

ФОРМИРОВАНИЕ НАБОРА ДАННЫХ ДЛЯ МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ

А.Е. Бочарова, Н.А. Воронцовская
г. Томск (Томский политехнический университет)
e-mail: aeb14@tpu.ru

PREPARATING A DATA SET FOR MEDICAL RESEARCH

A.E. Bocharova, N.A. Voronetskaya
Tomsk (Tomsk Polytechnic University)

The article focused on the data search and data analysis problems of questioning data on studying of biological markers at diseases of nervous system. The existing information system, which was created in Oracle APEX, was supplemented with a module that allows exporting data to an Excel file. It allows making the process of data analysis less time consuming.

Key words: Medical research, information system, Oracle APEX, data analysis, data export, SQL, PL/SQL.

В настоящее время наблюдается значительный рост объемов данных во всех сферах жизнедеятельности человека, что значительно усложняет процессы сбора, хранения, поиска и анализа. Данная проблема стоит и перед исследователями в области медицины

Сибирский государственный медицинский университет г. Томск (СибГМУ) проводит множество исследований на основе данных анкетирования по изучению биологических маркеров при заболеваниях нервной системы. Данный опрос производится в бумажном виде, что существенно увеличивает временные затраты на сбор, хранение, обработку данных и проведение качественного анализа, а значит, затрудняет выявление ключевых закономерностей и тенденций для борьбы с болезнями.

Ранее авторами была показана проблема учета и обработки анкетных данных, а также были рассмотрены механизмы, которые использует исследовательская группа при работе с данными о пациентах, выявили проблему и причины ее возникновения [1]. Далее спроектировали информационную систему, позволяющую автоматизировать сбор и хранение данных. В настоящее время перед авторами стоит задача рассмотрения вопроса трудоемкости поиска и анализа данных.

Основная проблема заключается в том, что количество самих анкет и количество параметров, хранимых в системе чрезвычайно велико. Поиск закономерностей часто требует детального изучения конкретных параметров, и присутствие лишних, на данном этапе, данных увеличивают временные затраты и снижают качество. Более того, чтобы выявлять тенденции необходимо визуализировать информацию, это позволит существенно снизить вероятность упущения важных фактов, что важно для такой сферы деятельности как медицина [2]. Параметры, по которым необходимо делать выборку данных, меняются в зависимости от идеи исследователя, а это значит, что поиск и анализ нужных данных является весьма трудоемким и затратным по времени процессом даже с использованием настоящей версии информационной системы.

Таким образом, возникла необходимость дополнительного модуля системы, отвечающего за экспорт данных, который позволил бы назначить необходимые атрибуты и условия поиска, осуществить выборку и сформировать файл в удобном для анализа формате. Таким форматом для группы исследователей СибГМУ является формат электронных таблиц Excel. Формирования файла такого типа позволит осуществлять необходимый уровень визуализации, и удобство представления данных [3].

При реализации данного модуля мы столкнулись с техническими проблемами:

1. Необходимость создания универсального механизма, который позволил бы производить выборку по любым желаемым атрибутам с любым желаемым количеством условий

отбора. Решить задачу используя только SQL не представляется возможным, поэтому в ходе работы был использован процедурный язык PL/SQL, позволяющий создать процедуру, которая динамически сгенерирует SQL запрос, а значит количество атрибутов и количество условий перестало нас волновать. Далее для формирования запроса необходимо передать в процедуру все необходимые параметры.

2. Необходимость передачи неопределенного количества атрибутов для выборки и неопределенного количества разнотипных параметров для условий отбора от приложения к базе. Решение этой задачи возможно следующими способами:

- с помощью формирования двух строк на стороне приложения при фиксации параметров, первая из которых содержала бы список выбранных атрибутов, разделенные огороженным разделителем, вторая – список выбранных параметров и их значений для условий выборки в указанном виде. Далее эти строки передавались бы процедуре в качестве входных параметров.

- с помощью формирования строки во временной таблице для каждого выбранного атрибута или условия на стороне приложения при фиксации параметров. Временная таблица содержала бы id, колонку, которая отображала бы тип этого параметра (атрибут или условие), имя таблицы, имя параметра и его значение, если это условие.

Первый способ имеет такие недостатки как: во-первых, перегрузка процедуры, потому что необходимо разобрать сформированные строки, прежде чем динамически формировать запрос, а во-вторых, в ходе формирования и парсинга строк возможны ошибки, которые пришлось бы обрабатывать отдельно. Поэтому более приемлемым является второй способ.

Таким образом, для модуля «Экспорт» на стороне базы данных была создана временная таблица, содержащая вышеописанные атрибуты, и хранимая процедура, динамически формирующая запрос. На стороне приложения была реализована форма, содержащая две области параметров: атрибуты выборки и параметры условий с их значениями, кнопку, позволяющую зафиксировать выбранные параметры и осуществить вставку во временную таблицу, а также кнопка формирования и выведения файла средствами Oracle APEX на основе сформированного хранимой процедурой запроса.

Систематизация и анализ данных в любых сферах деятельности в настоящий момент является трудоемким процессом [4], и разработанный модуль экспорта позволит исследователям значительно снизить временные затраты для подготовки данных к анализу и сделать сам процесс выявления закономерностей гораздо более удобным и менее трудоемким.

ЛИТЕРАТУРА

1. Бочарова А.Е., Воронцовская Н.А. Проблемы учета и обработки анкетных данных медицинских исследований // Информационные технологии в науке, управлении, социальной сфере и медицине. Сборник научных трудов II Международной конференции. Национальный исследовательский Томский политехнический университет. — 2015. — С. 846-847.
2. Марухина О.В., Мокина Е.Е., Берестнева Е.В., Применение методов Data mining для выявления скрытых закономерностей в задачах анализа медицинских данных // Фундаментальные исследования. — 2015. — № 4-0. — С. 107-113.
3. Мокина Е.Е., Марухина О.В., Шагарова М.Д. Подходы к разработке информационной системы поддержки формирования документов при оказании высокотехнологичной медицинской помощи // Фундаментальные исследования. — №2-9, — 2015. — С. 1857-1861.
4. Мещеряков Р.В., Балацкая Л.Н., Чойнзонов Е.Л. Специализированная информационная система поддержки деятельности медицинского учреждения // Информационно-управляющие системы. — 2012. — № 5. — С. 51-56.