

К ПРОБЛЕМЕ ОРГАНИЗАЦИИ РЕГИСТРОВ МЕДИЦИНСКИХ ДАННЫХ: АНАЛИЗ И ОПРЕДЕЛЕНИЕ СОВРЕМЕННОГО СОСТОЯНИЯ СРЕДСТВ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ РЕГИОНАЛЬНОГО ДИСПАНСЕРНОГО МОНИТОРИНГА

© Валентин Викторович Шаповалов, Юрий Михайлович Шерстюк, Екатерина Михайловна Кучинская

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина). 197376, Россия, Санкт-Петербург, ул. Профессора Попова, д. 5

Контактная информация: Валентин Викторович Шаповалов — д. т. н., профессор, профессор кафедры биотехнических систем СПбГЭТУ «ЛЭТИ». E-mail: svv@inprosys.ru

РЕЗЮМЕ. Научная задача исследования заключается в разработке методики построения средств информационной поддержки анализа данных медико-диспансерного мониторинга популяции региона на основе применения аналитических информационных технологий. Описывается подход к использованию аналитических технологий в качестве современного средства, которое позволит осуществить аккумуляцию, агрегацию и комплексный анализ данных регионального диспансерного мониторинга из лечебно-профилактических учреждений региона. Существующая система сбора и анализа данных регионального уровня крайне неэффективна, а комплексный анализ данных регионального диспансерного мониторинга с использованием только имеющихся информационных систем регистровой направленности весьма затруднен. Низкий уровень эффективности в данном случае обусловлен противоречием между требованиями по содержанию и характером задач комплексного анализа данных РДМ, оперативности их решения, составу и форме представления необходимых для их решения данных с одной стороны, и локальным характером автоматизированного сбора и хранения информационно несогласованных данных в разнородных автономных информационных системах учетного характера — с другой. Задачу обеспечения комплексного анализа данных регионального диспансерного мониторинга на базе аналитических технологий целесообразно рассматривать как двухэтапную — первый этап можно считать этапом создания информационной поддержки комплексного анализа данных регионального диспансерного мониторинга, а второй — аналитической.

КЛЮЧЕВЫЕ СЛОВА: аналитические технологии, математические методы, OLAP система, интеллектуальный анализ данных.

ON THE ISSUE OF THE MEDICAL DATA REGISTERS' ORGANIZATION: ANALYSIS AND DETERMINATION OF THE CURRENT STATUS OF FUNDS DATA STORAGE AND PROCESSING DISPENSARY REGIONAL MONITORING

© Valentin V. Shapovalov, Yuri M. Sherstyuk, Ekaterina M. Kuchinskaya

Saint Petersburg Electrotechnical University "LETI". ul. Professora Popova 5, 197376 St. Petersburg, Russian Federation.

Contact information: Valentin Viktorovich Shapovalov — doctor of technical sciences, Professor, Department of Biotechnical Systems. E-mail: svv@inprosys.ru

ABSTRACT: The scientific purpose of the research is the development of IT support's construction methods for regional outpatient population monitoring data analysis on the basis of analytical information technologies. The approach to the using of analytical technology as a modern tool that will allow the accumulation, aggregation and analysis of regional medical institutions dispensary monitoring data is presented. The present system of data collection and analysis is extremely inefficient at the regional level and comprehensive analysis of the regional dispensary monitoring data using only existing register-oriented information systems is very difficult. The low level of efficiency in this case

due to a contradiction between the requirements on the content and nature of the problems of complex analysis for regional outpatient population monitoring data, the efficiency of their solution, the composition and presentation required for their data solutions on the one hand, and the local character of the automated collection and storage of information inconsistent data across heterogeneous autonomous character information accounting systems — on the other. The purpose of ensuring an integrated analysis of the regional dispensary monitoring data based on analytical technologies should be viewed as a two-staged process — the first step can be considered as the creation of information support of complex analysis of regional dispensary monitoring data, and the second stage is analytical phase.

KEY WORDS: Analytical techniques, mathematical methods, OLAP system, data mining.

АКТУАЛЬНОСТЬ ПРОБЛЕМЫ

Оперативный и наиболее полный анализ данных регионального диспансерного мониторинга (РДМ) возможен только при автоматизации процессов их сбора, хранения и обработки на базе применения средств вычислительной техники, современных информационных и телекоммуникационных технологий. Однако существующая система сбора и анализа данных РДМ регионального уровня даже при наличии множества эксплуатируемых информационных систем (ИС) крайне неэффективна, а комплексный анализ данных РДМ с использованием только имеющихся (или им подобных) ИС регистровой направленности весьма затруднен. Низкий уровень эффективности в данном случае обусловлен противоречием между требованиями по содержанию и характером задач комплексного анализа данных РДМ, оперативности их решения, составу и форме представления необходимых для их решения данных с одной стороны, и локальным характером автоматизированного сбора и хранения информационно несогласованных данных РДМ в разнородных автономных ИС учетного характера — с другой. Автоматизированные комплексы диспансерного обследования АКДО выполняют только 10% диспансерных обследований [1].

Аналитические технологии могут выступать в качестве современного средства, которое позволит осуществить аккумуляцию, агрегацию и комплексный анализ данных РДМ из лечебно-профилактических учреждений (ЛПУ) региона. Задачу обеспечения комплексного анализа данных РДМ на базе аналитических технологий целесообразно рассматривать как двухэтапную — на первом этапе внедрить витрину данных (ВД) и online analytical processing (OLAP), а на втором — преобразовать ВД в хранилище данных (ХД) и дополнить OLAP средствами интеллектуального анализа данных (ИАД). Учитывая, что OLAP, в отличие от ИАД, не предполагает применения математических методов содержательной обработки данных (методов решения задач анализа и прогнозирования), первый этап можно считать этапом создания информационной поддержки комплексного анализа данных РДМ, а второй — аналитической.

Совокупность указанных факторов привела к осознанию возможности принципиально нового подхода к осуществлению процессов сбора и анализа информации в системах планирования и управления, который находит свое выра-

жение в создании и применении аналитических технологий, ориентированных на решение задач поддержки принятия решений [2].

К таким задачам относятся:

1. оценка текущего и прогнозируемого состояния объекта управления и (или) среды его функционирования;
2. обнаружение и исследование скрытых закономерностей, факторов, тенденций и взаимосвязей;
3. обобщение информации как агрегация и интеграция сведений различного характера;
4. формирование альтернативных решений и выбор «оптимального» в соответствии с заданным критерием, а также результатами анализа сценариев развития ситуаций;
5. моделирование процесса эволюции состояния объекта в нестационарной неоднородной среде и т. д.

В свою очередь, повышение качества процессов обобщения и анализа информации на основе аналитических технологий и их реализующих современных средств достигается автоматизированной реализацией процессов:

1. выявления скрытых закономерностей и факторов;
2. количественной оценки факторов влияния и угроз в сложившейся ситуации;
3. использования опыта на основе автоматизированного формирования и поиска прецедентов анализируемых ситуаций в массивах ретроспективных данных;
4. высокостепенного прогнозирования эволюции состояния объекта исследования, в т. ч. выявления предпосылок к скачкообразному изменению этого состояния.

Информационная поддержка анализа данных РДМ с рядом упрощений может быть представлена как формальная система IP вида

$$IP = (I_S, I_{DM}, Q_{DM}, S_{DM}, I_{MD}, Q_{MD}, S_{MD}) \quad (1)$$

где I_S — множество БД $\{I_{S_1}, I_{S_2}, \dots, I_{S_N}\}$ ИС, эксплуатируемых в ЛПУ региона (каждая БД I_{S_i} имеет некую логическую структуру I_{SL_i} , физическую структуру I_{SP_i} и содержит данные I_{SD_i}); I_{DM} — витрина данных как реляционная база данных, имеющая логическую структуру I_{DML} , физическую структуру I_{DMP} и содержащая данные I_{DMD} ; Q_{DM} — оператор переноса данных из БД всех ИС в ВД; $I_{DMD}(t) = Q_{DM}(I_{DMD}(t_{i-1}), I_{SD}(t))$; $I_{SD} = \cup I_{SD_i}$; S_{DM} — оператор выборки данных из ВД; I_{MD} — множество многомерных кубов данных (МКД) (каждый МКД I_{MDi} имеет некую логическую структуру I_{MDLi} , физическую структуру I_{MDPi} и содер-

жит данные I_{MDDi} ; Q_{MDi} — оператор построения i -го МКД, $I_{MDDi} = Q_{MDi}(I_{DMD})$; S_{MD} — оператор выборки данных из МКД.

ЗАДАЧА ИССЛЕДОВАНИЯ

Заключается в разработке методики построения средств информационной поддержки анализа данных медико-диспансерного мониторинга популяции региона на основе применения аналитических информационных технологий.

Исходя из (1), содержательно научную задачу можно определить как разработку методики построения средств автоматизации решения задачи синтеза

$$\{I_{SLi}, I_{SPi}\} \rightarrow \{(I_{DML}, I_{DMP}), Q_{DM}, \{I_{MDL}, I_{MDP}\}, Q_{MD}\}, \quad (2)$$

причем частные задачи синтеза структуры ВД $\{I_{SLi}, I_{SPi}\} \rightarrow (I_{DML}, I_{DMP})$ и синтеза оператора переноса данных в ВД $\{I_{MDL}, I_{MDP}\} \rightarrow Q_{DM}$ должны решаться при создании ВД и при каждом изменении множества I_S , а задача синтеза структуры МКД $(I_{DML}, I_{DMP}) \rightarrow (I_{MDL}, I_{MDP})$ — каждый раз при построении i -го МКД. Поскольку в соответствии с концепцией OLAP МКД могут создаваться в произвольные моменты времени, после решения задачи синтеза структуры МКД должна решаться и задача синтеза оператора построения МКД: $\{(I_{DML}, I_{DMP}), (I_{MDL}, I_{MDP})\} \rightarrow Q_{MD}$. В основу решения последней задачи должно быть положено использование модели M_{MD} процесса перехода от реляционных данных в ВД к МКД с вычислением агрегированных значений.

Операторы S_{DM} и S_{MD} синтезу не подлежат, так как выполняются штатными средствами систем управления базами данных и OLAP.

Если через $T(Z)$ обозначить временную оценку длительности выполнения оператора Z , то от задачи синтеза (2) можно перейти к следующей оптимизационной задаче, которую должна решать разрабатываемая методика:

$$\{(I_{DML}, I_{DMP}), Q_{DM}, \{I_{MDL}, I_{MDP}\}, Q_{MD}\} \rightarrow T_{min}, \quad (3)$$

где $T = T(Q_{DM}) + T(S_{DM}) + T(Q_{MD}) + T(S_{MD})$.

С учетом (2) и (3) для решения сформулированной научной задачи исследования необходимо решить следующие частные задачи:

1. анализ особенностей синтеза структуры витрин данных, заполняемых данными из информационно несогласованных разнородных ИС учетного типа, с учетом их отображения на модель жизненного цикла ОМ;
2. разработка методики синтеза структуры витрины данных РДМ, охватывающих разные аспекты и стадии жизненного цикла ОМ;
3. разработка алгоритма скорейшего заполнения витрины данных РДМ данными из разнородных информационно несогласованных ИС;

4. разработка модели процесса построения МКД по содержанию витрины данных;
5. формирование общей методики построения средств информационной поддержки анализа данных диспансерного мониторинга популяции региона на основе применения аналитических информационных технологий;
6. оценка эффективности применения разработанной методики.

ВЫВОД

Аналитические технологии могут выступать в качестве современного средства, которое позволит осуществить аккумуляцию, агрегацию и комплексный анализ данных РДМ в Медицинский информационно-аналитический центр (МИАЦ). Задачу обеспечения комплексного анализа данных РДМ на базе аналитических технологий целесообразно рассматривать как двухэтапную — на первом этапе внедрить ВД и OLAP, а на втором — преобразовать ВД в ХД и дополнить OLAP средствами ИАД. Учитывая, что OLAP, в отличие от ИАД, не предполагает применения математических методов содержательной обработки данных (методов решения задач анализа и прогнозирования), первый этап можно считать этапом создания информационной поддержки комплексного анализа данных РДМ, а второй — аналитической.

REFERENCES

1. Voroncov I. M., Shapovalov V. V., Sherstjuk Ju. M. Zdorov'e. Sozdanie i primeneniye avtomatizirovannykh sistem dlja monitoringa i skrinirovushhej diagnostiki narusheniya zdorov'ya [Health. Development and using of automatic systems for the monitoring of screening and diagnosis of health disorders]. SPb.: OOO «IPK Kosta»; 2006 (in Russian).
2. Kopylov D. S., Shapovalov V. V., Sherstjuk Ju. M. Agregaciya dannykh avtomatizirovannykh kompleksov [Aggregation of automated data systems]. Biomedicinskaja radioelektronika. 2013; № 11: 62–4 (in Russian).

ЛИТЕРАТУРА

1. Воронцов И. М., Шаповалов В. В., Шерстюк Ю. М. Здоровье. Создание и применение автоматизированных систем для мониторинга и скринирующей диагностики нарушения здоровья. СПб.: ООО «ИПК Коста»; 2006.
2. Копылов Д. С., Шаповалов В. В., Шерстюк Ю. М. Агрегация данных автоматизированных комплексов. Биомедицинская радиоэлектроника. 2013; № 11: 62–4.

