

Markeringsspråk og XML

```
<?xml version = "1.0" encoding="utf-8"?>
<eventyr>
Det var en gang...
...
og snipp, snapp, snute, så var det eventyret ute.
</eventyr>
```

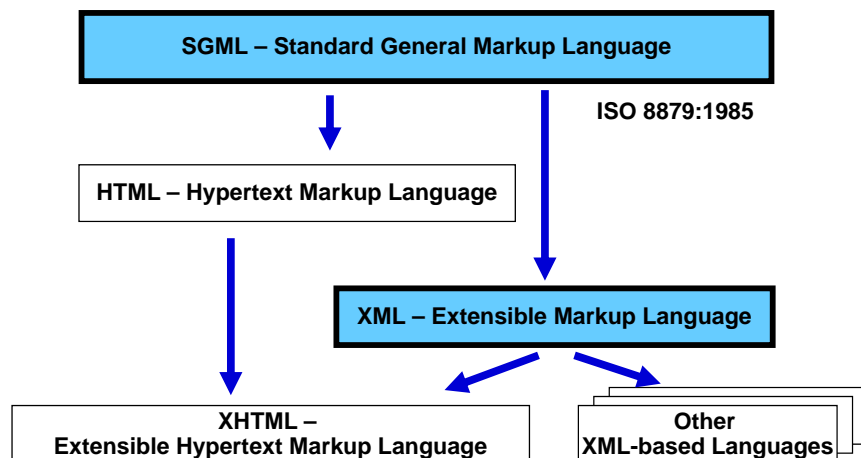
Kapittel 3

3. september 2008

Læringsmål – XML

- Forstå prinsippene bak XML og XHTML.
- Forstå hva XML kan brukes til – og hvordan.
- Forstå hva XML er – og hva det ikke er.
- Kunne lage XML-dokumenter.
- Kunne bruke dokumenttypedefinisjoner (DTDer) og XML skjemaer.

SGML-familien



Markering av tekst

- Spesifikk markering (eksempel fra WordStar)

```
.PL 66
.MT 6
.MB 9
^B Innhold INF1040 ^B
I INF1040 skal vi blant annet lære ...
```

page length 66 linjer
margin top = 6 linjer
margin bottom = 9 linjer
switch bold on and off

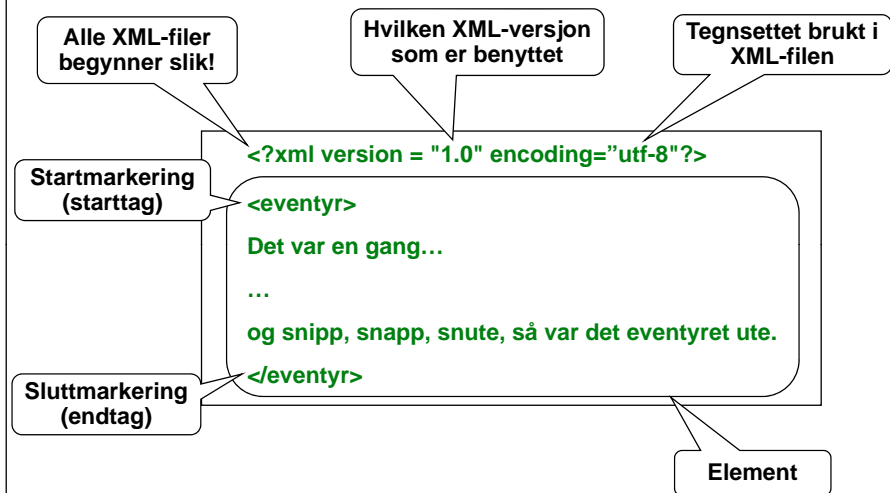
- Generalisert markering (eksempel fra XHTML)

```
<h1>Innhold INF1040</h1>
<p>I INF1040 skal vi blant annet lære...</p>
```

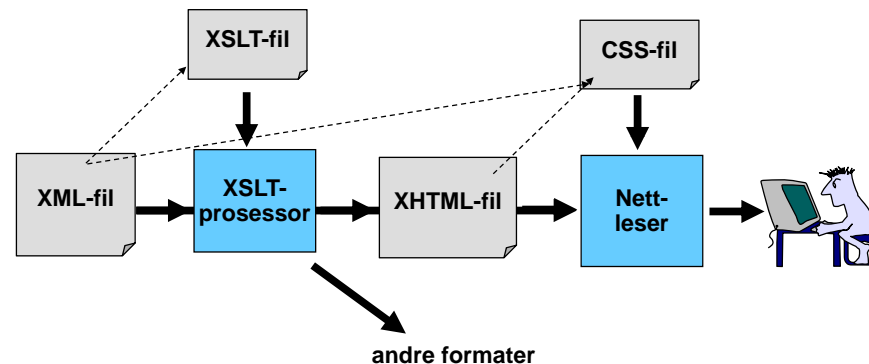
- Generisk markering (eksempel fra XML)

```
<kapittel>
<kapittelnavn>Innhold INF1040</kapittelnavn>
<avsnitt>I INF1040 skal vi blant annet lære... </avsnitt>
</kapittel>
```

Et meget enkelt XML-eksempel



Fremvisning av XML-dokumenter



Et litt mer avansert XML-eksempel

```
<?xml version = "1.0" encoding="ISO-8859-1"?>
<emne>
<emnekode>INF1040</emnekode>
<emnenavn>Digital representasjon</emnenavn>
<emnebeskrivelse>I INF1040 skal vi blant annet lære om hvordan vi i en
datamaskin representerer
<temaliste>
<tema>tekst</tema>
<tema>tall</tema>
<tema>lyd</tema>
<tema>bilder</tema>
</temaliste>
og hva dette innebærer for
<stikkord>lagring</stikkord>
<stikkord>overføring</stikkord>
<stikkord>fremvisning</stikkord>
</emnebeskrivelse>
</emne>
```

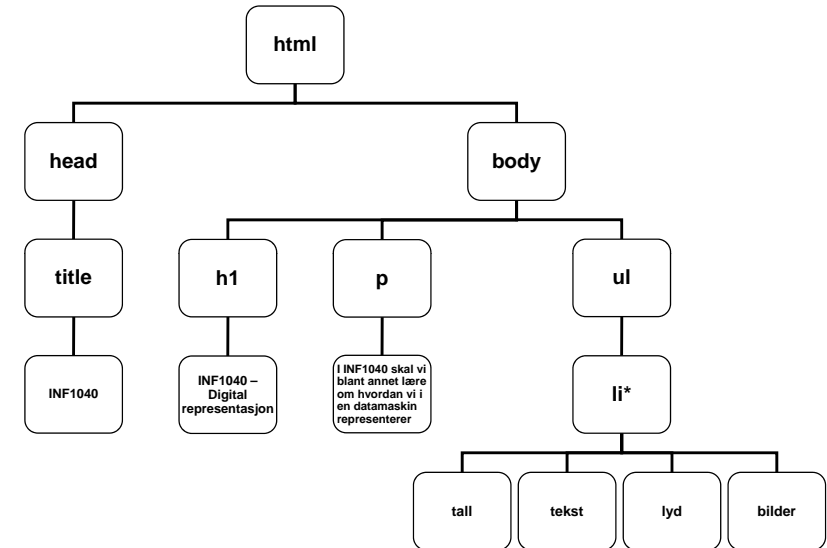
Elementene danner en hierarkisk struktur



XHTML – en spesialisering av XML

```
<?xml version="1.0" encoding="utf-8"?>
<!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>
  <title>INF1040</title>
</head>
<body>
  <h1>INF1040 – Digital representasjon</h1>
  <p> I INF1040 skal vi blant annet lære om hvordan vi en datamaskin
representerer </p>
  <ul>
    <li>tekst</li>
    <li>tall</li>
    <li>lyd</li>
    <li>bilder</li>
  </ul>
</body>
</html>
```

Strukturen i XHTML-eksempelet



Oppbyggingen av et element

- Hvert element starter med en startmarkering og avsluttes med en tilsvarende sluttmarkering:
`<elementnavn>elementinnhold</elementnavn>`
Eksempel fra XHTML:
`<h1>INF1040 – Digital representasjon</h1>`
- Hvis elementet er tomt kan start- og sluttmarkeringen slås sammen:
`<elementnavn />`
`
`
- Startmarkeringen kan inneholde ett eller flere attributter med tilhørende verdier:
`<elementnavn attributt="verdi">elementinnhold</elementnavn>`
`UiO`
``

Regler for elementnavn i XML

- Kan inneholde bokstaver, tall og andre tegn.
- Ikke-engelske bokstaver er tillatt, men det er ikke sikkert all programvare forstår dem.
- Kan ikke inneholde mellomrom (space).
- Kan ikke starte med tall, skilletegn eller xml, XML, Xml
- Kan starte med _ (underscore)
- Unngå tegnene – . :
 - De kan lett misforstås

Spesielle tegn

- Tegnene < > & har en spesiell betydning i XML (de brukes som "escape"-tegn)
- Hvis slike tegn inngår i teksten, må de kodes på en spesiell måte: &tegnnavn; eller &#kodepunkt;

□ altså:

<	<	<
>	>	>
&	&	&

x betyr at kodepunktet er angitt i heksadesimal!

Hva kan XML brukes til?

- XML brukes i dag hovedsaklig til
 - overføringsformat mellom systemer
 - lagringsformat for semistrukturerte dataer
- Kategorisering av XML-dokumenter
 - dokumentsentrerte XML-dokumenter
 - datasentrerte XML-dokumenter

Eksempel dokumenttypedefinisjon (DTD)

- Dokumenttypedefinisjon for emnebeskrivelser (emne.dtd):

```
<!ELEMENT emne (emnekode, emnenavn, emnebeskrivelse)>
  <!ELEMENT emnekode (#PCDATA)>
  <!ELEMENT emnenavn (#PCDATA)>
  <!ELEMENT emnebeskrivelse (#PCDATA|temaliste|stikkord)*>
  <!ELEMENT temaliste (tema*)>
  <!ELEMENT tema (#PCDATA)>
  <!ELEMENT stikkord (#PCDATA)>
```

Bruk av DTD

- På filen emne.xml:

```
<?xml version = "1.0" encoding="utf-8"?>
<!DOCTYPE emne SYSTEM "emne.dtd">
<emne>
  <emnekode>INF1040</emnekode>
  ...
  ...
</emne>
```

c

- DTD-en til XHTML strict finnes på <http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd>

Eksempel XML skjema

□ XML-skjema for emnebeskrivelser (emne.xsd):

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="emnekode" type="xs:string"/>
      <xs:element name="emnenavn" type="xs:string"/>
      <xs:element name="emnebeskrivelse">
        <xs:complexType mixed="true">
          <xs:element name="tema" type="xs:string"
            minOccurs="2" maxOccurs="10"/>
          <xs:element name="stikkord" type="xs:string"
            minOccurs="0" maxOccurs="unbounded"/>
        </xs:complexType>
      </xs:sequence>
    </xs:complexType>
  </xs:schema>
```

Bruk av XML-skjema

□ På filen emne.xml:

```
<?xml version="1.0" encoding="utf-8"?>
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
<emne xsi:schemaLocation="emne.xsd">
  <emnekode>INF1040</emnekode>
  ...
  ...
</emne>
```

Planen videre

- I dag har vi
 - gjort oss ferdige med **tegn** og **tekst**.
 - sett på **XML**.
- Neste onsdag ser vi spesielt på **overføring av data**.
- Plenumsgruppene går nå over til å først og fremst være gjennomgang av eksempler og oppgaver.