Математичні моделі й методи в системах автоматизованого управління, проектування

УДК 004.05:004.658.6

Т.И. Брагина, аспирант
Запорожский национальный технический университет
Bragina.zntu@gmail.com

Информационная технология оценки достоверности данных при интеграции web-ориентированных систем

В данной работе рассматривается предложенная риск-ориентированная информационная технология оценки достоверности данных при интеграции web-ориентированных информационных систем, позволяющая снизить тестовое покрытие при интеграции баз данных web-ориентированных информационных систем на основании априорного анализа рисков.
Ключевые слова: интеграция, web-технологии, тестирование, базы данных.

Введение

В настоящее время подавляющее большинство предприятий используют информационные технологии (ИТ). Однако в случае неупорядоченной автоматизации возникает проблема так называемой «лоскутной автоматизации», что подразумевает отсутствие единых автоматизированных бизнес-процессов и наличие обособленных информационных систем (ИС), в ряде случаев дублирующих друг друга [1]. Такая ситуация приводит к задаче интеграции распределенных ИС, а в связи с активным ростом Интернета и web-технологий все более актуальной становится интеграция web-ориентированных ИС (ВОИС).

При интеграции ВОИС необходим постоянный контроль над содержимым базы данных (БД) и корректностью использования данных методами системы, т.к. сложнее всего выявить неточности в логике приложения и обмене информацией между модулями и БД. Сложность такого контроля заключается в необходимости обработки и тестирования большого набора данных, который свойственен ВОИС. Поэтому актуальной задачей является разработка ИТ, позволяющей выполнить тестирования БД при интеграции ВОИС с одним или несколькими источниками данных, а именно:

– автоматизировать процесс оценки состояния интеграции в условиях неопределенности;

– автоматизировано сформировать тестовое покрытие и план тестирования на основании априорного анализа рисков для кортежей БД и методов ВОИС;

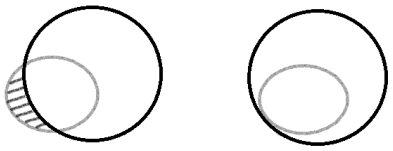
– обеспечить организованный процесс оценки качества на ранних этапах процесса разработки интегрированной ВОИС.

Постановка задачи

При интеграции ИС при анализе предметной области необходимо представить все слои взаимодействия между интегрированными ИС и БД для определения всех информационных процессов, которые происходят в системе, и, соответственно, возможных локализаций ошибок. Для этого необходимо выбрать архитектуру системы интеграции, наиболее распространенными из которых являются консолидация, федерализация, распространение данных и сервисный подход (Service Oriented Architecture) [2 – 5].

При интеграции ВОИС наиболее часто для хранения и распространения данных используют федеративный подход, т.е. единое виртуальное видение разнородных источников данных, которые хранятся фактически в разных по составу и структуре источниках.

При таком подходе в различных источниках данные могут не только дублироваться, но и конфликтовать друг с другом. Из-за этого возникает задача не только интегрировать схемы хранения и создать процессор федерализации для доступа к физически распределенным данным, но и иметь средства для поиска неактуальных, ошибочных и некорректных данных (рис. 1).



а) б)

Рисунок 1 а) – конфликт множества кортежей двух БД; б) – кортежи двух БД содержат данные, не конфликтующие между собой

Пусть задача интеграции данных ВОИС выглядит следующим образом:

$$DB_{ВОИС} = \{DB_1\} \cap \{DB_2\}, \{DB_1\} \subseteq \{DB_2\}, \quad (1)$$

где $DB_{ВОИС}$ – БД ВОИС, которая использует данные из двух (или более) БД $\{DB_1\}$ и $\{DB_2\}$, не конфликтующие между собой.

Следовательно, для достижения такого состояния необходимо обнаружить все кортежи данных, соответствие которым не обнаружено в другом источнике данных:

$$\{DB_{conf}\} = \{DB_1\} / \{DB_2\}, \quad (2)$$

т.е. все кортежи $\{DB_{conf}\}$ будут содержать искомые неактуальные, ошибочные или некорректные данных.

Исходя из того, что полное тестирование всех кортежей БД и модулей интегрированной системы, имеющих доступ к ним, является невозможным в связи с несопоставимыми трудозатратами на тестирование, предлагается использовать риск-ориентированный метод тестирования интегрированных баз данных [6] для определения наиболее критичных кортежей БД и модулей ВОИС, связанных с ними, а также механизмов выбора методов тестирования для них.

Используя предложенные в данном методе модификации, для решения задач (2) было решено разработать ИТ оценки достоверности данных при интеграции ВОИС, основанную на модели верификации БД [6, 7]:

$$M_{db} = < I, D, V_M, M, P, R, E, P_R, P_E, R_T >, \quad (3)$$

где M_{db} – модель БД, которую необходимо иметь для интегрируемых $\{DB_1\}$ и $\{DB_2\}$;

I – извлеченные знания из предметной области различными методами;

D – поля (атрибуты сущностей) БД;

M – совокупность методов, имеющих доступ к БД. В случае использования N различных приложений $M = \{M_1, M_2, \ldots M_N\}$, каждое из которых содержит множество методов $M_1 = \{m_{11}, m_{21}, \ldots m_{kl}\}$;

V_M – переменные, используемые в методах M, содержащие информацию из атрибутов БД;

P – план работ с d_i, где $d_i \in D$;

R – множество возможных рисков для $d_i (d_i \in D)$, $v^i_{m^k}$, $m_k (m_k \in M)$;

E – множество ошибок, которые могут возникнуть при работе с d_i, $E = \{E_{DB}, E_M, E_N\}$, где E_{DB} – ошибки, возникшие при обращении к БД, E_M – ошибки, возникшие в M, E_N – новые выделенные возможные ошибки, которые еще не отнесли к E_{DB} или E_M;

P_R – множество продукционных правил, отражающих план действий в случае возникновения ошибок при работе с d_i, где $d_i \in D$;

P_E – множество проверочных корректировочных тестов, содержащих план обработки ошибок при работе с d_i, где $d_i \in D$; элементы множества P_E имеют доступ к данным d_i, и могут менять их формат;

R_T – результат тестирования, исходя, из которого определяется необходимость в повторном запуске используемого проверочного теста либо изменении направления тестирования, т.е. применении других тестов.

Решение задачи

При интеграции ВОИС с дополнительной БД для контроля достоверности данных была разработана риск-ориентированная ИТ оценки достоверности данных при интеграции ВОИС, модель которой представлена на рисунке 2.

Предложенная ИТ состоит из 7 модулей, итерационно использующихся в процессе разработки. Рассмотрим подробнее использование каждого из них.

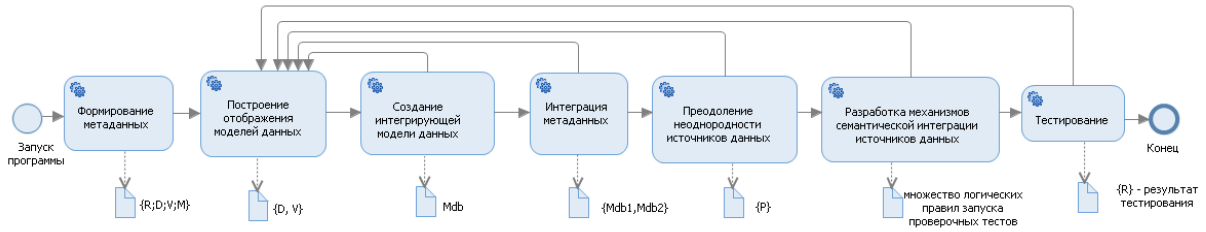


Рисунок 2 – Модель риск-ориентированной ИТ оценки достоверности данных при интеграции ВОИС

Модуль 1. В первом модуле выполняется формирование метаданных для источников данных и модулей ВОИС, которые их используют:

– записываются упорядоченные наборы имен полей БД $\{d_i\}$, которые используются в ВОИС;

– записываются упорядоченные наборы имен переменных $v^i_{m^k}$, которые используются методами ВОИС m^k;– проводится анализ рисков, которые могут возникнуть при работе с ВОИС, определяется связь рисков с выделенными объектами, переменными и методами в виде:

$$\{r_j; d_i; v^i_{m^k}; m^k\}; \quad (4)$$

– формируется множество e_i, $e_i \in E$ – множество ошибок, которые могут возникнуть при работе с d_i, $e_i \in \{E_{DB}, E_M\}$ в случае реализации риска.

Модуль 2. В модуле 2 происходит построение отображения моделей данных, для которого определяются объекты информационного пространства, их представление в БД (массив D) и отражение в слоях ВОИС (массив V_M) следующим образом в виде матрицы связей $\{d_i; v^i_{m^k}\}$:

– d_i – представление i–той характеристики объекта из множества I, хранящееся в БД;

– v^i_m – переменная, используемая в методе m^k, $m^k \in M$, и содержащая информацию из d_i; в случае, использования нескольких наборов методов для различных приложений $M = \{M_1, M_2, \ldots M_N\}$ для каждого атрибута d_i устанавливается связь с каждым набором методов.

Модуль 3. В модуле 3 создается интегрирующая модель данных (3), которая отличается от существующих наличием информации о рисках, связанных с данными БД, и их критичности, а также связью полей БД и переменных методов.

Модуль 4. В модуле 4 в случае использования нескольких источников данных интегрируются метаданные, выделенные в первом модуле для БД, а затем исключается дублирование полученной информации относительно метаданных о переменных в модулях ВОИС, которые имеют доступ к БД:

$$\mathrm{M}_{\mathrm{db\,ВОИС}} = \mathrm{M}_{\mathrm{db1}} \cup \mathrm{M}_{\mathrm{db2}}.$$

Модуль 5. Модуль 5 используется для преодоления неоднородности источников данных. Для этого в данном модуле создается:

–план тестирования p_i для проверки элементов d_i, v^i_m или m^i, для которых на предыдущем этапе были выделены риски r_j, $r_j \in R$.

– P_E, $p^j_A \in P_E$ – множество проверочных корректирующих тестов, содержащих план обработки ошибок e_i, $e_i \in \{E_{DB}, E_M\}$ при работе с d_i, где $d_i \in D$.

Проверочные корректирующие тесты создаются по следующему алгоритму:

1. Определение граничных значений для атрибута d_i, и связанной с ним переменной $v^i_{m^k}$;

2. Определение возможных рисков и ошибок, возникающих в результате реализации данных рисков, при работе с d_i и $v^i_{m^k}$; каждый риск априорно оценивается двумя характеристиками: вероятность возникновения p^i_r и влияние (критичность) c^i_r, а также определяется, может ли изменяться уровень этого риска в процессе тестирования s^i_r и на какую величину $\{a^i_r; e_j\}$, т.е. создается правило вида:

$$if\,(e_j \cap s^i_r) \Rightarrow p^i_r = p^i_r + a^i_r / k, \quad (5)$$

где k – количество протестированных записей между возникновениями ошибки e_j.

3. Составление автоматического теста, анализирующего состояние d_i или $v^i_{m^k}$ и корректирующего его формат для соответствия его требования.

4. Выполнение проверочного корректирующего теста.

5. Анализ эффективности разработанного теста.

6. Анализ возможности применения данного теста к другим атрибутам d_j и переменным $v^j_{m^k}$.

Модуль 6. В модуле 6 происходит разработка механизмов семантической интеграции источников данных. В качестве механизма семантической интеграции был выбран механизм логических правил следующего вида:

если во время выполнения теста p_i, возникла ошибка из множества E_{DB} и/или E_M, необходимо выполнить проверочный тест p^j_E:

$$if\,(result(p_i) = failed \cap \cap (e_i \in E_{DB} \cup e_i \in E_M) \Rightarrow p^j_A. \quad (4)$$

Модуль 7. В модуле 7 происходит непосредственно тестирование данных

интегрированной ВОИС и корректируется БД проверочных тестов следующим образом:

1. Если во время выполнения теста p_i, возникла непредвиденная ошибка e_N, т.е. не принадлежащая множеству E_{DB} или E_M, необходимо добавить новый вид ошибки в E_N. Через определенный период времени тестировщик должен рассмотреть данную ошибку, определить с какими элементами d_i, v^i_m или m^i она связана, перенести e_N из множества E_N в E_{DB} или E_M, написать новый проверочный тест p_j, создать новый риск r_j, затем связать новые элементы $\{r_j;\ d_i;\ v^i_{m^k}\ ;\ m^k\}$ и создать логическое правило вида:

$$if\,(result(p_i) = failed \cap (e_N \notin E_{DB} \cap e_N \notin E_M) \Rightarrow$$
$$\Rightarrow e_N \in E_N \Rightarrow$$
$$\Rightarrow (move(e_N, E_{DB}) \cup move(e_N, E_M)) \cap$$
$$\cap create(p^j_E, r_j, \{r_j; d_i; v^i_{m^k}; m^k\}) \Rightarrow p^j_E.$$

2. Если при прохождении теста p_i не возникли ошибки, необходимо занести в r^i_T статус «*checked*», иначе, в r^i_T заносится код полученной ошибки;

3. Если тест p_i проверяет кортеж данных, для которых установлено r^i_T статус «*checked*» и не выделено критичных рисков, выполнение данного теста пропускается:

$$if\,(r^i_T = checked\ \cap (d_i; r_i) \in 0) \Rightarrow$$
$$\Rightarrow result(p_i) = passed. \qquad (5)$$

4. Если тест p_i или метод m_i обновляет кортеж данных, для него необходимо установить r^i_T статус «*not tested*».

Риски, выделенные в первом и четвертом модуле, влияют на процесс тестирования и позволяют принимать управляющие решения для автоматизации составления плана тестирования.

Представленная ИТ на основе модели верификации БД (3) отличается наличием связей множества проверочных корректировочных тестов, содержащих план обработки ошибок P_E, с рисками для атрибутов сущностей БД. Данная связь позволяет автоматически определять наиболее критичные кортежи данных и методы тестирования, направленные на контроль наиболее вероятных рисков, с целью исправить некоторое множество известных ошибок $\{E_{DB}, E_M\}$, а также получить информацию о неисследованных ошибках.

Практическое применение

Предложенная ИТ была использована для тестирования корректности взаимодействия БД приемной комиссии с БД «Єдиної державної електронної бази з питань освіти» («ЄДЕБО») [8]. Задача состояла в контроле процесса интеграции БД, созданной с использованием системы управления БД MS Access, и БД, созданной на базе технологии Oracle. Общая структура инструментальной среды тестирования представлена на рис. 3. Составные части представленной инструментальной среды были полностью или частично автоматизированы.

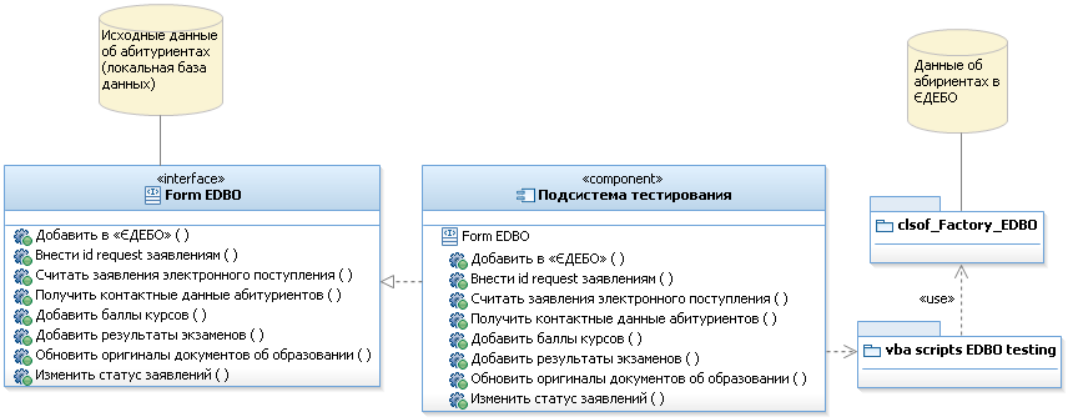


Рисунок 3 – Структурная схема инструментальной среды тестирования

Рассмотрим использование модулей ИТ в данном практическом примере.

При формировании метаданных для источников данных и модулей ВОИС, которые их используют:

– были выделены основные и вспомогательные сущности и оценена мощность множества тестируемых атрибутов D из (3) – $|D| = 450$, записаны имена полей БД сущностей «Абитури-

ент», «Заявка», «Специальности» и т.д., которые используются в ВОИС;

– записаны имена переменных $v_{m^k}^i$, которые используются методами локальной БД и «ЄДЕБО»;

– проведен анализ рисков, которые могут возникнуть при передаче и считывании данных с «ЄДЕБО», определяется связь рисков с выделенными полями, переменными и методами;

– сформировано множество ошибок e_i, $e_i \in E$, которые могут возникнуть при работе с данными, например для поля БД, содержащее номер паспорта абитуриента определены возможные ошибки e_{10}= {«значение не соответствует граничным условиям», «паспорт не найден», «персона с указанным паспортом уже существует»}.Было построено отображение моделей данных, для которого определяются объекты информационного пространства, их представление в БД (массив D) и отражение в слоях ВОИС (массив V_M) следующим образом в виде матрицы связей $\{d_i;\ v_{m^k}^i\}$между полями БД и переменными в методах, например, d_{10}= "*pasnum*" v_m^{10} ="PasportNumber".

Была создана интегрирующая модель данных (3), которая содержит выделенную информацию о рисках, связанных с данными БД, и их критичности, а также связи между полями БД и переменными методов, например, для d_{10} было выделено множество рисков {«некорректный ввод данных»; «потеря данных при передаче»; «отсутствие данных»}. Из данного множества наиболее критичным является риск «некорректный ввод данных», вероятность возникновения $p_r^{10} = 0{,}2$, критичность $c_r^{10} = 1$ (критичность была оценена по шкале [0; 1]) (рис.4).

Т.к. в рассматриваемом случае интегрируются два источника данных, были интегрированы метаданные для локальной БД и «ЄДЕБО», выделенные в первом модуле, и исключены повторы метаданных:

$$M_{db\ ВОИС} = M_{db1} \cup M_{db2}.$$

Для преодоления неоднородности источников данных созданы план тестирования и множество проверочных корректирующих тестов для каждого атрибута d_i. *Б*ыли определены граничные значения и условия на значение, а также переменные v_m^i, используемые в методах интегрированной ИС в следующем виде, например:

- $d_{10}['name'] =$" *pasnum*";

- 2 варианта граничных условий:

1) для граждан Украины: $d_{10}['\min'] = 000000$, $d_{10}['\max'] = 999999$;

2) для нерезидентов $d_{10}['\min'] = 0$, $d_{10}['\max'] = 9999999999$.

Следующим этапом было определение рисков и вероятных ошибок для атрибутов, например, для номера паспорта установлено $s_r^{10} = true$, что означает, что вероятность возникновения риска может изменяться в процессе тестирования на величину $a_r^{10} = 1$ в случае, если получена ошибка «значение не соответствует граничным условиям», т.е. было создано правило вида:

$$if\,(e_{10}[1] \cap s_r^{10}) \Rightarrow p_r^{10} = p_r^{10} + a_r^{10}/k.$$

Автоматические проверочные тесты были написаны с помощью языка *Visual Basic for Application* и запросов *SQL*. Например, для исключения вероятности некорректного переноса из текстового формата в числовой, было написано несколько тестов, среди них следующий запрос «*UpdatePassNum1*»:

```
UPDATE MAIN
SET MAIN.pasnum = "0" & [pasnum]
WHERE (((Len([pasnum]))=5));
```

Были разработаны механизмы семантической интеграции источников данных в виде логических правил, например:
если во время выполнения теста p_i, который проверяет соответствие номера паспорта граничным условиям возникла ошибка из множества E_{DB} и/или E_M, и при этом абитуриент является резидентом Украины, необходимо выполнить проверочный тест «*UpdatePassNum1*».
В **модуле 7** проводилось непосредственно тестирование данных интегрированной ВОИС и корректировалась БД проверочных тестов.

Например, после выполнения проверочных корректирующих тестов для верификации 700 записей в локальной БД были получены результаты R_T из (3), представленный тест «*UpdatePassNum1*» обнаружил и исправил 40 записей номеров паспортов.

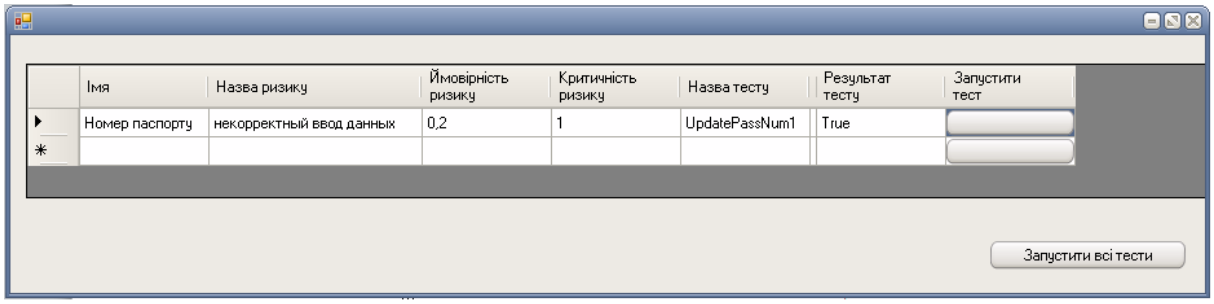


Рисунок 4 – Пример внесения информации для риск-ориентированного теста

При передаче данных в «ЄДЕБО» были обнаружены две ошибки $e_{10}[2]$ =«паспорт не найден». Следовательно, эффективность приведенного для примера теста составила 95%, т.е. в 95% случаев применение корректировочного теста привело к исчезновению ошибок e_{10}= {«значение не соответствует граничным условиям», «паспорт не найден», «персона с указанным паспортом уже существует»}. В остальных случаях, ошибкой являлась механическая опечатка при вводе цифр, что невозможно исправить автоматически.

Вероятность возникновения риска «некорректный ввод данных» увеличился до $p_r^{10} = 0{,}25$ для d_{10}, но уменьшен для v_m^{10} $p_r^{10}(v_m^{10}) = 0$, т.к. данные, связанные с этой переменной были проверены и по возможности исправлены, следовательно, нет необходимости их повторно тестировать.

Количество реализованных тестов для проверки информации в локальной БД равно 197, для проверки корректности данных полученных из «ЄДЕБО» – 253 и для проверки данных, используемых в методах для обмена информацией – 69. При тестировании 700 кортежей в основных сущностях локальной БД с использованием стандартных методов тестирования [8] тестовое покрытие составило 519 тестов и было затрачено 2,3 часа. При использовании предложенной ИТ количество тестовых сценариев в тестовом покрытии составило 482 теста при временных затратах 1,94 часа. Количество выявленных ошибок в обоих случаях составило 120 некорректных данных в значениях атрибутов локальной БД.

Из вышеописанного можно сделать вывод, что применение предложенной ИТ оценки достоверности данных при интеграции ВОИС позволило снизить временные затраты на 17% за счет сокращения тестового покрытия, сохранив уровень не выявленных ошибок на допустимом уровне (0,1% ошибочных данных по отношению к общему числу данных в БД).

Заключение

В данной работе рассматривается ИТ оценки достоверности данных при интеграции ВОИС, которая реализует риск-ориентированный метод тестирования интегрированных БД и заключается в формировании тестового покрытия и плана тестирования на основании априорного анализа рисков для кортежей БД и методов ВОИС, а также дает возможность корректировать начальную оценку рисков, используя результаты тестирования. За счет возможности автоматически формировать тестовое покрытие на основе анализа рисков сокращается количество используемых тестовых сценариев и как следствие, снижаются временные затраты на тестирование.

Предложенная информационная технология представляет несколько возможностей:

– использование информации о выделенных рисках позволит определить наиболее критичные модули и кортежи для тестирования;

– использование проверочных корректирующих тестов позволит автоматизировать трудоемкий процесс приведения данных к необходимому формату и непротиворечивости;

– постоянный контроль новых неизученных ошибок позволит расширить базу тестов и, как следствие, снизить одну из самых распространённых причин ошибок в приложении – появление непредвиденных и, следовательно, неисследованных комбинаций входных данных.

Автоматизация предложенного риск-ориентированного метода тестирования интегрированных БД в рамках ИТ оценки достоверности данных при интеграции ВОИС позволило снизить временные затраты на 17% за счет сокращения тестового покрытия, сохранив уровень не выявленных ошибок на допустимом уровне (0,1% ошибочных данных по отношению к общему числу данных в БД).

Список литературы

1. Чеснавский, А.А. Инструментальные средства интеграции контента унаследованных веб-приложений в единое информационное пространство предприятия [Текст]: автореф. дис. канд. екон. наук: 08.10.01 / Чеснавский Александр Александрович; Моск. инженерно-физич. институт. – М., 2009. – 18 с.
2. Черненко, Н.В. Методы и информационная технология интеграции баз данных корпоративных систем [Текст]: дис. канд. техн. наук / Н.В. Черненко. – Х., 2013. – 202 с.
3. IBM: Patterns: Implementing an SOA Using an Enterprise Service Bus [Электронный ресурс] / Режим доступа: http://www.redbooks.ibm.com/redbooks/pdfs/sg246346.pdf
4. Microsoft Integration Patterns [Электронный ресурс] / Режим доступа: http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=12474
5. Хоп Г. Шаблоны интеграции корпоративных приложений / Г. Хоп, Б. Вульф – М.: Издательский дом «Вильямс». – 2006. – 672 с.
6. Брагина, Т.И. Риск-ориентированный метод тестирования интегрированных баз данных /Т.И. Брагина, Г.В. Табунщик // «Электротехнические и компьютерные системы». – Одесса, 2014. – №13 (89). – С. 223-230.
7. Брагина, Т.И. Модифицированная риск-ориентированная модель верификации интегрированных баз данных [Текст] / Т.И. Брагина, Г.В. Табунщик // Радиоэлектроника и молодежь в XXI веке: тези доп. 17-ого Міжнар. молод. форума, ХНУРЕ–2014, квітень 14–16, 2014. В 6 т. – Х.:ХНУРЕ. Т. 6. 2014. – С. 172-173.
8. Брагіна, Т.І. Розробка засобів інтеграції / Т.І. Брагіна // Системи обробки інформації. Вип.8 (106). – 2012. – С. 127-130.

Надійшла до редколегії 12.03.2014

Т.І. БРАГІНА
Запорізький національний технічний університет

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ОЦІНЮВАННЯ ДОСТОВІРНОСТІ ДАНИХ ПРИ ІНТЕГРАЦІЇ WEB-ОРІЄНТОВАНИХ СИСТЕМ

У даній роботі розглядається запропонована ризик-орієнтована інформаційна технологія оцінювання достовірності даних при інтеграції web-орієнтованих інформаційних систем, що дозволяє знизити тестове покриття при інтеграції баз даних web-орієнтованих інформаційних систем на підставі апріорного аналізу ризиків.

Ключові слова: інтеграція, web-технології, тестування, бази даних.

T.I. BRAGINA
Zaporizhia National Technical University, Ukraine

INFORMATION TECHNOLOGY FOR DATA ACCURACY ASSESSMENT WITHIN INTEGRATION OF WEB-BASED SYSTEMS

In this paper the risk-oriented information technology for data accuracy assessment within integration of web-based systems was proposed by the author, which allows reducing test coverage within database integrating for web-oriented information systems based on a priori risk analysis.
Presented information technology based on a database verification model is characterized by having a plurality of parity relations of corrective tests containing the errors treatment plan with risks for database entities attributes. This connection allows identifying automatically the most critical data tuples and test methods aimed at controlling the most probabile risks in order to fix a set of known bugs, as well as get information about undiscovered errors.
The proposed information technology presents several possibilities:
–information on selected risks will allow determining the most critical modules and tuples for testing;
–screening corrective tests will automate the time-consuming process of bringing data to the required format and consistency;
– continuous monitoring of new unexplored errors will expand the base of tests and, consequently, reduce one of the most common causes of errors in the application – the emergence of unanticipated and hence unexplored combinations of inputs.
Automation of the proposed risk-based method of integrated databases testing within information technology for data accuracy assessment within integration of web-based systems reduced the time spent by 17% by reducing test coverage, maintaining the level of undetected errors at an acceptable level (0.1% of erroneous data in relation to the total number of data in the database).

Keywords: integration, web-technologies, testing, database.