

Воробьева Е.И. — рецензент Корначевский Я.И.
УНК “ИПСА” НТУУ “КПИ”

Реализация и разработка метаданных для структуризации хранилища данных

Развитие методов записи и хранения данных привело к бурному росту объемов собираемой и анализируемой информации. Для того, чтобы провести автоматический анализ данных, используется Data Mining.

Data Mining содержат форматированную информацию о БД: имена параметров, коды, свойства, форматы, адреса хранения и т.п. – все, что необходимо информационным технологиям, чтобы найти информационные ресурсы и правильно их обработать. Data Mining позволяют осуществлять поиск исходных данных на нескольких уровнях их физической организации в зависимости от квалификации пользователей, их прав, используемых инструментальных средств поиска информации и т.п. [1].

Верхний уровень – самый доступный, рассчитанный на широкого пользователя. Здесь находятся общие сведения о БД, имеющихся в различных организациях страны и за рубежом; сведения об организациях; проектах; экспертах; информационных ресурсах, имеющиеся на Web и др. При числе документов более 10 по каждому типу Data Mining можно использовать динамические средства организации доступа к информации с использованием языка XML. При этом появляется возможность стандартизации представления и поиска Data Mining с использованием Web – технологий [2].

Второй уровень Data Mining рассчитан на пользователей – специалистов, которые хотят оценить количество, качество, полноту той или иной БД. Поэтому здесь должны использоваться базы Data Mining, представляющие сведения о платформах, методах измерений, используемых приборах, подробные каталоги изученности того или иного района и т.п. Этот уровень Data Mining создается в среде СУБД с возможностями выдачи информации на Web в виде динамических страниц [2].

Третий уровень Data Mining предназначен для поиска исходной информации в базах данных и выдачи данных в удобной для профессионального пользователя форме. Эти Data Mining в основном предназначены для администратора БД. Создаваемые здесь базы Data Mining позволят по логическим характеристикам данных найти их физические адреса хранения или процедуры расчета. Этот уровень Data Mining создается в среде СУБД [2].

Существует общая тенденция к стандартизации форматов хранения метаданных для упрощения интеграции систем разных производителей. На сегодняшний день самым популярным форматом хранения метаданных является XML.

В ходе работы предложена разработка интеллектуального программного модуля на базе XML, который позволит обеспечить структуризацию хранилища экспериментальных данных и обеспечение применения стандартных методов технологии Data Mining.

Литература

1. Дэвид Сеппа XML. М.: Русская редакция, 2003. – 640 с.
2. Сахил Малик XML для профессионалов. СПб.: БХВ-Петербург, 2006. – 560 с.