ОРГАНИЗАЦИЯ ЗДРАВООХРАНЕНИЯ

УДК 004.041 Оригинальная статья

ВЫБОР КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ ДЛЯ АНАЛИТИЧЕСКОЙ ПОДДЕРЖКИ БАЗЫ ДАННЫХ КРУПНОМАСШТАБНЫХ МЕДИЦИНСКИХ ИНФОРМАЦИОННЫХ СИСТЕМ

А.П. Бирюков *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна — Федерального медико-биологического агентства», заведующий отделом радиационной эпидемиологии, профессор, доктор медицинских наук;* ***Е.В. Васильев*** *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна — Федерального медико-биологического агентства», инженер лаборатории информационно-программного обеспечения и технической поддержки баз данных;* ***С. М. Думанский*** *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна — Федерального медико-биологического агентства», старший научный сотрудник лаборатории радиационно-эпидемиологического анализа, кандидат экономических наук;* ***О.А. Тихонова*** *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна — Федерального медико-биологического агентства», старший научный сотрудник лаборатории радиационно-эпидемиологического анализа, кандидат медицинских наук;* ***Ю.А. Герт*** *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна — Федерального медико-биологического агентства», инженер лаборатории радиационно-эпидемиологического анализа;* ***Н.В. Капитонова*** *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна Федерального медико-биологического агентства», научный сотрудник.*

PREFERANCE OF COMPUTER TECHNOLOGY FOR ANALYTICAL SUPPORT OF LARGE DATABASE OF MEDICAL INFORMATION SYSTEMS

A.P. Biryukov *— State Scientific Research Center n.a. A.I. Burnazyan — Federal Medical Biophysical Center of Federal Medical Biological Agency, Head of the Department of Radiation Epidemiology, Professor, Doctor of Medical Science;* ***E.V. Vasiliev*** *— State Scientific Research Center n.a. A.I. Burnazyan — Federal Medical Biophysical Center of Federal Medical Biological Agency, laboratory information software and technical support database, engineer;* ***S. M. Dumansky*** *— State Scientific Research Center n.a. A.I. Burnazyan — Federal Medical Biophysical Center of Federal Medical Biological Agency, Laboratory of Radiation and epidemiological analysis, Candidate of Economic Sciences, Senior Researcher;* ***O.A. Tikhonova*** *— State Scientific Research Center n.a. A.I. Burnazyan — Federal Medical Biophysical Center of Federal Medical Biological Agency, Laboratory of Radiation and epidemiological analysis, Candidate of Medical Sciences, Senior Researcher;* ***Ju.A. Gert*** *— State Scientific Research Center n.a. A.I. Burnazyan — Federal Medical Biophysical Center of Federal Medical Biological Agency, Laboratory radiation- epidemiological analysis, Engineer;* ***N.V. Kapitonova*** *— State Scientific Research Center n.a. A.I. Burnazyan — Federal Medical Biophysical Center of Federal Medical Biological Agency, scientific researcher.*

Дата поступления — 28.11.2013 г. Дата принятия в печать — 16.12.2013 г.

Бирюков А.П., Васильев Е.В., Думанский С. М., Тихонова О.А., Герт Ю.А., Капитонова Н.В. **Выбор компьютерных технологий для аналитической поддержки базы данных крупномасштабных медицинских информационных систем // Саратовский научно-медицинский журнал. 2013. Т. 9, № 4. С. 983–987.**

Цель: изучение возможностей использования интеллектуальных технологий для аналитической поддержки баз данных крупномасштабных медицинских информационных систем. *Материал и методы.* В работе использованы методы объектно ориентированного проектирования программного обеспечения и проектирования баз данных. *Результаты.* На основании экспертного изучения моделей и алгоритмов анализа клинико-эпидемиологических данных и принципов представления знаний в крупномасштабных медицинских информационных системах схемы интеллектуального анализа данных были реализованы в программном комплексе Единого регистра ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России. Были определены области эффективного применения абстрактной модели данных EAV и процедур Data Maning при проектировании эволюционирующей базы данных биомедицинского регистра. *Выводы.* Использование интеллектуальной программной платформы, поддерживающей различные наборы программных интерфейсов и объектных моделей для различных операций в различных программных средах, позволяет выстраивать и поддерживать в масштабах всей информационной системы сквозные процедуры и процессы обработки разнородных биомедицинских данных.

Ключевые слова: бизнес-интеллектуальные технологии, информационно-аналитические системы, хранилища данных.

Biryukov A.P., Vasiliev E.V., Dumansky S. M., Tikhonova O.A., Gert Ju.A., Kapitonova N.V. **Preferance of computer technology for analytical support of large database of medical information systems // Saratov Journal of Medical Scientific Research. 2013. Vol. 9, № 4. P. 983–987.**

Aim: to study the use of intelligent technologies for analytical support of large databases of medical information systems. *Material and methods*. We used the techniques of object-oriented software design and database design. *Results*. Based on expert review of models and algorithms for analysis of clinical and epidemiological data and prin-

ciples of knowledge representation in large-scale health information systems, data mining schema were implemented in the software package of the register of Research Center n.a. A.I. Burnazyan of Russia. Identified areas for effective implementation of abstract data model of EAV and procedures Data Maning for the design of database of biomedical registers. *Conclusions*. Using intelligent software platform that supports different sets of APIs and object models for different operations in different software environments, allows you to build and maintain an information system through the procedures of data biomedical processing.

Key words: OLAP, Data Mining, Business Intelligence, Analysis Services, Data warehouse, Integration Services, Reporting Services.

Введение. Объективная оценка влияния на здоровье человека ионизирующего излучения является одним из важнейших направлений исследований современного здравоохранения. Особенно важным при этом является разработка методологии сбора и обработки первичной медицинской и дозиметрической информации для оценки влияния радиоэкологических факторов на состояние здоровья. Необходимым условием реализации эпидемиологических подходов к оценке воздействия ионизирующего излучения является объединение существующих информационных ресурсов с целью обеспечения долговременного, автоматизированного персонального учета лиц, подвергшихся радиационному воздействию в результате производственной деятельности или иных ситуаций. Задача объединения и обработки этих данных в интегрированном информационном пространстве осложнена тем, что существующие в настоящее время цифровые ресурсы представляются совокупностью гетерогенных информационных источников и баз данных. В связи с этим необходимо формирование интегрированной информационной среды, где соответствующая ей совокупность гетерогенных информационных источников и баз данных будет представляться пользователям непротиворечивым и интегрированным образом через единый интерфейс.

Таким образом, для объективной оценки данных крупномасштабных информационных систем требуется использование нетрадиционных наукоемких аналитических решений, к которым можно отнести бизнес-интеллектуальные технологии, практическому использованию которых в области радиационной медицины и посвящена данная работа.

Цель: изучение возможностей использования интеллектуальных технологий для аналитической поддержки баз данных крупномасштабных медицинских информационных систем.

Материал и методы. Для выполнения поставленных задач выявления закономерностей развития тех или иных заболеваний у подвергшихся облучению групп населения было принято решение использовать следующее программное обеспечение:

MS SQL Server 2005, для управления базами данных (БД), с входящими компонентами Analysis Services, Integration Services, Reporting Services;

MS Windows Server 2003, операционная система (ОС);

Visual studio 2005, среда разработки.

OLAP (On line Analytic Processing) в широком понимании — это технология обработки данных, организованных в многомерную Базу Данных (БД), часто называемую Хранилищем Данных (Data warehouse), логически представляемым в виде многомерного Куба (Cube). Основными объектами OLAP-куба являются Меры (Measures) и Измерения (Dimensions). В роли мер или фактов выступали таблицы реляционной БД, имеющие в своем составе интегрируемые значения. В качестве измерений использовали таблицы, определяющие группировки данных, по которым в дальнейшем определяли интегрируемые показатели. Таблицы мер и измерений были организованы по принципу «Звезды» или «Снежинки». Схема «Звезда» определялась в виде таблицы мер в середине, от которой отходят связи к таблицам измерений, причем между таблицами измерений не должно существовать связей. Схема «Снежинка» такая же, как «Звезда», но допускающая связи между таблицами измерений [1, 2].

Результаты. В ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России (далее — ФМБЦ) в течение многих десятилетий интегрировались данные радиационно-эпидемиологического характера, которые впоследствии были объединены по тематическим рубрикам в Единый регистр (ЕДрег), на базе которого спроектирован Центр хранения и обработки данных (ЦХОД). В основе архитектуры ЦХОД лежат принципы и концепции проектирования хранилищ данных (Data warehouse), позволяющие применять различного рода информационно-аналитические системы [3, 4].

На рис. 1 показана модель развернутого многомерного куба, спроектированная применительно к данным Единого регистра ФМБЦ им. Бурназяна, которая состоит из трех таблиц мер и девяти таблиц измерений. Как видно из рис. 1, одна и та же таблица может выступать в роли как меры, так и измерения. Например, в таблице PPP зафиксированы персональные данные пациентов: пол, возраст и т.д. В таблице TTT отражены данные о госпитализациях пациента: дата госпитализации, оценка поглощенной дозы и т.д. В таблице TTT_1 отмечены выявляемые впервые заболевания при очередной госпитализации. Куб спроектирован по схеме «Звезда». Таблицы с префиксом SPR и таблица ICD10 — это справочные таблицы данных, причем последняя таблица является международным иерархическим классификатором болезней в последней редакции. В разделах Measures и Dimensions перечислены соответственно таблицы мер и измерений, знак плюс указывает на признак узла, щелкнув по которому объект распадается на составляющие его атрибуты таблиц и иерархии атрибутов по принципу «класс заболеваний –> блок заболеваний –> заболевание».

Перетаскивая мышью из разделов Dimensions и Measures атрибуты и иерархии атрибутов на соответствующие области Drop Row Fields Here, Drop Column Fields Here и Drop Totals or Detail Fields Here, формируют требуемые отчеты, руководствуясь правилом первоначального перенесения от более агрегированных иерархий к менее агрегированным.

Под технологией Data Maning («добыча данных») понимается технология анализа больших объемов данных для обнаружения скрытых закономерностей. Службы Analysis Services Data Maning включают следующие типы алгоритмов:

алгоритмы классификации (осуществляют прогнозирование одной или нескольких дискретных переменных на основе других атрибутов в наборе данных);

Ответственный автор — Бирюков Александр Петрович
Адрес: 123182, Москва, ул. Живописная, 46.
Тел.: (499) 190-96-93
E-mail: abirukov@fmbcfmba.ru

регрессивные алгоритмы (осуществляют прогнозирование одной или нескольких непрерывных переменных, на основе других атрибутов в наборе данных);

алгоритмы сегментации (делят данные на группы или кластеры элементов, имеющих схожие свойства);

алгоритмы взаимосвязей (осуществляют поиск корреляции между различными атрибутами в наборе данных);

алгоритмы анализа последовательностей (обобщают часто встречающиеся последовательности в данных).

Работа со всеми типами алгоритмов унифицирована и состоит из трех главных этапов:

1) создается модель добычи данных с использованием определенного алгоритма, настроенная на обучающую выборку данных;

2) по обучающей выборке (в которой известны как исходные атрибуты, так и те атрибуты, которые мы собираемся предсказывать в будущем) производим обучение модели добычи данных;

3) после обучения на вход модели добычи данных подаются исходные атрибуты и рассчитываются выходные атрибуты.

Тема Microsoft Data Mining очень обширна, и здесь кратко представлены лишь два алгоритма: Microsoft Time Series и Microsoft Decision Trees. В SQL Server 2005 в алгоритме Time Series используется оригинальный алгоритм ARTXP, оптимизированный для краткосрочных прогнозов. Алгоритм ARTXP можно описать как модель дерева с авторегрессией для представления данных в периодических временных рядах.

На рис. 2 показаны график заболеваемости и график прогноза заболеваемости по всем нозологиям и новообразованиям, где в качестве источника данных выступает описанный OLAP-куб. Хорошо видно, что заболеваемость для данной когорты достигла максимума в 1997–1998 гг. и далее снизилась до уровня стабилизации, предсказываемого моделью.

Алгоритм Decision Trees (Майкрософт) представляет собой гибридный алгоритм, объединяющий различные методы для создания дерева и поддерживающий несколько аналитических задач, в том числе регрессию, классификацию и взаимосвязи. Алгоритм представлен в работе Meek C. и Hechrman D. [5].

Использованные аналитические подходы могут быть использованы не только в практике радиационной эпидемиологии. Так, одной из практических реализаций алгоритма была выполненная нами задача по прогнозированию структуры осложнений при оперативном вмешательстве — холецистэктомии, а также выраженности косметического эффекта и выбора типа операции. Рассматривали три типа операций в зависимости от числа используемых троакаров, а именно одно- (1 группа), двух- (2 группа) и четырехпортовый (3 группа) доступ.

На рис. 3 в графическом изображении представлено дерево решений по выбору типа операции, построенное на основании БД, содержащей персональные данные о больных с желчнокаменной болезнью (ЖКБ). При обработке данных учитывали такие параметры, как пол, возраст, индекс массы тела, величина желчного пузыря перед операцией по данным ультразвукового исследования, данные фиброгастроскопии, опыт хирурга, длительность операции и т.д. Технологии Data Maning позволяют с долей вероятности, выраженной в процентах, получить прогноз до операции об исходе операции и определить, какой вид вмешательства (в зависимости от числа используемых троакаров) использовать у того или иного больного с ЖКБ. На примере имеющейся базы данных нами было установлено, что величина желчного пузыря (выраженность остроты инфекционного процесса) влияет на число послеоперационных осложнений, таких, как серома в области троакарных ран и желчный перитонит. В то же время даже при увеличенном желчном пузыре, при большом личном опыте хирурга по выполнению лапароскопических холецистэктомий (200 и более операций) величина желчного пузыря не оказывает большого влияния на

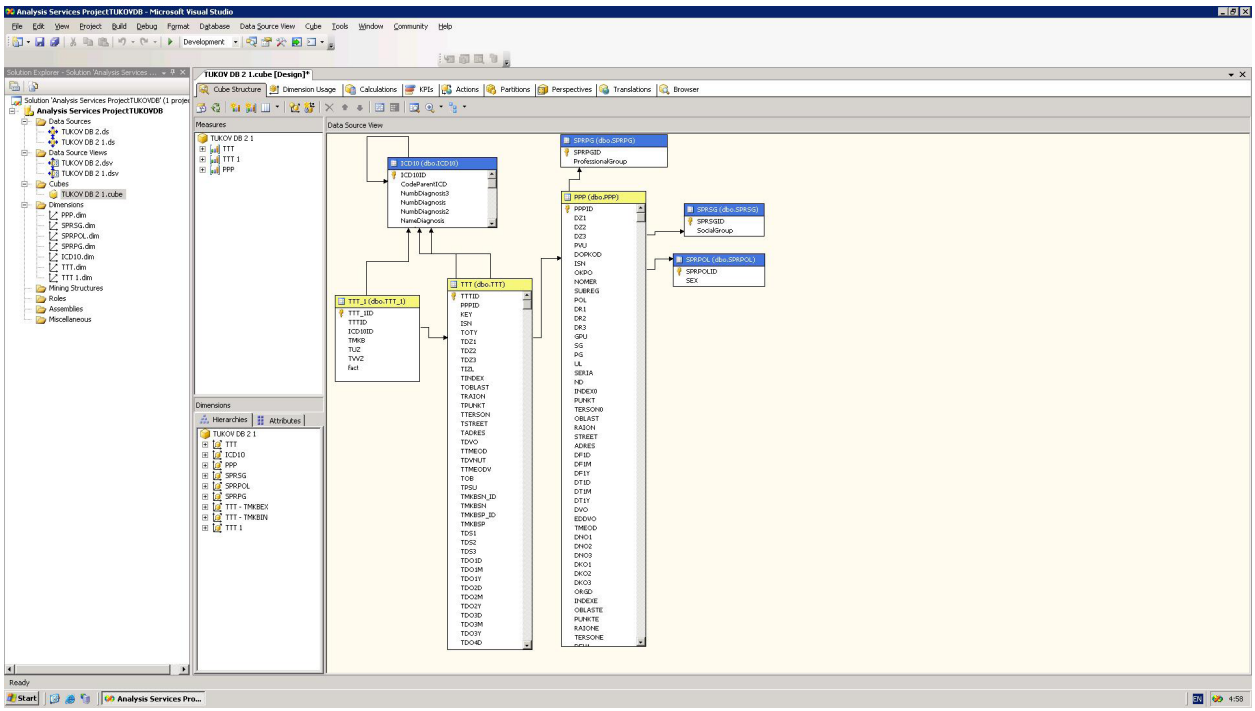

Рис. 1 Вкладка Cube Structure (светлые заголовки столбцов — таблицы фактов; темные — измерений)

число и структуру осложнений, а следовательно, и на косметический эффект.

Таким образом, для анализа данных крупномасштабных информационных систем предлагается хорошо зарекомендовавшая себя на практике технология использования процедуры интеллектуального анализа данных. Проведение этих мероприятий позволило обеспечить полную и объективную информацию на достаточно высоком уровне качества и достоверности. В аналогичных информационных системах (Национальный радиационно-эпидемиологический регистр [6] и Регистр острой лучевой болезни, зарегистрированной у работников предприятия атомной промышленности «Маяк» [7]) не были использованы подобные подходы. При этом используемые в перечисленных регистрационных системах

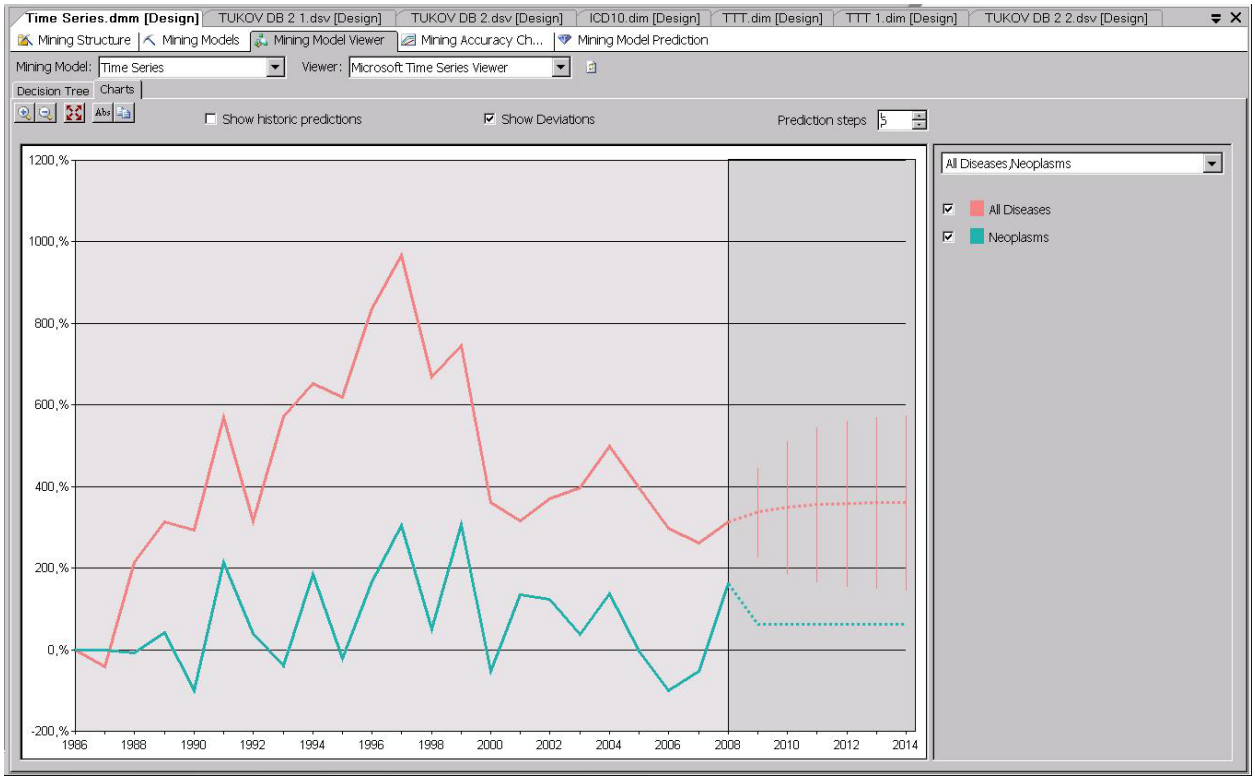

Рис. 2. Заболеваемость в относительных единицах

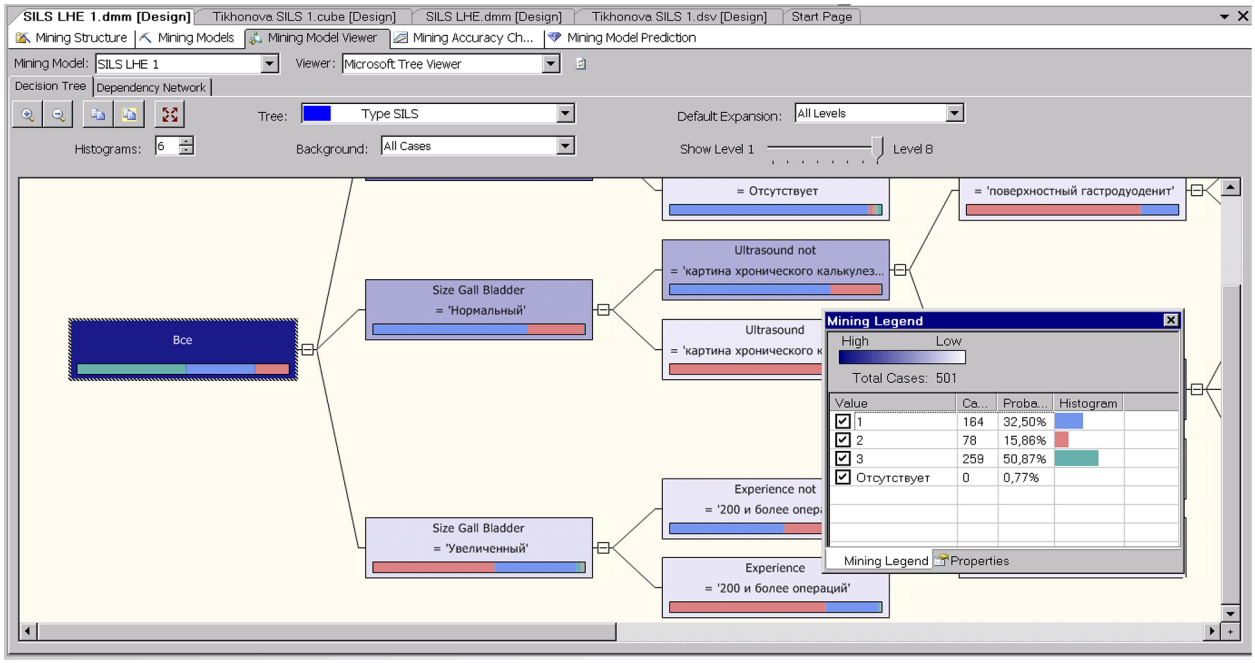

Рис. 3. «Дерево» выбора типа оперативного вмешательства

модели обработки медико-статистической информации рекомендуются для широкого применения на региональном и муниципальном уровнях, а сами автоматизированные системы оперативной информации и методы комплексной интегральной оценки являются основанием для принятия обоснованных управленческих решений. Тем не менее нам для выполнения задач, связанных со спецификой баз данных ФМБЦ, потребовалась разработка дополнительных мероприятий и использование бизнес-ителлектуальных технологий OLAP и Data Mining [8, 9]. При этом разработанные нами методические подходы были максимально приближены к практике российского здравоохранения и требованиям нормативных документов, определяющих лечебно-профилактическое обслуживание населения. В целом предложенный нами комплекс программно-аналитических мероприятий является новым методологическим подходом к решению проблемы качества информации, необходимой для проведения современного радиационно-эпидемиологического анализа.

Заключение. Решение ФГБУ ГНЦ ФМБЦ им. Бурназяна ФМБА России о переводе части задач обработки данных на технологическую платформу Microsoft — одного из лидеров в создании систем бизнес-аналитики, дает ряд преимуществ относительно ранее применяемых “лоскутных” решений и собственных разработок. В первую очередь следует учитывать, что Microsoft предлагает промышленную, поддерживаемую и развиваемую BI-платформу, глубоко интегрированную с другими технологиями и продуктами Microsoft, в частности, с Microsoft SQL Server (аналитические сервисы входят в комплект поставки старших версий SQL Server), Microsoft Office, Microsoft Visual Studio. BI-платформа Microsoft поддерживает различные наборы программных интерфейсов (API) и объектных моделей для различных операций, в различных программных средах, позволяет выстраивать и поддерживать в масштабах всей организации сквозные процедуры и процессы обработки, единые централизованные, аналитические модели и проекты и многое другое.

Конфликт интересов не заявляется

Библиографический список

1. Нильсен П. Microsoft SQL Server 2005: Библия пользователя / пер. с англ. М.: Диалектика, 2008. 1232 с.

2. Хаританх С., Куин С. Microsoft SQL Server 2005: Analysis Services и MDX для профессионалов / пер. с англ. М.: Диалектика, 2008. 834 с.

3. Алексеева Т.В., Амириди Ю.В., Дик В.В. Информационные аналитические системы. М.: МФПУ «Синергия», 2013. 384 с.

4. Туманов В.Е., Маклаков С. В. Проектирование реляционных хранилищ данных. М.: Диалог-МИФИ, 2007. 333 с.

5. Meek C., Hechrman D. Structure and Parameter Learning for Causal Independence and Causal Interaction Models // Research.Microsoft.com. URL http://research.microsoft.com/en-us/um/people/heckerman/INoisyOr.pdf

6. Программно-математический комплекс Российского государственного медико-дозиметрического регистра / А.Ф. Цыб, В.К. Иванов, С. А. Айрапетов [и др.] // Радиация и риск. 1992. № 1. С. 71–93.

7. Азизова Т.В., Мосеева М.Б., Осовец С. В., Сумина М.В. Регистр острой лучевой болезни, зарегистрированной у работников предприятия атомной промышленности «Маяк» // Мед.-биол. и соц.-психол. пробл. безопасности в чрезв. ситуациях. 2011. № 4. С. 14–19.

8. Рыженко Р. Личный опыт: история одного знакомства с OLAP // CNews. URL. http://corp.cnews.ru/text.shtm

9. Meek C., Chickering D.M., Hechrman D. Autoregressive Tree Models for Time-Series Analysis // Research.Microsoft.com. URL http://research.microsoft.com/en-us/um/people/dmax/publications/dmart-final.pdf

Translit

1. Nil’sen P. Microsoft SQL Server 2005: Biblija pol’zovatelja / per. s angl. M.: Dialektika, 2008. 1232 s.

2. Haritanh S., Kuin S. Microsoft SQL Server 2005: Analysis Services i MDX dlja professionalov / per. s angl. M.: Dialektika, 2008. 834 s.

3. Alekseeva T.V., Amiridi Ju.V., Dik V.V. Informacionnye analiticheskie sistemy. M.: MFPU «Sinergija», 2013. 384 s.

4. Tumanov V.E., Maklakov S. V. Proektirovanie reljacionnyh hranilishh dannyh. M.: Dialog-MIFI, 2007. 333 s.

5. Meek C., Hechrman D. Structure and Parameter Learning for Causal Independence and Causal Interaction Models // Research.Microsoft.com. URL http://research.microsoft.com/en-us/um/people/heckerman/INoisyOr.pdf

6. Programmno-matematicheskij kompleks Rossijskogo gosudarstvennogo mediko-dozimetricheskogo registra / A.F. Cyb, V.K. Ivanov, S. A. Ajrapetov [i dr.] // Radiacija i risk. 1992. № 1. S. 71–93.

7. Azizova T.V., Moseeva M.B., Osovec S. V., Sumina M.V. Registr ostroj luchevoj bolezni, zaregistrirovannoj u rabotnikov predprijatija atomnoj promyshlennosti «Majak» // Med.-biol. i soc.-psihol. probl. bezopasnosti v chrezv. situacijah. 2011. № 4. S. 14–19.

8. Ryzhenko R. Lichnyj opyt: istorija odnogo znakomstva s OLAP // CNews. URL. http://corp.cnews.ru/text.shtm

9. Meek C., Chickering D.M., Hechrman D. Autoregressive Tree Models for Time-Series Analysis // Research.Microsoft.com. URL http://research.microsoft.com/en-us/um/people/dmax/publications/dmart-final.pdf

УДК 614.29

Авторское мнение

ПЕРСПЕКТИВЫ РАЗВИТИЯ РЕГЛАМЕНТА ОБЯЗАТЕЛЬНЫХ МЕДИЦИНСКИХ ОСМОТРОВ

А.Ю. Бушманов *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна Федерального медико-биологического агентства», первый заместитель генерального директора, профессор, доктор медицинских наук;* ***А.С. Кретов*** *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна Федерального медико-биологического агентства», Центр профессиональной патологии, врач-профпатолог;* ***Н.Н. Мазитова*** *— ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр им. А.И. Бурназяна Федерального медико-биологического агентства», Институт последипломного профессионального образования, профессор кафедры медицины труда, гигиены и профпатологии, доктор медицинских наук.*