



eXtensible Markup Language, cz. 1

Marcin Gryszkalis, mg@fork.pl

Plan wykładu

- Wprowadzenie: historia rozwoju technik znakowania tekstu
- Motywacje dla prac nad XML-em
- Podstawowe koncepcje XML-a
- XML jako metajęzyk
- Podstawy składni XML-a
- XML a HTML
- Zastosowania XML-a
- Źródła informacji

XML 1 - Historia technik znakowania tekstu

- Prehistoria – znakowanie tekstu
 - ręczna redakcja
 - umówiony sposób oznaczania
 - wspólny język redaktora i zecera

XML 1 - Historia technik znakowania tekstu

- Znakowanie w dobie komputerów
 - rozdzielenie treści, znaczników i wyniku
 - uzgodniony (mniej lub bardziej precyzyjny) język znaczników
 - automatyczne przetwarzanie

XML 1 - Historia technik znakowania tekstu

- Przykładowe języki znaczników
 - RTF (Rich Text Format)
 - `{\b pogrubienie}`
 - PostScript
 - `(pogrubienie)W`
 - TeX/LaTeX
 - `\textbf{pogrubienie}`
 - HTML
 - `pogrubienie`
 - Wikipedia
 - `"pogrubienie"`

XML 1 - Historia technik znakowania tekstu

- Korzenie
 - Lata 60 dwudziestego wieku
 - idea rozdzielenia informacji od formatu (1967)
 - propozycja użycia uniwersalnych znaczników
 - projekt GenCode
 - projekt INTIME (IBM)
 - GML – Generalized Markup Language (IBM, 1969)

XML 1 – Motywacja – Programy i ich formaty

- Każda aplikacja wprowadza swój wewnętrzny format
- Nowe wersje aplikacji wprowadzają zmiany do formatu
 - problem kompatybilności wstecznej
 - typowo brak możliwości zapisu w starszym formacie
- Aplikacje dostarczają konwerterów
 - tylko do najpopularniejszych formatów
 - konwersja może powodować utratę części informacji

XML 1 – Motywacja – Standardy

- Często brak uznanych standardów
- Pod-standardy i pseudo-standardy w różnych dziedzinach
 - dokumenty biurowe – Microsoft Word
 - teksty naukowe – PostScript, TeX
 - internet – HTML
 - dane finansowe – Quicken Interchange Format
- Standard musi być
 - własnością publiczną
 - otwarty i jawny
 - niezależny od konkretnego producenta oprogramowania

XML 1 – Motywacja – Potrzeba struktury

- Cyfrowa informacja wymusza potrzebę istnienia struktury
 - Jeden format nie wystarczy dla wszystkich
 - Ale zginiemy wśród milionów niekompatybilnych formatów
- Historia wymiany danych
 - człowiek -> człowiek
 - człowiek -> aplikacja
 - aplikacja -> człowiek
 - aplikacja -> aplikacja

XML 1 - Historia technik znakowania tekstu

- Rozwój języków uogólnionego znakowania tekstu
 - 1969 – GML (Generalized Markup Language)
 - 1978 – ANSI rozpoczyna pracę nad normą
 - 1983 – SGML (Standard Generalized Markup Language), wersja 6
 - 1985 – SGML standardem Komisji UE i US DoD
 - 1986 – SGML jako norma ISO 8879:1986
 - 1991 – początek WWW
 - 1994 – HTML 2.0 jako aplikacja SGML
 - 1998 – XML 1.0 – eXtensible Markup Language (W3C)

XML 1 - W3C

- World Wide Web Consortium (W3C)
 - założone w 1994 – Tim Berners-Lee (MIT) we współpracy z CERN, DARPA i Komisją Europejską
 - Zrzesza kilkaset organizacji
 - Przygotowuje rozwiązania wspierające organizację dostępu do zasobów internetu
 - Publikuje oficjalne rekomendacje (standardy)
 - HTML
 - CSS
 - XML
 - WAI
 - SVG, RDF, WSDL, X*
 - Współpracuje z IETF, Unicode Consortium,

XML 1 – Idea SGML-a / XML-a

- Oddzielenie znaczenia tekstu od sposobu prezentacji

<osoba>Hamlet</osoba>

<wypowiedź>Być albo nie być.</wypowiedź>

XML 1 – Idea SGML-a / XML-a

- `<osoba>Hamlet</osoba>`
`<wypowiedź>Być albo nie być.</wypowiedź>`
- osoba = pogrubienie, do lewej
- wypowiedź = kursywa, wcięcie, do lewej

- efekt:

Hamlet

Być albo nie być.

XML 1 – Idea SGML-a / XML-a

- Stworzenie najodpowiedniejszej struktury do konkretnego zastosowania

```
<główny_bohater>Hamlet</główny_bohater>  
<wypowiedź>  
  <głośno>być</głośno>  
  <cicho>albo nie być</cicho>  
</wypowiedź>
```

XML 1 – Idea XML-a

- Stworzenie najodpowiedniejszej struktury do konkretnego zastosowania – przykłady
 - słownik - <słowo>, <wymowa>, <etymologia>
 - prawo - <rocznik>, <nr_akt>, <teza>, <artykuł>
 - ubezpieczenia - <nr_polisy>, <wartość_polisy>, <zniżka>
 - itd

XML 1 – XML jako metajęzyk

- Nie możemy mieć jednego języka
- Możemy mieć wspólny metajęzyk – język tworzący języki
 - znana składnia, gramatyka
 - znana metodologia
 - wspólne narzędzia
- Dowolnie wiele języków do specyficznych zastosowań

XML 1 – co to jest XML

- XML to *nie* jest język programowania
- XML to sposób zapamiętywania danych wraz z ich strukturą w dokumencie tekstowym
 - otwarty
 - elastyczny
 - bezpłatny
 - niezależny (od dostawcy, od platformy)
- XML to rama składniowa do tworzenia języków specyficznych dla konkretnych zastosowań (tzw. aplikacje XML)

XML 1 – jak wygląda XML

```
<?xml version="1.0" encoding="ISO-8859-2"?>
```

```
<OSOBA>
```

```
  <IMIE> Jan </IMIE>
```

```
  <DRUGIEIMIE/>
```

```
  <NAZWISKO> Kowalski </NAZWISKO>
```

```
  <URODZONY>
```

```
    <DATA format="dd.mm.rrrr">
```

```
      12.04.1980
```

```
    </DATA>
```

```
  <MIEJSCE>
```

```
    <MIASTO> Warszawa </MIASTO>
```

```
  </MIEJSCE>
```

```
</URODZONY>
```

```
<PESEL>80041201322</PESEL>
```

```
<UWAGI><!-- tu wpisz uwagi -->
```

```
</UWAGI>
```

```
</OSOBA>
```

- deklaracja XML
- element
- atrybut
- znacznik początkowy
- znacznik końcowy
- element pusty
- treść (tekst)
- komentarz

XML 1 – cechy XML-a

- Język tekstowy (w sensie Unikodu)
 - `<Данные>żółć</Данные>`
- Opisuje dane hierarchiczne
- Ścisłe reguły składniowe dają możliwość wydajnej obsługi
- Pozwala na automatyczne testowanie poprawności
- Niezależny od platformy
- Bardzo szerokie wsparcie (narzędzia, dostawcy)
- “Rozwlekła” składnia
- Brak wbudowanej definicji typów
- Samodokumentujący się (czasem tak - czasem nie)

XML – HTML a XML

- HTML
 - znaczenie elementów z góry określone
 - interpretacja określona standardem (w praktyce przez przeglądarki)
 - to co “poprawne” faktycznie też określone przez przeglądarki
- XML
 - znaczenie elementów określa użytkownik i/lub aplikacja
 - <p> może oznaczać akapit (*paragraph*), pomoc, poprawkę
 - poprawność ściśle określona przez specyfikację

XML 1 – zastosowania XML-a

- Zarządzanie dokumentami, treścią, wiedzą:
 - oryginalne zastosowanie SGML-a
 - dokumenty tworzone przez człowieka, dla człowieka (lub aplikacji)
- Elektroniczna wymiana danych, integracja aplikacji
 - nowe zastosowanie XML-a
 - dokumenty tworzone i przetwarzane w pełni automatycznie
 - dokumenty tworzone tylko na czas komunikacji
 - potrzeba dokładnej weryfikacji struktury i zawartości

XML 1 – zastosowania XML-a

- dokument tekstowy

```
<zeznanie>Wypadek dnia  
<data>01.01.2007</data> o  
godzinie  
<godzina>13:13</godzina>  
(<dzień>piątek</dzień>) miał  
miejsce nie z mojej winy.  
<poszkodowany>Alojzy  
N.</poszkodowany> ble ble  
ble...</zeznanie>
```

- baza danych

```
<zamówienie>  
<zamawiający id="3">  
  <imię>Adam</imię>  
  <nazwisko>Kowalski</nazwisko>  
</zamawiający>  
<specyfikacja>  
  <pozycja>  
    <nazwa>hamburger</nazwa>  
    <ilość>2</ilość>  
  </pozycja>  
</specyfikacja>  
</zamówienie>
```

XML 1 – odnośniki

- Konsorcjum W3C: <http://w3.org>
- Rekomendacje XML:
 - 1.0 <http://www.w3.org/TR/2006/REC-xml-20060816/>
 - 1.1 <http://www.w3.org/TR/2006/REC-xml11-20060816/>
- Wikipedia:
 - <http://en.wikipedia.org/wiki/XML>
 - <http://en.wikipedia.org/wiki/Category:XML>
 - http://en.wikipedia.org/wiki/Markup_language
- Portal O'Reilly <http://www.xml.com>