

Tilbakeblikk på to Semantisk Web prosjekter

Gjesteforelesning på Semantisk Web kurs ved
UiO, IFI
2010-05-04

david.norheim@computas.com



Agenda

- Min bakgrunn
- Sublima
- Mediasone

tilbakeblikk...





- M.Sc. datateknikk fra NTNU/NTH, 1996
- Overingeniør i Computas
- Ansvarlig for *Åpne data og Semantikk* i Computas.
- Aktiv i forskningsprosjekter, både i Computas og tidligere for bl.a. EU kommisjonen og UniK
- Internasjonal erfaring fra forskning og fri programvare i EU-kommisjonen, W3C medlem, Innovasjon norge (tidl. Eksportrådet) og egen virksomhet.
- Deltaker i en rekke semantiske prosjekter: Sublima, Mediasone, Seres, Sesam4, AKSIO, DIFI/FAD, Brønnøysundregisterne
- Co-chair "Semantiske Dager 2010" med Arild Waaler som chair.

- Sublima
 - Formålet med Sublima
 - Arkitektoniske valg og endringer
 - SKOS
 - Demo av Smil
 - SPARQL i Sublima

- DETEKTOR for offentlige skoler. Resurser er annotert av bibliotekarer. Inneholder ca 1850 emner og 4600 ressurser. Resulterer i ca 100.000 tripler



- SMIL er et skandinavisk samarbeid for å gi bedre kvalitetsikrede metadata og en samling av sider knyttet til helse, sykdommer og behandlinger. Partnere er bibliotek og sykepleiere fra alle nordiske land. Dagens Smil base har ca 8500 records som utgjør 250.000 tripler.

- Sublima er ment som et verktøy for å samle metadata om spesialiserte lenkesamlinger
- Aka "deli.cio.us on steroids"
- Løsningen er installert for to lenkesamlinger
 - Skandinavisk medisinsk informasjon for lekfolk (SMIL)
 - Detektor

- Består av et sett av metadata (dublin core) til ressurser (lenker)
 - Tittel
 - Beskrivelse
 - Subject
 - Godkjent dato etc.
- Og en beskrivelse av området (emner)

- Det fantes her kategorisering fra før i form av en ikke-formell taksonomi
- Valget falt derfor på SKOS
- Kjent med SKOS?
 - En W3C standard for beskrivelse av en thesauri
 - Alt er instanser
 - Broader, narrower, related to...
 - ≈ emnekart

SPARQL endpoint

- <http://smil.sys.wserver.no:8890/sparql>

Noen spørringer

```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
select ?s ?title
where {
  ?s a skos:Concept;
     skos:prefLabel ?title .
FILTER (lang(?title) = "no")
} LIMIT 10
```

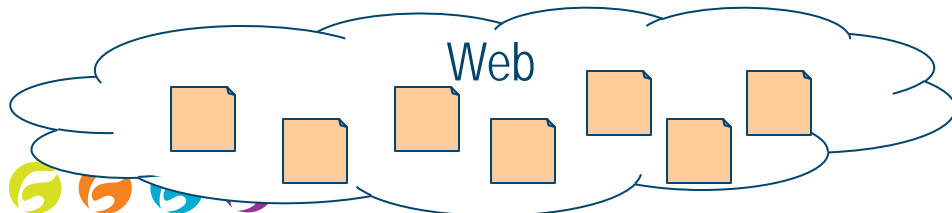
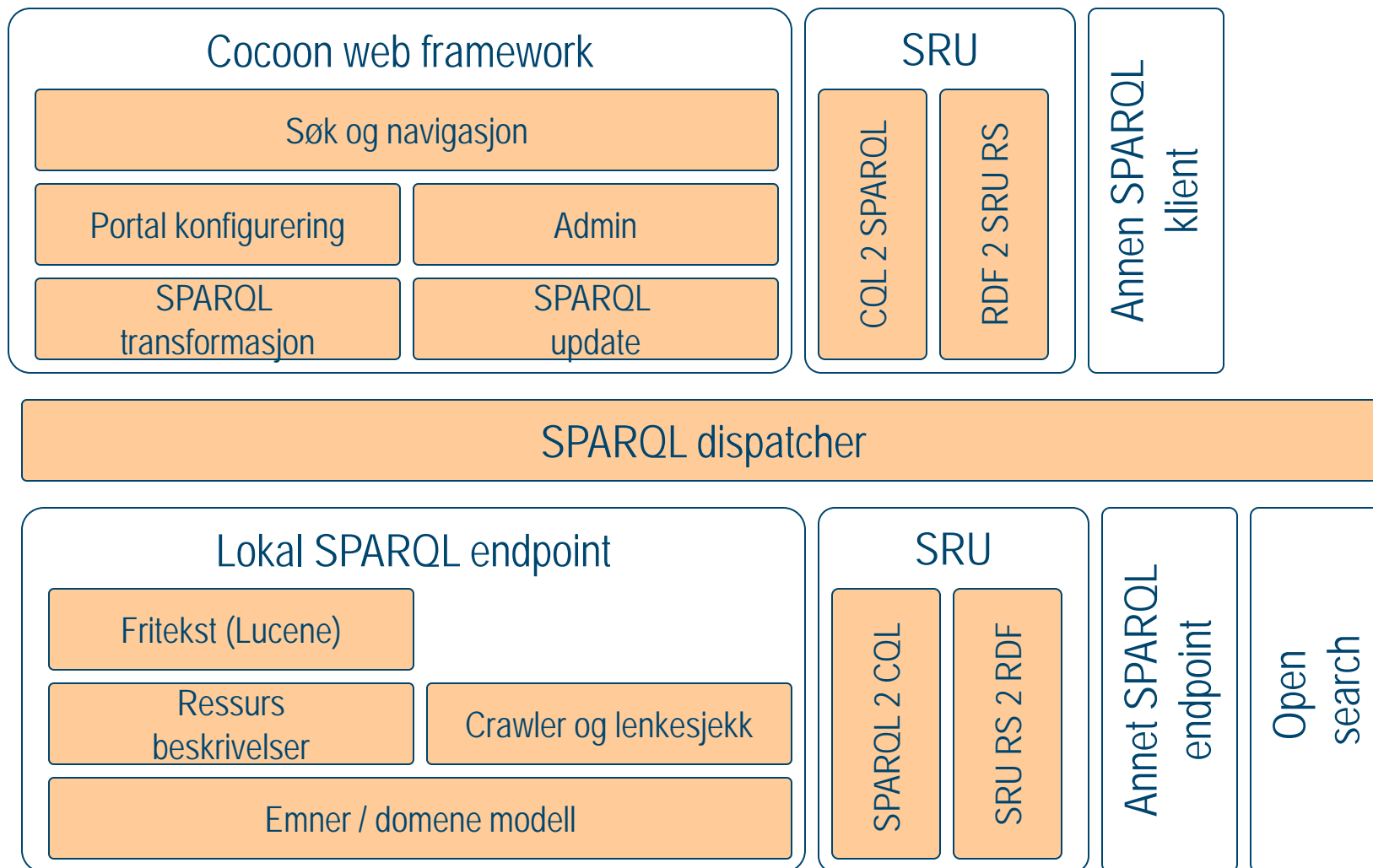
Noen spørringer

```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
select ?top ?title
where {
  ?top a skos:Concept;
      skos:prefLabel ?title .
OPTIONAL{?top skos:broader ?no}
FILTER (lang(?title) = "no")
FILTER (!bound(?no))
} LIMIT 10
```

Noen flere


```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
select ?top ?title ?sub ?titleSub
where {
  ?top a skos:Concept;
      skos:prefLabel ?title .
OPTIONAL{?top skos:broader ?no}
OPTIONAL{?top skos:narrower ?sub .
          ?sub skos:prefLabel ?titleSub . }
FILTER (lang(?title) = "no")
FILTER (lang(?titleSub) = "no")

FILTER (!bound(?no))
} LIMIT 10
```



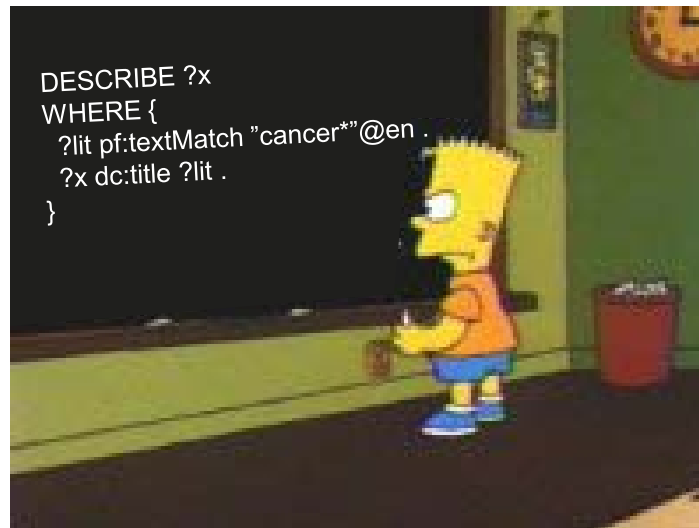
Sublima

- Ontolgoier (eller emneregister) gir en struktur for navigering i resultatene og støtter browsing og klassifisering
- In Sublima bruker vi SKOS til å
 - Navigere i resultatene, vise relasjoner mellom emnene ("fish eye")
 - Utvide søk med synonymer, feilstavelsler etc.
 - Fasettert filtrering basert på emner, men også andre metadata



The screenshot shows a web browser window displaying the Sublima portal. The page title is "Religion | Detektor - Mozilla Firefox". The address bar shows "http://sublima.computas.com". The main content area features a search bar with the text "Søk" and "Avansert søk". Below the search bar, there are filters for "Utgiver" (Nettdoktor (1), Tryggabarn.nu (1), Örebro läns landsting (1)), "Språk" (Dansk (1), Svensk (2)), and "Emne" (Barnesykdommer (3)). The search results show 3 hits, with the first result titled "Barnesykdommer" and a sub-heading "Barnesykdommer - en översikt". The text of the result describes various childhood diseases and their symptoms. A sidebar on the right titled "Se også" lists related terms like "Femte barnesykdom", "Fjerde barnesykdom", "Kikhoste", "Kusma", "Messlinger", "Røde hunder", "Skarlagensfeber", and "Vanrikopper". The footer contains contact information and a note about the project being supported by ABM Utvikling and Computas AS, 2008.

- We needed flexibility for users to add any metadata without touching code
- SPARQL SELECT loses the meaning returning only a binding, hence clients become static. SPARQL DESCRIBE extensively to maximize flexibility



- Formålet med Mediasone
- Arkitektoniske valg
- Datasettene
- OWL ontologier
- Integrasjonsontologi
- Domeneontologi
- Ressonering
- Demo av Mediasone
- Lessons learned
- Serialisering
- Open Source



Mediesone - Deichmanske bibliotek

Naviger gjennom flere multimedia-arkiv mens du sitter i en skulpturell installasjon – og opplev Mediesone. Installasjonene som er plassert i resepsjonsområdet i Deichmanske Bibliotek kalles poder og er laget i innovativt materiale. Programvaren er en løsning basert på semantisk teknologi som styres ved hjelp av berørings skjermene som er installert i podene.





Krav

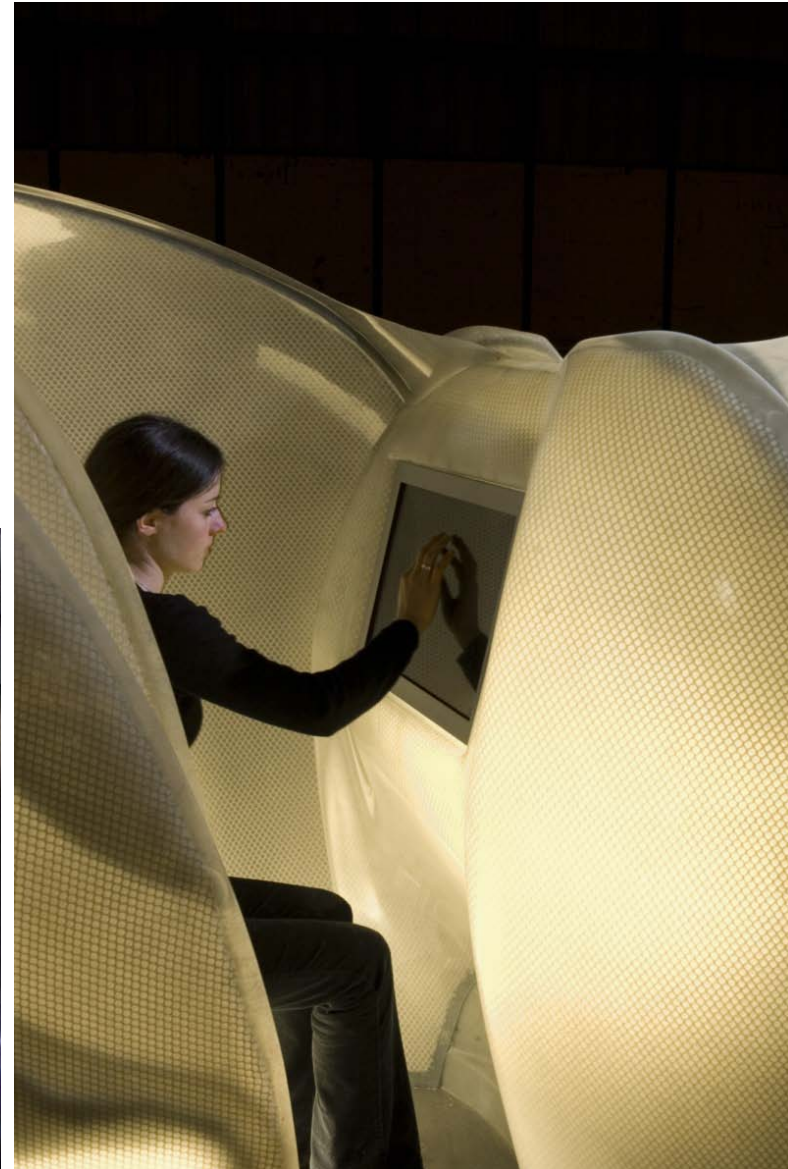
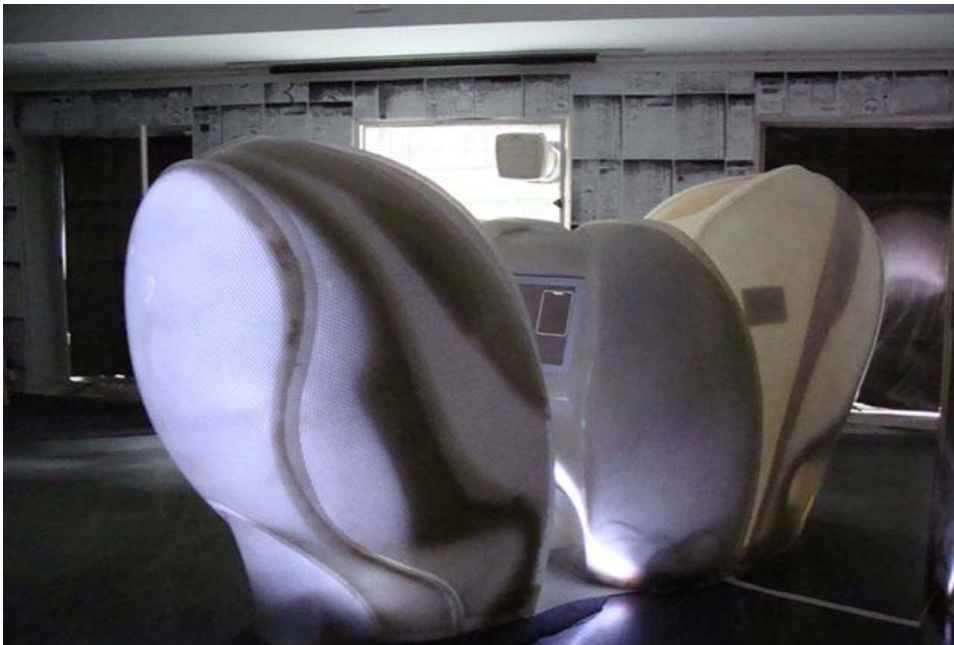
- Mange multimedia samlinger skal kunne sees på som én
- Dekke ulike mediatyper (tekst, film, lyd, bilde)
- Fleksible tematiske endringer

Hvordan?

- En arkitektur som informasjon fra mange ulike samlinger konsolideres og kun spørre én gang (RDF)
- Ontologier representerer ulike glass i samlingen



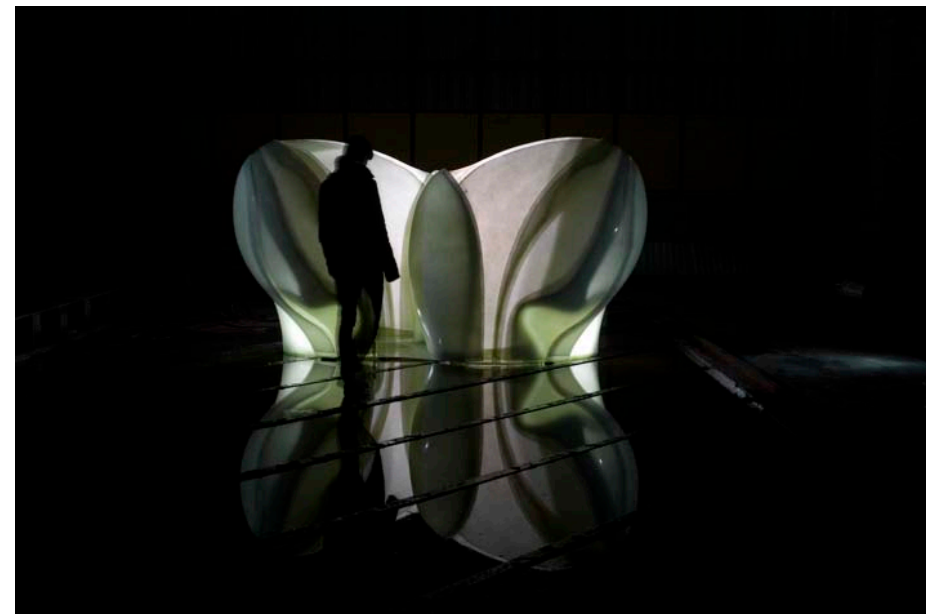
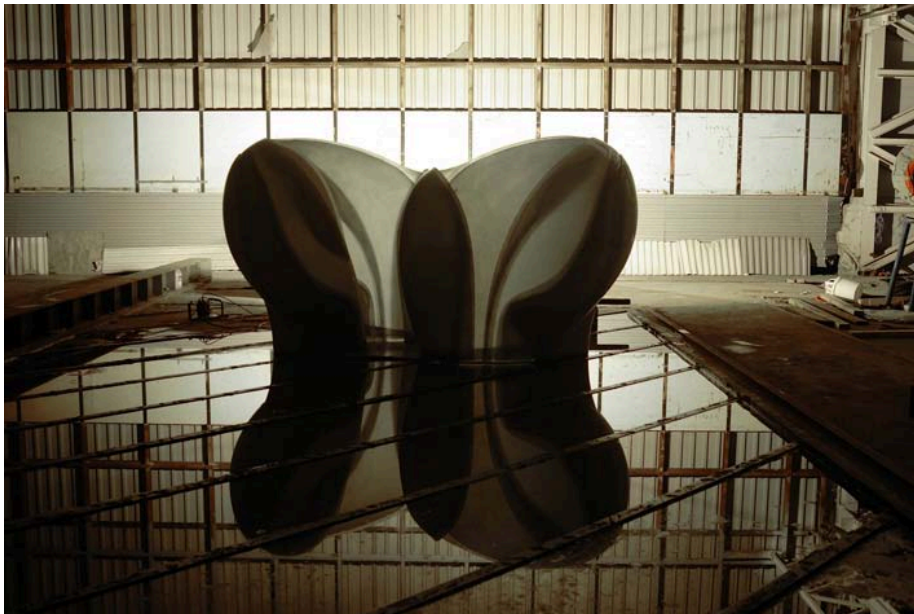
- Ferdig juni 2009
- “Bleeding edge” semantisk teknologi
- En touch screen UI som gir muligheter for
 - Søke, navigere, se og laste ned multimedia innhold
 - Filmer, musikk, tekster og bilder



MedieSone

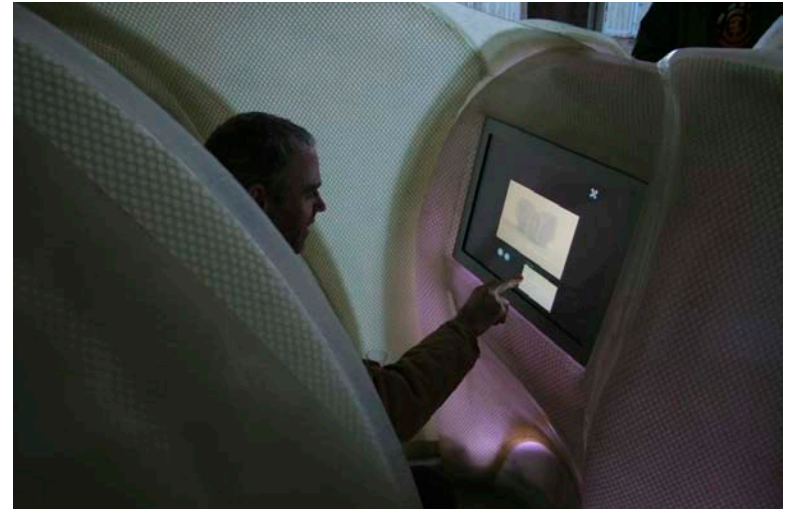
Offers a collection of rare archives to the public.
 Music, movies, audio books, speeches, articles, photos etc.
 Archives includes: NFI (Norsk Filminstitutt) and Phonofile
 More archives will be added continuously.

Futuristic chair made by Ocean Design Research, students from HFG Offenbach and Oslo School of Architecture



MedieSone

- Mac mini
- Custom 22 inch elotouch screen
- Arduino chip with a IR chip for motion detection
- Semantic web backend



QuickTime™ and a
decompressor
are needed to see this picture.

Playing a movie



Search

Komponist: etnisk jazz

Komponist: elektro-jazz

Jazzmusiker

Arrangør: jazz

Dirigent: jazz

Klassisk jazz

Elektro-jazz

Komponist: jazz

Komponist: ambient jazz

Komponist: impro-jazz

Impro-jazz

Jazz/lyrikk

Dirigent: klassisk jazz

Jazz/foik

Etnisk jazz

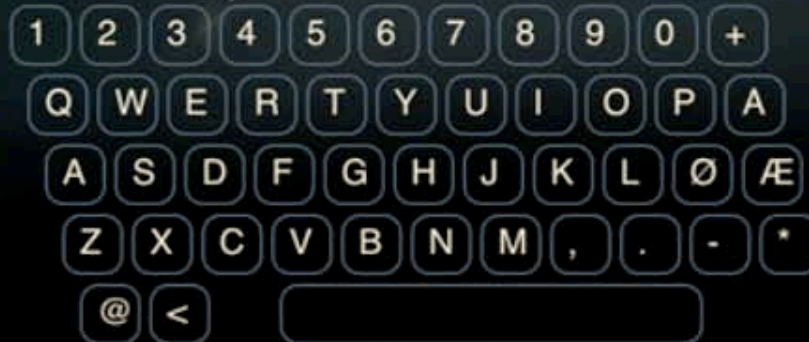
Arrangør: klassisk jazz

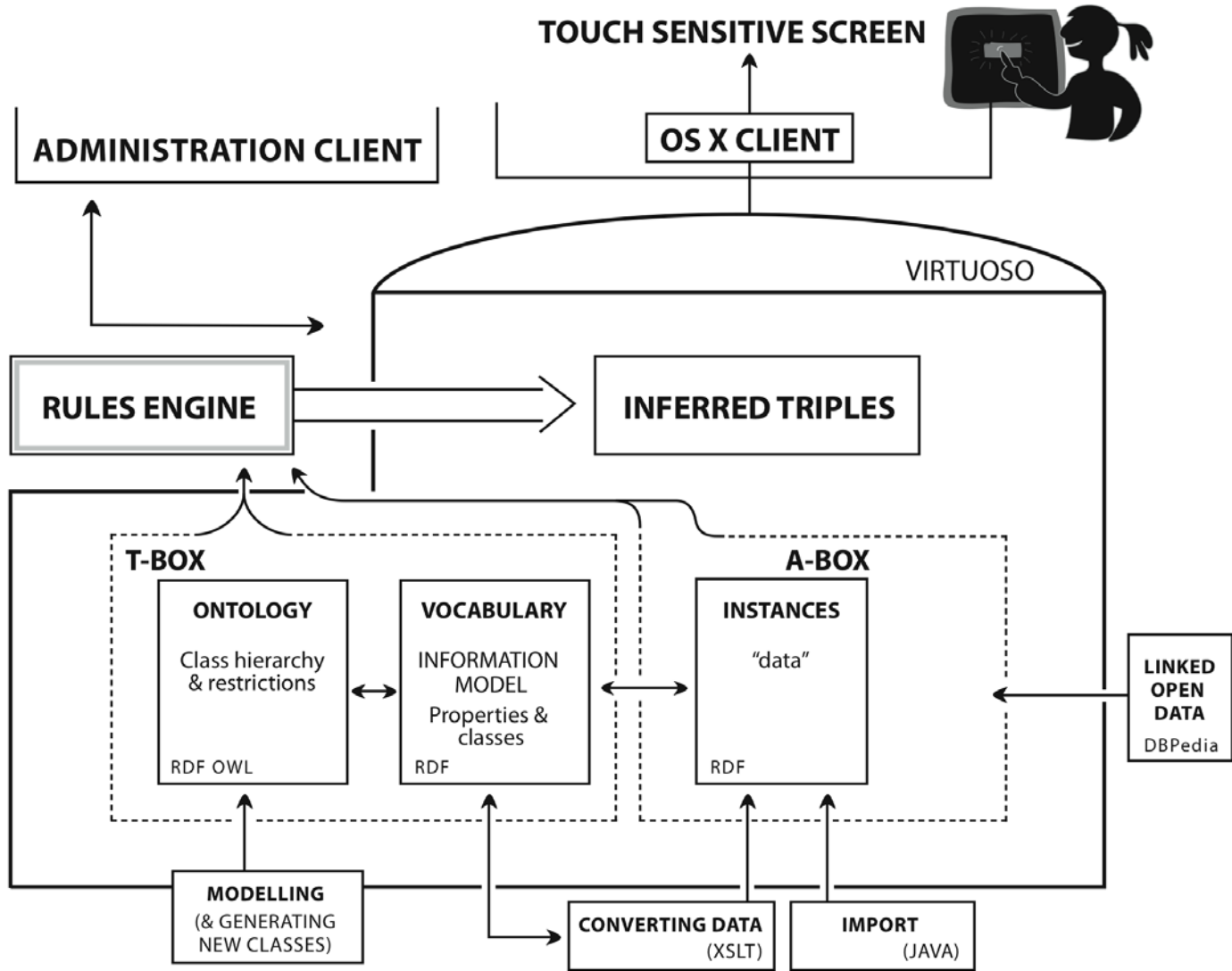
Jazz

Ambient jazz

JAZZ

SØK





- Open source
 - Sublima - prosjektet (tomcat)
 - Jena
 - Pellet
 - Protegé
- Open standards
 - RDF/Turtle
 - SPARQL
 - OWL
- Open licences
 - Creative Commons
- Open vocabularies:
 - Foaf
 - MusicOntology
 - Skos
 - Dct
 - GeoNames
- Open sources:
 - LOD/DBPedia

- Touch screen
- Hele ontologien lastes og visualiseres grafisk og dynamisk
- RDF parser
- Spøringer etter instanser genereres når man trykker på en klasse
- Trykk på en instans medfører spøringer etter klasse tilhørighet og metadata
- Søk mot en indeks gir treff i både klasser og instanser

Konsekvenser av å vise en ontologi som et tre?

- Filer: Uttrekk fra ulike embeddede metadataformater (EXIF, ID3, OGG)
 - Mye teknisk informasjon men lite brukbart (fotograf, lisens...)
- XML: oversent filer fra ulike arkiv (Filminstituttet, Musikk online)
- Metadata konvertert til Turtle (RDF syntaks) i batch
 - Først identifisering av ressursene (tildeler dem URler)
 - Identifisering av artister, album, ... bnode?
 - Legger på metadata i henhold til vokabularer
 - Kodelister lastes fra ekstern hold (land, språk)
 - Serialisering til Turtle

Problemstilling: Ikke unike nøkler for artister, antar vi at artist navnet unik og genererer nøkkel fra det? Konsekvenser?



- Informasjon som manglet blir hentet fra dbPedia, f.eks. om et emneord (klasse) ikke har noen beskrivelse
- Ingen disambiguering

- Integrasjonsontologi
 - Gjenbruk av kjente vokabularer (dublin core, foaf, sioc etc)
- Tematiske ontologier (OWL DL)
 - Skulle kunne bygges av Deichmanske
 - Generell
 - Hamsun
 - Barn
- Resonerings over ontologien
 - Pellet; forward chaining
 - Ønsker å lage et tre av ontologien
 - Ønsker å fjerne klasser uten instanser
 - Ønsker å finne et instans klasses tilhørighet
 - Ting tar tid: ressonere over deler av datasettet

Mediasone-ontologi.ttl



Hamsun-ontologi.ttl



- Mye DESCRIBE - “gi meg alt du vet om...”
- Vi ville ha semantikken med oss til klienten
 - Krever RDF parsing i klienten
 - Desverre er DESCRIBE normativt spesifisert
 - Bytte av database skapte store problemer
 - RDF/XML sucks - men vi hadde låst oss til den i admin grenesnippet (RDF/XML->XSLT->XHTML)
- Issues rundt ulik bruk av tekst matching, ikke standardisert

Hva returnerer en
DESCRIBE <<http://some-url.com>>?



- Redigere ressursene (SUBLIMA)
- SPARQL/Update (SPARQL 1.1)

<http://msonetest.computas.no:8890/sparql>



DESCRIBE <<http://www.wizardis.com.au/2005/12/MAVIS#TitleWorkkey43977>>



[prefix sub: <http://xmlns.computas.com/sublima#>](http://xmlns.computas.com/sublima#)
[prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>](http://www.w3.org/1999/02/22-rdf-syntax-ns#)
[prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>](http://www.w3.org/2000/01/rdf-schema#)
[prefix owl: <http://www.w3.org/2002/07/owl#>](http://www.w3.org/2002/07/owl#)
[prefix xsd: <http://www.w3.org/2001/XMLSchema#>](http://www.w3.org/2001/XMLSchema#)
[prefix yago: <http://dbpedia.org/class/yago/>](http://dbpedia.org/class/yago/)
[prefix dct: <http://purl.org/dc/terms/>](http://purl.org/dc/terms/)
[prefix foaf: <http://xmlns.com/foaf/0.1/>](http://xmlns.com/foaf/0.1/)
[prefix cc: <http://creativecommons.org/ns#>](http://creativecommons.org/ns#)
[prefix mo: <http://purl.org/ontology/mo/>](http://purl.org/ontology/mo/)
[prefix mm: <http://musicbrainz.org/mm/mm-2.1#>](http://musicbrainz.org/mm/mm-2.1#)
[prefix mediasone: <http://www.computas.com/mediasone#>](http://www.computas.com/mediasone#)
[prefix mv: <http://www.computas.com/mediasone-vokabular/>](http://www.computas.com/mediasone-vokabular/)
[prefix mi: <http://www.computas.com/mediasone-instanser/>](http://www.computas.com/mediasone-instanser/)
[DESCRIBE <http://www.phonofile.com/XMLSchema#TrackID_NOX9X3531704>](http://www.phonofile.com/XMLSchema#TrackID_NOX9X3531704)
[FROM <http://msone.computas.no/graphs/instance/test>](http://msone.computas.no/graphs/instance/test)
[FROM <http://msone.computas.no/graphs/instance/nfi>](http://msone.computas.no/graphs/instance/nfi)
[FROM <http://msone.computas.no/graphs/instance/testtest>](http://msone.computas.no/graphs/instance/testtest)
[FROM <http://msone.computas.no/graphs/instance/mo>](http://msone.computas.no/graphs/instance/mo)
[FROM <http://msone.computas.no/graphs/instance/msone-362>](http://msone.computas.no/graphs/instance/msone-362)
[FROM <http://msone.computas.no/graphs/instance/qarchive>](http://msone.computas.no/graphs/instance/qarchive)



Browsing multimedia archives while seated in this depicted pod is the Media Zone experience. The physical pods are made of innovative material and are placed in the reception area of Deichmanske bibliotek / Oslo Public Library. The software runs on a computer connected to a touch screen installation in the pods.

Technical and visual innovation in a public space

BACKGROUND

As a modern library Deichmanske bibliotek / Oslo Public Library is also focusing on digital assets. The Media Zone enables visitors to search and view multimedia material from national archives which are not accessible to the general public through the web. Deichmanske bibliotek / Oslo Public Library has in cooperation with The Oslo School of Architecture and Design and HFG Offenbach in Germany and Ocean Design Research developed media stations to be placed in the libraries in Oslo.

The software project started autumn 2007. The first archives available in the pods are historical movies from the Norwegian Film Institute and from Musikonline.no.

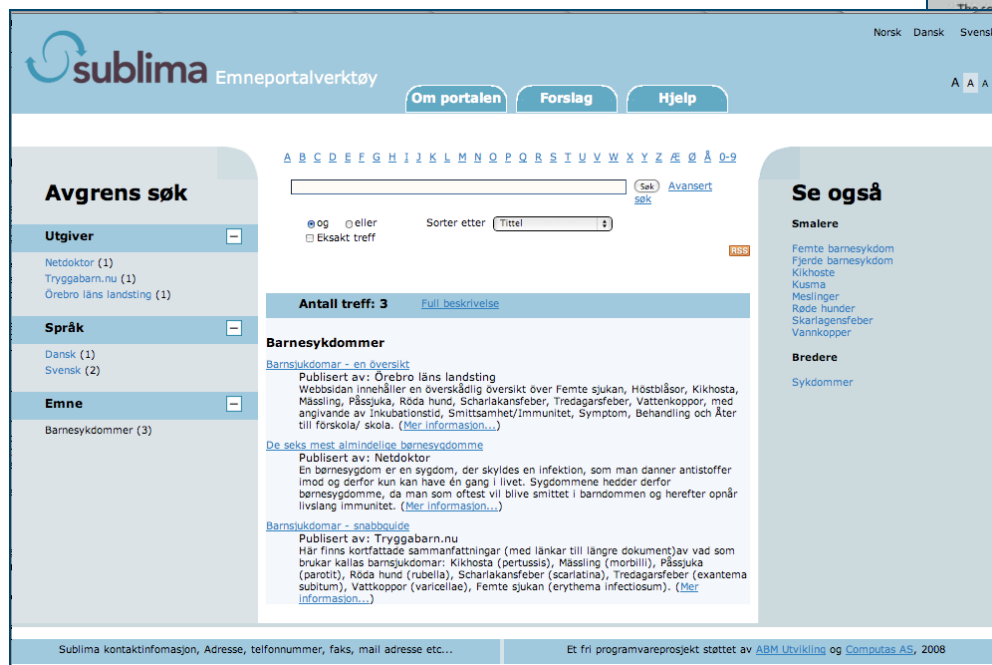
The Media Zone Project

The Media Zone project is a digital library project for making vast amounts of multimedia resources available to the general public. Specially designed pods/kiosks located in the libraries allow library visitors to navigate across archives and watch resources while sitting comfortably inside the pods. Users control the application through a touch screen interface to find material from selected archives. Users can view material and licence permitting, also copy material to a memory stick.

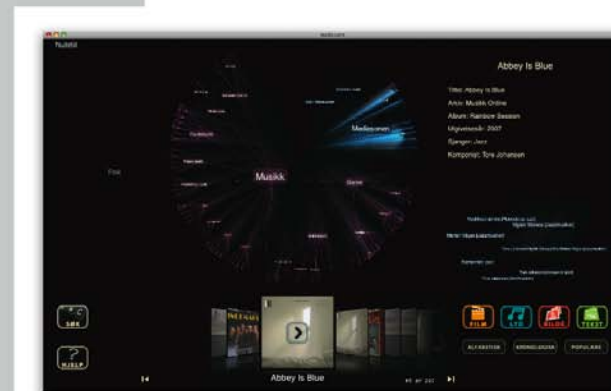
The system is built on a metadata driven state-of-the-art findability solution from Computas based on the World Wide Web Consortium Semantic Web standards. Metadata records from the archives are imported into the system and the multimedia files are streamed to the pods on demand. Dynamically generated topic hierarchies enable easy and powerful navigation and the search results are presented in a coverflow.

Ensuring high quality in metadata, enriching metadata from trusted sources, distribution management and metadata management is of prime concern.

The motivation and focus was to enable thematic search and navigation across media types and enable public access to national archives that are restricted due to copyright issues.



The screenshot shows the Sublima Emneportalverktøy website. At the top, there are navigation tabs for 'Om portalen', 'Forslag', and 'Hjelp'. Below this is a search bar with a dropdown menu for 'Avgrens søk' (Advanced search) and a search button. The search results are displayed in a list format, with the first result being 'Barnesykdommer' (Childhood diseases) from Örebro läns landsting. The page also features a 'Se også' (See also) section with related items like 'Femte barnesykdrom' and 'Fjerde barnesykdrom'. The footer contains contact information and a note that the program is supported by ABM Utvikling and Computas AS, 2008.



Lessons learned

Våre prosjekter har vært
prosjekter
med kunder

som har betalt noe for utviklingen,
har forventinger
til resultatet
og at det virker...

Smertefullt med ny teknologi




- Vi har satt opp en Oracle RDF server som vi vil fylle med offentlige data
- Vil dere være med å fylle den?



Spørsmål?

computas 

Computas 25 år  1985 - 2010
Share your knowledge

Kontakt david.norheim@computas.com

*Computas AS
Lysaker Torg 45, pb 482
1327 Lysaker
Norway*

*Tel +47 67 83 10 00
Fax +47 67 83 10 01
Org.nr: NO 986 352 325 MVA
www.computas.com*

