

УДК 004.42

А.Н. Кокоулин, Р.И. ЮжаниновПермский национальный исследовательский политехнический университет,
Пермь, Россия**МНОГОМЕРНЫЙ АНАЛИЗ ДАННЫХ ПО ОБРАЩАЕМОСТИ
В ЛЕЧЕБНЫЕ УЧРЕЖДЕНИЯ С ПОМОЩЬЮ
СРЕДСТВ ORACLE OLAP**

В статье описывается решение задачи анализа больших объемов медицинских данных с использованием технологии OLAP. Для практической реализации была выбрана платформа Oracle OLAP – встроенная в Oracle Database технология, позволяющая осуществлять быстрый доступ к агрегированным данным. Немаловажным достоинством является то, что выборка данных может выполняться даже неквалифицированными специалистами и не требует знаний MS SQL. Дальнейшее развитие технологий многомерного анализа необходимо проводить в направлении интеграции построенных и планируемых к реализации OLAP-кубов в Oracle Fusion Middleware, что позволит быстро создавать приложения, использующие OLAP-данные в расчетах, реализовать экспорт данных в MS Excel, а также в направлении исследования технологии Oracle Data Mining, позволяющей производить статистический анализ данных с построением моделей.

Ключевые слова: OLAP, интеллектуальный анализ данных, база данных.

A.N. Kokoulin, R.I. Yuzhaninov

Perm National Research Polytechnic University, Perm, Russian Federation

**MULTIDIMENSIONAL DATA ANALYSIS IN MEDICAL
INSTITUTIONS USING ORACLE OLAP**

The article describes the analyze process of large volumes of medical data using OLAP technology. For practical implementation was chosen platform Oracle OLAP – built-in Oracle Database technology which enables fast access to aggregated data. An important advantage is that the data sampling can be performed even by unskilled professionals and requires no knowledge of SQL. Further development of technologies multivariate analysis should be carried out towards the integration of built and planned for implementation OLAP-cubes in the Oracle Fusion Middleware, which will allow you to quickly create applications that use the OLAP-data in the calculations, to implement data export to MS Excel, as well as in the direction of technology research Oracle Data Mining, allowing to make a statistical analysis of the data with the construction of models.

Keywords: OLAP, Data Mining, Data Base.

Крупные информационные системы (ИС), к которым можно отнести и систему обработки медицинских данных, как правило, содержат приложения, предназначенные для комплексного многомерного анализа данных, выявления их динамики и тенденций [5, 8]. Предложить или проверить какую-либо гипотезу невозможно без накопления и обработки необходимой для этого исходной информации. С этой целью в информационной системе информация должна содержаться в виде хранилищ данных (Data warehouses), автоматизирующих процесс сбора, отсеивания и предварительной обработки данных с целью предоставления результирующей информации пользователям для статистического анализа [1, 3]. Основные требования, которые обычно предъявляются к хранилищам данных:

- поддержка высокой скорости получения данных из хранилища;
- поддержка внутренней непротиворечивости данных;
- возможность получения и сравнения срезов данных (slice and dice);
- наличие удобных утилит просмотра данных в хранилище;
- полнота и достоверность хранимых данных.

Типичное хранилище данных, как правило, отличается от обычной реляционной базы данных. Во-первых, обычные базы данных предназначены для того, чтобы помочь пользователям выполнять повседневную работу, тогда как хранилища данных предназначены для выполнения анализа данных за большой период [4]. Например, ввод информации о помещении пациента в стационар или ввод анализов в ИС учреждения производится с использованием реляционной базы данных, а анализ динамики обращений пациентов по видам заболеваний или по территориальным признакам за несколько лет – с помощью хранилища данных [5].

Во-вторых, обычные базы данных подвержены постоянным изменениям в процессе работы пользователей, а хранилище данных относительно стабильно: данные в нем обычно обновляются согласно расписанию (например, еженедельно, ежедневно или ежечасно – в зависимости от потребностей). В идеале процесс пополнения представляет собой просто добавление новых данных за определенный период времени без изменения прежней информации, уже находящейся в хранилище [7].

И, в-третьих, обычные базы данных чаще всего являются источником данных, попадающих в хранилище.

OLAP-технология. Системы поддержки принятия решений обычно обладают средствами предоставления пользователю агрегатных данных для различных выборок из исходного набора в удобном для восприятия и анализа виде. Как правило, такие агрегатные функции образуют многомерный набор данных (куб данных), оси которого содержат параметры, а ячейки – зависящие от них агрегатные данные. Вдоль каждой оси данные могут быть организованы в виде иерархии, представляющей различные уровни их детализации. Благодаря такой модели данных пользователи могут формулировать сложные запросы, генерировать отчеты, получать подмножества данных [6].

Таблицы измерений содержат неизменяемые либо редко изменяемые данные. В подавляющем большинстве случаев эти данные представляют собой по одной записи для каждого члена нижнего уровня иерархии в измерении. Таблицы измерений также содержат, как минимум, одно описательное поле (обычно с именем члена измерения) и, как правило, целочисленное ключевое поле (обычно это суррогатный ключ) для однозначной идентификации члена измерения. Если будущее измерение, основанное на данной таблице измерений, содержит иерархию, то таблица измерений также может содержать поля, указывающие на «родителя» данного члена в этой иерархии.

Каждая таблица измерений должна находиться в отношении «один ко многим» с таблицей фактов (рис. 1). Скорость роста таблиц измерений должна быть незначительной по сравнению со скоростью роста таблицы фактов [2].

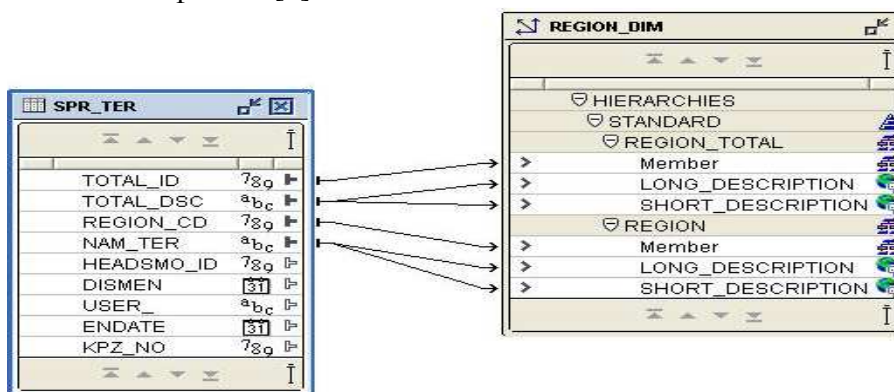


Рис. 1. Пример измерения куба – измерение «Территории»

Данная таблица содержит сведения о месте проживания пациента. Остальные таблицы содержат сведения о возрасте больного, об общих сведениях, связанных с временем (время обслуживания пациента, время его обращения в медучреждение, длительность лечения и т.д.), сведения о диагнозе и сведения о результатах лечения (рис. 2, 3). Эти таблицы в будущем будут являться измерениями куба, и по ним (и их полям) можно будет делать срезы данных и производить различные подсчеты и запросы.

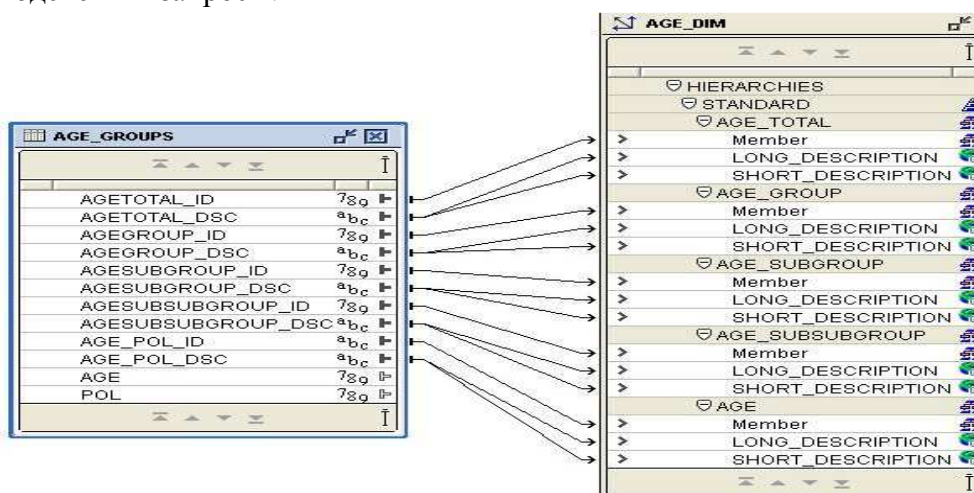


Рис. 2. Пример измерения куба – измерение «Половозрастная группа»

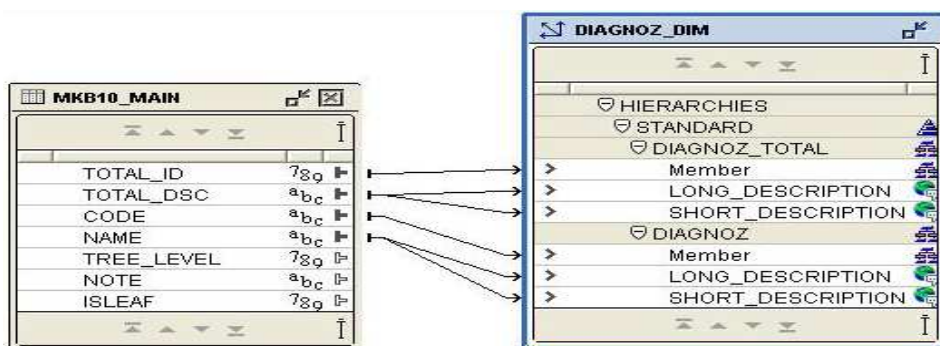


Рис. 3. Измерение «Диагноз», связанное со справочной таблицей стандартных обозначений диагнозов

Одно измерение куба может содержаться как в одной таблице (кстати, и при наличии нескольких уровней иерархии), так и в нескольких связанных таблицах, соответствующих различным уровням

иерархии в измерении. Если каждое измерение содержится в одной таблице, такая схема хранилища данных носит название «звезда» (star schema). Пример такой схемы приведен на рис. 1–3.

Если же хотя бы одно измерение содержится в нескольких связанных таблицах, такая схема хранилища данных носит название «снежинка» (snowflake schema) [3]. Дополнительные таблицы измерений в такой схеме, обычно соответствующие верхним уровням иерархии измерения и находящиеся в соотношении «один ко многим» в главной таблице измерений, соответствующей нижнему уровню иерархии, иногда называют консольными таблицами (outrigger table). Пример схемы «снежинка» приведен ниже (рис. 4).

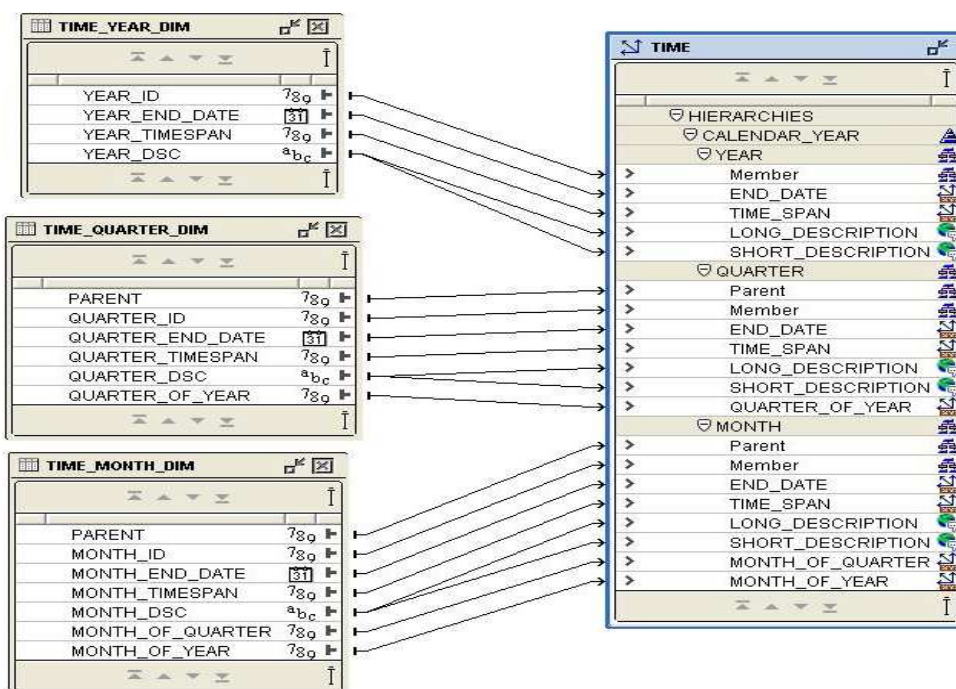


Рис. 4. Измерение «дата посещения», построенное по схеме «снежинка»

Создание OLAP-кубов. Как и измерение, куб можно создать с помощью соответствующего мастера или непосредственно в редакторе кубов в Oracle Analytics Workspace Manager [2, 9, 10]. Ниже приведен пример куба OLAP, построенного из таблиц с данными, накапливаемыми в ходе обследований населения внутри ФГУН ФНЦ УРЗН, и совместно используемыми деперсонифицированными данными (рис. 5).

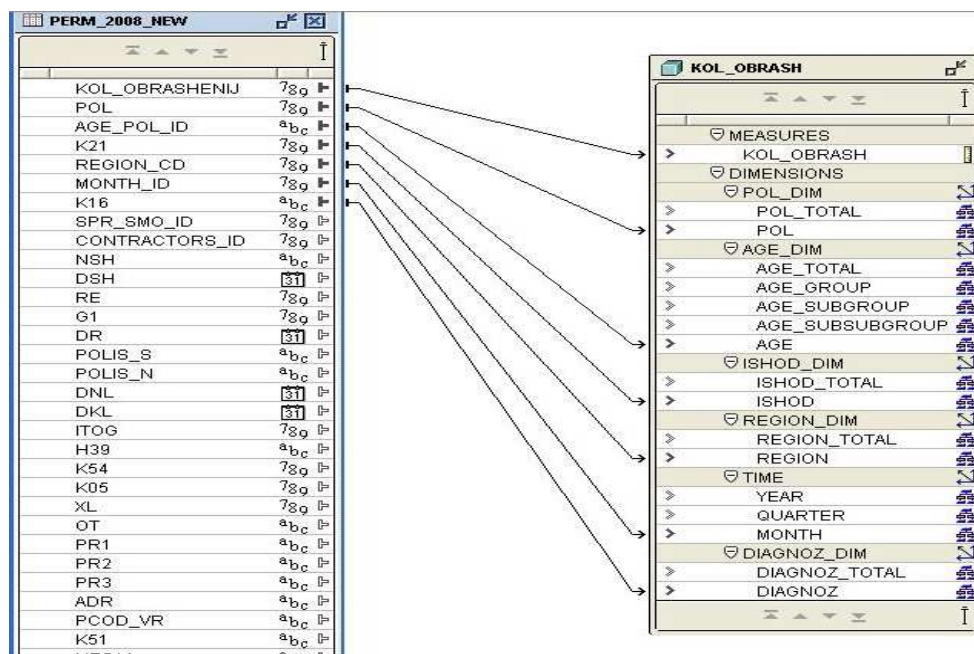


Рис. 5. Связь таблицы фактов (Perm_2008_new) с измерениями и мерами куба

Как можно видеть, измерениями куба являются метаданные, содержащиеся в ранее рассмотренных таблицах. Таблица фактов взята практически без изменений и относится к деперсонифицированным данным – фактам обращений жителей Перми в медицинские учреждения за 2008 год. Гораздо больше усилий было приложено к организации таблиц с измерениями. Так, например, была создана таблица с половозрастной группировкой населения, в которой по возрасту и полу выделялись сначала уровни «дети и подростки», «трудоспособное население» и «население пенсионного возраста», а затем в каждом из уровней выделялись подуровни, такие как «дети до года», «дети 1–3 лет» и т.д., всего 5 уровней. То же самое можно сказать и про таблицу с датами обращений: необходимо было создать несколько уровней группировки: год, квартал, месяц и день. Остальные таблицы измерений являются стандартными для системы здравоохранения и содержат перечень диагнозов, исходов заболеваний, кодов учреждений, территорий (районов проживания) и т.д.

Примеры срезов данных. Ниже приведены несколько срезов данных из построенного куба (рис. 6–7), которые наглядно иллюстрируют возможности технологии OLAP.

	Q1-08			Q2-08			Q3-08		
	ВСЕ ПОЛ	Женский	Мужской	ВСЕ ПОЛЫ	Женский	Мужской	ВСЕ ПОЛЫ	Женский	Мужской
ВСЕ ВОЗРАСТЫ	156 201,00	97 139,00	59 062,00	154 881,00	98 463,00	56 418,00	133 190,00	85 935,00	47 255,00
население пенсионного возраста свыше 54(59) лет	51 141,00	35 997,00	15 144,00	48 231,00	34 482,00	13 749,00	38 424,00	27 675,00	10 749,00
население пенсионного возраста свыше 54(59) лет	51 141,00	35 997,00	15 144,00	48 231,00	34 482,00	13 749,00	38 424,00	27 675,00	10 749,00
дети и подростки 0-17 лет	27 645,00	13 506,00	14 139,00	27 704,00	13 825,00	13 879,00	25 407,00	12 734,00	12 673,00
дети 0-14 лет	22 760,00	11 030,00	11 730,00	23 074,00	11 296,00	11 778,00	21 287,00	10 506,00	10 781,00
дети 0 лет	2 155,00	1 001,00	1 154,00	2 204,00	1 047,00	1 157,00	2 086,00	1 042,00	1 044,00
дети 10-14 лет	5 899,00	2 870,00	3 029,00	5 738,00	2 874,00	2 864,00	5 102,00	2 552,00	2 550,00
дети 1-2 лет	3 483,00	1 664,00	1 819,00	3 572,00	1 692,00	1 880,00	3 665,00	1 740,00	1 925,00
дети 3-6 лет	7 122,00	3 470,00	3 652,00	7 111,00	3 509,00	3 602,00	6 551,00	3 245,00	3 306,00
дети 7-9 лет	4 101,00	2 025,00	2 076,00	4 449,00	2 174,00	2 275,00	3 883,00	1 927,00	1 956,00
подростки 15-17 лет	4 885,00	2 476,00	2 409,00	4 630,00	2 529,00	2 101,00	4 120,00	2 228,00	1 892,00
трудоспособное население 18-54(59) лет	77 415,00	47 636,00	29 779,00	78 946,00	50 156,00	28 790,00	69 359,00	45 526,00	23 833,00

Рис. 6. Выборка количества обращений по возрасту, полу и дате

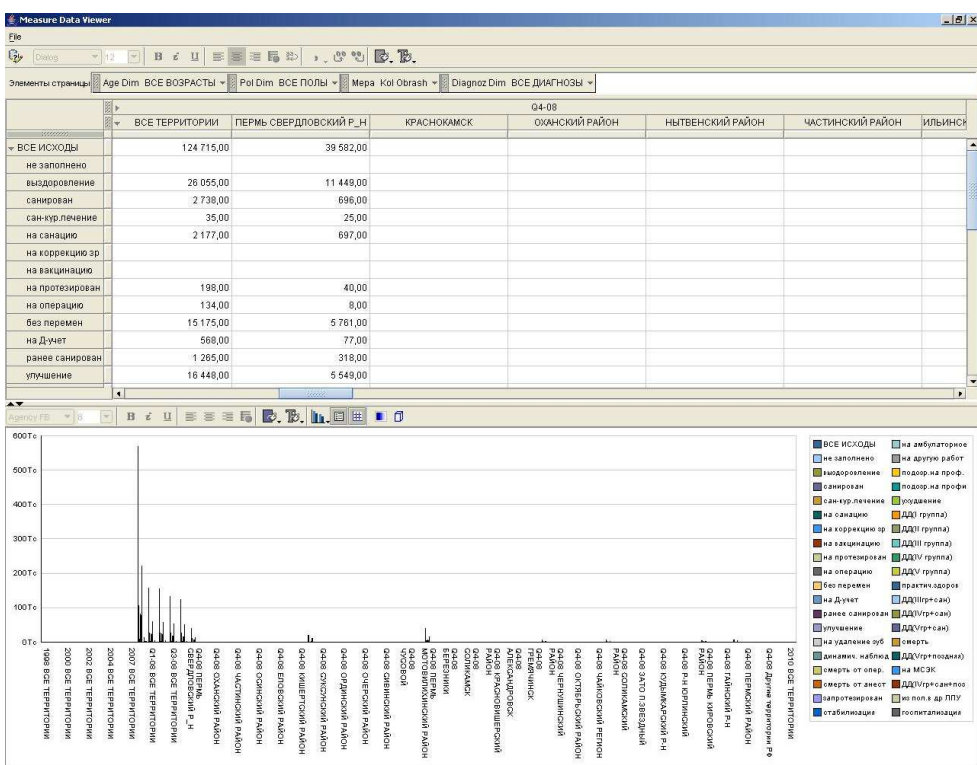


Рис. 7. Выборка количества обращений по исходу лечения, району проживания и дате с автоматическим построением диаграммы

Заключение. Oracle OLAP – встроенная в Oracle Database технология, позволяющая осуществлять быстрый доступ к агрегированным данным. Немаловажным достоинством является то, что выборка данных может выполняться даже неквалифицированными специалистами и не требует знаний SQL.

Дальнейшее развитие технологий многомерного анализа в ФГУН ФНЦ УРЗН необходимо проводить в направлении интеграции построенных и планируемых к реализации OLAP-кубов в Oracle Fusion Middleware, что позволит быстро создавать приложения, использующие OLAP-данные в расчетах, реализовать экспорт данных в MS Excel, а также в направлении исследования технологии Oracle Data Mining, позволяющей производить статистический анализ данных с построением моделей.

Библиографический список

1. Microsoft SQL Server 2005 Analysis Services. OLAP и многомерный анализ данных / А. Бергер [и др.]. – СПб.: БХВ-Петербург, 2007. – 928 с.
2. Документация по Oracle Database 10g [Электронный ресурс]. – URL: www.otn.oracle.com/documentation
3. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: БХВ-Петербург, 2004. – 336 с.
4. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.
5. Голоскоков А.Е., Савич М.В. Разработка базы знаний системы диагностирования пациента с обострением бронхиальной астмы (на примере детей младшего возраста) // Вестник НТУ «ХПИ»: сб. науч. тр. Темат. вып. «Системный анализ, управление и информационные технологии». – Харьков: Изд-во НТУ «ХПИ», 2009. – № 4. – С. 76–79.
6. Берестнева О.Г., Пеккер Я.С. Выявление скрытых закономерностей в сложных системах // Известия Томского политехнического университета. – 2009. – Т. 315. – № 5. – С. 138–143.
7. От хранения данных к управлению информацией / ред.: Г. Сомасундарам (Сому), А. Шривастава; ЕМС; пер. с англ. В. Воронинцева [и др.]. – СПб.: Питер, 2010. – 544 с.
8. Голенищев Э.П. Информационное обеспечение систем управления: учебное пособие для вузов – Ростов-н/Д: Феникс, 2010. – 315 с.
9. Кайт Т. Oracle для профессионалов: пер. с англ. – 2-е изд. – М.: Торгово-издат. дом «DiaSoft», 2004. – Кн. 1: Архитектура и основные особенности. – 662 с.

10. Oracle для профессионалов. Кн. 2. Расширение возможностей и защита / Т. Кайт. – 2-е изд. – К.; М.; СПб.: ООО «ТИД “ДС”», 2004. – 831 с.

References

1. Berger A. [et al.] Microsoft SQL Server 2005 Analysis Services. OLAP i mnogomernyi analiz dannykh [Microsoft SQL Server 2005 Analysis Services. OLAP and multidimensional data analysis]. Saint Petersburg: BKhV-Peterburg, 2007. 928 p.

2. Dokumentatsiia po Oracle Database 10g [Documentation on a Oracle Database 10g], available at: www.otn.oracle.com/documentation (accessed: 15 December 2014)

3. Barsegian A.A., Kupriianov M.S., Stepanenko V.V., Kholod I.I. Metody i modeli analiza dannykh: OLAP i Data Mining [Methods and data mining models: OLAP and Data Mining]. Saint Petersburg: BKhV-Peterburg, 2004. 336 p.

4. Barsegian A.A., Kupriianov M.S., Kholod I.I., Tess M.D., Elizarov S.I. Analiz dannykh i protsessov: uchebnoe posobie [Data analysis and processes: manual]. Saint Petersburg: BKhV-Peterburg, 2009. 512 p.

5. Goloskokov A.E., Savich M.V. Razrabotka bazy znaniia sistema diagnostirovaniia patsienta s obostreniem bronkhial'noi astmy (na primere detei mladshego vozrasta) [Development of the knowledge base of system of diagnosing of the patient with peaking of bronchial asthma (on the example of children of low age)]. *Vestnik Natsional'nogo tekhnicheskogo universiteta "Khar'kovskii Politekhnikeskii Institut": sbornik nauchnykh trudov. Tematicheskaiia vypiska "Sistemnyi analiz, upravlenie i informatsionnye tekhnologii"*. 2009, no. 4, pp. 76-79.

6. Berestneva O.G., Pekker Ia.S. Vyiavlenie skrytykh zakonomernostei v slozhnykh sistemakh [Detection of the hidden regularities in difficult systems]. *Izvestiia Tomskogo politekhnicheskogo universiteta*, 2009, vol. 315, no. 5, pp. 138-143.

7. Somasundaram (Somu) G., Shrivastava A., V. Vorotintsev. Ot khraneniia dannykh k upravleniiu informatsiei [From data storage to information management]. Saint Petersburg: Piter, 2010. 544 p.

8. Golenishchev E.P. Informatsionnoe obespechenie sistem upravleniia: uchebnoe posobie dlia vuzov [Information support of management systems: manual for higher education institutions]. Rostov na Donu: Feniks, 2010. 315 p.

9. Kait. Т. Oracle dlia professionalov: perevod s angliiskogo. Arkhitektura i osnovnye osobennosti [Oracle for professionals: translation from English. Architecture and main features]. Kiev, Moscow, Saint Petersburg: Torgovo-izdatel'skii dom DiaSoftOracle, 2004, vol. 1. 662 p.

10. Kait. Т. Oracle dlia professionalov. Kniga 2. Rasshirenie vozmozhnostei i zashchita [Oracle for professionals. Book 2. Extension of opportunities and protection]. Moscow [et al.]: Torgovo-izdatel'skii dom DiaSoftOracle, 2004, vol. 2. 831 p.

Сведения об авторах

Кокоулин Андрей Николаевич (Пермь, Россия) – кандидат технических наук, доцент кафедры автоматизации и телемеханики Пермского национального исследовательского политехнического университета (614990, г. Пермь, Комсомольский пр., 29, e-mail: a.n.kokoulin@gmail.com).

Южанинов Роман Игоревич (Пермь, Россия) – студент Пермского национального исследовательского политехнического университета (614990, г. Пермь, Комсомольский пр., 29, e-mail: sok-da-vad@mail.ru).

About the authors

Kokoulin Andrey Nikolaevich (Perm, Russian Federation) is Ph.D. in Technical Sciences, Associate Professor at the Department of Automation and Telemechanics Perm National Research Polytechnic University (614990, Perm, 29, Komsomolsky pr., e-mail: a.n.kokoulin@gmail.com).

Uzganinov Roman Igorevich (Perm, Russian Federation) is a student Perm National Research Polytechnic University (614990, Perm, 29, Komsomolsky pr., e-mail: sok-da-vad@mail.ru).

Получено: 20.02.2015