

# XML Technologies and Applications

- harjoitustyö -

TURUN YLIOPISTO  
Informaatioteknologian laitos  
Tietojenkäsittelytiede  
Harjoitustyö  
Helmikuu 2009

Jyri Lehtonen (72039)  
jyri.lehtonen@utu.fi  
(yksin tehty harjoitustyö)

# 1. Kuvaus

Kehitin aikaisemmin keväällä yhteisöllisen internet-sivuston, jossa on mahdollista jakaa opiskelijoiden/kavereiden kesken heidän itse tekemiä kuvia, ääniraitoja tai kouluraportteja/dokumentteja. Toistaiseksi tämä toimi pelkästään siten, että siirretty objekti ilmestyy latausvalikkoon pelkkänä linkkinä. Tämä oli toteutettu käytännöllisestä näkökulmasta, jotta sivuston vain saisi kylmästi toimimaan. XML Technologies and Applications -harjoitustyön on tarkoituksena mahdollisesti muuttaa tilannetta ja näyttää se, miten siirrettyjen objektien tulisi muodostua.

Ratkaisu olisi mahdollista toteuttaa erilaisilla ohjelmointikielillä, mutta koska en itse ole koskaan aikaisemmin tehnyt mitään XML/XSL/DTD kokoonpanolla, oli tämä oiva mahdollisuus kehittää tuntemusta näissä, sekä kehittää yhteisösivustoani. Kun sain harjoitustyön ohjaajaltani esimerkkikappaleita siitä, millaisia töitä XML-yhdistelmällä on mahdollista tehdä, valitsin heti mediagallerian. Tämän työn tarkoituksena on siis luoda pohjaa sille, miltä yhteisösivustolleni ladatut objektit näyttävät. Jos en käytä tätä projektia suoraan sivustolle, ainakin otan sen ulkomuodosta mallia.

XML:n hyötynä työssäni on varmasti se, että sitä on varmasti useasti käytetty vastaavanlaiseen toimintaan. Uskon, että XML-pohjaisia ratkaisuja mediagallerioihin on varmasti olemassa komentosarjakielten ohessa tai yhteistyössä. Itse koin ainakin sen, että tällä yhdistelmällä (XML/XSL/DTD) on vaivatonta mallintaa tietty taulukkoalue, joka tulee sisältämään erilaista dataa, ja monistaa sitä.

## 2. DTD (Document Type Definition)

DTD alkaa ensimmäisen elementin määrittelyllä, joka on `galleria`. Tämä elementti sisältää `gobjekti`-elementtejä nolla tai enemmän. `Galleria` on siis korkeimman tason elementti, jonka sisältä löytyy kaikki sen toimintaan liittyvät elementit ja attribuutit. Itse `galleria`-elementtiin liittyy attribuutti `type`, joka on pakollinen ja sen on oltava jokin kolmesta vaihtoehdosta: `image`, `audio` tai `text`. Tämä tarkoittaa sitä, että gallerioita on mahdollista toistaiseksi kehittää kolmea tyyppiä: kuvia varten, ääniraitoja varten sekä dokumentteja varten.

Siirryttäessä seuraavalle tasolle, on suurin painoarvo elementin `gobject` määrittelyllä. `Gobject` (Gallery Object) on jokin tietty objekti, joka sisältyy galleriaan. `Gobject` sisältää seuraavat objektien määrittelyt: `id`, `date`, `thumb`, `title`, `authorinfo`, `comment` ja `rating`. Elementtiin liittyvät myös kolme erikoistyyppiä, joista vain yksi voi esiintyä objektissa: `resolution`, `length` ja `pages`. Lisäksi elementissä ovat vielä kaksi hallinnointiin liittyvää elementtiä `delete` ja `edit`. `Gobject`-elementtiin liittyy myös attribuutti `otype` (Object Type), joka määrää sen, että attribuutin on oltava joko `image`, `audio` tai `text` -tyyppiä.

`Gobject`in sisältämät elementit ovat suhteellisen tyhjentyviä. `id` määrittää objektille ainutlaatuisen numeraalisen tunnuksen, jolla tämä tietty objekti tunnustetaan. `date` on se päivämäärä, jolloin kyseinen objekti ladattiin palvelimelle. `thumb` on pienoiskuvake, joka on oleellisin kuvagallerioissa. Ääniraita- sekä dokumenttigallerioissa pienoiskuvakkeen lisääminen on mahdollista, tekijän halutessa personalisoida tuotostaan. Pienoiskuvakkeen on oltava tiettyä määrättyä kokoa, joka on määrätty attribuutissa `width` ja `height`. `title` on objektin nimi. `authorinfo` sisältää tarvittavat tiedot sen tekijästä: nimen ja sähköpostiosoitteen. `comment` on tekijälle varattu kenttä, jossa hän voi kertoa työstään. `rating` on arvioijalle varattu kenttä, jossa arvioija voi antaa objektille tietyn numeraalisen arvon. Elementtiin liittyy myös attribuutti `rate`, joka on valinnainen, mutta oltava arvojen 0-5 väliltä. `filesize` on yksinkertaisesti objektin koko kovalevyllä.

Jokaisella objektilla on yksi erikoiselementti, jonka on tarkoitus näkyä vain oikean tyyppin gallerialla. Näitä elementtejä ovat `resolution` (`image`), `length` (`audio`) sekä `pages` (`text`). Erikoiselementtien lisäksi jokaisella objektilla on vielä kaksi hallintoelementtiä: `delete` ja `edit`. Nämä ovat tarkoitettu järjestelmänvalvojan ja objektin siirtäjän välineiksi hallita tuotoksia.

Alla on koko `definition.dtd`-tiedoston sisältö.

```
<!-- DTD -->

<!-- Galleria -->
<!ELEMENT gallery (gobject)*> <!-- Voi sisältää nolla tai enemmän objekteja-->

<!--Elementit-->
<!ELEMENT gobject (id, date, thumb, title, authorinfo, comment?, rating?, filesize, resolution,
length, pages, delete, edit)> <!-- Objektin arviointi tai kommentointi ei ole pakollista -->

<!ELEMENT id (#PCDATA)> <!-- PC: Vain tekstiä, ei muita elementtejä-->
<!ELEMENT date (#PCDATA)> <!-- Päiväysleima, koska objekti on lisätty-->
<!ELEMENT thumb (#PCDATA)> <!-- Vaihtoehtona EMPTY, mahdollistaa <img />. (siis tyhjät tagit) -->
<!ELEMENT title (#PCDATA)> <!-- Objektin nimi-->
<!ELEMENT authorinfo (author)+> <!-- Voi sisältää yhden tai enemmän authoreita-->
    <!ELEMENT author (name, email?)> <!-- Voi sisältää sähköpostin, mutta ei ole pakko-->
    <!ELEMENT name (#PCDATA)>
    <!ELEMENT email (#PCDATA)>
```

```

<!ELEMENT comment (#PCDATA)> <!-- Objektiin liitettävä vapaaehtoinen kommentti-->
<!ELEMENT rating (#PCDATA)> <!-- Objektiin liitettävä vapaaehtoinen arviointi-->

<!ELEMENT filesize (#PCDATA)> <!-- Objektin koko kovalevyllä-->

<!ELEMENT delete (#PCDATA)> <!-- Omistajan/Järjestelmänvalvojan oikeus poistaa -->
<!ELEMENT edit (#PCDATA)> <!-- Omistajan/Järjestelmänvalvojan oikeus muokata -->

<!-- Eri gallerioiden erikoistyyppit -->
<!ELEMENT resolution (#PCDATA)> <!-- Kuvagalleria-->
<!ELEMENT length (#PCDATA)> <!-- Äänigalleria-->
<!ELEMENT pages (#PCDATA)> <!-- Tekstigalleria-->

<!--Attribuutit-->
<!ATTLIST gallery type (image|audio|text) #REQUIRED>
<!-- Gallerian on oltava joko kuva/ääni/teksti-->
<!ATTLIST gobject otype (image|audio|text) #REQUIRED>
<!-- Gallerian objektin on oltava kuva/ääni/teksti-->
<!ATTLIST thumb width CDATA #FIXED "100"
                height CDATA #FIXED "100">
<!-- Kuvakkeiden on oltava kokoa 300x330-->
<!ATTLIST rating rate (0|1|2|3|4|5) #IMPLIED>
<!-- Arvioinnin on oltava aseikkoa 0-5 -->

<!--Entiteetit (testausta)-->
<!ENTITY site "Archives">

```

### 3. XML (eXtensible Markup Language)

Ensimmäinen galleriaesimerkki on kuvagalleria. XML-tiedosto määrittää aluksi sen käyttämään definition.dtd määritelmäpakettia, joka on esiteltynä edellisessä luvussa. Tyylimäärittelyt käsitellään seuraavassa luvussa. Nämä määrittelyt ovat harjoitustyön jokaisessa XML-tiedossa, vaikka niitä ei tässä enää erikseen mainita.

Galleria on tyyppiä kuva (attribute: type="image"). Galleria sisältää viisi objektia, jotka käsittelevät kuvia (attribute otype="image"). Jokaisella objektilla on oma ID (numerot 1 - 5). Päivämäärä on harjoitustyön vuoksi myös jokaiseen valittu eri. Pienoiskuvake, joka esittää gallerian kuvaa, on sijoitettu kansioon thumbpic, ja kuvien nimet alkavat test1:stä eteenpäin. Kuvien otsikkokenttä/nimikenttä voi olla mielivaltainen, kuten myös tekijän nimi sekä sähköpostiosoite ja kommentti. Arvioinnin on oltava välillä 0-5 ja objektin arviointi on rating-elementin attribuutissa rate. Objektin erikoistyyppielementti resolution esittää pikseleiden määrän kuvassa. Loput elementit ovat itsestään selviä.

Gallerioiden välillä ei ole suuria eroja. Ääniraitagalleriassa galleria-elementin attributtina on type="audio", objektin attribuuttina otype="audio" sekä resoluution tilalla pituus-elementti length. Dokumenttigalleriassa vastaavasti galleria-elementin attribuuttina on type="text", objektin attribuuttina on otype="text" ja resoluution/pituuden tilalla on sivujen määrä -elementti pages. Seuraavalla sivulla on yksi objekti jokaisesta galleriasta esimerkkinä.

```

<?xml version="1.0" encoding="utf-8"?>
<?DOCTYPE gallery SYSTEM "definition.dtd"?>
<?xml-stylesheet type="text/xsl" href="styles.xsl"?>

<!--Gallerian tyyppi-->
<gallery type="image">                                <!--Galleria on tyyppiä kuva-->

<!--Gallerian objektit-->
<gobject otype="image">                                <!--Objekti on tyyppiä kuva-->
  <id>1</id>                                            <!--Ensimmäinen objekti-->
  <date>1.1.2009</date>                                <!--Päivämäärä leima-->
  <thumb>thumbpic\test1.jpg</thumb>                  <!--Kuvan osoite (josta muodostetaan kuvake)-->
  <title>Vortex of Whee</title>                        <!--Objektin otsikko-->
  <author>                                              <!--Objektin tekijä sekä hänen sähköpostiosoite-->
    <name>Jay</name>
    <email>JayJayStar@gmail.com</email>
  </author>
  <comment> A vortex (pl. vortices) </comment>          <!--Objektiin liitettävä kommentti-->
  <rating rate="5">/5</rating>                        <!--Attribuutissa oleva vaihtuva arvo-->
  <resolution>1024 x 768</resolution>                 <!--Objektin pituus (kuvassa pikseleitä-->
  <filesize>1 MB</filesize>                            <!--Objektin suuruus kovalevyllä-->

  <delete>Delete</delete>                             <!--Omistajan/Järjestelmänvalvojan-->
  <edit>Edit</edit>                                   <!--Omistajan/Järjestelmänvalvojan-->
</gobject>

<!--Gallerian tyyppi-->
<gallery type="audio">                                <!--Galleria on tyyppiä audio-->

<!--Gallerian objektit-->
<gobject otype="audio">                                <!--Objekti on tyyppiä audio-->
  <id>1</id>                                            <!--Ensimmäinen objekti-->
  <date>1.1.2009</date>                                <!--Päivämäärä leima-->
  <thumb>thumbaudio\test1.jpg</thumb><!--Kuvan osoite (josta muodostetaan kuvake)-->
  <title>Ambient Dub</title>                          <!--Objektin otsikko-->
  <author>                                              <!--Objektin tekijä sekä hänen sähköpostiosoite-->
    <name>JaySan</name>
    <email>JaySan@gmail.com</email>
  </author>
  <comment> Ambient Dub was a phrase first ... -Wikipedia</comment>
  <rating rate="5">/5</rating><!--Attribuutissa oleva vaihtuva arvo-->
  <length>3.20min</length>                            <!--Objektin pituus (äänessä minuuttia)-->
  <filesize>8 MB</filesize>                            <!--Objektin suuruus kovalevyllä-->

  <delete>Delete</delete>                             <!--Omistajan/Järjestelmänvalvojan-->
  <edit>Edit</edit>                                   <!--Omistajan/Järjestelmänvalvojan-->
</gobject>

<!--Gallerian tyyppi-->
<gallery type="text">                                <!--Galleria on tyyppiä teksti-->

<!--Gallerian objektit-->
<gobject otype="text">                                <!--Objekti on tyyppiä teksti-->
  <id>1</id>                                            <!--Ensimmäinen objekti-->
  <date>1.1.2009</date>                                <!--Päivämäärä leima-->
  <thumb>thumbdoc\test1.jpg</thumb><!--Kuvan osoite (josta muodostetaan kuvake)-->
  <title>Essee: Dokumentin määritelmä</title><!--Objektin otsikko-->
  <author>                                              <!--Objektin tekijä sekä hänen sähköpostiosoite-->
    <name>JayEr</name>
    <email>JayEr@gmail.com</email>
  </author>
  <comment> A document (noun)... -Wikipedia</comment>
  <rating rate="5">/5</rating><!--Attribuutissa oleva vaihtuva arvo-->
  <pages>8s.</pages>                                  <!--Objektin pituus (tekstissä sivua)-->
  <filesize>600 kB</filesize><!--Objektin suuruus kovalevyllä-->

  <delete>Delete</delete>                             <!--Omistajan/Järjestelmänvalvojan-->
  <edit>Edit</edit>                                   <!--Omistajan/Järjestelmänvalvojan-->
</gobject>

```

## 4. XSL (eXtensible Stylesheet Language)

Sivuston taustan on oltava musta, koska yhteisösivustossakin on taustana ollut aina musta. XSL aloittaa toimintansa käymällä läpi kaikki gallerian objektit, ja tulostaa ne laskevassa id järjestyksessä sivulle. Tämä on toteutettu käyttämällä XSL:n komentoa `<xsl:for-each select="ELEMENTTI">` ja tämän jälkeen komentoa `<xsl:sort select="ELEMENTTI" order="descending"/>`. Tällöin uusin tiedosto (korkein ID) on ensimmäisenä. Tämän toiminnan sisällä on kaksi toimintoa. Ensimmäinen toiminto käy läpi kaikki parittomat ID:t, ja tulostaa niiden taustan vaaleampana kuin parillisten. Tämä on toteutettu XSL:n komennolla `<xsl:if test="(ID mod 2)=1">`. Toinen toiminto käy läpi kaikki parilliset ID:t, ja tulostaa niiden taustan tummempana kuin parittomien. Tämä on toteutettu komennolla `<xsl:if test="(ID mod 2)=0">`. Mitä nämä toiminnot pitävät sitällään, käsitellään seuraavaksi.

Jokaisesta gallerian objektista luodaan taulukko. Taulukossa on yhteensä kolme riviä, ja kolme saraketta. Ensimmäisellä rivillä ovat ID, THUMB sekä TITLE -elementit. Elementit haetaan komennolla `<xsl:value-of select="ELEMENTTI" />`. Tällä tavalla ovat haettu ID sekä TITLE elementit. Thumb-elementissä on tarkoituksena se, että käyttäjä voisi painamalla pienkuvaketta avata alkuperäisen kuvan. Tämä onnistuu toteuttamalla pienoiskuvasta linkki sijoittamalla thumb-elementtiin kuvan osoitteen, ja hakemalla tämä sitten XSL:n komentojen `<xsl:attribute name="ATTRIBUUTTI">` sekä aikaisemmin mainittu `<xsl:value-of select .../ >` avulla. Alla on esimerkki toiminnasta:

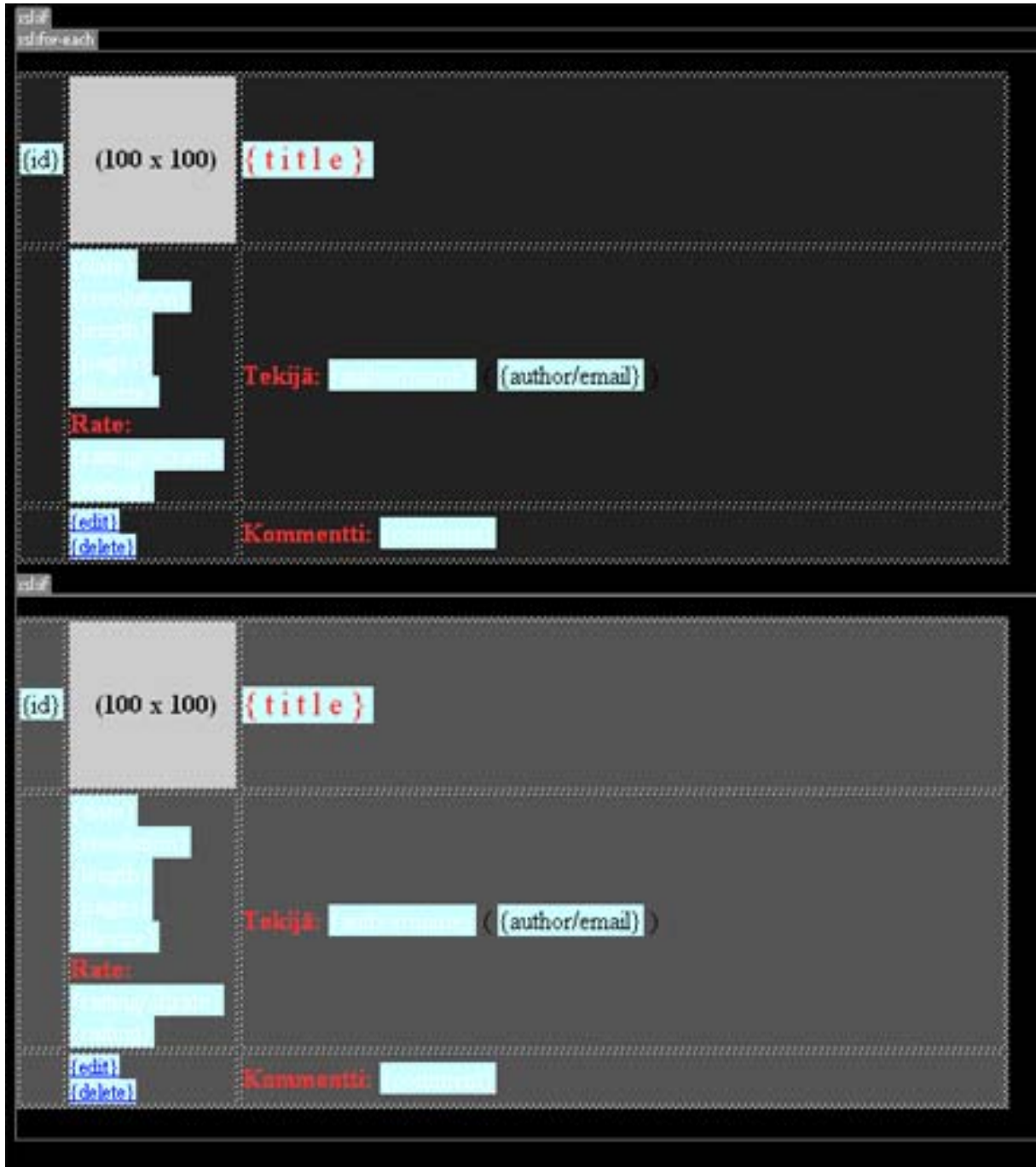
```
<a><xsl:attribute name="href">
  <xsl:value-of select="thumb"/>
</xsl:attribute>

  <img width="100" height="100" style="border:none">

    <xsl:attribute name="src">
      xsl:value-of select="thumb"/>
    </xsl:attribute>
  </img></a>
```

Toisella rivillä ovat kaikki tietoa antavat elementit (date, resolution, length, pages, filesize, rate) sekä tekijän tiedot `authorinfo -> name, email`. Toisella rivillä ei ole uutta XSL-toimintaa. Rivillä on käytetty CSS:n yksinkertaisia koristeluita kuten `<strong>` sekä `<span style="color...">`.

Viimeisellä rivillä ovat hallintaelementit `edit` ja `delete` sekä tekijän laatima kommentti objektista. Tämä kaikki on esitettyä alla olevassa kuvassa. Selvää syytä sille, miksi joka toinen objekti olisi erivärinen taustaltaan, ei ole. Kyseessä on vain esteettinen ratkaisu.



```
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:output method="html" encoding="utf-8" doctype-public="-//W3C//DTD XHTML 1.0 Transitional//EN"
doctype-system="http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"/>
<xsl:template match="/">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
</head>
```

```

<!--Sisältö alkaa-->
<body bgcolor="#000000">

<!--Käy läpi kaikki gallerian objektit, ja tulostaa ne laskevassa järjestyksessä-->
<xsl:for-each select="gallery/gobject">
<xsl:sort select="id" order="descending"/>

<!--Käy läpi parittomat, ja tulostaa niiden taustan vaalempana kuin parillisten-->
<xsl:if test="(id mod 2)=1">
<table width="60%" border="0" bgcolor="#232323">
<tr><!--tulostaa ensimmäiseen riviin: ID, kuvakkeen ja objektin nimen-->
<td width="3%"><xsl:value-of select="id" /></td>
<td width="14%"><a><xsl:attribute name="href">
<xsl:value-of select="thumb"/>
</xsl:attribute>
<img width="100" height="100" style="border:none">
<xsl:attribute name="src">
<xsl:value-of select="thumb"/>
</xsl:attribute>
</img></a></td>
<td width="83%">
<span style="color:#FF3333; font-size:larger; letter-spacing:0.10cm">
<xsl:value-of select="title"/></span><br />
</td>
</tr>

<tr>
<td></td>

<!--Tulostaa toiseen riviin Nippelitiedot, arvioinnin ja tekijän-->
<td><span style="color:#FFFFFF"><xsl:value-of select="date"/><br />
<xsl:value-of select="resolution"/>
<xsl:value-of select="length"/>
<xsl:value-of select="pages"/>
<br /><xsl:value-of select="filesize"/></span><br />
<span style="color:#FF3333"><strong>Rate: </strong></span>
<span style="color:#FFFFFF"><xsl:value-of select="raiting/@rate"/>
<xsl:value-of select="raiting"/></span></td>
<td><span style="color:#FF3333"><strong>Tekijä: </strong></span>
<span style="color:#FFFFFF"><xsl:value-of select="author/name"/></span>
( <xsl:value-of select="author/email"/> )<br /></td>
</tr>

<tr>
<td></td>
<!--Tulostaa kolmanteen riviin hallinta vaihtoehdot sekä kommentit-->
<td><span style="color:#0000FF; font-size:smaller; text-decoration:underline">
<xsl:value-of select="edit" /><br />
<xsl:value-of select="delete" /></span></td>
<td><span style="color:#FF3333"><strong>Kommentti: </strong></span>
<span style="color:#FFFFFF"><xsl:value-of select="comment"/></span><br
/></td>
</tr>

</table>
<br />
</xsl:if>

<!--Käy läpi parilliset, ja tulostaa niiden taustan tummenpana kuin parittomien-->
<xsl:if test="(id mod 2)=0">
<table width="60%" border="0" bgcolor="#555555">
<tr>
.
. <!--Jatkuu samana kuin yllä-->
.
</xsl:if>
</xsl:for-each>

</body>
</html>

</xsl:template>
</xsl:stylesheet>

```



## 5. Lopputulos

Testasin XML-tiedostoni palvelulla: [http://www.w3schools.com/dom/dom\\_validate.asp](http://www.w3schools.com/dom/dom_validate.asp), ja kyseinen palvelu antoi palautteeksi: No errors found!.

Internet-selaimen näkymä: (kuvagalleria)

The screenshot shows a web gallery interface. At the top, there is a navigation bar with "Main", "Pictures | Upload a file", "Sound | Upload a file", and "Other | Upload a file". The main content area displays a post titled "More Vortex Action" with a red and white vortex image. The post includes a date (4.1.2009), dimensions (1024 x 768), size (0.5 MB), and a rating (4/5). The author is listed as "Tekijä: JayYen (JayYen@gmail.com)". A comment section follows, containing a paragraph of text about vortices and their properties. Below the main post, there is a preview of another post titled "4D Vortex" with a blue and white vortex image.

Internet-selaimen näkymä: (yksi audio-objekti)

The screenshot shows a web gallery interface displaying an audio object. The post is titled "Organic ambient music" and features a square image of a vortex. The post details include a date of 2.1.2009, a duration of 4.54min, and a size of 12 MB. The author is "Tekijä: JayYee (JayYee@gmail.com)" and the rating is 4/5. A comment section follows, containing a paragraph of text about organic ambient music and its characteristics. The comment mentions various artists in the sub-genre.