



Semantiska webben och sökmotorer

Semantic web and search engines

Datum: 2010-06-07

Version: Slutversion

Författare

Amir Haj-Bolouri 850420

amirhajbolouri@gmail.com

Handledare

Professor Per Flensburg

Examinator

Ulrika Lundh Snis

C-Uppsats

Informatik 15 hp

Abstract

This report deals with the definitions and terms that relates to the semantic web. The main purpose has been to investigate how the semantic web affects search engines on the web. This has been done through an investigation consisting of ten different search engines. Nine of these search engines are considering being semantic search engines, and the last one being the most used one on the web today. The study is conducted as a descriptive and quantitative study. A literature review has also been implemented by the relevant sources about the semantic web and search engines. The conclusions drawn where that the semantic web is multifaceted with its definitions and that the result of how concrete search engines implements semantic web principles can vary depending on which search engine one interacts with

Sammanfattning

Den här rapporten behandlar definitioner av begrepp som är kopplade till den semantiska webben. Syftet är att undersöka hur den semantiska webben påverkar sökmotorer på webben. Detta sker genom en undersökning av tio olika sökmotorer där nio är semantiskt sådana och den tionde är den mest använda sökmotorn idag. Studien är genomförd som både en deskriptiv och kvantitativ studie. En litteraturundersökning har också genomförts om den semantiska webben och sökmotorer. Slutsatserna av den här studien är att den semantiska webben är mångfacetterad med dess definitioner, och att resultatet kring hur konkreta sökmotorer tillämpar semantiska webbprinciper kan variera beroende vilken sökmotor man interagerar med.

Nyckelord: Semantic web, Semantiska webben, Semantik, Informatik, Web 2.0, Internet, Search engines, Sökmotorer

Högskolan Väst
Institutionen Ekonomi och IT
C-uppsats Informatik 15 hp
Haj-Bolouri, Amir

Förord

Jag vill ta tillfället i akt och tacka min flickvän Maria som har stått ut med mig under den intensiva period det varit att skriva den här uppsatsen. Jag vill också tacka min handledare Professor Per Flensburg för adekvat, ärlig samt konsekvent kritik. Och det är även honom jag har att tacka för hans bidragande klokhet under de senaste fyra åren av varierande studier.

Tack!

Amir Haj-Bolouri

Innehållsförteckning

ABSTRACT	II
SAMMANFATTNING	II
FÖRORD	III
INNEHÅLLSFÖRTECKNING.....	IV
1. INLEDNING	1
1.1. INLEDANDE OM DEN SEMANTISKA WEBBEN	1
1.2. VARFÖR SKRIVA OM DEN SEMANTISKA WEBBEN?	1
1.3. PROBLEMRÅDE OCH FRÅGESTÄLLNING	1
1.4. SYFTE	2
1.5. AVGRÄNSNINGAR	2
1.6. MÅLGRUPP	2
1.7. DISPOSITION	2
2. METODER	3
2.1. ÖVERGRIPANDE METOD SAMT METODANSATS	3
2.1.1. TYP AV STUDIE	3
2.2. INFORMATIONSSAMLINGSMETOD	4
2.3. KVANTITATIV UNDERSÖKNING	5
2.3.1. URVAL OCH AVGRÄNSNING AV UNDERSÖKNINGSENHETER	5
2.3.2. SÖKFRÅGOR	5
2.3.3. UNDERSÖKNINGENS UPPLÄGGNING	6
2.4. STUDIENS VALIDITET OCH RELIABILITET	7
3. TEORETISK REFERENS RAM SAMT DEFINITIONER	8
3.1. DATA, INFORMATION OCH KUNSKAP	8
3.2. KORT OM WEBBEN	9
3.3. KORT OM W3C	10
3.4. WEB 2.0 I ETT NÖTSKAL	10
3.4.1. VAD ÄR AJAX?	11
3.5. DEN SEMANTISKA WEBBEN	12
3.5.1. VAD ÄR SEMANTIK?	12

3.5.2. VAD ÄR DEN SEMANTISKA WEBBEN FÖR NÅGOT?	13
3.5.2.1. METADATA	16
3.5.2.2. VAD ÄR ONTOLOGIER?	16
3.5.2.3. SEMANTISKA WEBBENS GRUNDLÄGGANDE SPRÅK	17
3.5.2.5. Vad är RDF för något?	18
3.5.2.6. Vad är RDFS för något?	19
3.5.2.7. Vad är OWL för något?	19
3.6. SÖKMOTORER	20
3.6.1. VAD ÄR EN SÖKMOTOR?	20
3.6.2. HUR FUNGERAR EN SÖKMOTOR?	20
3.6.2.1. Spindeln	20
3.6.2.2. Indexeraren	21
3.6.2.3. Frågebehandlaren	21
3.7. SEARCH 2.0	22
3.7.1. GOOGLE SEARCH	22
3.8. SÖKMOTORER I DEN SEMANTISKA WEBBEN	23
3.8.1. SEMANTISKA SÖKPRINCIPER	23
3.8.2. SWOOGLE	24
3.8.3. HAKIA	25
3.8.4. LEXXE	28
3.8.5. KOSMIX	29
3.8.6. COGNITION SEARCH	29
3.8.7. POWERSET	29
3.8.8. SENSEBOT	30
3.8.9. FACTBITES OCH EXALEAD	30
4. RESULTAT	31
4.1. SAMMANFATTNING AV DEN SEMANTISKA WEBBEN I BEGREPP OCH DEFINITION	31
4.2. RESULTAT FÖR SÖKTESTET	32
4.2.1. HAKIA	32
4.2.2. SWOOGLE	32
4.2.3. EXALEAD	33
4.2.4. POWERSET	33
4.2.5. SENSEBOT	33
4.2.6. FACTBITES	34
4.2.7. COGNITION SEARCH	34
4.2.8. LEXXE	34
4.2.9. KOSMIX	35
4.2.10. GOOGLE SEARCH	35
5. ANALYS OCH DISKUSSION	36
5.1. DEN SEMANTISKA WEBBENS ROLL OCH SYFTE	36
5.2. SÖKTESTET UTIFRÅN ETT ANVÄNDARPERSPEKTIV	38
5.3. VAD ÄR SKILLNADEN MELLAN SEMANTISKA SÖKMOTORER OCH GOOGLE SEARCH?	39
5.4. HAR DEN SEMANTISKA WEBBEN LYCKATS?	40

Högskolan Väst
Institutionen Ekonomi och IT
C-uppsats Informatik 15 hp
Haj-Bolouri, Amir

5.5. MIN DEFINITION AV DEN SEMANTISKA WEBBEN.....	41
6. SLUTSATSER.....	43
7. FORTSATT FORSKNING.....	43
8. REFERENSER.....	44

1. Inledning

Webben går idag under begreppet "Web 2.0". Begreppet myntades år 2004 av Tim O'Reilly samt *MediaLive International* under en webbkonferens (Vossen & Hagemann, 2007) och omfattar webben som den ser ut idag. Upphovsmännen till webben som Tim Berners Lee, anser dock att webben är på väg mot ett nytt spektrum, nämligen den semantiska webben.

Under nedanstående rubriker diskuteras den semantiska webben samt ges en vidare inledande inblick i problemområde och ämnesval.

1.1. Inledande om den semantiska webben

Den semantiska webben är ett koncept som drivs av Tim Berners-Lee och World Wide Web Consortium (W3C, 2010), vilka även anses vara skaparna av *World Wide Web* (Semantic Web Org, 2009). Konceptet har till syfte att möjliggöra för användare att dela med sig innehåll samt information, som sträcker sig bortom de begränsningar som dagens applikationer och hemsidor har. Andra definitioner menar på att den semantiska webben har som primärt syfte att gynna "maskiner" över webben, genom att låta de tillhandahålla bättre integrerat utbyte av information (W3C, 2010). Detta innebär med andra ord att det behövs teknologi som bidrar med, och upprättar ett enhetligt språk, vilket webbapplikationer sedan skall kunna kommunicera med och genom.

1.2. Varför skriva om den semantiska webben?

Som en flitig användare av webben, så är jag även en granskande sådan. Därför är webben och dess potentiella utveckling en självklar kategori för mig att skriva om. Detta val har dock även påverkats av tidigare undersökningar vid föregående kurstillfällen samt diskussioner kring Web 2.0 och den semantiska webben. Mina tidigare studier kring den semantiska webben som ett koncept, har resulterat i att det finns flera aspekter av konceptet att studera. En aspekt är hur semantiska webbprinciper skulle kunna påverka dagens sökmotorer, där söktid och sökprecision står i fokus för förbättring (Vossen & Hagemann, 2007).

1.3. Problemområde och frågeställning

Det problemområde som jag har till syfte att undersöka samt studera om, är den semantiska webbens sökmotorer. Nedan följer problemställningarna i uppsatsen:

- **Vad är den semantiska webben?**
- **Hur betar sig sökmotorer i relation till semantiska sökprinciper?**

1.4. Syfte

Syftet är att besvara på problemställningarna genom att utföra en deskriptiv studie om den semantiska webben, och ett test på hur semantiska sökmotorer förhåller sig till semantiska sökprinciper. Studien utgörs ur ett användarperspektiv, där jag som användare är nyfiken på vad den semantiska webben är, och hur semantiska sökmotorer fungerar i praktiken.

1.5. Avgränsningar

För att sätta definitionerna och innebörden av den semantiska webben i en kontext, har jag valt att avgränsa studien till hur semantiska webbprinciper skulle kunna påverka sökmotorer. Avgränsningen sker mot 9 stycken semantiska sökmotorer, samt *Google Search*.

1.6. Målgrupp

Målgruppen för den här rapporten är studenter samt forskare som eventuellt skulle kunna finna intressanta uppslag och infallsvinklar till vidare studier i ämnet. Tanken och förhoppningen med den här uppsatsen är att den skall kunna fungera som en omfattande beskrivning av multipla definitioner samt exemplen som behandlar den semantiska webben och sökmotorer.

1.7. Disposition

Kapitel 1 Inledning

Problemformulering och syfte, avgränsningar, målgrupp och disposition.

Kapitel 2 Metoder

Det här kapitlet redogör övergripande val av metoder, samt sekundära val av metoder som datainsamlingsmetod, urvalsmetod samt analysmetod.

Kapitel 3 Teoretisk referensram samt definitioner

I det här kapitlet redogörs för definitioner kring ämnesvalet, samt sekundära definitioner som är sammanhängande med den primära definitionen av den semantiska webben. Det här kapitlet redogör för vad som har framkommit i de litteraturstudier som den här rapporten delvis omfattar. Konkreta begrepp och koncept som det här kapitlet redogör är bland annat: data, information, kunskap, semantik, ontologier, den semantiska webben, xml, rdf, etc.

Kapitel 4 Resultat

Här redovisas det sammanställda resultatet av själva studien.

Kapitel 5 Analys och diskussion

I det här kapitlet analyseras samt diskuteras resultatet utifrån den teoretiska sektionen.

Kapitel 6 Slutsatser

I det här kapitlet redovisas vilka slutsatser som har dragits av själva analysen.

Kapitel 7 Rekommendationer

Här redovisas potentiella rekommendationer för vidare studier utifrån det framkomna resultatet.

Kapitel 8 Referenser

En sammanställande källförteckning över de angivna referenserna i rapporten.

2. Metoder

2.1. Övergripande metod samt metodansats

Holme & Solvang (1997) beskriver en metod som ett redskap, ett sätt att lösa problem och komma fram till ny kunskap, där allt som kan bidra till att uppnå dessa mål är en metod. Vidare beskriver de att metodläran i sig ger forskaren grunden för ett systematiskt och planmässigt arbete kring frågor som rör vem, vad, hur, och varför beträffande samhällliga problem, dock så är enbart metoden i sig ett redskap och därför genererar den inte några svar på dessa frågor.

2.1.1. Typ av studie

Deskriptiva studier används för att bestämma forskningsobjektets egenskaper, genom insamling och systematisering av data, som man sedermera använder för att bestämma värden på variabler och samband (Wallén, 1996).

En **förklarande studie** har å andra sidan till syfte att ta upp "varför-problem", dvs. vad för typ av förklaring är relevant: avsikt-effekt, orsak-verkan, systemeffekter samt mekanismer.

Wallén (1996) beskriver en **explorativ studieansats** som en studie, för att få grundläggande kunskaper om problemets vad, när, var och i vilket sammanhang, hur, samt bestämning av vad som behöver undersökas: Det vill säga, vad är relevanta variabelt, begrepp? Vad hör inte till problemet, vilka alternativ kan uteslutas?

Jag valde en deskriptiv studieansats för att besvara på den första frågan i problemställningen. Detta innebär att jag samlade in information som handlar om den semantiska webben, och sedan systematiserade informationen för att kunna redogöra dem.

2.2. Informationssamlingsmetod

Holme & Solvang (1997) beskriver skillnaden mellan "information" och "data" som begrepp där det sistnämnda allt för ofta förknippas med datorer och datoranvändning, medan det förstnämnda betecknar ett oprecist begrepp som innefattar kunskaper om företeelser, personer samt övriga tänkbara undersökningsenheter. Sedermera beskriver författarna även att, vilken information man önskar att samla in bestäms helt och hållet av frågeställningen, med andra ord skall frågeställningen och problemformuleringen klargöra vilka enheter det är som undersökningen skall ta hänsyn till, samt vilka attribut hos enheterna som man önskar att få upplysningar om. Författarnas definitioner av "information" och "data" skiljer sig dock med de definitioner som är anpassade till den här rapporten. Detta påpekas av Flensburg & Friis (1999) längre ned i rapporten.

Definitioner om vad data, information, kunskap, semantik, ontologi, den semantiska webben, sökmotor etc. är för någonting, samt hur de förhåller sig gentemot varandra förtydligas i den här rapporten. För att kunna göra detta, bör man enligt Backman (2008) ta del av tidigare dokumentation som visar hur begrepp inom ett specifikt område definieras, preciseras och används.

Därför utfördes informationsinsamlingen först och främst som litteraturstudier. Sökandet efter information skedde genom följande sökord:

- "Den semantiska webben"
- "The semantic web"
- "Semantic web"
- "Semantiska webben"
- "Sökmotorer"
- "Search engines"

Litteraturstudien skedde mot litteratur som behandlar den semantiska webben och sökmotorer. I första hand lade jag fokus på boken: *Unleashing Web 2.0* (Vossen & Hagemann, 2007) som behandlar information om sökmotorer och den semantiska webbens påverkan mot dem.

Följande databaser och tidsskrifter sökte jag i för att hitta relevant information:

- *IEEE Xplore*
- *Emerald*
- *Science direct*

2.3. Kvantitativ undersökning

Undersökningen om hur semantiska sökmotorer förhåller sig gentemot specifika söktermer och sökresultat var kvantitativ. Därför att kvantitativa undersökningar har delvis till syfte att ge ringa information om många undersökningsenheter, samt beskriva och förklara informationen (Holme & Solvang, 1997). Detta var nödvändigt för att kunna utföra ett funktionstest med samtliga sökmotorer. Därför utförde jag en kvantitativ undersökning på hur tio stycken sökmotorer behandlar sex stycken sökfrågor. Dock så låg fokus på att testa funktionaliteten hos sökmotorerna, därför var det inte intressant att presentera de faktiska svaren som sökmotorerna genererade i resultatet. Undersökningen bedrivs av mig som användare.

2.3.1. Urval och avgränsning av undersökningsenheter

Följande sökmotorer kommer att utvärderas:

- **Hakia**
- **Swoogle**
- **Exalead**
- **Powerset**
- **Sensebot**
- **Factbites**
- **Cognition Search**
- **Lexxe**
- **Kosmix**
- **Google Search**

Enligt Radhakrishnan (2009) kommer nio stycken semantiska sökmotorer att förändra sökinteraktionen på webben. Dessa nio sökmotorer är detsamma som de nio första ovan i listan. Anledningen till att jag valde dessa är för att testa och se ifall jag som användare upplever kvalitén lika tillfredsställande som (Radhakrishnan, 2009) uttrycker sig om dem. *Google Search* tog jag med i testet för att det är den mest populära sökmotorn idag (Wikipedia: *Web Search Engine*, 2010). Därför valde jag att jämföra de föregående nio sökmotorer med den mest populära, för att se ifall de håller samma typ av kvalitet i sökträffarna eller ej.

2.3.2. Sökfrågor

Konkreta sökfrågor som testet kommer att bygga på är:

1. ***Which university is the oldest in the world?***
2. ***When did Platons Akademia come to an end?***
3. ***Who is Barack Obama?***
4. ***What is the Semantic Web?***
5. ***When was the semantic web first mentioned?***

Sökfrågorna är konstruerade som frågor och inte enskilda meningar eller begrepp. Anledningen till detta var för att kunna avgöra hur undersökningsenheterna behandlar konkreta frågor och dess betydelse.

Jag valde ju ovanstående frågor på grund av deras olika karaktäristika. Den första frågan är en generell sådan som inte har något säkert svar. Den andra frågan tog jag med för att testa språk och grammatik. Den tredje frågan är en fråga vars svar är självklar i praktiken, därför tog jag med den för att se ifall sökmotorerna upplever den som enkel eller ej. De två sista frågorna i listan tog jag med för att de dels är definierande och dels faktaorienterade.

Sökfrågorna kommer ställas enbart på engelska därför att sökmotorerna är primärt gjorda för det språket. Sökfrågorna ställades även utan citationstecken för att sökmotorerna delvis skulle kunna svara explicit på frågorna, samt generera träffar som var relaterade till de enskilda begreppen som sökfrågorna består av.

2.3.3. Undersökningens uppläggning

Undersökningens praktiska tillvägagångssätt bestod huvudsakligen av två faktorer:

- **Testa** – Testa tio stycken sökmotorer genom ett antal ekvivalenta söktermer. Söktermerna var konkreta frågor som respektive sökmotor fick bearbeta.
- **Evaluera** – Evalueringen på resultatet gjordes utifrån följande faktorer. Samtliga faktorer utgör även de mått som jag grundar utvärderingen av sökträffarna på:
 - **Träffsäkerhet** – behandlar sökprecision och relevans hos sökträffar som sökmotorer genererar i förhållande till angivna söktermer (Fransson, 2007). I testet använde jag mig av tre stycken alternativa mätfaktorer för sökträffarnas träffsäkerhet:
 - **”Mycket tillfredsställande”** – Sökträffar med fem eller fler träffar som svarade explicit korrekt per sökfråga. Med explicit menar jag konkreta svar på sökfrågorna.
 - **”Tillfredsställande”** – Sökträffar med fem eller fler än fem relevanta svar, det vill säga sökträffar med relevant information till sökfrågorna. Relevant information innebär
 - **”Mindre tillfredsställande”** – Sökträffar med mindre än fem relevanta
 - **”Icke tillfredsställande”** – Noll sökträffar.

- **Allmängiltighet** – behandlar hur omfattande antalet indexerade sidor är. Även här använde jag mig av fyra stycken mätfaktorer för att mäta allmängiltigheten på sökträffarna:
 - **"Hög allmängiltighet"** – Åtta till tio av tio stycken indexerade sökträffar refererar till relevant information gentemot sökfrågorna.
 - **"Medel allmängiltighet"** – Fem till sju av tio stycken indexerade sökträffar refererar till relevant information gentemot sökfrågorna.
 - **"Låg allmängiltighet"** – Två till fyra av tio stycken indexerade sökträffar refererar till relevant information gentemot sökfrågorna.
 - **"Ingen allmängiltighet"** – Noll av tio stycken indexerade sökträffar.

2.4. Studiens validitet och reliabilitet

Holme & Solvang (1997) säger att studiens reliabilitet bestäms av hur mätningarna utförs samt hur noggrann man är vid bearbetningen av informationen, medan validiteten är beroende av vad vi mäter, samt ifall det är inkluderat i själva frågeställningen. Vidare beskriver författarna att hög reliabilitet uppstår om olika och oberoende mätningar av ett och samma fenomen ger samma eller ungefärligen samma resultat.

Reliabilitet och validitet i förhållande till de källor som behandlar väsentlig information gentemot huvudfrågeställningen, bygger på källornas pålitlighet samt min förmåga att på ett tillförlitligt sätt bearbeta dem.

3. Teoretisk referensram samt definitioner

Den semantiska webben anses kunna förbättra samt optimera aspekter av dagens webbapplikationer (Vossen & Hagemann, 2007). En aspekt av detta är sökmotorer, dess precision och sakförhållande gentemot sökfrågor. Men först, en redogörelse om vad "data", "information" och "kunskap" innebär.

3.1. Data, information och kunskap

Definitionen av data kan beskrivas som följande, exemplet är hämtat ifrån (Flensburg & Friis, 1999):

- **Data** – "Data" definieras som symboler utan mening, kontext eller innehåll. Exempel på data kan vara "31", "45", "Härlanda", "november", "focularin". Data är med andra ord per definition: kontextlösa, innehållslösa och meningslösa.
- **Information** – När väl data sätts i struktur, vilket kan vara flera typer av strukturer, i det här fallet är det dock aktuellt med en syntaktisk struktur, dvs. i en mening, omvandlas data till information. Datauppgifterna ovan kan då sättas i en grammatikalisk syntax: "Den 31 november satt 45 foculariner på Härlanda fängelse". Detta är en mening, utsaga, det är med andra ord information.
- **Kunskap** – Information som av en människa tolkas i ett sammanhang blir kunskap. Data kan sättas i olika strukturer, samtidigt som information kan sättas i olika kontexter, vilket kan tolkas olika och ge upphov till kunskap: ifall samma resonemang om foculariner ovan finns i en tidning som behandlar information om antalet fångar i ett fängelse, kan informationen uppfattas och tolkas på ett sätt. Men skulle samma information sättas in i en kontext där nyheter rörandes Anticimex behandlas, skulle informationen tolkas på ett annat sätt, vilket skulle ge upphov till en ny typ av kunskap.

Författarna vill med ovanstående resonemang visa skillnaden mellan form och innehåll, mellan tolkning och presentation, mellan information och kunskap. Denna begreppsanalys är nödvändig för att kunna behandla hur termerna används i samband med den semantiska webben.

3.2. Kort om webben

Webben är ett system baserat på och av hyperlänkade dokument (Vossen & Hagemann, 2007), och det var Sir Tim Berners Lee som år 1989 presenterade ett förslag om att skapa det här systemet (Wikipedia: *World Wide Web*, 2010). Systemet bestod då, och även idag av ett antal nyckelelement för att det skall vara implementerbart:

- **Webbläsare** – En webbläsare är i dess konkretaste mening ett program som används för att hämta, tolka samt återge hemsidor (Wikipedia: *Webbläsare*, 2010).
- **Webbserver** – Webbservern är ett datorprogram, samt den del av webben som tillhandahåller hemsidorna och möjliggör de för användarna (Wikipedia: *Webbserver*, 2010).
- **Klient/Server principen** – Den här principen är aktuell ännu idag, och den går ut på att en klient, vilket i det här fallet är en dator, skickar iväg en begäran (Request) till en webbserver om att få tillgång till en hemsida. Webbserverns reaktion är att den svarar (Response) klienten med att återge hemsidan i webbläsaren (Vossen & Hagemann, 2007).
- **HTML och http** – HTML står för Hypertext Markup Language och är kortfattat ett webbspråk som Tim Berners Lee skrev för att beskriva webbsidor typografiskt. http står för Hypertext Transfer Protocol och är det protokoll som skickar iväg en begäran från en klient till webbservern, och vice versa med responsen (Vossen & Hagemann, 2007).
- **Hyperlänkar** – Hyperlänkar är den delen av webben som möjliggör för en användare att navigera från en plats till en annan. Hyperlänkar kan vara externa samt interna (Vossen & Hagemann, 2007), med externa länkar menar man länkar som leder en användare från hemsida A till hemsida B, medan interna länkar refererar till länkar som leder användaren från plats A till plats B på en och samma hemsida.

Ovanstående punkter sammanfattar fem stycken fundamentala faktorer av webben som (Vossen & Hagemann, 2007) anser vara viktigast för att webben skall sammanhålla. Jag har däremot utelämnat aspekten kring hårdvara, detta därför att det inte är lika väsentligt i förhållande till huvudfrågeställningen.

3.3. Kort om W3C

W3C står för *World Wide Web Consortium* och är ett industrikonsortium grundat år 1994, där över femhundra medlemmar från ledande industrier, forsknings- och utvecklingsinstitut, standardiseringsorgan och regeringar samt EU, finansierar verksamheten tillsammans med statliga bidrag (Wikipedia: *World Wide Web Consortium*). W3C arbetar primärt med att utveckla nya webbstandarder, tekniska protokoll samt programvaror för webben. Exempel på kända standarder som W3C har utvecklat är följande (W3C, 2010):

- **CSS** – *Cascading Style Sheets*
- **HTML** – *HyperText Markup Language*
- **RDF** – *Resource Description Language*
- **SOAP** – *Simple Object Access Protocol*
- **XML** – *eXtensible Markup Language*
- **XHTML** – *eXtensible HyperText Markup Language*
- **WSDL** – *Web Services Description Language*
- **UDDI** – *Universal Description, Discovery and Integration*

Chef för konsortiumet är sedan starten Tim Berners-Lee, och det är även han tillsammans med W3C som bedriver projektet om och för den semantiska webben.

3.4. Web 2.0 i ett nötskal

Enligt (Wikipedia: *Web 2.0*) myntades för första gången "Web 2.0" som begrepp av (DiNucci, 1999) i artikeln "*Fragmented Future*". I artikeln beskrivs begreppet primärt ur ett webbdesignperspektiv där informationen först och främst är riktad mot webbdesigner som målgrupp. (Vossen & Hagemann, 2007) menar dock på att det var (O'Reilly, 2005) som myntade begreppet först under en webbkonferens med O'Reilly Media.

Web 2.0 kan enligt O'Reilly's definition sammanfattas genom följande i förhållande till en webbsida, och dess krav för att webbsidan skall vara av Web 2.0-karaktär:

- **Större användardeltagande** – Användaren skall själv kunna vara med och bidra till en webbsidas innehåll.
- **Större kontroll från användaren** – Den här principen återger tanken att användaren skall ha större kontroll över sin information på en webbsida.
- **Rikare applikationer** – Går den här principen ut på att webbdesignen för hemsidor skall vara så fylliga, interaktiva och användarvänliga som möjligt.

(Vossen & Hagemann, 2007) har sammanfattat följande koncept, utvecklingsteknologier samt webbapplikationer som en del av Web 2.0:

- **Google Search**
- **Google Map**
- **Asynchronous Javascript And XML (AJAX)**
- **Mash Ups**

Författarna poängterar betydelsen av rik-interaktion hos webbapplikationer som en förutsättning för en större grad av användarvänlighet. En viktig utvecklingsteknologi för att uppnå detta, är punkt nummer åtta i listan ovan, nämligen *AJAX*.

3.4.1. Vad är *AJAX*?

AJAX står för *Asynchronous Javascript And XML*. Termen myntades år 2005 av Jesse James Garrett, och är ett samlingsnamn för olika tekniker som främjar för rikare webbapplikationer. Wikipedia (*AJAX*, 2010) beskriver samtliga teknikers roll som följande:

- ***XMLHttpRequest*** – Är det objekt som tillåter JavaScript på en sida att göra asynkrona anrop till en webbserver. Detta innebär att anropen sker mot en webbserver utan att sidan laddas om.
- ***DOM (Document Object Model)*** – Har som roll att tillåta JavaScript att manipulera element och innehåll på en webbsida, för att skapa dynamiska interaktionsmedel på webbsidan.
- ***XHTML*** – Beskriver tillsammans med CSS utseende på webbsidan.
- ***XML*** – Används för utbyte, manipulation samt för att visa data på webbsidan.

Ballard (2006) beskriver också att *AJAX* upprättar ett lager mellan webbsidan och servern. Detta lager kallas oftast för *Ajax-motor* och eller *Ajax-ramverk*, och har till syfte att fånga upp användarens förfrågningar och hanterar kommunikationen med servern både asynkront och diskret. Kända applikationer som tillämpar den här teknologin är *Google Mail*, *Youtube*, *Facebook* samt *Google Search*.

Resonemanget ovan om Web 2.0, dess sammanhängande komponenter och *AJAX* är vad (Vossen & Hagemann, 2007) kallar för ett potentiellt bidragande till något nytt och delvis revolutionerande över webben, nämligen *Den Semantiska Webben*.

3.5. Den semantiska webben

3.5.1. Vad är Semantik?

Första definitionen samt citatet är hämtad från Nyblom & Torrkulla (2009).

”(av gr. *Semainein*, betyda, beteckna, ytterst av sema tecken, signal), studiet av språkliga uttrycks betydelseaspekter, i motsats till deras syntax, dvs. den strukturella uppbyggnaden, och deras pragmatik, dvs. drag hos uttryckens användning i olika kontexter.”

Det andra exemplet samt citatet är hämtad från Bydén & Ebbesen (2000) där de översätter den grekiske filosofen Aristoteles och hans syn på begreppet *Semantik* .

”Ett namn är en talenhet som fungerar som tecken enligt överenskommelse, utan bibetydelse av tid, varav ingen del fungerar som tecken då den tas separat. I namnet ”Fagerlund” betecknar ju ”lund” ingenting i sig självt, vilket samma talenhet gör i satsen: det är en fager lund. Men det är inte heller så att det som är fallet i fråga om enkla namn också är fallet i fråga om sammansatta namn: i de förra fungerar delarna inte på något sätt som tecken, i de senare har de visserligen denna bestämmelse, men betecknar faktiskt ingenting då de tas separat, till exempel ”-kryssare” i ”slup-kryssare”.

Det tredje samt sista exemplet av hur semantik kan definieras är hämtad från Wikipedia (*Semantics*, 2010).

”In computer science, where it is considered as an application of mathematical logic, formal semantics of programming languages reflects the meaning of programs or functions. In this regard, semantics permits programs to be separated into their syntactical part (grammatical structure) and their semantic part (meaning)... the Semantic Web refers to the extension of the World Wide Web through embedding of additional semantic metadata”.

Begreppet definieras generellt som läran om språkliga betydelser (Nationalencyklopedin, 2010). Exempel på detta kan vara begreppet ”Ford” och dess betydelse. Gustavsson (2003) menar på att samma begrepp kan användas som efternamn samt som bilmärke, men tolkningen av det kan arta sig olika beroendes på kontexten som begreppet används i. Ett exempel på detta är webben och semantikens roll gentemot informationshantering av webbapplikationer.

3.5.2. Vad är den semantiska webben för något?

Följande citat är hämtad från Vossen & Hagemann (2007), där de citerar Berners-Lee et al. (2001):

"...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. 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."

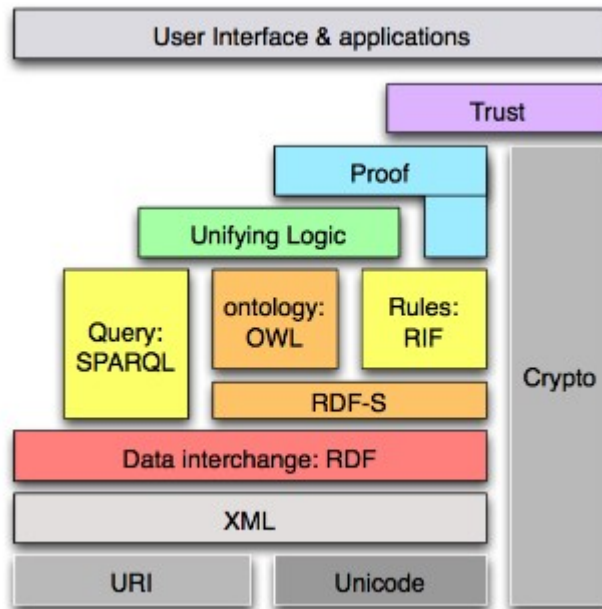
Förutom föregående definition, citerar (Bakke, 2007) två stycken definitioner om den semantiska webben och dess primära roll som följande:

"... to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data. Facilities to put machine-understandable data on the Web are quickly becoming a high priority for many organizations, individuals and communities."

The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently. The Semantic Web Activity is an initiative of the World Wide Web Consortium (W3C) designed to provide a leadership role in defining this Web. The Activity develops open specifications for those technologies that are ready for large scale deployment, and identifies, through open source advanced development, the infrastructure components that will be necessary to scale in the Web in the future." (W3C, 2001-2007c)

Ovanstående citater är än så länge generella och abstrakta, därför behövs det en djupare studie om själva begreppet och dess sammanhängande komponenter.

Nedanstående bild illustrerar en konceptuell bild av hur den semantiska webbens grundläggande arkitektur är uppbyggd. Bilden är hämtad från (W3C, 2006):



För att konkretisera modellen ett steg närmare så beskriver (Vossen & Hagemann, 2007) arkitekturen steg för steg som den semantiska webbens grundläggande struktur:

- **XML** tillhandahåller en elementär syntax för innehållsstrukturen i dokument, dock utan att associera någon semantik till innehållet. Detta innebär att **XML** integreras tillsammans med komponenterna **URI** och **Unicode** i strukturen, för att strukturera, representera samt referera korrekt data för webbapplikationer.
- Ovanpå **XML** i arkitekturen finner man **RDF**, vilket enligt författarna är den initierande fasen för den semantiska webben. **RDF** är en datamodell och en generell metod för att uttrycka information
- **RDFS** används i det här sammanhanget för att beskriva attribut, egenskaper och klasser tillhörandes en resurs som återfinns i **RDF**-format.
- **OWL** och **Ontologier** har den mest centrala rollen inom den semantiska webbens grundläggande struktur. **Ontologier** kan i den här kontexten användas till att: beskriva en specifik sektion av verkligheten tillsammans med all dess innehåll, representera en terminologi eller enhetligföra data från semantiska webbmiljöer. **OWL** används som det språk med högst uttryckbara vokabulär till att beskriva egenskaper och klasser.

- **SPARQL** står för **SPARQL Protocol and RDF Query Language** och används till att ställa frågor till databaser som tillhandahåller av *RDF*-data.
- **RIF** är en förkortning på *Rules Interchange Format* och anses enligt (Wikipedia: *RIF*) vara en standard som fortfarande är under utveckling inom den semantiska webben hos W3C. *RIF* är i konkreta termer ett format för utbyte av regler i regelbaserade system i den semantiska webben. En regel kan härledas till vanliga "IF", "OR", "THEN"-satsar där respektive sats skapar ett villkor kring hur kod sedermera bör exekveras.
- För att specificera meningen med de termer och komponenter som används i en ontologi, behövs det en logisk teorimodell. Den generella semantiken definieras i en ontologi, genom det som kallas för *Unified Logic* i den semantiska webbstrukturen. Modellen har till syfte att skapa logiska regler för hur semantiken skall byggas upp i en ontologi, samt för att skapa en korrektare struktur kring det data som en ontologi beskriver.
- Den komponent i strukturen vilket representeras av begreppet *Trust*, har som termen antyder till syfte att upprätta pålitligheten kring den informationen som den semantiska webben skall behandla över webben. Först och främst ur ett användarperspektiv, där användaren skall lita på informationen som han/hon interagerar med.
- Sist men inte minst behöver respektive webbapplikation ett användargränssnitt. Än så länge är det dock inte avgränsat till vilken typ av teknologi som skall användas för det här syftet. Detta lagret kommer att vara beroendes på vad det är för applikationen.

Vossen & Hagemann (2007) beskriver ovanstående punkter som nödvändiga komponenter inom ramen för den semantiska webben. Dock så beskriver även författarna att det är nödvändigt med en explicit syntax och semantik för att en optimalare informationsbearbetning skall kunna fungera över webben. Effekten av ett sådant åstadkommande kopplas till frågan om hur sökmotorer skall kunna optimeras med hjälp av semantiska webbprinciper. Ett exempel på detta är: hur skall bearbetningen av metadata optimeras i förhållande till sökresultat? Eller hur skall relevansen och precisionen kring söktermer och sökresultat effektiviseras? Det här resonemanget leder studien vidare till en viktig del av den semantiska webben, nämligen definitionen av *metadata*.

3.5.2.1. Metadata

Rent generellt beskriver (Wikipedia: *Metadata*, 2010) *metadata* som data om data, eller information om data. Med andra ord används metadata för att beskriva innehållet och/eller strukturen för en viss datarepresentation och samling. Exempel på områden där metadata kan användas är:

- **HTML-filer** – innehåller metadata för att beskriva hemsidan ur ett internt perspektiv genom så kallade "meta-taggar". "meta-taggar" används tillsammans med HTML-dokument för att beskriva meta-information om själva dokumenten. Med andra ord har meta-taggar till syfte att beskriva information om den informationen som finns lagrad i HTML-dokumentet (Vossen & Hagemann, 2007).
- **Sökmotorer** – Som bland annat använder metadata tillhörande HTML-filer för att ge högre relevans kring de sökträffar som sökmotorn genererar.

Metadata kan med andra ord användas som attribut för en resurs vilket gör den unik. Detta kan vara resurser som hemsidor, filmer, ljudfiler och bilder.

3.5.2.2. Vad är Ontologier?

Begreppet *Ontologi* definieras ur ett filosofiskt perspektiv som: "Läran om det varande" (Wikipedia: *Ontologi*, 2010). Begreppet är ursprungligen hämtat från den grekiske filosofen Aristoteles då han letade efter ett sätt att klassificera alla entiteter i kategorier. Därav blev resultatet begreppet: "Onta" (vara) och "Logos" (vetenskap), det vill säga *ontologi*.

När det gäller att beskriva entiteter, dess egenskaper och relationer med övriga entiteter, har ontologier samma funktion inom filosofin som de har inom informatik (Wikipedia: *Ontology (Information Science)*, 2010).

Ontologier används inom informatik för att upprätta ett homogent språk för olika typer av informationssystem (Vossen & Hagemann, 2007). Detta beskriver Gustavsson (2003) som en metod för att reducera kommunikationssvårigheter mellan olika typer av webbapplikationer. Detta är också ontologiers primära roll i förhållande till den semantiska webben. (För mer om ontologier och dess syfte i den semantiska webben; läs avsnitt [3.5.2.7](#))

3.5.2.3. Semantiska webbens grundläggande språk

Enligt Allemang & Hendler (2008) följer de språk vilket den semantiska webben bygger på, en hierarki där det minst uttryckbara språket befinner sig högst upp och det mest uttryckbara befinner sig längst ned. Med "uttryckbara" menar författarna språkens förmåga att kunna beskriva semantiken av den informationen som de behandlar:

- **XML**
- **RDF**
- **RDFS**
- **OWL**

Därför har jag valt att definiera samt förklara respektive begrepp utifrån ovanstående ordning.

3.5.2.4. Vad är XML för något?

XML står för *Extensible Markup Language* och är enligt Vossen & Hagemann (2007) en W3C-rekommendation samt ett subset av språket *SGML* (*Standard Generalized Markup Language*). *XML* används bland annat till bättre utbyte och integrering av data i informationssystem. Språket används även inom webbsammanhang då en utökning av *HTML* vid beteckningen *XHTML* är baserad på *XML*. På så vis används även *XML* till att beskriva webbsidor, samt att utväxla information i olika system, vanligast i affärssystem. Skillnaden mellan *XML* och *HTML* är att det sistnämnda används till att åskådliggöra och presentera data, medan *XML* används för att strukturera och organisera data/information (Wikipedia: *XML*, 2010).

3.5.2.5. Vad är *RDF* för något?

(Wikipedia: *RDF*) beskriver *RDF* som en samling tekniker som är framtagna av W3C. Samma källa beskriver även *RDF* som en datamodell med en ursprunglig mening till att uttrycka metadata samt information.

(Vossen & Hagemann, 2007) beskriver *RDF* utifrån följande tre kärnaspekter:

- **Resurser** – En resurs i den här kontexten kan vara ett subjekt eller objekt, vilket exempelvis skulle kunna vara en bildfil. En resurs kan även beskrivas som ett abstrakt koncept. Exempel på detta kan vara ett företag. Det som dock är ekvivalent för samtliga två typer av resurser, är att bägge två identifieras unikt genom en så kallad *URI* (Uniform Resource Identifier). En *URI* är en sträng av tecken som används primärt i webbsammanhang för att kunna unikt identifiera resurser. En *URL* (Uniform Resource Locator) är dock den faktiska länken som representerar en hemsida.

En *URL* är med andra ord en *URI*, som förutom att identifiera en resurs även beskriver hur man når resursen och var den på webben finns.

- **Egenskaper** – Varje resurs i ett *semantiskt nätverk* består av en till flera egenskaper med ett värde som beskriver resursen som sådan. Exempelvis så kan resursen "artist" bestå av egenskapen "namn" och dess värde vara "John Lennon".
- **Påståenden** – Påståenden används för att kombinera en egenskap hos resursen med ett värde. Detta görs som enklast genom en så kallad "*Subjekt-Predikat-Objekt*"-relation, eller i andra termer *Triples*. Exempel på detta kan vara:
 - "En fågel (Subjekt), har (Predikat), vingar (Objekt)"
 - "Ett djur (Subjekt), äter (Predikat), mat (Objekt)"

En viktig aspekt kring resonemanget ovan är att, varje subjekt och objekt identifieras unikt genom en *URI*, medans predikatet istället länkar ihop subjektet med objektet.

Ett exempel på ett unikt *URI* på nätet är *URL*:en: <http://sv.wikipedia.org/>, som symboliserar wikipedias startsida på svenska.

3.5.2.6. Vad är *RDFS* för något?

RDFS står för *Resource Description Framework Schema*. Dess primära roll och syfte i förhållande till den semantiska webben är enligt (Allemang & Hendler, 2008) att schematisera regler för hur data och information skall representeras i *RDF*-format. Den här företeelsen förekommer även i *XML* där ett *XML*-schema används för att beskriva strukturen av ett *XML*-dokument. (Allemang & Hendler, 2008) menar även på att *RDFS* skapar en större mening kring det data som representeras i *RDF*, och att det sker genom märkspråk med en ekvivalent syntax som används i *RDF*, och som även är utifrån dess strukturella aspekt, analogt med syntaxen i *XML*. Detta innebär att *RDFS* har samma funktion gentemot *RDF*-dokument som *XMLS* har gentemot *XML*-dokument.

3.5.2.7. Vad är *OWL* för något?

OWL (*The Web Ontology Language*) utvecklades av *W3C* Ontology Working Group och används primärt som ett språk för att specificera *ontologier*. *Ontologier* beskriver (Vossen & Hagemann, 2007) ur ett tekniskt perspektiv som en samling av termer eller koncept som är associerade med verkliga entiteter. Ett exempel på detta kan vara följande tes: "En fågel är ett djur", eller "En fågel har vingar". I det första resonemanget generaliserar man tesen genom att låta en fågel tillhöra djurkategorin.

Detta kallas för en *generaliseringsrelation*, där meningen: "är ett" representerar ontologierna. I det andra resonemanget representerar tesen: "har" en ontologi, och där "vingar" ses som en egenskap till subjektet "fågel".

Idag existerar det tre stycken subspråk till *OWL*, som tillsammans beskriver språkets funktionalitet:

- ***OWL Lite*** – är det mest grundläggande språket och har till syfte att definiera klasshierarkier samt enkla mekanismer för att beskriva begränsningen i den kardinalitet som existerar mellan klasserna.
- ***OWL DL*** – är en utökning av *OWL Lite*, och baseras på *Description Logics (DL)*. *DL* är enligt (Wikipedia: *Description Logics*) viktigt för den semantiska webben på grund av dess logiska formalism som den bidrar med till ontologier.
- ***OWL Full*** – till skillnad från de *OWL Lite* och *OWL DL* så har *OWL Full* en viss kompatibilitet med *RDFS* samtidigt som den anses innehålla samtliga föregående två språkens möjligheter.

3.6. Sökmotorer

Sökmotorer är enligt Fransson (2007) den mest använda typen av söktjänster. Att söka information via Google har till och med blivit ett verb – ”att googla”. Men vad är egentligen en sökmotor för något, och hur ser anatomin för en sökmotor ut? Nedanstående underrubriker kommer att behandla vad en sökmotor är för något samt hur den generella konstruktionen ser.

3.6.1. Vad är en sökmotor?

”En sökmotor är en söktjänst som automatiskt indexerar webbsidor och andra filtyper i viss utsträckning. Sökmotorer är datorprogram som läser webbsidor på Internet och sparar ned dem i en databas. När du söker i en sökmotor så söker den i sin databas efter relevant material som den sedan presenterar för dig” (Fransson, 2007).

3.6.2. Hur fungerar en sökmotor?

Enligt Fransson (2007) och Vossen & Hagemann (2007) består en sökmotor av tre stycken nyckelaspekter:

- **Spindeln** – Hittar och samlar in webbsidor.
- **Indexeraren** – Indexerar och lagrar webbsidor.
- **Frågebehandlaren** – Genererar sökresultat

3.6.2.1. Spindeln

En ”spindel” i den här kontexten är ett datorprogram som bläddrar igenom webben utifrån en given algoritm (Wikipedia: *WebCrawler*, 2010). Med andra ord har spindeln som uppgift att utforska webben samt konstant kopiera sidor från webben, som den sedan levererar vidare till indexeraren för analys. Spindeln kan ladda hem över tusentals sidor samtidigt, och inte enbart en åt gången, vilket de flesta webbläsare gör för användaren. En annan viktig aspekt enligt Fransson (2007) är att spindeln inte enbart hittar nya webbsidor, utan den återbesöker också regelbundet de sidor som redan finns i sökmotorns index för att hålla sin databas uppdaterad.

3.6.2.2. Indexeraren

Indexeraren extraherar termer och begrepp från de webbsidor som spindeln levererar till den och därefter lagrar den dessa i en databas (Vossen & Hagemann, 2007). I indexet lagras på vilken webbsida ordet återfinns samt vart i sidan de står, oavsett om det handlar om brödtext, rubriker eller länkar som lagras. (Fransson, 2007) jämför indexet för en sökmotor med indexet som finns längst bak i en bok, där varje ord hänvisar till de sidor där de återkommer. Vanligtvis brukar begrepp och termer indexeras tillsammans med den URL (Uniform Resource Locator), vilket är länken till den webbsida som orden återfinns i.

3.6.2.3. Frågebehandlaren

Frågebehandlaren är den faktiska delen av sökmotorn som användaren interagerar med. Den behandlar sökfrågan (användarens sökord), beräknar relevans och sedan presenterar resultaten, vilket i många fall är en lista med potentiella sidor relaterade till sökordet (Fransson, 2007). Ett exempel på en frågebehandlare och sökmotor är *Google Search* (Vossen & Hagemann, 2007), som använder AJAX för att leverera potentiella sökord, beroendes på hur användaren inleder sitt sökargument i sökfältet (Ballard, 2006). Därefter väljer användaren det sökord som han/hon finner lämpligast, vilket i sin tur genererar en lista med sökträffar.

3.7. Search 2.0

Ezzy (2006) jämför sökmotorer av karaktären *Search 2.0* med traditionella sökmotorer genom att kategorisera dem enligt följande:

1. **Generation 1** – Sökmotorer rankar sidor genom innehållet av en webbsida. Exempel på detta är sökmotorer som: *Altavista* samt tidiga *yahoo.com*
2. **Generation 2** – Smartare beteenden hos sökmotorer utvecklas för att analysera samt ranka hemsidor beroendes av antalet länkar från, och till en specifik webbsida. På så vis börjar sökmotorerna ta hänsyn till webbstrukturen, och exempel på söktjänster som tillämpar detta är: *Google Search* samt *Overture*.
3. **Generation 3** – Sökmotorer av den tredje generationen kombinerar skalbarheten av andra generationens sökmotorer med nya samt förbättrade modeller som behandlar sökrelevansen i en sökträff. Detta görs genom tillämpade Web 2.0-principer: kollaborering, användarens personliga preferenser, smartare användning av den kollektiva intelligensen över webben, rikare användargränssnitt samt övrig möjligheter som gör informationen över webben mer produktiv.

3.7.1. Google Search

Enligt (Wikipedia: *Google Search*, 2010) är Google's sökmotor den mest besökta platsen på hela Internet. *Google Search* räknas som en Search 2.0-applikation och består i huvudsak av följande tekniska aspekter (Vossen & Hagemann, 2007):

- **PageRank** – Är en algoritm för att mäta relativt informationsvärde på webbsidor. Algoritmen utgör grundmetoden för hur sökmotorn väger olika webbplatsers informationsvärde mot varandra.
- **Googlebot** – Är den faktiska sökroboten som används för att hämta dokument från Webben, för att bygga upp en databas åt *Google Search*. Sökroboten finns i två olika versioner:
 - **Freshbot** – Den här versionen letar efter nytt material på webben genom att besöka enskilda webbsidor beroende på hur mycket av sidans innehåll som förändras.
 - **Deepbot** – Följer samtliga länkar som den hittar och laddar ned så mycket den kan till Googles indexerare.
- **Google Suggest** – Implementerar AJAX och är den faktiska funktionen som gör att sökmotorn ger förslag på sökord utifrån ett få tal bokstäver. Funktionen visar även tillsammans med sökförslagen hur många resultat man kommer att få av sökningen.

3.8. Sökmotorer i den semantiska webben

Följande användarscenario som beskrivs är hämtad från Vossen & Hagemann (2007) där de genom att referera till (Pellegrini & Blumauer, 2006) vill framhäva vilka problem som en användare stöter på idag när han/hon skall söka efter information på Internet, samt hur scenariot skulle kunna optimeras:

- **Sökscenario 1:** En kvinna vid namnet Ms. Maler söker efter en homeopatisk läkare närmast hennes hemstad, vilket i det här fallet är staden Graz i Österrike. Hon inleder sökbefarandet genom att mata in söktermerna: "doctor", "homeopathic" och "Graz" i sökfältet tillhörandes hennes mest uppskattade sökmotor, och får en lång lista med länkar som sökträff. Därefter granskar Ms. Maler igenom listan efter en läkare med den expertis som hon söker, de besökstider som ser bäst ut i förhållande till hennes schema samt ifall läkaren är nåbar genom offentlig transport. Det sistnämnda tar hon reda på genom att konsultera en stadskarta och för att på så vis stämma av samtliga läkares adresser. Slutligen granskar hon de utvärderingar som har gjorts gentemot läkarna, och efter tjugo minuter återstår det enbart tre stycken läkare som hon finner tillfredsställande gentemot hennes behov och krav.
- **Sökscenario 2:** Det här scenariot behandlar frågan ifall sökscenario 1 hade kunnat implementeras effektivare samt optimalare i förhållande till Ms. Maler och hennes sökintentioner. Svaret är enligt (Vossen & Hagemann, 2007) ja, genom att tillämpa semantiska webbprinciper som innehåller intelligentare metoder för att lokalisera samt precisera sökträffar gentemot söktermerna. I det här fallet så skulle sökmotorn ha kunnat lokalisera sökträffar vars fokus ligger på läkare som bor i staden Graz tillsammans med deras specifika adressuppgifter, vilket sedermera en karttjänst hade kunnat behandla för att visualisera exakt vart på kartan respektive läkare befinner sig. Från Ms. Malers sida hade det varit tidseffektivare samt intuitivare att använda en karttjänst integrerat med information som behandlar hennes sökfrågor, en så kallad *Mash-Up* som integrerar information från olika tjänster.

3.8.1. Semantiska sökprinciper

Semantiska sökprinciper har enligt Vossen & Hagemann (2007) till syfte att återge korrektare sökträffar i förhållande till angivna söktermer. Nuvarande sökmotorer indexerar HTML-sidor och återger sökträffar som kan upplevas som totalt irrelevanta gentemot användarens intentioner. Det här problemet kommer enligt Vossen & Hagemann (2007) att reduceras genom en implementation av semantiska sökprinciper.

Principerna har till syfte att optimera följande aspekter av en sökmotors beteende:

- **Precision** – Återge relevantare sökträffar, genom att identifiera samtliga resurser genom dess unika *URI*. Exempel på detta kan vara namnet "Benny", där namnet i programmeringstermer anses vara av strängtyp (Gustavsson, 2003). I en semantisk kontext bör hellre "Benny" identifieras genom ett *URI*: <http://server.com/benny>. På så vis blir namnet unikt över webben till skillnad från själva namnbeteckningen "Benny". Detta gör att sökmotorer kan återge relevantare sökträffar i förhållande till söktermer, beroendes på hur söktermerna artar sig. Ställer man exempelvis sökfrågan: "Vem är Benny Gustavsson som skrev artikeln "Den semantiska webben"?", så skall respektive del i frågan analyseras och återge sökträffar som bygger på frågans olika *URI*'er. Dessa *URI*'er kopplas ihop och tolkas av sökmotorn som *Triples*: *URI.1* (Benny Gustavsson), *URI.2* (Skrev), *URI.3* (Den semantiska webben). Detta är också det som kallas för *Metadata* i en semantisk kontext. Där information beskrivs av annan information, i det här fallet information som identifieras av dess unika *URI*.
- **Relevans** – Ifall precisionen kring sökträffarna ökar, så ökar även relevansen gentemot användarens sökintentioner. Detta förklarar (Vossen & Hagemann, 2007) som bättre integration av data samt metadata. Där sökmotorer skall kunna avgöra sökresultatets relevans genom den explicita beskrivningen som en webbsidas metadata beskriver om informationsinnehållet.

Nedan följer det information som har till syfte att beskriva de semantiska sökmotorerna i testet.

3.8.2. Swoogle

Enligt (Wikipedia: *Swoogle*) är *Swoogle* en icke-kommersiell sökmotor för den semantiska webben som biquity Research Group (Swoogle, 2004) på University of Maryland, Baltimore County har utvecklat för att behandla semantiska webbontologier, dokument, termer samt data som finns publicerade över webben. Detta görs genom ett system av *spindlar* vilket *Swoogle* integrerar för att hitta *RDF*-dokument och *HTML*-dokument med ett inbäddat *RDF*-innehåll (Finin et al. 2005). Enligt författarna består *Swoogle*'s arkitektur av följande fyra komponenter:

1. **Discovery** – Den här komponenten hittar samt återbesöker semantiska webbdokument (SWD) per automatik genom ett antal integrerade *spindlar*.
2. **Digest** – Nästa steg är att beräkna metadata för respektive dokument samt term som är av den semantiska webbkaraktären. Den här komponenten gör detta, samt identifierar relationen mellan dokumenten och termerna.
3. **Analysis** – Analyskomponenten använder cachade SWDs och dess metadata för att erhålla analytiska rapporter. Dessa rapporter kan behandla systematiserade

ontologier bland SWD samt rankningen av dokumenten beroendes av och på deras betydelse.

4. **Service** – Den sista komponenten i arkitekturen stödjer både mänskliga samt digitala mjukvaruagenter genom konventionella webbgränssnitt samt SOAP-baserade webbtjänsters *API*. Servicekomponenten består av två nyckelaspekter:
 - a. **Swoogle search** – En tjänst som söker efter semantiska webbdokument genom en restriktion på deras *URL*, dvs. länkadresser. Tjänsten söker även efter webbsidor som är värdar för dokumenten, samt de klasser och egenskaper som används eller är definierade av dokumenten.
 - b. **Ontology dictionary** – Är en tjänst som söker efter semantiska webbtermer (SWT) samt deras relation gentemot övriga SWT och SWD.

Ett typiskt användarscenario kan vara följande vid interaktion med *Swoogle*:

1. Användaren väljer ifall han/hon vill söka på kategorin:
 - a. **Ontology** – Söker efter en mindre samling webbdokument som endast består av semantiska webbontologier.
 - b. **Document** – Söker efter alla semantiska webbdokument.
2. Användaren anger söktermen "mat" i sökfältet och klickar på "Search".
3. Swoogle analyserar samt tolkar sökbegreppet genom att indexera de webbdokument vars semantik återspeglar söktermen bäst. Semantiken kan i det här fallet vara webbdokumentens unika URI eller övrig metadata som beskriver webbdokumentet.
4. Sökträffarna sammanställs i listformat, där användaren själv får sortera efter "datum" eller "triples".

3.8.3. Hakia

Hakia tillämpar semantiska webbprinciper genom att tillhandahålla samt möjliggöra för användaren att ställa frågor, eller skriva fullständiga meningar i sökfältet. När sökningen initieras använder sig sökmotorn av två mekanismer vid beteckningen *Semantic Rank Algorithm* samt *QDEXing* vilket i sig är tillämpningar av och på semantiska webbprinciper (Vossen & Hagemann, 2007). Algoritmen har till syfte i sökkontexten att förfina resultat, samt framförallt framhäva innebörden kring sökresultaten baserat på innehållet av söktermen. Ett typiskt användarscenario kan vara följande vid interaktion med *Hakia*:

1. Användaren anger söktermen "semantic web" i Hakia's sökfält och klickar på "search"

2. Hakia's sökmotor behandlar söktermen och kategoriserar sökträffarna utifrån följande kategorier
 - a. **Wikipedia** – Sökresultatet bygger på artiklar som primärt behandlar själva söktermen i sig.
 - b. **Web** – Prioriterade källor över webben som behandlar söktermen som ämnestema. Detta innebär att Hakia's rankningsmekanism prioriterar källorna utifrån de kriterier som mekanismen själv sätter upp och prioriterar.
 - c. **Credible** – Den här kategorin behandlar sökträffar som anses av Hakia själva vara de mest relevanta i förhållande till söktermen. Detta innebär i praktiken att källor som behandlar termen i en längre mening, eller artiklar som omfattar en större omfattning kring termen används som primära källor för sökresultatet.
 - d. **Pubmed** – En avgränsning till information som huvudsakligen behandlas inom medicinska discipliner.
 - e. **News** – Som titeln själv påpekar så behandlar den här kategorin information som återges bland nyhetssidor över webben.
 - f. **Blogs** – Diverse olika typer av bloggar som diskuterar frågor om hur och vad den semantiska webben är för någonting, samt hur den kan användas för att förbättra webben.
 - g. **Twitter** – Kategorin återger sökträffar som är länkade till den sociala nätverkssidan *Twitter*, där söktermen oftast är bifogad tillsammans med övrig personrelaterad information.
 - h. **Images** – Behandlar huvudsakligen bilder över webben som är relaterad till söktermen.
 - i. **Video** – Diverse olika videoklipp som behandlar information om den semantiska webben återges under den här kategorin.
 - j. **Gallery** – Återger en galleri fylld med underkategorier rörandes olika element av söktermen. Exempelvis så strukturerar galleriet resultatet kring söksträngen "Beatles" efter underkategorier som: "Mest kända låt", "Populäraste hitlåt", "Bästa plattan", "Pressbilder", "Fan-bloggar" etc.

Ovanstående sökscenariot visar att Hakia kategoriserar sina sökträffar efter kategorier. Samtliga kategorier använder sig dock av två olika mekanismer för att behandla samt återge information som är relaterad till söktermen. Under följer en noggrannare granskning kring dessa mekanismers grundläggande koncept och syfte (Hakia, 2010):

- **QDEX** – *QDEX* står för *Query Detection and Extraction* och har till uppgift att analysera samt lagra innehåll från olika webbsidor i form av kunskap. Detta görs genom att *QDEX* analyserar en webbsidas totala innehåll, och sedan använder en algoritm som extraherar alla möjliga användarfrågor, dvs. söktermer som kan vara kopplade till innehållet. Söktermerna kan vara olika beroendes av längden på söksträngen samt söktermens karaktär. Fördelen med den här metoden är att data som är av semantisk webbkaraktär blir flexiblare att hantera samt återge i

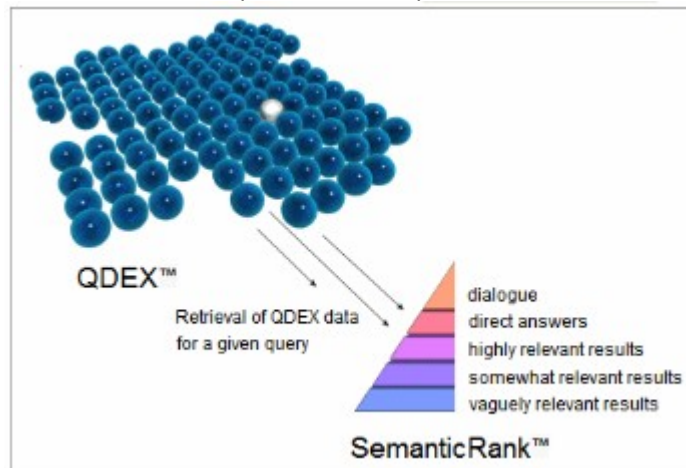
form av sökresultat. Kritiskt är dock själva tillvägagångssättet som *QDEX* använder för att analysera samt återge meningsfulla meningssekvenser. Exempelvis så kan en mening beståendes av åtta stycken signifikanta ord, återge över 40 000 olika sekvenser där enbart ett fåtal upplevs som vettiga av användaren. Den här problematiken löser dock Hafia genom följande mekanism.

- **Commercial Ontology (CO)** – CO kan metaforiskt beskrivas som en byggnad bestående av två våningar (Hafia, 2010), där den första våningen består av en karta med objekt (substantiv) som inkluderar konkreta saker så som substantiv. Kartan innebär att givna ord, meningar, fraser eller text identifieras genom de element som bestämmer vart de angivna sökorden hör hemma i förhållande till substantiven.

Den andra våningen består av en karta som behandlar händelser (verb). Denna hierarki innehåller fragment, grupper samt kluster som nödvändigtvis inte hör ihop. Istället är vissa typer av händelser kopplade med den första våningen där de hör ihop med vissa typer av substantiv. Exempel på detta kan vara händelsen "testing" i vidare bemärkelse "stress testing" vilket hör ihop med materialen från första våningen.

Dock så kan även den här händelsen höra ihop med den andra våningen, på grund av att konceptet är tillämpligt gentemot "det mänskliga kroppssystemet". Resonemanget kring våningarna leder slutsatsen till att CO används för att lokalisera till vilken våning en söksträng hör hemma hos. Och i det här fallet så kan det både finnas inslag av söktermen på våning ett och två.

- **Semantic Rank** – *Semantic Rank* är en algoritm som har till syfte att ranka sökresultaten utifrån dess relevans i förhållande till den angivna söktermen. Följande bild har till syfte att illustrera förhållandet mellan *QDEX* och *Semantic Rank*. Bilden är hämtad från (Hafia, 2010):



Bilden visar hur bägge koncepten interagerar med varandra, där *QDEX* levererar data utifrån den givna söktermen, och där *Semantic Rank* avgör den slutgiltiga relevansen genom att analysera meningen för att matcha konceptet kring söktermen med mest relevanta data som *QDEX* levererar.

Hakia kategoriserar varje sökträff som den genererar i separata kategorier. Nedan visas det vilka kategorier som *QDEX* och *Semantic Rank* underhåller:

- **QDEX**
 - a. *News*
 - b. *Blogs*
 - c. *Credible*
 - d. *Hakia Galleries*

- **Semantic Rank**
 - a. *Web*
 - b. *Video*
 - c. *Images*

Vilken sökträff som hamnar i vilken kategori avgörs genom sökträffens karaktäristika (Hakia, 2010)

3.8.4. Lexxe

Enligt Radhakrishnan (2009) använder *Lexxe* lingvistiska metoder för att behandla söktermer. *Lexxe* (2005) själva beskriver sökmotorn som en sökmotor för den tredje generationens Internet, där den i första hand fokuserar sig på att behandla sökfrågor istället för söktermer.

Sökmotorn behandlar sökfrågorna på följande vis (Lexxe, 2005):

- **Handling Questions** – Sökmotorn hanterar sökfrågorna som bäst när frågan består av tio ord, eller mindre. Framförallt genererar den explicita svar på frågor som består av korta svar. Exempelvis: "Who is Tony Blair?".
- **Key Word Query Processing** – Ifall sökfrågan inte konstrueras som en fråga men istället som en sökfras, behandlar *Lexxe* frasens olika begrepp och genererar träffar som återger begrepp ur frasen. Exempelvis frasen: "the lord of the rings assistant director". Ur den här frasen genererar sökmotorn träffar som matchar meningarna: "the lord of the rings" och "assistant director".
- **Presenting the results** – Sökmotorn kategoriserar sökträffarna under följande kategorier:
 - **Answers** – Explicita svar på sökfrågor presenteras i den här kategorin.
 - **Clusters** – Sökträffar med en stark koppling till deras semantik samlas i den här kategorin. Om användaren exempelvis anger söktermen

”Thunderbird”, behandlar och samlar sökmotorn träffarna i ett kluster med träffar vars semantik är relaterade till söktermen.

- **Web Page Snippets** – I den här kategorin genererar sökmotorn de träffar som är bäst rankade och mest relevanta gentemot sökfrågan

3.8.5. Kosmix

Kosmix (2008) beskriver sökmotorn och dess funktion som en guide till webben, genom att skapa en röd tråd mellan sökträffarna. Konkretare sagt fokuserar sig *Kosmix* på att generera sökträffar som beskriver information som handlar om användarens angivna sökord. Informationen som *Kosmix* genererar, kategoriserar sökmotorn i olika kategorier beroende av söktermens karaktäristika. Exempel på kategorier är: **Videos, Photos** och **News**. Och det är just kategorierna som skapar den röda tråden mellan träffarna.

3.8.6. Cognition Search

Cognition Search kombinerar lingvistiska algoritmer med semantiska representationer för att beskriva vad sökord, eller meningar betyder (Cognition Search, 2008). Med andra vill man beskriva semantiken av de sökbegrepp eller meningar som användaren anger när han/hon söker. Därför klassar Radhakrishnan (2009) sökmotorn som en semantisk sådan.

Sökmotorn behandlar sökbegrepp och frågor utifrån fyra olika kategorier (Cognition Search, 2008):

1. **Legal** – Returnerar träffar som behandlar juridiska inslag.
2. **Health** – Fokuserar på träffar som handlar om medicin och hälsa.
3. **Wiki** - Söker efter engelska artiklar på *Wikipedia*.
4. **Bible** – Återger träffar som är kopplade till bibeln.

För att användaren skall kunna få ut så optimala sökträffar som möjligt, bör han/hon välja vilken kategori som lämpar sig bäst mot sökfrågorna.

3.8.7. Powerset

Powerset genererar sökträffar som är baserade på artiklar från *Wikipedia*. Sökmotorn behandlar frågor, begrepp och meningar och genererar därefter träffar som oftast är organiserade i kategorier. Dessa kategorier beskriver Powerset (2008) som ”Factz”, vilket är representationer av information som extraheras från sökbegreppen. *Powerset* tillämpar ontologier för att beskriva representationerna.

3.8.8. SenseBot

SenseBot (2007) beskriver sökmotorn som en semantisk sådan. Sökmotorn genererar en summering i respons till sökbegreppen, istället för en samling länkar som sökträffar. *SenseBot* analyserar de mest väsentliga resultaten på webben, och sammanställer dessa i en summering av text (Sensebot, 2007). Teknologier som sökmotorn använder för att åstadkomma detta är:

- **Text Mining** – Används för att analysera webbsidor som återger den mest relevanta informationen i förhållande till söktermen.
- **Multidocument Summarization** – Är den faktiska metoden som används för att generera koherenta sökträffar

3.8.9. Factbites och Exalead

Radhakrishnan (2009) beskriver *Factbites* som en sökmotor med målet att returnera betydelsefulla meningar som sökträffar. Detta görs genom att sökmotorn söker efter matchande meningar och information på webben, för att generera träffar som nödvändigtvis inte nämner den använda söktermen i resultatet, utan istället har resultaten fokus på semantiken av termen.

Exalead är enligt Radhakrishnan (2009) en optimal sökmotor för bildsökning på webben. Exalead (2008) beskriver att sökmotorn använder semantiska teknologier för att strukturera, skapa mening samt precisera sökresultat. Sökmotorn samlar data från samtliga källor på webben, oavsett formattyp. Sökresultaten genereras sedan i olika kategorier och dataformat.

4. Resultat

Enligt (Backman, 2008) behandlar resultatdelen i en uppsats data i sammanfattad form, och endast sådana data som har direkt anknytning till frågeställningen/problemställningen ska redovisas.

Nedan följer först en sammanfattning av viktiga begrepp och definitioner om vad den semantiska webben är för något. Därefter presenteras resultatet av det faktiska söktestet.

4.1. Sammanfattning av den semantiska webben i begrepp och definition

I teoriavsnittet kom det fram en hel del teori som behandlar begrepp och definitioner om den semantiska webben. Det är dock klart konstaterat att den semantiska webben, dess begrepp och definitioner har till syfte att gynna webben till dess fullständiga potentialer. De viktigaste begreppen och definitionerna som är av relevans till mina frågeställningar presenteras nedan:

begrepp	Definition
<i>Ontologier</i>	Kommunikationsmedel för sökmotorer som bygger på "Triples"
<i>Triples</i>	Använder en "Subjekt-Predikat-Objekt"-relation för att beskriva relationen mellan olika begrepp.
<i>Semantik</i>	Beskriver innebörden av begrepp.
<i>RDF</i>	En datamodell som bygger på XMLs syntax
<i>RDFS</i>	Ett schemaspråk för RDF
<i>OWL</i>	Ett språk för ontologier inom den semantiska webben
<i>XML</i>	Ett språk som beskriver informationsstrukturer
<i>Sökmotor</i>	Söktjänst som läser webbsidor på Internet och sparar ned dem i en databas
<i>Semantiska sökprinciper</i>	Har till syfte att optimera sökträffar i förhållande till precision relevans
<i>Metadata</i>	Data som beskriver annan typ av data
<i>Relevans</i>	Ju högre precision sökträffarna har, desto högre relevans får de gentemot användarens sökintentioner
<i>Precision</i>	Återge träffsäkra sökträffar gentemot söktermer

De två sista begreppen i tabellen är intressanta i förhållande till resultatet av söktestet. Som det kom fram i metoddelen så använde jag mig av mätfaktorer som "Träffsäkerhet" och "Allmängiltighet" för att mäta sökresultatet, och det är dessa två mätfaktorer som motsvarar: "Precision" och "Relevans".

4.2. Resultat för söktestet

Resultatet av det faktiska söktestet bygger på information som återspeglar samtliga undersökningsenheters beteende i förhållande till sökfrågorna: semantiska sökmotorers beteende kontra webbens mest använda sökmotor, *Google Search*.

Nedan följer en redogörelse av resultatet i tabeller, där varje sökmotor och dess sökresultat presenteras i en tabell var för sig.

4.2.1. Hakia

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Mycket tillfredsställande	Hög allmängiltighet	8	6
2	Icke tillfredsställande	Låg allmängiltighet	0	0
3	Mycket tillfredsställande	Hög allmängiltighet	10	10
4	Mycket tillfredsställande	Hög allmängiltighet	10	10
5	Mindre tillfredsställande	Låg allmängiltighet	4	0

4.2.2. Swoogle

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Icke tillfredsställande	Ingen allmängiltighet	0	0
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Mindre tillfredsställande	Låg allmängiltighet	2	0
4	Icke tillfredsställande	Ingen allmängiltighet	0	0
5	Icke tillfredsställande	Ingen allmängiltighet	0	0

4.2.3. Exalead

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Icke tillfredsställande	Ingen allmängiltighet	0	0
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Icke tillfredsställande	Ingen allmängiltighet	0	0
4	Icke tillfredsställande	Ingen allmängiltighet	0	0
5	Icke tillfredsställande	Ingen allmängiltighet	0	0

4.2.4. Powerset

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Tillfredsställande	Medel allmängiltighet	5	2
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Mycket tillfredsställande	Hög allmängiltighet	10	10
4	Mycket tillfredsställande	Hög allmängiltighet	10	6
5	Mindre tillfredsställande	Medel allmängiltighet	4	0

4.2.5. Sensebot

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Icke tillfredsställande	Ingen allmängiltighet	0	0
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Icke tillfredsställande	Ingen allmängiltighet	0	0
4	Icke tillfredsställande	Ingen allmängiltighet	0	0
5	Icke tillfredsställande	Ingen allmängiltighet	0	0

4.2.6. Factbites

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Mindre tillfredsställande	Medel allmängiltighet	4	0
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Mycket tillfredsställande	Hög allmängiltighet	10	10
4	Mycket tillfredsställande	Hög allmängiltighet	10	10
5	Icke tillfredsställande	Låg allmängiltighet	2	0

4.2.7. Cognition Search

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Mindre tillfredsställande	Låg allmängiltighet	3	0
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Tillfredsställande	Medel allmängiltighet	5	3
4	Tillfredsställande	Medel allmängiltighet	5	2
5	Icke tillfredsställande	Låg allmängiltighet	1	0

4.2.8. Lexxe

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Mindre tillfredsställande	Låg allmängiltighet	3	2
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Mycket Tillfredsställande	Hög allmängiltighet	10	10
4	Mycket Tillfredsställande	Hög allmängiltighet	10	10
5	Tillfredsställande	Medel allmängiltighet	6	0

4.2.9. Kosmix

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Mindre tillfredsställande	Låg allmängiltighet	3	3
2	Icke tillfredsställande	Ingen allmängiltighet	0	0
3	Mycket Tillfredsställande	Hög allmängiltighet	10	10
4	Mycket Tillfredsställande	Hög allmängiltighet	10	7
5	Mindre Tillfredsställande	Låg allmängiltighet	2	0

4.2.10. Google Search

Sökfråga	Träffsäkerhet	Allmängiltighet	Relevanta träffar	Explicita svar
1	Mycket tillfredsställande	Hög allmängiltighet	10	10
2	Mindre tillfredsställande	Låg allmängiltighet	2	0
3	Mycket Tillfredsställande	Hög allmängiltighet	10	10
4	Mycket Tillfredsställande	Hög allmängiltighet	10	10
5	Tillfredsställande	Medel allmängiltighet	6	0

5. Analys och diskussion

I detta avsnitt analyseras och diskuteras resultatdelen samt det viktigaste i den teoretiska referensramen som behandlar avsnittet om den semantiska webben.

5.1. Den semantiska webbens roll och syfte

Tim Berners-Lee och hans definition av den semantiska webben, som (Vossen & Hagemann, 2007) citerar, kan tolkas som om att: webben och dess fullständiga potentialer är ej ännu uppnådda. Men med hjälp utav den semantiska webben kommer webben att finna ett optimalare tillstånd för vad gällande utbyte av information mellan webbapplikationer, vilket kan leda till större och smartare automatisering av informationsbehandling över webben.

Bakke (2007) beskriver Berners-Lee's definition av den semantiska webbens primära mål som ett försök att kamouflera de teknologisk deterministiska aspekterna av den semantiska webbevolutionen. Detta görs av Berners-Lee genom att argumentera om evolutionens teknologiska fördelar för slutanvändarna. Dock så ser Bakke (2007) en kontradiktiv mening i det här resonemanget, då teknologin som (Berners-Lee et al. 2001) beskriver är egentligen en specifikation-driven evolution som i huvudsak kommer att utveckla nya teknologistandarder för teknologins skull och inte användarens. Resonemanget ovan kan jag som användare uppleva som korrekt. Speciellt i förhållande till det resultat som jag kom fram till i undersökningen. Där hälften av sökmotorerna inte riktigt levererar som de skall göra.

Bakke (2007) fortsätter sin diskussion om att den semantiska webben mycket väl skulle kunna vara ett namn eller koncept som omfattar utvecklingen av ett par specifika teknologier. Dessa teknologier redogjordes i teoridelen av den här uppsatsen, och utgjorde även den semantiska webbens grundläggande struktur som (Vossen & Hagemann, 2007) illustrerade som konceptets teknologiska arkitektur. Det finns dock kritik riktad mot den här strukturen och dess komponenter. Kritiker som Peter Norvig, direktör för forskning på Google, kritiserar användningsbilden av ontologier och metadata samt dess komplexitet i den semantiska webben. Detta framförallt ur ett användarperspektiv, där nödvändig kompetens och insikt är viktig för att förstå hur den semantiska webben fungerar. Ett konkret exempel på detta är *RDF* och kritiken riktad mot det. (Bakke, 2007) påstår att *RDF* är ett betydligt mycket mer komplexare märkspråk än vad exempelvis *HTML* är. Detta på grund av dess anpassning till webbutvecklare. Författaren ger ett exempel på den varierande kompetensen som finns inom webbutveckling, genom att deklarerat problemområdet för utvecklare som döper sina webbsidor till: *untitled.html*. Denna indikation ser författaren som ett tecken på omogenhet inom utvecklarvärlden, omogenhet för bruk av nya teknologier som exempelvis:

RDF, RDFS och *OWL*. Dessa språk omfattar dock den semantiska webbens mest fundamentala språk, och ifall inte webbutvecklare är redo att börja utveckla semantiska webbapplikationer med hjälp av de språken, på vilket sätt skulle slutanvändarna kunna vara redo för den semantiska webben?

För att besvara på ovanstående fråga kan det vara nödvändigt att återvända till definitionen som (Bakke, 2007) citerar angående den semantiska webben och dess primära mål. Följande del av det andra citatet kan vara intressant ur ett kritiskt perspektiv att diskutera om:

“The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people.” (W3C, 2001)

Ovanstående citerades år 2001. Citatet är en del av definitionen på den semantiska webben och dess syfte. Dock så menar man på att vissa kriterier skall uppfyllas för att webben skall uppnå sina fulla potentialer. Ett av dessa är nödvändigheten av datadelning och behandling. Men vad innebär detta i praktiken? Är det ännu en ursäkt för en reducerad restriktion på hur användarrelaterad information skall behandlas över webben?

En artikel från CIO Sweden (2005) beskriver ett önsketänkande scenario på hur den semantiska webben skulle kunna påverka sökagenter över webben:

”För att ta ett enkelt exempel – du kan skicka ut en sökagent som letar reda på den billigaste resan till Paris från Stockholm. Agenten kan hitta priser från flera olika flygbolag för de aktuella datumen. Men mer än så blir svårt.

Du kan exempelvis inte skicka ut en agent som tar hänsyn till hur din kalender ser ut eller ändras. Att resa från Stockholm kan dessutom vara ett vitt begrepp.” (CIO Sweden, 2005)

Likheten mellan det ovanstående exemplet och W3C's resonemang är likartade ifall man analyserar vissa fragment av citatet. Ovanstående citat resonerar att dagens sökagenter inte tar någon hänsyn till hur en användares kalender ser ut för att kunna avgöra vilket resedatum som passar användaren bäst. W3C påstår indirekt att detta skulle kunna lösas med hjälp av en optimalare datadelning och behandling över webben. Detta skulle ju i praktiken kunna innebära att webbapplikationer behöver åtkomst till personrelaterad information, dvs. åtkomst till en användares kalender, för att sedan kunna behandla informationen och ge feedback. Detta medför frågetecken om hur konsekvenserna ser ut kring en sådan implementering. Problem som kan uppstå tillsammans med en sådan ansats är enligt kritiker: politiska och ekonomiska sådana (CIO Sweden, 2005), där företag kommer att slåss på grund av olika viljor kring hur semantiska webb principer skall implementeras. Detta innebär dock inte att samtliga idéer om den semantiska webben bör förkastas och glömmas bort. Exempel på webbapplikationer som tillämpar semantiska webbprinciper och som dessutom presterar tillfredsställande finns exemplifierade från det faktiska söktestet i den här uppsatsen.

5.2. Söktestet utifrån ett användarperspektiv

Jag som författare var den enda användaren som deltog i söktestet. Anledningen till detta var för att lägga fokus på hur sökmotorerna beteende skiljer sig åt, och inte lägga så mycket vikt på hur övriga användare skulle kunna uppleva interaktionen med sökmotorerna. Jag som användare upplevde dock att hälften av de semantiska sökmotorerna inte riktigt levererade som jag ville att de skulle göra. Resultatet kring detta artade sig olika från sökmotor till sökmotor, och sökfråga till sökfråga. Men den sammanfattande upplevelsen av resultatet stämmer inte riktigt överrens med vad (Radhakrishnan, 2009) skriver om dem. Författaren skriver att de mer eller mindre skall vara revolutionerande i den så kallade sökvärlden. Titeln av själva artikeln säger ju det mesta hur författaren resonerar: "9 Semantic Search Engines That Will Change The World Of Search" (Radhakrishnan, 2009). På vilket stämmer titeln av artikeln ihop med verkligheten? Frågan kan delvis besvaras med hjälp av en analys av hur söktestet artade sig.

De enskilda söktesten visade hur varje sökmotorerna förhöll sig mot respektive sökfråga. *Hakia* levererade godtyckligt nog en tillfredsställande prestation, där tre av fem frågor besvarades med tillfredsställande träffsäkerhet och hög allmängiltighet. Anledningen till att *Swoogle* inte presterade lika bra som *Hakia* kan bero på att *Swoogle* inte är en kommersiell sökmotor. *Swoogle* tillsammans med *Exalead*, *Sensebot* och *Cognitive Search* var de sökmotorer som presterade minst tillfredsställande i söktestet. Ingen av dessa sökmotorer förutom *Cognitive Search* levererade några som helst explicita svar. Hur kan detta komma sig? Kan detta vara en konkret indikation på att det existerar en brist på kompetens från utvecklarnas sida? I sådana fall har Bakke (2007) en poäng i det han säger om att utvecklare av den semantiska webben inte är tillräckligt mogna för att utveckla semantiska webbapplikationer. Detta stämmer dock helt och hållet inte, med tanke på att *Hakia*, *Lexxe*, *Powerset*, *Factbites* och *Kosmix* levererade godtyckliga svar på sökfrågorna. Jämför man exempelvis *Hakia* och *Google Search* i söktestet, finner man mindre marginaler mellan antalet explicita svar som bägge sökmotorer levererade:

<i>Google Search</i> Explicita svar
10
0
10
10
0

<i>Hakia</i>
Explicita svar
6
0
10
10
0

Google Search levererade alltså fyra stycken fler explicita träffar än vad *Hakia* gjorde. Detta kan bero på att *Google Search* levererar betydligt fler sökträffar än vad *Hakia* gör, dock var söktestet avgränsat till att undersöka de tio första sökträffarna och inte samtliga sökträffar, och eftersom att *Google Search* rankar de mest relevanta sökträffarna som de tio första på resultatlistan så kan det vara en anledning till den marginella skillnaden.

5.3. Vad är skillnaden mellan semantiska sökmotorer och Google Search?

Som det visade sig i resultatdelen av den här uppsatsen, var det ytterst få semantiska sökmotorer som tillsammans med *Google Search* genererade flest träffar på sökfrågorna. *Swoogle* tillsammans med *SenseBot* och *Exalead* var den sökmotorn i testet som var minst funktionell i förhållande till sökfrågorna. Faktum att *Swoogle* enbart genererade sammanlagt 2 st sökträffar och att hälften av dessa träffar var brutna länkar, kan vara ett tecken på att sökmotorn inte är tillräckligt moget ännu. Men att *SenseBot* och *Exalead* inte genererade några träffar överhuvudtaget är inte särskilt lyckat! Vilken uppfattning bidrar dessa sökmotorer med till användare av den semantiska webben? Det finns ju en risk och sannolikhet för att man som användare inte väljer att tro på konceptet och visionen när man interagerar med konkreta implementationer som *SenseBot* och *Exalead*.

Enligt Ezzy (2006) tillhör *Google Search* den tredje generationens sökmotorer vilket i praktiken rankas som en Web 2.0-applikation (Vossen & Hagemann, 2007). Men på vilket sätt skiljer sig semantiska sökmotorer med andra generationens sökmotorer som exempelvis *Google Search*? Vad är skillnaden mellan de semantiska sökmotorerna och *Google Search*?

Det är ju precis den här skillnaden som kommer fram mellan *Swoogle*, *SenseBot* och *Exalead* som semantiska sökmotorer, och *Google Search* som en andra generations sökmotor, det vill säga aspekten kring hur funktionalitet smälter samman med användarvänligheten. En märkbar skillnad mellan de semantiska sökmotorerna i testet och *Google Search* är hur den sistnämnda sökmotorn tillämpar *AJAX* och *Google Suggest* för att generera potentiella sökförslag till användaren. Och med tanke på att *Google Search* är den mest använda sökmotorn över webben idag, kan man fråga sig ifall tillämpningen av *AJAX* på något vis kan ha påverkat det? Hur skulle *Hakia*'s användarsiffror och trafik artat sig ifall sökmotorn tillämpade en liknande funktionalitet som *Google Suggest*? Eller vice versa, ifall *Google Search* började tillämpa semantiska sökprinciper?

Att det finns en skillnad mellan samtliga sökmotorer i testet kan framhäva sig genom att de tillämpar teknologier och standarder på olika sätt, men hur interaktionsaspekten skiljer sig åt mellan dem är nog slutanvändarna på webben mest representerade att svara på. *Hakia* kategoriserar exempelvis sina sökträffar, vilket *Google Search* också numera gör. Innebär detta att *Swoogle* och de övriga semantiska sökmotorerna som inte har den funktionaliteten, är mindre representerade som semantiska sökmotorer?

5.4. Har den semantiska webben lyckats?

Gustavsson (2003) skriver varför han tror att den semantiska webben kommer att misslyckas. Detta skrev författaren år 2003, det vill säga för sju år sedan. Han oroar sig för att "vi" som användare och utvecklare är lata i den meningen att "vi" inte har någon lust att producera komplexa syntaxer i *RDF* och *OWL*, när "vi" i själva verket har problem med att uppdatera våra hemsidor i ren *HTML*. Andra aspekter som författaren beskriver att det fungerar fint idag när man skall söka efter information över webben. Det sistnämnda visade sig stämma bra överrens med undersökningen i den här uppsatsen, där *Google Search* var den sökmotorn som genererade flest antalet sökträffar. Men skall man föredra kvantitet över kvalitet i en sådan kontext? Visserligen genererade *Google Search* inte enbart flest antal sökträffar, men samtliga träffar förhöll sig tillfredsställande nog gentemot sökfaktorer som: *Träffsäkerhet*, *Hastighet* och *Allmängiltighet*.

En annan aspekt som Gustavsson (2003) drar upp angående den semantiska webben är frågan om varför auktoriteter och utvecklare inte satsar mer på att få datorer att processa naturligt språk och dra slutsatser av det, istället för att skapa teknikberoende lösningar som kräver tekniska standarder?

W3C är de ansvariga för att skapa nya webbstandarder, detta omfattar även standarder som kommer att användas tillsammans med den semantiska webben. Nya standarder kräver nya insikter om hur de skall brukas. Bakke (2007) och hans kritik mot att utvecklarvärlden inte är riktigt redo för bruk av nya standarder, kan ha en poäng i det.

Men samtidigt så är webben under en konstant utveckling, och detta bör inte heller stagneras på lång sikt. Berners-Lee (2006) beskriver att den semantiska webben kommer att vara en del av "Web 3.0".

I dagsläget befinner sig webben i "Web 2.0" (Vossen & Hagemann, 2007), men är verkligen slutanvändarna redo för en ny webbevolution eller ej? Eller är det redan för mycket frågetecken över de fenomen som webben består av idag? Ifall slutanvändarna inte förstår sig på hur den semantiska webben fungerar, kommer den heller kanske inte ha lyckats, såvida begreppet "lyckas" och dess semantik är nödvändigtvis lika med nya webbstandarder.

5.5. Min definition av den semantiska webben

För att återknyta ett svar på vad den semantiska webben är för något enligt i förhållande till vad upphovsmän och övriga forskare säger, har jag valt att definiera en egen definition av vad den semantiska webben.

Den semantiska webben enligt mig är ett begrepp vars semantik bidrar med fler frågetecken än svar. Begreppet kan enligt mig inte rationaliseras och endast bidra till en konkret definition av den semantiska webben egentligen är, och därför skapar det förvirring kring konceptet som sådan. Varför kallar man inte istället konceptet för: "Den perfekta webben", "Den utökande webben", "Web 3.0", "Den revolutionerande webben"? För det är ju mer eller mindre det visionen av konceptet antyder på, det vill säga: ett nytt kapitel av webben som skall bidra med standarder som förhoppningsvis fulländar webben till dess fulla potentialer. Detta innebär att det sker en revolution som mycket väl skulle kunna få användare på webben att uppleva den som "perfekt", vilket skapar en tillfällighet för upphovsmän och auktoriteter på webben att döpa om webben till "Web 3.0" istället för "Web 2.0".

Bydén & Ebbesen (2000) citerade Aristoteles syn på hur semantik kan definieras, och inledningsvis finner jag citatet intressant betraktandes på att den delvis har en filosofisk klang och innebörd som väcker en del tankar på hur hans syn på semantik, kan sedermera förhålla sig med definitionen av den semantiska webben:

"Ett namn är en talenhet som fungerar som tecken enligt överenskommelse, utan bibetydelse av tid, varav ingen del fungerar som tecken då den tas separat. I namnet "Fagerlund" betecknar ju "lund" ingenting i sig självt, vilket samma talenhet gör i satsen: det är en fager lund..."

Jag återknyter nu citatet ovan till "Den semantiska webben" som begrepp, där begreppet "Den semantiska webben" bidrar med en tolkning om att: "Det är en semantisk webb", eller "En webb med semantik", vars innebörd i sig kan uppfattas som ekvivalent. Men vad antyder begreppet "Semantik" om i den här kontexten? Min tolkning är att webben den semantiska webben har till syfte att lägga betoning på semantik. Men i vilken kontext?

Semantik är ju som sagt kopplat till begrepp och ord, därför är det väsentligt att sökmotorer skall nyttja semantiska sökprinciper för att optimera användarens sökning. Den här ansatsen upplever jag som en självklar sådan i teorin, med tanke på hela konceptet och dess titel. Detta är en förbättringsaspekt som upphovsmän formulerade för nio år sedan (Berners-Lee et al. 2001), men mitt söktest visade ju att detta inte stämmer hundra procentigt. Visserligen var min studie avgränsad till tio stycken sökmotorer där nio av dessa var semantiska, och i teorin mycket lovande. Dock så förhöll ju enbart hälften av dem tillfredsställande nog gentemot mina sökintentioner som användare.

Jag tror att den semantiska webben är en grundläggande anledning till att börja utveckla nya webbstandarder och teknologier. Detta ser jag både på gott och ont, eftersom nya standarder bidrar med nya möjligheter för att utvecklare att utveckla applikationer, eller som Tim Berners-Lee uttrycker det som "Maskiner" (Berners-Lee et al. 2001). Men samtidigt så måste ju utvecklare lära sig dessa standarder och utvecklingsteknologier, vilket i sig tar tid och energi. Därför tror jag att den semantiska webben är ett begrepp, koncept, vision samt ett projekt som behöver mogna. Efter nio års tid har det enligt mig inte riktigt gjort det. Författare som Vossen & Hagemann (2007) skriver om den semantiska webben som om att det egentligen vore något nytt, detta sex år efter att W3C (2001) har definierat vad den semantiska webbens primära syfte och roll är. Detta ser jag som ett tecken på att den semantiska webben inte riktigt har mognat ännu. Därför ser jag att den semantiska webben som ett begrepp som bidrar med fler frågetecken än svar om vad det egentligen är för något. Samt att dess betydelse mognar med tiden, vilket innebär att framtiden kommer att få avgöra hur den semantiska webben egentligen företer sig som ett fenomen i förhållande till webben som den är idag.

6. Slutsatser

Det som har framkommit är att det finns en mångfacetterad definitionsarsenal om den semantiska webben. Den semantiska webbens primära mål är som Berners-Lee och W3C beskriver att utöka webben så att den når dess fullständiga potentialer. Detta kräver dock en ny ansats av teknologistandarder som W3C själva utvecklar. Det råder dock delade meningar om vad konsekvenserna av nya standarder blir för webbutvecklare. Kritiker riktar kritik mot den semantiska webbens fundamentala teknologiarkitektur och menar på att utvecklarvärlden inte är redo för nya standarder och språk. Trots detta så pågår det ett faktiskt projekt som W3C bedriver för att implementera semantiska webbstandarder och verktyg som övriga utvecklarforum över webben skall kunna använda.

En viktig aspekt av den semantiska webbens syfte är att bidra med principer som sökmotorer kan utnyttja för optimalare informationsbehandling. Detta visade sig dock arta sig olika i sökundersökningen, där hälften av de semantiska sökmotorerna presterade tillfredsställande och de övriga inte.

Slutligen ser jag den semantiska webben som ett begrepp, koncept, vision och ett projekt som behöver mogna. Jag ser även begreppet som ett bidragande till fler frågetecken än svar på vad den semantiska webben är för något, och att framtiden får avgöra hur den semantiska webben företer sig i förhållande till webben som den är idag.

7. Fortsatt forskning

Den här uppsatsen kan ses som en sammanställning och beskrivning om vad den semantiska webben är och hur den kan arta sig gentemot sökmotorer över webben. Denna konstellation av definitioner kommer förhoppningsvis att underlätta förståelse om den semantiska webben som ett begrepp, koncept och implementation.

Större utbud av semantiska webbapplikationer hade kunnat påverka insikten om vilka skillnaderna är mellan applikationer som tillämpar semantiska webbprinciper och inte. Potentialerna kring dessa principer hade man sedan kunnat koppla och se hur de hade kunnat påverka befintliga affärsmodeller över webben. På så vis skapar man ett större perspektiv på hur den semantiska webben kan bidra med fördelar. Att söka på webben eller via de angivna källor som jag tidigare har angett i den här uppsatsen, kan vara en rekommendation för att finna artiklar och övrig information som handlar om den semantiska webben

Avslutningsvis kan man använda den här rapporten för att få en grundläggande uppfattning av vad den semantiska webben är för något, samt hur effekten av det kan påverka sökmotorer för undersökning av information på webben.

8. Referenser

Nyblom Martin & Torrkulla Göran, (2009). Filosofisk ordbok [Elektronisk] Tillgänglig:
<http://filosofia.fi/se/ordbok/S>

Backman, Jarl, (2008). Rapporter och uppsatser, 2:a upplagan. Lund: Studentlitteratur.

Holme Magne Idar & Solvang Krohn Bernt, (1997). Forskningsmetodik - Om kvalitativa och kvantitativa metoder. 2:a upplagan. Lund: Studentlitteratur.

Wallén, Göran, (1996). Vetenskapsteori och forskningsmetodik. Upplaga 2:13. Lund: Studentlitteratur.

Bydén Börje & Ebbesen Sten, (2000), De interpretatione & Om sofistiska vederläggningar, Thales Stockholm

Vossen Gottfried & Hagemann Stephan, (2007), Unleashing Web 2.0: From concepts to creativity, Morgan Kaufmann Publishers

Flensburg, Per & Friis, S, (1999), Mänskligare datasystem: utveckling, användning och principer, Lund: Studentlitteratur

Fransson, Jonas, (2007), Effektivare informationssökning på webben: Handbok i konsten att söka information, Hexa Förlag

Ballard, Phil, (2006), Kom igång med Ajax, Pagina förlags

Allemang, Dean & Hendler, James, (2008), Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL, Morgan Kaufmann Publishers

World Wide Web Consortium, (2010), Homepage. [Elektronisk]. USA. Tillgänglig:
<http://www.w3.org/>

The Semantic Web Project, (2007), [Elektronisk]. USA. Tillgänglig:
http://semanticweb.org/wiki/Main_Page

Gustavsson, Benny , Den Semantiska Webben, (Senast ändrad 2003-10-04). [Elektronisk]. Swe. Tillgänglig: (<http://home.swipnet.se/semanticweb/swe/>)

Wikipedia, Den Semantiska Webben, (Senast ändrad 2009-06-01). [Elektronisk]. Swe. Tillgänglig: (http://sv.wikipedia.org/wiki/Semantiska_webben)

Högskolan Väst
Institutionen Ekonomi och IT
C-uppsats Informatik 15 hp
Haj-Bolouri, Amir

Radhakrishnan, Arun, 9 Semantic Search Engines That Will Change The World of Search, (Senast ändrad 2009-04-19). [Elektronisk]. Tillgänglig: (<http://www.searchenginejournal.com/semantic-search-engines/9832/>)

Hakie Search Engine, 2010, [Elektronisk]. USA. Tillgänglig: (<http://www.hakia.com/>)

Kosmix, 2008, [Elektronisk]. USA. Tillgänglig: (<http://www.kosmix.com/>)

Cognition Search, 2008, [Elektronisk]. USA. Tillgänglig: (<http://www.cognition.com/>)

Lexxe, 2005, [Elektronisk]. USA. Tillgänglig (<http://www.lexxe.com/>)

Powerset, 2008, [Elektronisk]. USA. Tillgänglig (<http://www.powerset.com/>)

SenseBot, 2007, [Elektronisk]. USA. Tillgänglig (<http://www.sensebot.com/>)

Factbites, 2005, [Elektronisk]. USA. Tillgänglig (<http://www.factbites.com/>)

Exalead, 2008, [Elektronisk]. USA. Tillgänglig (<http://www.exalead.com/>)

Google Search, 2010, [Elektronisk]. USA. Tillgänglig: (<http://www.google.com/>)

IEEE Xplore, 2010, [Elektronisk]. Tillgänglig:
<http://ieeexplore.ieee.org.ezproxy.server.hv.se/Xplore/dynhome.jsp>

Emerald, 2010, [Elektronisk]. Tillgänglig:
http://www.emeraldinsight.com.ezproxy.server.hv.se/Insight/menuNavigation.do;jsessionid=65A83CA92A810C4C30006447573C848A?hdAction=menu_ca_sb_quick_search

Science Direct, 2010, [Elektronisk]. Tillgänglig:
http://www.emeraldinsight.com.ezproxy.server.hv.se/Insight/menuNavigation.do;jsessionid=65A83CA92A810C4C30006447573C848A?hdAction=menu_ca_sb_quick_search

Wikipedia, World Wide Web (Senast ändrad 2010-04-28) [Elektronisk]. Tillgänglig:
http://sv.wikipedia.org/wiki/World_Wide_Web

Wikipedia, Webbläsare (Senast ändrad 2010-04-09) [Elektronisk]. Tillgänglig:
<http://sv.wikipedia.org/wiki/Webbl%C3%A4sare>

Wikipedia, Webbserver (Senast ändrad 2010-05-16) [Elektronisk]. Tillgänglig:
<http://sv.wikipedia.org/wiki/Webbserver>

Högskolan Väst
Institutionen Ekonomi och IT
C-uppsats Informatik 15 hp
Haj-Bolouri, Amir

Wikipedia, World Wide Web Consortium (Senast ändrad 2010-05-02) [Elektronisk].
Tillgänglig: http://sv.wikipedia.org/wiki/World_Wide_Web_Consortium

Bakke, Sturla, (2007), Waiting for the Semantic Web: Some comparative and sociotechnical aspects regarding the evolvement of the World Wide Web, [Elektronisk].
Tillgänglig: <http://www.duo.uio.no/publ/informatikk/2007/69029/Bakke.pdf>

Wikipedia, Web 2.0 (Senast ändrad 2010-04-22) [Elektronisk]. Tillgänglig:
http://sv.wikipedia.org/wiki/Web_2.0

Wikipedia, Web 2.0 (Senast ändrad 2010-04-22) [Elektronisk]. Tillgänglig:
http://sv.wikipedia.org/wiki/Web_2.0

DiNucci, Darcy, (1999), Design & New Media: Fragmented Future, [Elektronisk].
Tillgänglig: <http://www.cdinucci.com/Darcy2/articles/Print/Printarticle7.html>

O'Reilly, Tim, (2005), What is Web 2.0: Design Patterns and Business Models for The Next Generation of Software, [Elektronisk]. Tillgänglig:
<http://oreilly.com/web2/archive/what-is-web-20.html>

Wikipedia, AJAX (Senast ändrad 2010-04-22) [Elektronisk]. Tillgänglig:
<http://sv.wikipedia.org/wiki/AJAX>

Natinoalencyklopedin, Semantik, (2010) [Elektronisk]. Tillgänglig:
<http://www.ne.se/semantik>

Gustavsson, Benny (Senast ändrad 2003-10-04) [Elektronisk]. Tillgänglig:
<http://home.swipnet.se/semanticweb/swe/>

W3C, Semantic Web Technology Stack, (2006) [Elektronisk]. Tillgänglig:
<http://www.w3.org/DesignIssues/diagrams/sweb-stack/2006a.png>

Wikipedia, metadata (Senast ändrad 2010-05-11) [Elektronisk]. Tillgänglig:
<http://sv.wikipedia.org/wiki/Metadata>

Wikipedia, ontologi (Senast ändrad 2010-04-03) [Elektronisk]. Tillgänglig:
<http://sv.wikipedia.org/wiki/Ontologi>

Wikipedia, xml (Senast ändrad 2010-04-12) [Elektronisk].
Tillgänglig: <http://sv.wikipedia.org/wiki/XML>

Wikipedia, resource description framework (Senast ändrad 2010-01-21) [Elektronisk].
http://sv.wikipedia.org/wiki/Resource_Description_Framework

Högskolan Väst
Institutionen Ekonomi och IT
C-uppsats Informatik 15 hp
Haj-Bolouri, Amir

Wikipedia, WebCrawler (Senast ändrad 2010-04-01) [Elektronisk].
<http://en.wikipedia.org/wiki/WebCrawler>

Ezzy, Ebrahim, (2006), Search 2.0 vs Traditional Search, [Elektronisk]. Tillgänglig:
http://www.readwriteweb.com/archives/search_20_vs_tr.php

Wikipedia, Google (Senast ändrad 2010-05-21) [Elektronisk]. Tillgänglig:
http://sv.wikipedia.org/wiki/Google_search

Wikipedia, Swoogle (Senast ändrad 2010-04-18) [Elektronisk]. Tillgänglig:
<http://en.wikipedia.org/wiki/Swoogle>

Finin et al. (2005), Swoogle: Searching for Knowledge on the Semantic Web,
[Elektronisk]. Tillgänglig: <https://www.aaai.org/Papers/AAAI/2005/ISD05-007.pdf>

Wikipedia, Academy (Senast ändrad 2010-05-01) [Elektronisk]. Tillgänglig:
<http://en.wikipedia.org/wiki/Academy>

Berners-Lee, T & Hendler, J & LASSILA, O (2001), The Semantic Web, Scientific
American

W3C, (2001-2007c), W3C Semantic Web Activity Statement, [Elektronisk]. Tillgänglig:
<http://www.w3.org/2001/sw/Activity>

CIO Sweden, (2005), Semantik gör webben smartare, [Elektronisk]. Tillgänglig:
<http://cio.idg.se/2.1782/1.84420>

Wikipedia, Semantics (Senast ändrad 2010-05-19) [Elektronisk]. Tillgänglig:
<http://en.wikipedia.org/wiki/Semantics>

Wikipedia, Ontology (information science) (Senast ändrad 2010-05-20) [Elektronisk].
Tillgänglig:
[http://en.wikipedia.org/wiki/Ontology_\(information_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science))