



Semantisk interoperabilitet

Datalog Foreningen, København

6. september 2005

Forfatter, seniorkonsulent og Ekstern Lektor,
Henrik Hvid Jensen

DEVOTEAM
FISCHER & LORENZ

C o n n e c t i n g B u s i n e s s & T e c h n o l o g y

Overordnet budskab

Internettet bevæger sig fra et browserbaseret medie primært rettet mod manuelt forbrug, til et allestedsnærværende netværk der forbinder samtlige computere og muliggør dynamisk integration af information/viden, funktionalitet, og forretningshændelser.

Dagsorden

- Internettet i dag
- Udfordringer ved Semantisk interoperabilitet
- Hvorfor er Web Services og XML ikke nok
- Opbygningen af Semantic Web
- Opsummering

Internettet er i dag fundamentalt et medie til offentliggørelse.

- Et sted hvor man gemmer og deler billeder og tekst.
- Søgninger forbigår al den information, der ligger i dynamisk genererede sider
- Svært at bruge i stor skala
 - Der er ikke et globalt system til offentliggørelse af data, på en sådan måde, at det let kan blive tilgængeligt af andre.

Hvis man for eksempel beder en bibliotekar om et kort over Sønderjylland, som det så ud på tidspunktet for kampene ved Dybbøl Mølle, vil man sandsynligvis få en bog, der indeholder et kort fra det tidspunkt.

En søgning på en søgemaskine vil inkludere mange resultater med tekst der indeholder "Kort over Sønderjylland" og "Dybbøl Mølle", nogle indeholder måske/måske ikke faktiske kort. Derudover vil formuleringer som ikke matchede tidspunktet specificeret i søgningen blive forbigået.

Tilsvarende vil søgninger efter netværks-sikkerhedsbegivenheder i Københavns-området ikke fange en anti-spam debat på Scandic Hotel i Lyngby, fordi relationen mellem netværkssikkerhed og anti-spam og at Lyngby ligger tæt på København ikke er fundamentale associationer på internettet.

- Aktiviteterne omkring Semantic Web er målrettet mod at fylde hullerne indenfor data-associationer og kollektiv forståelse.

Datas heterogenitet og distribution

- To forskellige perspektiver
 - velformede strukturerede opbygninger i relationelle databaser,
 - som typisk er meget præcise
 - Den relativt dårligt strukturerede verden af dokumenter
 - uden samme præcision
 - e-mails, hjemmesider og dokumenter, lyd og video.
 - Samlet giver de et mere komplet billede af virksomhedens situation.
- Datas distribution
 - Data kan ikke forventes at være placeret på en logisk server (som det ofte er i et datawarehouse),
 - Distribueret på multiple maskiner i forskellige interne og eksterne organisationer.
 - Data skal integreres fra et ukendt antal kilder placeret overalt på internettet.

Datas brug til konkurrencemæssige fordele

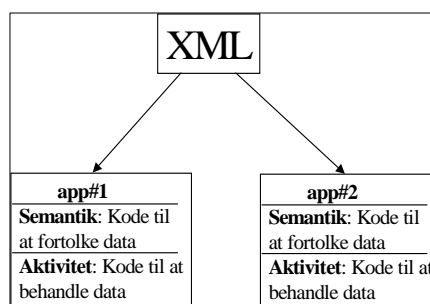
- Disse distribuerede og heterogene data skal sammenstilles i reel tid
 - I en form der er relevant for forretningsbehovet
 - Kombineret med information fra en bred vifte af kilder, der samlet giver den nødvendige synlighed i forretningen.
 - Uafhængigt af datas oprindelse.
 - Uden behov for at opsætte en hel ny infrastruktur
- Stigende mængde af struktureret og ustruktureret data,
 - Muligheden for at beslutningstagerne kan undersøge data bliver vanskeligere,
 - Dataanalyser, der arbejder på tværs af enheder af data, bliver vigtigere.

Dagsorden

- Internettet i dag
- Udfordringer ved Semantisk interoperabilitet
- Hvorfor er Web Services og XML ikke nok
- Opbygningen af Semantic Web
- Opsummering

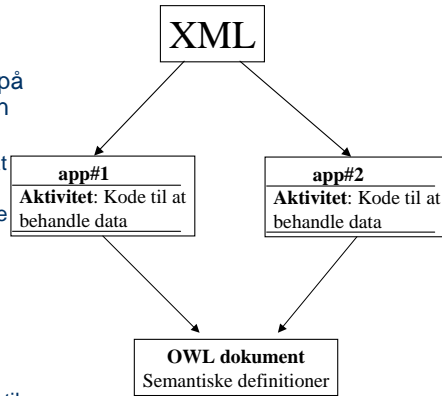
Hvorfor er Web Service standarderne ikke nok?

- Web Services-standarderne fokuserer på applikationsinteroperabilitet
 - Sikrer at information troværdigt kan flyde fra kilde til destination
 - Fokuserer på den fysiske forbindelse og infrastrukturen til at flytte data det rigtige sted hen.
 - XML og XML Schema angiver syntaksen af den udvekslede information
- Datas betydning indlejret i logikken i programmerne
 - Ingen fælles forståelse af informationen.
 - Applikation kan indeholde dets egen fortolkning
 - Dobbeltarbejde
 - Dynamisk tilpasning ikke muligt



Semantikken placeres i et løst koblet lag udenfor applikationen

- Ikke indlejret i logikken i de enkelte programmer.
- Semantic Web-standarderne fokuserer på den logiske informationsinfrastruktur i en virksomhed
 - Meningsfuld udveksling mellem separat udviklede informationer
 - Angiver betydningen af den udvekslede information
- Udtrykkes i en standardiseret ontologi
 - Maskinlæsbar og maskinforståelig begrebsmodel
 - Hvordan skal data fortolkes
 - Præcis beskrivelse af specifikke dataelementer og hvordan de relaterer til andre dataelementer.
- Fælles datasyn



Web Services og Semantic Web har forskelligt formål

- Web Services aktiviteterne svarer til de standardiseringer, der skulle til for at sikre, at alle telefoner i hele verden kunne forbindes.
- Informationsinteroperabilitet fokuserer på, at man også kan forstå den person, der tager røret, uafhængigt af hvilket sprog denne person taler.

Dagsorden

- Internettet i dag
- Udfordringer ved Semantisk interoperabilitet
- Hvorfor er Web Services og XML ikke nok
- Opbygningen af Semantic Web
- Opsummering

Eksempel på sammenhængende digital forvaltning

Peter og Nina er på deres søndagstur i Nordsjælland kommet forbi et hus, der er til salg.

De stiger ud af bilen og Peter henter via sin mobiltelefon et kort frem over grunden.

Han checker de historiske kort, for at sikre, at der ikke har ligget en sø før.

Han trykker på en knap og får tinglysningerne frem. *"Jeg kan se, at der er tinglyst garanteret adgang til stranden via den sti derhenne"* siger han til Nina og får den samtidig fremhævet på kortet.

Han klikker på højen, der er afbildet på den vestlige nabogrund og får afbildet hvordan højen kaster skygger på grunden, på de forskellige tidspunkter af året. *"Det er vist kun i december og januar at den høj vil skygge på terrassen om aftenen, så det ser også fint ud"*.

Nina smiler for sig selv: *"Endelig ser det ud til, at de har fundet et hus, som kan klare alle Peters prøver, nu mangler jeg kun at høre om trafiktætheden i myldretiden"*

Et komplet billede vil også kræve informationsinteroperabilitet

- Offentligretlige servitutter har i almindelighed gyldighed mod enhver uden tinglysning
- Samme forhold gør sig gældende i relation til den offentlige planlægning af rammerne for anvendelse af fast ejendom.
- Lokalplaner er således bindende, når der er foretaget offentlig bekendtgørelse
- En række rådighedsindskrænkninger stiftet på offentligretligt grundlag tinglyses slet ikke.
 - *beskyttede naturtyper*, herunder heder, moser, ferske enge mv. og *beskyttelseslinjer* omkring søer, skove, vandløb. Det gælder endvidere sten- og jorddiger samt fortidsminder efter museumsloven.
- Tinglysningsudvalget anbefaler ligeledes at der følges med i, hvilke data om "særlige byrder" der med tiden forankres i andre tilgængelige og pålidelige offentlige registre, og som derfor kan udskilles fra tingbogen.

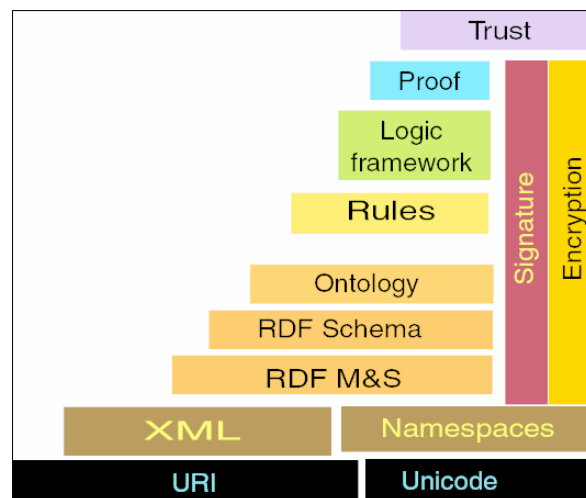
Målet med Semantic Web

- Semantic Web er en infrastruktur, til at lette samarbejdet mellem services og applikationer på internettet.
 - Semantic Web er en samling af standarder
 - Fjerner forståelsesbarriere mellem applikationer
 - Understøtter integrationen af data fra mange forskellige uafhængige kilder
- Værktøjer til at sikre fælles forståelse af information indenfor et domæne
 - F.eks. Information hos offentlige forvaltninger
 - Indenfor et industridomæne
- Løst koblet - Uafhængig af underliggende platform
 - Passer ind i det eksisterende it-økosystem.

Udvalgte principper

- Alt identificerbart er på nettet
- Alle skal kunne sige alt om alle ting
- Forvent ikke global konsistent af data
- Tillad effektiv kombination af uafhængige data
- Gør de simple ting simpelt og de komplekse ting muligt
 - Tilgængeligt for menigmand
- Der er behov for
 - En standardiseret betegnelse for alt
 - En standardiseret måde at beskrive sin viden på
 - En standardiseret måde at definere begreber og viden indenfor et vidensdomæne

Semantic Webs lagkage



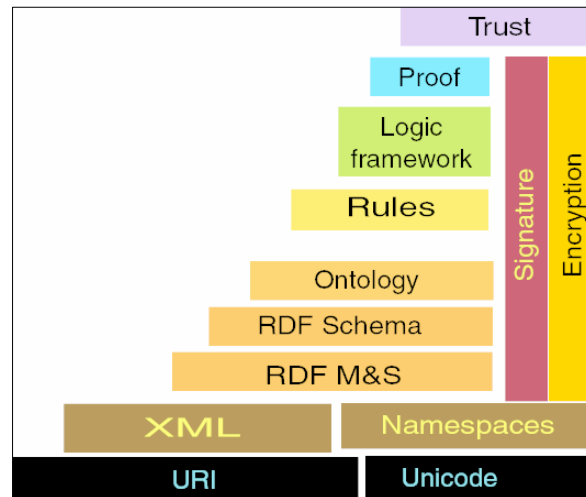
URI – standardiseret betegnelse

- Man skal kunne identificere
 - De ting, man ønsker at beskrive
 - De egenskaber, man ønsker at kunne give disse ting
 - Værdien af disse egenskaber.
- Vores velkendte URL er en delmængde af URI
- Alle kan selv definere deres begreber
 - Veldefinerede organisation omkring URI
 - Netværkstilgængelige ting
 - Et dokument, et billede, en service (KRAK)
 - Ikke Netværkstilgængelige ting
 - Mennesker, firmaer, biler, landområder
 - Abstrakte koncepter
 - Være en dogme film, spille golf, afholde seminar
- Alt identificeret unikt - Ved præcis hvad man snakker om
 - Ting er forskellig: Mørke – Mørke
 - Ting er det samme: Nordsøen, Vesterhavet, North Sea
- Leverer kun en identifikation

URI Eksempler

- Steder
 - kms.dk/øer#Samsø
 - kms.dk/byer#Nakskov
 - kms.dk/station#ØsterportStation
- Personer
 - cpr.dk/personer#123456-7890
- Personlige ting
 - henrikvid.dk/bil#TV12345
- Virksomheder
 - cvr.dk/firma#114141556
- Adresser
 - bbr.dk/adresse#8315351
- Værdier
 - opengeospatial.org/4cmkort
 - Kms.dk/atlasblade
 - Kms.dk/daniaekort
 - Job.dk/kartograf
- Begreber
 - kms.dk/ejendomssammenlægning
 - domstol.dk/tinglysning
 - kms.dk/udstykning
 - kms.dk/arealoverførsel
 - kms.dk/datoer#kortFra
 - Bibliotek.dk/lavetAf
 - Bibliotek.dk/harNavn
 - Bibliotek.dk/harType
 - Bibliotek.dk/harBeskæftigelse
- Industrispecifikke
 - opengeospatial.org/multipolygon
 - opengeospatial.org/point
 - opengeospatial.org/xyzCoordinates
 - opengeospatial.org/hasSpatialLocation
- Virksomhedsspecifikke
 - Virksomhedens informationsafdeling

Semantic Webs lagkage



Devoteam Fischer & Lorenz

19

RDF – Beskrivelse af betydningen

- En **Ressource** er alt som kan have en URI.
- En **egenskab** er en ressource, som har et navn og kan bruges som en egenskab f.eks. forfatter eller titel
- Et **udsagn/sætning** er en triple,
 - Består af en kombination af en ressource, en egenskab og en værdi.
 - Disse dele er kendt som en sætnings subjekt/grundled, prædikat/udsagnsled og objekt/genstandsled

Devoteam Fischer & Lorenz

20

Eksempel på RDF-tripler

Grundled/subjekt	Udsagnsled/prædikat	Genstandsled/objekt
Kortforsyningen	Sælger	Historiske kort
www.kortforsyningen.dk	Bibliotek.dk/sælger	Bibliotek.dk/kort#historiske
Kortet på www.kms.dk/kort1234	er et	Daniae kort
www.kms.dk/kort1234	Bibliotek.dk/harType	Kms.dk/daniaekort
Johannes Janssonius	Er	kartograf
MapHist.com/persons#JohannesJanssonius	Bibliotek.dk/harBeskæftigelse	Job.dk/kartograf
Kortet på www.kms.dk/kort1234	Er lavet af	Johannes Janssonius
www.kms.dk/kort1234	Bibliotek.dk/lavetAf	MapHist.com/persons#JohannesJanssonius
Kortet på www.kms.dk/kort1234	Er fremstillet	1629
www.kms.dk/kort1234	Bibliotek.dk/fremstillet	"1629"

Devoteam Fischer & Lorenz

21

Struktureret info. i DB offentliggøres på internettet via RDF

En af de drivende faktorer omkring Semantic Web har været ønsket om at gøre den store mængde af relationel databaseinformation tilgængelige på internettet, så de kan behandles af maskiner

	Egenskab	
Emne	Værdi	

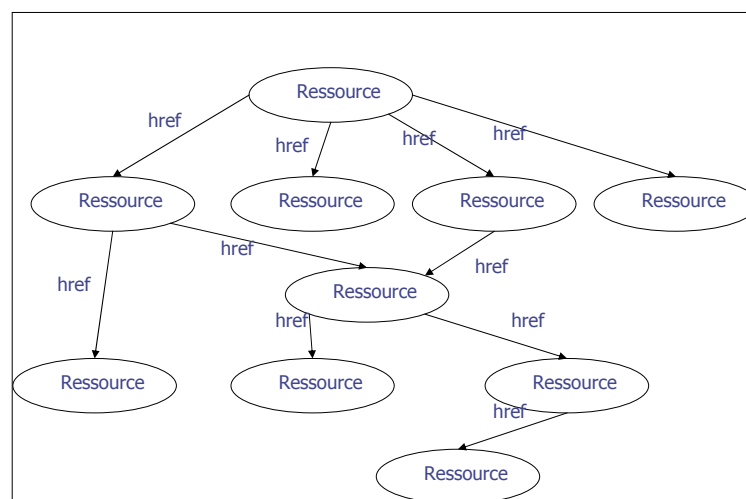
Devoteam Fischer & Lorenz

22

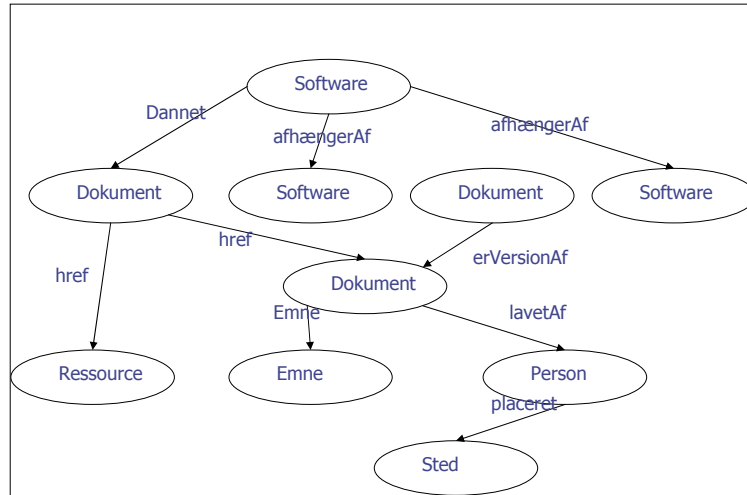
Integration af offentlige databaser til en samlet virtuel database

- På statistik.politi.dk offentliggør politiet statistik over kriminalitet i Danmark,
 - på <http://www.statistikbanken.dk> offentliggøres en mængde statistikker fra Danmarks Statistik
 - f.eks. ledighed, personbeskatning og fertilitet,
 - på DMI.dk ligger statistik over vejret osv.
 - Hos KMS.dk ligger information om hvor stederne er placeret
 - Hver enkelt løsning kun bygget til deres egen verden
 - Webmasteren bestemmer kombinationsmuligheder
- Burde også være tilgængeligt via RDF
 - Mulighed for individuel kombination af data fra uafhængige databaser.
- Når Semantic Web teknologierne har forenet beskrivelsen af data, kan internettet betragtes som del af en stor virtuel database.
 - Fremvisning og kombination af fodboldstatistik fra alle europæiske ligaer

Den nuværende web mangler forståelse for indholdet



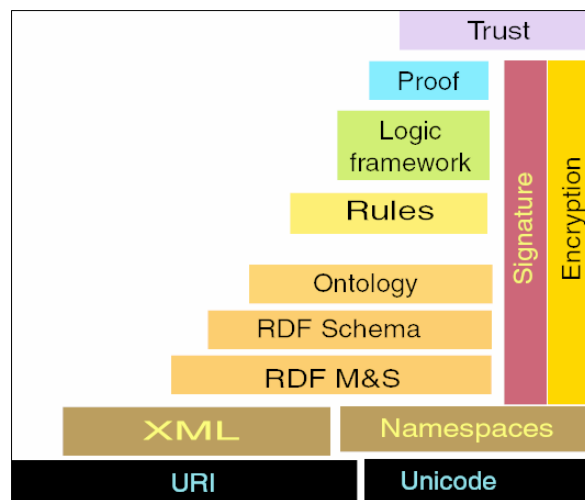
Semantiske relationer fanger betydningen af link og ressourcer



Devoteam Fischer & Lorenz

25

Semantic Webs lagkage



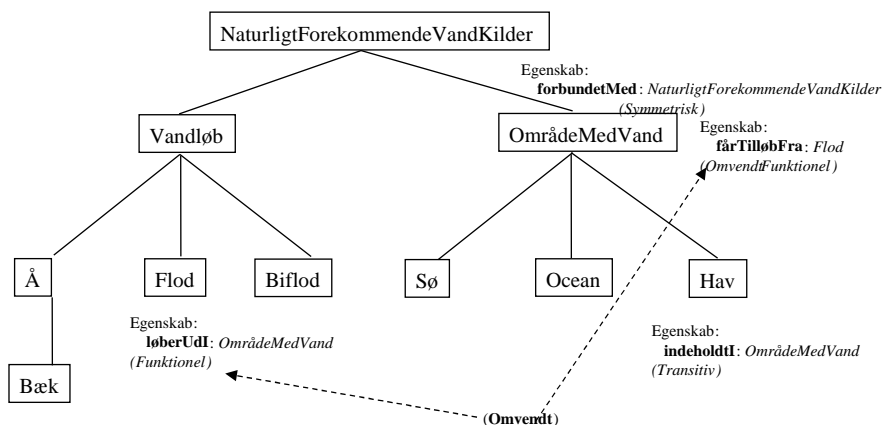
Devoteam Fischer & Lorenz

26

Behovet for ontologier

- Deling af information og viden indenfor et domæne kræver
 - Domæne er et specifikt vidensområde
 - Fælles sæt af termer/begreber
 - Fælles forståelse af de termer
- Ontologien indeholder viden om et domæne
 - Koncepter og deres relationer
- Maskinlæsbart og maskinforståeligt

Sammenhæng udtrykkes i Ontologier



Udlede slutninger

```
<Flod rdf:ID="Gudenå"  
  <løberUdI rdf:resource="http://kms.dk/hav#Kattegat"/>  
  <forbundetMed rdf:resource="http://kms.dk/sø#Julsø"/>  
</Flod>
```

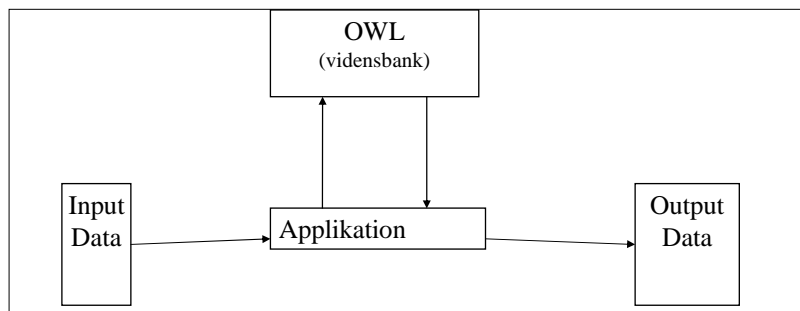
- Kattegat har tilløb fra Gudenåen
 - harTilløbFra er det omvendte at løberUdI
- Julsø er forbundet med Gudenåen
 - Forbundet med er symmetrisk
- Kattegat er et Område med vand
 - Range af løberUdI er OmrådeMedVand
- Julsø er en naturligt forekommende vandkilde
 - Range af forbundetMed er NaturligtForekommendeVandkilde

OWL gør viden genbrugelig

- Når man fortæller en anden person noget, kan denne kombinere den nye viden med gammel og fortælle en noget nyt.
- Når man fortæller en computer noget i XML, kan det måske fortælle en noget nyt, men det er kun på grund af noget software, det bruger og som ikke er del af XML-specifikationen.
 - Kan være implementeret forskelligt på forskellige computere
 - Kan derfor få forskellige svar fra disse systemer.
 - For at udnytte XML til at generere ny data, er det nødvendigt at inkludere viden i noget kode et eller andet sted i stedet for at angive det eksplicit som i OWL.
- Når man fortæller en computer noget nyt i OWL, kan den give en ny information, baseret alene på OWL-standarden fordi
 - en mængde af OWL-sætninger, tillader selv at konkludere andre OWL-sætninger.

OWL bruges som vidensbank

- Ontologier designes, så man kan dele viden indenfor et specifikt domæne.
 - Angiver relationer mellem data
 - Bruges til at dele information indenfor et domæne
 - Computer-brugbare definitioner af koncepter
- Kan betragtes som en vidensbank, som applikationen kan udnytte helt på linie med databaser og lignende
 - Hov! her er noget jeg ikke forstår, lad os se om OWL-Ontologien kan hjælpe?
 - Hvad kan udledes af disse informationer?



Devoteam Fischer & Lorenz

31

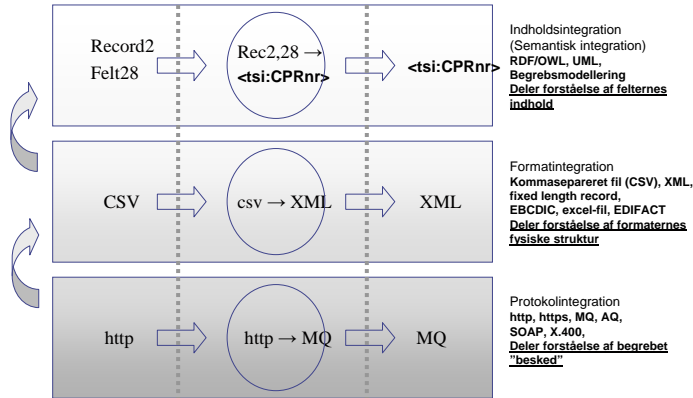
Størst betydning, er indenfor vertikale domæner

- Semantic Web giver alle mulighed for at lave deres egne ontologier
 - Løser lokale integrationsudfordringer
 - Kontoplaner, sundhedsdata osv.
 - Helt i internettets ånd, hvor alle skal kunne sige alt om alle ting.
- For den enkelte virksomhed og dets værdikæde
 - Gå struktureret frem
 - Etablerer overordnede ontologier (øvre-ontologi)
 - Fungerer som fundament for mere specifikke domæneontologier.
- Øvre-ontologi for Danmark
 - En offentlig opgave
- Sundhed.dk baserer sig på den og laver en øvre-ontologi for sundhedsområdet
 - Elektronisk patientjournal, medicinprofilen
- E&S, DLG, Told og Skat osv.
- Novo Nordisk vil kombinere relevante øvre farmaceutiske ontologier med danske og internationale forretningsontologier

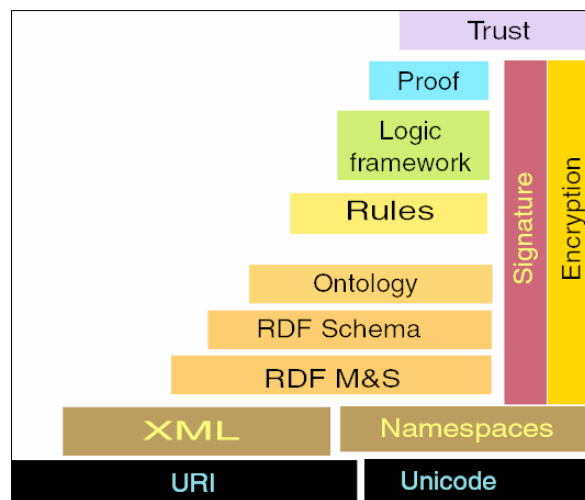
Devoteam Fischer & Lorenz

32

Integrationsniveauer fra Told og Skats udbud (Fra Morten Hass præsentation)



Semantic Webs lagkage



Rules

- Indførelse af regler
- Semantic Web Rule Language (SWRL) er W3Cs bud på en standard
 - Baseret på RuleML standarden
- Udvidelse af OWL
- SWRL beskriver implikationen mellem en forudsætning (body) og en konsekvens (head)
 - "Når en betingelse specificeret i forudsætningen holder, så må betingelserne i konsekvensen også holde.

Regel for at være en onkel

```
<ruleml:imp>
<ruleml:_rlab ruleml:href="#example1"/>
<ruleml:_body>

  <swrlx:individualPropertyAtom swrlx:property="hasParent">
    <ruleml:var>x1</ruleml:var>
    <ruleml:var>x2</ruleml:var>
  </swrlx:individualPropertyAtom>

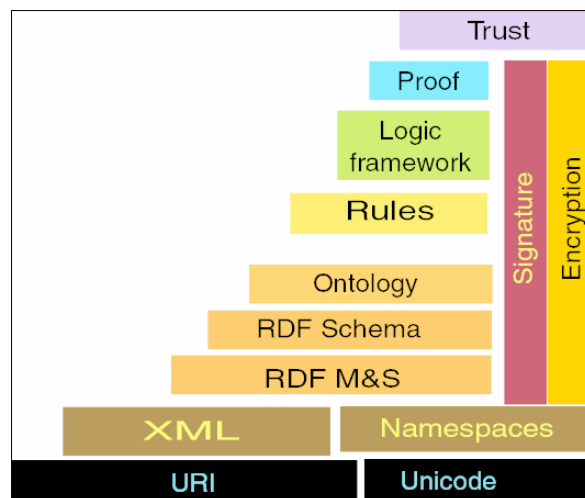
  <swrlx:individualPropertyAtom swrlx:property="hasBrother">
    <ruleml:var>x2</ruleml:var>
    <ruleml:var>x3</ruleml:var>
  </swrlx:individualPropertyAtom>

</ruleml:_body>

<ruleml:_head>
  <swrlx:individualPropertyAtom swrlx:property="hasUncle">
    <ruleml:var>x1</ruleml:var>
    <ruleml:var>x3</ruleml:var>
  </swrlx:individualPropertyAtom>
</ruleml:_head>
</ruleml:imp>
```

■ METALOG Eksempel

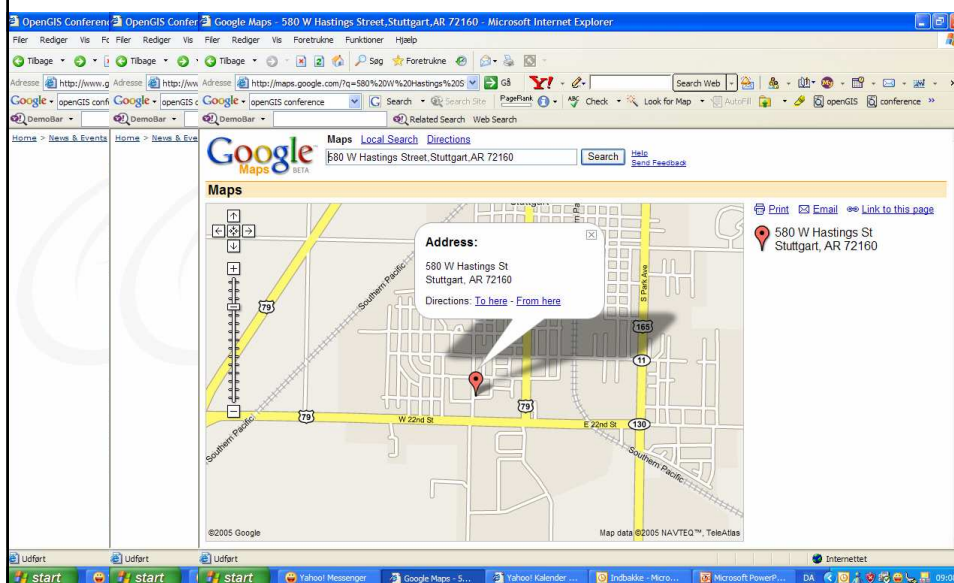
Semantic Webs lagkage



Dagsorden

- Internettet i dag
- Udfordringer ved Semantisk interoperabilitet
- Hvorfor er Web Services og XML ikke nok
- Opbygningen af Semantic Web
- Opsummering

Googles spæde start på Semantic Web



Når data opmærkes og identificeres bedre når borgerne lettere relevant information

The screenshot shows the website nettidende.dk for the organization 'Foreningen Tilgængelighed for Alle'. It features a calendar for August 2005 and a table of tax rates for the municipality of Fredensborg-Humlebæk.

	2003	2004	2005
Kommuneskat	20,10	20,10	20,10
Kirkeskat	0,54	0,54	0,54
Armskommuneskat	11,60	11,60	11,60
Grundslyd til kommunen	9,5 o/oo	9,5 o/oo	9,5 o/oo
Grundslyd til amtet	10,0 o/oo	10,0 o/oo	10,0 o/oo

Devoteam Fischer & Lorenz 41

Det er kun beslutningsstøtte ikke svaret på alt!

The screenshot shows a search results page for 'Tom Cruise' on the TAP website. It includes a search bar, a list of search results, and a table of movie ratings.

Movie Title	Rating
War of the Worlds	71%
H.O. Welter's War of the Worlds: An Historical Perspective of the H. O. Wells Classic Book	
Collateral	
Shattered Glass	
Best: Risky Business	
Worst: L.A. Confidential	

Devoteam Fischer & Lorenz 42

Fokuseret søgning

- På Semantic Web vil siderne ikke kun gemme indholdet, som en samling af ikke relaterede ord i et dokument, men også deres mening og struktur.
- Fortsættelse af eksempel
 - *Data associeret med kort vil være rigere (f.eks. vil en dato være tilføjet samt identificeret som en dato). Derved kan intelligente søgninger foretages ved brug af fleksible repræsentationer af datoer, som vil inkludere forskellige datarepræsentationer (April 1864; 1864 eller 1860'erne) såvel som associationer omkring koncepter (såsom "Slaget ved Dybbøl", "Dannevirke", "1864-krigen" som selv kan have datoer associeret)*
 - *På samme måde med anti-spam seminariet i Lyngby, skal konceptet omkring steder indeholde associationer, som kan placere f.eks. byer indenfor mere fleksible afgrænsede områder (København og omegn), ligesom begrebet anti-spam skal have en betydningsmæssig nærhed til netværks-sikkerhed.*
- Fokuseret søgning på tværs af ESDH-systemer

Semantic Web understøtter det offentlige princip om adgang til data

- *"offentlige myndigheders adgang til data må grundlæggende hvile på et princip om ret til andre myndigheders data."*
- Offentlige organisationer må sikre at data kan deles
- Organisationer har forskellig terminologi og definitioner
- Informationsdeling kræver fortolkning af datas betydning i forskellige situationer
- Semantisk teknologier
 - Leverer et rammeværk for at forbinde distribuerede data
 - Beskriver betydning af information, der udveksles
 - Tillader uafhængighed mellem de kommunikerende parter
 - Dynamisk kommunikation mellem informationsdomæner
 - Kan bibeholde de interne terminologier og definitioner
- Brug af fleksible informationsmodeller ikke ufleksibel programkode

Principper og forudsætninger for visionen I (Fra Helle Pilsgårds Miljøportal præsentation)

Miljødata og datamodeller er standardiserede i en sammenkoblet struktur



- ansvaret for at **definere basisdata og fastlægge standarder for data, datamodeller, udvekslingsformater og tjenester** er centralt placeret - og varetages i **dialog** med aktørerne i sektoren

- ansvaret for **kvalitet, dokumentation og placering** af de enkelte miljødata er klart defineret,



- myndighederne forpligter sig til at gøre deres ikke-fortrolige **miljødata tilgængelige** for andre,



- data lagres, så de kan **genfindes og udtrækkes** på en let måde - og så vidt muligt **omkostningsfrit**.

Principper og forudsætninger for visionen II (Fra Helle Pilsgårds Miljøportal præsentation)



- data kan **integreres i sagsbehandlingssystemerne**, så de brugervenligt kan anvendes i løsningen af miljøopgaverne i kommuner, regioner og i staten,



- de sammenkoblede databaser er tilgængelig for myndighederne via én indgang - **Danmarks Miljøportal**,



- Danmarks Miljøportal giver mulighed for **struktureret og standardiseret formidling** af ajourførte, kvalitetssikrede miljøinformationer og miljødata til offentligheden – borgere, erhvervsliv, organisationer, forskningsverden og presse

Principper og forudsætninger for visionen III (Fra Helle Pilsgårds Miljøportal præsentation)



- Danmarks Miljøportal er **koordineret med andre portaler** som ois.dk og virk.dk og skal kunne tilbyde et relevant udbud af specifikke dataudtræk og tjenester til andre IT-systemer og fagsystemer

- virksomheder og andre skal **kun indberette** de enkelte miljøinformationer til det offentlige **én gang**, hvorfra det stilles til rådighed for de relevante modtagere,



- borgere og erhvervsvirksomheder har mulighed for at se de miljøinformationer, som miljømyndighederne har **registreret om den enkelte borger/erhvervsvirksomhed**

Opsummering

- Ikke kun løs kobling af applikationerne
 - også løs kobling af det semantiske niveau.
 - Ikke indlejret i logikken i de enkelte programmer.
- Semantic Web er ikke en applikation men en infrastruktur
 - Er en vigtig ingrediens i en veldesignet SOA,
 - Fortolker informationens underliggende forretningsforståelse på en fælles måde
 - Vilkårige applikationer deler fælles strukturer sammen med deres egen applikationsspecifikke datastruktur.
- Tidskrævende opgave at opbygge ontologier
- Evolutionær
- Begynd at anvende principperne
 - interne applikationer og med faste partnere



Tak for
opmærksomheden

Henrik.hvid@devoteam.dk

DEVOTEAM
FISCHER & LORENZ

C o n n e c t i n g B u s i n e s s & T e c h n o l o g y