



XML – metakieli

- Extensible Markup Language
- rakenteellinen esitystapa tiedon vaihtoon, talletukseen, yhdistämiseen ja julkaisemiseen
- yleinen rakenteenkuvauskieli, eli nk. metakieli
- toimintaperiaate, ei työkalu
- "standardiperheen" ydin
- pohjana SGML

XML = Extensible Markup Language



- monia käyttötarkoituksia ja –mahdollisuuksia kun yhdistetään muihin standardeihin, kieliin tai tuotteisiin
- sekä ihmisen että koneen luettavissa:
 - helppo lukea vaikkei varsinaista koodia tulkittaisikaan millään parserilla (kuten IE6:ssä olevalla)
- riippumaton alustasta tai ohjelmointikielistä
- W3C suositus (recommendation)
 - versio 1.0
 - 6.10.2000



W3C World Wide Web Consortium

- vuodesta 1994
- standardoi www-avaruutta
- kuten HTML - standardit "Recommendations"
- "vapaa" organisaatio, asianharrastajien perustama, yli 400 jäsentä, suuryritykset ja yliopistot
- <http://www.w3.org/Consortium/Member/List>
- suositukset ilmaisia, vapaaehtoisia



XML – dokumentti-ilmentymä

```
<!-- Dokumentti-ilmentymän esimerkki (osa) -->
<university>
  <department>
    <name>
      Humanistinen tiedekunta
    </name>
    <address>
      Fabianinkatu
    </address>
  </department>
</university>
```



"XML is for structuring data"

- rakenteellinen tieto: laskentataulukot, asennusparametrit, kirjastokortit, vrt. tietokannat
- XML antaa puumallin mukaisen rakenteen tiedon järjestämiseen
- XML ei ole ohjelmointikieli, mutta standardiperheeseen kuuluu XSLT
- laajennettava (extensible), tukee kansainvälisyyttä ja paikallistamista = täysin Unicode-yhteensopiva



XML:n käyttöalueita

- Dokumentaatio (alkuperäinen)
 - Tekninen dokumentointi: manuaalit, termipankit, varaosaluettelot, kieliversiot
- Julkaiseminen
 - documentit, DocBook
 - metadata, RSS, uutisjakelu (syndication & podcasting)
- Semanttinen tieto, ontologiat (semantic web)
- Uudelleenkäyttö:
 - sama informaatio voidaan välittää eri kanavien kautta: monikanavajulkaiseminen
 - WWW, PDA, mobiililaitteet, DVD, paino

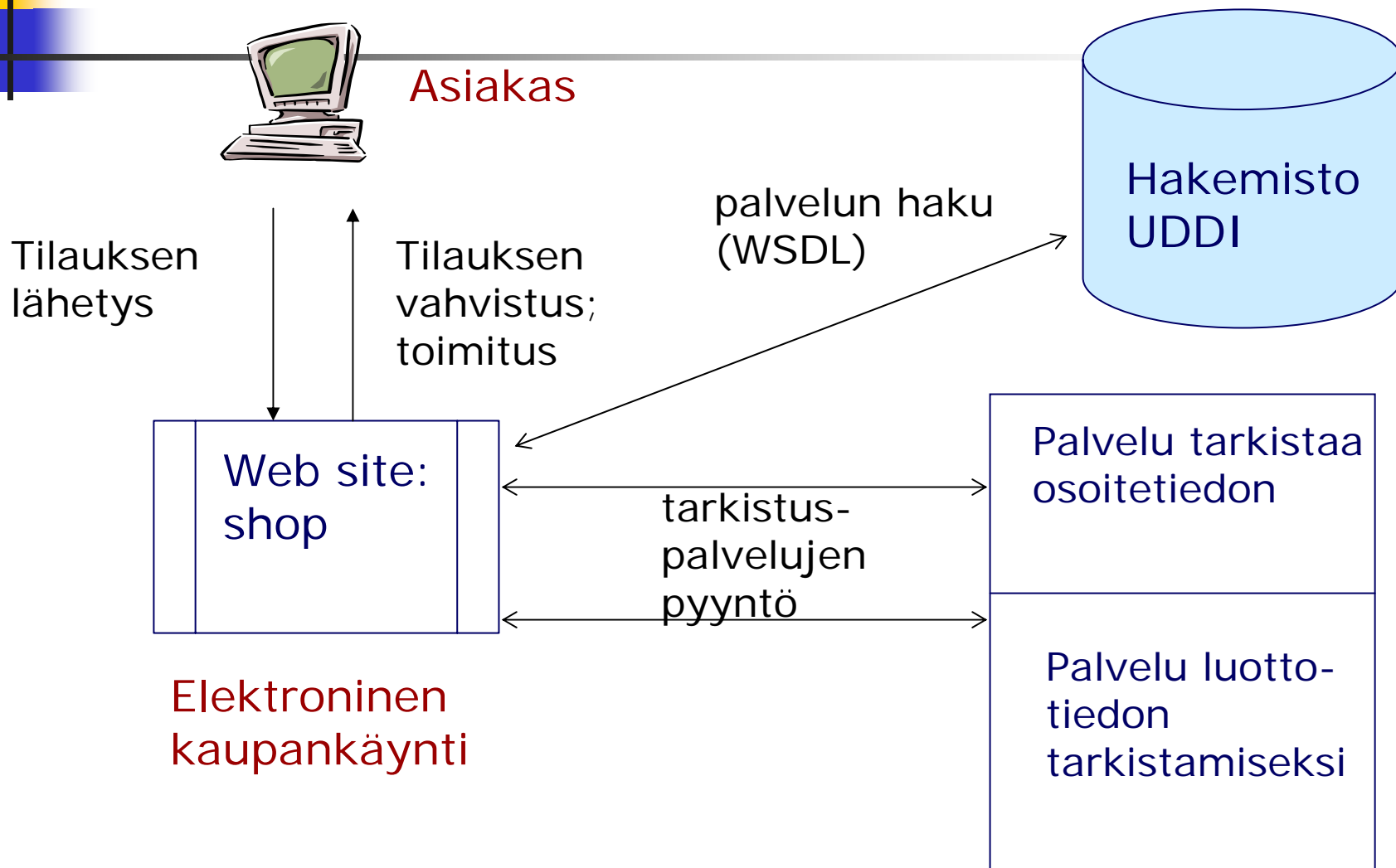


XML sovellusalueita

Multimedia

- Voice XML, SVG, SMIL, X3D
- Elektroninen kaupankäynti, EDI (Electronic Data Interchange)
 - ebXML: electronic business XML, UN/CEFACT ja OASIS
 - pääosat: ebXML storage, CPP – Collaboration Protocol Profile, CPA – Collaboration Partner agreement
 - BizTalk, WebServices
 - Sähköinen laskutus
- Paikkatieto: GIS

Web Services tapausesimerkki





XML:n käyttöalueita

- Nykyään ?
- Selainten sisäinen esitysmuoto
- Microsoft: .NET ja Office sisäinen esitysmuoto
- Ajax ja XHTML, Googlemaps



XML ulkoasu

XML näyttää samanlaiselta kuin HTML

- Elementit (tagit) ja attribuutit
- `<p>` tarkoitus käyttäjän valittavissa

XML on tekstimuotoista, mutta tarkoitettu sovellusten käyttöön

- tarkkaan määritelty syntaksi,
- tavoitteena yksiselitteinen tulkinta
- ei ole pyritty tiiviYTEEN



XML teknologiaperhe

- XML 1.0
- DTD:t ja Skeemat
- nimiavaruudet (Namespaces)
- Xpath – kieli XML-dokumentin osien paikallistamiseen ja valintaan
- XSLT muunnokset
- (CSS tai) XSL-Formatting Objects
- DOM sovellusliittymästandardi



Semanttinen verkko

- XML:n perustalle on luotu RDF ja semanttinen Web
- vrt. RSS
- tarkoituksena on luoda luokitteluja ja sanastoja
- tiedon löytyminen määriteltyjen hakusanojen avulla, ei vapaatekstihaulla
- eri alojen sanastotyöryhmiä
- esim. Dublin Core ja RosettaNet



XML vs. HTML

- HTML on luotu tiettyyn tarkoitukseen
- XML ei ole tarkoitettu vain yhteen tarkoitukseen, vaan melkein mihin vain
- HTML:llä on rajattu määrä tageja
- XML:llä ei ole ennalta määrättyjä tageja
- XML on tiukasti hierarkkinen ja sitä lukevat parserit tarkkoja siitä



XML –dokumentti-ilmentymä

```
<?xml version="1.0" ?>
<!-- Dokumentti-ilmentymän esimerkki (osa) -->
<university>
  <department>
    <name>
      Humanistinen tiedekunta
    </name>
    <address>
      Fabianinkatu
    </address>
  </department>
</university>
```



Dokumentti-ilmentymä

- sisältää dokumentin varsinaisen sisällön, joka on merkattu sovitulla tavalla
- itseään kuvaavia tunnisteita
- auttaa tiedon tulkinnassa
- elementit ja alielementit
- tekstit ja kommentit



XML -prosessorit

- XML-jäsennin (=parser)
 - Löytää virheet
 - Tuottaa tietoa muille sovelluksille
- Entiteettien (dokumenttiosien) hallinta
 - Koostaa XML-dokumentit
 - Myös eri tiedostoista



Elementit

- Dokumentin (looginen) rakenneosa
- erottavat tunnisteet
 - alkutunniste
 - lopputunniste
- elementin nimi
- elementin sisältö
 - voi olla muita elementtejä, tai
 - tekstiä

```
<capital>Helsinki</capital>
```

```
<country>
```

```
<cname>Finland</cname>
```

```
<capital>Helsinki</capital>
```

```
</country>
```



Elementtien sisäkkäisyys: säännöt

- aina p.o. vastaavat alku- ja lopputunnisteet
- isot ja pienet kirjaimet merkitseviä
 - <otsikko> eri kuin <Otsikko>
- elementtien nimien täytyy noudattaa XML:n sääntöjä
- elementin täytyy sisältyä täysin toiseen elementtiin (ei limittäin!)
- elementtihierarkia
 - juuri = dokumenttielementti - yksi ainoa!
 - Puumuotoinen rakenne
- **Hyvinmuodostuneisuus (= well-formed)!**



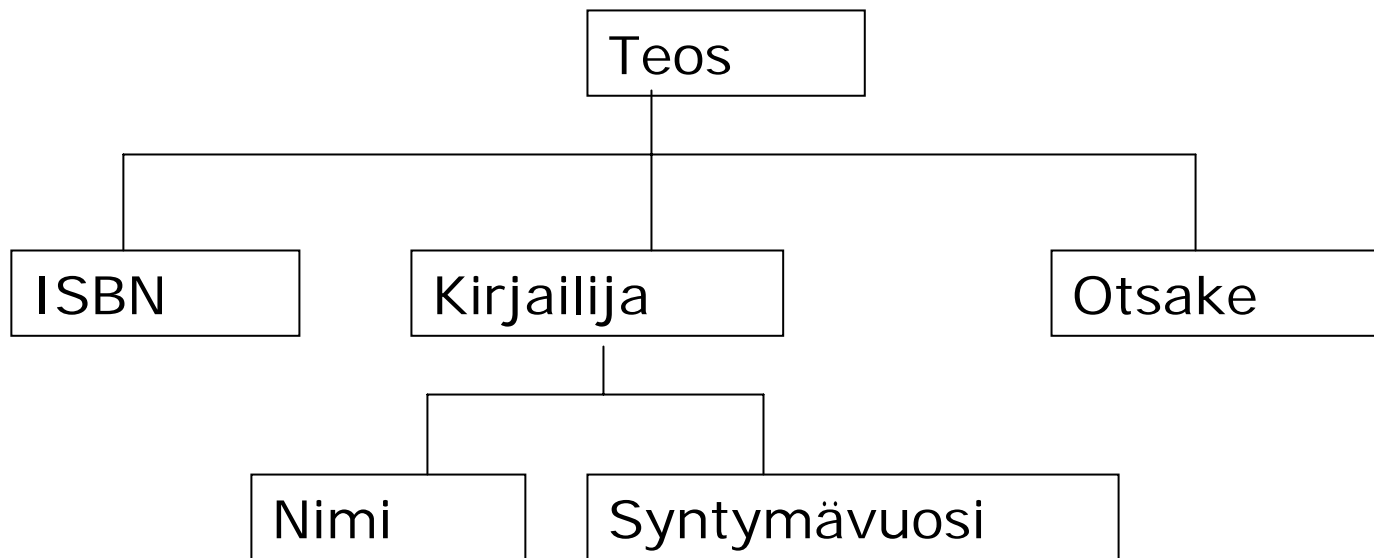
Sisäkkäiset elementit

```
<pieni_esimerkki>  
  <first>sisäkkäisyys</first>  
  <second>oikein</second>  
</pieni_esimerkki>
```

```
<small_example>  
  <first>sisäkkäisyys<second>  
  </first>aivan väärin</second>  
</small_example>
```



Dokumenttipuu





Elementtien nimet, säännöt

- alkavat kirjaimella tai alaviivalla _
- muualla sallitaan myös numerot, väliviiva - ja piste .
- nimissä ei saa olla välilyöntejä, ei kaksoispistettä, eikä yhtäläisyysmerkkiä tms. välimerkkejä
- nimen alussa ei saa olla XML, xml, Xml
- alussa ei välilyöntiä <merkkejä >, mutta lopussa voi olla

- mitkä seuraavista ovat sallittuja?
 - <first.name> <xml-juuri> <123> <iso omena>
 - <p>kappale</P>



Elementtien sisällöt

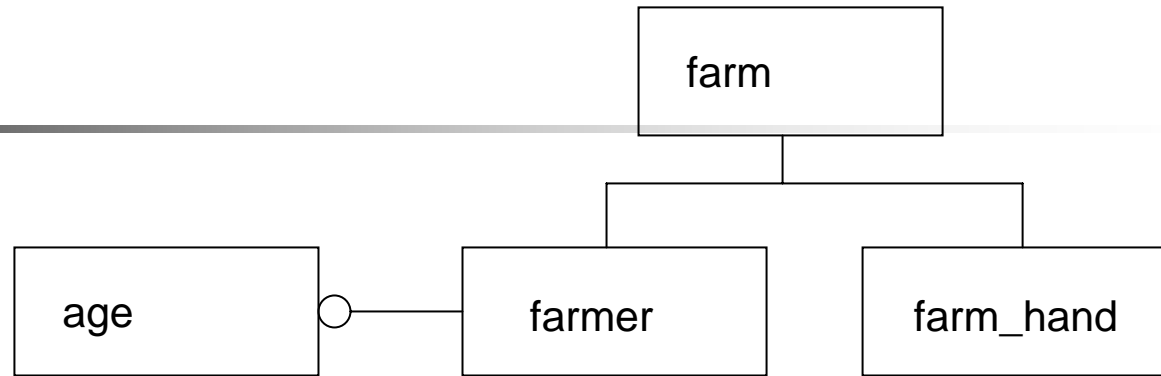
- Elementti voi sisältää
 - muita elementtejä,
 - tekstiä, tai
 - molempia
- Elementti voi myös olla tyhjä
 - Elementillä ei ole sisältöä
 - `<nothing></nothing>`
 - `<turhaa/>`
- miksi: sisältö voi olla esim. muualla
 - tyhjä elementti sisältää viittauksen
 - `<image file="pict.jpg"/>`



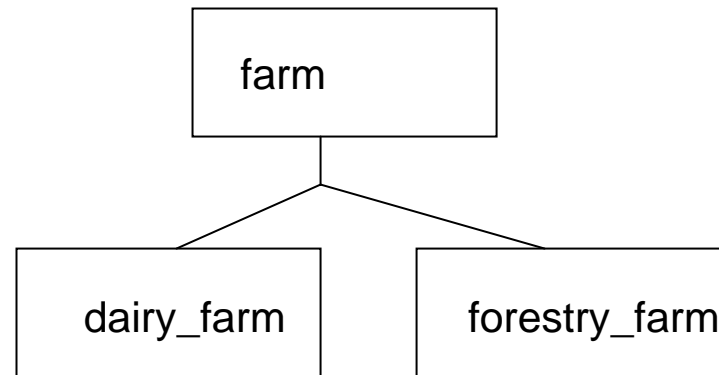
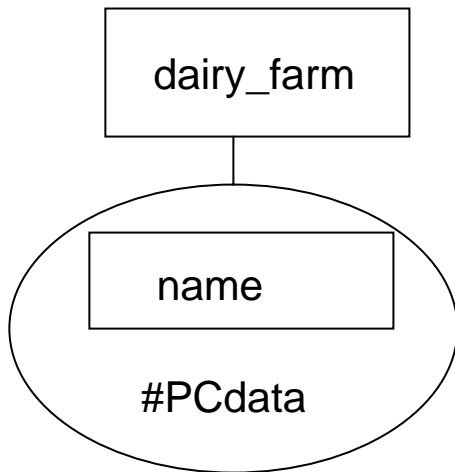
Attribuutit (määritteet)

- Elementin ominaisuuden tai sisällön tarkenne
- Liitetään alkutunnisteisiin (tai tyhjien elementtien tunnisteisiin)
 - attribuutin nimi
 - attribuutin arvo
- aina vain yksi arvo
- esim. `<tekijä rooli="toimittaja">Ahti</tekijä>`

Puukuvauksia



<!ELEMENT farm (farmer, farm_hand)>



<!ELEMENT farm (dairy_farm | forestry_farm)>



XML - metakielen kielioppi

- Käyttäjä määrittelee merkkauskielen
 - elementtien nimet
 - elementtien järjestyksen ja hierarkian
 - tietyt varatut sanat
- DTD tai skeema
= kielioppi

<!-- Dokumenttityypin esittely (DTD) esimerkki (osa) -->

```
<!ELEMENT university (department+)>
```

```
<!ELEMENT department (name, address)>
```

```
<!ELEMENT name (#PCDATA)>
```

```
<!ELEMENT address (#PCDATA)>
```



DTD tai XML Schema

- Dokumenttityypin esittely eli rakenne-esittely
- yksi sääntö/elementti
 - nimi
 - sisältö
- dokumentti-ilmentymien "kielioppi"
- "säännölliset lausekkeet"
- (ei välttämätön)



Skeeman hyvät puolet

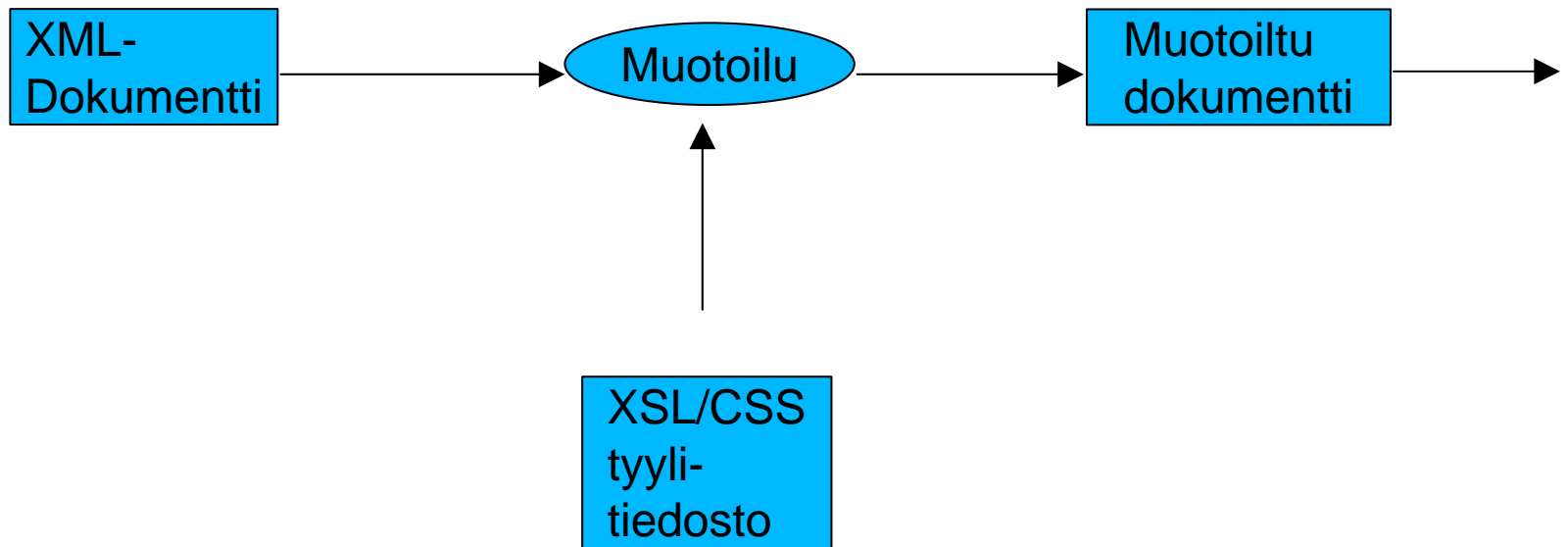
- validoiva jäsennin
 - tarkistaa että dokumentti on skeeman mukainen
- tunnisteiden johdonmukainen käyttö
- eri sovellusten DTD-standardit ja skeemat
 - yhteinen sanasto



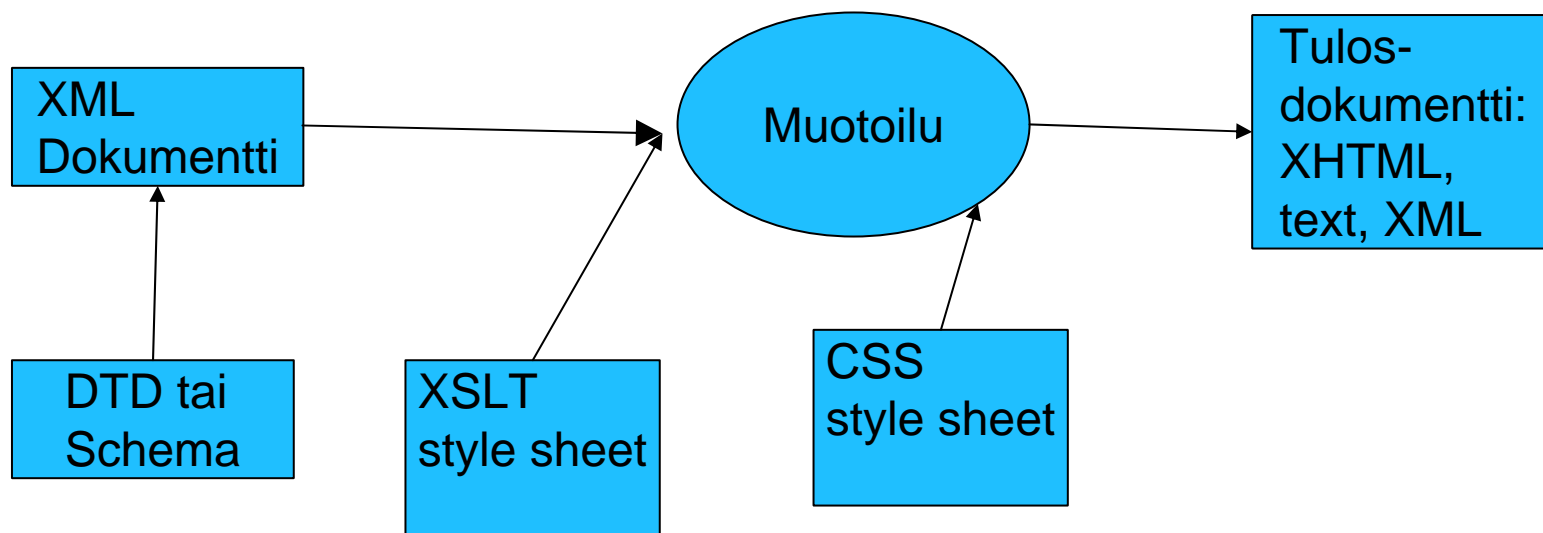
Tyylitiedostot

- Tulosmuotoilua varten
- Useampia tyylitiedostoja/skeema ja/tai dokumentti
- Cascading Style Sheets (CSS)
- XML Stylesheet Language (XSL)

Julkaisuprosessi



XML -standardit käytössä





SGML

- Standard Generalised Markup Language
- standardi 1986 (ISO)
- sisälsi DTD:n ja tyylitiedostot
- monimutkainen
- työkalut monimutkaisia ja hankalia tehdä, kalliita ja vähän
- XML on SGML: osajoukko (melkein)
- Käytetään edelleen, esim. lentokoneiden ja paperikoneiden dokumentaation hallinta

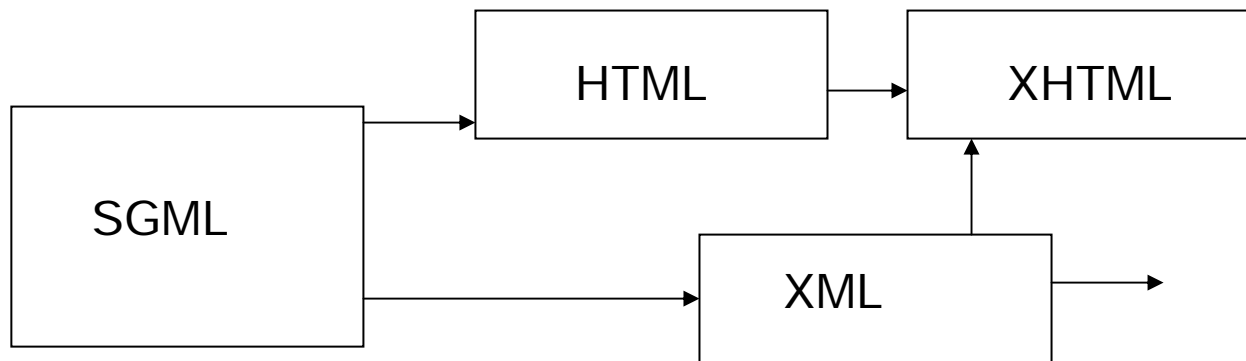


HTML

- Hypertext Markup Language (ja HTTP protokolla)
- perustui SGML:ään
- iso menestys
- ei-standardoidut laajennukset
- paljon työkaluja
- ennaltamääritelty muoto
 - hyvä kun ainoastaan halutaan esittää tietoa
 - kauniisti selaimessa

XML - SGML - HTML

- XML yhdistää SGML:n ja HTML:n piirteitä
- paljon työkaluja
- XHTML, XML-suosituksen mukainen HTML
- ei ratkaisua kaikkiin ongelmiin
- kaikkia kolme kieltä (XML, HTML, SGML) tarvitaan





XML:n suunnittelutavoitteet I

1. XML shall be straightforwardly usable over the Internet.
2. XML shall support a wide variety of applications.
3. XML shall be compatible with SGML.
4. It shall be easy to write programs which process XML documents.
5. The number of optional features in XML is to be kept to the absolute minimum, ideally zero.



Suunnittelutavoitteet II

6. XML documents should be human-legible and reasonably clear.
7. The XML design should be prepared quickly.
8. The design of XML shall be formal and concise.
9. XML documents shall be easy to create.
10. Terseness in XML markup is of minimal importance