

XPath

Michał Gałka

XML

- eXtensible Markup Language
- Język znaczników (mniej więcej jak HTML)
- Z założenia ma być samoopisywalny
- Zaprojektowany do przechowywania i wymiany danych
- Dokumenty XML tworzą struktury drzewiaste

XML - przykład

```
<message>  
  <from>bob@example.com</from>  
  <to>alice@example.com</to>  
  <subject>Reminder</subject>  
  <body>Remember the milk</body>  
</message>
```

XML to nie HTML

- Tagi XMLa nie są predefiniowane
 - Twórca dokumentu XML tworzy tagi
 - Nowe tagi mogą zostać dodane do dokumentów
- XML został stworzony do transportu danych, a HTML do ich prezentacji

XML - przykład

```
<message>
  <timestamp>2019-03-14 15:45:09Z</timestamp>
  <from>bob@example.com</from>
  <to>alice@example.com</to>
  <subject>Reminder</subject>
  <body>Remember the milk</body>
</message>
```

Struktura dokumentu

- Element XML to wszystko począwszy od otwierającego, a skończywszy na zamykającym tagu
 - Element może zawierać:
 - tekst
 - atrybuty
 - inne elementy
 - złożenie wszystkich powyższych

```
<?xml version="1.0" encoding="UTF-8"?>
<message lang="pl">
  <timestamp>2019-03-14 15:45:09Z</timestamp>
  <from>bob@example.com</from>
  <to>alice@example.com</to>
  <subject>Reminder</subject>
  <body>Remember the milk</body>
</message>
```

Struktura dokumentu

- Reguły nazewnictwa elementów:
 - Wielkość znaków ma znaczenie
 - Nazwa musi zaczynać się od znaku "_", albo litery
 - Nazwa nie może zaczynać się od ciągu xml (o dowolnej wielkości liter)
 - Nazwy mogą zawierać litery, cyfry, myślniki, podkreślenia, kropki
 - Nazwy nie mogą zawierać spacji
- Atrybuty zawsze muszą być ujęte w cudzysłów (pojedynczy lub podwójny)

```
<?xml version="1.0" encoding="UTF-8"?>
<message lang="pl">
  <timestamp>2019-03-14 15:45:09Z</timestamp>
  <from>bob@example.com</from>
  <to>alice@example.com</to>
  <subject>Reminder</subject>
  <body>Remember the milk</body>
</message>
```

- Nazwy elementów XML są definiowane przez deweloperów
- Chcąc połączyć dokumenty XML różnych aplikacji mogą pojawić się konflikty nazw

Przestrzenie nazw - przykład

- Problem: Identyczne nazwy elementów
- Definicja tabeli

```
<table>  
  <tr>  
    <td>Apples</td>  
    <td>Bananas</td>  
  </tr>  
</table>
```

- Opis stołu

```
<table>  
  <name>African Coffee Table</name>  
  <width>80</width>  
  <length>120</length>  
</table>
```

Przestrzenie nazw - przykład

- Rozwiązanie: prefiksy przestrzeni nazw

```
<root>

<h:table xmlns:h="http://www.w3.org/TR/html4/">
  <h:tr>
    <h:td>Apples</h:td>
    <h:td>Bananas</h:td>
  </h:tr>
</h:table>

<f:table xmlns:f="https://www.w3schools.com/furniture">
  <f:name>African Coffee Table</f:name>
  <f:width>80</f:width>
  <f:length>120</f:length>
</f:table>

</root>
```

Przestrzenie nazw - przykład

```
<root xmlns:h="http://www.w3.org/TR/html4/"  
xmlns:f="https://www.w3schools.com/furniture">
```

```
<h:table>
```

```
<h:tr>
```

```
<h:td>Apples</h:td>
```

```
<h:td>Bananas</h:td>
```

```
</h:tr>
```

```
</h:table>
```

```
<f:table>
```

```
<f:name>African Coffee Table</f:name>
```

```
<f:width>80</f:width>
```

```
<f:length>120</f:length>
```

```
</f:table>
```

```
</root>
```

Przestrzenie nazw - prawdziwy przykład

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

<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

<xsl:template match="/">
<html>
<body>
  <h2>My CD Collection</h2>
  <table border="1">
    <tr>
      <th style="text-align:left">Title</th>
      <th style="text-align:left">Artist</th>
    </tr>
```

Przestrzenie nazw - prawdziwy przykład

```
<xsl:for-each select="catalog/cd">
  <tr>
    <td><xsl:value-of select="title"/></td>
    <td><xsl:value-of select="artist"/></td>
  </tr>
</xsl:for-each>
</table>
</body>
</html>
</xsl:template>

</xsl:stylesheet>
```

XPath

- XML Path Language
- Pozwala na użycie “ścieżek” przy nawigacji pomiędzy elementami dokumentu XML
- Ma ponad 200 wbudowanych funkcji

Węzły (nodes)'

- Typy węzłów:
 - element
 - atrybut
 - tekst
 - przestrzeń nazw (namespace)
 - instrukcja przetwarzania (processing-instruction)
 - komentarz
 - dokument

Zależności między węzłami

- Rodzic (parent)
 - Każdy węzeł ma jednego rodzica
- Dzieci (children)
 - Każdy węzeł-element ma zero lub więcej dzieci
- Rodzeństwo (siblings)
 - Węzły posiadające tego samego rodzica
- Przodkowie (ancestors)
 - Rodzice, rodzice rodziców itd.
- Zstępni (descendants)
 - dzieci, ich dzieci itd.

Przykład dokumentu

```
<cinema>
  <movie>
    <title lang="en"> Good Bad and Ugly</title>
    <director>Sergio Leone</director>
    <year>1966</year>
    <stars>
      <actor>Clint Eastwood</actor>
      <actor>Eli Wallach</actor>
      <actor>Lee Van Cleef</actor>
    </stars>
  </movie>
</cinema>
```