$ ($z_j \geq b_j$) – это требование, накладываемое на j -ю целевую переменную, – ее значение не должно быть меньше (больше) заданного порога b_j . При необходимости любое ограничение вида $z_j \leq b_j$ можно привести к эквивалентному ограничению вида $-z_j \geq -b_j$, и наоборот (как и любое двустороннее ограничение можно привести к двум односторонним);

г) максимизации/минимизации (для z_v^*): $z_v \rightarrow \max$ ($z_v \rightarrow \min$) – это требование означает, что целевая переменная z_v вектора Z должна быть максимальной (минимальной). В рассматриваемой задаче следует максимизировать или минимизировать (в зависимости от предназначения) определенные атрибуты качества, при обязательном выполнении, предъявляемых к ним же ограничений в виде равенств и неравенств. При этом переход от задачи максимизации к задаче минимизации осуществляется путем изменения знаков коэффициентов целевой функции на противоположный.

С учетом сказанного вектор Z^* имеет следующее представление:

$$Z^* : \begin{cases} z_i^* : z_i = \varphi_i(Y') = a_i, & (i = 1, \dots, l_1); \\ z_j^* : z_j = \psi_j(Y') \leq b_j, & (j = 1, \dots, l_2); \\ z_v^* : z_v = \eta_v(Y') \rightarrow \text{extr}, & (v = 1, \dots, l_3), \end{cases} \quad (2)$$

где $l_1 + l_2 + l_3 = \Theta$; φ_i , ψ_j , η_v – некоторые функционалы, определяющие связь между состоянием наблюдаемого объекта управления и соответствующим параметром (атрибутом качества): z_i , z_j , z_v .

– R – временной, финансовый, людской ресурсы ($R = (r_t, r_f, r_p)$), выделяемые на реинжиниринг РБД ИСОУ.

Решение задачи

Решение задачи реинжиниринга РБД ИСОУ (адаптации БД к изменениям условий функционирования) заключается в нахождении на основании полученной информации об изменениях в ПрО (X'), состоянии ОУ (Y'), цели управления Z^* и ограниченных выделяемых ресурсов R такого управляющего воздействия U^* , с помощью которого стал бы возможным перевод ОУ в искомое состояние Y^* :

$$\langle X', Y', Z^*, R \rangle \rightarrow U^* \rightarrow Y^* \quad (3)$$

Для нахождения U^* необходимо определить оператор (алгоритм) A , преобразующий исходную информацию в управление:

$$U = A(X', Y', Z, R). \quad (4)$$

Задачу синтеза оператора A обычно декомпозируют на две [3]:

1) синтез модели F объекта управления:

$$Y' = F(X', U); \quad (5)$$

2) синтез управления с помощью этой модели.

Синтез модели

В общем случае оператор F определяется некоторым алгоритмом (инструкцией, правилом), который указывает, как, располагая информацией об X' и U , определить выход Y' .

В рассматриваемой задаче в качестве основы такого алгоритма целесообразно взять последовательность основных этапов классической методологии проектирования реляционных баз данных (концептуального, логического и физического проектирования). Тогда оператор F можно представить в виде FEO-диаграммы (*For Exposition Only*) последовательности технологических операций с уточнением содержания их основных этапов и применяемых разработчиками соответствующих CASE-средств (рис. 2).

Действия по внесению требуемых изменений в концептуальную модель ПрО, логическую и физическую схемы БД ИСОУ осуществляются соответствующими разработчиками, экспертами ПрО с помощью доступных инструментальных средств ($U_4 \in U$) в соответствии с существующими правилами описания, структурирования данных и видами ограничений их целостности, принятыми в «расширенных» (термин, введенный

Дейтом, для обозначения моделей, используемых при семантическом моделировании [4]) – $U_1 \in U$ и реляционной – $U_2 \in U$ моделях данных, а также правилами и процедурами разработки объектов физической схемы БД ИСОУ ($U_3 \in U$).

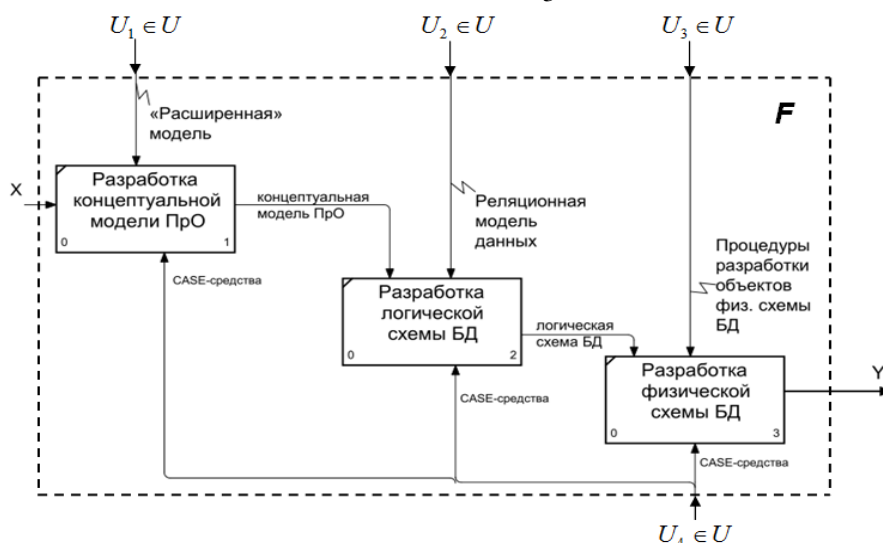


Рис. 2. Схема оператора F

В результате применения оператора F к входной информации и U ($U = U_1 \cup U_2 \cup U_3 \cup U_4$), как правило, изменяются структуры (структурные компоненты):

– концептуальной модели ПрО:

$$S_{sd} = \langle Ent, Rel \rangle, \quad (6)$$

где $Ent = \{Ent_1, \dots, Ent_n\}$ – множество сущностей рассматриваемой ПрО, Ent_n – n -я сущность: $Ent_n = (e_{n1}, \dots, e_{nm})$, e_{nm} – m -е свойство n -й сущности; $Rel = \{Rel_1, \dots, Rel_i\}$ – множество связей между сущностями (и их роли), Rel_i – i -я связь: $Rel_i = rel_{kj}$, где rel_{kj} – связь между k -й и j -й сущностями, которая тоже может описываться набором свойств;

– логической схемы БД:

$$S_{ls} = \langle O, H \rangle, \quad (7)$$

где $O = \{O_1, \dots, O_g\}$ – множество отношений, O_g – g -е отношение: $O_g(o_{g1}, \dots, o_{g\lambda})$, $o_{g\lambda}$ – λ -й атрибут g -го отношения; $H = \{H_1, \dots, H_\alpha\}$ – множество ключей, в том числе определяющих связи между отношениями (внешних ключей);

– физической схемы БД ИСОУ как собственно структуры управляемого объекта (под физической схемой БД будем понимать реализацию логической схемы в конкретной СУБД):

$$S_{ps} = \langle V, K, D \rangle, \quad (8)$$

где $V = \{V_1, \dots, V_\mu\}$ – множество базовых отношений (таблиц), V_μ – μ -я таблица: $V_\mu = (v_{\mu 1}, \dots, v_{\mu \psi})$, $v_{\mu \psi}$ – ψ -й столбец μ -й таблицы; $K = \{K_1, \dots, K_\sigma\}$ – множество ключей, в том числе определяющих связи между таблицами (внешних ключей), $D = \{D_1, \dots, D_\theta\}$ – множество иных объектов схемы, обеспечивающих эффективность выборки (индексы), целостность и безопасность данных (триггеры, функции, процедуры и т. д.).

Синтез управления

Суть этапа синтеза заключается в определении управления U^* , реализация которого в объекте дает возможность добиться заданных целей управления Z^* (выражение (2)).

Подставляя в выражение (2) полученную модель F , приходим к многокритериальной задаче оптимизации:

$$\eta_v(F(X', U)) \rightarrow \underset{U \in \Omega}{extr}, \quad (v = 1, \dots, l_3), \quad (9)$$

где Ω – множество допустимых управлений, определяемое следующими соотношениями:

$$\Omega: \begin{cases} \varphi_i(F(X', U)) = a_i, & (i = 1, \dots, l_1); \\ \psi_j(F(X', U)) \leq b_j, & (j = 1, \dots, l_2); \\ U \in R. \end{cases} \quad (10)$$

Для решения этой задачи, прежде всего, следует осуществить свертку экстремальных целей (9). Однако ввиду того, что все атрибуты качества, входящие в выражение (9), имеют различные меры (единицы физической величины) и, как следствие, не могут непосредственно объединяться и сопоставляться, все эти атрибуты в первую очередь необходимо привести к единой шкале оценки. Один из возможных подходов приведения к такой единой безразмерной шкале оценки – от 0 до 1 приведен в работе [5]. В соответствии с данным подходом рассчитываются относительные значения этих атрибутов качества – как отношение фактического значения ($z_v^{факт}$, где $v = (1, \dots, l_3)$) к базовому (эталонному) значению ($z_v^{эталон}$):

$$S_v = z_v^{факт} / z_v^{эталон}. \quad (11)$$

Эталонное значение соответствует значению данного атрибута у лучшего образца БД из числа аналогов. К аналогам относятся ранее созданные БД ИСОУ, отвечающие современному уровню развития, подобные по функциональному назначению и условиям применения, что и оцениваемая БД. Кроме того, данный подход предполагает наличие сведений о важности этих атрибутов качества, обычно учитываемых с помощью специальных коэффициентов весомости, определяемых экспертным путем [5]. В результате осуществления таких преобразований был получен интегральный показатель качества создаваемой в процессе реинжиниринга БД ИСОУ, определяемый как среднее взвешенное значение от заданных относительных значений атрибутов качества с учетом коэффициентов их весомости (важности):

$$J = \sum_{v=1}^{l_3} k_v \times S_v, \quad (12)$$

где k_v – весовые коэффициенты, учитывающие важность (весомость) v -го атрибута качества; S_v – относительное значение v -го атрибута качества.

Ограниченность выделяемых ресурсов R заставляет разработчика искать наиболее эффективный вариант решения поставленной задачи. А это предполагает, как отмечалось в работе [6], необходимость сравнения того, что дает и во что обходится в рассматриваемом случае создаваемая в процессе реинжиниринга БД ИСОУ. То есть необходимость сравнения, не только исходя из результативности, расчет которой осуществляется в соответствии с выражением (12), но и эффективности как степени удовлетворения потребностей и достижения целей, соотношенной с соответствующими затратами ресурсов. Об этом также указывалось в работах [7, 8]. При этом затраты ресурсов R можно привести и выразить в стоимостном

эквиваленте либо в ином виде [9], например в виде трудозатрат (финансовые затраты являются тривиальной функцией от трудоемкости [10]), значения которых далее, согласно изложенному выше подходу, привести к безразмерной величине. То есть к так называемому значению относительных трудозатрат R_n , определяемому как отношение реальных трудозатрат на обеспечение соответствующих значений атрибутов качества создаваемой в процессе реинжиниринга БД ИСОУ $P_{факт}$, к эталонному значению полученных ранее трудозатрат на создание лучшего образца БД из числа аналогов – $P_{эталон}$.

В результате соотнесения интегрального показателя качества (12) со значением относительных трудозатрат R_n был получен показатель эффективности, учитывающий рациональность использования ресурсов. Показатель эффективности процесса реинжиниринга БД ИСОУ – есть отношение среднего взвешенного значения от заданных относительных значений атрибутов качества, с учетом коэффициентов их весомости, к относительному значению трудозатрат на обеспечение этих качеств:

$$E = \frac{1}{R_n} \times \sum_{v=1}^{l_3} k_v \times S_v, \quad (13)$$

где

$$R_n = P_{факт} / P_{эталон}. \quad (14)$$

В результате проведенных преобразований получаем оптимизационную задачу:

$$E(X', U) = \frac{1}{R_n} \times \sum_{v=1}^{l_3} k_v \times S_v \rightarrow \max_{U \in \Omega} \Rightarrow U^*, \quad (15)$$

где Ω – определяется в соответствии с (10), U^* – оптимальное управление.

Решение данной задачи при незначительной сложности вносимых структурных изменений (с учетом всех схем различного уровня) и низкой частоте изменений условий функционирования (состояния предметной области, требований к функциональности, значениям атрибутов качества БД ИСОУ, расходованию ресурсов) не вызывает принципиальных затруднений. Однако при увеличении частоты изменения условий функционирования и сложности вносимых структурных изменений, обусловленных сегодня достаточно частыми обновлениями нормативно-законодательной базы, стремлением компаний к увеличению поддерживаемой функциональности систем в связи с обострившейся конкуренцией, учитывая сложность формализации правил применения моделей, методов и средств как составных элементов управляющего воздействия соответствующих технологий и характер их зависимости (разрывность, нелинейность), решение полученной оптимизационной задачи существенно затрудняется. Кроме того, эта задача для сложных объектов управления, к которым относится и БД ИСОУ, имеющих сложные законы их описания, обычно является многоэкстремальной и овражной [3]. Все это в совокупности заставляет искать и использовать для ее решения специальные методы поисковой оптимизации.

Выводы

1. Показана актуальность проблемы выбора технологии при реинжиниринге баз данных ИСОУ в условиях необходимости своевременного завершения соответствующих проектов в рамках запланированного бюджета с заданными характеристиками качества.

2. Сформулирована и формализована задача реинжиниринга БД ИСОУ в виде известной из общей теории адаптивных систем задачи структурной адаптации, которая в результате преобразований была приведена к оптимизационной задаче.

3. Отмечена сложность решения полученной оптимизационной задачи, обусловленная трудностью формализации правил применения моделей, методов и средств как составных элементов управляющего воздействия при использовании соответствующей технологии, характера их зависимости (разрывности, нелинейности), сложностью закона описания объекта управления, и приводящая к необходимости находить и использовать специальные методы поисковой оптимизации.

4. Полученное выражение для определения показателя эффективности (формула (13)) может быть практически использовано при расчете относительной (сравнительной) эффективности процесса реинжиниринга БД ИСОУ для определения конкретной количественной оценки преимущества одной информационной технологии перед другой для обоснованного выбора ее применения в проекте.

Список литературы: 1. *Chaos Manifesto 2013: Think Big, Act Small* online version. The Standish Group [Electronic resource]. – Access mode : <http://www.versionone.com/assets/img/files/ChaosManifesto2013.pdf>. 2. *Standish Group 2015 Chaos Report – Q&A with Jennifer Lynch* [Electronic resource]. – Access mode : <https://www.infoq.com/articles/standish-chaos-2015>. 3. *Расстригин, Л. А.* Адаптация сложных систем / Л. А. Расстригин. – Рига : Зинатне, 1981. – 375 с. 4. *Дейт, К. Дж.* Введение в системы баз данных ; 8-е изд. ; пер. с англ. / К. Дейт. – Москва : Изд. дом "Вильямс", 2005. – 1328 с. 5. *Андон, Ф. И.* Основы инженерии качества программных систем ; 2-е изд., перераб. и доп. / [Ф. И. Андон, Г. И. Коваль, Т. М. Коротун, Е. М. Лаврищева и др.]. – К. : Академперіодика, 2007. – 672 с. 6. *Захаров, В. Н.* Системы управления. Задание. Проектирование. Реализация / В. Н. Захаров, Д. А. Поспелов, В. Е. Хазацкий. – Москва : Энергия, 1977. – 423 с. 7. *Архипенков, С.* Лекции по управлению программными проектами / С. Архипенков. – Москва, 2009. – 128 с. 8. *Липаев В. В.* Качество программных средств : метод. рекомендации / В. В. Липаев ; под общей ред. проф., д.т.н. А. А. Полякова. – Москва : Янус-К, 2002. – 400 с. 9. *Юркова, Т. И.* Термин «эффективность процесса» и его экономическая интерпретация / Т. И. Юркова // *Современные проблемы науки и образования.* – 2014. – № 1. 10. *Макконнелл, С.* Сколько стоит программный проект / С. Макконнелл. – Москва : Русская редакция ; СПб. : Питер, 2007. – 297 с.

*Харьковский национальный университет
имени В.Н.Каразина*

Поступила в редколлегию 20.04.2017