

# XML

eXtensible  
Markup  
Language



# Plan wykładów

- Wprowadzenie do XML
- Struktura dokumentów XML
- Języki zapytań dla XML
- XML w bazach danych

# Plan wykładu

---

- Geneza XML
- Rekomendacja XML 1.0 – składnia języka
- Zastosowania XML
- Modele dostępu do dokumentów XML

# Plan wykładu

- Geneza XML
- Rekomendacja XML 1.0 – składnia języka
- Zastosowania XML
- Modele dostępu do dokumentów XML

# Geneza XML

- 1986 r. – Przyjęcie normy ISO 8879 – Standard Generalized Markup Language (SGML)
- 1994 r. – HTML 2.0 zostaje zdefiniowany jako zastosowanie SGML (RFC 1866)
- 1998 r. – XML staje się oficjalną rekomendacją W3 Consortium

# Rekomendacja XML 1.0

## Założenia towarzyszące tworzeniu XML

- XML powinien być prosty w wykorzystaniu w sieci Internet
- XML powinien wspierać szerokie spektrum aplikacji
- XML powinien być kompatybilny z SGML
- Pisanie programów przetwarzających dokumenty XML powinno być proste

# Rekomendacja XML 1.0

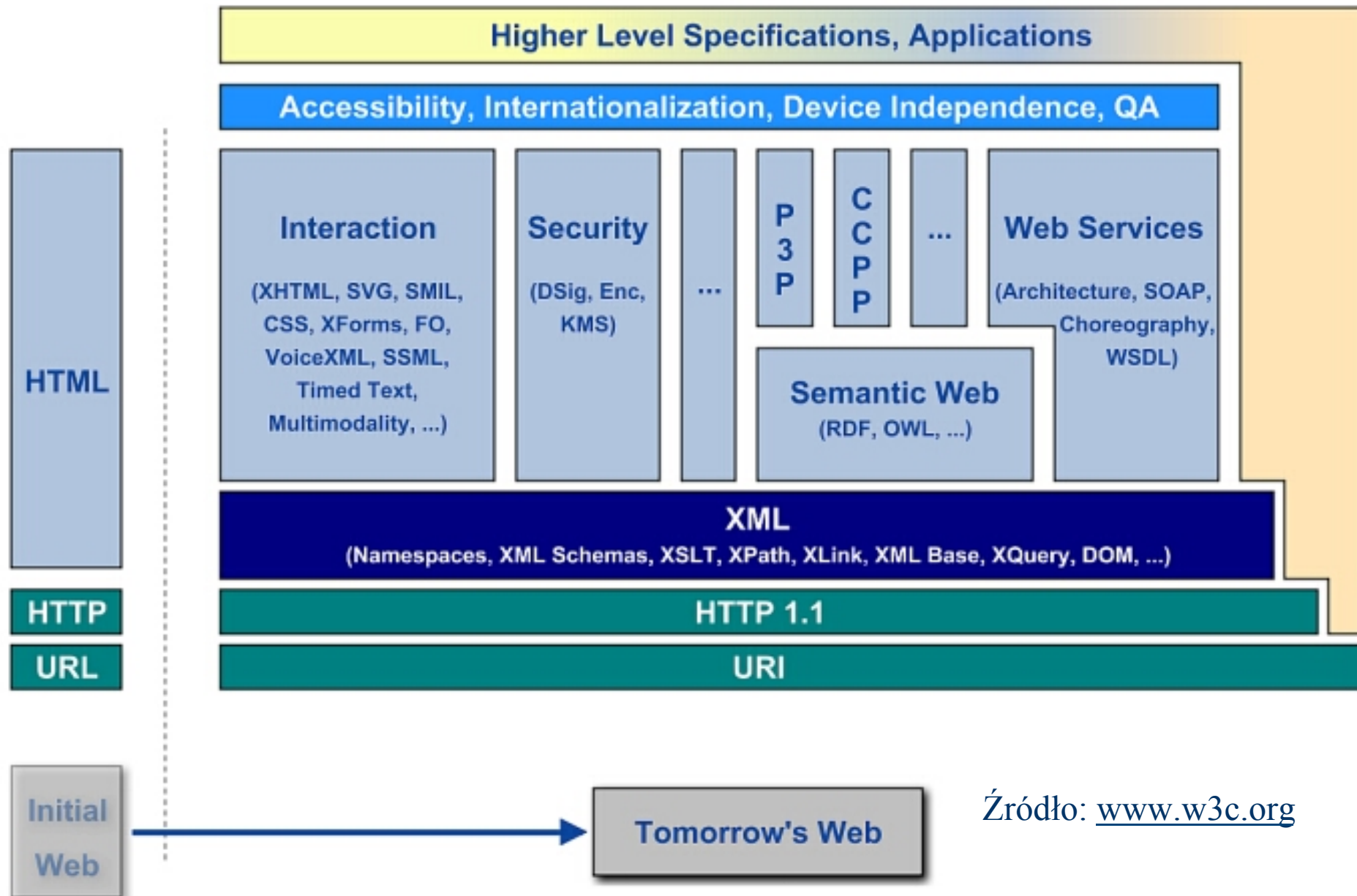
## Założenia towarzyszące tworzeniu XML

- Liczba opcji w języku powinna być ograniczona do minimum
- Dokumenty XML powinny być jasne i czytelne dla człowieka
- Tworzenie dokumentów XML powinno być proste
- Zwięzłość znaczników XML ma niewielkie znaczenie

# XML 1.1

- Równocześnie z czwartą odsłoną wersji 1.0, W3C opublikowało drugą odsłonę wersji 1.1. Wprowadza ona zmiany w zestawie dopuszczanych znaków, co ma związek z modyfikacjami standardu Unicode przeprowadzanymi już po publikacji wersji 1.0.
- .





Źródło: [www.w3c.org](http://www.w3c.org)

# Prace nad technologią XML w W3C

Prace na XML są realizowane przez 6 grup roboczych W3C:

1. ***XML Coordination Group***: Koordynacja pomiędzy grupami zajmującymi się XML oraz współdziałanie z innymi grupami. Nie tworzy własnych specyfikacji.
2. ***XML Core Working Group***: Specyfikacja XML oraz najbliższe z samym językiem związane: Namespaces, XML Information Set, oraz XInclude.
3. ***XSL Working Group***: Definicje stylów i transformacje obiektów.
4. ***XML Linking Working Group***: Łączy hipertekstowe dla XML. Obecnie nie jest aktywna.
5. ***XML Query Working Group***: Język zapytań dla XML.
6. ***XML Schema Working Group***: Definicje i opis struktury, zawartości oraz w pewnym zakresie również semantyki dokumentów XML.

# Plan wykładu

- Geneza XML
- Rekomendacja XML 1.0 – składnia języka
- Zastosowania XML
- Modele dostępu do dokumentów XML

# Dokument XML

- Dokument XML:
  - dane poprawnie opisane przy pomocy XML
- Dokument XML poprawny składniowo (ang. *well-formed*):
  - zawiera przynajmniej jeden element
  - elementy są poprawnie zagnieżdżone
  - spełnia warunki omówione w dalszej części wykładu

# Przykład

**<Pracownicy>**

**<Pracownik>**

**<Nazwisko> Kowalski</Nazwisko>**

**<PESEL> 56022812335 </PESEL>**

**<Data\_ur> 28.02.1956 </Data\_ur>**

**<Data\_przyjecia> 01.10.1980 </Data\_przyjecia>**

**<Dzieci> Brak </Dzieci>**

**<Zespól> R&D </Zespól>**

**</Pracownik>**

**</Pracownicy>**

# Struktury logiczne dokumentu XML

## Element

- Złożony ze znacznika początkowego i końcowego  
**<nazwa\_elem> zawartość </nazwa\_elem>**
- Znaczniki otwierający i zamykający elementu muszą być takie same
- Pusty  
**<element\_pusty/>**

# Struktury logiczne dokumentu XML

## Element – nazwa

- może zawierać litery, cyfry, podkreślenie, kropkę
- nie powinna zaczynać się od liczby lub kropki
- nie powinna zaczynać się od liter xml
- nie może zawierać spacji
- wielkość liter ma znaczenie

# Struktury logiczne dokumentu XML

## Element – zawartość

Element może zawierać:

- element
- tekst
- zawartość mieszana (tekst i element)
- zawartość pustą



# Przykład

```
<korzen>
  <element>
    <inny_element> Zawartość tekstowa (prosta)
    </inny_element>
  </element>
  <element>
    Zawartość mieszana – tekst oraz
    <inny_element> ważniejszy tekst </inny_element>
  </element>
  <pusty_element> </pusty_element>
  <kolejny_pusty_element/>
</korzen>
```

# Struktury logiczne dokumentu XML

## Zagnieżdżanie elementów

- Znaczniki początkowe i końcowe elementów nie mogą się przeplatać
- Hierarchia rodzic-dziecko

# Przykład

```
<korzen>  
  <element>  
    <inny_element> Zawartość  
  </inny_element>  
  </element>  
</korzen>
```

# Przykład

**<Pracownicy>**

**<Pracownik>**

**<Nazwisko> Kowalski</Nazwisko>**

**<PESEL> 56022812345 </PESEL>**

**<Data\_ur> 28.02.1956 </Data\_ur>**

**<Data\_przyjcia> 01.10.1980 </Data\_przyjcia>**

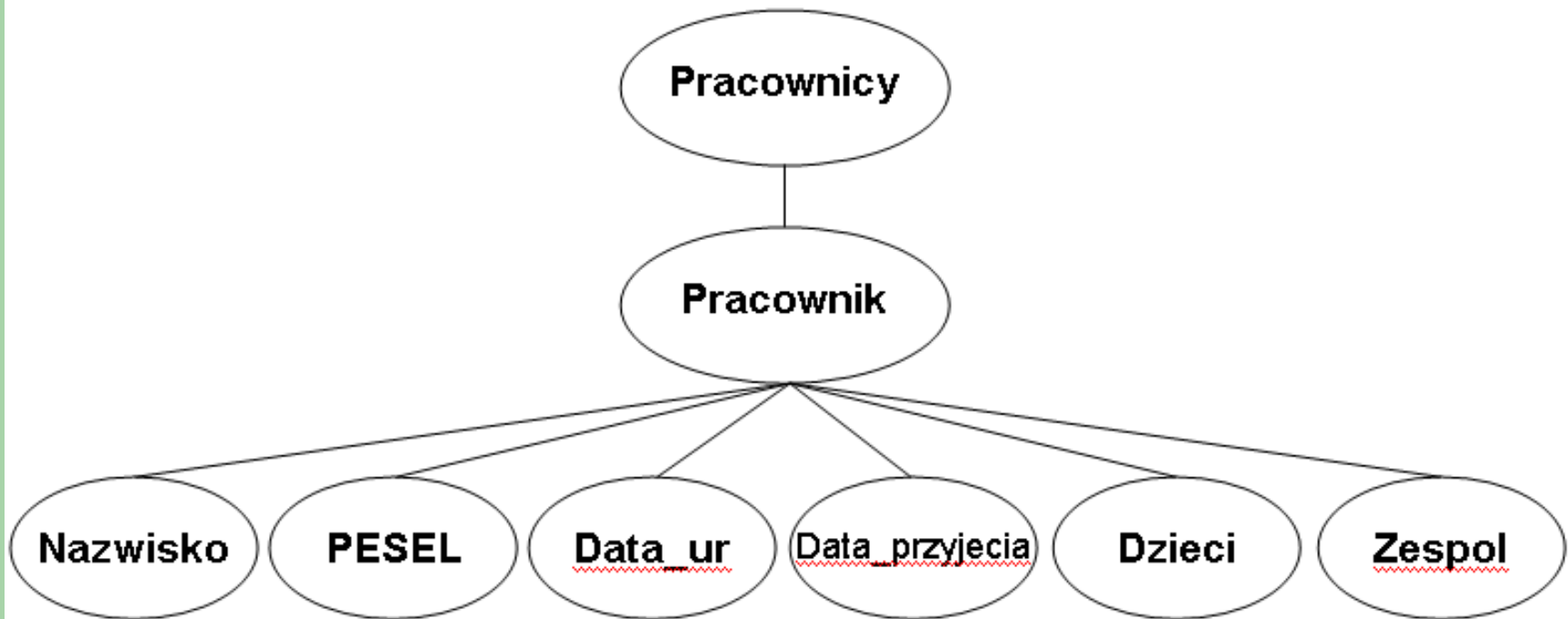
**<Dzieci> Brak </Dzieci>**

**<Zespól> Projektowania </Zespól>**

**</Pracownik>**

**</Pracownicy>**

# Przykład



# Struktury logiczne dokumentu XML

## Element – atrybuty

- Znacznik otwierający i pusty mogą zawierać atrybuty
- Nazwa danego atrybutu może wystąpić tylko raz wśród atrybutów danego elementu
- Wartość atrybutu musi być umieszczona w cudzysłowie " " lub apostrofach ' '

**<ELEMENT id="1"> zawartość </ELEMENT>**

# Przykład

**<Pracownicy>**

**<Pracownik**

**nazwisko="Kowalski"**

**PESEL="56022812345"**

**data\_ur="28.02.1956"**

**data\_przyjecia="01.10.1980"**

**dzieci="Brak"**

**zespol="Projektowania">**

**</Pracownik>**

**</Pracownicy>**

# Struktury logiczne dokumentu XML

## Elementy potomne, czy atrybuty?

- Atrybuty nie mogą zawierać wielu wartości
- Atrybuty nie są łatwo rozszerzalne
- Atrybuty nie opisują struktury



# Przykład

```
<Pracownicy>
  <Pracownik>
    <Nazwisko> Grabski </Nazwisko>
    <PESEL> 56063112345 </PESEL>
    <Data_ur> 31.06.1956 </Data_ur>
    <Data_przyjcia> 01.10.1980 </Data_przyjcia>
    <Dzieci>
      <Dziecko id="1"> Małgosia </Dziecko>
      <Dziecko id="2"> Jaś </Dziecko>
    </Dzieci>
    <Zespol> Piece </Zespol>
  </Pracownik>
</Pracownicy>
```

# Dokument XML

## Deklaracja dokumentu XML – nagłówek

- Występuje opcjonalnie na początku dokumentu
- Zawiera numer wersji XML
- Może zawierać typ kodowania

**<?xml version="1.0" encoding="ISO-8859-1"?>**

# Dokument XML

## Sekcja CDATA

&lt;	<
&gt;	>
&amp;	&
&apos;	'
&quot;	"

**<DOCHODY> netto &lt; 1200 </DOCHODY>**

# Dokument XML

## Sekcja CDATA

- Pole tekstowe nie podlegające analizie parsera
- Może zawierać znaki specjalne np. <

**<![CDATA[ zawartość ]]>**

**<DOCHODY>**

**<![CDATA[ netto < 1200 ]]>**

**</DOCHODY>**

# Dokument XML

## Komentarze

- Nie może znajdować się przed nagłówkiem
- Może znajdować się w dowolnym miejscu
- Nie może zawierać znaków „--”

**<!-- komentarz -->**

# Dokument XML

## Instrukcje przetwarzania (ang. processing instructions)

- Interpretowane przez aplikację przetwarzającą dokument

**<?moja\_apl instrukcja\_sterujaca?>**

# Struktury fizyczne dokumentu XML

## Encje

Encją może być:

- Właściwy dokument XML
- Zewnętrzny podzbiór definicji typu dokumentu
- Zewnętrzny plik
- Ciąg znaków stosowany jako rozwinięcie

# Struktury fizyczne dokumentu XML

## Rodzaje encji

- Ogólna – tekst XML, inne tekstowe lub nietekstowe dane
- Parametryczna – tekst XML wstawiony do definicji typu dokumentu
- Wewnętrzna – ciąg znaków (np. rozwijany w dokumencie)
- Zewnętrzna – zawarta w oddzielnym pliku



# Struktury fizyczne dokumentu XML

## Rodzaje encji cd.

- Przetwarzana – dane XML wstawiane w miejsce odwołania się do encji, analizowane przez procesor tak samo jak cały dokument
- Nie przetwarzana – dowolne dane nie analizowane przez procesor XML

# Plan wykładu

---

- Geneza XML
- Rekomendacja XML 1.0 – składnia języka
- Zastosowania XML
- Modele dostępu do dokumentów XML

# Zastosowania XML

- Strukturalny opis danych
- Elektroniczna wymiana danych
- Komunikacja między aplikacjami

# Zastosowania XML



```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!--
<!--      W4F: Copyright Arnaud Sahuguet and Fabien Azavant, 1998-99
<!--      URL: http://db.cis.upenn.edu/W4F
<!--
<!DOCTYPE W4F_DOC (View Source for full doctype...)>
- <W4F_DOC>
-   <Movie>
       <Title>Matrix, The</Title>
       <Year>1999</Year>
-   <Directed_By>
       <Director>Andy Wachowski</Director>
       <Director>Larry Wachowski</Director>
       </Directed_By>
-   <Genres>
       <Genre>Action</Genre>
       <Genre>Thriller</Genre>
       <Genre>Sci-Fi</Genre>
       <Genre>(more)</Genre>
       </Genres>
-   <Cast>
-       <Actor>
           <FirstName>Keanu</FirstName>
           <LastName>Reeves</LastName>
           </Actor>
-       <Actor>
           <FirstName>Laurence</FirstName>
           <LastName>Fishburne</LastName>
           </Actor>
```

# Zastosowania XML

- Strukturalny opis danych F1
  - Zbieranie i przechowywanie danych w postaci XML
  - Dane różnego typu (liczby, dźwięki, video)
  - Dostępność danych przez Internet



# Zastosowania XML

- RSS – opis informacji publikowanych przez portale internetowe np. CNN, BBC

```
<item>
```

```
  <title>Car bomb kills 30 in Iraq</title>
```

```
    <link>http://www.cnn.com/rssclick/2005/WORLD/meast/11/24/iraq.main  
    section=cnn_world</link>
```

```
  <description>Seven people were killed and 13 wounded Thursday when a  
  car bomb detonated outside a hospital in the town of Mahmudiya, about  
  35 kilometers (22 miles) south of Baghdad, according to a police official  
  with the Iraqi Interior Ministry.</description>
```

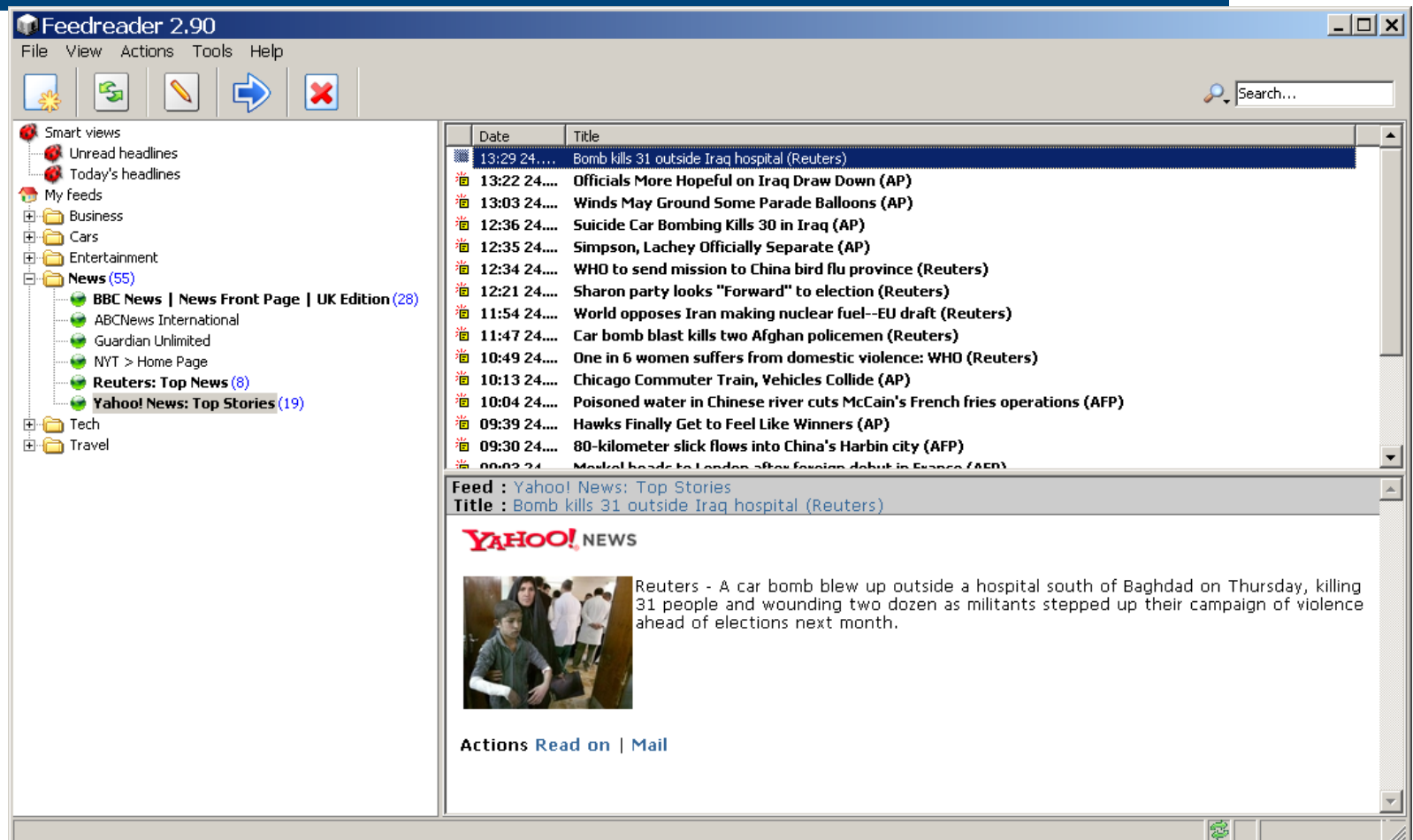
```
  <pubDate>Thu, 24 Nov 2005 05:29:22 EST</pubDate>
```

```
</item>
```

```
<item>
```

```
  <title>EU row looms as Merkel visits UK</title>
```

# Zastosowania XML



```

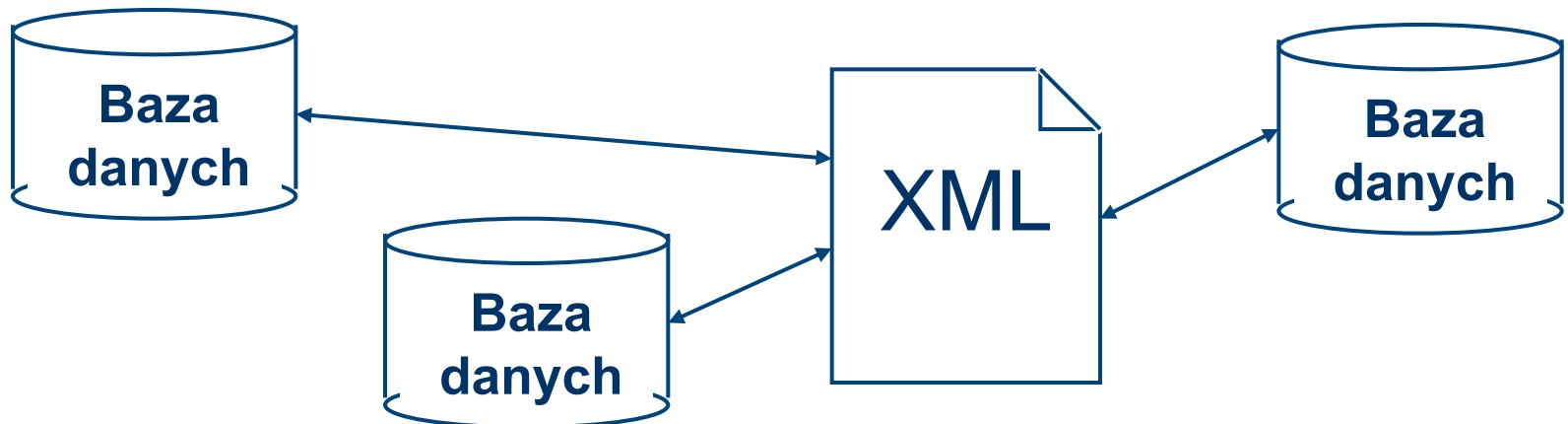
1 <?xml version="1.0" encoding="utf-8"?>
2 <manifest xmlns:android="http://schemas.android.com/apk/res/android"
3     package="com.sample.someapp"
4     android:versionCode="1"
5     android:versionName="1.0">
6     <uses-sdk android:minSdkVersion="7" />
7
8     <uses-permission android:name="android.permission.CAMERA"></uses-permi
9
10    <application android:label="@string/app_name" android:description="@st
11        android:icon="@drawable/quizicon" android:debuggable="true">
12        <activity android:name=".QuizSplashActivity"
13            android:label="@string/app_name">
14            <intent-filter>
15                <action android:name="android.intent.action.MAIN" />
16                <category android:name="android.intent.category.LAUNCHER"
17            </intent-filter>
18        </activity>
19        <activity android:name="QuizMenuActivity"></activity>
20        <activity android:name="QuizHelpActivity"></activity>
21        <activity android:name="QuizScoresActivity"></activity>
22        <activity android:name="QuizSettingsActivity"></activity>
23        <activity android:name="QuizGameActivity"></activity>
24
25        <activity android:name="About"
26            android:label="@string/about_title"
27            android:theme="@android:style/Theme.Dialog" />
28
29        <activity android:name="Prefs"
30            android:label="@string/settings_title" />
31
32        <activity android:name="Game"
33            android:label="@string/game_title" />
34
35        <activity android:name="Keypad"
36            android:label="@string/keypad_title" />
37
38    </application>
39 </manifest>

```



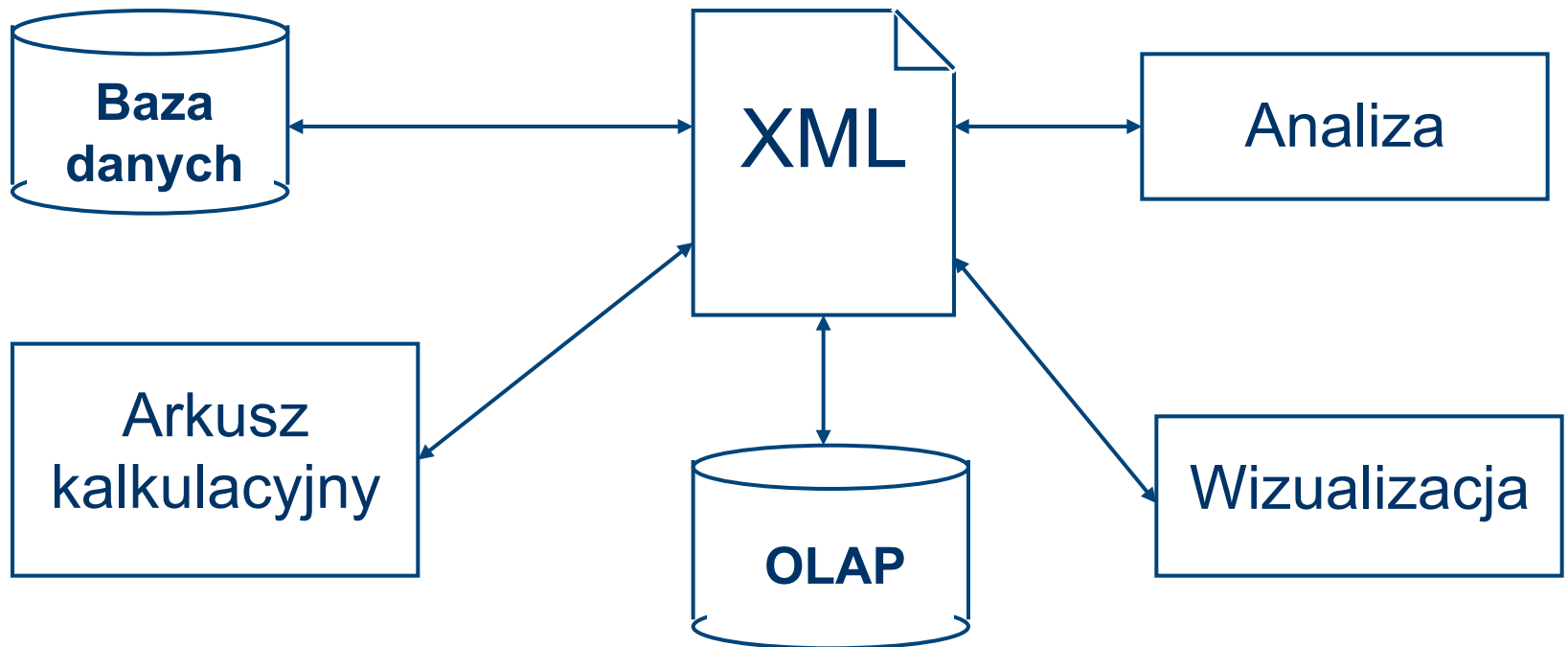
# Zastosowania XML

- Archiwizacja bazy danych
- XML jako format pośredniczący w bazodanowym środowisku heterogenicznym



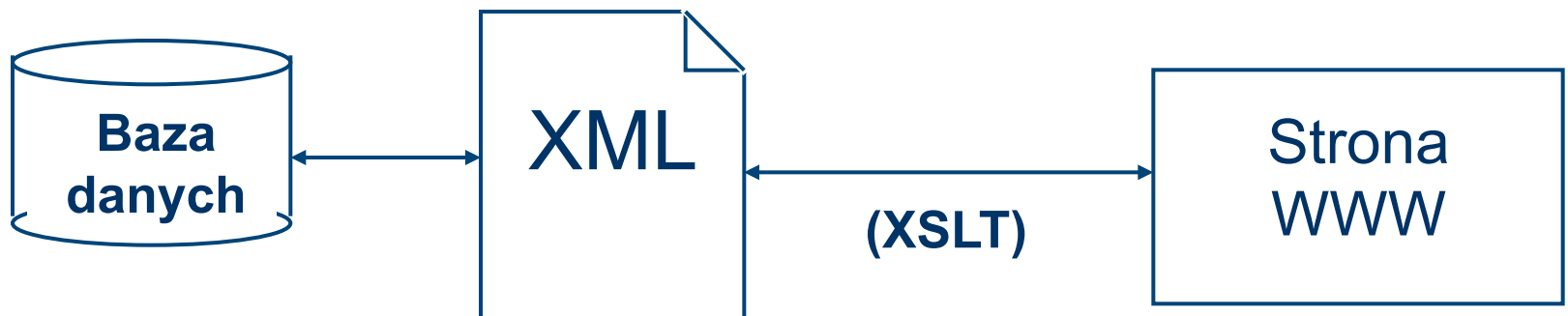
# Zastosowania XML

- XML jako format wymiany pomiędzy aplikacjami

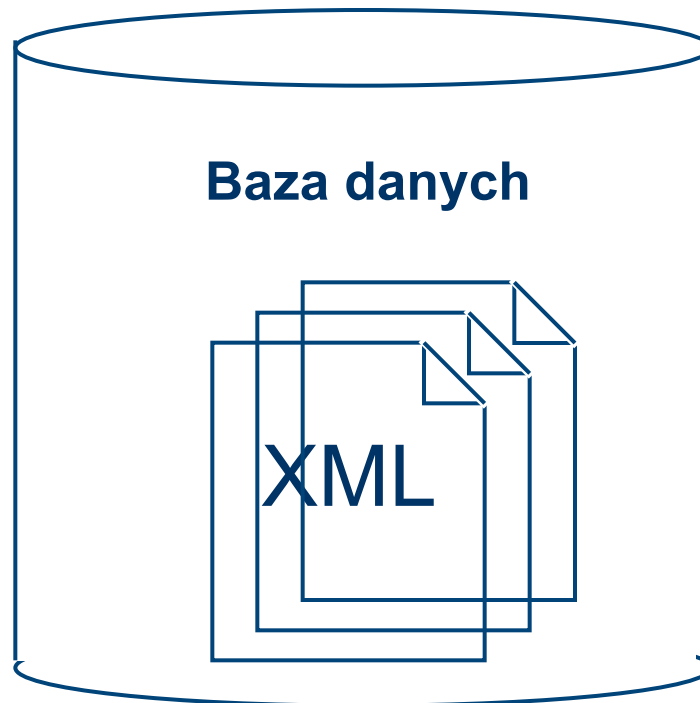


# Zastosowania XML

- XML jako format pośredniczący przy tworzeniu stron WWW



# XML w bazach danych



# Plan wykładu

---

- Geneza XML
- Rekomendacja XML 1.0 – składnia języka
- Zastosowania XML
- Modele dostępu do dokumentów XML

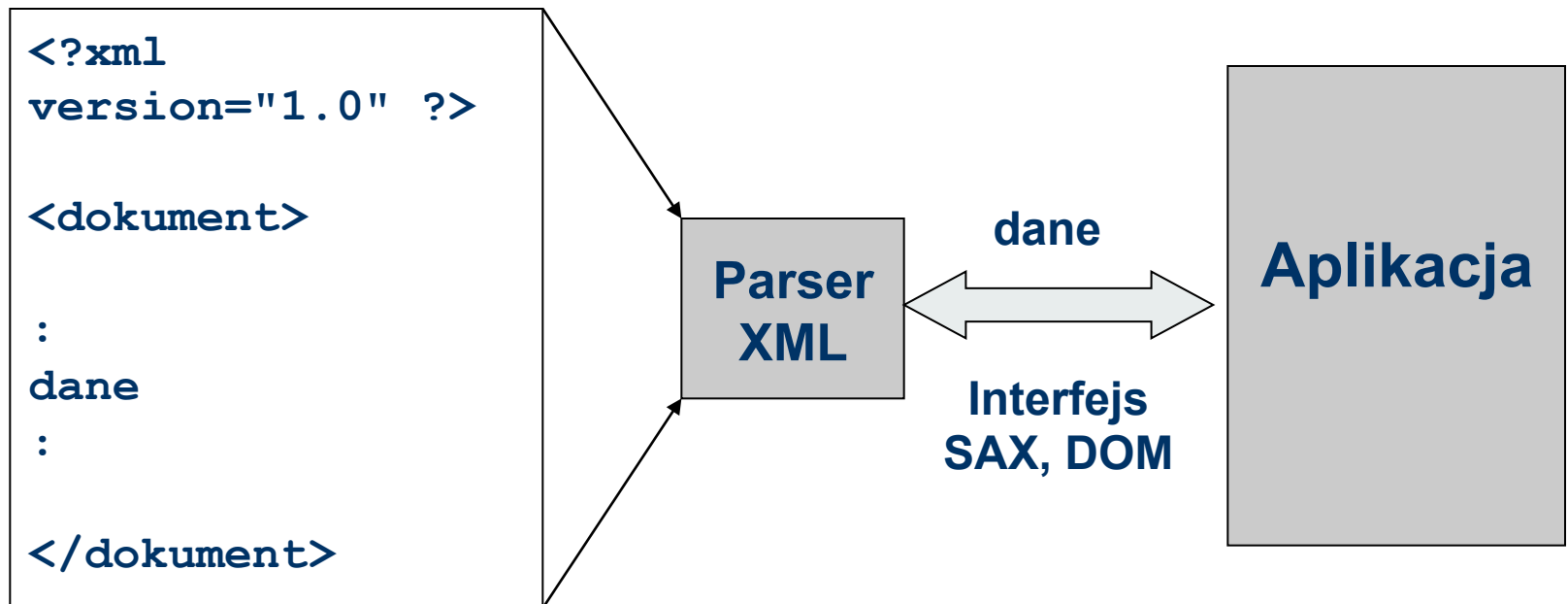
# XML w aplikacjach

## Elementy potrzebne do wykorzystania dokumentów XML w aplikacji:

- Dokument XML – strukturalny opis danych
- Parser XML – moduł analizujący dokument XML i udostępniający aplikacji jego zawartość
- Standard komunikacji między aplikacją, a parserem

# XML w aplikacjach

## Modele dostępu do dokumentów XML



# XML w aplikacjach

## Modele dostępu do dokumentów XML

- SAX – Simple API for XML
- DOM – Document Object Model




# SAX

- Interfejs (API) pozwalający na analizę dokumentów XML
- Wykorzystuje model zdarzeniowy
- Specyfikacja SAX została stworzona w języku Java

# SAX

```
<?xml version="1.0"?>  
<doc>  
  <para>  
    Hello, world!  
  </para>  
</doc>
```

start document  
start element: doc  
start element: para  
characters: Hello, world!  
end element: para  
end element: doc  
end document



# DOM

- Interfejs (API) pozwalający na przetwarzanie dokumentów XML i HTML
- DOM odzwierciedla strukturę modelowanego dokumentu w postaci struktury obiektów przypominającej drzewo
- Specyfikacja DOM została stworzona w języku OMG IDL (Interface Description Language)

# DOM

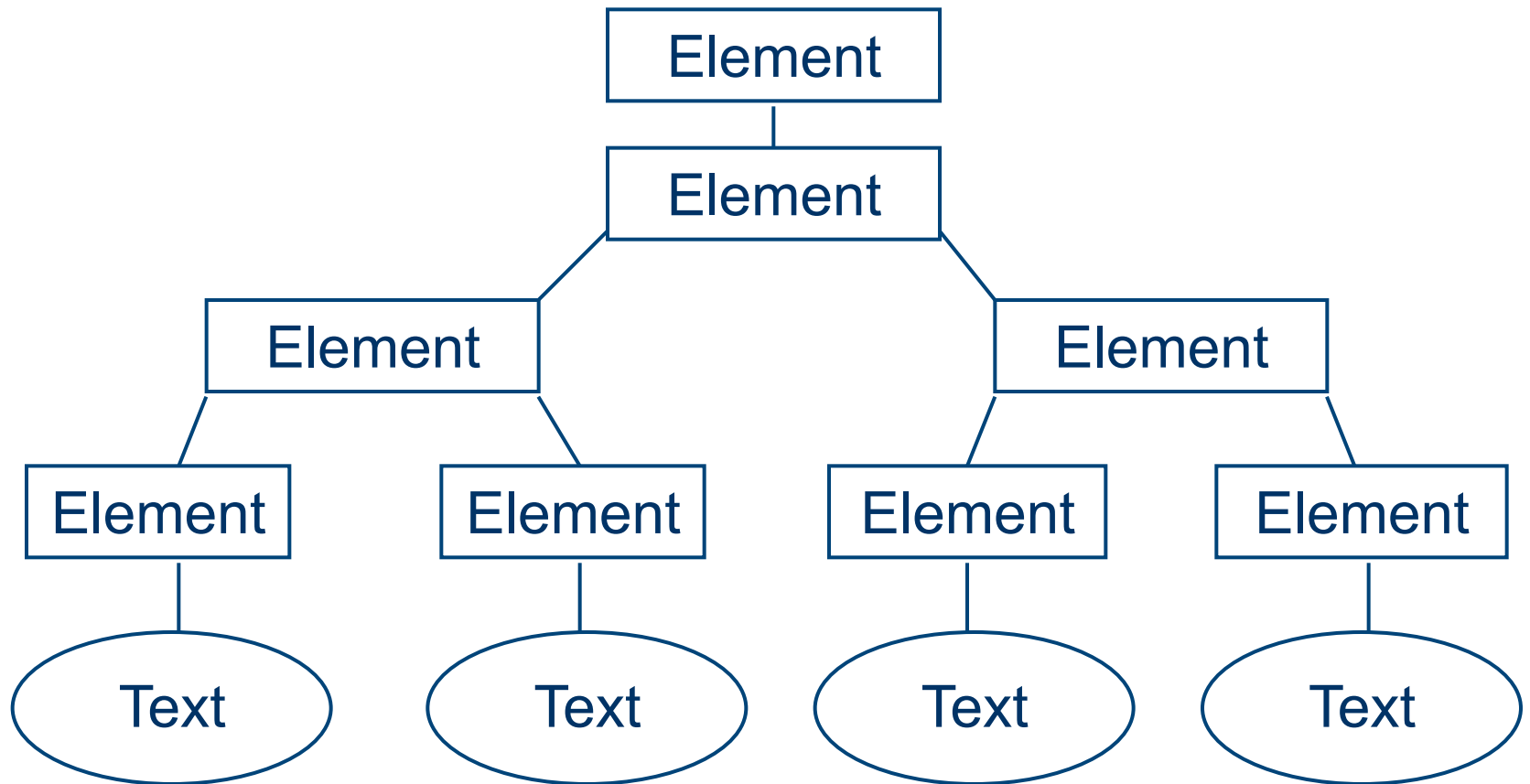
## DOM umożliwia:

- tworzenie dokumentów
- nawigowanie po strukturze dokumentu
- dodawanie elementów
- modyfikację elementów
- usuwanie elementów i ich zawartości

# DOM

```
<TABLE>
  <TBODY>
    <TR>
      <TD>Roxanne</TD>
      <TD>The Police</TD>
    </TR>
    <TR>
      <TD>One</TD>
      <TD>U2</TD>
    </TR>
  </TBODY>
</TABLE>
```

# DOM



# DOM

**Operacje na dokumentach i ich zawartości można przeprowadzać poprzez interfejs w postaci:**

- hierarchicznej struktury klas obiektów odzwierciedlającej powiązania typu dziedziczenie pomiędzy obiektami
- płaskiego widoku elementów dokumentu

# DOM

## Struktura hierarchiczna:

- **Document** – Element, ProcessingInstruction, Comment, DocumentType
- **DocumentFragment** – Element, ProcessingInstruction, Comment, Text, CDATASection, EntityReference
- **DocumentType** – brak dzieci
- **EntityReference** – Element, ProcessingInstruction, Comment, Text, CDATASection, EntityReference
- **Element** – Element, Text, Comment, ProcessingInstruction, CDATASection, EntityReference



# DOM

## Struktura hierarchiczna:

- **Attr** – Text, EntityReference
- **ProcessingInstruction** – brak dzieci
- **Comment** – brak dzieci
- **Text** – brak dzieci
- **CDATASection** – brak dzieci
- **Entity** – Element, Text, Comment, ProcessingInstruction, CDATASection, EntityReference
- **Notation** – brak dzieci

# DOM

## Widok płaski:

- Node
- NodeList
- NamedNodeMap

# XML w aplikacjach – podsumowanie

<b>SAX</b>	<b>DOM</b>
Model zdarzeniowy	Model obiektowy
Prostota	Złożoność
Dostęp jedynie do bieżącego fragmentu dokumentu	Dostęp do całego dokumentu
Brak dostępu do komentarzy i dużej części DTD	Pełniejsze odzwierciedlenie dokumentu
Brak możliwości modyfikacji dokumentu	Możliwość tworzenia nowych dokumentów

# Referencje

- [www.w3.org](http://www.w3.org) – rekomendacje, odnośniki do materiałów
- <http://www.saxproject.org/> - SAX
- [www.rpbouret.com](http://www.rpbouret.com) – opracowanie dotyczące XML i baz danych
- [exist.sourceforge.net/](http://exist.sourceforge.net/), [www.softwareag.com](http://www.softwareag.com), [www.x-hive.com](http://www.x-hive.com), ...
- [www.oracle.com](http://www.oracle.com), [www.ibm.com](http://www.ibm.com), ...