

# Web servisai

XML

Osvaldas Grigas

# XML

- W3C specifikacija
  - Pilnas pavadinimas : Extensible Markup Language
- Aprašo labai lankstų tekstinį duomenų formatą
- Naudojamas dokumentams koduoti ir duomenims serializuoti
- Markup = ženklėjimas
  - visa tai, kas tarp < ir >

# XML ir Unicode

- XML yra tekstinis formatas, ir visa informacija jame yra užrašoma Unicode simboliais
- Unicode yra standartas, aprašantis apie 100 000 simbolių iš beveik visų pasaulio kalbų (įskaitant mirusias kalbas, breilio raštą, matematinius simbolius ir kt.)
  - Kiekvienas simbolis turi unikalų kodą (*codepoint*). Kodas paprastai užrašomas šešioliktainiu skaičiumi taip: U+XXXX, pavyzdžiui:
    - lotyniškos raidės "A" kodas: U+0041
    - lietuviškos raidės "Ą" kodas: U+0104
    - simbolio "<" kodas: U+003C
    - skorpiono ženklą "♏" kodas: U+264F

# XML ir koduotės

- Unicode yra abstraktus simbolių rinkinys. Konkreti koduotė (*encoding*) paverčia simbolius baitais
  - XML procesoriai privalo palaikyti UTF-8 ir UTF-16 koduotes
  - Bet XML dokumentai gali naudoti bet kurią kitą koduotę (pvz, windows-1257)
  - Koduotė nurodoma XML deklaracijoje (pačioje dokumento pradžioje):

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

# XML dokumentų pavyzdžiai

# XML dokumentų pavyzdžiai

## Laiškas

```
<letter>
  <to>Marytė</to>
  <from>Joniukas</from>
  <body>
    <paragraph>Labas, Caca,</paragraph>
    <paragraph>Čia rašo Rembo.</paragraph>
    <paragraph>Kaip gyveni?</paragraph>
  </body>
</letter>
```

## Laiškas, užrašytas naudojant kitokį dokumento modelį

```
<message to-person="Marytė" from-person="Joniukas">
  Labas, Caca,
  Čia rašo Rembo.
  Kaip gyveni?
</message>
```

## Daugiakalbis laiškas su kinišku ženkliniu

```
<?xml version="1.0" encoding="UTF-16"?>
<咨文 向="Marytė" 形="Joniukas">
  Привет, Цаца,
  Это - Рембо.
  كيف حالك ?
</咨文>
```

# XML dokumentų pavyzdžiai

## XHTML tinklalapis

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
    <title>XHTML 1.0 pavyzdys</title>
  </head>
  <body>
    <h1>Paprastas tinklalapis</h1>
    <p>
      Lorem ipsum dolor sit amet, consectetur adipisicing elit,
      sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
      Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris
      nisi ut aliquip ex ea commodo consequat.
    </p>
    <p>
      
    </p>
  </body>
</html>
```

# XML dokumentų pavyzdžiai

## Vektorinės grafikos paveikslėlis (SVG)

```
<?xml version="1.0"?>
<svg xmlns="http://www.w3.org/2000/svg" width="300" height="200">
  <circle cx="150" cy="100" r="50"/>
  <rect x="30" y="20" width="250" height="170" rx="20"
    style="fill: #ff0000; stroke: #000000; stroke-width: 2px;"/>
</svg>
```



# XML dokumentų pavyzdžiai

## SVG vektorinė grafika XHTML tinklalapyje

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
    <title>XHTML+SVG pavyzdys</title>
  </head>
  <body>
    <h1>Tinklalapis su SVG grafika</h1>

    <svg xmlns="http://www.w3.org/2000/svg" width="300" height="200">
      <circle cx="150" cy="100" r="50"/>
      <rect x="30" y="20" width="250" height="170" rx="20"
        style="fill: #ff0000; stroke: #000000; stroke-width: 2px;"/>
    </svg>

  </body>
</html>
```

# XML dokumentų pavyzdžiai

## OpenDocument (OASIS ir ISO standartas ofiso dokumentams)

```
<?xml version="1.0" encoding="UTF-8"?>
<office:document-content version="1.0"
  xmlns:office="urn:oasis:names:tc:opendocument:xmlns:office:1.0"
  xmlns:style="urn:oasis:names:tc:opendocument:xmlns:style:1.0"
  xmlns:fo="urn:oasis:names:tc:opendocument:xmlns:xsl-fo-compatible:1.0"
  xmlns:text="urn:oasis:names:tc:opendocument:xmlns:text:1.0">

  <office:automatic-styles>
    <style:style family="paragraph" name="P0">
      <style:paragraph-properties fo:text-align="center"/>
    </style:style>
    <style:style family="text" name="T0">
      <style:text-properties fo:font-size="9pt" fo:font-weight="bold"/>
    </style:style>
  </office:automatic-styles>

  <office:body>
    <office:text>
      <text:p style-name="P0">
        <text:span style-name="T0">Paprastas dokumentas</text:span>
      </text:p>
    </office:text>
  </office:body>
</office:document-content>
```

# XML dokumentų pavyzdžiai

## Java servlet deployment descriptor

```
<web-app version="2.4"  
  xmlns="http://java.sun.com/xml/ns/j2ee"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="http://java.sun.com/dtd/web-app_2_3.dtd">  
  
  <servlet>  
    <servlet-name>hello</servlet-name>  
    <servlet-class>test.HelloServlet</servlet-class>  
  </servlet>  
  
  <servlet-mapping>  
    <servlet-name>hello</servlet-name>  
    <url-pattern>/hello</url-pattern>  
  </servlet-mapping>  
</web-app>
```

# XML dokumentų pavyzdžiai

Akcijų biržos Web serviso iškvietimas (naudojant SOAP protokolą)

```
<?xml version="1.0"?>
<Envelope xmlns="http://www.w3.org/2001/12/soap-envelope">
  <Body>
    <GetStockPrice xmlns="http://www.example.org/stock">
      <StockName>GOOG</StockName>
    </GetStockPrice>
  </Body>
</Envelope>
```

# XML dokumentų pavyzdžiai

## Amazon.com web serviso atsakymas (naudojant REST protokolą)

```
<?xml version="1.0" encoding="UTF-8"?>
<ItemLookupResponse xmlns="http://webservices.amazon.com/AWSECommerceService/2007-01-17">
  <Items>
    <Item>
      <ASIN>B000080E6I</ASIN>
      <ItemAttributes>
        <Feature>4.0 megapixel sensor captures enough detail to create
          photo-quality prints up to 11 x 14</Feature>
        <Feature>3x optical zoom and 3.6x digital zoom (10.8x total)</Feature>
        <ListPrice>
          <Amount>44999</Amount>
          <CurrencyCode>USD</CurrencyCode>
          <FormattedPrice>$449.99</FormattedPrice>
        </ListPrice>
        <Manufacturer>Canon Cameras US</Manufacturer>
        <NumberOfItems>1</NumberOfItems>
        <ProductGroup>Photography</ProductGroup>
        <Title>Canon PowerShot S400 4MP Digital Camera w/ 3x Optical Zoom</Title>
        <UPC>013803023961</UPC>
        <Weight Units="pounds">0.41</Weight>
        <Width Units="inches">3.43</Width>
      </ItemAttributes>
    </Item>
  </Items>
</ItemLookupResponse>
```



# XML Infoset

# XML Infoset

- W3C specifikacija (kaip ir XML)
  - Pilnas pavadinimas : XML Information Set
- Aprašo *abstrakty* XML dokumento modelį
- Palyginimui - XML:
  - XML yra konkretus duomenų formatas
  - XML aprašo informacijos kodavimo sintaksę
  - XML nurodo, kaip turi elgtis XML procesoriai, skaitydami XML dokumentą
- Tuo XML Infoset skiriasi nuo XML:
  - XML Infoset yra abstrakti duomenų struktūra
  - XML Infoset aprašo informacijos vienetus (elementus, atributus ir t.t.) bei ryšį tarp jų
  - XML Infoset gali būti *serializuojamas* į XML arba kitokį (dvejotinį) formatą

# Informacijos vienetai

Pagal XML Infoset specifikaciją XML dokumentą gali sudaryti 11 skirtingų informacijos vienetų (*information items*):

| Informacijos vienetas       | Atitikmuo serializavus į XML  |
|-----------------------------|---|
| Document                    | <i>Pats XML dokumentas</i>  |
| Element                     | <code>&lt;elementas /&gt;</code>                                      |
| Attribute                   | <code>atributas="reikšmė"</code>                                      |
| Processing instruction      | <code>&lt;?php echo "Labas, \$name"; ?&gt;</code>                     |
| Unexpanded entity reference | <code>&amp;entityreference;</code>                                    |
| Character                   | A   |
| Comment                     | <code>&lt;!-- komentaras --&gt;</code>                                |
| Document Type Declaration   | <code>&lt;!DOCTYPE html PUBLIC "-//W3C ...."&gt;</code>               |
| Unparsed entity             | <code>&lt;!ENTITY embedded-picture SYSTEM "image.png" PNG&gt;</code>  |
| Notation                    | <code>&lt;!NOTATION PNG SYSTEM "Portable Network Graphics"&gt;</code> |
| Namespace                   | <code>xmlns:mano="http://www.mano.lt/mano_xml_schema"</code>          |



# XML Infoset serializavimas

