

## Luento 9: DocBook

AS-0.110 XML-kuvauskielten perusteet

Janne Kalliola

### DocBook

- DocBook pähkinänkuoressa
  - ominaisuudet
  - käyttökohteet
- Dokumentin rakenne
  - leipätekstin rakenteet
  - lohkot ja rivin sisäiset rakenteet
  - taulukot
  - listat
  - kuvat
  - indeksit
- DocBookin käyttö
  - ongelmat

## Kirjan luvut

- Kirja ei käsittele DocBookia
- Verkosta löytyy Norman Walshin kirja DocBook: The Definitive Guide ilmaiseksi
  - linkki kirjaan löytyy osoitteesta [www.docbook.org](http://www.docbook.org)
  - tämä kirja ei kuulu tenttimateriaaliin - mutta on suureksi hyödyksi, jos tarvitset DocBookia
  - Walshin kirja keskittyy suurelta osin esittelemään elementtien syntaksia
    - ei opasta kovin hyvin DocBookin järkevään käyttöön

## DocBook pähkinäkuoressa

- DocBook on XML-pohjainen kieli tekniseen dokumentointiin
  - kyseessä siis DTD-kokoelma
  - olemassa myös SGML-versio
- Kieli on varsin yleisessä käytössä sekä yrityksissä että Open Source -projekteissa
  - tällä hetkellä käytössä on versio 4.3
- Lisää tietoa: [www.docbook.org](http://www.docbook.org)

## DocBookin ominaisuudet

- Kieli on suunniteltu erityisesti tietokoneisiin liittyvän dokumentaation laadintaan
  - sanaston painotus on kovin ohjelmisto- ja laitteistopainotteista
  - tästä huolimatta kieltä voi käyttää hyvin myös mihin tahansa dokumentaatioon
- DocBook on suunniteltu sisältöä kuvaavaksi kieleksi
  - ei ota juurikaan kantaa sisällön lopulliseen esitykseen
  - tämä on jätetty tyylisivujen ja muiden esitysjärjestelmien huoleksi
  - samasta dokumentista voidaan laatia useita erilaisia loppudokumentteja erilaisiin tarpeisiin

## Käyttökohteet

- DocBookia voi periaatteessa käyttää minkä tahansa kirjan tai artikkelin laatimiseen
  - soveltuu erityisesti (tieto)tekniisiin teksteihin sanastonsa puolesta
  - osa elementeistä hyvinkin laajasti ymmärrettäviä
    - käyttökohdekohtaista tietoa voidaan lisätä tietyillä attribuuteilla
- DocBook on laajennettavissa
  - elementtejä voidaan lisätä tai poistaa käyttötarpeen mukaan
  - tässä on syytä olla varovainen
    - muutokset eivät saisi olla liian suuria, ettei yhteensopivuus kärsisi turhaan

## Dokumentin rakenne

- DocBook-dokumentti voi olla joko kirja (book) tai artikkeli (article)
  - artikkeli on kirjaa pienempi kokonaisuus ja vertautuu pitkälti yhteen kirjan kappaleeseen
- Kirja (book) koostuu osista (part) ja luvuista (chapter)
  - kirjassa ei välttämättä tarvitse olla erillisiä osia, vaan se voi suoraan jakaantua luvuiksi
  - kirjan luvuiksi jakautuvat edelleen osioihin (sect1-5 tai section), jotka sisältävät tekstikappaleet (para)
- DocBook sisältää tuen myös erilaisille luetteloille ja indekseille

## Harvemmin käytetyt rakenteet

- Edellisellä kalvolla mainittujen osien lisäksi DocBook määrittelee
  - liitteet (appendix)
  - johdannon, alkulauseen, saatesanat (preface)
  - takakannen esittelytekstin (colophon)
  - omistustekstin (dedication)
  - referenssisivun (refentry)
    - muistuttaa Unixissa käytettyjä man-sivuja
- Kirjan yläpuolinen elementti on myös olemassa
  - joukko (set)
    - sisältää joukon kirjoja

```
<book>
  <title>Esimerkkidokumentti</title>
  <chapter>
    <title>Esimerkkiluku</title>
    <para>Kappale 1</para>
    <para>Kappale 2</para>
    <sect1>
      <title>Luvun osio</title>
      <para>Kappale 3</para>
    </sect1>
  </chapter>
</book>
```

- DocBook tukee normaalin tekstin lisäksi muutamia yleisempiä rakenteita ja suurta joukkoa tekstin tarkentimia
  - rakenteita ovat esimerkiksi taulukot ja listat
    - molempia löytyy useita tyyppejä
  - tekstin tarkentimilla annetaan lisätietoa jollekin yksittäiselle sanalle tai lauseelle
    - tietyllä tavalla verrattavissa metatietoon, mutta koskettaa vain osaa dokumentista
    - tarkentimilla voidaan määrittää joku sana esimerkiksi yrityksen tai tuotteen nimeksi
    - tarkentimet ovat kovin tietotekniikkapainotteisia (guimenuitem, interface, keycode, jne.)
    - tarkentimien käyttö ei välttämättä näy lainkaan näytettävässä tai tulostettavassa dokumentissa
    - koneellinen käsittely voi hyödyntää tarkentimia

## Lohkot ja rivin sisäiset rakenteet

- DocBookin kaikki elementit voidaan jakaa kahteen luokkaan:
  - lohkorakenteet (block)
    - näytetään tai tulostetaan kappaleina
    - molemmille puolille tulee rivin tai kappaleen vaihto
    - dokumentin rakenne muodostuu sisäkkäisistä lohkoista
  - rivin sisäiset rakenteet (inline)
    - eivät aiheuta rivin tai kappaleen vaihtumista
    - käytetään yleensä korostuksiin ja lisäarvon määrittämiseen joihinkin sanoihin

## Taulukot (1/2)

- DocBookin taulukot noudattavat CALS-mallia
  - taulukko määritellään geometrisesti
  - taulukko koostuu kolmesta osasta
    - ylätunniste (thead), runko (tbody), alatunniste (tfoot)
  - jokaisessa osassa on rivejä (row), jotka sisältävät soluja (entry)
  - solut sisältävät taulukossa esitettävät tiedot
  - solut voivat olla usean rivin tai sarakkeen mittaisia
  - sarakkeita ei välttämättä tarvitse määritellä
    - tämä on kuitenkin mahdollista

## Taulukot (2/2)

- Malli muistuttaa hyvin paljon HTML:n taulukkomallia
  - on kuitenkin rikkaampi ja siksi myös monimutkaisempi
  - taulukkoon kuuluu myös käsite taulukkoryhmä (tgroup)
    - yksi taulukko voi sisältää yhden tai useamman ryhmän, jotka sisältävät ylätunnisteen, rungon ja alatunnisteen
    - tälle käsitteelle ei ole vastinetta yleisimmissä esityskielissä

## Esimerkkitaulukko

```
<table>
  <title>Esimerkkitaulukko</title>
  <tgroup>
    <thead>
      <row>
        <entry>Otsikko 1</entry>
        <entry>Otsikko 2</entry>
      </row>
    </thead>
    <tbody>
      <row>
        <entry>Solu 1</entry>
        <entry>Solu 2</entry>
      </row>
    </tbody>
  </tgroup>
</table>
```

## Listat

- DocBook tukee useita erilaisia listatyyppejä
  - numeroitu (orderedlist)
    - numerointi voidaan määrittellä
  - numeroimaton (itemizedlist)
  - määrittelylista (variablelist)
    - yhtä tai useampaa termiä (term) määrittellään
    - muistuttaa kaksisarakkeista taulukkoa
  - segmentoitu lista (segmentedlist)
    - ei ole ahtaasti katsottuna lista, vaan toinen tapa kuvata taulukkoa
    - listassa määrittellään joukko otsikoita ja näiden alle tulevat arvot rivi kerrallaan

## Esimerkkilista

```
<orderedlist>
  <listitem><para>Yksi</para></listitem>
  <listitem><para>Kaksi</para></listitem>
  <listitem><para>Kolme</para></listitem>
</orderedlist>

<itemizedlist>
  <listitem><para>Yksi</para></listitem>
  <listitem><para>Kaksi</para></listitem>
  <listitem><para>Kolme</para></listitem>
</itemizedlist>
```



## Kuvien käyttö

- DocBookissa on useita elementtejä kuvien ja muiden ulkopuolisten elementtien liittämiseen dokumenttiin (imageobject)
  - on mahdollista sekä viitata kuvatiedostoon (imagedata fileref="") että sisällyttää tiedoston sisältö dokumenttiin (imagedata entityref="")
- Kieli ei ota kantaa sallittuihin tiedostomuotoihin
  - ongelma jätetty esitysjärjestelmän harteille
- Kuvien lisäksi on mahdollista liittää muita mediaolioita (mediaobject)
  - esimerkiksi animaatioita, elokuvaa (videoobject) tai ääniä (audioobject)
  - voivat olla varsin käytännöllisiä esimerkiksi online-dokumentaatioissa
- Kuvat voivat olla rivien sisäisiä tai erillisiä lohkoja (figure)

## Esimerkkikuva

```
<figure>
  <title>Esimerkkikuva</title>
  <mediaobject>
    <imageobject>
      <imagedata fileref="esimerkki.png"
        contentwidth="2cm" contentheight="2cm" />
    </imageobject>
  </mediaobject>
</figure>
```

## Muita käteviä lohkoelementtejä

- Huomautukset (admonition)
  - vinkit (tip)
  - varoitukset (warning)
  - huomautukset (note)
- Esimerkit (example)
- Näyttökaappaukset (screenshot, screen)
- Ohjelmalistaukset (programlisting)
- Kaavat (equation)
- Alaviitteet (footnote)

## Käteviä rivin sisäisiä elementtejä

- Lyhenteet (abbrev, acronym)
- Korostukset (emphasis)
- Kielen vaihdot (foreignphrase)
- Ylä- ja alaviitteet (superscript, subscript)
- Merkityksen lisääminen
  - sähköposti (email)
  - literaali (literal)
  - merkinne (markup, token)
  - SGML-elementti (sgmltag)
- Merkityksen poisto (wordasword)

## Muodolliset ja epämuodolliset osat

- DocBookissa on määritelty melkein kaikille dokumentin osille sekä muodolliset (formal) ja epämuodolliset (informal) elementit
  - muodollisissa elementeissä täytyy olla aina otsikko (title)
  - epämuodollisissa elementeissä ei ole koskaan otsikkoa
- Yleensä kirjan luvut ovat muodollisia ja tekstikappaleet epämuodollisia
  - kirjasta riippuen taulukot, listat ja kuvat voivat olla muodollisia tai epämuodollisia

## Ristiinviittaukset

- Ristiinviittaukset ovat tärkeitä teknisessä dokumentaatioissa
  - lisätiedon haku
- Docbook määrittää ristiinviittauselementin (xref), linkin (link), ulkoisen linkin (ulink - viittaus URI:lla)
  - elementti voi viitata mihin tahansa DocBookin elementtiin
  - viittaus kohdistuu elementtien id-tribuuttiin
    - id-tribuutin arvon on oltava yksikäsitteinen koko dokumentissa
- DocBook ei huolehdi ristiinviittausten tarkistamisesta tai toiminnasta
  - jätetty käsittelyohjelman huoleksi
  - Validointi DTD:ta vastaan tosin tarkistaa viittausten eheyden
  - muutetaan esimerkiksi linkeiksi tai sivunumeroviittauksiksi

## Metatieto

- Kirjoihin tai artikkeleihin on syytä liittää metatietoa
  - esimerkiksi kirjoittaja, julkaisu yms.
  - DocBook tarjoaa joukon elementtejä metatiedon määrittämiseen
  - käytössä oleva malli on DocBookin itsensä määrittelemä
    - mitään metatietostandardia ei valitettavasti käytetä
    - tällaiset metatietosirpaleet voidaan tietysti liittää dokumenttiin myös dokumentin ulkopuolella
- Dokumentin sisällä voi olla myös metatietoa
  - esimerkiksi joku nimi voidaan merkitä yrityksen tai organisaation nimeksi
    - lopputuloksessa tämä merkintä ei välttämättä näy lainkaan
    - helpottaa asiasanojen hakua tekstistä
  - DocBookissa on hyvin rikas määrittelyelementistö
    - esimerkiksi osoitteissa voidaan määrittellä jokainen komponentti, kuten postinumero, erikseen

## Indeksit

- DocBook sisältää elementtejä erilaisten indeksien, kuten sisällysluettelon ja hakemiston, laadintaan
  - yleensä näitä indeksejä ei kuitenkaan tehdä käsin
  - indeksit ovat ulkonäköriippuvaisia
    - sivunumerot saattavat vaihdella riippuen sivujen ja tekstin koosta sekä käytetyistä kirjasimista
    - kaikissa medioissa ei ole sivunumeroja
- Indeksien generoinnissa voidaan toimia kahdella tavalla
  - laatia valmis indeksi loppudokumenttiin
  - luoda DocBook-tiedosto, joka sisältää indeksin määrittelyn
    - tämä tiedosto voidaan edelleen liittää muuhun dokumentaatioon
    - koko dokumentaatio muokataan kerralla esityskuntoon

## Käytettävissä olevat indeksit

- DocBook tukee seuraavia indeksityyppejä
  - sisällysluettelo (toc)
  - hakemisto (index)
  - sanasto (glossary)
  - bibliografia (bibliography)
- Indeksit voivat olla koko kirjan laajuisia tai sisältää vain yhden osan tai luvun tiedot

## DocBookin suhde ulkonäköön (1/2)

- DocBook on tarkoitettu pelkästään dokumentin sisällön kuvaamiseen
  - ei ole olemassa elementtejä ulkonäköseikkojen asettamiseen
  - dokumentin laatijan ei pitäisi tarvita välittää lopullisesta ulkonäöstä
- Raja on valitettavan häilyvä
  - moni elementti sisältää ulkonäköä määritteleviä attribuutteja
    - tarkalleen ottaen nämä pitäisi hoitaa käsittelyohjeina
    - standardi ei kuitenkaan sisällä tällaisia käsittelyohjeita tai edes suosituksia niistä
  - kyse on aina kompromissista

## DocBookin suhde ulkonäköön (1/2)

- Osa elementeistä määrittelee ulkonäköä tai siihen liittyviä asioita liiankin tarkoin käytännön kannalta
  - esimerkiksi, lähdedokumentissa ei ole järkevää viitata numerolla sivunumeroihin
    - tämä on mahdollista DocBookissa
    - pitäisi olla tunnistisiin tai vastaaviin perustuva tapa

## DocBook-dokumentin laadinta

- DocBook-dokumentteja voidaan laatia millä tahansa editorilla, jolla voidaan tuottaa XML-dokumentteja
  - normaali tekstieditori siis käy
    - käyttö voi olla hankalaa
  - XML-editori parempi valinta
    - varsinkin, jos DocBookin DTD:n voi ladata käyttöön dokumenttia laadittaessa
  - osaa XML-kyvykkäistä tekstinkäsittely- ja taitto-ohjelmista, kuten Adobe FrameMaker+SGML, voidaan käyttää DocBook-dokumenttien laadintaan
    - yleensä tarvitaan DTD:n lisäksi muuta tietoa dokumentin visuaalisen ulkoasun laatimiseksi
- Dokumentti voidaan myös validoida erikseen DTD:ta vastaan

## DocBook-dokumentin käyttö

- Valmis DocBook-dokumentti ei sinällään ole suoranaisesti käytettävissä
  - ihmisten hankala lukea tekstiä XML-merkintöjen seasta
  - ei ole valmiita ohjelmia dokumenttien näyttämiseen
- Dokumentti muunnetaan toiseen muotoon XSLT:lla
  - tyylisivun voi laatia itse
  - verkossa on useita valmiita tyylisivuja
    - tuottavat HTML:a tai XSL-FO-koodia
    - osa näistä tyylisivuista mahdollistaa omien määritysten käytön

## DocBookin ongelmat

- DocBook on hyvin laaja standardi
  - elementtejä on paljon ja oikean elementin löytäminen voi olla vaikeaa
- Sanasto sopii vain ohjelmistojen ja järjestelmien kuvaamiseen
  - tietoteknisillä termeillä liikaa painoarvoa
  - vastaavasti yleiskäyttöisiä lisäarvoa antavia elementtejä kovin vähän
    - nämäkin hyvin erikoistuneita (postinumero, katuosoite, jne.)

## Simplified DocBook

- Simplified DocBook on nimensä mukaisesti yksinkertaistettu versio DocBookista
  - suuri joukko elementtejä karsittu
  - haluttu madaltaa kynnystä aloittaa kielen käyttö
- Myös tämän kielen määrittymiset löytyvät osoitteesta [www.docbook.org](http://www.docbook.org)

## Yhteenveto

- DocBook on XML-pohjainen kieli tekniseen dokumentointiin
  - DocBook on sisältöä kuvaava kieli
  - kieli on suunniteltu erityisesti tietokoneisiin liittyvän dokumentaation laadintaan
- DocBook-dokumentti voi olla joko kirja tai artikkeli
  - kirja koostuu osista ja luvuista
- DocBook tukee normaalin tekstin lisäksi muutamia yleisempiä rakenteita ja suurta joukkoa tekstin tarkentimia
  - DocBookin kaikki elementit voidaan jakaa lohkorakenteisiin ja rivin sisäisiin rakenteisiin
- Käyttöä varten dokumentti muunnetaan toiseen muotoon XSLT:lla



**Kysymyksiä? Kommentteja?**