
Бурый А.С., Морин Е.В. Структурирование информационных данных при сертификации программных продуктов // Информационно-экономические аспекты стандартизации и технического регулирования, 2017. № 5(39).

УДК 004.05:004.93

СТРУКТУРИРОВАНИЕ ИНФОРМАЦИОННЫХ ДАННЫХ ПРИ СЕРТИФИКАЦИИ ПРОГРАММНЫХ ПРОДУКТОВ

Бурый А.С., д.т.н., Российский научно-технический центр информации по стандартизации, метрологии и оценке соответствия, директор департамента общероссийских классификаторов (ФГУП «СТАНДАРТИНФОРМ»)
Морин Е.В., соискатель Российский научно-технический центр информации по стандартизации, метрологии и оценке соответствия, директор департамента общероссийских классификаторов (ФГУП «СТАНДАРТИНФОРМ»)

В статье рассматривается структуризация данных о результатах испытаний программных продуктов при их сертификации в виде гиперкуба. Показывается возможность применения OLAP-технологий для хранения информации о показателях качества программной продукции.

Ключевые слова: испытательные лаборатории; качество программных продуктов; программные средства; сертификация программных средств; структуризация данных; OLAP-технология.

UDC 004.05:004.93

QUALITY ASSESSMENT AT THE STAGE OF CERTIFICATION OF SOFTWARE

Buryu A.S., doctor of technical sciences, Director of the Department of all-Russian classifiers, FSUE «STANDARTINFORM»
Morin E.V., the applicant of the Director of the Department of all-Russian classifiers, FSUE «STANDARTINFORM»

The article discusses the structuring of data about the results of tests of software products in their certification in the form of a hypercube. Shows the applications of OLAP-technologies for the storage of heterogeneous information on the indicators of quality of software products.

Key words: testing laboratories; quality software products; software; certification software; structuring data; OLAP technology.

Несмотря на сложности в экономике, обусловленные рядом причин, рынок информационных, компьютерных, телекоммуникационных технологий продолжает тенденцию роста, чему способствует развитие коммуникационной инфраструктуры, систем распространения и использования информационных ресурсов и программных средств [1]. Успехи технологий виртуального пространства (AR/VR – технологии), performance – технологий с учетом социальной направленности пользователей сети интернет, развитие рекламных технологий на базе programmatic – технологий, технологий машинного обучения и многие другие направления, в том числе и на базе облачных сервисов [2, 3], требуют существенных изменений, как программных продуктов, так и оборудования.

На фоне коммерциализации интернет пространства, активно развивается архитектура открытых систем. На базе этой архитектуры разрабатываются правительственные профили по федеральным органам исполнительной власти и службам, а также обеспечивающих их систем, с большим разнообразием протоколов обмена данными, прикладных программ, функционирующих в едином информационном пространстве, в том числе и в рамках Федеральной целевой программы «Электронная Россия» [4]. Указанные особенности акцентируют необходимость обеспечения совместимости программных средств (ПС) различных производителей. *Совместимость* – «пригодность изделий или их систем к совместному использованию при определенных условиях для выполнения соответствующих требований, которая не вызывает при этом нежелательных последствий» [5].

Контроль качества готового программного продукта осуществляется в ходе сертификационных испытаний [5], которые проводятся силами испытательных лабораторий (ИЛ), аккредитованных на проведение только тех испытаний, которые предусмотрены в их нормативных документах.

Причинами, снижающими качество программных продуктов, часто выступают [6]:

«низкая технологичность производства компонентов и комплексов программ;

недостаточно эффективные средства защиты информационных и программных ресурсов;

несоответствие реальных и декларируемых функциональных характеристик разрабатываемых компонентов и комплексов программ;

несоответствие требованиям стандартов, влекущее за собой невозможность взаимодействия, совершенствования и развития систем;

реализованные алгоритмы обработки информации, неспособны обеспечить в течение жизненного цикла ПС надежное и своевременное представление полной, достоверной, актуальной и конфиденциальной информации для функционального использования». Это подтверждает тот факт, что указанная совокупность свойств не обеспечивает *качество* функционирования программного продукта.

При этом *качество функционирования* программного продукта (ПП) неразрывно с понятием качества функционирования информационной системы (на примере компьютерной системы), как совокупности свойств «качества системы/программного продукта, включая функциональную пригодность, уровень производительности, совместимость, удобство пользования, надежность, защищенность, сопровождаемость и переносимость (мобильность)» [7]. Процесс функционирования информационной системы регулируется определенными нормами, стандартами и соответствует целевому предназначению надсистемы [5, 8].

На рис. 1. испытательная лаборатория представлена в виде совокупности испытательных участков, включающих оборудование, документацию, общие методики испытаний, а также персонал (специалисты,

эксперты, технологи) формируемый для испытания определенных видов программной продукции.

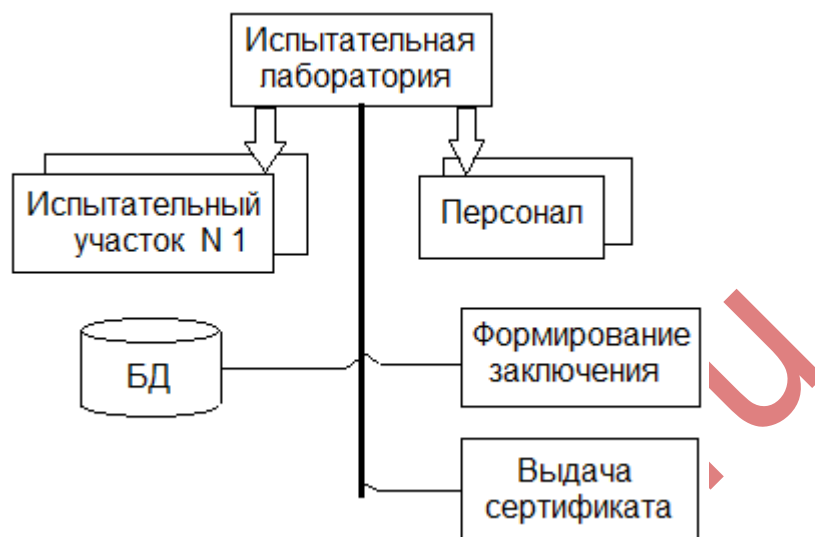


Рис. 1. Общая структура взаимодействия в ходе сертификации

База данных (БД) используется для информационного обеспечения испытаний на этапах их подготовки, проведения, записи результатов и регистрации итоговых документов.

Важным этапом предварительной обработки содержательной информации является ее структурирование, т.е. организация логического построения информационных массивов и отдельных данных для решения целевых задач на этапах манипулирования, а также представление данных в определенной форме и формате. Результаты испытаний образцов программных средств, первоначально группируются в неструктурированной форме, когда текст, цифровые данные существуют в неупорядоченном виде в ходе сбора данных от различных источников, этапов проверки (тестирования), или когда необходима переработка существующих систем управления базами данных, которые первоначально использовались для поиска и хранения только текстовых файлов [9].

Формирование подобных информационных структур требует организации взаимосвязей между рядом показателей, часть из которых имеет

цифровую форму представления, часть – качественное представление, а также в виде нечетких лингвистических переменных. При таком многообразии входных данных, уместно говорить о семантическом образе объекта или явления, когда возможны описания свойств объекта на естественном языке [10], когда в обработку могут попасть данные в формате изображений, видео- или аудио-файлы и т.д.

Государственная метрологическая служба России (ГМС) представляет собой территориально-распределенную информационную систему, в составе которой функционируют испытательные лаборатории, ряд из которых осуществляет сертификационную деятельность программной продукции.

Результаты проведенных испытаний образцов ПП фиксируются в общей базе данных, где отображается номер сертификата, наименование продукта, которому он соответствует, разработчик данного программного продукта, основные характеристики ПП, а также параметры (характеристики качества), которые были зафиксированы в результате испытаний.

Анализ ряда территориально распределенных систем сбора данных показывает, что зачастую первичные измерительные данные в силу множества причин, могут иметь пропуски, т.е. характеризуются неполнотой данных [11, 12]. При этом *данные* будем рассматривать, как продукт информационных систем, в ходе сбора, поиска, преобразования, накопления, хранения и других действий над данными [13]. Применительно к системам измерений, участвующим в процессе испытаний ПП *данные* – есть результаты измерений или результаты обработки измерительной информации [14].

Одним из направлений представления слабоструктурированной информации выступают модели представления данных на основе разреженных гиперкубов [10, 11, 15]. При этом оси гиперкуба являются параметрами, а ячейки – агрегированными данными. Гиперкубическое (многомерное) представление данных является основой технологий

оперативной аналитической обработки данных *OLAP* (On-Line Analytical Processing) и интеллектуального анализа данных (Data Mining – технологии выявления скрытых взаимосвязей внутри больших баз данных), используя алгоритмы регрессионного анализа, кластеризации, прогнозирования и др. *OLAP*-концепция продолжает активно развиваться на протяжении последних десятилетий, когда появились их разновидности: *MOLAP* (для баз данных с многомерной структурой данных), *ROLAP* (с реляционной структурой) и *HOLAP* (с гибридной структурой).

Гиперкуб представляет собой многомерную логическую модель данных в виде упорядоченного набора ячеек (Рис. 2), каждая из которых содержит одно измерение или результаты измерения по одному объекту (явлению). Таким образом, измерение образует одну из граней гиперкуба. На рис. 2 представлен трехмерный гиперкуб. Любой элемент измерения представляет собой координату по данному измерению, а совокупность таких координат образует кортеж, который идентифицирует ячейку гиперкуба C_{ijk} с координатами (d_{1i}, d_{2j}, d_{3k}) для трехмерного куба [16], при этом общему числу изменений соответствует множество $D = \{d_1, d_2, \dots, d_n\}$. Таким образом, n будет задавать размерность гиперкуба.

С ячейкой ассоциируются m показателей качества, характеризующих результаты испытания ПП, проведенные в конкретном испытательном центре или лаборатории, т.е. (f_1, f_2, \dots, f_m) .

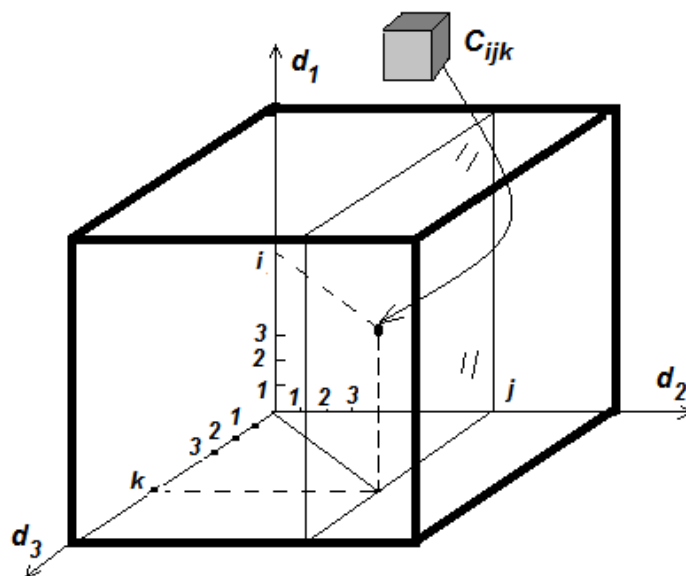


Рис. 2. Гиперкуб результатов данных испытаний

Под срезом гиперкуба будем понимать плоскость $Sr_j, j \in d_2$, составленную из ячеек, или ее элементы, когда фиксируются значения мер хотя бы по двум измерениям, параллельную плоскости $d_1 0 d_3$, образованной осями d_1 и d_3 , где 0 – начало системы координат. Меры – это числовые значения, позволяющие создавать срезы, выполнять операции вращения, свертки, агрегирования.

Измерения (координаты) представляют собой совокупность данных, например, испытательная лаборатория (ИЛ), ДАТА (дата проведения испытаний), ПРОГРАММНЫЙ ПРОДУКТ (ПП) (объект испытания), требуемая оперативная память (ТОП) для функционального применения ПП и т.д. По запросу можно получить аналитику: сколько в ИЛ №01084 в 2016 г. сертифицировано ПП «Аист» и значения ТОП для каждого пакета.

Для формирования базы данных (БД), представления связей между отдельными сущностями вводят понятие отношения, которые традиционно представляются в виде таблиц, где строки – сущности, а столбцы – атрибуты или измерения (оси гиперкуба). Множество отношений и, соответственно определенное на них множество атрибутов запишем в виде:

$$R = \{R_1, R_2, \dots, R_l\}, A = \{A_1, A_2, \dots, A_g\},$$

где g – количество атрибутов, а l – количество отношений или мощность множества R [17].

В табл.1 представлены примеры идентификаторов атрибутов, которые применим для составления отношений, исходя из технологических особенностей формирования запросов.

Таблица 1

Примеры идентификации атрибутов базы данных

Идентификатор атрибута	Имя атрибута	Идентификатор атрибута	Имя атрибута
A ₁	Код ИЛ	A ₈	ФамИмяОтч
A ₂	Название ИЛ	A ₉	Должность
A ₃	Контакты ИЛ	A ₁₀	Смена
A ₄	ТабНомерСотрудник	A ₁₁	ХарКачества
A ₅	РегНомерПП	A ₁₂	ХарЗащиты
A ₆	Назв ПП	A ₁₃	ХарУнивер
A ₇	Автор_Разработчик	A ₁₄	ХарФункц

Испытательная лаборатория = $R_1(A_1, A_2, A_3)$; *Персонал* = $R_2(A_4, A_8, A_9, A_{10})$;

Программный продукт = $R_3(A_5, A_6, A_7)$; *Карточка ПП* = $R_4(A_5, A_6, A_{11})$;

Характеристики качества ПП = $R_5(A_5, A_{12}, A_{13}, A_{14})$.

На основании данных отношений можно проанализировать, например штатный состав ИЛ, какие ПП проходят контроль в ИЛ определенного региона и т.д.

Обращение к OLAP–системам позволяет исследовать информационные массивы большего объема, интегрировать данные различных информационных систем, решать задачи прогнозирования отдельных показателей и объектов исследования по комплексу показателей.

Список использованных источников и литературы

1. Терехов А. Прогноз развития digital-рынка на 2017 год // COSSA.RU: информационный портал о маркетинге и коммуникациях в цифровой среде. [Электронный ресурс]. Режим доступа: <http://www.cossa.ru/152/154262/>.

2. Бурый А.С. Тенденции развития распределенных информационных систем на основе облачных технологий // Транспортное дело России. – 2013. № 6. – С. 160-162.

3. Бурый А.С. Информационное пространство сетевого взаимодействия в клиентской среде // Транспортное дело России. – 2011. № 8. – С. 156-158.

4. Ломакин М.И., Докукин А.В. Функции единой информационной системы по техническому регулированию в рамках концепции электронного государства // Перспективы науки. – 2011. № 12 (27). – С. 230-233.

5. Ананьева, Т.Н. Стандартизация, сертификация и управление качеством программного обеспечения: учеб. пособие / Т.Н. Ананьева, Н.Г. Новикова, Г.Н. Исаев. – М.: ИНФРА-М, 2017. – 232 с.

6. Липаев В.В. Сертификация программных средств. Учебник. – М.: СИНТЕГ, 2010. – 348 с.

7. ГОСТ ИСО/МЭК 25010-2015. Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов. – М.: Стандартинформ, 2015. – 36 с.

8. Ломакин М.И., Докукин А.В. Интеграция российских инновационных предприятий в мировую экономику на основе развития информационного обеспечения стандартизации // Российское предпринимательство. – 2012. № 2(200). – С. 109-119.

9. Черняк А. Аналитика неструктурированных данных // Открытые системы. СУБД. – 2012. № 6. – С. 30-34.

10. Сизов А.С., Шельдешов С.И. Формализация многомерных слабо-структурированных данных на основе разреженных гиперкубов // Инновации в информационно-аналитических системах: сб. науч. трудов. – Курск: Наукком, 2013. Вып. 5. – С. 7-15.

11. Заботнев М.С. Методы представления информации в разреженных гиперкубах данных [Электронный ресурс]. Режим доступа: <http://www.olap.ru/basic/theory.asp>.

12. Бурый А.С. Отказоустойчивые распределенные системы переработки информации / А.С. Бурый. – М.: Горячая линия – Телеком, 2016. – 128 с.

13. ГОСТ Р 51170-98. Качество служебной информации. Термины и определения. – М.: ГОССТАНДАРТ РОССИИ, 1998. – 8 с.

14. ГОСТ Р 8.654-2015. Государственная система обеспечения единства измерений. Требования к программному обеспечению средств измерений. Основные положения. – М.: Стандартинформ, 2016. – 8 с.

15. Ахрем А.А., Макаров И.М., Рахманкулов В.З., Ровкин И.О. Повышение производительности и масштабируемости реляционных хранилищ данных и OLAP-систем // Информационные технологии и вычислительные системы. – 2007. № 3. – С. 59-70.

16. Миронов В.В., Макарова Е.С. Агрегация показателей в OLAP-кубе при сведении по зависимым измерениям // Вестник УГАТУ. – 2012. Т. 16. № 3(48). – С.180-186.

17. Зыкин С.В. Формирование гиперкубового представления данных со списочными компонентами // Информационные технологии и вычислительные системы. – 2010. № 4. – С. 38-46.

© Бурый А.С.

© Морин Е.В.