

## **Реферат**

магистерской аттестационной работы на тему:

«Разработка и реализация метаданных для структуризации хранилища данных»

Воробьевой Елены Ивановны

### **Актуальность работы**

Популярность хранилищ данных в последние годы существенно возросла. Конкуренспособные организации находятся на пути построения хранилищ данных, либо расширения, перепроектирования и усовершенствования уже имеющихся. Метаданные считаются ключевым фактором успеха в проектах по внедрению хранилищ. Они содержат всю информацию, необходимую для извлечения, преобразования и загрузки данных из исходных систем, а также для последующего использования и интерпретации содержимого хранилищ данных.

### **Цель работы**

Целью работы являлось создание программного продукта для структуризации хранилища экспериментальных данных и обеспечения применения стандартных методов современной технологии Data Mining с использованием среды разработки Delphi 7. Рассмотреть основные форматы отображения метаданных, такие как HDF, NetCDF, FITS и стандартные форматы метаданных. Для структурирования хранилища данных использовать XML- документ формата метаданных. Апробировать на примере хранилища данных, полученных при экспериментальных исследованиях характеристик плазменного источника ультрафиолета.

### **Задачи, которые решались в работе**

Для достижения целей в работе решались следующие задачи:

- обзор существующих баз данных
- структура хранилищ данных
- data mining и его представление с помощью языка XML разметки
- основные форматы отображения метаданных
- создание программного продукта, для структуризации хранилища данных

## **Достигнутые результаты**

Достигнув результата в решении поставленных задач автор защищает:

- анализ развития баз данных
- преимущества языка XML разметки для структуризации хранилищ данных
- анализ и перспективы развития хранилищ данных
- анализ форматов данных для отображения метаданных
- создание программного продукта для структуризации хранилища данных.

## **Научная новизна работы**

Научной новизной работы есть анализ программного обеспечения для структуризации хранилищ данных и перспективы развития технологий для B2B приложений. Это дало возможность сделать выводы о перспективе развития метаданных и хранилищ данных.

Так же сюда можно отнести разработанный программный продукт для структуризации хранилища данных.

## **Практическая ценность работы**

На основании сделанного анализа был создан программный продукт для структуризации хранилища экспериментальных данных и обеспечения применения стандартных методов современной технологии Data Mining с использованием среды разработки Delphi 7.

## **Выводы**

На сегодняшний день объемы информации, которые хранятся на серверах баз данных измеряются терабайтами. Очень мало компаний, предприятий интересуются «сырыми данными». Гораздо важнее получать знания. Таким образом, необходимо, чтоб разработчики ПО поставили перед собою задачу встроить в сервера баз данных методы интеллектуального анализа данных (а лучше комбинацию из нескольких методов, в зависимости от задач), чтоб в режиме реального времени можно было проводить анализ данных, и при изменении или добавлении начальной информации пользователям тут же выводились новые знания полученный из этого. Именно таким образом пользователь

сможет перестать думать о физической структуре таблиц и колонок. Для работы с хранилищами данных разработчики ПО должны предложить новую технологию, которая в отличие от классических BI-систем, в которых входящие данные сначала загружаются в хранилище, очищаются и агрегируются и только после этого становятся доступны для исследования при помощи созданной заранее многомерной модели, позволит начать обработку данных уже в момент их поступления. Эта технология должна будет пропускать входящие данные через систему фильтров, выделяя только значимую для конкретного бизнес-процесса информацию. В результате общий объем хранимой исторической информации сократится без потери качества. Было бы хорошо, если бы технология могла одновременно анализировать данные из большого числа источников, реализуя технологию комплексного анализа событий (Complex Event Processing, CEP).

Работа содержит 110 стр., 22 рис., 1 табл., 20 источников.

**Ключевые слова:** DATA MINING, ХРАНИЛИЩА ДАННЫХ, ЯЗЫК  
РАЗМЕТКИ XML, БАЗЫ ДАННЫХ