

2006 Semantic Technology Conference - några noteringar¹

(2006-03-13)

Stig Berild

(Santa Anna IT Research Institute AB)

¹ Rapport framtagen under e-Society II-projektet (www.skriver.nu/esociety) vid Santa Anna IT Research Institute (www.santaanna.se) samt med stöd från Stiftelsen SISU.

Årets konferens var den andra i ordningen under rubriken "Semantic Technology Conference". Den genomfördes även denna gång i Silicon Valley. (Reserapport från förra årets konferens finns att hämta på e-Societys webbplats www.skriver.nu/esociety.)

Intrycken från konferensen ges för enkelhets skull i punktform och utan någon som helst inbördes ordning:

- Mycket välskött konferens. Dokumentationen på CD. De allra flesta presentationerna fanns med till skillnad från vad som tyvärr ofta brukar vara fallet vid större breda konferenser som inte tillämpar reviewing (som mer akademiska konferenser gör). Dessutom fylldes resterande dokumentation på efterhand på en webbplats för betalda deltagare. Varje konferensdag skickade dessutom arrangörerna ut e-post med uppdateringar, kommentarer, tillkommande information, med mera. Snyggt.
- Titeln på konferensen var representativ för innehållet. Inte så mycket diskussion om "the Semantic Web" (SW) utan snarare breda semantiska frågeställningar. Kanske inte heller så mycket teknologi även om en del produkter presenterades i en mindre utställningshall.
- I år deltog ca 600 personer på konferensen vilket är drygt dubbelt så många som förra året (ungefär samma plats och tid). Ett tecken på ökat intresse och förmodligen ökad förståelse för semantikens betydelse i ett affärsnyttoperspektiv. Första fasen med XML-taggar har kompletterats med fas två där man inser att taggarna inte räcker till, att det behövs något mer.
- Intresset från svensk sida är fortfarande närmast obefintligt. Författaren var enligt närvarolistan ensam svensk representant. Norge däremot kunde noteras för ett antal deltagare.
- Finland och Nokia exponerades genom Ora Lassila som keynote speaker (en av de tre författarna till "The Semantic Web", den numer välkända artikeln från 2001 i Scientific American). Smart av Nokia. Visar på vilja och förmåga att befinna sig i den teknologiska utvecklingens framkant, inte bara nästa kvartal utan i ett längre, mer utmanande perspektiv. Visar dessutom att man är villig att bidra med egen kunskap inom detta område. Ger tilltro, respekt till liten kostnad.

Nokia utvecklar dessutom tillämpningar inom området. Kanske ger dylika satsningar inte avgörande "ringar på vattnet", men ett definitivt bidrag som säkert enkelt kan räknas hem ekonomiskt. Dessutom befinner man sig med automatik i kunskapsströmmarnas centrum.

- En intressant synpunkt från Lassila i anslutning till agenter: När agenter i större utsträckning kommer att representera oss på nätet blir det agenterna som kommer i kontakt med de annonser som idag utgör affärsidén för många webbaktörer. Men agenterna kanske varken vill se eller ta till sig dessa budskap, i alla fall inte på samma

sätt som människor. Vilket kan komma att resultera i att helt nya affärsmodeller kommer att behöva etableras i webbmiljön.

- Digitala agenter diskuterades flitigt för några år sedan i anslutning till SW men har inte materialiserats nämnvärt. Förmodligen på grund av en inneboende komplexitet i konceptet, men förmodligen lika mycket för att tiden ännu inte är mogen på bred front. Det finns så mycket mer jordnära att bekymra sig om just nu som prestanda, säkerhet, integrering av legacy system, avtal-juridik i anslutning till informationssamverkan, med mycket mera. För att inte tala om de semantiska fallgroparna.
- Egentligen har inte så mycket hänt sedan förra konferensen. Fortfarande diskuteras semantik utifrån ett antal olika infallsvinklar och med olika uppfattningar om hur semantiken representeras, alltifrån XML-taggar till det mest invecklade regelbaserade språk. Entusiasmen är det inget fel på. Nya, mindre företag tycks dyka upp om än inte i någon nämnvärd takt. Forskning, utvecklingsprojekt är vanligare företeelser.
- Deltagarna representerar många olika kompetensområden, bl.a. lingvistik, databaser, konceptuell modellering, artificiell intelligens, systemering, dokumenthantering, SOA, Det blir något av en smältdegel där det som förenar är intresse för semantik. Givetvis blir resultatet stundom att man ibland pratar förbi varandra, men glöden i intresset för semantik och viljan att föra fram semantik, som central dimension i de flesta sammanhang, förenar. Här samlas både forskare, utvecklare, leverantörer och allmänt intresserade. Med någon övervikt för den förstnämnda kategorin.
- Bekymret med det så kallade semantiska området kan vara just att det inte är något enkelt identifierbart område. Semantik kommer in överallt, oavsett vad som ska beskrivas och varför. Det kommer in i vårt mänskliga umgänge genom de tolkningar vi gör av utbyte i vardagslag genom vårt modersmål, Det dyker (förhoppningsvis) upp i anslutning till de mest djupstrukturerade bitströmmar som digitala objekt skickar, tar emot och lagrar.

I grunden är det fråga om att om möjligt nå fram med budskap eller utsagor eller att etablera någon form av kunskap. Var behövs inte det? Men som sagt, kunde 20000 deltagare, redan de första åren i slutet av 1990-talet, lockas till konferenser som specifikt behandlade språket Java så är det inte konstigt att RDF, OWL med flera så kallade semantiska språk lockar 600 personer till SW-konferenser.

- På tal om RDF och OWL; OWL har etablerats med stabilt rykte. Språket kommer säkert att användas inom allt flera sammanhang. Redan anses flera tusen ontologier finnas definierade i OWL. Det verkar som om OWL vinner terräng på RDFs bekostnad, samtidigt som RDFs enkelhet fortfarande lockar. Man pratar gärna om de möjligheter till slutsatsdragning som formuleringar i OWL erbjuder. Dock visas sällan realistiska praktikfall på detta, bara små enkla exempel.

- RDFs och OWLs förtjänster och brister diskuterades alltså ofta och gärna. Men bekymret är idag sällan bristen på modelleringsspråk och deras uttryckskraft. Bekymret är att komma överens om semantik, tolkning, mening när det gäller termer och begrepp i den egna verksamheten, samt möjligtvis att formulera det man kommer fram till i något språk. Hur många saknar den möjlighet till regeldefinition som finns i OWL och inte finns i något annat språk? Ytterst få av de modelleringsinriktade presentationerna pekade på behovet. I alla fall inte i dagsläget.
- Värt att notera i sammanhanget är att UML, ER-modelleringsspråk, och liknande som funnits och använts i många år knappast nämns. Lustigt. Den uttryckskraft som UML erbjuder för datamodellering klarar det mesta av vad som diskuteras på konferensen kring semantiska uttrycksbehov. Kanske lockas man tro att UML är något väsensskilt bara för att det utgår från den objektorienterade paradigmen som tillåter specifikation även av klassers beteenden. Kanske upplever man att den semantik som har relatering till webben är väsensskild annan semantik och därför behöver egna språk.

Upprinnelsen var ju behovet av att beskriva digitala resurser på webben, en beskrivning som man av någon anledning kommot att kalla metadata. På vilket sätt är semantik för metadata olik semantik för annan information? Inte alls, bara ett annat tillämpningsområde än vad det administrativa tillämpningsområdet normalt jobbar med. Därmed ingen kritik mot OWL. OWL har förtjänster som inte UML kan matcha och som kan vara användbara i vissa typer av tillämpningar. Dit hör till exempel möjligheten att samtidigt arbeta med relateringar över förekomst - typnivåer.

- Någon noterade att USA var tidigt ute inom området med DARPA-stödda aktiviteter men att det nu är mer fart i Europa med de satsningar som görs inom det 6:e ramprogrammet (se www.sdk-cluster.org) och andra projekt. Tycker nog inte att detta med tidigt ute riktigt svarar mot verkligheten.

I Europa och inte minst i Sverige har begreppsmodellering tillämpats sedan 1970-talet. Begreppsmodelleringsspråket "Stanli" definierades för mer än 15 år sedan inom ett svenskt standardiseringsprojekt med samma namn (numer under SIS) och då utgående från en framsynt forskning under något decennium. Under denna period var det relationsmodellen som gällde på andra sidan Atlanten, möjligtvis kompletterad med någon Entity-Relationship-modellering. Dock hela tiden med fokus på att fånga strukturen för en databas. Nåväl, detta är historia men förtjänar att påpekas.

- På tal om Europa; Semantic Web Services som ges en framträdande plats i europeiska projekt kring den semantiska webben diskuterades knappast. Borde givits större utrymme. Enligt egen uppfattning är komplettering av SOA och då specifikt Web Services Architecture med utökad semantisk uttryckskraft ett angeläget tillskott som hittills inte givits den plats det förtjänar. Förmodligen på grund av bristande intresse hos de teknikorienterade företrädare som driver utvecklingen.

Möjligtvis upplever man fortfarande XML vara tillräckligt. Kanske har inte heller behoven tillräckligt starkt uttryckts från användarhåll. Kanske har man helt enkelt

ännu inte konfronterats tillräckligt hårt med bekymren. Erfarenheter av större interoperativa satsningar är ännu begränsade. Men att dessa kommer i sinom tid på bred front råder knappast något tvivel om.

Sedan återstår det förstås för de europeiska projekten (WSMO – www.wsmo.org - bland annat) att lägga ambitionsnivån på en realistisk nivå och inte enbart förtjusas över de häftiga, mer forskningsbetonade utmaningar som finns att hämta. Dessa må leva i akademisk miljö. Men det behövs rimliga ambitionsnivåer som kan sättas i drift snarast möjligt, helst redan nu.

- Någon noterade att ”Semantic Web” är vilseledande. De flesta relaterar ”web” med någon browser, alltså snarast till presentation av information medan semantik har med innebörd att göra, alltså med mycket annat än presentation. Kanske är därför ”Semantic Technology” eller liknande beteckningar att föredra i fortsättningen.
- På tal om ”web”-delen; det är helt okontroversiellt att påstå att vi den vägen översköjs av information eller snarare data. Följaktligen blir det allt svårare att hitta det vi vill, om ingenting görs. Metadata och mer semantik är ett sätt att hjälpa till att öka precisionen i utsökningar. Intelligent urval utifrån innehåll ett annat (som Google förordar).
- På tal om detta; Årets konferens innebar den definitiva döden för ”conceptual models”. Ingen använde beteckningen. Det är ”ontologies” som gäller. I alla fall på denna arena.
- Ontologier hade som sagt självfallet en framträdande plats på konferensen och har varit något av ett underförstått mantra eller fundament för den semantiska webben. Något förra årets keynote speaker Peter Norvig från Google fick erfa när han förde fram bedömningen att formell strukturering av information i ontologier i många fall inte hinner med utvecklingen. Informationen strömmar för snabbt och är kontinuerligt under förändring i inriktning och betydelse, även inom ett och samma ämnesområde.

Bättre då att utgå ifrån informationen själv och därifrån finna de semantiska samband som sedan kan utnyttjas för intelligenta klassificeringar och utsökningar. Vilken hädisk tanke ansåg ontologiföreträdarna då. Någon ifrågasatte öppet vad denne keynote speaker hade på konferensen att göra. Sökmotorer var helt enkelt undermåliga på att hantera semantik. De utgick endast från termer och i bästa fall syntax.

Ett år senare tycks inställningen vara förändrad. Tankegången är rumsren, i alla fall inom vissa tillämpningsområden. Mer avancerade analysstrategier är under utveckling. Möjligheten att fånga semantik i så kallad ostrukturerad information blir allt bättre. IBMs UIMA (Unstructured Information Management Architecture) är ett intressant och som det verkar mycket ambitiöst exempel (utöver Googles egna satsningar). Inte minst för webbmiljön är detta med att fånga semantiken utifrån informationsmassan ofta ett tilltalande alternativ. Kravet på 100% träffsäkerhet finns

sällan här. Tillräckligt bra, räcker ofta. I andra miljöer återigen fyller förstås ontologier centrala roller.

- Det av Object Management Group (OMG) specificerade ODM (Object Definition Metamodel) presenterades. Rörigt, mycket och en ansenlig portion formalism. Vilket någon i publiken också konstaterade. Allt är nog gott och väl för den lilla skara som håller på att ta fram denna standard. Men för den breda massan är det för komplext och över huvudet på folk. Därmed blir heller inte utvärderingen av det som tas fram effektivt. Användarna orkar inte. Ser heller ingen affärsnytta än så länge. Web Services Architecture är svårt nog. De flesta har bekymmer nog med XML. För att inte tala om att ta fram den enklaste ontologi. Att då börja bekymra sig om språk för formell definition av modelleringspråk, mappning emellan dessa m.m. känns inte angeläget i dagsläget.

Däremot är ODMs genomgång och jämförelse mellan ett antal modelleringsspråks uttryckskraft intressant att ta del av för den nyfikne. På något vis upplevs OMG inte vara i händelsernas centrum varken när det gäller SOA eller Semantic Web. Man tycks befinna sig i en situation där det snickras ihop alternativ till sådant som redan har drivkraft och de kommersiella aktörernas öra. Därmed inte sagt att OMGs specifikationer är sämre. Tvärtom, många gånger, dock ibland med fel timing på något sätt. Men kraftfälten inom de facto-standardisering är högst dynamiska. Något nytt samarbete eller någon viktig kompletterande standard kan snabbt förändra läget. Kanske blir tiden mogen även för ODM inför en bredare publik.

- Oracle presenterade en satsning på RDF i anslutning till Oracle Database 10g. Ett white paper delades ut till alla deltagare. Man figurerade också relativt flitigt med flera föredragshållare. Ska det tolkas som en strategisk vilja att figurera som aktör inom området? Förmodligen.
- Mills Davis höll en uppskattad avslutningspresentation under rubriken ”Semantic Wave”. Vilket engagemang, vilken glöd kring ett relativt abstrakt och svårfångat ämne. Den intresserade kan hämta hans rapport i ämnet på konferensens webbplats www.semantic-conference.com. Väl värd att läsas! Borde övertyga den mest skeptiske om vikten av semantik i alla upptänkliga sammanhang och om att det är ”mycket på gång”.

Tyvärr tog presentationen ett för tidigt slut i och med att ett intensivt brandlarm (ljus och ljud) tog över intresset. Lyckligtvis meddelades efter ett tag att larmet berodde på något fel, men vid det laget hade de flesta redan lämnat lokalen och valt att anträda hemfärden.

- Om man får tro samme keynote speaker (Mills Davis) blir *semantic agents* nästa års ”big thing”, vad nu detta betyder. Låter dock spännande. Det behövs något mer visionärt (än bara semantik och webb en gång till) för att hålla liv i konferensintresset. Digitala agenter har legat ”på lur” en längre tid. Är det dags nu? Kanske i samverkan med spelindustrin?