

SVEINS BLOGG

BLOGG FOR SVEIN ØLNES. MEST OM IT, MEN OGSÅ OM ANDRE INTERESSER SOM JAZZ OG BILAR. DET SOM BLIR PUBLISERT HER STÅR FOR MI EIGA REKNING.

FREDAG 27. MAI 2011

WIMS'11 - dag 3

Ashwin Ram: Open Social Learning Communities

Open Social Learning, what's the problem?

- access (we're not able to meet the demand for education)
- dropouts (the education is not interesting enough)

The Long Tail of Education

- "Short Head" dominated by major players
- Long Tail consists of among others a range of online courses (MIT, iTunes U, Open Yale

Problem with online courses is the one-way delivering of information an no interaction

We have to learn from online games and other popular web activities for young people, and build educational systems like that.

Open Study is an initiative that tries to do this.

- Open Study is a social platform for learners who want to help each other study
- two important aspects: **social** and **gamelike**
- launched in Sept 2010 and has already > 50 000 users

Technologies behind:

- Really Real-time Collaboration
- AI Recommendation Engine
- Social Media Analysis
- Social Capital Engine

Focusing on the features that really helps the user and throw away the rest (typically less features for every new update)

Platform

MongoDB

Scala Framework

Lift web framework (Scala)

- all this is hosted on Amazon's Cloud System

Economy

- online education is a very large market
- the business models are rooted in the old world of education

SISTE KOMMENTARAR

Treme står på lista. Eg såg eit glimt av både Bunk... - Svein Ø

Har du ikkje kasta deg på Treme då, slik som eg? M... - Arild Nybø

> Akkurat kor opne standardar som blir > ny... - Lars Marius Garshol

Det er dessverre ikke så mye som forundrer meg i o... - Roy Allan Hansen

På dette punktet er The Wire svakare enn Sopranos.... - Svein Ø

ARKIV

▼ 2011 (23)

▼ mai (7)

WIMS'11 - dag 3

WIMS'11 - dag 2

WIMS'11 - dag 1

Arrogant Skatteetat II

Skatt Verst?

Statoil og oljesand

Feilmeldingar med eit smil

▶ april (2)

▶ mars (5)

▶ februar (4)

▶ januar (5)

▶ 2010 (72)

▶ 2009 (70)

▶ 2008 (70)

▶ 2007 (102)

▶ 2006 (92)

▶ 2005 (13)

- most of today's online courses are funded by Foundations
- have to look to certificates (be able to grant certificates for courses)
- students are also asking for services that could be provided for payment
- universities as we know today will have a quite different role - they will not disappear but must adapt to an online education reality

Marko Grobelnik: Many Faces of Text Processing

Different approaches to text processing:

- Computational Linguistics (language)
- web 2.0 (community)
- Sem Web (interoperability)
- Text Mining (analytics)
- Machine Learning (statistics)
- Information Retrieval (search)
- Social Networks Analysis (graphs/networks)

Methodological approaches:

Top-down (Sem Web, KRR)

Bottom-up (Machine Learning, Data Mining)

Collaborative approach (web 2.0, Social Computing)

Levels of text representation:

- Lexical (character, words, phrases, part-of-speech tags, taxonomies/thesauri)
- Syntactic (vector-space model, language models, full-parsing, cross-modality)
- Semantic (collaborative tagging/web 2.0, templates/frames, ontologies/first order theories)

The majority of the market actors today works on the syntactic (and lexical) level.

Marko then goes on to demonstrate the different approaches to text processing by showing a range of demos.

- Enrich: Enriching
- Searchpoint: Crawls the next result pages (a few hundred results) and clusters them in different categories on the fly, or classifying them with dmoz.org categories (you should try it! searchpoint.ijs.si)
- News reporting bias
- News visualization (News Explorer - will be available on the web soon)
- Knowledge based summarization
- Question and Answering: Demo showing how triples are extracted from documents given the NLP search and the result shows the triples, the corresponding text and the text repr. of the triples
- The Cyc Ontology ("The Ontology of the World")
 - 15 000 predicates (this is the hard part!)
 - 300 000 concepts
 - 3,2 mill assertions

More information on VideoLectures.net

SKRIVEN AV SVEIN Ø DEN 27.5.11 0 KOMMENTARAR

TORS DAG 26. MAI 2011

WIMS'11 - dag 2

GODE OFF. NETTJENESTER

A List Apart

Anders Brenna

Boxes and Arrows

Gisle Hannemyr

Gode off. nettjenester (Roy Allan Hansen)

Jon Udell

Kuttisme

Kvartventilen (Paal Fosdal)

Larry Lessig

Lars Marius Garshol - Larsblog

Martin Bekkelund

Netlife Research - I_all_enkelhet

NRKbeta

O'Reilly Radar

Per-Erik Skramstad:

Webkommunikasjon

ReadWriteWeb

Vestlandsforskning

OM MEG



SVEIN Ø

Eg arbeider som forskar innan IT ved Vestlandsforskning, eit forskingsinstitutt med

hovudfokus på IT og transport/miljø. I tillegg driv eg ein liten gard med sauer og frukt.

VIS HELE PROFILEN MIN



Sveins blogg av Svein Ølnes er lisensiert under ein Creative Commons Namngiving 3.0 Unported Lisens.

Basert på eit verk på adresse sv.eino.blogspot.com.



Peter Mika, Yahoo! Research

2001: The birth of the Semantic Web

- Scientific American article
- Web Working Symposium at Stanford
- Semantic Web standard
- EU funding (OnToKnowledge) and DAML funding
- the Semantic Web starts a career.... and so do I

2004-2006: Reality sets in

- engineers are not logicians
- humans will have to do most of the job
- no funding

2007: A Second Chance

- data first, schema second... logic third
- Linked Data

Why Semantic Search?

Improvements in IR are harder and harder to come by

- machine learning using hundreds of features
- heavy investment in computational power

Remaining challenges are not computational, but in modeling and user recognition

Poorly solved information needs

- multiple interpretations ('Paris Hilton')
- Long tail queries (George Bush - the beer brewer in Arizona)
- multimedia search
- imprecise or overly precise searches
- searches for descriptions (probably the most important)

Searching for 'roi blanco' [colleague of Peter] gives fairly good organic results, but the advertisements attached are quite bad

Don't solve the sparsity problem where it doesn't exist!

Why Semantic Search? (Part II)

Swoogle - the first semantic web search engine (2007)

- The Semantic Web is now a reality (not in 2007)
- end users use keyword queries, not SPARQL

Novel search tasks

- aggregation of search results (e.g. price comparison across websites)
- analysis and prediction (e.g. world temperature by 2020)
- semantic profiling

- semantic log analysis
- support for complex tasks (search apps)

Why Semantic Search? (Part III)

There is a model

- publishers are (increasingly) interested in making their content searchable, linkable and easier to aggregate and reuse

-

rNews: RDFa vocabulary for news articles

Facebook's Like and the Open Graph Protocol (also mentioned by Jim H.)

Semantic Search

Definition:

- makes use of the structure of the data
- exploits this understanding at some part of the search process

Data on the web

Two solutions:

- extraction using Information Extraction techn.
 - NLP
 - extr of triples
 - filling web forms aut.
 - extr from HTML tables
 - wrapper induction

Have to teach your search engine how to crawl the semantic web

- linked data
- RDFa
- SPARQL endpoints (problem of discovering the endpoints)

Data fusion

- ontology matching
- entity resolution
- blending

Query interpretation

- provide a higher level representation of queries in some conceptual space
- interpretation happens before the query is executed

Query interpretation in Semantic Search

- "Snap to grid"
- larger user involvement
 - guiding the user in constructiong the queries (e.g. Freebase Suggest)

Indexing

- matching and ranking
- indexing for speeding up matching
- type of index depends on the query language to support

Semantic Search evaluation

- critical component in developing IR systems
- keyword search over RDF data
- focus on relevance, not efficiency
- real queries and real data
- TREC style evaluation

Search interface

- snippet generation (enriched search results for pages that contain microformat or RDFa)
- adaptive and interactive pres.
- aggregated search
- query and task templates

Sigma - Semantic Information Mashup (DERI)
Time Explorer

Future work in Semantic Web Search

- semi-automated ways of metadata creation (how to go from 5 % metadata to 95 %?)
- data quality (how to assess the quality of data?)
- reasoning (aut. ontology mapping, instance mapping and blending)
- scalability
- ontology reuse (how to get people to reuse ontologies?)
- display (how to aut. generate effective displays for data we don't understand, or only partially understand?)

In 2011

- the Semantic Web is still evolving (finally, a JSON syntax for RDF!)
- leaner and meaner
 - bottom-up approaches
- we get some credit in other fields
 - RDF data management is now a topic at VLDB, others

The Semantic Web has finally grown up - and so have I!

Sören Auer (Universität Leipzig): Creating Knowledge out of Interlinked Data

Based on the EU FP7 project *LOD2*

Why the Semantic Web won't work

- reasoning does not scale on the web (web scalable DL reasoning is out-of-sight)

"What is the only former Yugoslav republic in the EU?"

- this question can still not be answered by IBM's Watson

But we can do **what works already now:**

A global, distributed platform for data, information and knowledge

integration

Linked Data Lifecycle:

...-> Interlinking/Fusing -> Classification/Enrichment -> Quality anal. -> Evolution/Repair -> Search/Browse/Explore -> Extraction -> Storage/Querying - Manual revision/authoring ->...

Extraction

- recently launched DBpedia Live (the most important part of Wikipedia are the Infoboxes which are highly interlinked in the original source)

DBpedia Live <http://live.dbpedia.org/sparql> (DBpedia Live is constantly updated while the "official" DBpedia SPARQL endpoint is only updated about twice a year)

Mappings Wiki - <http://mappings.dbpedia.org>

DBpedia inline: use the typed links approach to provide more information about internal links

Semantic wikis: Currently does not scale to Wikipedia's needs, but is perfectly ok for smaller wikis

LinkedGeoData - revealing the data behind OpenStreetMaps

OpenStreetMaps: "Wikipedia for GeoData"

- extremely rich source of data

- LinkedGeoData will try to exploit this rich information source with Linked Data technology

Important work is the ongoing work on standardization of how to map RDB to RDF

- W3C RDB2RDF Working Group

From unstructured data:

- deploy existing NLP apps (OpenCalais, Ontos API) - NLP2RDF

From semi-structured sources

- efficient bi-directional synchronization

From structured sources

- declarative syntax and semantics

RDF Data Management

- still 5-50 times slower than RDBMS

- performance increases steadily

- a little performance decrease is acceptable, but not too much

DBpedia Benchmark

- uses DBpedia as data and a selection of 25 frequently executed queries

- ranking between different systems as expected, but the differences were bigger than other benchmarks

The performance gap between RDB and RDF must be reduced
More realistic benchmarks

Authoring

Two kinds of Semantic Wikis:

- semantic (text) wikis (edit the text)

OntoWiki - a semantic data wiki

- serves as LInked data/SPARQL endpoint

- semantic data wikis (edit triples)

RDFaCE - RDFa Content Editor (rdface.aksw.org)

- especially targeted for rNews

LOD Linking

- automatic

- semi-aut. (SILK, LIMES)

- manual

Interlinking challenges: Only 5 % of the information on the Data Web is actually linked

Enrichment

Linked Data is mainly instance data

Quality Analysis

Challenges: Establish measures for assessing the authority, provenance, reliability of Data Web resources

Evolution

Exploration

LOD Lifecycle supported by Debian based LOD2 stack (to be released in September)

- but will also be available through a web interface directly accessible from the web (without downloading or pre-installing)

Use cases:

- especially suited for governmental data (open data)

- Publicdata.eu

- scoreboard.lod2.eu

Michael Hausenblas (DERI, Univ. of Galway): Utilising Linked Open Data in Applications

Six steps for utilising LOD in applications:

1. Data awareness

- opendata.ie

- LOD cloud

2. Modeling

- Neologism
- Data Cube
- prefix.cc (established and controlled by DERI)

3. Publishing

- Google refine plug-in
- RDB2RDF/D2R

4. Discovery

- VoiD/DCat
- Sindice
- CKAN

5. Integration

- LATC
- Sig.ma

6. Use cases

- DERI in-house
- CSO/schools pilot

Workshop part II (Sören Auer)

Linked Data provides a global data-space with a uniform API (due to RDF as the data model)

Access methods:

- dereference URIs via HTTP GET
- SPARQL
- data dumps (RDF/XML)

When to use RDB/SQL and when to use RDF/SPARQL:

RDB/SQL:

- well defined RDB schema that won't change very much
- performance is an important issue

RDF/SPARQL

- getting more information out of your data
- highly dynamic, frequently changes in structure/schema

All in all the two technologies should be seen as complementing each others and not competing.

SKRIVEN AV SVEIN Ø DEN 26.5.11 0 KOMMENTARAR

ONSDAG 25. MAI 2011

WIMS'11 - dag 1



Liv e-blogg frå WIMS'11 -
konferansen

Konferansen mottok 170
paper og programkomiteen valde ut 44 full paper og 11 posters for
presentering.

Jim Hendler: "The Semantic Web 10th Year Update"

Først ei oppmoding til deltakarane om å sjå seg rundt i det fantastiske
landskapet ("certainly one of the most beautiful places I have been to").

Vil fokusera meir på kva som skjer for tida enn på kva som ikkje har
skjedd.

Viser først utkastet (kladd) som vart sendt til Scientific American:

I Semantic Web Vision

II What are the enablers

- Screen Scraping

- Data on Web

Zip code link

Ontology Independence

"Then, a miracle occurs"

III Bringing it all together

Viktige begrep:

Web 3.0

Semantic Web

Linked Data

Hugs at for 10 år sidan var søk ganske eksotisk (Alta Vista var den
leiande, Google knapt synleg).

Facebook var langt frå oppstart og han som etablerte Twitter budde
heime hjå foreldra..

Facebooks Open Graph Protocol (OGP) - annonsert i april 2010

- det er eit veldig enkelt vokabular (mellom 6 og 8 klassar og rundt 20
førehandsdefinerte ty par)

. basert på RDFa

- 10-15 % av alle som brukar 'FB Likes', brukar OGP (men det er 10-15
% av meir enn 3 mill. 'likes' pr. dag!

OGP er utvilsamt den mest brukte ontologien på web-en i dag

Kvifor har FB satsa på RDF(a)?

- for å få meir presis beskrivelse av eksakt **kva** den enkelte brukaren
likte på den spesielle FB-sida.

- det gir FB "labeled links"

- "the network of likes is where their money is made!"

Før 2006 var porno den industrien som genererte mest inntekter på
nettet, etter den tida er det annonsering.

Tredjepartsfirma brukar OGP og lagar sine egne utvidingar og

påbygg (eksempelvis SocialWire). FB og OGP blir eit økosystem for vidare utvikling.

Bitte lite semantikk brukt på eit enormt domene gir etter kvart betydelege resultat.

Best Buy opplevde ein auke på over 30 % i click-through ved å leggja på semantisk informasjon på produkta slik at dei framstod i Google-søk med meir utfyllande informasjon.

Kva har endra seg no samanlikna med rundt 10 år sidan?

- semantisk søk
- annonsering driv web-marknadar
-

Sjølv om semantisk web-stakken ikkje lenger blir oppdatert ("for komplisert") ser vi at fundamentet (dei nedste klossane) no begynner å komma på plass.

SPARQL har vore viktig både for standardisering av søk mot triplar, men også for å laga hybrid-løysingar som fungerer mot standard relasjons-databasar.

Det er ein bitte liten del av OWL i fundamentet ("skal seia noko om OWL seinare, men må vera litt forsiktig").

Web-en handlar om lenker, ikkje sider (links, not pages).

Typed links er ikkje noko nytt, det kom veldig fort opp etter at web-en var oppfunnen.

Kva skjedde med ontologiar?

- frå 2001 til 2006 vart det investert veldig mykje i ontologi-utvikling
- det er dyrt og ROI må vera høg for å forsvare investeringa og vedlikehaldet
- det er grunnen til at "expert systems revolution" ikkje skjedde
- analogien er pre-web hypertekst-verd (som på mange måtar var langt meir avansert enn web-ens hyperlenker)
- web-ens løysing er typisk 80/20: "Heavy use of a small amount of RDFS and a tiny bit of OWL suits the web much better" -> "A Little Semantics Goes a Long Way"

Hendler meiner dei var inne på det rette for 10 år sidan, men undervurderte kor lite (kor enkelt) det måtte vera for å ta av.

SKRIVEN AV SVEIN ØDEN 25.5.11 0 KOMMENTARAR

ONSDAG 18. MAI 2011

Arrogant Skatteetat II



Skatteetaten gir seg ikkje. Nytt brev i dag, denne gangen i ei anna sak om eit manglande kontonummer. Saka gjeld i korte trekk skifte av kontonummer i eit selskap eg er kasserar for. Endring av

kontonummer skal meldast Skatteetaten (det var nytt for meg), og eg

fekk brev frå etaten 29.04 om at nytt kontonummer måtte sendast dei.

Eg sende det nye kontonummeret pr. brev like etter, men det har tydelegvis ikkje Skatteetaten oppdaga. Det er ein stor etat må vita, og omorganisert seg har dei også for å bli meir effektive (til kva??), så det er ikkje rart dei ikkje klarer å halda oversikt over kommunikasjonen med brukarane.

I dag fekk eg altså nytt brev om manglande kontonummer, med mellom anna følgjande ordlyd:

... Vi manglar opplysningar om verksemda sitt bankkontonummer og ber difor om, for 3. gong at de sender inn denne informasjonen...

For det første er det ikkje tredje gongen eg blir beden om å senda inn kontonummeret, men andre gongen. For det andre er nytt kontonummer alt sendt inn, tydelegvis utan at dei har registrert det. Det er vel ulike avdelingar som handterer dette..

Som i forrige sak saknar eg ei meir audmjuk haldning og ei opning for at også etaten kan ha gjort feil. Det er vanleg elles i næringslivet og samhandling med offentleg sektor at ein opnar for at meldingar/innbetalingar kan ha funne stad etter at første påminning er gitt.

SKRIVEN AV SVEIN Ø DEN 18.5.11 0 KOMMENTARAR

TIRSDAG 17. MAI 2011

Skatt Verst?



Eg har fått brev frå Skatt Vest om omsetningsoppgåva for 2010 og utrekna av gift. Omsetningsoppgåva er ein oversikt over meirverdiavgift inn og meirverdiavgift ut, og oppsummert om ein har til gode eller er skuldig av gift. For jordbruket er det ei årleg rapportering og fristen er 10. april i påfølgjande år.

Eg leverte omsetningsoppgåva for 2010 på papirskjema for di Buy pass-kortet mitt var utgått då eg skulle rapportera, og skjemaet vart levert innan fristen og på originalskjema som kravet tilseier. Nokre veker etter fekk eg svar frå Skatt Vest om at oppgåva ikkje kunne lesast maskinelt og at det kunne vera fleire grunnar til det. I mellomtida hadde eg fått nytt Buy pass-kort og sende omsetningsoppgåva inn på nytt via Altinn.

Så får eg brevet frå Skatt Vest med utrekna av gift for 2010, og der skriv dei:

Omsetningsoppgaven deres ble mottatt etter fristen for innlevering. Fristen var i dette tilfellet 11.04.2011. Når omsetningsoppgaven ikke er kommet inn i rett tid eller ikke er

utfylt på riktig måte, kan avgiften økes uten varsel med inntil tre prosent (min. 250 kroner, maks. 5000 kroner) etter merverdiavgiftsloven § 21-2. Denne gangen har vi ikke økt avgiften, men det kan bli aktuelt dersom dette gjentar seg.

Hallo? Eg lev erte oppgåva innan fristen så ikkje kom her med den tonen takk!

Det slår meg at Skatteetaten kan gjera ein god del med standardbrev a sine, ikkje minst jekka ned tonen og og gjera den litt meir audmjuk for at feil også kan førekomma på deira side. Men nei då, Staten gjer aldri feil må vita.

I tillegg er brevet skrive på bokmål, og eg har vel krav på å få svar på mi eiga målform sidan det er ein offentleg etat som svarer?

SKRIVEN AV SVEIN Ø DEN 17.5.11 0 KOMMENTARAR

FREDAG 13. MAI 2011

Statoil og oljesand



Det er generalforsamling i Statoil 19. mai og som dobbelteigar i selskapet (gjennom Staten sine 67 % og gjennom egne aksjar) støttar eg forslaget om at selskapet snarast

avviklar oljesandprosjektet sitt i Canada. Det er svært dårleg reklame for Norge. Sidan eg ikkje møter på generalforsamlinga sjølv, har eg gitt Greenpeace fullmakt til å stemma for meg. Greenpeace har ein kampanje der dei samlar fullmakter for å stemma for forslaget om å trekkja seg ut av oljesand-prosjektet.

I Stavanger vil også "**Besteforeldre mot oljesand**" vera på plass for å demonstrera. For ein månads tid sidan rykte dei inn ein heilsides annonse i Edmonton Journal, den største avisa i delstaten Alberta der oljesand-utvinninga føregår. I annonsen beklaga dei på vegne av bekymra besteforeldre i Norge, at den norske staten er involvert i det miljø- og klimamessig svært tvilsame prosjektet.

Dette er litt utanfor tema eg vanlegvis skriv om, men eg kommenterte også samanslåinga av Statoil og Hydro i 2006, ei samanslåing eg meinte var uheldig. Det store problemet med Statoil er at dei er involverte i **tvilsame prosjekt** (som oljesand i Canada, gass-skifer i USA) og i land med **tvilsame regime** (Angola, Nigeria, Irak, Aserbadsjan ..). Eg forstår ikkje at dette skal vera ei oppgåve for den norske staten. Viss Statoil var eit privat selskap ville det vera heilt ok, då vart det eit forhold mellom selskapet og dei lokale styresmaktene. Saka stiller seg annleis når det er nasjonale interesser bak.

Generalforsamlinga kan også følgjast via webcast.

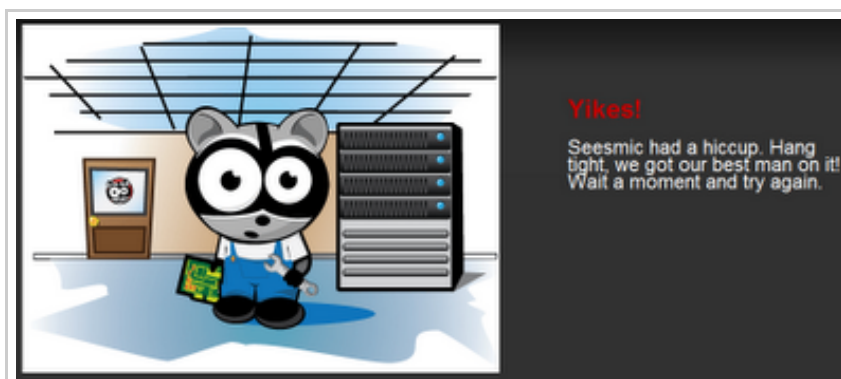
SKRIVEN AV SVEIN Ø DEN 13.5.11 0 KOMMENTARAR

TIRSDAG 3. MAI 2011

Feilmeldingar med eit smil

Buypass har som tidlegare nemnt elendige feilmeldingar, og er ikkje åleine om det. Men om det er vanskeleg å gi presise feilmeldingar, er humor eit bra alternativ.

Framfor alt har web 2.0-bølga gitt oss meir avslappa haldningar til feil gjennom tidlege lanseringar (beta-lanseringar har vore ein farsott som eg håpar snart er på retur) og eit anna syn på feil og feilmeldingar. Den avslappa stilen er ofte kommunisert med humor, og det fungerer. Det er ganske avvæpnande når avsendaren viser både sjølvironi og humor. Her er først eit eksempel frå Twitter-klienten *Seismic*:



Og Firefox har skjont at det hjelper å leggja seg flate:



SKRIVEN AV SVEIN Ø DEN 3.5.11 0 KOMMENTARAR

[Start](#)

[Eldre innlegg](#)

Abonner på: [Innlegg \(Atom\)](#)