

Toolbox, включенный в программный комплекс Matlab, при котором на вход подавался вектор из 30 составляющих и классифицирующий полученные образы фонемно.

#### ЛИТЕРАТУРА

1. *Алдошина И.А.* Основы психоакустики // *Звукорежиссер*. 2004. № 1-10.
2. SPESOM-2006: тр. науч. конф. / СПИРАС при РАН РФ. СПб., 2006.

3. *Бекешу.* Проблемы физиологии слуха // *Успехи физических наук*. 1982. Т. XV.
4. *Baumgarde F., Ferekidis C., Fuchs H.* A Nonlinear Psychoacoustic Model Applied to the ISO MPEG Layer 3 Coder. <http://www.int.uni-hannover.de/project/coding/audio/perception/publications.html>. 1995.
5. *Robinson D.J.M., Hawksford M.O.J.* Psychoacoustic models and non-linear human hearing // *Proceedings of IEEE*. 2000. V. 88. № 4.
6. *Rabiner L., Jang B.* Fundamentals of speech recognition. CLSU, 1993.

Поступила в редакцию 17 октября 2006 г.

## КОНТРОЛЬ КАЧЕСТВА ДАННЫХ ПРИ ПОСТРОЕНИИ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ

© А.А. Ильин

Ilyin A.A. Quality control of data at formation of an information analytical system/

Большинство потребителей информационных технологий оперируют с большим объемом данных, которые необходимо анализировать. На сегодняшний день разработан ряд программных средств, предназначенных для облегчения задачи анализа информации. Одним из классов таких программных средств являются информационно-аналитические системы. За последние десятилетия были разработаны методологии построения систем этого класса. Разработанные методики способствовали решению ряда важных задач, таких как быстрый доступ к необходимой информации и возможность наиболее полного анализа данных. При этом одним из основных факторов риска, определяющим успешность проекта по созданию информационно-аналитической системы, является проблема качества данных [1].

Многие исследователи определяют качественную информацию, как обладающую определенным набором свойств. Наиболее полный список свойств, характеризующих качественную информацию, для систем поддержки принятия решений приводится в работе [2]:

- **Корректность.** То есть все значения, содержащиеся в информационно-аналитической системе, являются достоверными и безошибочными.

- **Недвусмысленность.** То есть любая запрошенная информация должна иметь единственное значение, так чтобы она не могла быть истолкована различными пользователями по-разному.

- **Согласованность.** То есть вся информация должна соответствовать единой нотации.

- **Полнота.** Существуют два аспекта полноты:
  1. Обеспечение того, чтобы все необходимые величины содержали непустые значения.
  2. Обеспечение контроля попадания в информационно-аналитическую систему всех необходимых записей.

Была поставлена задача разработать методологию контроля качества данных, удовлетворяющую следующим требованиям:

1. контроль качества данных на всех стадиях создания информационно-аналитической системы;

2. возможность настройки производительности системы;

3. использование архитектурных особенностей информационно-аналитических систем.

В результате были определены пять типов проверок качества данных, требующихся при разработке информационно-аналитических систем, и предложена методика применения данных проверок, позволяющая реализовать систему контроля качества данных, удовлетворяющую всем сформулированным требованиям:

1. контроль значений колонок таблицы;
2. контроль наличия в таблице всех необходимых значений;
3. контроль дубликатов;
4. контроль правила «трех сигм»;
5. контроль качества информации с помощью прогнозирования.

Также была поставлена задача разработать программное средство, позволяющее контролировать качество данных в рамках данной методологии.

На рис. 1 представлена архитектура разработанного программного комплекса, осуществляющего контроль качества данных. Программный комплекс состоит из трех частей: 1) средство определения правил проверки качества данных; 2) инструмент ETL (Extract Transform Load) – используется для наполнения информационно-аналитической системы данными; 3) серверная часть системы, осуществляющая проверку качества данных.

Разработчик определяет ограничения, накладываемые на данные, и сохраняет их в текстовом файле. Затем добавляет в процедуры ETL компоненты проверки качества данных, в качестве параметра которых указывает имя файла, содержащего ограничения. При выполнении процедур ETL происходит вызов необходимых проверок качества данных, результат выполнения которых определяет последующие действия процедур ETL (продолжение загрузки данных либо предупреждение администратора системы о проблемах с качеством данных).

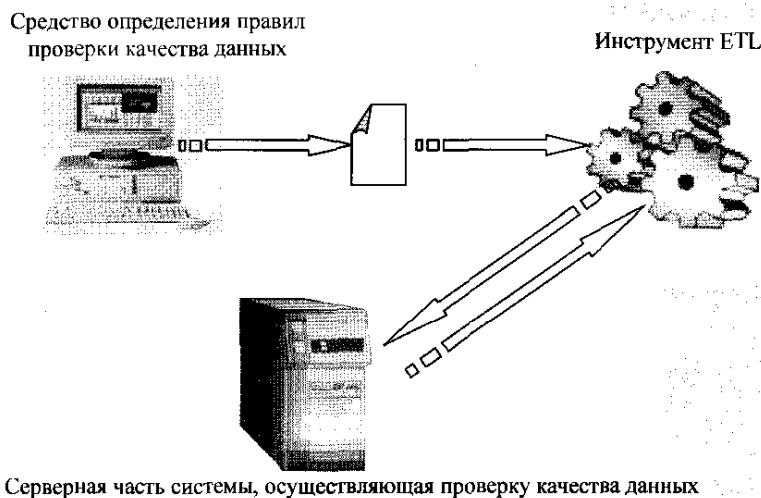


Рис. 1. Архитектура разработанного программного комплекса

Для осуществления взаимодействия между ETL-средством и серверной частью системы был выбран один из универсальных способов интеграции нескольких приложений – web-сервисы, т. к. технология web-сервисов поддерживается большинством ETL-средств.

Таким образом, был разработан и реализован программный комплекс, позволяющий производить контроль качества данных в рамках разработанной методики. Реализована возможность определения пяти типов правил контроля качества информации, которые могут быть применены на любом этапе построения информационно-аналитической системы. При этом множество типов проверок качества данных может быть расширено. Для этого новый вид проверок должен быть реализован в виде web-сервиса, принимаю-

щего в качестве входного параметра имя файла, содержащего ограничения, накладываемые на анализируемую информацию. Реализованный таким образом новый вид проверок может быть использован аналогично тому, как было описано выше.

#### ЛИТЕРАТУРА

1. Olson J. Data Quality Accuracy Dimension. Morgan Kauffmann Publishers, 2003. 293 p.
2. Kimball R., Caserta J. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Confirming and Delivering Data. Wiley, 2004. 525 p.

Поступила в редакцию 20 октября 2006 г.

## МЕТОДЫ МОДЕЛИРОВАНИЯ СОЦИАЛЬНЫХ СИСТЕМ

© О.А. Соломина

Solomina O.A. The methods of the social systems modeling.

В последние годы быстрыми темпами развивается идея использования моделирования для исследования социальных процессов. Применение моделей позволяет проводить контролируемые эксперименты в ситуациях – весьма характерных для социального познания – где экспериментирование на реальных объектах является практически невозможным или по каким-то причинам (экономическим, нравственным и т. д.) нецелесообразным.

Целью данной работы является проведение сравнительного анализа методов математического и компьютерного моделирования в социальных системах.

Методология построения моделей социальных явлений существенно отличается от методологии моделирования в естественных науках. В последних, как

правило, основу составляет набор твердо установленных экспериментом и практикой зависимостей (законов), которые, будучи однажды найденными, всегда остаются справедливыми в границах своей применимости.

В социальных науках ситуация намного сложнее. При построении моделей социальных явлений необходимо учитывать изначальную неточность задания всех данных, отсутствие четкого математического описания переменных и параметров, используемых при моделировании, социально-психологические факторы (такие, как соотношение личных и групповых интересов, особенности индивидуальной и национальной психологии при принятии решений и др.).