

ПРИМЕНЕНИЕ XSLT ПРЕОБРАЗОВАНИЙ В АНАЛИТИЧЕСКОЙ СУБД

Д.Д. ЧЕРКАСОВ¹, О.В. ЗАБРОДИН²

¹Санкт-Петербургский государственный политехнический университет
ул. Политехническая, 5, 195251, г. Санкт-Петербург, Россия
daniil-cherkasov@mail.ru

²Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
ул. Профессора Попова, 5, 197376, г. Санкт-Петербург, Россия
zolegzabrodin@gmail.ru

Рассматривается применение языка преобразований XSLT в аналитике СУБД в контексте стандарта SQL/MM. Выполнен анализ реализации функционала аналитической подсистемы стандарта SQL/MM с использованием XML технологии. Показана структура типов аналитической подсистемы, определяющая интерфейс взаимодействия с моделью.

Ключевые слова: извлечение знаний, аналитическая платформа, унифицированный интерфейс.

Введение

Тенденция последних лет в развитии интеллектуальных систем заключается в интеграции средств аналитической обработки, алгоритмов извлечения знаний, управления метаданными и визуализации результатов на одной программной аналитической платформе. Проведенный анализ показал, что стандарт SQL/MM [1], согласованный с форматом представления моделей извлечения данных PMML [2], достаточен для разработки архитектуры аналитической платформы и удовлетворяет основным требованиям: организация процесса аналитической обработки данных, унификация интерфейсов, единое соглашение по хранению и передачи моделей DataMining.

Начало XML технологии было положено в 1996 году. Консорциум всемирной паутины (W3C) обозначил задачу создания такого языка разметки, который обеспечивал стандартный способ структурирования информации, т.е. отделение данных от их представления и передачу ее из одной программы в другую. Такой язык получил название extensible Markup Language (XML) – расширяемый язык разметки. Спустя некоторое время возник вопрос о возможности преобразования данных представленных в виде XML-документа, эта предпосылка определила язык преобразований XSLT, что означает eXtensible Stylesheet Language: Transformations (расширяемый язык таблиц стилей: Преобразования).

Предусмотренное стандартом SQL/MM представление алгоритмов в виде PMML моделей основано на языке разметки XML. Такое решение позволяет интегрировать в систему аналитики новые разработки при сохранении действующего функционала.

Основными методами стандарта SQL/MM, выполняют функции преобразования одного XML документа в другой без вычислительных процессов, и XSLT – идеальный инструмент для этой работы. Рассмотрим пример выполнения метода DM_getClasTask() класса DM_ClasModel.

Описание метода:

1) Результат выполнения метода *DM_getClasTask()* определяется следующим образом:

а) Если *DM_ClasModel* не содержит значение *DM_ClasBldTask* которое было использовано для построения модели тогда возвращается значение null.

b) В противном случае возвращается значение *DM_ClasBldTask* которое было использовано для вычисления модели.

Согласно описанию метода входной XML-документ должен быть представлен типом *DM_ClasModel*, который должен соответствовать заданной XMLсхеме:

1) *DM_ClasBldTask* является абстракцией всей информации, которая определяет задачу классификации, в частности содержит входные данные для обучения и валидации (*trainData*, *validationData*) и параметрические настройки (*settings*)

2) Тег *PMML* содержит *PMML*-модель. Применение стандарта *PMML* дает целый ряд важных технологических и организационных преимуществ для нейросетевой технологии.

3) Атрибуты и пара тегов: *RankQuality*, *PredAccuracy* заданные в схеме являются качественными характеристиками построенной модели.

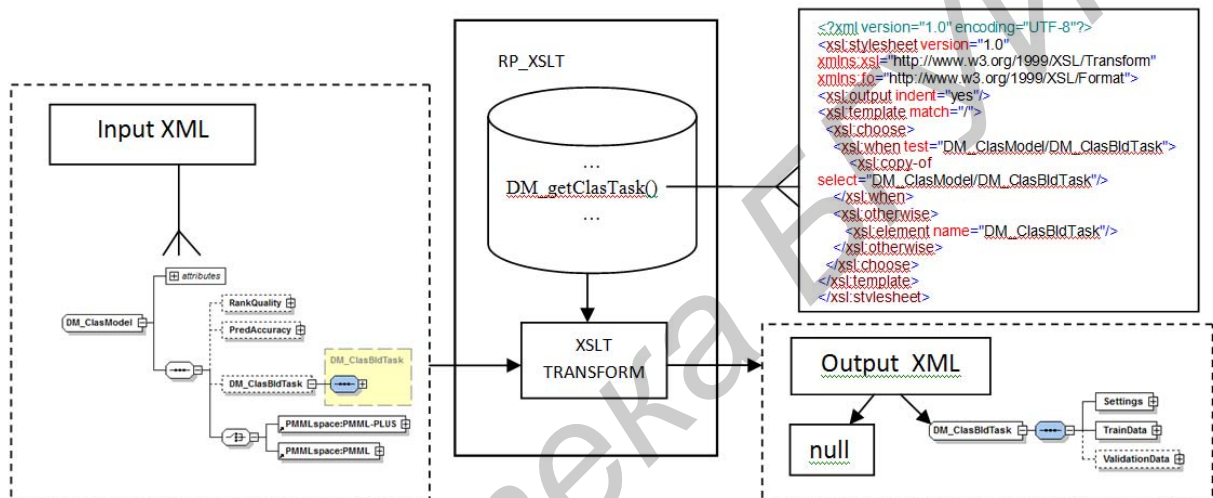


Рис. 1. Структурная схема XSLT преобразования в рамках SQL/MM стандарта

Структурная схема XSLT преобразования (рис. 1) определяет процесс XSLT преобразования в контексте стандарта SQL/MM. Состоит из трех блоков: входной XML-документ; функциональный блок XSLT преобразования; выходной XML-документ. Основным блоком данной схемы является функциональный блок и представляет из себя две составные части: системное хранилище, поддерживающее инфраструктуру аналитической подсистемы и XSLT преобразователь или XSLT-процессор. Выходной XML-документ определяется двумя возможными случаями, описанными в методе, и может возвращать значение null или *DM_ClasBldTask*.

Подобным образом реализуются все функции стандарта SQL/MM.

Список литературы

1. ISO/IEC 13249-6-2006, SQL/MM Part 6.
2. PMML Version 4.1, 2012, Data Mining Group (DMG) <http://www.dmg.org/>