

Język XML w aplikacjach z bazami danych — po roku

Tomasz Traczyk
ttraczyk@ia.pw.edu.pl



Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

Język XML w aplikacjach z bazami danych — po roku

- Wprowadzenie
- Przypomnienie — co jest XML
- Rozwój składników XML
- Przetwarzanie dokumentów w XML
- Zastosowania XML
- XML a Oracle
- Podsumowanie



Język XML w aplikacjach z bazami danych

2

Wprowadzenie

Co było przed rokiem

- Spore nadzieje
- Pierwsze wersje specyfikacji
- Eksperymentalne narzędzia
- Pierwsze zastosowania
- Brak narzędzi dla baz danych
- Brak wsparcia przez Oracle

Co się zdarzyło przez ten rok

- „Moda” na XML
- Postęp w pracach standaryzacyjnych
- Pierwsze standardy *de facto*
- Pierwsze powszechnie dostępne aplikacje i przeglądarki
- Nowe propozycje wzbogacające język
- Wiele nowych profesjonalnych zastosowań
- Pierwsze narzędzia dla baz danych
- Pierwsze narzędzia w ofercie Oracle

- Pierwsze próby wykorzystania XML na WEITI PW
- Referat na PLOUG

- Działająca aplikacja XML w systemie wspomagania dziekanatu WEITI PW
- Kontynuacja referatu na PLOUG



Język XML w aplikacjach z bazami danych

3

Wprowadzenie

- Wprowadzenie
- Przypomnienie — co to jest XML
- Rozwój składników XML
- Przetwarzanie dokumentów w XML
- Zastosowania XML
- XML a Oracle
- Podsumowanie



Język XML w aplikacjach z bazami danych

4

Przypomnienie — co to jest XML

Co to jest XML

(*Extensible Markup Language*)?

- Metajęzyk
 - umożliwia definiowanie języków znakowania
 - służy do zapisu różnorodnych dokumentów i struktur danych
 - przeznaczony m.in. do użycia w WWW
- Znakowanie z znaczeniowe
- Do formatowania wizualnego — *style-sheets*

Co zapewnia XML?

- Możliwość definiowania
 - struktury dokumentu
 - znaczników (*tags*)
- Łatwość interpretacji dokumentu
 - przez przeglądarki
 - nawet bez definicji struktury dokumentu
- Możliwość tworzenia połączeń (*links*)
 - między dokumentami
 - przez sieć

Geneza XML

- Podzbiór SGML (*Standard Generalised Markup Language*)
- Uproszczony przez usunięcie trudnych i niekoniecznych konstrukcji



Język XML w aplikacjach z bazami danych

5

Przykład dokumentu w XML

```
<?xml version="1.0" encoding="ISO-8859-2"?>
<!DOCTYPE eres_konspekty SYSTEM "konspekty.dtd">
<?xml-stylesheet type="text/xsl" href="konspekty.xsl"?>
<eres_konspekty>
  <przedmiot id="KBD2" wersja="1">
    <slowo_kluczowe>bazy danych</slowo_kluczowe>
    <slowo_kluczowe>Oracle</slowo_kluczowe>
    <konspekt>
      <czesc_konspektu id="Streszczenie">
        <P> Monograficzny przedmiot poświęcony bazie danych i
        narzędziom Oracle. </P>
      </czesc_konspektu>
      <czesc_konspektu id="Treść przedmiotu">
        <P> Omawiane są podstawowe zagadnienia związane z wykorzystaniem
        RDBMS Oracle7 i <I>Oracle8</I> oraz administrowaniem nimi.</P>
        <P> Przedstawiane są także narzędzia do budowy aplikacji: </P>
        <UL>
          <LI> Oracle Forms, </LI>
          <LI> Oracle Reports. </LI>
        </UL>
      </czesc_konspektu>
    </konspekt>
  </przedmiot>
</eres_konspekty>
```



Język XML w aplikacjach z bazami danych

6

Przykład DTD

```
<!ELEMENT eres_konsepky |przedmiot|+ >
<!ELEMENT przedmiot |{slowo_kluczowe}*, konspekt| >
<ATTLIST przedmiot id ID #REQUIRED
wersja CDATA #IMPLIED >
<!ELEMENT slowo_kluczowe |#PCDATA| >
<!ELEMENT konspekt |czesc_konspektu|+ >
<!ELEMENT czesc_konspektu |P|UL|+ >
<ATTLIST czesc_konspektu id ID #REQUIRED >
<!ELEMENT P |#PCDATA|I|* >
<!ELEMENT I |#PCDATA| >
<!ELEMENT UL |LI|+ >
<!ELEMENT LI |#PCDATA| >
```

Poprawność dokumentu w XML

W sensie well-formed

- Warunki
 - niepusła zawartość
 - wszystkie konstrukcje poprawne w sensie well-formed (wg specyfikacji języka)
 - wszystkie konstrukcje rozpoczęte muszą być jawnie zakończone
 - każda konstrukcja zawarta w innej musi być w niej zawarta w całości
 - tylko jedna konstrukcja (element główny) nie zawarta w żadnej innej
 - brak odwołań do niestandardowych definicji nie zawartych w tym dokumencie
- Nie musi mieć DTD (dokument *standalone*)
- Mimo braku DTD może być wyświetlany przez przeglądarki
- Przeznaczenie: tylko wyświetlanie

W sensie valid

- Warunki
 - jest well-formed
 - powiązany z istniejącym DTD
 - pełna zgodność zawartości z DTD
 - wszystkie konstrukcje poprawne w sensie *valid* (wg specyfikacji języka)
- Przeznaczenie: przetwarzanie
- Do przetwarzania na ogół niezbędny jest dostęp do DTD

- Wprowadzenie
- Przypomnienie — co to jest XML
- Rozwój składników XML
- Przetwarzanie dokumentów w XML
- Zastosowania XML
- XML a Oracle
- Podsumowanie

Rozwój składników XML

- „Klasyczne” składniki
 - dokument
 - DTD (*Document Type Definition*)
 - *style-sheets* (XSL)
 - łączniki
- Nowe składniki
 - przestrzenie nazw
 - schematy
 - *data islands*

Łączniki

XLL — łączniki między dokumentami

- Łączniki — znaczniki zawierające atrybut `xml:link`
- Lokatory: adresy URI (*Uniform Resource Identifier*):
 - adres sieciowy
 - zapytanie
 - identyfikator fragmentu (XPointer)
- Możliwości łączenia
 - proste — jak A z HTML
 - złożone (*extended*), np. z wieloma lokatorami
- Zachowania (*behaviors*) sterowane atrybutami
 - `show: replace|embed|new`
 - `activate: user|auto`
- Specyfikacja XLL (XLink) jeszcze nie całkiem gotowa
- Brak standardu *de facto*

XPointer — adresowanie w dokumencie

- Adresowanie oparte na drzewie elementów
- W dokumencie nie są konieczne żadne specjalne oznaczenia
- Przykłady:
 - `root||.child|1,przedmiot|`
– pierwszy element typu `przedmiot`
 - `id|KBD2|.child|2,#element|`
– drugi element podrzędny elementu o identyfikatorze `KBD2`
- Specyfikacja jeszcze nie całkiem gotowa
- Wygląda na to, że standard *de facto* będzie zawierał elementy XQL (*XML Query Language*)

Przestrzenie nazw (*namespaces*)

Problemy

- Tworzenie standardowych słowników znaczników
 - uniwersalnych
 - dziedzinowych
- Uniknięcie konfliktów nazw
- Umożliwienie użycia standardowych słowników w nowo tworzonej definicjach dokumentów

Rozwiązanie

- Użycie przestrzeni nazw
 - nazwy znaczników i atrybutów poprzedzone prefiksami
 - prefiksy określone z powołaniem się na URI domeny zarządzającej słownikiem
- Przykład: zastosowanie przestrzeni nazw `eres`

```
<eres:konsepky xmlns:eres="http://www.elka.pw.edu.pl/eres">
<eres:konsepky>
  <eres:slowo_kluczowe>
    ...
```


DOM

Problem

- Jak łatwo i w standardowy sposób przetwarzać dokumenty XML w programach

Rozwiązanie

- Standardowe API do typowych języków programowania
- Propozycje:
 - SAX (*Simple API for XML*)
 - DOM (*Document Object Model*)

DOM

- Specyfikacja analogiczna jak dla HTML
- Obiektowy model dostępu do struktury dokumentu
- Zbiór klas i metod do manipulacji dokumentami
- Dostępny dla języków Java, ECMAScript (JavaScript), VBScript, C++
- Wbudowany w przeglądarkę MSIE 5.0

- Wprowadzenie
- Przypomnienie — co to jest XML
- Rozwój składników XML
- Przetwarzanie dokumentów w XML
- Zastosowania XML
- XML a Oracle
- Podsumowanie

Podstawowe narzędzia

Przeglądarki

- Microsoft Internet Explorer 5.0 zawiera
 - generyczny format do wyświetlania XML bez arkuszy stylizacyjnych
 - walidację dokumentów w oparciu o DTD i schematy (wg specyfikacji XML-Data)
 - obsługę XSL z wzorcami rozszerzonymi wg specyfikacji XQL
 - API zgodny z DOM
 - *XML Data Islands* z dostępem z DHTML

- Istnieje spore prawdopodobieństwo, że MSIE wyznaczy standardy *de facto*

- Konkurencja też pracuje...

Edytory

- Microsoft XML Notepad
 - prosty edytor do dokumentów XML
 - zawiera sprawdzanie poprawności w sensie *valid*
 - związany z MSIE 4 lub 5
- Pojawiają się narzędzia typu *shareware/freeware*



Wykorzystanie w programach powszechnego użytku

MS Office 2000

- HTML równorzędny formatem zapisu dokumentów
 - zapisu dokumentów w HTML
 - możliwość ponownego odwzorzenia dokumentów z HTML bez utraty informacji
- Konieczne uzupełnienie HTML o brakujące informacje
 - wyspy danych w XML
 - dodatkowe pliki z grafiką i z danymi w XML
 - grafika wektorowa w VML
- Możliwości
 - bezpośredniej prezentacji w WWW (MSIE 5)
 - dalszej edycji dokumentów dystrybuowanych przez WWW
 - odczytu przez inne oprogramowanie
 - odczytu dokumentów z nowszych wersji przez wersję w cześniejszej

VML (*Vector Markup Language*)

- Służy do definiowania i prezentowania grafiki wektorowej
- Zapisywane są w nim rysunki OfficeArt z MS Office 2000
- Przeglądarka MSIE 5.0 potrafi bezpośrednio wyświetlać rysunki VML

SVG (*Scalable Vector Graphics*)

- Standard proponowany przez W3C
- Scala kilka propozycji specyfikacji grafiki wektorowej
- Prace standaryzacyjne w toku

Specjalizowane struktury danych

Motywacja

- Tworzenie specjalnych formatów dziedzinowych
- Umożliwienie łatwej wymiany danych

Stan obecny

- Bardzo wiele propozycji
- W większości dziedzin jeszcze brak powszechnie uznanych standardów

Przykłady

- Zastosowania naukowe, np.
 - MathML (*Mathematical Markup Language*)
 - CML (*Chemical Markup Language*)
- Modelowanie systemów, np.
 - PIF-XML (*Process Interchange Format*)
 - UXF (*UML eXchange Format*)
 - XMI (*XML Metadata Interchange*)
- EDI (*Electronic Data Interchange*)
 - liczne projekty
- Finanse i bankowość, np.
 - OFX/OFE (*Open Financial Exchange*)
 - BIPS (*Bank Internet Payment System*)
- Multimedia, np.
 - SMIL (*Synchronized Multimedia Integration Language*)
 - PGML (*Precision Graphics Markup Language*)
 - VML

XML w systemach z bazami danych

Stan

- XML zauważony przez producentów systemów baz danych
- Pojawiają się
 - narzędzia wspomagające
 - specjalizowane serwery informacyjne i aplikacyjne

Przykłady

- eXcelon firmy Object Design
 - serwer aplikacyjny
 - przetwarza dokumenty XML na struktury obiektowej bazy danych i odwrotnie
 - wykorzystuje XQL
- Tamino firmy Software AG
 - serwer informacyjny dla zastosowań internetowych
 - X-Machine — składnica danych w XML, przechowywanych bez konwersji
 - obsługa XML w środowisku serwera aplikacji Bolero
 - generowanie Mas na podstawie XML
 - interfejs zgodny z SAX

- Wprowadzenie
- Przypomnienie — co to jest XML
- Rozwój składników XML
- Przetwarzanie dokumentów w XML
- Zastosowania XML
- XML a Oracle
- Podsumowanie

25

PLSXML

Co to jest

- Zestaw pakietów PL/SQL
- Przeznaczony dla
 - Oracle w. 7+
 - OWA Server. 2.1+
 - MSIE 5.0
- Zawiera
 - pakiety DBXML, DBDOM, DBXSL
 - przykłady

DBXML

- Formatuje wyniki zapytania SQL w postaci XML
- Procedura *Query*
 - prosty interfejs
 - za pytanie SQL przekazywane jako tekst
 - ograniczone możliwości
- Procedura *Get*
 - złożone wywołanie
 - znacznie większe możliwości

DBDOM

- Programowe tworzenie dokumentów w XML
- Rozbór (*parsing*) dokumentów XML
- Zapis/odczyt dokumentów do/z
 - pliku
 - zmiennej znakowej
 - strumienia HTTP
- Wyszukiwanie prostych wzorców (typu *XSL pattern*) w dokumencie XML
- Brak sprawdzania poprawności (nawet *well-formed*)

DBXSL

- Dynamiczne generowanie arkusza stylizacyjnego
- Prezentacja wyników zapytania w postaci rozwijalnego drzewa

26

Oracle Core XML Support

Co to jest

- Podstawowe mechanizmy obsługi XML dla Oracle8i
- Wykorzystują wbudowaną w serwer *BI* maszynę wirtualną *Oracle Java VM*
- Mogą być uruchamiane na serwerze bazy danych lub na OWS

Składniki

- *Oracle XML Parser*
 - zgodny z DOM i SAX
 - dostępny w 3 warstwach architektury
 - można używać w procedurach składowanych napisanych w Javie
- Wsparcie dla XML w opcji IFS (*Internet File System*)
 - automatyczna strukturalizacja dokumentów ładowanych do IFS
- Mechanizmy wyszukiwania w XML dostępne w opcji *Context*

Narzędzia dodatkowe

- Parsery dla języków PL/SQL, C, C++
- *XML Class Generator for Java*
 - generuje definicje klas do manipulowania dokumentem na podstawie DTD
- *XML SQL Utility for Java*
 - zestaw gotowych klas
 - do generowania dokumentów w XML na podstawie zapytań w SQL
 - do wczytywania dokumentów w XML do struktur bazy danych
- *Oracle XSQL Servlet*
 - umożliwia wstawianie zapytań SQL wprost do dokumentów w XML
 - dynamicznie wykonuje te zapytania
 - transformuje dokumenty w XML na podstawie skryptów w XSL

27

- Wprowadzenie
- Przypomnienie — co to jest XML
- Rozwój składników XML
- Przetwarzanie dokumentów w XML
- Zastosowania XML
- XML a Oracle
- Podsumowanie

28

Stan prac nad XML

- Gotowe specyfikacje (*W3C Recommendation*):
 - XML
 - XML Namespaces
 - DOM
- Zaawansowane prace (*W3C Working Drafts*):
 - XSL
 - XLL
 - XPointer
- Inne specyfikacje w fazie roboczej (*W3C Note*)
- Duże szanse narzucenia standardów *de facto* przez producentów przeglądarek
- Pierwsze narzędzia wspomagające wykorzystanie XML w systemach z bazami danych

29

Perspektywy rozwoju XML

- Świetlane ☺
- Bardzo duże zainteresowanie
- Rzeczywiste wsparcie ze strony znaczących producentów oprogramowania, np. Microsoft i Oracle
- Pewny kierunek rozwoju aplikacji inter- i intranetowych
- Niemal pewny znaczny wpływ na sposób projektowania systemów informacyjnych
- Prawdopodobny duży wpływ na rozwój systemów baz danych

30

Język XML w aplikacjach z bazami danych — po roku

- Wprowadzenie
 - Przypomnienie — co to jest XML
 - Rozwój składników XML
 - Przetwarzanie dokumentów w XML
 - Zastosowania XML
 - XML a Oracle
 - Podsumowanie
-

