

РАЗРАБОТКА ПОДСИСТЕМЫ ХРАНЕНИЯ XML-ДАННЫХ СЕРВЕРА XSS

В современных информационных системах широко применяется XML представление данных. Одним из достоинств XML формата данных является то, что это самоописываемый формат данных, что значительно облегчает работу с ним. Популярны ранние форматы, такие как файлы с разделяемыми запятой значениями (CSV), или подходят для табличных данных и плохо описывают псевдоструктурированные данные, или, как RTF, слишком специализированы для псевдоструктурированных текстовых документов. Также стоит отметить наличие развитых инструментальных средств для работы с XML данными, поддерживающих языки запросов XPath и преобразования XSLT.

В рамках формата XML возможно представление как линейных, так и иерархических наборов данных. XML формат характеризуется гибкостью, открытостью, расширяемостью, платформо-независимостью и поддерживает локализацию, так как полностью совместим с Unicode. Поскольку XML — это текстовый формат представления данных, то при возникновении необходимости XML-документы можно читать и редактировать, используя стандартные инструменты редактирования текстов. С помощью XML формата можно представлять данные практически любой предметной области, используя стандартные языковые средства.

Но XML имеет ряд недостатков. Один из них - избыточность данных, увеличивающая объем, и сложность процесса разбора, что накладывает жесткие требования на аппаратные ресурсы и не позволяет использовать XML представление при работе с большими объемами данных. В W3C нотации XML определены ссылки XLink и XPointer, позволяющие на основании нескольких файлов строить единый XML документ и переиспользовать элементы. Но существующие реализации ссылок основаны на полном разборе и построении всех документов одновременно, что не позволяет получить выигрыш производительности при разбиении больших документов на множество маленьких.

В рамках разрабатываемой подсистемы хранения (ПСХ), в составе проекта XSS фирмы ООО «Деловые Консультации СПб», было предложено реализовать итеративную подгрузку XML документов по мере обращения к ним, кэширование уже подгруженных документов в общем дереве (ОД), с организацией механизмов выполнения XPath запросов над ним. Таким образом, ПСХ решает следующие задачи:

- Хранение множества XML документов
- Обработка ссылок между документами с поддержкой
 - Итеративной загрузки частей XML-документа при обращении к линкованным элементам (Итеративный XLink);
 - Поддержки работы с общим деревом посредством XPath.
- Реализация функций для работы с XML документами
 - get_elem, get_subtree, get_sublist;
 - set_elem, set_subtree;
 - create_elem, create_subtree;
 - delete_elem, delete_subtree.

Архитектура подсистемы хранения приведена на рис. 1.

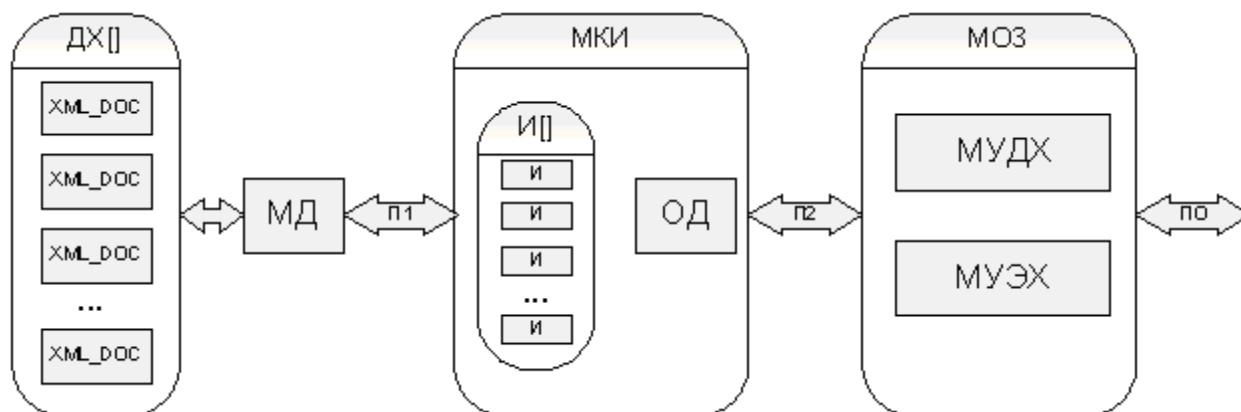


Рис. 1. Архитектура подсистемы хранения

Подсистема состоит из трех модулей:

- МД (модуль доступа, обеспечивает взаимодействие с физическим хранилищем XML документов);
- МКИ (модуль кэширования и индексирования, обеспечивает контроль ссылочной целостности, контроль типов, построение общего дерева);
- МОЗ (модуль обработки входящего запроса по протоколу взаимодействия ПО (объединенный протокол), в соответствии с поступающими командами выполняет последовательности элементарных команд в терминах МКИ, МОЗ состоит из следующих элементов: модуль запроса/ответа – МЗО, модуль управления деревьями хранения – МУДХ, модуль управления элементами хранения - МУЭХ).

ПСХ решает задачи хранения XML деревьев. В качестве физического хранилища могут выступать РБД или файловая система. На уровне ПСХ реализуется функциональность разбора ссылок и построения общего XML дерева, над которым реализуется механизм обеспечения ссылочной целостности данных, операции создания, удаления, чтения и коррекции над элементами и деревом элементов, выполнение XPath запроса над поддеревом элементов.

ПСХ может использоваться как непосредственный источник XML данных (локальный вариант), так и источник данных для ПСМД (централизованного многопользовательского хранения). ПСХ обеспечивает взаимодействие по ПО для использования локальной и централизованной версии XSS без модификации взаимодействующих подсистем.

На данный момент, система находится на стадии тестирования макета с ограниченной функциональностью.