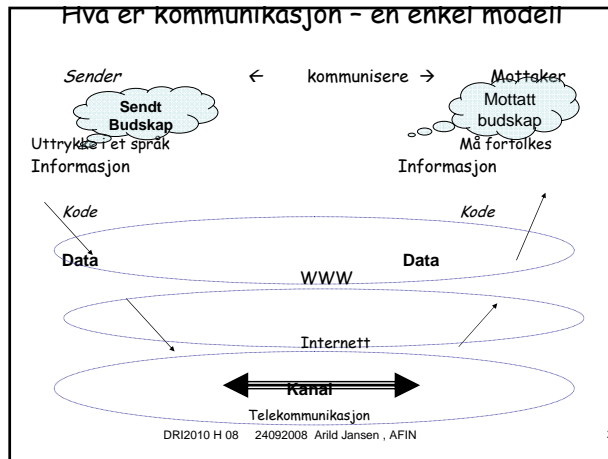


DRI 2010

Internet - fra TCP/IP til semantisk web

- Grunnleggende om Internett
 - Arkitektur -prinsipper
 - TCP/IP
 - Litt om utviklingsfilosofien
- WWW - Hva er hemmeligheten
- Fra databaser til XML og semantisk web

1



Hva er datakommunikasjon ?

- Utrveksling av data i digital form mellom 2 eller flere datamaskiner i henhold til en eller annen standard som partene er enige om

Dette forutsetter (blant annet)

- Felles representasjonsform av data
- Felles protokoller for styre kommunikasjonsprosessene
- At datamaskinene er knyttet sammen i et (eller flere nettverk)
- At to datamaskinprogrammer (applikasjoner) kan 'snakke' sammen (utveksle data)

3

Ulike typer datanettverk

Bedrifts-interne datanettverk (lokalnett- LAN&WLAN)
 Nasjonale og internasjonale nett datanett (f eks. for elektronisk pengeformdling, osv)

Internett

- Ett nett av nettverk som kan kommunisere ved TCP/IP-protokollene, og et felles adresserom definert ved 32-biters IP-adresser (som kan angis ved DNS), administrert av ICANN

Internett benytter seg av flere typer kanaler eller bærere (carriers), f eks.

- Fastnett (..ISDN, ASDL,..)
 - Mobilnettene (GSM, GPRS, UMTS,...)
 - Satellitt kommunikasjon
 - Andre medier (radioforbindelser, fiberkabler, ...)

4

Noen begreper

Protokoll

En *protokoll* er et sett med regler for utveksling av data mellom samvirkende, uavhengige systemer
 - Eks. : IP, TCP, SMTP, HTTP, UMTS... men ikke HTML, XML (som bare beskriver formater)

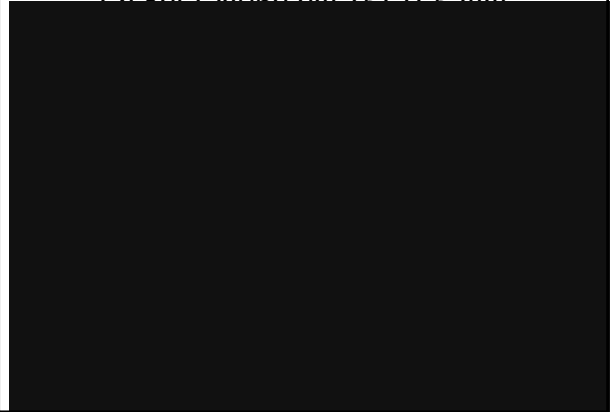
Internet baseres seg på "pakkeveksling", dvs. at all datatrafikk skjer ved oversendelse av datapakker:

IP- Pakke (IP-packet), som er en innkapslet sekvens av data med IP-adresser (sender og mottaker) + kontrollinformasjon

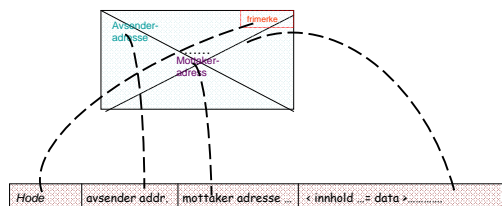
Ruter funksjon (router)

En utstyrsenhet/ program som sørger for å sende pakkene til riktig nett og holde lokaltrafikken lokalt

En kort video om IP/TCP mm



Hva er en IP-pakke



IPv4-protokollen : Struktur og innhold

- Faste og valgfrie felter : Veldefinert betydning av hvert informasjonselement (Utenfor datadelen)
- Min. lengde 20 tegn (hodet) maksimum er 65535 tegn
- Struktur:

Id.<..>	Avsenderadr.	Mottakeradr.	.<annet.>	...data	Kontroll
---------	--------------	--------------	-----------	---------------	----------

Se f eks. http://no.wikipedia.org/wiki/IPv4#IPv4_hode-format

Er syntaks viktig?

- Er formatet viktig?
- Er innholdet viktig
- Kan denne brukes til andre formål

Internett modellen

- TCP/IP "tilnærmingen"
 - Utgangspunkt i behovet for å knytte sammen ulike nett
 - Basert på lagdeling og horisontal integrasjon
 - Balansere standarder og fleksibilitet
 - Åpenhet og dynamikk
 - Minimale standarder, dvs
 - Løse et problem av gangen
 - Foreløpige, gradvis forbedringer
 - Ende-til-ende prinsippet
 - Overlate mest mulig til brukernes ansvar
 - Gjøre nettverket så enkelt som mulig med minimum av kontroll
 - Men gjør også nettet mer sårbart

Noen sentrale prinsipper

- Lagdeling :
 - Programmene er delt i *vertikale* lag som bygger på hverandre, men utveksler data gjennom enkle formater
 - Eks. E-post (SMTP) bygger på TCP, som bygger på IP,...
- Modularisering
 - Funksjoner og tjenester atskilles i programmer som arbeider selvstendig
 - Eks: e-post, filoverføring og WWW er helt adskilte tjenester
- Minimumsløsninger
 - Hver tjeneste er så enkel som mulig

Ulike tilnæringer til utvikling av nettverk

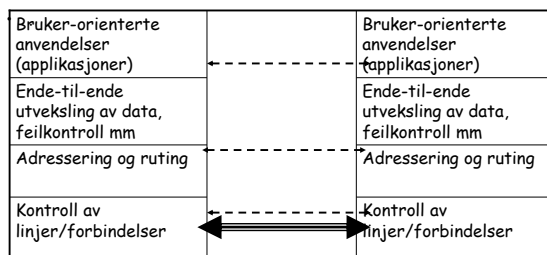
Topp-styrt utvikling

- Målet er den rasjonelle og fullstendige, perfekte løsningen
 - Søker å inkludere alle gode ønsker og behov
- Spesifikasjonsdrevet: fossefallmetoden
 - 'Lukket' verden og sentralisert kontroll
 - Monolittisk organisasjon
- Som regel implementert revolusjonært (alt på en gang)

Nedenfra-opp utvikling

- eksperimentell og iterativ
- Involverer flere bruker og miljøer
- Starter med minimumsløsninger
 - Akseptere kompromisser og søker pragmatiske løsninger
 - "Det gode er ikke det bestes fiende"

Ulike lag i en datakommunikasjonsmodell-forenkelt

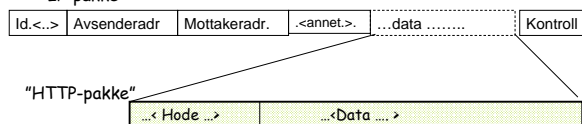


Verdensveven - World Wide Web

- Samme grunnleggende prinsipper som IP/TCP, kontrollen ligger hos brukerne
 - Bygger på IP-protokollen adresse -struktur :
 - En URL = <IP-adresse > + <fil-adresse>
 - eks : http://www.afin.uio.no/om_enheten/omafin.html, som ligger på fila Z: (afin-www)\om_enheten/omafin.html
- "Nettet" (dvs. TCP/IP ++) bryr seg ikke noe om hva som ligger på www-filer
 - Nettleserne er kjernen i verdensveven, og forutsetter dokumentene overføres ved HTTP-protokollen .
 - Krever at dokumenter lagres på formater som kan gjenkjennes av nettleseren (ved typebetegnelsen)

HTTP-protokollen : Struktur og innhold

- IP-pakke :



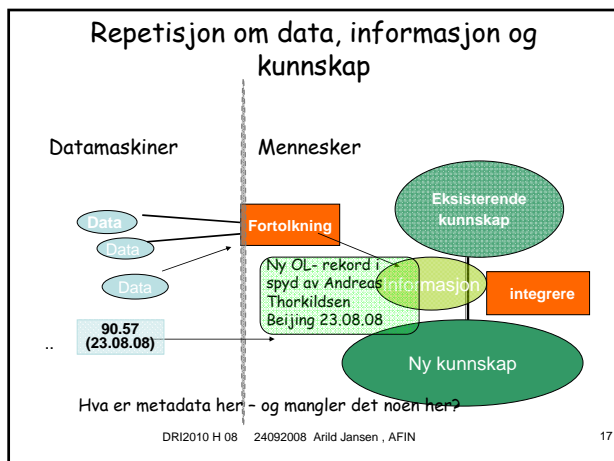
Se f eks. http://en.wikipedia.org/wiki/HyperText_Transfer_Protocol

Noen ulike binære representasjonsformer

- Tekstdokumenter
 - Ren tekst , (.txt)
 - Word-dokument (.doc)
 - RTF :rich text format (.rtf)
 - PDF-format (Portable Document Format) : .pdf
 - HTML-format (.html)
- Bilder :
 - GIF (Graphics Interchange Format)
 - JPEG (Joint Photographic Experts Group)

Semantisk web - Hva er problemet ?

- Semantic web is ...the Web of data (Ivan Herman , W3C)
- Tilbyr er felles rammeverk for å utveksle data på tvers av applikasjoner
- Hjelpemidler for å forstå informasjonsinnholdet i data, og derved muliggjøre bruk av felles data
- En ny "AI-bølge" : SW gjør datamaskinene smartere (gjennom å kunne fortolke og "forstå" hva slag informasjon (kunnskap) data representerer ?



Hva forstår dere med dette dette?

- 0110111010111011
- OLA HANSEN
- HANSEN, KARI
- OLA BJØRN
- 11032003
- 031103
- 11/03/03
- 13/03/45

Hva kreves for å forstå (få mening ut av dette)?

DRI2010 H 08 24092008 Arild Jansen , AFIN 18

Koding, syntaks og innhold

- Vi må bli enige om
 - Hvor mange bits i en byte (lengde : 6, 8, 16, ...)
 - Hvilke byte som representerer hvilke tegn (ASCII, 'UTF-8', ...)
 - Språket jeg vil bruke (norsk, engelsk,
 - Struktur - organisering
 - Syntaks - grammatikken
 - Datamodellen (skjemaet)
 - "Navn" eller "Fornavn" + "etternavn"
 - Semantikken (hva opplysningene betyr)

Hvordan blir vi enige om alt dette ?

- Avtaler (mellom hvem, hvor mange, ...)
- Standarder (hvilke, hvor universelle) Eksempel XML:
 - <?xml version="1.0" encoding='UTF-8'>

DRI2010 H 08 24092008 Arild Jansen , AFIN 19

Syntaks, semantikk og grammatikk og ontologi

- **Syntaks** er læren om hvordan ord settes sammen til større enheter, dvs. setninger og fraser. Ulike språk har *ulike regler* for hvordan dette skjer på både med hensyn til ordstilling og syntaktiske avhengigheter
- **Semantikk** er den grenen av språkvitenskapen som studerer *ordenes betydning*, både enkeltvis og i sammenstilte strukturer (setninger).
- **Grammatikk** er studiet av reglene som gjelder et språk. Reglene som styrer et spesielt språk kalles språkets grammatikk
- **Ontologi** En spesifisering av hvordan man presenterer objekter, begreper og andre "ting" som eksisterer i et interessefelt og relasjonene mellom disse. Både W3C (RDF) og emnekart (XMT) kan beskrive ontologier (se f eks. <http://home.hio.no/~toreh/blog/arkiv/docs/Ontologien%20for%20Kulturnett%203.doc>)

DRI2010 H 08 24092008 Arild Jansen , AFIN 20

Litt om organisering av [tekstlig] informasjon

Et tekstlig dokument kan (blant annet) karakteriseres ved:

- **Innhold:** Hva teksten uttrykker/formidler,
 - Eks: Roman, dikt, fagstoff, lovtekst, offentlig rundskriv, brosjyre,
- **Struktur:** Måten innholdet er organisert,
 - Eks. Bind, kapitler, avsnitt, nummerering, referanser,...
- **Form/utseende (Layout, "design")**
 - Skrifttyper/størrelser, farger/grafikk, sidestørrelse, spalter, bokser,...
- **Koding og språk** (ikke tema denne)

- Disse er ikke uavhengige av hverandre

Hva er viktigst av disse for bøker ??

DRI2010 H 08 24092008 Arild Jansen , AFIN

21

Data, datadefinisjon og metadata

Dataelement: Enhet av data som er udelelig, f eks. f. navn, e.navn, p.nr, telefonnr. ...

- **Datadefinisjon:** *Type og formatbeskrivelse* av et dataelement
- **Metadata:** Data om dataelementer, inkl. datadefinisjon, dataeierskap, tilgangsrettigheter,.....
 - Metadata brukes både i tradisjonelle (relasjons) databaser og andre typer databaser, f eks. XML-baserte databaser.

Metadata omfatter mer enn [rene]datadefinisjoner :

- Bidrar til å opprette logiske sammenhenger, der de ikke finnes fra før
- Bidrar til å gi opplysninger entydige egenskaper
- Bidrar til å knytte informasjon til informasjonens tilhørende sammenheng

DRI2010 H 08 24092008 Arild Jansen , AFIN

22

Eksempel på enkel (tabellbasert) database

Arild Johan Jansen, Hofstadgate , 1384 Asker
Dag Wiese Schartum, Harald Løvenskiolds v , 0760 Oslo

Prnr	Efternavn	Fornavn	Gate/veinavn	Postnr	Poststed
002	Jansen	Arild Johan	Hofstadgata	1384	Asker
01	Schartum	Dag Wiese	H. Løvenskiold vei	0760	Oslo

Poster (record) En 'linje' i tabellen som inneholder verdier i de enkelte feltene

Primær-nøkkel entydig identifikasjon for alle poster

Databaser kan brukes ves f eks. SQL

DRI2010 H 08 24092008 Arild Jansen , AFIN

23

Hvordan representeres dokumenter

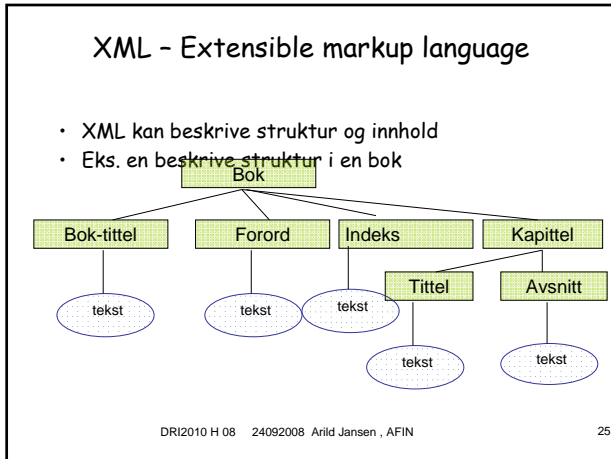
- Dokumenter har blant annet
 - En **identifikasjon** ('navn', kap i en bok, forfatter/dato,
 - **Innholdet**, dvs. teksten vi er interessert i
 - **Utseende** - formatet eller layout, dvs. slik det framstår (presenteres) på skjerm eller papir

Et dokument kan ha ulike typer fysisk *representasjon*, f eks. skriftlig, nedkopier på mikrofilm, digital på disk, disketter, Cd-rom,

Et dokument kan *presenteres* (visualiseres) på ulike måter: på papir, på skjermen, på film osv..

DRI2010 H 08 24092008 Arild Jansen , AFIN

24



Eksempel på XML-kode, inkludert HTML-kode

```
<?XML versjon="1.0" Encoding="ISO-8859-1"?>
<book>
  <description>
    <title> Fra kjernen og ut, fra skallet og inn </title>
    <author>
      <first-name> Gerhard </first name>, <Last-name Skagestein</last-name>
    </author>
  </description>
  <body>
    <Forord > I denne boka vil jeg...</forord>
    <chapter title ="Innledning" >
      <p> I dette kapitlet ser vi på .....
    </chapter >
    Chapter title ="systemutviklingprosessen"
  </body>
</book>
```

DRI2010 H 08 24092008 Arild Jansen , AFIN 26
(fra Skagestein, kap. 17 forenklet. Fargene er for å synliggjøre teksten)

- ### Noen forskjeller mellom HTML og XML
- HTML beskriver bare utseende - ikke hva dataene betyr
 - HTML har en løs syntaks (*feil oppdages ikke lett*)
 - HTML har et begrenset sett av fast definerte *markeringer* og tilhørende attributter (*egenskaper*)
 - XML kan beskrive både struktur og utseende
 - XML har en strengere syntaks
 - Dette gjør at feil kan oppdages før et program brukes
 - XML tillater *egendefinerte* markeringer og attributt-navn
- DRI2010 H 08 24092008 Arild Jansen , AFIN 27

