



Semantic Web - en översikt¹

(2004-05-31)

Stig Berild

(Santa Anna IT Research Institute AB)

¹ Rapport framtagen i samarbete mellan e-Society II-projektet (www.skriver.nu/esociety) vid Santa Anna IT Research Institute (www.santaanna.se) och Serviam-projektet (www.serviam.se) samt med stöd från Stiftelsen SISU och Vinnova.

1 Bakgrund

Till en början utnyttjade vi Internet för att utbyta information i form av dokument, bilder, video. Det sker regelmässigt människor emellan, det sker företag emellan. Så småningom kom XML och med XML möjligheten till utbyte av information vars innehåll har en definierad struktur och där varje element märks med dess betydelse – en anvisning om hur det ska tolkas. Inte minst inom affärssamverkan är precision på betydelsen av den information som utbyts av avgörande betydelse.

Parallellt med XML-utbyte kom Internet att användas för olika slags e-handelsstillämpningar. Så småningom expanderade e-handel till e-affärer, d.v.s. till all typ av affärs- och/eller produktionssamarbete som lämpligen kunde bedrivs över Internet. Till en början rörde det sig fortfarande om informationsutbyte. Men informationsutbyte är i dessa sammanhang sällan isolerade företeelser. De ingår som steg i mer integrerade affärsprocesser. Alltså är det viktigt att både kunna beskriva affärs/produktionprocesser och helst även kunna exekvera affärs/produktionsprocesser i enlighet med dokumenterade beskrivningar. Inte minst viktigt eftersom dessa affärsprocesser i allt ökande utsträckning sker i teamwork mellan ett antal självständiga företag - företag som kan befinna sig varhelst på denna jord. Varje företag utför sin del av det gemensamma arbetet.

Ett sådan avgränsad arbetsinsats brukar kalla Web Service (webbtjänst). För att få till stånd alla delar av en sammansatt affärsprocess kanske den ansvarige behöver kunna söka efter om en viss webbtjänst finns tillgänglig, av vem, var och under vilka villkor. När företaget hittat lämplig webbtjänst vore det ju effektivt om villkoren för utnyttjande av denna tjänst kunde avtalas elektroniskt, momentant. Det vore ju också finurligt att som komplement till beskrivningen av webbtjänstens kapacitet även få kunskap om vilka tekniska krav för ianspråkstagande som gäller. Allra bäst vore det om dessa tillgängliga teknikvillkor skulle kunna nyttjas för att med automatik etablera kontakt med och direkt inlemma webbtjänsten i exekvering. Allt utan mänsklig inblandning.

Fram växer en **dynamisk webbmiljö**. Vi är inte där än, men på väg. De flesta mjukvaruleverantörer erbjuder idag under begreppet ”**Web Services**” någon form av teknikstöd inom detta tema, om än i högst begränsad form.

2 Semantisk webbmiljö

Av avgörande betydelse för att åstadkomma en dynamisk webb är att såväl den information som de processer någon aktör (företag, privatperson, intelligent agent, ...) på Internet erbjuder finns beskrivet på ett sätt som både är rimligt fullödigt för att vara av intresse enligt ovan och har en sådan semantisk precision i beskrivningen att missförstånd undviks. Denna beskrivningsinformation brukar ofta kallas **metadata**.

Riktigt spännande blir det om informationen kan utgöra underlag för olika typer av slutsatsdragningar, intelligenta bedömningar, automatiska analyser. Som stöd behövs standardiserade språk, modeller, regelbaserad uttrycks kraft med mycket mera som är generellt

accepterade av alla aktörer. Det behövs en **semantisk webbmiljö**. De märkord som HTML och XML använder sig av räcker inte till som metadata. Ambitionsnivån måste höjas.

Begreppet ”**Semantic Web**” (den semantiska webben) har myntats för denna framtida webbmiljö som tänks bli baserad på såväl dynamik som semantik. Vissa anser att den semantiska webben kommer att bli nästa stora utvecklingstrend inom webben. Stora ord som revolution och paradigmskifte förekommer, dock utan nämnvärd precisering.

3 Begreppet ”Semantic Web” tar form

Semantic Web myntades av en av webbens förgrundsfigurer Tim Berners-Lee i en rapport från 1998. När en artikel om den semantiska webben, skriven av Berners-Lee, Hendler, Lassila, kom med i Scientific American, maj 2001, tog intresset snabbt fart. Här fanns pirrande visioner, här kunde anas hur nästa generation av webben skulle komma att formas. Vilka härliga godbitar för media att föra vidare – gärna i fantasieggande formuleringar. När mediaintresset strax därefter svalnade blev det snabbt ganska tyst.

De flesta kände nu visserligen till något som hette ”Semantic Web” (SW) och som var framtiden. Men få hade en aning om vad det egentligen stod för ”under ytan”. Och de som sökte tyckte sig upptäcka mycket luft och lite substans. Det hjälpte knappast att Berners-Lee och det av honom ledda World Wide Web Consortium (W3C) bildade arbetsgrupper och talade varmt för SW på konferenser. Om möjligt var det kommersiella intresset än mindre.

Däremot såg universitetsforskningen återigen en chans att komma med i det allmänna utvecklingsracet när nu de stora kommersiella aktörerna och större standardiseringsorganen tagit kommandot kring de ny webbarkitekturerna (SOA). SW var ju något mer långsiktigt framtidsbetonat och i och med detta en betydligt längre och lugnare väg att vandra. Dessutom innehöll SW utmanande frågeställningar, lämpligen ämnade för forskningsinsatser och andra djupare funderingar.

Inte minst företrädare för det under en längre tid inte speciellt vibrerande AI-området (Artificiell Intelligens) tyckte sig finna en ny intressant plattform att arbeta utifrån. Alltså mutade forskningsintressen in SW-området för egen del, inte officiellt förstås, men indirekt genom att ge sig själva ett slags tolkningsföreträdare. Vilket fick till följd att de kommersiella aktörerna tills vidare valde att utebli på denna arena till förmån för Web Servicesområdet.

Visionerna formuleras som sagt i form av högst vaga, mångtydiga utsagor. Här följer ett par som fångats upp på W3Cs webbplats under Semantic Web:

“The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries.”

“The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.”

I artikeln i Scientific American sägs att "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Metadata måste alltså kunna beskrivas på ett sätt som gör att semantiken framgår med god precision. Denna semantik tänks bli formulerad i så kallade ontologier. Likheten mellan det man här kallar ontologi och det man av hävd i datamodelleringsammanhang kallar begreppsmodeller tycks påtaglig, ja näst intill fullt överensstämmande. En förfäktad skillnad är att ontologier kan omfatta såväl förekomster som dess abstraktioner. Något säkert andra anser att även en begreppsmodell klarar av.

Oavsett vilket; ontologier/begreppsmodeller måste kunna uttryckas i modelleringsspråk med tillräcklig uttryckskraft. Alltså anses det vara angelägna uppgifter för SW att ta fram såväl språk som principer och metoder för ontologiframställning. Kanske även att ta fram specifika ontologier inom lämpligt generella tillämpningsområden.

Nästa ambitionsnivå är att med AI-influens utveckla kompletterande regelbaserade språk som dessutom erbjuder intressanta slutledningsförmågor av olika slag.

Så är ännu läget i stora drag.

Samtidigt tycks det i ökande grad finnas falanger inom SW som strävar efter att nå ut till den mer tillämpningsorienterade världen. Vad är i det avseendet bättre än att söka kontakt med de mer avancerade tankegångarna inom Web Serviceområdet (WS)? Inom dess ramverk finns ju skikt som är inriktade mot beskrivning av Services (tjänster), dess egenskaper och gränssnitt. Även kompletterande uppgifter som ägare, ekonomiska, juridiska och andra villkor, mm finns företrädda. Lägg till detta skikt för att beskriva sammansatta processer. Komplettera med specifikationer av den information som avses utbytas mellan Services. Tala om metadata!

Dessa metadata måste ju kunna beskrivas på ett sätt som erbjuder semantisk distinktion både för människor och i förlängningen – i en mer automatiserad värld – för digitala autonoma services. Den semantiska profilen på WS-arkitekturen måste ges ökat utrymme. Anser med rätta SW-folket. Dags för samarbete mellan SW och WS alltså. En påtaglig notering av denna omständighet är det nyligen myntade **Semantic Web Services (SWS)**. En god blandning av respektive falangs ord. Mer om SWS i nästa avsnitt.

I stort kan alltså, i mycket grova drag, tre grenar av SW identifieras:

- Modelleringsspråk och Metadata.
- Regelbaserade språk och slutledningsförmåga.
- Semantic Web Services.

Både inom, mellan och utanför dessa grenar är åsikterna många, något som får anses vara normalt inom nya utvecklingsområden. I förlängningen förmodligen också fruktbart att inte ansätta en stabil form alltför tidigt. Tankar och idéer måste få ha sin frisvängande tid.

Helt klart tycks viljan och intresset för SW återigen vara i stigande. Här finns vittring på nästa generation av webben. Att webben är under snabb utveckling råder knappast någon tvekan om. Både användar- och teknikdriven sådan. Vilken form och riktning utvecklingen kommer

att ta är dock mindre klart. Att de semantiska aspekterna förtjänar en mer framträdande plats tycks numer ändå de flesta vara överens om.

Men står vi verkligen vid startsträckan för en så kallad webbrevolution? Eller borde vi använda webbevolution som ett mer representativt begrepp för det som nu pågår? Kanske är det i grund och botten inte alls något speciellt märkvärdigt vid närmare påseende – men viktigt ändå? Tänk om idéerna inte ens håller eller får tillräckligt genomslag? Tänk om webbutvecklingen tar helt andra riktningar? Eller tar riktningar som är betydligt mer visionärt utmanande? Frågorna är många – och intressanta!

4 Dagens Semantic Web i mer konkreta termer, tack

Egentligen finns inget svar på denna önskan eftersom SW är en generell beteckning på en oklar profil för en framtida webb. Den som så önskar må ha sin uppfattning. Om vi utgår ifrån vad W3Cs arbetsgrupp inom området fokuserar på blir det inte speciellt kittlande. XML har anammats som ett utbytesformat för modelldata. Modelleringspråket Resource Description Framework (RDF) ses därutöver som en central hörnsten.

En vidareutveckling av RDF för att inkludera mer uttryckskraft har resulterat i språket OWL. Båda språken antogs den 10 februari 2004 som Recommendations av W3C, d.v.s. som genomarbetade standarder. För den som går in på W3Cs webbplats (www.w3c.org) finner att respektive språk i realiteten beskrivs i ett flertal dokument.

Därutöver arbetar man med ett frågespråk mot RDF-definierade data och lite annat.

Man anser sig, genom dessa två språk nu i skick som Recommendations, ha nått en viktig milstolpe. W3C deklarerar att det är dags att nå ut ur den mer akademiska enklaven och börja använda språken på bred front. En nästan nödvändig förutsättning för att detta ska ske är att stödjande verktyg utvecklas till kommersiella produkter. Där finns fortfarande en hel del kvar att göra. En förmodligen viktigare uppgift kommer att bli att hitta seriösa tillämpningsområden, att finna ett par så kallade ”killer applications” för att skapa genuint intresse. Samt att övertyga alla dem, som nu arbetar med andra modelleringspråk, om att dessa inte är lämpliga för metadata.....

Intressantare är den utveckling som äger rum inom ramen för vad som man valt kalla Semantic Web Services (SWS). En hel del överlappning med de ansträngningar som görs inom det (mer konventionella) Web Servicesområdet (WS) kan noteras. Dock kompletterat med en hel del innovativt tänkande som det mer kommersiellt orienterade WS valt att ännu inte ta tag i eller ännu inte fångat in. Genom några projekt, nyligen startade (januari 2004) och massivt finansierade av EUs 6:e ramprogram, kommer förhoppningsvis en hel del nytt spännande att tas fram inom SWS. I vad mån eller i vilken form de förväntade innovationerna kommer att nå den kommersiella arenan är, som vanligt med dessa typer av projekt, oklart.

Projekten heter DIP (Data, Information and Process Integration), SEKT (Semantic Knowledge Technologies) och Knowledge Web. Projekten samordnas under något som kallas SDK. Bokstavskombinationen står för respektive projekts förstabokstav. Se www.sdk-cluster.org och dess länkar för vidare information om de olika projektens förehavanden. Inte minst WSMO (Web Services Modelling Ontology) innehåller en hel del spännande

tankegångar. Motsvarande ansträngningar görs i USA genom det av Defense Advanced Research Projects Agency (DARPA) stödda programmet DAML (DARPA Agent Markup Language). Se vidare på deras webbplats www.daml.org.

Även om SWS tycks vara den ”räddningsplanka” på vilken SW huvudsakligen kommer att kunna motivera sitt fortsatta berättigande, tycks upplägget idag företrädesvis kännetecknas av en hel del ”finlir” i form av detaljerade specifikationer av olika slag. Med hjälp av dessa kan man säkerligen beskriva det mesta av information och tjänster/services för exponering. Dock saknas nästan helt diskussion kring hur denna information kommer till användning. För vad? Under vilka förutsättningar? Inom vilka dynamiska förlopp? Med vilken QoS? Vilka risker? Hur kopplas dessa nya möjligheter till existerande behov hos individer, företag och organisationer? Med mera. Var finns de mer övergripande visionerna?

5 Visionen om ett E-samhälle

På tal om vision; Ordning och reda, stabilitet och struktur genomsyrar det mesta av vad moderna Service Oriented Architectures (SOA) står för och försöker erbjuda. Inte minst WS är ett typiskt exempel. Men vi lever i en snabbföränderlig värld. Kanske har vi inte alltid tid att lära känna varandra innan vi samarbetar. Kanske ser vi ändå en fördel att påbörja samarbetet med förhoppningen att vi allteftersom ska kunna förstå varandra bättre baserat på erfarenheter, indikationer, signaler, tolkningar. Kanske bedömer vi det helt enkelt värt att ”chansa” eftersom omständigheterna inte erbjuder andra alternativ.

Samma osäkra förutsättningar kring exempelvis betydelse och kvalitet kan visa sig gälla för den information vi av olika skäl är intresserade av men ändå vill försöka nå och förstå. Hur kan vi bygga upp och hantera miljöer som har att leva under osäkra förutsättningar? Hur kan dessa miljöer på egen hand successivt bygga upp och nyttiggöra den erfarenhet som det faktiska samspelet med omgivningen trots allt ger? Hur kan de förmås formulera kloka beslut baserade på osäkerhet men inklusive viss erfarenhet? Och så vidare.

Egentligen är inte detta att leva med osäkra förutsättningar någonting nytt. Det är snarast ett perspektiv på livet självt. Ändå tycks vårt samhälle i stort fungera på ett rimligt sätt. Har vi måne inte en hel del att lära från våra dagliga livsmönster när det gäller realisering av IT-lösningar. Trots allt är ju våra livsmönster och vårt samhällssystem baserat på en synnerligen lång erfarenhetshistoria. Varför inte basera vårt tänkande kring webbaserade affärsprocesser, samverkan, avtal, upphovsrätt, tilltro, kunskap, m.m. mer direkt på dessa erfarenheter? Varför inte försöka efterlikna det verkliga livet i e-miljön? Inte som en simuleringsmodell för nöjes skull utan för att åstadkomma ett rikare mer integrerat och flexibelt IT-stöd som på ett mer nyanserat och effektivt sätt kan svara upp mot våra behov och förväntningar.

Resultatet skulle kunna bli ett ekosystem eller en dynamisk e-livsmiljö befolkad av element som föds, lever och dör, som agerar i olika roller under olika perioder av sitt liv, tillhör olika grupperingar (”communities”) av legala eller intressemotiverade skäl, som under mer eller mindre reglerade förutsättningar utför arbetsuppgifter enskilt eller i samverkan, som kanske bedöms eller kontrolleras av element i andra roller, som

Visst låter det både spännande, svårt, kanske lite kusligt. Dock inte nödvändigtvis så avlägset som det låter. Inom mer avgränsade problemställningar pågår redan diskussioner, utbyts idéer,

etableras ramverk med kreativiteten i högsätet. Bland tidiga exempel kan nämnas eCo Framework framtaget av Commercenet (www.commercenet.com). Förmodligen kommer en spretig flora av framtidsscenarier och stödjande teknik kring ekosystem att växa fram. Kanske kommer de att så småningom inordnas under den semantiska webbens paraply. Kanske kommer ett parallellt spår att växa fram.

Många frågetecken, många kanske, ja ett helt spektrum av frågeställningar från det mest spektakulära till det mest jordnära. Men är det inte just det framtiden alltid erbjuder? Är det inte just detta som också stimulerar, lockar till kreativitet?

Helt klart lever vi i en snabbföränderlig värld, en värld som ständigt fylls av ny teknik, erbjuder nya möjligheter, nyttjas för nya tillämpningar och tjänster, genererar nya utmaningar, skapar nya hotbilder. För näringsliv och organisationer är det helt avgörande för överlevnad och utveckling att hänga med både inom e-affärer och alla andra områden som på olika sätt har sin kontakt med Internet och webben och ofta givits prefixet "e-". Även som individer är vi på olika sätt och till olika grad direkt eller indirekt inflyttade i och beroende av de nya livsmiljöer som skapas. I förlängningen hyser vi förstås förhoppningen att effekterna ska bli en långsiktigt uthålligare livsmiljö och ge rikare livskvalitet. Eller kanske något annat mer jordnära formulerat, eftersträvansvärt utfall.

6 Avslutande funderingar

Än så länge är det svårt att hitta lockande resultat som föranleder kommersiella aktörer att gå in i SW-området. Inte heller finns några speciellt visionära idéer med lyskraft. SWs starkaste tillgång, ja kanske den enda av dignitet i dagsläget, är dess namn. "Web" är intressant och står för något dynamiskt framåtsträvande. "Semantic" tolkas som något förfinat, något som förädlar webben. Möjligtvis något orättvist formulerat, dock inte alldeles missvisande.

Ett par kommentarer i anslutning till de tre grenarna:

- Modelleringspråk och Metadata

RDF och OWL är i sina huvuddelar högst konventionella modelleringspråk som hämtat det mesta av sin uttrycks kraft från andra modelleringspråk, där de senare funnits och använts under många år. Något av "uppfinna hjulet på nytt" kan tyckas. Visserligen säger man sig företrädesvis hantera webbresurser och av den anledningen uppleva sig arbeta med modeller för metadata unikt, men det finns inget som hindrar andra etablerade modelleringspråk att göra detsamma. Dessutom finns mycket att säga om "metadata". Se gärna rapporterna "Metadata – Vad, När, Hur, Varför?" och "Mycket 'meta' är det" för en diskussion.

- Regelbaserade språk och slutledningsförmåga

Kommer detta med slutledningsförmåga att vara en eftertraktad del av den kommande webben? Kommer det att fungera i en öppen värld där alla möjliga aktörer ska kunna prata om i princip allting och göra det med olika kulturella, sociala, kunskaps- och syftesrelaterade referenser? Kommer det ens att vara möjligt att i rimlig omfattning etablera förankrade ontologier? Om så, ligger det i rimlig närtid? Kritiska kommentarer

saknas inte. Ta exempelvis http://www.shirky.com/writings/semantic_syllogism.htm. Knappast någon mer vetenskaplig analys, men rolig. Är det möjligtvis så att de som företräder denna gren just nu gör SW en otjänst genom att ge det en något oförtjänt flumstämpel?

- Semantic Web Services.

Förhoppningsvis kommer de ansehliga summor som EUs 6:e ramprogram lagt ut på de tre nyligen initierade Semantic Web-orienterade projekten att bära frukt. Även i USA pågår på olika håll utvecklingsaktiviteter. På webbplatsen www.daml.org finns en omfattande lista på institutioner och forskare med beröring till DAML.

Skillnaden mellan SWS och WS är knappast av en art som motiverar uppdelning i två olika insatsområden. Den nya beteckningen SWS kan för närvarande minst lika väl stå för de drivkrafter som finns inom WS som för de som finns inom SW. WSDL och UDDI är ju alldeles påtagliga begreppsmodeller även om de uttrycks i ett semantiskt torftigt språk som XML. Även om det inte direkt är fråga om något ”uppfinna hjulet på nytt”, SWS bygger på en något mer vidareutvecklad idébas, så är överlappningen stor. De resultat som SWS förhoppningsvis kommer att producera, kommer lika förhoppningsvis att utgöra inspiration för olika delområden inom WS. Snarare än att ta över där WS gör halt.

Det riktiga lyftet för SW låter ännu vänta på sig. Kommer det eller kommer saker och ting att sakta avsomna, att sugas upp av andra trender? I vilken mån kommer SDK-projekten och DAML att kunna hävda sina intressen, att spela en aktiv roll även utanför den mer lokala gruppen entusiaster? Kommer WS att sluka de delar som det finner användbart? WS är ju under snabb vidareutveckling och underställt ansehliga resurser. Bakom dessa resurser finns i förlängningen ett intensivt tryck på någon form av ekonomisk avkastning, ett tryck som knappast de EU-finansierade projekten känner av i lika stor utsträckning. Vilket förstås både kan vara en fördel och en nackdel.

Förresten, är det så alldeles självklart att WS kommer att lyckas fullt ut med sitt ramverk i långa loppet? Kommer något helt annat att växa fram i dess ställe? Någoting baserat på tillämpningsmönster ingen idag har någon aning om?

SW genererar som sagt mest frågor för tillfället. Låt oss hoppas att några av dem så småningom kan vändas till utropstecken. Den närmaste treårsperioden kommer att bli avgörande. Levererar SDK- och DAML-aktiviteterna substantiella, attraktiva resultat finns ett fortsatt berättigande. Då finns en plattform ur vilken nya, intressanta utvecklingsgrenar kan ta form. Annars är nog ”loppet kört” för det självständiga revir SW tillåtits profilera sig inom idag. Men i det läget finns säkert andra alternativa ”revir” att flyta in i.

eSociety-projektet avser att i görligaste mån ta del av, följa och rapportera om dessa projekts framskridanden.