

ВЪЗМОЖНОСТИ ЗА ПРИЛОЖЕНИЕ НА СЪВРЕМЕННИ XML БАЗИРАНИ ТЕХНОЛОГИИ В МИННАТА ПРОМИШЛЕНОСТ

Николай Янев¹

Минно-геоложки университет "Св. Иван Рилски", 1700 София, E-mail niki@mgu.bg

РЕЗЮМЕ. През последните 10 години огромна популярност придобиха XML и базираните на него технологии. Те се наложиха в различни сфери на индустрията, включително и в минното дело. Минния софтуер в България изостава от тази тенденция. Целта на статията е да представи XML базирани технологии и приложения използвани за целите минната промишленост.

OPPORTUNITIES FOR IMPLEMENTATION OF MODERN XML BASED TECHNOLOGIES IN THE MINING INDUSTRY

*Nikolay Yanev*¹

¹ *University of Mining and Geology "St. Ivan Rilski", 1700 Sofia, e-mail: niki@mgu.bg*

ABSTRACT. For the recent 10 years XML and the based on it technologies have become enormous popularity. They have affirmed themselves and have been applied in many sectors of industries including mining. The mining software in Bulgaria lags behind from these tendencies. The main aim of the current paper is to present the XML based technologies and applications for the use purposes of the mining industry.

Въведение

Три са направленията в които през последните години XML се наложи:

- Инструмент за разработване на описателни езици;
- Средство за реализиране на web услуги;
- Стандарт за пренос и съхранение на данни.

Разработване на описателни езици

За целите на минната индустрия се използват следните XML базирани езици:

Geography Markup Language (GML) - XML базиран стандарт използван за предаване и съхраняване на географска информация. Той съдържа универсален набор от средства за описание на пространствено положение на обекти с оглед на техните топологически отношения. Пространственото положение на големи обекти на цифрова топографическа карта може да се представи във вид на полигон, линейни обекти или точки.[1] Тази спецификация дефинира XML Schema синтаксис [2][3], чиито механизми и конвенции са:

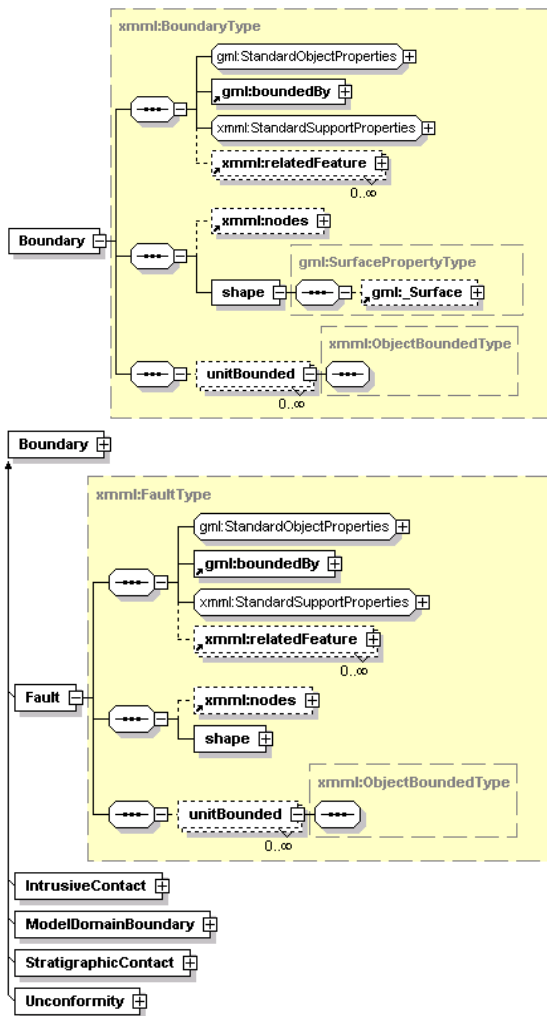
- Осигуряват отворена и независима рамка за дефиниране на географски схеми и обекти.
- Поддържат описание на географските схеми за специализирани области.
- Позволяват създаване и управление на свързани

географски схеми и набори от данни.

- Предават и съхраняват географските схеми и набори от данни.
- Увеличават способността на организациите да споделят географска информация.

Както при XML, GML представя географската информация в текстов вид, което позволява съхраняването му в традиционна база от данни. GML Стандарта се развива, като актуалната версия на GML е 3-та. GML може да се импортира и експортира от и в програми като AutoCAD Map 3D.

XMML - Въпреки, че GML е разработен за работа с географски данни, то той не е особено подходящ за представяне на геоложки данни, тъй като не поддържа специфичните за сондирането характеристики. XMML [4] е приложна схема на GML. XMML използва основните характеристики на GML като добавя допълнителни, които са специфични за геологията. На практика това се постига чрез добавяне на подходящи геометрични и не геометрични свойства към GML схемата.



Фиг. 1 – XMML Геоложки граници

International Rock Excavation Data Exchange Standard (IREDES) е индустриален стандарт за обмен на данни между минно оборудване (машини) и компютърна система [5]. За първи път е представен на петата конференция ISMMA and Telemip през 1999, като в момента се поддържа официално от един от най-големите производители на изкопно оборудване Atlas Copco.

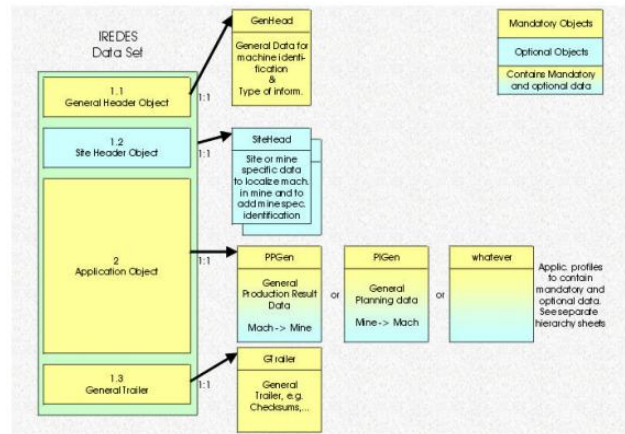
IREDES дефинира общ електронен език за комуникация между изкопните машини и централен компютър или друго оборудване. Той се базира на XML и използва DTD за описанието на данните.

Информацията обменена с IREDES се структурира в „набор от данни“ (Data Sets), като един набор от данни представлява цялата информация предадена наведнъж (фиг. 2).

Използването на стандарти като IREDES стандарт има редица важни предимства:

- Съществено намаляване на разходите за разработване, използване и поддръжка на софтуер и оборудване;
- Създава предпоставки за концентрация на усилията върху развитието на машини функции и използването на информацията, вместо постоянно да се преработва интерфейса;

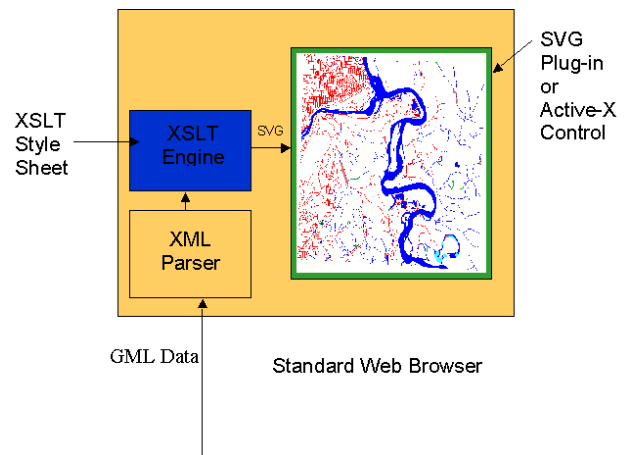
- IREDES се основава на установени международни стандарти;
- Използването на IREDES средства осигурява безпроблемно интегриране на машини от различни доставчици в корпоративната ИТ инфраструктура.



Фиг. 2 - Структура на IREDES Data Set

Scalable Vector Graphics (SVG) е XML базиран език за описание на двуизмерни графични данни във векторен формат, които са готови за публикуване в Интернет или други приложения. SVG използва няколко прости форми, при комбинирането на които се получават сложни фигури [6].

Geographic Information System (Географска информационна система) има много специфични изисквания: богати графични характеристики, поддръжка на векторно и растрено съдържание и възможност за манипулиране на голямо количество данни. SVG предлага добри възможности за този пазар и много GIS системи осигуряват SVG експорт. SVG е идеално допълнение към GML формата на OpenGIS консорциум. Чрез използване на XML базирани инструменти, като XSLT, сравнително лесно GML документ може да бъде конвертиран в SVG (фиг. 3).



Фиг. 3 – Конвертиране GML-SVG

Keyhole Markup Language (KML) – XML базиран описателен език за представяне на тримерни, гео-

пространствени данни в Google Earth и частично в Google Maps [7].

KML използва таг базирана структура, с вложени елементи и атрибути. KML файлът се обработва от Google Earth по начин, подобен на обработването на HTML и XML файлове от web браузърите. Така Google Earth действа като браузър за KML файлове.

KML е международен стандарт поддържан от Open Geospatial Consortium

XML бази данни

Повечето сървъри за бази от данни се развиха по направление, което доведе до сблъсък с XML. XML накара доставчиците на софтуер за бази от данни да преосмислят фундаментално стратегиите си. Водещите доставчици на СУБД въведоха XML технологии в своите продукти. Oracle базира решенията си на XDB, Microsoft на SQL XML, IBM предлага XML Extender за своята DB2.

Преходът към XML е тясно свързан с новия подход при електронната обработка на информацията - внедряване на Web програми с възможности за управление на бази от данни, работещи на сървъри за приложения и заместващи самостоятелни продукти. XML е в основата на повечето web услуги.

XML документите могат да бъдат съхранявани в текстови файлове, XML хранилища, или бази от данни. Има две основни причини, поради които много компании предлагат те да бъдат съхранявани в бази от данни [8]:

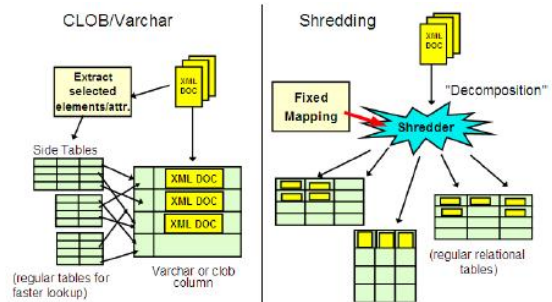
- Управлението на големи количества XML данни е проблем, характерен за базите от данни. XML данните са данни както всички останали, само че в различен цялостен формат. Причините за съхраняване на релационни данни в бази от данни се отнасят и за XML данните: базите от данни предлагат ефективно търсене и извличане, добра поддръжка за устойчивост на данните, архивиране и възстановяване, поддръжка на транзакции, производителност и скалируемост.
- Интеграция: съхранявайки релационни данни и XML документи заедно, можете да интегрирате нови XML данни във вече съществуващи релационни данни и да комбинирате SQL с XPath или XQuery в една заявка. Освен това, релационните данни могат да бъдат публикувани като XML и обратно. Посредством интеграция, базите от данни могат да осигуряват по-добра поддръжка на Уеб приложения, SOA и Уеб услуги.

Обикновено в XML формат се съхраняват консервативни данни, като например описание на минни машини.

Три са основните типа XML бази данни.

XML Enabled Database - XML-enabled базите от данни използват релационния модел за същинското съхранение на данните. Това изисква средства за регламентиране на съответствие между XML (йерархичният модел на данните) и релационния модел на данните или, иначе

казано, съхраняване на XML данни като голям символен обект. Макар това да се счита за остаряла технология, все още се използва от много производители на бази от данни. Фигура 1.7 обяснява по-подробно двата варианта на XML-enabled бази от данни.



Фиг. 4 - Варианта на XML-enabled бази от данни

Лявата страна на фигурата показва метода "CLOB и Varchar" за съхраняване на XML документи в база от данни. Използвайки този метод, XML документ се съхранява като непрекъснат низ или като CLOB (Character large object), или като varchar в базата от данни. Ако XML документът се съхранява като низ, то когато искаме да извлечем част от XML документа, програмата ни трябва да извлече низа и да извърши парсване, за да открие, каквото ни трябва. Методът не е особено гъвкав.

Другият вариант на XML-enabled бази от данни се нарича разделяне или декомпозиция и е илюстриран в дясната част на фигура 4. При този метод, XML документът се разделя на части, които се съхраняват в таблици. Освен това йерархичният модел на XML документа се преобразува в релационен модел. Това също води до ограничаване на гъвкавостта: промяна в XML документа трудно се отразява в съответните таблици и може да се наложи създаването на много нови таблици. Този метод също не е добър и откъм бърздействие: ако ни се наложи да възстановим оригиналния XML документ, трябва да извършим скъпа SQL операция, която става дори още по-скъпа, когато трябва да се обединят повече таблици;

Native XML бази от данни - Native XML базите от данни използват йерархичния XML модел за вътрешно съхраняване и обработка на XML. Форматът при съхраняване е същият като този при обработка: няма преобразуване към релационен модел и XML документите не се съхраняват като изображения. Когато се използват XPath или XQuery изрази, те се обработват вътрешно (natively) от механизма и не се преобразуват в SQL. Затова тези бази от данни се наричат native XML бази от данни;

Hybrid XML Database (HXD) - при тях данните могат да бъдат третирани и като Native XML Database и като XML Enabled Database.

Езици за XML данни

За извличане на XML данни се използва езика XQuery. Той се базира на системата типове XML Schema и е съвместим със стандартите XSL и XPath, свързани с XML. XQuery бе възприет и от водещи компании като Oracle, IBM и Microsoft. XQuery позволява формирането на изрази за

обхождане на йерархическата структура на XML документа, от друга страна „чистия“ SQL (без XML разширение) не допуска такава възможност. XQuery поддържа както типизирани, така и не типизирани данни, а за данните в SQL винаги се определя конкретен тип. В XQuery няма нулеви стойности – в XML документите липсващите или неизвестните данни се пропускат, а SQL използва null за представянето на такива данни. XQuery връща резултат в XML формат. SQL връща резултат, състоящ се от различни типове данни [9].

В отговор на очакванията на потребителите за добавяне в SQL на възможности за обработка на XML данни в стандарта SQL 2006 бе добавено разширение SQL/XML за обработване на XML документи. В таблица 1 са представени някои от SQL/XML функциите, включени в стандарта SQL 2006.

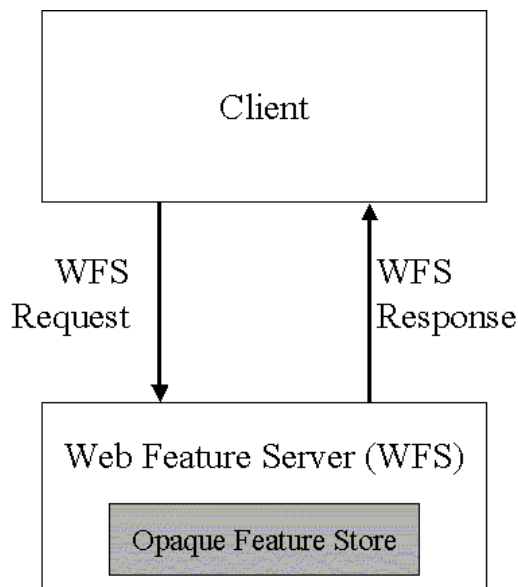
Таблица 1

Функция	Описание
XMLPARSE	Парсва бинарни данни от тип character или large object и връща резултат от тип XML
XMLSERIALIZE	Превръща данни от тип XML в бинарни данни от тип character или large object
XMLVALIDATE	Валидира данните от тип XML спрямо дадена XML схема
XMLEXISTS	Определя дали дадена Xquery заявка връща резултат (например, последователност от един или повече елемента)
XMLQUERY	Изпълнява XQuery заявката и връща резултатната последователност
XMLTABLE	Изпълнява XQuery заявката и връща резултатната последователност като релационна таблица, ако е възможно
XMLCAST	Преобразува в или от тип XML

Web услуги и протоколи

Simple Object Access Protocol (SOAP) - XML е полезен с универсалността си и с това, че чрез него различни компютърни системи могат да споделят данни. SOAP е XML базиран протокол, който предоставя достъп до отдалечени обекти (и техните методи), Web services и сървъри. Чрез SOAP е възможно едно приложение да извиква метод, който да се изпълни на отдалечен сървър, и да получи резултата, като информацията (във вид на XML) се предава по HTTP или друг протокол. Предимството е, че това са отворени и свободни за използване технологии. Простотата и липсата на ограничения правят SOAP предпочитан и полезен. С него все по-широка е и употребата на Web services. Така, чрез SOAP, XML осигурява по-лесен обмен на данни и споделяне на ресурси.[10]

Web Feature Service (WFS) е услуга която позволява публикуването и визуализирането на географски обекти в Интернет. Обменът на данни между потребителя и сървъра се осигурява посредством GML [11].



Фиг. 5 - Web Feature Service

Изискванията за WFS са:

1. Интерфейсите трябва да бъдат XML базирани.
2. GML трябва да се използва за описание на функциите в интерфейса.
3. Най-малко един WFS трябва да бъде в състояние да представи функции чрез GML.
4. Предикатите или филтриращия език трябва да бъдат описани в XML и да произлизат от CQL, както е определено в OpenGIS Catalogue Interface Implementation Specification.
5. Хранилищата за данни, използвани за съхранение на географски характеристики, трябва да са непрозрачни за клиентските приложения и да бъдат разглеждани единствено чрез интерфейса на WFS.
6. Използване на подмножество на XPath изрази.

1.0.0 версия на спецификацията WFS изисква използването на GML версия 2.1.2, а 1.1.0 версия на спецификацията на WFS изисква използването на GML версия 3.1.1.

Други приложения

RDF (Resource Description Framework) – Разработка на W3C. RDF е модел на метаданни и често се използва при съхранение и транспортиране на части от документа отнасящи се към символното му множество, езика и др. В много случаи RDF разчита на XML стандарта [12]. Използва Uniform Resource Identifier (URI) за маркиране на обекти и предикати.

За запис и предаване използва няколко формата, в това число:

- RDF/XML — запис във вид на XML документ;
- RDF/JSON — запис във вид на JSON данни;
- RDFa (*RDF in attributes*) — запис като вътрешни атрибути на HTML или XHTML документ.

PMML (Predictive Model Markup Language) - XML базиран език за описание на статистически, data mining модели и изкуствени невронни мрежи. Разработката и внедряването на езика се осъществява от IT консорциума Data Mining Group [13].

PMML предоставя възможност за различни видове преобразуване на данни като нормиране, дискредитация, създаване на карта на стойности и др.

Сред основните предимства на езика са:

- Улеснения обмен на данни и модели данни;
- Достъпен е – в основата на езика е XML стандарта;
- Намалява продължителността на работния цикъл на статистическия модел;
- Лесен за внедряване;
- Позволява осъществяване на DM в самото хранилище на данните.

PMML се използва в продукти като DB2, SPSS, Zementis ADAPA и др.

Geographically Encoded Objects for RSS (**Географски Кодирани Обекти за RSS**)

Geographically Encoded Objects for RSS осигурява начин за добавяне на географски данни към RSS или ATOM емисия използвайки GML или опростени GeoRSS стандарти [14]. В GeoRSS, местоположението се състои от географски точки, линии и полигони. GeoRSS бюлетините са проектирани да бъдат консумирани от географската карта на софтуера, като генератори. Има и специални програми за създаване на GeoRSS feed.

*Препоръчана за публикуване от
Катедра «Информатика», МЕМФ*

В Picasa и Flickr GeoRSS са вградени, а изображенията се кодират с помощта на маркер в географска карта. Всеки албум или емисия има GeoRSS feed за публикуване в Google Maps и KML емисия за публикуване в Google Earth.

GeoRSS е имплементиран и в WordPress. Използва се плъгин за GeoPress. С негова помощ може да се добавя Google Maps директно във всеки пост с една команда както и да се създава динамичен маркер за всяка публикация. Плъгина е достъпен за сваляне в GeoRSS.

Литература

- <http://www.opengis.net/gml/>
- <http://schemas.opengis.net/gml/>
- <http://schemas.opengis.net/gmlsfProfile/2.0/>
- <http://xmml.arcc.csiro.au/>
- <https://www.iredes.org/>
- <http://www.w3.org/Graphics/SVG/>
- <http://code.google.com/apis/kml/documentation/>
- Чонг, Р., И. Хейкс, Р. Хужа. 2008. Въведение в DB2 Express-C, IBM Corporation
- <http://www.w3.org/TR/xquery/>
- <http://www.w3.org/TR/soap/>
- <http://cite.opengeospatial.org/OGCTestData/wfs/1.0.0/specs/wfs/1.0.0/02-058.html>
- <http://www.w3.org/RDF/>
- <http://www.dmg.org/>
- <http://www.georss.org>
- Иванов, К., Д. Белчевски. 2004. Съвременни методи за обмен на данни в минната промишленост, *Минно-геоложки университет "Св. Иван Рилски", Годишник, том 47, свитък III, Механизация, електрификация и автоматизация на мините*, С., 125-130