

# Markeringspråk og XML

## Nettsider og XHTML

```
<?xml version = "1.0" encoding="ISO-8859-1"?>  
<eventyr>  
Det var en gang...  
...  
og snipp, snapp, snute, så var det eventyret ute.  
</eventyr>
```

Læreboka kapittel 3 og 4

5. september 2007

# XML og XHTML

- ❑ **XML = EXtensible Markup Language**
  - Laget for å beskrive informasjon.
  - Brukes for
    - » Lagring av semistrukturerte data.
    - » Overføring av data mellom systemer.
  
- ❑ **XHTML = EXtensible HyperText Markup Language**
  - Laget for å vise frem informasjon.
  - Brukes for
    - » Markering av tekster som skal vises frem av en nettleser.
  
- ❑ **Standardisert av World Wide Web Consortium – se <http://www.w3.org/>**

# Markering av tekst

## ❑ Spesifikk markering (eksempel fra WordStar)

.PL 66	page length 66 linjer
.MT 6	margin top = 6 linjer
.MB 9	margin bottom = 9 linjer
^B Innhold INF1040^B	switch bold on and off
I ^BINF1040^B skal vi blant annet lære ...	

## ❑ Generalisert markering (eksempel fra XHTML)

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

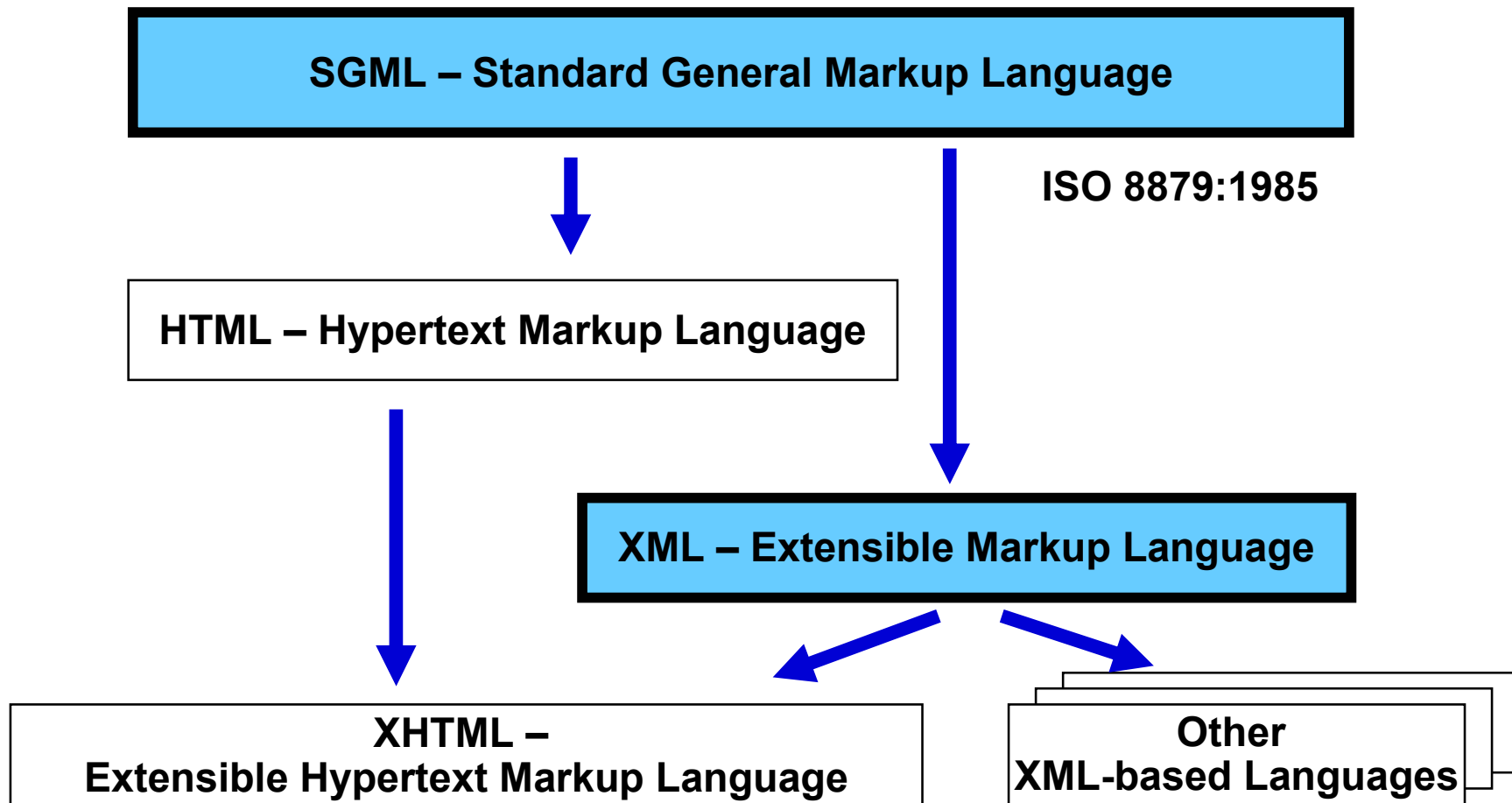
## ❑ Generisk markering (eksempel fra XML)

```
<chapter>  
<chapter name>Innhold INF1040</chapter name>  
<paragraph>I <span>INF1040</span> skal vi blant annet lære...  
</paragraph>  
...  
</chapter>
```

# Direkte redigering av markert tekst

- ❑ **Trinn 1: Generalisert eller generisk markering**  
Hver enkelt del av teksten deklarereres til å være av en bestemt type (f.eks. overskrift, brødtekst, sitat, liste...)
- ❑ **Trinn 2: Formatering**  
Hver av teksttypene tilordnes en kombinasjon av formater (ofte kalt en stil).
- ❑ **Prinsipp: Skill mellom innhold (hva er teksten) og visualisering (hvordan vise frem teksten).**

# SGML-familien



# Et meget enkelt XML-eksempel

Alle XML-filer  
begynner slik!

Tegnsettet brukt  
i XML-filen

Startmarkering  
(starttag)

```
<?xml version = "1.0" encoding="ISO-8859-1"?>
```

```
<eventyr>
```

Det var en gang...

...

og snipp, snapp, snute, så var det eventyret ute.

Sluttmarkering  
(endtag)

```
</eventyr>
```

Element

# Et litt mer avansert XML-eksempel

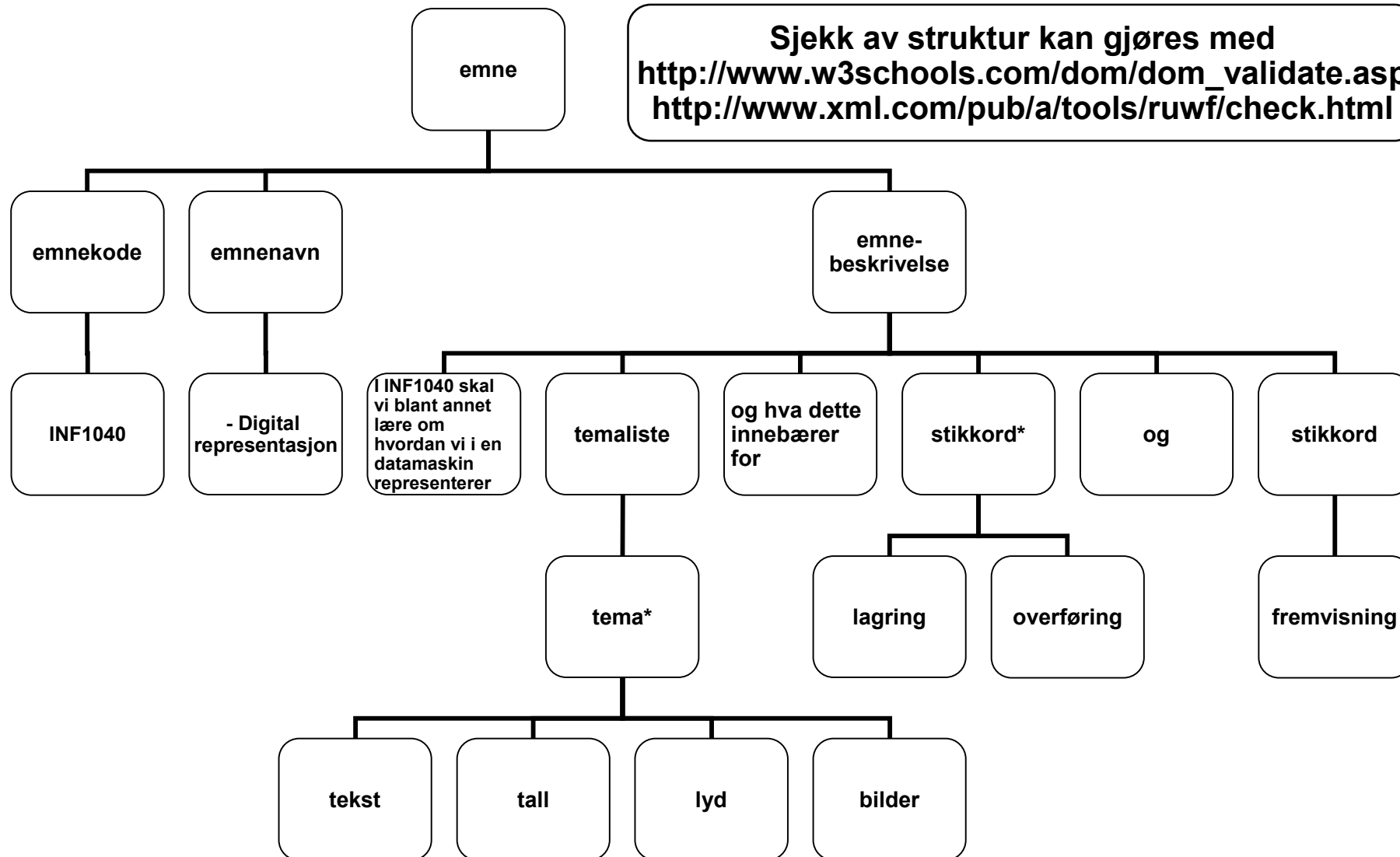
```
<?xml version = "1.0" encoding="ISO-8859-1"?>
<emne id="ifi12345">
<emnekode>INF1040</emnekode>
<emnenavn>Digital representasjon</emnenavn>
<emnebeskrivelse>I INF1040 skal vi blant annet lære om hvordan vi i en
    datamaskin representerer
<temaliste>
    <tema>tekst</tema>
    <tema>tall</tema>
    <tema>lyd</tema>
    <tema>bilder</tema>
</temaliste>
og hva dette innebærer for
    <stikkord>lagring</stikkord>,
    <stikkord>overføring</stikkord> og
    <stikkord>fremvisning</stikkord>.
</emnebeskrivelse>
</emne>
```

*Vi snakker gjerne om et "XML-dokument", selv om det ikke er meningen å skrive det ut på papir.*

*Terminologien understreker imidlertid at XML er et tekstbasert format.*

# Elementene danner en hierarkisk struktur

Sjekk av struktur kan gjøres med  
[http://www.w3schools.com/dom/dom\\_validate.asp](http://www.w3schools.com/dom/dom_validate.asp)  
<http://www.xml.com/pub/a/tools/ruwf/check.html>





# Oppbyggingen av et element

- ❑ Et element må ha både en start- og en slutt-markering  
`<markering>elementinnhold</markering>`
- ❑ Elementer kan ha attributter i startmarkeringen: `attributt="verdi"`
- ❑ Tomme elementer kan ha en kombinert start- og sluttmarkering `<markering/>`
- ❑ Regler for elementnavn
  - 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 (men med \_)
  - 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:

<code>&lt;</code>	<code>&amp;lt;</code>	<code>&amp;#x3C;</code>
<code>&gt;</code>	<code>&amp;gt;</code>	<code>&amp;#x3E;</code>
<code>&amp;</code>	<code>&amp;amp;</code>	<code>&amp;#x26;</code>

*x betyr at kodepunktet er angitt i heksadesimal!*

- ❑ For konvertering fra Unicode-tegn til heksadesimal, bruk for eksempel <http://pioneer.stereo.lu/converter.html>

# To anvendelser av XML

- ❑ **Som et “lingua franca” mellom systemer**
  - Endelig en “standard” !
  - Emneorienterte markeringer må defineres  
(begge systemer må ha samme oppfatning av <eventyr>)
  - Er tekstlig format alltid hensiktsmessig?
  - Stor datamessig “overhead”  
Men: Store muligheter for komprimering
- ❑ **Lagring av XML-dokumenter i en database**
  - “Native” XML-database
  - Trenger vi enda en type database?

# Eksempel på en Dokumenttypedefinisjon - DTD

Dokumenttypedefinisjon på filen 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)>
```

Parced Character  
Data, dvs vanlig tekst

På filen emne.xml

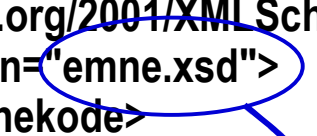
```
<?xml version = "1.0" encoding="ISO-8859-1"?>
<!DOCTYPE emne SYSTEM "emne.dtd">
<emne>
<emnekode>INF1040</emnekode>
...
...

```

Dokumenttypedeclarasjon

# Et XML-skjema

```
<?xml version = "1.0" encoding="ISO-8859-1"?>
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
<emne xsi:schemaLocation="emne.xsd">
<emnekode>INF1040</emnekode>
```

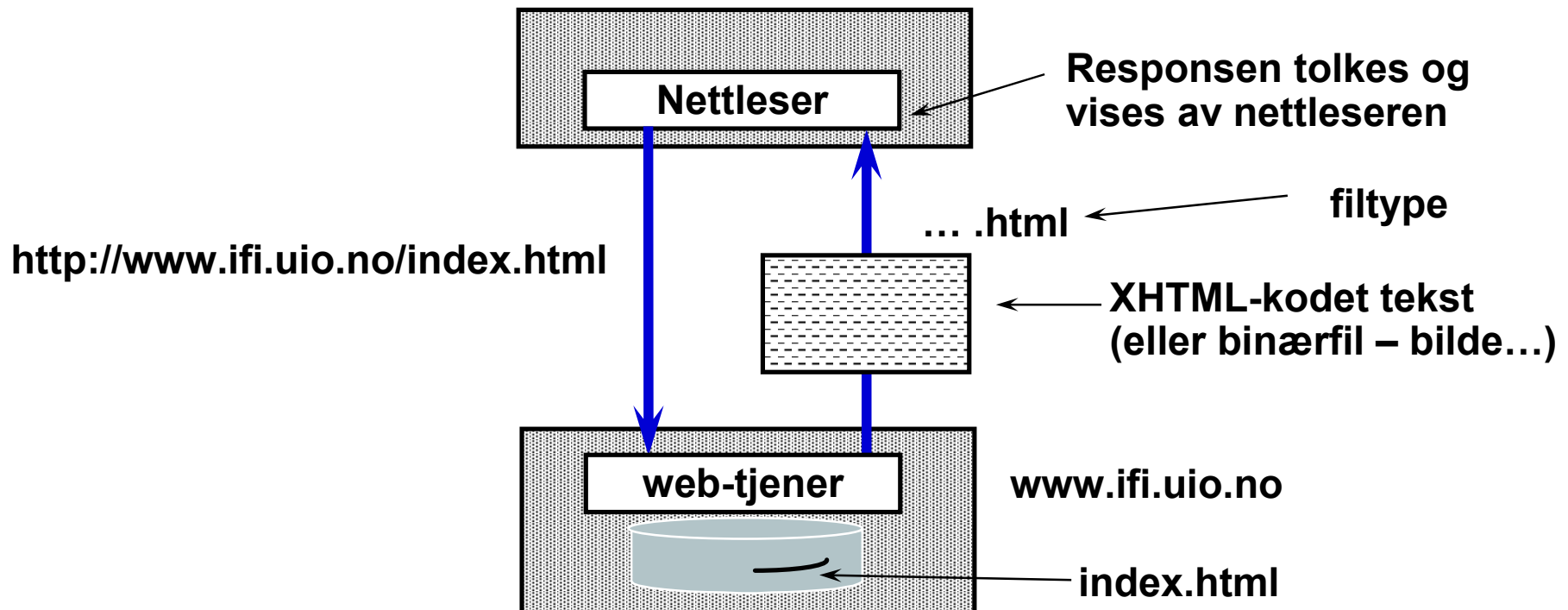


På filen emne.xsd:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<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>
```

# XHTML – Extensible Hypertext Markup Language

- ❑ XML med en spesiell DTD egnet for utforming av nettsider  
se <http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd>
- ❑ Erstatning for HTML – Hypertext Markup Language
- ❑ Viktig bestanddel av Hypertext Transfer Protocol – HTTP



# Tre varianter av XHTML

For XHTML kan vi velge mellom tre ulike DTDer

## Transitional

Litt romslig DTD som bl.a. tillater elementer og attributter som har med visuell presentasjon å gjøre.

Egnet for XHTML som er oppussede versjoner av tradisjonell HTML

## Frameset

DTD som tillater bruk av rammer ("frames").

Omdiskutert teknologi.

## Strict

DTD som utelukker elementer og attributter som har med visuell presentasjon å gjøre – dette skal uttrykkes i stiler og stilark

***Vi skal bruke strict!***

# Hva INF1040 ikke dekker

- Dynamiske nettsider (INF1050)**  
Nettsider som bygges opp "on the fly" på web-tjeneren, ofte på grunnlag av data som hentes fra en database
- Forms (INF1050)**  
Formularer som brukeren av nettstedet kan fylle ut
- Image-maps**  
Klikkbare bilder
- Klient-programmering**  
Bruk av skriptspråk for å skape liv i nettsiden
- Java-Applets og ActiveX-controls**  
Små programmer som lastes ned fra web-tjeneren og som utføres på klienten
- Sikkerhetsaspekter**



# Den minimale XHTML nettside

```
<?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">
```

```
  <head>
```

```
    <title> En minimal nettside </title>
```

```
  </head>
```

```
  <body>
```

```
    <p>Hallo, verden!</p>
```

```
  </body>
```

```
</html>
```

Angivelse av  
DTD

Angivelse av  
navnerom

# Publisering av nettsiden (på Ifi)

- ❑ Legg opp en mappe med navn `www_docs` på brukerområdet ditt (hvis den ikke er der fra før), og gi den tilgangstillatelse 755
  - > `chmod a+rx www_docs`
- ❑ Legg XHTML-teksten du vil publisere i mappen `www_docs`, sørg for at filtypen er `html` eller `htm`, og gi filen tilgangstillatelse 644
  - > `chmod a+r fil.html`
- ❑ Nå er nettsiden tilgjengelig under URLen `http://folk.uio.no/~brukernavn/filnavn.html`

*I det øyeblikk du har satt tilgangstillatelsene, er nettsiden offentlig tilgjengelig ved hjelp av nettlesere over hele Internett.*

*Tenk over hva du publiserer!*