



eXtensible Markup Language, cz. 2

Marcin Gryszkalis, mg@fork.pl

XML 2 - Poprawność składniowa

- Dokument XML musi być poprawny pod względem składniowym (syntaktycznym)
- W specyfikacji oryginalnej: *well-formed*
- Poprawność składniowa jest warunkiem koniecznym dla dopuszczenia dalszego przetwarzania dokumentu XML

XML 2 – przykładowy dokument 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 2 – Deklaracja XML

- Dokument XML powinien (nie musi) zaczynać się deklaracją XML, np.:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
```

- atrybut `version` służy odróżnieniu wersji standardu XML, w tej chwili 1.0 lub 1.1
- atrybut `encoding` informuje o użytym schemacie kodowania (opcjonalny, domyślnie Unicode)
- atrybut `standalone` (dopuszczalne wartości "yes" i "no") określa czy istnieje zewnętrzny dokument DTD dla danego pliku (opcjonalny, domyślnie "no")

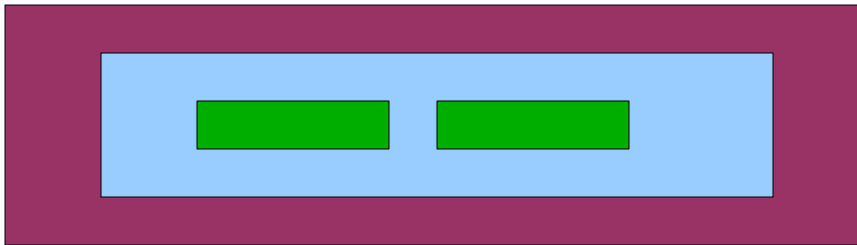
XML 2 – Elementy

- wyznaczone są za pomocą znaczników
 - znacznik otwierający element rozpoczyna się i kończy nawiasem trójkątnym, np. `<nazwisko>`
 - znacznik zamykający element dodatkowo zawiera symbol `/` zaraz za pierwszym nawiasem trójkątnym `</nazwisko>`
- element niepusty obowiązkowo składa się ze znacznika otwierającego i zamykającego
`<nazwisko>Kowalski</nazwisko>`
- element pusty może składać się z jednego znacznika zawierającego znak `/` przed nawiasem kończącym znacznik:
 - postać zwykła: `<drugie_imie></drugie_imie>`
 - postać skrócona: `<drugie_imie />`

XML 2 – Elementy

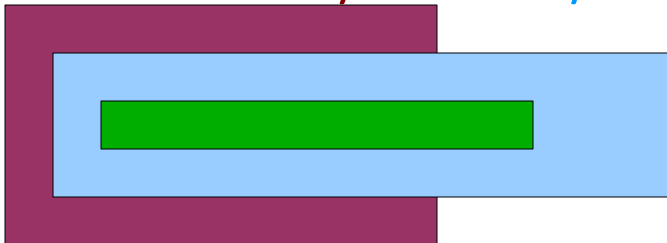
- Elementy muszą być poprawnie zagnieżdżone, tak aby dokument posiadał strukturę hierarchiczną
- dobrze:

```
<a> <b> <c> </c> <c> </c> </b> </a>
```



- źle:

```
<a> <b> <c> </a> <c> </b>
```



XML 2 – Element główny

- Element główny (*root tag*) jest to najbardziej zewnętrzny element dokumentu
 - wszystkie inne elementy znajdują się wewnątrz elementu głównego
 - na zewnątrz mogą znajdować się:
 - komentarze
 - deklaracja xml
 - instrukcje przetwarzania

XML 2 – Atrybuty

- atrybut to para
`nazwa="wartość"`
dołączona do znacznika otwierającego element (lub do znacznika elementu pustego)
- wartość zawsze ograniczona jest cudzysłowem "a" lub apostrofem 'a'
`<lekcja początek="8:15" koniec='9:00'>XML</lekcja>`
`<tekst język="polski" />`
- element może zawierać tylko jeden atrybut o danej nazwie
`<tekst autor="Kowalski" autor="Nowak">...</tekst>`

XML 2 – Nazwy w XML-u

- Nazwy elementów i atrybutów podlegają ograniczeniom:
 - wielkość liter w nazwach jest rozróżniana
 - mogą zawierać dowolne litery i cyfry (dowolnych alfabetów, nie tylko łacińskiego) oraz trzy znaki interpunkcyjne: łącznik “-”, podkreślenie “_” i kropkę “.”
 - nie mogą zawierać innych znaków interpunkcyjnych (cudzysłowy, apostrofy, znak dolara, procent, średnik, ukośnik)
 - nie mogą zawierać białych znaków (odstęp, tabulator)
 - dwukropek ma znaczenie specjalne w nazwach (patrz: przestrzenie nazw, *namespaces*)
 - mogą rozpoczynać się tylko od liter lub podkreślenia (nie od cyfr, łącznika lub kropki)

XML 2 – Znaki specjalne

- Znaki “<”, “>” oraz cudzysłów, apostrof i znak ampersand “&” są traktowane specjalnie
- Chcąc użyć ich jako znaków zwykłych w miejscach gdzie byłoby to niejednoznaczne, należy użyć predefiniowanych symboli (nazywanych encjami):
 - `>`; (> - greater than)
 - `<`; (< - less than)
 - `"`; (“ - quote)
 - `'`; (' - apostrophe)
 - `&`; (& - ampersand)

XML 2 – Sekcja CDATA

- Większe fragmenty tekstu zawierające znaki specjalne mogą być wstawiane za pomocą sekcji CDATA
`<![CDATA[inny tekst, może zawierać < > & i cudzysłowy]]>`
- w sekcji CDATA oczywiście nie może być symbolu zakończenia sekcji, tj. `]]>`

XML 2 - Komentarze

- komentarze obejmują tekst od znaku `<!--` do `-->`
`<a>xxx <!-- to jest komentarz -->`
- podwójny łącznik nie może pojawić się wewnątrz komentarza
- komentarz może zawierać niezakodowane znaki specjalne (nie są one wtedy interpretowane)
`<!-- <a>xxx -->`
- Komentarze nie mogą występować wewnątrz znaczników
`<a <!-- test --> />`

XML 2 – Instrukcje przetwarzania

- Zawierają instrukcje dla aplikacji przetwarzających dokument XML (np. przeglądarka)
- zaczynają się od `<?nazwa` i kończą `?>`
- mogą zawierać atrybuty
- Przykład – zewnętrzny arkusz styli
`<?xml-stylesheet href="ark.css" type="text/css" ?>`

XML 2 – weryfikacja składni

- Przeglądarki rozumiejące XML (MSIE, Firefox, Opera)
- Edytory ze wsparciem dla XML-a
 - kxmleditor (KDE)
 - Microsoft XML Notepad (Windows)
 - XML Fox (Windows)
- Inne narzędzia
 - xmllint (z libxml2, <http://www.xmlsoft.org/>)
 - SAXCount (xerces-j, <http://xerces.apache.org/>)