

Semantisk Web

Hvad er meningen?



“First I searched for Larry in Yahoo, then Lycos, Excite and Infoseek. Eventually, I found him in the bathroom.”

Dette projekt er udarbejdet af:

Janus Hove Jørgensen

Helle Bendsen

Magnus Vestureid

Morten Møller Tanderup Jacobsen

Peter Grønkjær

Rasmus Peter Precht

5. semester

Humanistisk Datalogi

Aalborg Universitet

Vejleder: Henrik Schärfe

INDLEDNING	1
Problemformulering	4
Analyse af problemformulering	4
INFORMATIONSSØGNING I DAG	6
Informationssøgning på biblioteket	6
Informationssøgning på Internettet	7
Brugeren og søgning	8
Ottetrinsmodellen.....	9
Søgeredskaber	11
VIDENSREPRÆSENTATION	15
Mening	15
Charles Sanders Peirce	16
Conceptual Graphs	17
Ontologibygning	18
Hvorfor ontologier?.....	18
Ontologiers rolle	19
Ontologiers opbygning	20
Cyc	21
Er det realistisk at repræsentere al <i>common sense</i> ?.....	22
Sandhed	24
DET SEMANTISKE WEB	26
eXtensible Markup Language.....	26
Ressource Description Framework.....	27
Universal Ressource Identifiers.....	28
RDF-skema.....	29
DAML-OIL og OWL.....	30

UBOTS ONTOLOGIER	3 3
Ontologierne.....	33
AEROSWARM - ANALYSE	3 6
SEMANTISK WEB I BRUGSKONTEKST	4 3
DISKUSSION	4 8
Derivationsmodellen.....	50
Semantik.....	51
Mening og betydning.....	52
Kan computeren forstå?	52
Grænser for formalisering.....	54
Forståelse i interaktionen	58
Værktøj eller interaktionspartner.....	59
KONKLUSION	6 1
TEMARAMMEN	6 4
Kurserne	65
PROCESBESKRIVELSE	7 0
Møder	71
Indbyrdes kontakt, styring af dokumenter.....	72
Tidsplan og planlægning	72
Vejleder	73
Empiriindsamling.....	73
Ansvarsområder	74

L I T T E R A T U R	7 6
Bilag 1	I
Bilag 2	VII
Bilag 3	XIV
Bilag 4	XXVIII

Indledning

Inden for de sidste ti år er Internettet vokset i popularitet. I december 1995 havde Internettet 16 millioner brugere. Dette antal er i oktober 2004 vokset til 812 millioner brugere over hele verden (Internet 1). Ligesom antallet af brugere nærmest er eksploderet indenfor det sidste årti, er også udbuddet af information på nettet blevet større og samtidig mere uoverskueligt. Dette medfører et voksende behov for effektive redskaber til at søge efter information (Baeza-Yates, 1998). De søgemaskiner, der findes i dag, har store problemer med, at de finder for mange irrelevante links og ikke nok relevant information (Li, 2002). Et af problemerne er, at søgemaskinerne udelukkende er i stand til at søge på konkrete ord og ikke på mening, sagt med andre ord søger de på form frem for indhold (Walker & Janes, 1999, p. 108). Ligeledes er det tankevækkende, når man i artiklen *From E-Sex to E-Commerce: Web Search Changes* kan læse, at brugerens adfærd i forbindelse med søgning har ændret sig meget lidt siden 1997, selvom Internettet er blevet væsentligt større og mere uoverskueligt. Der drages i artiklen følgende konklusion:

“We need a new generation of Web searching tool based on a more thorough understanding of human information behaviors” (Spink et al., 2002)

Det Semantiske Web kunne være en løsning, som kan gøre informationssøgning mere effektiv og danne grundlag for en ny generation af søgemaskiner.

Begrebet *The Semantic Web* blev introduceret af Tim Berners-Lee (2001) i artiklen af samme navn. Heri beskriver Berners-Lee sin vision om Webbet udvikling i fremtiden:

"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and "understand" the data that they merely display at present." (Berners-Lee, 2001).

I 2001 mente Berners-Lee altså, at computere "i den nærmeste fremtid" ville blive i stand til at "forstå" data frem for blot at vise dem. I dag, tre år senere, er Semantisk Web stadig på et tidligt udviklingsniveau. Den almindelige internetbruger har stadig ikke set noget til den "nye funktionalitet" det semantiske web skulle tilføje Webbet. Berners-Lee påpeger da også selv nogle områder, som mangler at blive udviklet eller fungere, før "Semantisk Web"-teknologien kan tages i brug: Vidensrepræsentation, ontologi og agenter er de vigtigste. Mens der er mange spændende problemstillinger forbundet med agenter, vil fokus i dette projekt ligge på de to førstnævnte. Her vil vi vende blikket mod John F. Sowa's *Knowledge Representation* (2000), der beskriver vidensrepræsentation og ontologi som en gren af forskningen indenfor kunstig intelligens, men samtidig forbinder begreberne med logikken, inden for hvilken diskussionen omkring vidensrepræsentation har stået på i to et halvt årtusinde (Sowa, 2000, p. 1). Sowa behandler også de filosofiske spørgsmål omkring vidensrepræsentation og ontologier og er dermed et oplagt udgangspunkt for vores diskussion af, hvorvidt vi "i den nærmeste fremtid" endelig skal høste frugterne af 2500 års forskning og af, hvordan man skal forstå de gåseøjne Berners-Lee sætter omkring ordet "forstå". Det er netop spørgsmålene om, hvordan man repræsenterer viden, og hvad der skal til, for at man kan hævde at en computer "forstår", vi især finder interessante. Berners-Lee har som nævnt store visioner med det semantiske web, bl.a. at det skal kunne indgå som en integreret del af hverdagen ved hjælp af agenter, som f.eks. kan kommunikere med ens kalender og derefter aftale en tid hos lægen. I dette projekt vil vi dog ikke komme nærmere ind på, hvor-

vidt denne drøm kan blive til virkelighed. Vi vil i stedet koncentrere os om den formentlig lidt nærmere forestående udvikling, nemlig det projekt at berige dokumenter på World Wide Web med semantisk betydning og helt konkret, hvad denne berigelse vil betyde for informationssøgning på Internettet. Berners-Lee har selv omtalt problemet med informationssøgning Internettet, som det ser ud dag:

Web searches today typically turn up innumerable completely irrelevant "hits," requiring much manual filtering by the user. If you search using the keyword "cook," for example, the computer has no way of knowing whether you are looking for a chef, information about how to cook something, or simply a place, person, business or some other entity with "cook" in its name. The problem is that the word "cook" has no meaning, or semantic content, to the computer (Berners-Lee, 2001).

Ifølge Berners Lee (2001) er det semantiske web en udbygning af og ikke en erstatning for World Wide Web, som vi kender det i dag. Provokeret af det høje ambitionsniveau, der kommer til udtryk i artiklen, vil vi diskutere hvordan ideerne og teknologierne bag Semantisk Web vil kunne forbedre vilkårene for informationssøgning i forhold til traditionelle teknologier.

Ovenstående leder os frem til følgende problemformulering:

Problemformulering

Med afsæt i en analyse af søgesituationen i dag og tankerne og teknikken bag det Semantiske Web, undersøges mulighederne for at forbedre vilkårene for informationssøgning.

Analyse af problemformulering

For at skærpe forståelsen af vores problemformulering vil vi i det følgende redegøre for betydningen af de enkelte dele.

Med afsæt i en analyse af søgesituationen i dag...

Med formuleringen "Med afsæt i..." mener vi, at vi bruger analysen af søgesituationen som et springbræt til at diskutere det semantiske web. Denne analyse vil altså ikke være hovedfokus i vores opgave. Dens formål er at belyse, hvilke problemer brugeren oplever, når denne skal finde information på Internettet i dag, og i hvor høj grad nutidens søgereedskaber afhjælper eller bevirker disse problemer. Disse problemer skal belyses for at klargøre, hvordan det Internet ser ud, som det semantiske web forsøger at forbedre og hvilke områder, der især er problematiske.

...og tankerne...(bag det semantiske web)

En stor del af udviklingen af det semantiske web er stadig på tankestadiet, og meget af Berners-Lees artikel *Semantic Web* er fremtidsvisioner og tanker om, hvad det semantiske web kan udvikle sig til. En vigtig del af projektet bliver derfor at diskutere/filosofere over, hvorvidt disse tanker kan realiseres på baggrund af forskellige teorier.

...og teknikken bag det semantiske web...

Det er dog ikke kun tanker, der ligger bag det semantiske web:

“Two important technologies for developing the Semantic Web are already in place: eXtensible Markup Language (XML) and the Resource Description Framework (RDF)” (Berners-Lee, 2001)

Selvom semantisk web endnu ikke er virkelighed, er der udviklet en række standardsprog, som man forestiller sig kan blive drivkraften bag det semantiske web. Bag navne så kryptiske som RDF, OWL, XML etc. ligger en række retningslinjer for, hvordan man beskriver webindhold semantisk. Disse retningslinjer fokuserer hovedsageligt på syntaks, men deres form kan også have konsekvenser for det semantiske webs funktion. Analysen af Ubots output skal give os en fornemmelse af, hvor langt man er nået i den tekniske udvikling af det semantiske web.

...undersøges mulighederne for...

Da det semantiske web endnu ikke er en realitet, er vores ambition med dette projekt at overveje og diskutere *mulige* aspekter af det semantiske web, ikke at komme med endelige svar.

...at forbedre vilkårene for informationsøgning.

Målet med dette projekt er at undersøge, hvordan informationsøgningen i dag kan forbedres, dvs. hvordan bliver det nemmere for brugeren at finde relevant materiale på Internettet? Vi mener, at en realisering af Semantisk Web ville kunne bidrage med en sådan forbedring, og derfor består en stor del af vores projekt af overvejelser og diskussioner om, hvad der skal til, for at drømmen om det semantiske web kan blive opfyldt, og hvordan det konkret kan forbedre informationsøgning på Internettet.

Informationssøgning i dag

Vi vil i dette afsnit redegøre for hvordan mulighederne og principperne for informationssøgning har udviklet sig, efterhånden som mængden af tilgængelig information er vokset

Informationssøgning på biblioteket

Før Internettet og World Wide Web blev tilgængeligt for den brede befolkning, foregik informationssøgning hovedsageligt på biblioteket via informationssøgningssystemer, som søger i forskellige databaser. Disse databaser kan indeholde metainformation, dvs. information om de bøger, der findes på biblioteket, repræsenteret ved f.eks. bibliografiske parametre og indholdsanalyser. Databaserne kan også indeholde fuldtekstdokumenter som f.eks. avisartikler og statistikker. Uanset om søgesystemet skal søge i databaser, som indeholder metadata eller fuldtekst, er det designet til en bestemt målgruppe, som har bestemte informationsbehov, og det er så systemets opgave at oplyse brugeren om, hvor denne information kan findes eller levere den direkte i form af fuldtekstdokumenter (Chowdhury, 1999, p. 2). Den information, som findes i disse databaser, er produceret, editet, organiseret og indekseret af professionelle, og brugeren kan derfor have stor tiltro til kvaliteten af det materiale, som databaserne repræsenterer (Walker & Janes, 1999, p. 108). Der findes flere forskellige systemer, som kan søge efter information i disse databaser. De søge-

sprog, de benytter, er forholdsvis avancerede at lære, fordi de stammer fra 70'erne, hvor de færreste mennesker havde adgang til computere, og det at søge information i databaser derfor ikke var noget almindelige personer nødvendigvis skulle være i stand til at lære på kort tid. Til gengæld er disse søgesystemer meget effektive, når man først har lært at bruge dem rigtigt.

Vi vil ikke komme nærmere ind på, hvordan disse søgesystemer fungerer. Grunden til, at vi nævner dem i denne forbindelse, er, at vi mener, det er givtigt i forhold til at diskutere, hvad der er vundet, og hvad der er gået tabt med hensyn til at finde relevant information af høj kvalitet, efter at informationssøgning i højere grad er blevet en aktivitet, der finder sted på Internettet.

Informationssøgning på Internettet

Selvom Internettet er en af de største og lettest tilgængelige kilder til information, der findes i dag, er det ikke dermed sagt, at det er blevet lettere at finde den information, man leder efter.

I artiklen *Searching the web: Challenges and partial solutions* beskriver Ricardo A. Baeza-Yates (1998) de væsentligste problemer i forbindelse med at søge på Internettet. Der er bl.a. problemer med den måde, hvorpå data distribueres på Internettet: Den store mængde af flygtigt data er et problem, som bl.a. gør det vanskeligt at genfinde data, når fil- og domænenavne skifter. Internettets eksponentielle vækst gør mængden af data uoverskuelig og svær at håndtere. Data på Internettet er ofte ustrukturerede, da der ikke er nogen bagvedliggende model for, hvordan data skal struktureres. Hvis man betragter Internettet som et formidlingsmedie, er der den afgørende forskel i forhold til de konventionelle medier, at det mangler den redaktionelle instans, så indholdet kan være dårligt skrevet eller direkte forkert (Baeza-Yates, 1998, p. 1).

Internettets decentrale, anarkistiske karakter medfører at alle kan distribuere information på Internettet. Ingen holder øje med indholdet på nettet, og der er ingen regler for, hvad man må bruge Internettet til, medmindre man foretager sig noget, som er direkte kriminelt. I modsætning til bibliotekernes afgrænsede databaser, er der altså ingen garanti for at det, man finder på Internettet, besidder nogen som helst form for kvalitet eller relevans (Walker & Janes, 1999, p. 108).

Et andet problem med at søge på Internettet i forhold til i databaser er, at HTML ikke er beregnet til at strukturere data, men blot præsentere dem.

Hvor det i databaser er muligt at søge på f.eks. en forfatter, et udgivelsesår eller et emne, kan man på Internettet kun søge på, hvordan informationen er repræsenteret, f.eks. et billede, et link eller en adresse. Walker og Janes skriver i forbindelse med søgning på Internettet overfor søgning i databaser: "Typically we will be able to search on what something *is* (an image, a link, an address) rather than what it *means* (an abstract, a subjekt heading, an author)" (Walker & Janes, 1999, p. 108).

Brugeren og søgning

Alle disse forhold medfører, at der stilles store krav til brugeren i forbindelse med informationssøgning på Internettet. Baeza-Yates (1998, p. 2) deler disse problemer op i to hovedområder: Hvordan man søger, og hvordan man efterfølgende håndterer resultatet af søgningen.

Det faktum, at man som regel kun kan søge på de konkrete ord, som indgår i et dokument og ikke på indhold som f.eks. emne, medfører, at det bliver vanskeligere at finde det, man søger efter. En søgning på et eller flere almindelige ord vil grundet den enorme mængde information på Internettet medføre et uoverskueligt antal resultater. Ofte vil mange af søgeresultaterne være irrelevante grundet problemet om flertydighed. Skal man f.eks. finde en side, om hvor hurtigt en jaguar (katten) kan løbe, og søger på "jaguar" og "speed", vil de fleste af siderne indeholde information om biler (Baeza-Yates, 1998, p. 5). Problemet opstår, fordi computeren ikke ved noget om den kontekst, som søgningen foregår i, og fordi ordene i et dokument dækker ikke hele dokumentets semantiske betydning (Baeza-Yates, 1998, p. 4). Eksemplet med jaguaren kan forekomme lidt naivt, fordi brugeren jo kan begrænse sin søgning væsentligt ved f.eks. at skrive "jaguar speed NOT car". (Kan variere alt efter hvilken søgemaskine man benytter) Problemet er bare, at ofte er brugeren ikke klar over at denne mulighed eksisterer, eller også gider brugeren ikke ofre den nødvendige tid, det tager at overveje, hvordan en søgning formuleres bedst muligt. I artiklen *From e-sex to e-commerce: Web Search Challenges* (Spink et al., 2002) fremlægges en undersøgelse om brugeres adfærd i forbindelse med søgning på søgemaskinen Excite. Undersøgelsen viser at knap 60 procent indtaster 2 søgeord eller derun-

der og kun 10 procent benytter sig af Boolean-søgning dvs. brug af AND, OR, eller NOT. Desuden viser undersøgelsen at mere end 70 procent nøjes med at se på de første to resultater af søgningen eller færre. Dette kan være et udtryk for, at brugeren ikke har tålmodighed til at se flere resultater igennem, hvis de første resultater ikke er relevante (Spink et al., 2002).

For at belyse denne søgeproces nærmere har vi valgt at inddrage Walker & Janes ottetrinsmodel.

Ottetrinsmodellen

Når man søger efter information, gennemgår man mere eller mindre bevidst nogle faser både under selve søgningen, og inden man kommer frem til det, man vil søge efter. En model som kan illustrere denne proces, er Walker og Janes ottetrinsmodel (Walker & Janes, 1999, p. 75ff), som vi her vil tage udgangspunkt i. Selvom denne model tager udgangspunkt i søgning i organiserede databaser, kan den også bruges til at illustrere de faser, man går igennem ved søgning på Internettet. Desuden vil vi senere forholde modellen til det semantiske web.

De otte trin: (Walker & Janes, 1999, p. 108)

1. Find ud af, hvad der skal søges efter. (Read the query)

Dette punkt afspejler, at modellen også dækker søgninger i f.eks. bibliotekssystemer, hvor en bibliotekar skal finde noget for låner. Derfor skal bibliotekaren først finde ud af, hvad det er, låneren leder efter. Ser man på modellen som et udtryk for, hvilke faser den, som søger, gennemgår, er dette punkt mindre relevant, da man jo oftest er klar over, hvad man vil søge efter, inden man går i gang.

2. Find begreber, der dækker det, der skal søges efter. (Identify the major concepts in the query)

Dette punkt ser vi som en uddybning af punkt et, hvor man gør sig klart hvilke begreber, man kan knytte på det, man søger efter.

3. Find ord, der passer på begreberne. (Identify potential terms to correspond to those concepts)

Som søgemaskinerne ser ud i dag, får man sjældent et godt resultat ud af at søge på begreber, fordi de er alt for brede. Det er derfor nødvendigt at knytte mere konkrete ord på disse begreber for at få et brugbart søgeresultat.

4. Find alternativer til ordene. (Select alternative terms to use if the original strategy needs help)

Vi mener, at dette foregår mere eller mindre bevidst. Når man søger på et eller flere ord, har man ofte alternativer i baghovedet. Disse kan være nyttige, hvis den første søgning ikke giver det ønskede resultat.

5. Find logiske sammenhænge imellem ordene. (Determine logical relationships between terms)

Dette punkt hjælper til at præcisere søgningen yderligere. Det kan være ved brug af *og* eller *eller*, trunkering eller udelukkelse af ord f.eks. sko –støvler. De fleste søgemaskiner har disse funktioner, men det kræver, at brugeren er villig til at bruge lidt tid på at sætte sig ind i, hvordan reglerne er defineret i den enkelte søgemaskine.

6. Udfør søgningen. (Begin the search)

Dette punkt omhandler selve søgningen, hvor søgemaskinen skal gøre sit arbejde.

7. Se nogle af søgeresultaterne igennem. (Have a look at a few documents)

Efter søgningen er foretaget skal brugeren danne sig et overblik over, om der er nogle af søgeresultaterne, som kan bruges. Brugbarheden kan sjældent vurderes ud fra blot at se på resultatoversigten. Ofte bliver man nødt til at klikke på et par af linkene for at danne sig et indtryk af, om de søgeord, man har benyttet, har givet det ønskede resultat.

8. Revidér søgningen ud fra resultatet. (Revise and refine the search based on those initial results)

Hvis søgningen ikke har fundet, hvad man ønsker, må man vurdere, hvad der er galt og prøve igen f.eks. ved at gøre søgningen bredere, snævrere eller prøve med nogle af de alternative ord, man har tænkt på i trin 4.

De første faser går ud på at "gætte" med hvilke ord, det man leder efter, er beskrevet. Man skal så at sige vide, hvad det er, man leder efter for at finde det. De sidste faser er selve søgningen og vurdering af, om man har fundet det man leder efter. Hvis ikke man finder det, man leder efter, kan modellen fra trin 2, afhængig af hvor meget søgningen skal ændres, bruges som en iterativ proces, man gennemgår indtil man opnår et tilfredsstillende resultat.

Vi har medtaget modellen for at illustrere, at søgning ofte kan være en lang proces, hvor brugeren skal gøre et stort stykke arbejde med at finde de rigtige søgeord, se resultater igennem og vurdere deres relevans, især når man medtager betragtningerne i efterfølgende afsnit, som omhandler søgemaskinernes formåen i forhold til at finde relevant information.

Desuden er det vores hensigt at inddrage modellen i forhold til det semantiske web, så vi kan undersøge om, og eventuelt hvordan, søgefaserne vil ændre sig, hvis det semantiske web bliver en realitet. Kunne man f.eks. forestille sig, at nogle af punkterne vil blive foretaget af computeren? Dette spørgsmål vil vi vende tilbage til.

Men først vil vi se nærmere på de søgeredskaber, vi har til rådighed i dag.

Søgeredskaber

Søgning på Internettet afhænger ikke kun af, om man kan finde ud af at beskrive det, man leder efter, på en passende måde, men også af hvordan søgemaskiner udfører søgninger og præsenterer søgeresultaterne. Derfor vil vi se på, hvordan forskellige søgemaskiner fungerer overordnet, og hvordan de adskiller sig fra hinanden f.eks. beskrive hvordan, de vurderer relevansen af information.

Søgning på Internettet forgår som nævnt ofte ved hjælp af søgemaskiner. Disse søgemaskiner registrerer hjemmesider på nettet og indekserer dem. Dette foregår automatisk og er ikke afhængigt af, hvad der søges på. Søgning via en søgemaskine foregår så ved, at man søger i indekset over hjemmesider (Internet 5).

De fleste søgemaskiner finder mere eller mindre hjemmesiderne på samme måde ved at følge linkene imellem de enkelte sider. Det, der adskiller søgemaskinerne fra hinanden, er hvordan, de vurderer relevansen af informationen på siderne.

Google bruger et system kaldet PageRank. (Internet 6) Dette går ud på at vurdere en sides relevans afhængig af hvor mange andre sider, der linker til den pågældende side. Så jo flere andre sider der linker til en side, jo mere relevant menes informationen på siden at være. Yderligere vægter links fra andre sider med mange links til højere. Denne måde at vurdere en hjemmesides relevans på adskiller sig markant fra en side som f.eks. Hotbot. Hotbot vurderer relevansen af indholdet på en hjemmeside efter, hvor mange gange ordet, der søges efter, forekommer på en side, og hvor mange gange det forekommer i forhold til, hvor meget der ellers står på siden. (Internet 7) Yderligere vurderes en side højere, hvis søgeordet forekommer i titlen på siden frem for kun i teksten. Spørgsmålet er så, hvorvidt nogen af disse metoder reelt siger noget om indholdets relevans i en søgning. Bare fordi de sider, hvor søgeresultatet står, linkes til af mange sider, behøver siden det ikke at være relevant i den aktuelle søgesituation, og en side bliver heller ikke mere relevant, fordi ordene står flere gange på en side.

En anden måde at søge efter information på er at bruge et directory som f.eks. Yahoo!'s. Et directory er en hierarkisk opbygning af kategorier som hjemmesider er inddelt i. Her søger man ved at browse igennem nogle kategorier, som starter bredt og efterhånden, som man bevæger sig ned i hierarkiet, bliver mere specifikke. Man ender til sidst med nogle hjemmesider, der behandler emner i den valgte kategori. Fordelen ved at søge information på denne måde er, at hvis man finder noget om det, man leder efter, kan man være ret sikker på, at det er relevant, fordi hjemmesiderne i et directory er registreret manuelt af personer. Ulempen ved dem er, at fordi de er registreret på denne måde, er der ikke så mange hjemmesider at søge i, da det tager lang tid at registrere dem manuelt (Baeza-Yates, 1998, p. 4). Desuden er kategoriseringerne ikke altid lige indlysende, og det kan derfor også her være svært at finde, hvad man søger.

De søgemaskiner, man har i dag, dækker cirka en tredjedel af de hjemmesider, der findes på Internettet (Google dækker over 8 milliarder). Dette enorme antal kan som før nævnt medføre, at brugeren får et stort og uoverskueligt søgeresultat. Men der er også fordele ved, at Internettet er så stort. Har man et konkret spørgsmål, kan man forsøge at søge på spørgsmålet i den form, man forventer, at det er repræsenteret f.eks. "The Eifel Tower was built by". Der er så pga. af Internettets enorme størrelse god sandsynlighed for, at svaret findes et eller andet sted i netop den form, man har indtastet. AskMSR (Banko et. al., 2002) er et program, som udnytter dette.

Her kan man indtaste sit spørgsmål, og programmet opstiller så spørgsmålet i forskellige formuleringer med henblik på at finde det i en af de mulige formuleringer. Dette program løser altså nogle af brugerens problemer. Sammenligner man med ottetrinsmodellen, er det trin 4 og 5, som automatiseres. Men AskMSR løser stadig ikke problemerne med flertydighed og kan også kun finde svaret, hvis spørgsmålet er formuleret i en af de alternative formuleringer, den kommer med. Den søger altså kun på form og ikke indhold.

Helt overordnet kan man sige, at alle søgemaskinerne har problemer med at opfylde kravene om *precision* og *recall* (Li, 2002, p. 1). *Precision* betyder, at alle de søgeresultater, som søgemaskinen finder, er relevante for brugeren. *Recall* betyder, at al relevant information bliver fundet. En optimal søgning vil altså medføre, at brugeren får søgeresultater, som omfatter al den relevante information, som ligger på Internettet og ikke andet. Almindelige søgemaskiner har især problemer med kravet om *precision*. De finder mange irrelevante dokumenter, hvilket bl.a. skyldes deres metoder til at vurdere relevansen af en side og problemerne med tvetydige ord. Søger man i et directory, er det mere sandsynligt, at kravet om *precision* er opfyldt, da de sider, directoriet indeholder, som før nævnt er kategoriseret af mennesker. Til gengæld er sandsynligheden for at kravet om *recall* bliver opfyldt meget lille, da directories kun indeholder en lille del af den information, som findes på Internettet.

I starten af dette afsnit omtalte vi søgning i databaser, og vi mener, at både søgning på Internettet og søgning i databaser besidder nogle problemer, som det semantiske web måske kunne løse: Databasesøgning gør det muligt at søge på metadata, og søgeresultaterne er af høj kvalitet og relevans. Men denne form for søgning er avanceret og kræver, at brugeren er villig til at bruge tid på at lære teknikkerne, og samtidig er udbuddet af information i databaserne begrænset. Internettet tilbyder søgemaskiner, som er meget intuitive og lette at gå til, og mængden af information på Internettet er enorm og vokser dag for dag. De store mængder information gør det desværre også sværere at finde, hvad man søger, og søgemaskinerne er endnu ikke gode nok til at sortere i denne information. Desuden er der ingen garanti for, at det, man finder, besidder nogen form for kvalitet. Som vi ser det, er det en af det semantiske webs opgaver at tage det bedste fra begge situationer, uden at alle ulemperne følger med. Dvs. man skal kunne søge på metadata ligesom i bibliotekernes databaser, hvilket medfører bedre og mere relevante søgeresultater, samtidig skal man

have let adgang til informationen hjemme fra sin egen computer uden at skulle sætte sig ind i besværlige og omfattende søgeteknikker.

Dette problem skaber en efterspørgsel efter en bedre metode til at finde relevant information, og det semantiske web er et forslag til en løsning. Et helt centralt begreb i forbindelse med semantisk web er vidensrepræsentation. Derfor vil dette begreb danne udgangspunkt for det følgende afsnit.

Vidensrepræsentation

For at kunne analysere og diskutere hvorvidt Semantisk web er muligt at udvikle, så er det nødvendigt, at danne et billede af de elementer som semantisk web dækker over. Da begrebet mening er et flygtigt begreb, og meningstilskrivelse er et kerneområde i semantisk web, så vil vi i det følgende adressere flertydigheden af dette begreb. Herefter vil vi redegøre for de to teorier, vi tager afsæt i. Teorierne illustrerer på den ene side en kognitiv forståelse af meningsdannelse og på den anden side en af meningsbegrebet.

Mening

Hvordan forklarer man mening? Filosoferne har ikke fået afklaret spørgsmålet endnu. Politikens bog om moderne videnskabsteori omtaler "Humpty Dumpty-teorien om mening":

"...ord betyder det, taleren har til hensigt at udtrykke med dem."

(Føllesdal et al.,1999, p. 82)

Senere i samme bog står der:

"mening, betydning og intension bruges normalt om det samme".

(Føllesdal et al.,1999, p. 176)

Disse to citater illustrerer på den ene side, hvilken subjektiv karakter mening har og på den anden side, hvor løst begrebet bliver brugt. Bogen tilbyder et alternativ til begrebet "mening", som har en konkret form, nemlig at mening har en *reference*. Dog er det problematisk at bruge denne form for mening i et system, da to ord kan have samme reference. Et eksempel herpå er "aftenstjernen" og "morgenstjernen", som er to ord med forskellig mening, men de refererer begge til planeten Venus. (Sowa, 2000, p. 176).

Charles Sanders Peirce

Peirces treleddede tegnbegrebsforståelse er her illustreret med en figur (Drotner, 2002, p. 186). Tegnbegrebet defineres som følger:

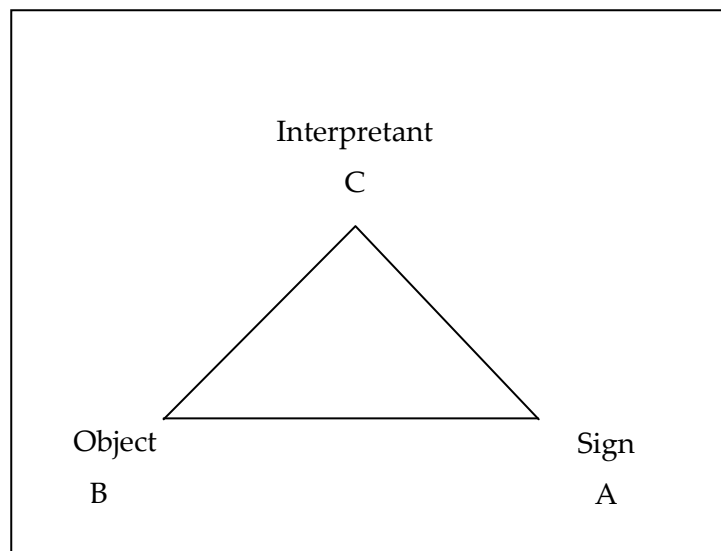
"a sign is something, A, which denotes some fact or object, B, to some interpretant thought, C." (Collected Papers of Charles Peirce, 1903, 1-346)

Denne tegnforståelse er en kognitiv proces, der forløber, såfremt vi står overfor et tegn, som vi skal forholde os til. Tegnet kan eksempelvis være et ord

eller billede. *Interpretanten* er de sammenhænge, der opstår i forbindelse med tegnet, mens *sign* er det, der repræsenterer *objektet*. I princippet er Peirces tegnforståelse uendelig, da tegn henviser til andre tegn og så videre (semiosis).

Ifølge bogen *Peirce's Doctrine of Signs* fremgår det, at Peirce mener, at et tegns mening opstår med tegnets interpretant, og der er ingen garanti for, at to personer deler samme sæt interpretanter for samme tegn (Colapiedro, 1996, p. 25).

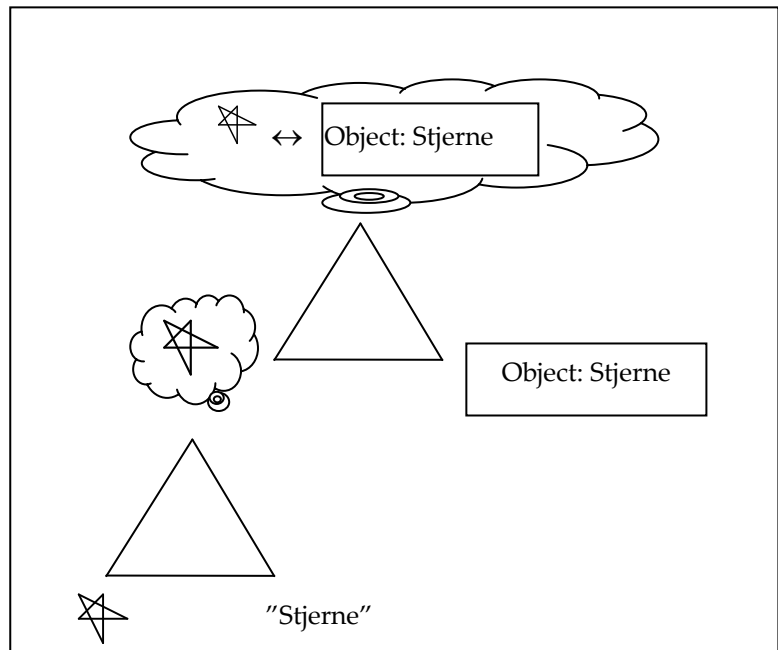
For vores problemstilling betyder dette, at med udgangspunkt i denne teori vil det blive svært at formalisere koncepter (Sign), da koncepter har denne individuelle karakter (Interpretant).



Figur 1: Figur der illustrer hvordan begreberne i Peirce tegnforståelses teori, står i forhold til hinanden.

Conceptual Graphs

Med afsæt i Peirces tegnforståelse opdeler Sowa meningsbegrebet i tre dele: Den virkelighed, man fokuserer på (Object), symbolet, der refererer til denne virkelighed (Sign) og konceptet om objektet (Interpretant). Denne trekant udvides med yderligere en trekant, som illustrerer tilblivelsen af et koncept. I Figur 2 illustreres overgang fra Peirces tegnforståelse til skabelsen af et koncept. Ved at sammensætte to trekantede finder Sowa frem til et koncept, som refererer til et objekt. Konceptet "Stjerne" defineres således i figuren som et objekt.



På denne måde opdeler Sowa meningsbegrebet og konkretiserer det

Figur 2

igen i form af en konceptrepræsentation. (Sowa, 2000, p. 191-194) Denne konceptrepræsentation er et led af det, som Sowa kalder for *Conceptual Graphs*.

Et andet element i Sowas *Conceptual Graphs* er det, Sowa kalder for *Three representational primitives*, som er strukturer, der ligger til grund for alle typer relationer, der kan laves mellem koncepter:

- Coreference: To koncepter referer til samme Objekt.
- Prehension: To koncepter står overfor hinanden ved relationen "har"; en mor har et barn.
- Containment: En struktur, der danner et abstrakt koncept ud fra andre koncepter. "The concept is an abstract container...that some entity x has a description p". Dette betyder, at dette abstrakte koncept eksisterer i kraft af, at det er beskrivende for andre koncepter. Eksempelvis dækker "Measurements" over måleenhederne meter og tomme. Sowa kalder det for "fact of togetherness".

(Sowa, 2000, p. 85)

Ontologibygning

Hvorfor ontologier?

Problemet med at få computeren til at behandle (dvs. søge i eller "forstå") sprog er som nævnt ikke så meget at få computeren til at analysere hver sætning på det syntaktiske niveau; Der er ifølge Sowa i dag maskiner, der ved hjælp af grammatikregler kan analysere en enkelt sætning på en brøkdel af et sekund, men disse regler kan kun analysere på formen og ikke indholdet. Dette gør sig gældende for vores problemstilling, da det i en søgesituation er nødvendigt for computeren at kunne analysere sig frem til relevant materiale, hvis ønsket er at gøre en computer i stand til at give en meningsfuld tilbagemelding på en forespørgsel efter materiale med et bestemt indhold. Problemet er konteksten "to represent the vast amount of background knowledge necessary to understand those languages" (Sowa, 2000, p. 179). Som det næsten allerede er sagt, er det her vidensrepræsentationen kommer ind i billedet. Ontologiens opgave er så at repræsentere den del af baggrundsviden, der er definition af og relationer imellem begreber; *entities*. Nogle ontologibyggere, som f.eks. Push Singh fra *The Open Mind Common Sense Project* hos MIT's MediaLab, bruger begrebet *common sense* om det, der her er blevet kaldt baggrundsviden:

"The real problem is that computers do not know anything about us! Our machines lack *common sense* -- all that ordinary knowledge that people in our society share, things we understand so well we hardly even realize we know them. Computers do not know what we look like, how we typically behave, or what we are capable of. They do not know anything about the patterns of people's lives, the places we spend our time, or the kinds of relationships we have with each other. They know nothing of our hopes and fears, the things we like and the things we loathe, or the feelings and emotions that motivate and underlie everything we do. By giving computers ways to represent and reason about knowledge about people, they can be made to become helpful and eager participants in the human world." (Singh, 2002)

Her forestiller man sig altså, i lighed med for eksempel cyc-projektet, at man kan "berige" computeren med den fælles *common knowledge*, som man forestiller sig eksisterer imellem mennesker. I det følgende vil vi kort forklare ontologiers rolle i forbindelse med vidensrepræsentation, samt redegøre for de traditionelle filosofiske og praktiske betragtninger i forbindelse med ontologibygning. På baggrund heraf vil afsnittet diskutere det perspektiv, der kommer til udtryk i denne måde at kategorisere på, samt problemer og begrænsninger i forbindelse hermed.

Ontologiers rolle

Den videnskabelige disciplin "Ontologi" er "the general theory of what there is" (Mautner 2000, p. 401). Udover om videnskaben om *alting* bruges ordet ontologi mere konkret om begrebshierakier: "it is the study of existence, of all the kinds of entities – abstract and concrete – that make up the world (...) the selection of categories determines everything that can be represented in a computer application" (Sowa 2000, p. 51). Endnu mere konkret bruges begrebet i forbindelse med OWL på w3c's hjemmeside:

"An ontology defines the terms used to describe and represent an area of knowledge. Ontologies are used by people, databases, and applications that need to share domain information (a domain is just a specific subject area or area of knowledge, like medicine, tool manufacturing, real estate, automobile repair, financial management, etc.). Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them" (Heflin et. al. 2002)

Ordet ontologi kan således tillægges mange forskellige betydninger afhængigt af perspektiv. I forbindelse med vidensrepræsentation er de to sidste af ovenstående definitioner mest interessante. De minder om hinanden og er i hvert fald enige om, at ontologier er den del af vidensrepræsentation, der skaber et vokabular og definitioner af kategorier eller koncepter, uanset at det ene citat taler om alt, hvad verden består af, og det andet om specifikke vidensdomæner.¹

¹Forskellen i de to citater er ikke udtryk for en uoverensstemmelse mellem Sowa's og W3C's opfattelse af ontologier. Tværtimod skriver Sowa (2000 p. 492) også at en ontologi er "a catalog of the types of things that are assumed to exist in a domain of interest".

Ontologiers opbygning

Petersen et. al. (2003) nævner fire måder at definere typer på i et hierarki: ekstension, intension, ved hjælp af aksiomer og definition ved andre typer + differentiae. De fire måder vil blive gennemgået i det følgende, der imidlertid ikke er tænkt som en udtømmende forklaring eller teorigennemgang, men som oplæg til den efterfølgende diskussion.

Ekstension

Hvis en type defineres ved ekstension, dens denotation, vil definition være en liste af alle dens instanser – eksempelvis er ekstensionen af "hest" alle heste (Sowa, 2000, 99).

Intension

Definition ved intension står i modsætning til ovenstående. Her defineres en type ved de specielle kendetegn, der inkluderer eller ekskluderer instanser fra typen. I praksis ligner denne type definition subtyperelationen (Petersen et al., 2003).

Aksiomer

Ved et aksiom forstås "a formula (i.e. a proposition, or a well-formed expression which on interpretation yields a proposition), that belongs to an axiomatic system, without being derived from any other formula in that system. Traditionally, propositions taken as axioms were so selected because they were thought to be self-evidently and indubitably true: Neither capable of proof nor requiring any" (Mautner, 2004, p. 56). Som eksempel nævner Petersen et al. En række aksiomer fra matematikken: "If a is a natural number then so is $a+1$.", og andet af samme skuffe. Aksiombegrebet tillægges imidlertid også en anden betydning i forbindelse med vidensrepræsentation. Copeland (1993, p. 111) beskriver følgende eksempel (hvor han trækker på det såkaldte *Yale shooting problem* (Hanks & McDermott)): "If a live human is shot in the head with a loaded gun, they die (immediately)". I forbindelse med dikotomien fysisk/abstrakt, der udgør den øverste top i mange ontologier, giver Sowa (2000, p. 76) blandt andet følgende eksempel på et aksiom: "If x is physical, then x has a location in space-time. If x is abstract, then x has no physical

location, but x might be assigned abstract coordinates in some imaginary space." og angiver desuden, at "The categories at every level of the hierarchy can be characterized by axioms and definition. But unlike the the categories Physical and Abstract, the definitions of the other top-level categories and their subtypes depend on the observer's viewpoint". Aksiomer kan altså betragtes som regler, der kan være indiskutable, givne eller at betragte som *common sense*. Som Sowa også antyder i ovenstående citat, er dette imidlertid ikke ensbetydende med, at definition ud fra aksiomer ikke er udtryk for et perspektiv.

Subtyperelationen

Den mest almindelige relation i en ontologi er *subtyperelationen*: "If $A \leq B$ (read as "A is a subtype of B"), then either: A is B, or A is a *specialization* of B." (Petersen et. al., 2003, Part II, 6.2). Hermed er princippet om *nedarøning* næsten også forklaret. Hvis $A \leq B$, og dermed en specialisering, arver A alle egenskaber fra B (og dermed fra alle dennes supertyper). Ifølge Aristoteles kan man således definere en type (species) som genus (supertypen) + differentiae, hvor differentiae er "The properties or features that distinguish a type from other types that have a common supertype" (Sowa, 2000, p. 492). I praksis kan den opgave, det er at bygge en ontologi ud fra disse principper, være ret omfattende, og mere konkrete og detaljerede filosofiske principper kan være ønskelige.

Cyc

Cyc, som vi har nævnt tidlige, er en vidensbase, det vil sige mere end en ontologi. Udviklerne beskriver Cyc som: "a formalized representation of a vast quantity of fundamental human knowledge: facts, rules of thumb, and heuristics for reasoning about the objects and events of everyday life." Cycs ontologi er i øvrigt også en af de ontologier, som Ubot kan trække på. Man kan altså sige, at Cyc-projektet antager en vision, der ligger tæt på den oprindelige betydning af ordet ontologi, nemlig den at lave en base, der omfatter al viden eller fælles *common sense* blandt mennesker.

Er det realistisk at repræsentere al *common sense*?

I forbindelse med dette spørgsmål, om Cyc og andre² er realistiske projekter, kan det være relevant at inddrage Sowa's ide om "knowledge soup" (2000, p. 348ff):

"Some of the knowledge in people's heads may be represented in proposition, more of it in imagelike forms, and the rest of it in habits, vague intuitions, and "gut feelings" that are never verbalized or visualized. Whatever its form, the knowledge is far too complex and disorganized to be called a knowledge base. Its fluid, heterogeneous, ever changing and often inconsistent nature could be better characterized as *knowledge soup*. The soup may contain many small chunks (...) The chunks should be internally consistent, but they may be inconsistent with one another". Sowa erkender altså, at det er problematisk at betragte menneskelig viden som en vidensbase, eller bare som noget repræsenterbart – selv om problemerne er tydeligst i natursprog, løser det ikke problemerne at repræsentere dem i logik: "A logical predicate like *bodyPart(x)* would solve nothing, the intermediate cases of hair and fingernails would raise the same questions of truth and falsity as the english noun phrase *body part*". For at løse problemet med at konsistens i globale vidensbaser er næsten umulige at opnå, foreslår Sowa (1990) en struktur på to niveauer: Et stort reservoir af løst organiseret leksikalsk viden (Videnssuppen) og en række små, stramt organiserede, internt konsistente teorier eller mikroverdener, som man så skal forestille sig "flyder" rundt i suppen. Menneskelig viden betragtes altså som inkonsistent. Inkonsistensen kan skyldes mange faktorer. Sowa nævner blandt andet følgende aspekter af vores måde at ræsonnere på som mulige årsager til vores videns inkonsistente karakter:

Generalizations that omit "obvious" assumptions: Birds fly. But what about penguins?

Abnormal conditions: If you have a car, you can drive from New York to Boston. But what if the battery is dead?

²Som nævnt omfatter begrebet vidensbase mere end blot ontologi. Når de efterfølgende overvejelser alligevel er interessante for denne teksts fokus, nemlig ontologierne, er det fordi ontologi er en del af vidensrepræsentationen, og at også den kan betragtes som en slags repræsentation af common sense. Det interessante her er altså ikke så meget Cyc eller lignende, men derimod ideen om at repræsentere "common sense"

Incomplete definitions: An oil well is a hole drilled in the ground that produces oil. But what about a dry hole? A hole that has been capped?

Unanticipated applications: The parts of the human body are described in anatomy books. But is hair a part of the body? Hair implants? A wig? A wig made from a person's own hair?

Conflicting defaults: Quakers are pacifists, and Republicans are not. But what about Richard Nixon, who is both a Quaker and a Republican?

(Sowa, 1990)

Sowa taler her om menneskelig viden i den form, som den efter hans mening antager, og ikke direkte om formel repræsentation eller ontologi. Men ovenstående ræsoneringsformer kan alligevel forsøges forholdt til de tidligere nævnte fire fremherskende principper for ontologibygning:

Subtyperelationen og definition ved intension indeholder et element af generalisering. Udsagnet "fugle flyver" (der også kan betragtes som et "common-sense-aksiom") kunne være et muligt differentiae eller "særligt kendetegn", der adskiller fugle fra andre dyr. Det ses af Sovas spørgsmål "men hvad så med pingviner?", at denne definition hurtigt giver problemer, da "abnormal conditions" gør sig gældende, eller definitionen er "incomplete". Spørgsmålet er så, hvad der er normalt og unormalt, eller hvornår en definition er komplet. Også common-sense-aksiomer kan give tydelige problemer. Selv om noget betragtes som *common sense* eller aksiomatisk, vil man sjældent helt kunne udelukke, at der kan herske konflikter mellem forskellige aksiomer eller måske unormale forhold. De fleste vil vel kunne godtage det som et aksiom, at en bold placeret på et skråtstillet bræt vil trille mod den laveste ende – alligevel er der steder, hvor det modsatte er tilfældet [www.mysteryspot.com]. Man kan så tro på det eller lade være, men det er egentlig ikke det væsentlige. Nogle mennesker vil efter at have besøgt førnævnte perle blandt verdens turistattraktioner utvivlsomt revidere deres "vidensbase" til ikke engang at opfatte tyngdelovene som aksiomatiske. Man kan herudfra diskutere, om det overhovedet giver mening at tale om fælles *common sense*.

Sowa (2000) påpeger følgende modsætning (mismatch):

- Languages (natural or artificial) are made up of discrete symbols organized in well-defined syntactic structures.
- The world is made up of an endless variety of things, forms, substances, gradations, changes, and continuous flows with imperceptible transitions from one to another.

Man kan altså yderligere diskutere, om det er muligt at formalisere en verden, som ikke er formel. Sowa bemærker, at ingen symbolsk notation kan beskrive verdens fulde rigdom. Alle klassifikationer er tilnærmelser. I det følgende vil vi se på de sandhedskriterier, der ligger til grund for ontologibygning

Sandhed

For at komme nærmere den meningsopfattelse, der kommer til udtryk i ontologier, vil vi kort komme med nogle overvejelser omkring de sandhedskriterier, der ligger til grund for de forskellige tankegange og måder at bygge ontologi på. Afsnittet vil tage udgangspunkt i Sovas (2000, p. 97ff) redegørelse for "sets, collections, types and categories".

Ser man på de tidligere gennemgåede metoder for definition af begreber i et hierarki, kan disse sammenstilles med forskellige sandhedskriterier: Definition ved ekstension, der jo som nævnt vil sige at definere en type ved dens instanser. Typen defineres altså ved den måde, hvorpå den er korrespondent med en empirisk virkelighed. Definitioner ved intension, altså ved de særlige kendetegn, der inkluderer eller ekskluderer instanser fra den pågældende type, kan være, og er sandsynligvis ofte, empirisk baserede, men med den forskel, at definitionen indgår som en del af et koherent system, hvor typen i kraft af sine differentiae adskiller sig fra sin genus i det system. Således kan man også have typer, der ikke har nogen ekstension. Sowa (2000, p. 99) nævner "Enhjørning" som eksempel. Men der eksisterer en beskrivelse af dyret, en intension, der sagtens kan være koherent med ontologien, trods det at

den ikke er korrespondent med "den objektive virkelighed". De sandhedskriterier, der gør sig gældende, er også dem, der ligger til grund for subtyperelationen, der er den typiske i et typehierarki. Aksiomer skal, for at kunne defineres som sådan, være en del af et aksiomatisk system, af hvilket der også kræves koherens. Et projekt som Cyc, der er baseret på aksiomer, accepterer imidlertid alligevel inkonsistens i vidensbasen. Dermed er der altså ikke global koherens i basen, men blot koherens internt i småteorier, jf. "knowledge soup"-begrebet. Men er der tale om et bagvedliggende korrespondenskriterie? Korrespondens kan defineres som følger:

"At sige om det, der er , at det ikke er, eller om det, der ikke er, at det er, er falsk, mens at sige om det, der er, at det er, og om det, der ikke er, at det ikke er, er sandt."

(Føllesdal et. al., 1992, p. 16)

Denne definition dækker over den fysiske verden, så vel som den intellektuelle verden. Dog har definitionen problemer med paradoksale sætninger som "denne sætning er falsk". På den ene side kan man acceptere, at sproget altid vil indeholde paradokser, eller bruge et løsningsforslag, der går ud på at differentiere mellem objektsprog og metasprog. Dermed er det muligt at definere sandhed i objektsproget, men metasproget kan ikke i sin helhed defineres ud fra sandhed (ibid. p. 208ff.). Da hverdagssprog kan karakteriseres som metasprog, så må man ud fra denne definition acceptere paradokser.

Spørgsmålet er nu: hvilken type ontologier kan fungere? Kan man sammenknytting i relationer, der er evigt gyldige i alle sammenhænge? Eller står man tilbage med det faktum, at man kun kan lave domænebegrænsede ontologier, der skaber en bestemt kontekst, og at disse ontologier pr. definition aldrig vil kunne håndtere alle situationer.

Det semantiske web

Udfordringen for det semantiske web, før det kan fungere, er at skabe et sprog, der udtrykker både data og regler for følgeslutninger om disse data, og som desuden gør det muligt at eksportere regler for alle eksisterende vidensrepræsentations-systemer til Internettet. Følgende er en gennemgang af de teknologier, der skal til for at skabe sådan et sprog.

eXtensible Markup Language

Der findes to væsentlige teknologier til udviklingen af det semantiske web: XML (eXtensible Markup Language) og RDF (Ressource Description Framework). I XML er det muligt at oprette egne koder. Disse koder optræder som skjulte etiketter eller kommentarer til websites eller tekstafsnit på en side f.eks <navn> eller <company> (Berners-Lee, 2001).

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```

```
<CATALOG>
```

```
--<CD>
```

```
<TITLE>Empire Burlesque</TITLE>
```

```
<ARTIST>Bob Dylan</ARTIST>
```

```
<COUNTRY>USA</COUNTRY>
```

```
<COMPANY>Columbia</COMPANY>
```

```
<PRICE>10.90</PRICE>
```

```
<YEAR>1985</YEAR>
```

```
</CD>
```

(Internet 2)

Andre programmer kan så anvende den kode, der er blevet knyttet til en side, men den, der er forfatter til programmet, skal vide, hvad forfatteren af siden har brugt de forskellige koder til. XML gør det muligt for brugeren at tilføje sine dokumenter en vilkårlig struktur, men fortæller ikke noget om, hvad denne struktur betyder. I eksemplet ovenfor kan "company" f.eks. både opfattes som noget militært og som et kommercielt foretagende.

For at udtrykke betydningen benytter man sig af RDF.

Ressource Description Framework

RDF bygger videre på XML-koden. RDF indeler i tre enheder (triple), der kan sammenlignes med subjekt, objekt og prædikat i almindelige sætninger. Disse tre enheder beskrives ved hjælp af XML-kode. I RDF antager et dokument, at bestemte ting (objekter f.eks. personer, websites osv.) har forskellige egenskaber (prædikater f.eks. - er søster til, - forfatter til) med bestemte værdier (subjekter f.eks. en anden person, et andet website) (Berners-Lee, 2001).

RDF anvender følgende grundlæggende begreber:

1. En *resurse* (*resource*): Data, som kan identificeres på Internettet. En resurse kan være en hjemmeside, et billede, et program, et element i et XML-dokument mm.

2. En *egenskab (property)*: Resurse, som har et navn og kan anvendes som egenskab. F.eks. kan en rapport have egenskaber som forfatter, editor, udgiver, antal sider, udgivelsesdato etc.
3. En *sætning (statement)*: Sammensætning af en resurse, en egenskab og en værdi, (triple). En værdi kan også være en resurse.

(Navaretta, 2004)

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.recshop.fake/cd">
<rdf:Description
  rdf:about="http://www.recshop.fake/cd/Empire_Burlesque">
  <cd:artist>Bob Dylan</cd:artist>
  <cd:country>USA</cd:country>
  <cd:company>Columbia</cd:company>
  <cd:price>10.90</cd:price>
  <cd:year>1985</cd:year>
</rdf:Description>
</rdf:RDF>
```

I eksemplet ovenfor kan man således læse, at : [http://www.recshop.fake/cd/Empire Burlesque](http://www.recshop.fake/cd/Empire_Burlesque) har en egenskab beskrevet som < company > med værdien Columbia (Internet 3). Eller som mennesker normalt ville udtale det Cd'en Empire Burlesque er udgivet af kompagniet Colombia

Det, at RDF beskriver forholdene mellem de enkelte XML-tags, løser imidlertid ikke problemet med forståelsen af de enkelte ord, som har dobbelt betydning. (Eksemplet med company) Endvidere mangler man i eksemplet ovenfor en forståelse af hvilken møntfod, der menes i forbindelse med < cd : price > , hvad der definerer < cd : country > osv. Problemet løses ved brug af såkaldte URIs (Universal Resource Identifiers).

Universal Resource Identifiers

URI definerer objekter, subjekter og verber, hvilket vil sige, at URI'erne sikrer, at begreberne ikke blot opfattes som ord, men er knyttet til en entydig beskrivelse. Det

er muligt selv at oprette en URI på Internettet, men opretter man sin egen URI, vil den imidlertid ikke være anvendelig for ret mange andre end en selv. I stedet kan henvises til en allerede eksisterende URI, som definerer de begreber, man anvender og knytter dem til XML-tags, eller med andre ord linke til en allerede eksisterende ontologi (Kristensen, 2003). Den mest almindelige URI er URL (Uniform Resource Locator), som identificerer en internetdomæneadresse, hvor URI'en er beskrevet. I vores eksempel ovenfor er det den fjerde linje, der er URL'en:

```
xmlns:cd="http://www.recshop.fake/cd">
```

På denne side vil betydningen af artist, country, company, price, and year være defineret.

Fra linje 5 og 6 og ned beskrives, ved hjælp af RDF, hvilke ord, de definerede URI knytter sig til på netop dette site .

```
<rdf:Description
  rdf:about="http://www.recshop.fake/cd/Empire_Burlesque">
  <cd:artist>Bob Dylan</cd:artist>
```

RDF-skema

RDF-modellen indeholder imidlertid ingen muligheder for at beskrive forholdet mellem de enkelte egenskaber og ressourcer. Egenskabernes indbyrdes betydning defineres derfor i RDF-Skemaer (RDFS). Ressourcer, som er beskrevet i RDF, bliver i RDFS organiseret i frames eller klasser, som er hierarkisk struktureret, og som har tilknyttet egenskaber.

Medlemmer af en klasse arver fælles karakteristika fra deres overklasser, mens nye egenskaber er tilknyttet den enkelte klasse.

Det er derfor med RDFS muligt at skabe vokabularer, der definerer i hvilke kombinationer og klasser og egenskaber kan benyttes (Carlsen, 2003)

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf= "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
```

```

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xml:base= "http://www.animals.fake/animals">

<rdfs:Class rdf:ID="animal" />

<rdfs:Class rdf:ID="horse">
<rdfs:subClassOf rdf:resource="#animal"/>
</rdfs:Class>

```

I eksemplet ovenfor ses at RDFS er en viderebygning på RDF. Således kan man se, at URI'en eller klassen "horse" er en underklasse af klassen animal. (Internet 4)

DAML-OIL og OWL

Defense Advanced Research Projects Agency under det amerikanske forsvarsministerium har udviklet sproget DARPA Agent Markup Language (DAML) for at formalisere ontologier. Sproget er senere blevet flettet sammen med et andet sprog: Ontology Inference Layer (OIL). Denne sammenfletning har resulteret i sproget DAML-OIL. Sproget er en videreudvikling af RDFS og kan udtrykke mere komplekse relationer indenfor ontologierne. F.eks. er det i DAML-OIL muligt at udtrykke, at to klasser eller egenskaber er identiske, samt at nogle egenskaber kan have flere værdimængder. Disse restriktioner kan ikke udtrykkes i RDFS. I DAML-OIL er klassebegrebet fra RDFS også videreudviklet, og man kan udtrykke restriktioner om kombinationer af klasser.

Web Ontology Language (OWL) er blevet udviklet af W3C med DAML-OIL som model og er et standardsprog til at beskrive Webaserede ontologier. Som i DAML-OIL anvendes XML syntaks. OWL bruger elementer fra RDF, RDFS og XML. Det giver mulighed for at udtrykke komplekse beskrivelser af klasser, egenskaber og relationerne mellem dem. Dette svarer til, hvad man kan repræsentere i de mest avancerede videnrepræsentationssprog.

OWL består af tre undersprog:

- OWL Lite
- OWL DL (indeholder OWL Lite)
- OWL Full (Indeholder OWL DL)

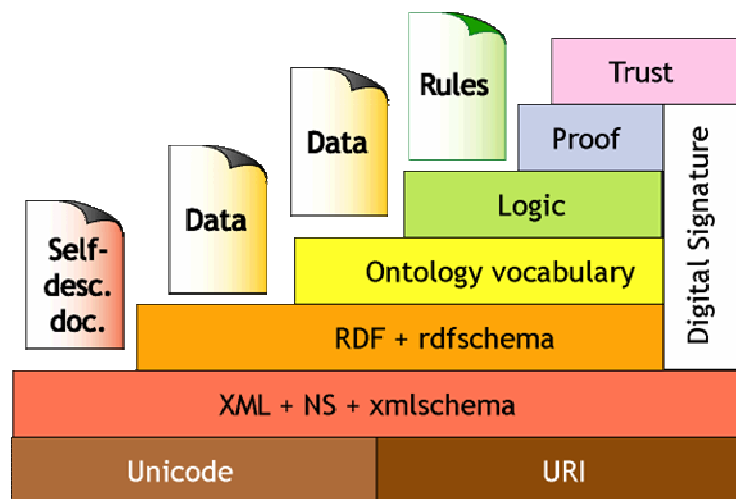
Forskellen på de tre udgaver ligger i de mulige syntaktiske specifikationer af klasser og egenskaber. Inddelingen sker også for, at man bedre kan tilpasse ontologien til det projekt, man skal arbejde med (Navaretta, 2004).

Tim Berners-Lee har fremstillet en model, der viser "lagene" i det semantiske web.

De fire nederste lag er beskrevet ovenfor, mens de tre øverste bygger videre på de fire lavereliggende lag. *Logic* ligger ovenpå ontologi. Her er det meningen, at påstanden indsamlet på Internettet kan bruges til at aflede ny viden. Således skal computere kunne ræsonnere sig frem til nye påstande baseret på kendte påstande. Et eksempel kunne være, hvis originalsproget i et dokument er Y og er skrevet af X, så medfører det, at X kender sproget Y.

De højere lag bruger logik til at udveksle *Proof*, så man kan opnå *Trust* (Li, 2002)

Tim Berners-Lees tanke med modellen er, at de sprog, som ligger på et højere niveau, arver og bruger syntaksen og semantikken fra det lavereliggende sprog.



Figur 3 : The Semantic Web "layer cake"
 (lagkagemodellen) as presented by tim berners-Lee.
 (Hendler 2001)

Men hvad gør, at semantisk web er semantisk? – man kan ikke sige, at tags i sig selv giver semantik (og arver de andre lag så noget "semantik"); er det i stedet, når man tager skridtet til URI, RDF, eller når man implementerer ontologier, der er afgørende for om Internettet er beriget med semantik? Kan man overhovedet sætte fingeren direkte på det punkt, hvor semantik opstår? Spørgsmålet er, hvad det menneskelige intellekt er, og hvordan det semantiske web passer ind i den sammenhæng. Er sø-

gen efter mening en idé i semantisk web – eller en social konstruktion, der foregår ved f.eks. interaktion. Er det derfor et spørgsmål om definitioner på hvor eller hvordan, betydning, mening eller semantik opstår. Disse overvejelser vil vi belyse senere i projektet.

Ubots ontologier

Ontologierne

I det følgende vil vi kort beskrive de tre ontologier, vi fra projektstart tog udgangspunkt i.

OpenCyc Ontology som ønskes, at blive en ontologi, der er bygget ud fra *common sense* (Internet 8)

SUMO Ontology er en mere overordnet ontologi, som er begrænset til koncepter, der er meta, generisk, abstrakt og filosofiske. Målet med SUMO er bl.a. at fremme datas samkøringsevne, informationssøgning og fremskaffelse og naturlig sprogbehandling. (Internet 8).

AeroSWARM indeholder koncepter og relationer, der er af generel betydning i flere domæner. AeroSWARM kan desuden linke to personer sammen og indeholder således en af de mest avancerede relationer, der findes i øjeblikket (Internet 8). Ontologien er en videreudviklet af AeroDAML, som blev udviklet til flyindustrien. AeroSWARM har desuden den fordel, at når outputtet skal læses af mennesker, så er det sat pænt op i tabelform. Når der f. eks klikkes på "vonRichthofen" under *Persons*, så fås nedenstående output.

Denne mulighed rummer de to andre ontologier ikke, da deres output vises i ren XML.

Person: vonRichthofen	
Also Known As:	von Richthofen (source)
Associates With:	Kunigunde (source)
Same As:	AlbrechtvonRichthofen (source)

Da AeroSWARM oprindeligt blev bygget til et begrænset domæne (fly) og samtidig påstås at kunne finde relationer mellem personer, vil det være oplagt at undersøge, hvad der kan fås ud af en side om fly, der samtidig beskriver flere forskellige relationer mellem flere forskellige personer. Da det er AeroSWARM, der siges at indeholde de mest avancerede relationer, er det oplagt at tage udgangspunkt i denne ontologi, men først har vi dog undersøgt, om der er hold i denne påstand.

Påstanden om, at AeroSWARM er den ontologi, der indeholder de mest avancerede relationer, der findes, styrkes af en sammenligning af de tre ontologiers output. Denne viser, at ontologierne finder de samme *Same as*-relationer, men udover disse finder AeroSWARM bl.a. også en *Associates With*-relation. Denne sammenligning kan ses på bilag 1. Da denne (lidt overfladiske) komparative analyse bekræfter, at AeroSWARM indeholder mere avancerede relationer end de øvrige ontologier, tager vi udgangspunkt i denne, da vores mål er at finde ud af, hvor langt man er nået med ontologibygning til det semantiske web.

Den side om fly, vi kikker på i vores analyse, består af ca. 3½ sider artikel/biografi om "Den Røde Baron". I artiklen står der både noget om Manfreds (Den Røde Baron) far, bror og mor. Hvis AeroSWARM virkelig kan linke to personer sammen, så burde den kunne finde de passende relationer af den ene eller anden slags. Desuden er det spændende at se, om ontologierne kan kende forskel på, hvornår et tal skal læses som antal, og hvornår de indgår i en dato.

Generelt om alle ontologierne kan nævnes, at de kun kan finde relationer på engelsksprogede sider.

Vi har udvalgt en side om Den Røde Baron, som vi vil bruge til at undersøge, hvor funktionel AeroSWARM-ontologien er til at finde semantisk betydning på hjemme-

sider. På bilag 2 findes der en kopi af Internetsiden, som den vises for brugeren, dog tilpasset A4-formatet.

AeroSWARM - analyse

For at finde ud af, hvor langt man er nået med semantisk web, tager vi udgangspunkt i et stykke eksisterende software, der til en vis grad er i stand til at finde semantisk betydning på hjemmesider. Det anvendte software ligger frit tilgængeligt på Internettet og anvendes online, dvs. man ikke behøver downloade noget for at bruge softwaren. Softwaren (AeroSWARM) udvikles under projektnavnet "DAML Ubot (UML Based Ontology Toolset) Project" og kan findes på URL:

<http://ubot.lockheedmartin.com/ubot/hotdaml/aeroswarm.html>. Da der også findes en ontologi ved navn AeroSWARM, tillader vi os at kalde denne software for Ubot for at tydeliggøre, hvornår vi omtaler henholdsvis software og ontologi.

Ubot tager en hjemmeside som input, og outputtet bliver XML. I AeroSWARMS tilfælde er outputtet HTML. Inputtet bestemmes af brugeren, som indtaster adressen på den hjemmeside, der ønskes anvendt som input. Outputtet bestemmes af hvilken ontologi, Ubot anvender. Der stilles tre kendte ontologier til rådighed, men brugeren kan også vælge, at outputtet skal bestemmes af en ontologi fra tredjepart, evt. en brugeren selv har lavet. Der tages udgangspunkt i AeroSWARM, da det er denne ontologi, der har vist sig at have de mest avancerede relationer.

Følgende vil vi se på, hvordan meningsstilskrivelse udtrykkes af Ubot's output med assistance af AeroSWARM-ontologien. Vi vil først sætte outputtet og teorien om konceptdannelse (der bygger på Peirce) op imod hinanden. Derefter vil vi ved hjælp af Sowas teorier om *three representational primitives* se på relationerne, AeroSWARM danner. Dette gør vi for at kunne lave en bedømmelse af hvor lang vej, der er tilbage, før ontologien er på et brugbart stadie.

Først vil vi se på, hvordan outputtet arrangerer de koncepter, som udledes af hjemmesiden om Den Røde Baron. Alle overskriftkoncepterne i outputtet må siges at passe til *containment*-definitionen. De *containment*-koncepter (supertyper), som AeroSWARM finder er: **Persons, Events, Measurements, Time Values, Organizations, Nationalities, Titles, Vehicles, Weapons** og **Locations**. Underordnet disse er der i skemaet forneden (jvf. Tabel 1) listet en række koncepter i form af subtyper, samt instanser. Ser man nærmere på disse koncepter, så er det kritisk, at "III" bliver karakteriseret som en person. Dette tal kan komme fra tre forskellige steder i teksten; enten "Kaiser_Alexander III", "Jasta III" eller "Albatros D.III". Dog er det ikke til at se, hvilket af de tre steder "III" stammer fra, men vi ved, at der ikke optræder nogen person ved navn "III" i teksten. Derfor har Ubot fejlagtigt fundet en person ved navn "III". Yderligere er det problematisk, at systemet kategoriserer "Legend" til at være af typen **Vehicles**. Det er rigtigt, at der er en bilmodel hos Honda, der hedder "Legend", men her refereres der til bogen "Richthofen: Beyond the Legend of The Red Baron". Dette er et endnu tydeligere eksempel på, at der er problemer med kontekstforståelsen. Der er ikke nogen ord i sætningen, der indikerer, at der er tale om en bil. I denne situation skal systemet selv kunne analysere, hvordan ordet skal forstås i denne sammenhæng, da det skal bruges til automatisk markup af hjemmesider.

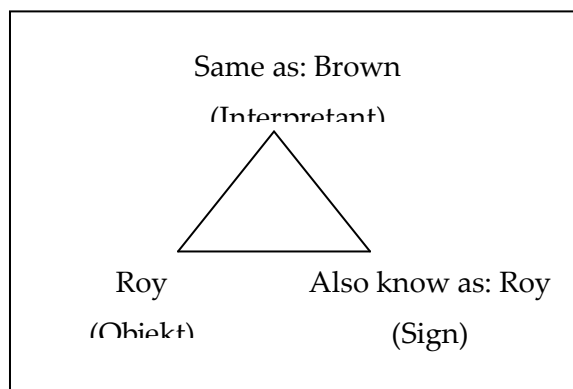
En anden *containment*-dannelse, systemet laver, er "JohnWiley", som står til at være en "**Organizations**". Dette er til dels rigtigt, men det korrekte er "John Wiley & Sons, Inc.", hvilket er et forlag. Dette er yderligere et eksempel på, at AeroSWARM laver et output, der er mangelfuldt.

Tabel 1 illustrerer den del af AeroSWARMs output, hvor kan man se, at ontologien har fundet koncepter i teksten og derefter placeret de passende koncepter i den rette sammenhæng.

Roy	Also know as: Roy Same as: Brown
RoyBrown	Also know as: Roy Brown Same as: Brown
LotharVonRichthofen	Also know as: Lothar Von Richthofen Same as: Lothar
VonRichthofen	Also know as: Von Richthofen Same as: AlbrechtvonRichthofen
ManfredvonRichthofen	Also know as: Manfred von Richthofen Same as: Manfred
KaiserAlexanderIII	Also know as: Kaiser Alexander III Same as: III
vonRichthofen	Also know as: von Richthofen Associates with: Kunigunde Same as: AlbrechtvonRichthofen
Richthofen	Also know as: Richthofen Same as: Lothar

Tabel 1

En måde hvorpå man kan illustrere den måde AeroSWARM sammenknytter koncepterne, er ved at bruge Peirces tegnforståelse. Ser vi på hvordan Roy bliver bearbejdet i Tabel 1, og opstiller dette som illustreret her til højre, så kan man på denne måde argumentere for, at AeroSWARM tilknytter "Roy" (*Object*) til variabelen "Also know as; Roy" (*Sign*). Herefter kan vi se, at AeroSWARM har fundet



Figur 4

"Brown" (Interpretant) i teksten, og sat "Brown" i forbindelse med "Also know as: Roy" og dermed også "Roy". Her sker der noget, man kan betragte som en betydningssammenhæng. AeroSWARM betragter et objekt, som bliver repræsenteret af en variable (Also know as), ligesom man i Peirce's tegnforståelse har en repræsenta-

tion af et objekt. Derefter er der en forbindelse med "Brown", som ganske rigtigt står i en association til "Roy", da Roy faktisk har efternavnet Brown. På denne måde er det lykkedes AeroSWARM at lave meningsfulde tilbagemeldinger om, hvordan nogle koncepter står i meningsfuld relation til hinanden.

Efter at have set på hvordan meningsstilskrivelsen mellem koncepter opstår i outputtet, vil vi nu se nærmere på, hvordan denne meningsstilskrivelse ligger som fundament for diverse slags relationer. De relationer, der opstår i systemet, er *Also know as*, *Same as*, *Associates with* og *Personal Contact*.

Vi har allerede været inde på, hvordan meningsstilskrivelsen opstår via Peirce, men en anden måde at anskue dette, som er mere rettet på det operationelle niveau, er at se på *Also know as* og *Same as* som *co-reference*-relationer. De fremstår som to koncepter, der refererer til samme objekt. Dette er som før nævnt den måde Sowa operationaliserer (formaliserer) Peirces meningsforståelse.

De andre relationer, der nævnes i AeroSWARM outputtet, er de to relationer *PersonalContact* og *AssociatesWith*. Disse kan karakteriseres som *prehension*-relationer, da disse imiterer en *has*-relation.

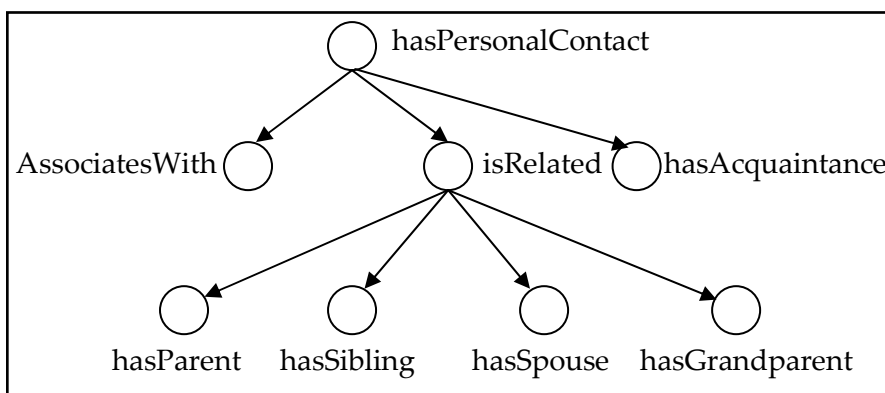
NormanFranks	Also know as: Norman Franks Personal Contact: HalGiblin
VonRichthofen	Also know as: von Richthofen Associates with: Kunigunde Same as: AlbrechtvonRichthofen

Tabel 2

Med *PersonalContact*-relationen finder ontologien en forbindelse, der er betydningsfuld. Norman Franks og Hal Giblin er medforfattere af samme tekst (Dog ikke "Red Baron"-teksten, men af en tekst "Red Baron"- teksten referer til). *Associates with*-relationen er meget interessant, da denne er relationen, ontologien vælger ud fra nedenstående tekst:

"...Major Albrecht von Richthofen, a Prussian nobleman, and his wife, Kunigunde"

AeroSWARMs ontologioversigt illustrerer, at AeroSWARM indeholder en *hasSpouse*-relation, og når man sammenligner *hasSpouse* og *AssociatesWith* med hinanden i oversigten, så er det uklart, hvorfor den vælger den ene frem for den anden. Da det er umuligt for os at se, præcist hvilke ord ontologien laver relationer imellem, så er det svært med sikkerhed at analysere sig frem til, hvorfor det ikke er lykkedes Ubot at lave *hasSpouse*-relationen. Dog kan vi påpege, at dette er et spørgsmål om præcision, fordi hvis vi ser på begrebshierarkiet i OWL-koden (Bilag 3), så kan vi se, at *hasPersonalContact* er supertype til *isRelated*, som igen er supertype til *hasSpouse*. For en mere overskuelig oversigt over ontologiens koncepter og relationer henvises til bilag 4. Bilag 4 viser dog kun de koncepter og relationer der findes og ikke deres indbyrdes forhold.



Figur 5

Dette illustrerer, hvordan relationerne står i forhold til hinanden i et hierarki. Hermed kan vi se, at det vi står overfor, er en relation, som skal præciseres. Systemet laver ikke en decideret fejl, men den udtrykker ikke hele sandheden, da relationen kan præciseres. Vi ser dette som en *prehension*-problematik.

På trods af at en af de relationer, AeroSWARM siges at kende til, er *hasParent*, og at der i artiklen står "*Manfred von Richthofen was born the son of Major Albrecht von Richthofen, a noble Prussian nobleman, and his wife, Kunigunde*", så finder den ingen relationer mellem Manfred og Albrecht. Hermed må vi igen påpege, at der er problemer med *prehension*-relationerne i AeroSWARM.

Ud fra vores analyse kan vi konkludere, at AeroSWARM har problemer med at finde alle de relationer, den siges at kunne klare. Der er tvivlsomme valg, der foretages med hensyn til *prehension*-relationerne, og der sker fejl i *containment*-strukturene.

Dette kunne tyde på, at problemet med AeroSWARM er manglende slutningsevne i mellem *Ontology Vocabulary*-laget og *logic*-laget i lagkage-modellen.

Følgende har vi fundet i teksten, der ikke er fundet af AeroSWARM.

Persons	Time Values	Kaiser
Red Baron	May 2, 1892	Emperor
Douglas Cunnel	September	
2nd Lt. Albert Woodbridge	Alle datoerne under årstal, eksempelvis 1916 Sep 24	Vehicles
Wilfredmay	January 12 th	Nieuport
(Hermann) Göring		Albatros D. II biplane
Erwin Böhme	Nationaliteter	Fokker Dr. I triplane
	Uhlans	Albatros D. III
Organizations	Frence	Sopwith Triplane
Jagdgeschwader 1	Preussian	Albatros D. V.
Jasta 11	Roman	
		Locations
Measurements	Titles	
Flyenes mål i tommer (")	Baron	Breslau, Germany
Max speed 103 mph	Rittmeister	Wroclaw, Poland
Twenty	2 nd Lt.	Vaux sur Somme, France
Eityh	Capt.	Oppy
16 th	Officer	Bray
	Leutnant	

Tabel 3

Dette er en del instanser, den ikke finder, og disse ligger på linje med de instanser, ontologien rent faktisk finder. Ydermere har AeroSWARM problemer med at lave tilstrækkeligt specificerede relationer. Mange af de relationer, ontologien knytter mellem koncepterne, er af en generel karakter, som gør, at sammenhængen mellem koncepterne fremstår uklart. Det må derfor siges, at ontologien fungerer i et begrænset omfang. Man får fornemmelsen af, at AeroSWARM er i stand til at lave noget, der minder om en logisk slutning. Præcis hvordan denne logiske slutning fremkommer, har vi har ikke formået at skaffe indsigt i. Sammenlignet med den

forventning om at ontologier skal være kontekstskabende, så er denne ikke indfriet fuldt ud. Det, vi kan se, er at der mangler en del finjustering af systemets slutningsmekanisme, men hvorvidt denne slutningsmekanisme er en del af Ubot eller AeroSWARM, ved vi ikke. Når AeroSWARM nu er i stand til at fremkomme med det ovennævnte resultat, mener vi som førnævnt, at programmet på nuværende stadie befinder sig et sted mellem ontologilaget og det logiske lag i Tim Berners-Lees kagemodel. Ontologien er med til at skabe en meningsfuld sammenhæng, men ontologien tilfører ikke sig selv ny viden. Det at programmet kan identificere to personer som den samme person, er udelukkende tegn på klassificeringer og regler for koncepternes indbyrdes forhold.

Semantisk web i brugskontekst

Vi har i forrige afsnit lavet en analyse af Ubot for at få en ide om, hvor langt man er nået med hensyn til automatisk meningsberigelse af tekster. Vi forestiller os, at Ubot ligner de programmer som Tim Berners-Lee omtaler, når han forklarer, hvordan han forestiller sig, at meningsberigelse af hjemmesider skal finde sted:

“Instead these semantics were encoded into the Web page when the clinic’s office manager (who never took Comp Sci 101) massaged it into shape using off-the-shelf software for writing Semantic Web pages along with resources listed on the Physical Therapy Association’s site.” (Berners-Lee, 2001)

Berners-Lees vision er at meningsberigelse af hjemmesider skal være noget enhver webmaster kan klare. Vi mener, at Ubot lever op til denne brugervenlighed, da man blot indtaster hjemmesidens URL i Ubot, hvorefter den beriger hjemmesiden med mening, samt finder relationer mellem ordene. Man kan så efterfølgende diskutere, hvor langt Ubot er fra at kunne udføre denne meningsberigelse tilfredsstillende i forhold til Berners-Lees ambitioner.

For at drømmen om det semantiske web kan opfyldes, er det dog ikke nok at meningsberige hjemmesider. Det kræver også, at de søgemaskiner, man benytter, bliver i stand til også at søge i metatekst frem for blot den primære tekst.

Som en søgning foregår nu, søger man direkte i teksten på en hjemmeside, det vil sige, den tekst, man søger i, står i naturligt sprog, men den måde, man søger på, er ved at søge på enkelte ord, som man prøver at udlede en logisk sammenhæng af.

Derimod vil man på Semantisk Web kunne formulere søgningen i natursprog, og det, man søger i, er metasprog, som meningsberiger det naturlige sprog, der findes på en hjemmeside. Dette er den tydeligste og en af de væsentligste forskelle ved søgning på henholdsvis traditionelt og semantisk web set ud fra brugerens perspektiv.

Følgende vil vi overveje, hvordan en søgemaskine på det semantiske web vil fungere i forhold til traditionelle søgemaskiner.

Da vi i afsnittet *Informationssøgning i dag* gennemgik ottetrinsmodellen, nåede vi frem til, at søgning på nettet i dag kan være en lang og besværlig proces, hvor brugeren skal gøre et stort stykke arbejde med at sortere i store mængder af ofte irrelevant information. Vi vil i følgende afsnit foretage nogle overvejelser om, hvordan en søgning vil foregå, hvis det semantiske web bliver en realitet. Altså hvordan kan man forestille sig, at en søgemaskine på det semantiske web³ vil fungere, og hvilken indflydelse vil det have på de krav, der stilles til brugeren? Vi vil derfor inddrage ottetrinsmodellen igen, da vi mener, at den kan bruges til at illustrere, hvordan en stor del af arbejdet med at søge og revidere søgningen bliver flyttet fra brugeren til computeren. Vi har ikke nogen endelig oversigt over, hvordan en søgning på det semantiske web vil foregå, men nedenstående afsnit illustrerer, hvordan vi forestiller os, at den ideelt set kunne foregå, når man sammenligner ottetrinsmodellen med Berners-Lee's (2001) lagkagemodel.

For overskuelighedens skyld præsenteres ottetrinsmodellen igen nedenfor:

1. Find ud af, hvad der skal søges efter. (Read the query)
2. Find begreber, der dækker det, der skal søges efter. (Identify the major concepts in the query)

³ Benævnes fremover; en semantisk søgemaskine

3. Find ord, der passer på begreberne. (Identify potential terms to correspond to those concepts)
4. Find alternativer til ordene. (Select alternative terms to use if the original strategy needs help)
5. Find logiske sammenhænge imellem ordene. (Determine logical relationships between terms)
6. Udfør søgningen. (Begin the search)
7. Se nogle af søgeresultaterne igennem. (Have a look at a few documents)
8. Revidér søgningen ud fra resultatet. (Revise and refine the search based on those initial results)

Vi mener ikke, det vil give mening at sammenligne ottetrinsmodellens enkelte trin med lagkagemodellens enkelte lag, bl.a. fordi ottetrinsmodellen er en proces og lagkagemodellen er en arkitektur. Vi vil i stedet betragte ottetrinsmodellen som et udtryk for, hvad en semantisk søgemaskine ideelt set skal være i stand til. Når brugeren har indtastet et spørgsmål i naturligt sprog skal den semantiske søgemaskine altså kunne formalisere dette spørgsmål. Denne formalisering sker, for at søgemaskinen kan matche forespørgslen med mulige resultater. En del af denne formalisering går ud på at tage ordene i forespørgslen med XML. Ydermere skal den kunne finde ud af, hvilken mening ordene og begreberne i spørgsmålet har. Dette gøres ved, at der tilknyttes en URI til de enkelte tags. (trin 1,2,3).

Den skal også kende alternativer til ordene, så den f.eks. på en søgning om graviditet ved, at dette er det samme som eksempelvis frugtsommelighed (trin 4). Den skal med andre ord have kendskab til synonymy. Teknisk set benytter man sig f.eks. af RDF-skemaer, som rummer muligheden for at beskrive forholdene mellem de enkelte egenskaber og ressourcer.

Den skal kunne finde de korrekte logiske sammenhænge mellem ordene (trin 5). De tekniske arbejdsfelter på dette trin knytter sig til *logic*-laget i lagkagemodellen. Computeren skal her være i stand til at lave overvejelser i stil med eksemplet med jaguaren, hvor søgemaskinen ud fra ordet "løber" skal kunne drage den logiske slutning, at jaguaren er et dyr og ikke en bil. Når søgemaskinen har alle disse forhold på plads, vil den ideelt set være i stand til at finde alle de relevante resultater (Trin 6). Dvs. at den vil have opfyldt kravet om *recall*, i forhold til de hjemmesider søgemaskinens har i sit indeks, men ikke nødvendigvis *precision*, da nogle af tek-

sterne kan omhandle det rigtige emne, men være irrelevante i kraft af useriøsitet eller lign. Dette problem løses i trin 7 og 8, hvor den tjekker søgeresultaterne for relevans ved at bruge de øverste lag *logic*, *proof* og *trust* i lagkagemodellen. Disse tre lag bruger logik til at udveksle beviser på påstande eller udsagn fra websider. Nogle sider vil applikationen automatisk have tiltro til (Li, 2002).

Set fra brugernes perspektiv vil den semantiske søgemodel altså kun bestå af 2 trin: Brugeren skal skrive en forespørgsel og derefter udvælge de bedste resultater. De resterende trin er blevet flyttet fra at være overvejelser, som brugeren skal foretage, til at være processer, der foregår i den semantiske søgemaskine.

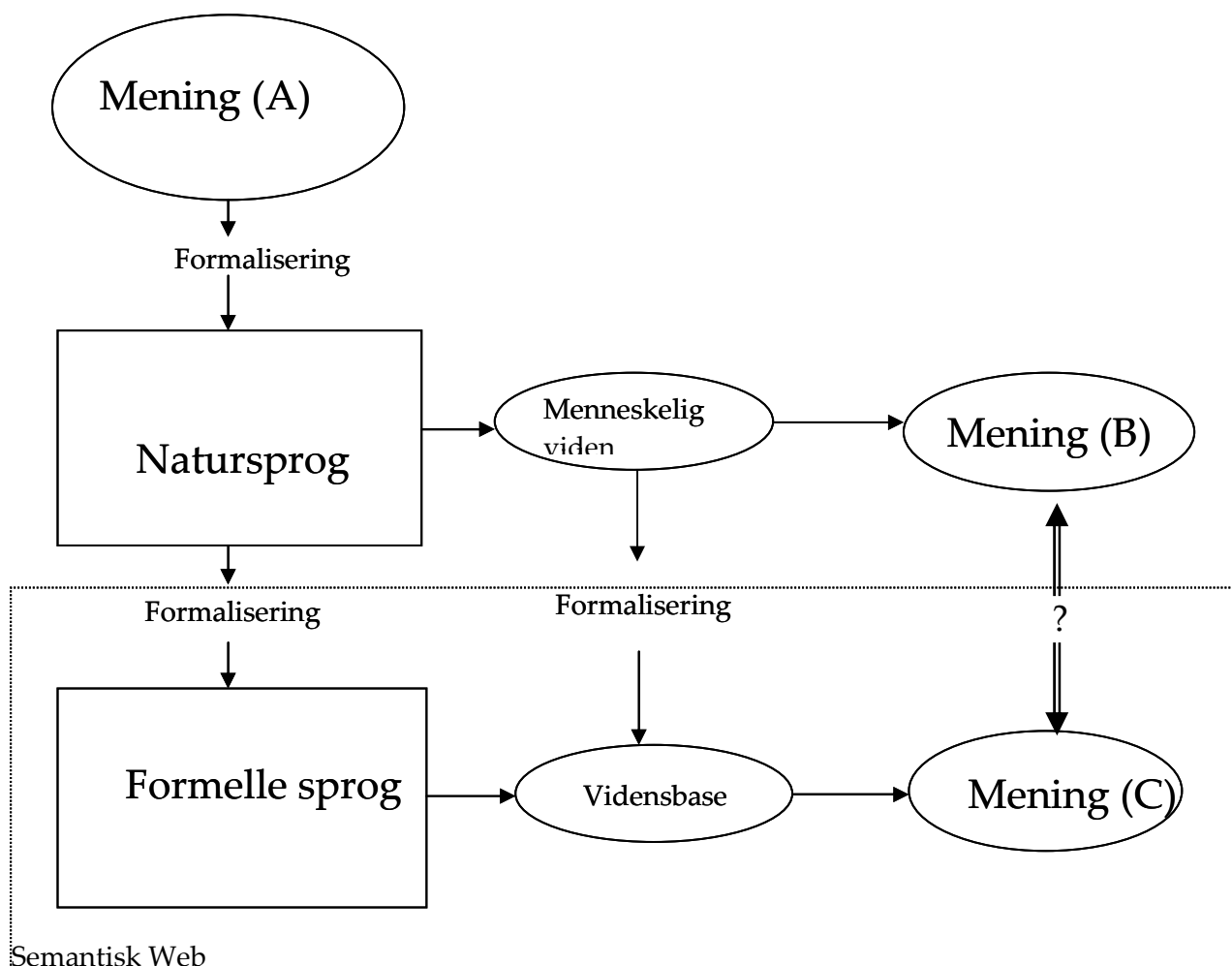
Det er dog ikke nødvendigvis sikkert, at en søgemaskine, som kan reducere brugers rolle til at blot at skulle indtaste søgesætninger, vil være den mest optimale eller realistiske løsning. Vi mener ikke, man kan give fuldt udtryk for, hvad man søger i en enkelt formulering, hverken på skrift eller i tale, fordi dette udtryk altid kun vil være en repræsentation af det, man tænker på (Argumentation for dette vil vi komme nærmere ind på i næste afsnit). Derfor behøver man heller ikke at se det som et ideal at stræbe efter, at en søgemaskine skal kunne finde *alle* relevante dokumenter og *kun* relevante dokumenter ved en enkelt søgning. Et mere realistisk bud kunne være, at brugeren i højere grad skal finde de relevante resultater gennem interaktion med computeren.

Man kunne forestille sig, at når søgemaskinen leder i sin ontologi og opdager, at "Jaguar" er et flertydigt ord, så stiller den nogle afklarende spørgsmål i form af: Når du skriver "Jaguar", mener du så "Kattedyr" eller "Bil"? Desuden kunne den semantiske søgemaskine måske hjælpe med at konstruere spørgsmål, rette stavning (dette kan mange søgemaskiner allerede) m.m. Dette vil sikkert også blive nødvendigt i forhold til, at spørgsmålene, selvom de skrives i naturligt sprog, sandsynligvis skal skrives i en bestemt grammatisk form, så søgemaskinen bliver i stand til at formalisere dem. I det hele taget er det sandsynligt, at det interface, som vi kender fra traditionelle søgemaskiner i dag, vil blive ændret radikalt, hvis den "semantiske søgemaskine" bliver en realitet, og at det i langt højere grad vil blive muligt for brugeren at indgå i en interaktion med søgemaskinen.

Vi vil i diskussionsafsnittet diskutere, hvornår forståelse opstår, om computeren forstår, og hvad forståelse i det hele taget er. Hvis ovenstående bliver en realitet vil forståelsen netop opstå gennem interaktion mellem menneske og computer. Computeren vil ikke nødvendigvis selv forstå, men den vil være med til at skabe en forståelse for brugeren. I og med at computeren gennem et spørgsmål som "mener du *kat* eller *bil*?" bliver i stand til at udføre en mere korrekt søgning, kan man hævde, at computeren har fået en større forståelse for spørgsmålets indhold eller mening gennem interaktion med brugeren.

Diskussion

For at illustrere vores opfattelse af problemfeltet, har vi udfærdiget nedenstående model. Den tager i vid udstrækning udgangspunkt i den argumentation, vi har gennemført i de foregående afsnit, men kræver stadig en forklaring, som vi vil gennemføre umiddelbart under modellen. Herefter vil modellen danne afsæt for den efterfølgende diskussion og vil have status af den konceptualisering af vores syn på problemfeltet, som vi vil have i baghovedet gennem hele diskussionen.



Figur 6

Til grund for modellen ligger den antagelse, at der findes en mening A, bag hvilken der ligger en intension. Denne intension formaliseres i natursprog, som gennem en menneskelig viden eller *common sense* tolkes til mening B. En lignende proces foregår i den boks, vi har kaldt Semantisk Web. Som tidligere gennemgået er ideen her, at man vil formalisere natursproget, således at det kan bearbejdes elektronisk ved hjælp af en vidensbase. Denne vidensbase er et forsøg på at formalisere menneskelig viden eller *common sense*. Der er altid forskelle mellem A, B og C, da A er den oprindelige mening, og mening, som er blevet repræsenteret, aldrig vil være den samme mening, som den var oprindelig. Dette har vi tidligere beskrevet i afsnit *Vidensrepræsentation*.

Det er en pointe i de foregående afsnit, at C aldrig vil være den samme som A, da ingen repræsentation kan rumme "hele verdens rigdom" jf. Sowa. Men er B heller ikke det samme som A? Når vi til daglig snakker om at skabe mening eller betyd-

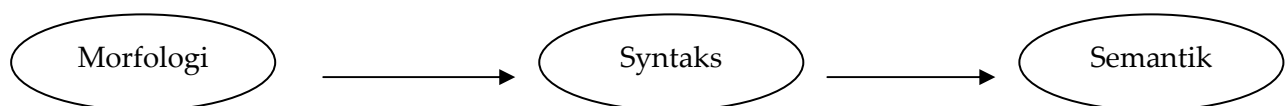
ning, så omtaler vi B, der heller ikke er A, men er en repræsentation af A. Det, der formaliseres i de forskellige sprog, og som bruges i forbindelse med semantisk web, er heller ikke en formalisering af A, men derimod en formalisering af en formalisering af A, der fortolkes gennem en formalisering af *common sense* eller menneskelig viden.

Vi vil mene, at B og C umiddelbart er forskellige meninger, da den proces, der sker for at nå til mening C, er en repræsentation af den proces, der sker for at nå til B. Men visionen bag vidensrepræsentation er at repræsentere menneskelig viden i en vidensbase, således at mening C bliver den samme som B.

Derivationsmodellen

For begge vandrette processer, i henholdsvis natursprogene og de formelle sprog, findes modeller, der er sammenlignelige. Disse modeller starter fra mindste instanser i syntaksen, går derefter til sammenhængen inden for syntaksen og slutter ved det betydningsskabende.

Inden for lingvistikken arbejder man med teorien om derivation. Derivation sættes lig med afledning. Modellen beskriver hvorledes, man går fra de mindste bestanddele, i form af de første ord, til at sætte ord sammen for til sidst at skabe betydning:



(enkelte ord → ord sammensætning → betydning)

Begrebet semantik er så centralt for den efterfølgende diskussion, at vi her vil forklare, hvad der kan forstås ved begrebet, og hvordan det kan forstås i forbindelse med semantisk web.

Semantik

Semantikken er studiet af mening kommunikeret gennem sprog og er tæt forbundet med filosofien og psykologien (Saeed, 2002). Videnskaben om semantik omfatter således mange forskellige teorier og dækker mange områder. Desuden findes der mange forskellige retninger indenfor feltet, og det er særdeles vanskeligt at give en entydig definition. Men i hvert fald handler semantik om, hvordan mennesker udleder mening ved hjælp ikke bare af tegn, men også af kontekst, modalitet, præsuppositioner m.v. Der er altså et kognitivt aspekt i ordet. Når vi vil diskutere hvornår og hvorvidt, semantik opstår på det semantiske web, taler vi ikke længere om teknologien, men om semantik som noget konkret og som resultatet af den proces, man studerer indenfor semantikken. Der må altså have fundet en kognitiv proces sted eller i hvert fald en meningstilskrivelse, der inddrager konteksten og "ekstern viden", før noget kan beskrives som semantisk. Denne "eksterne viden" kan have form af forskellige teorier, det kan være logik eller det, vi tidligere har kaldt *common sense*.

I teorien om derivation lægges der vægt på, at det er det sproglige sammen med en situation, samt de grammatiske regler, der knytter sig til situationen, der skaber betydningen.

Semantik + situationen = betydning

Hvis man f.eks. sidder ved middagsbordet, og én spørger: "Kan du nå saltet?" – Personen, der spørger, kan tydeligt se, at vedkommende, der spørges, kan nå saltet, men mener: "vil du række mig saltet?". Spørgsmålet kan i dette eksempel kun forstås fordi, der knytter sig en bestemt kontekst til netop denne situation (Glyngø, 2001).

Sammenligner vi ovenstående derivationsmodel med lagene i Berners-Lees lagkagemodel, kan man trække visse paralleller. Lagkagemodellen starter i de mindste bestanddele. Disse mindste bestanddele findes i form af tags, URI og grundkode. Det nederste lag kan vi sammenligne med morfologien. Her er der ingen struktur over de enkelte tags. Denne struktur bliver først sat på plads i det næste lag. Dette lag indeholder RDF og RDF-skemaer og er en syntaks for de mindre bestanddele. I

laget med ontologi træder man så ind i området for semantik eller betydningsdannelse (Park & Hunting, 2002).

Mening og betydning

Hvis vi ser på de to ovenstående modeller, opstår der problemer, som ikke kan forklares direkte med sproglige discipliner eller regler netop omkring det område, hvor vi skal beskrive, hvordan betydning eller mening tilskrives en sætning. I begge arbejdsfelter (lingvistik og det semantiske web) er der fokus på, at det er i samspillet mellem det sproglige og konteksten, at betydningen opstår. Kontekstproblemet er inden for det semantiske web bl.a. forsøgt løst med ontologier. Men kan man hævde, at der opstår mere forståelse ved indførelsen af ontologier? Godt nok vil det med stor sandsynlighed løse problemer som f.eks. problemet med flertydige ord. Her kan computeren nu søge i taksonomier over konteksten og dermed med stor sandsynlighed løse problemet. Men tager vi fat i ovenstående eksempel med udsagnet "*Kan du nå saltet?*", så er det vel ikke nødvendigvis at betragte som et tegn på meningsforståelse, hvis computeren via klassificering og systematisering vælger den rigtige betydning af sætningen. I hvert fald ikke hvis vi går ud fra den form for betydningsdannelse, som vi tænker på i forbindelse med natursprog. Ud fra Peirce's treleddede tegnbegrebsmodel har vi endvidere en anskuelse om, at forståelse er en kognitiv proces, der opstår individuelt, og at vi hver især tilskriver en betydning. Mens "betydningstilskrivningen" i det semantiske web vil blive et spørgsmål om, hvad der er sandt og falsk i forbindelse med de ontologier eller lignende, som en applikation benytter sig af. Denne betydningstilskrivelse er altså universel, fordi den samme tekst altid vil blive tilskrevet den samme mening. Derfor giver det ikke nødvendigvis mening at tale om en kognitionsproces i forbindelse med det semantiske web og heller ikke om egentlig semantik ud fra den forståelse.

Kan computeren forstå?

I diskussionen, om hvorvidt en maskine kan opnå forståelse, opsætter Searle (1980) tankeeksperimentet "Det kinesiske rum." I eksperimentet fremsættes en tese om, at en computer meget vel kan give et svar, som virker meningsfyldt, men af den grund

har computeren ikke nødvendigvis opnået nogen form for forståelse. Eksperimentet går ud på at give en person alle de instruktioner, der skal til for at udføre en bestemt type opgaver og dernæst give sig ud for at være en maskine med intelligens. Personen skal ikke vide, hvad indholdet af informationen er, men kun formen, det skal behandles på. Derved mener Searle, at bare fordi en maskine kan løse nogle opgaver, som får den til at se ud som om, den forstår informationen, så har den ikke nødvendigvis opnået forståelse af det, når nu en person kan udføre opgaverne uden at forstå dem (Searle, 1980).

Ud fra dette argument kan man mene, at det er forkert at sige, at en søgning på det semantiske web frem for det nuværende web foregår på baggrund af en forståelse af, hvad der søges på. Det er nærmere en simulation af forståelse. Samtidig er det semantiske web i stand til at tilskrive betydning på en måde, som det konventionelle web ikke er i stand til.

Spørgsmålet er, om sådan en sammenligning er særligt relevant i forbindelse med semantisk web. Når man bruger ordet semantik, hvor det knyttes til en menneskelig proces, rejser det et problem, da vi forsøger at bruge en menneskelig natursproglig kognitionsforståelse til at tilskrive forståelse til en computer. En af de spændende refleksioner omkring semantisk web er at finde ud af hvor og hvordan, semantik opstår i forbindelsen med teknologien. Ovenstående argumentation har modstridende perspektiver, da opfattelsen af ordet semantik er noget, der tilskrives den natursproglige disciplin, som en computer ikke benytter. Lad os tage det eksempel vi nævnte i afsnittet om informationssøgning. Hvis computeren kan få ud af spørgsmålet "Hvor hurtigt løber en jaguar", at der er tale om en kat og ikke en bil, kan man sige, at når computeren forstår, at jaguaren er et dyr, uden at ordet "dyr" forekommer i teksten, men derimod ud fra ordet "løber", foregår der en semantisk proces i computeren, som jo tilskriver teksten en mening, som ikke fremgår eksplicit. Denne meningstilskrivelse er i højere grad mulig på det semantiske web end på det traditionelle web. Som vi beskrev i forrige afsnit, forestiller vi os, at man med en semantisk søgemaskine kan søge på hele sætninger skrevet i naturligt sprog. Her bliver søgeordene sat i en kontekst i form af en hel sætning i modsætning til en traditionel søgemaskine, hvor de enkelte søgeord ofte står for sig selv. Som vi har beskrevet ovenfor, er det konteksten, som skaber betydning. Det semantiske web har netop denne kontekst at trække på, som er udledt af menneskelig viden jf. model-

len. Når den semantiske søgemaskine skal formalisere et spørgsmål og finde relevante dokumenter, bruger den de informationer, den kan finde i vidensbasen. De traditionelle søgemaskiner kan derimod ikke gøre brug af denne kontekst. De ord, som brugeren indtaster, har stort set ingen anden betydning for computeren, end at den kan finde ud af at matche dem med de samme ord i en søgetekst.

Det er svært at udlede mening af enkelte ord, og det må derfor siges at være en forbedring, hvis man kan søge på hele sætninger. Det kan dog stadig diskuteres hvor meget mening, der kan udledes af en enkelt sætning. Selvom den semantiske søgemaskine har en sætning at danne kontekst ud fra, er dette stadig en meget begrænset kontekst i forhold til de tekster i søgegrundlaget, som har mere indhold end enkelte sætninger og derved en større kontekst at danne mening ud fra. Ud fra denne betragtning kan man også diskutere, om der opstår et misforhold i at skulle matche en søgesætning med en hel tekst, når nu de ikke har samme grundlag for meningsdannelse.

Problemet med om en enkelt sætning kan skabe stor nok kontekst er dog stadig et mindre problem i forhold til det problem, det er, at få computeren at forstå en tekst.

Hvis computeren bliver i stand til at tilskrive semantisk betydning til en tekst, og der således, i hvert fald ud fra vores definition, kan påstås at opstå mening, er det så ensbetydende med, at der også er forståelse? Og hvad med den kognitive proces, der før blev nævnt som forudsætning for semantik?

Grænser for formalisering

Skal man komme nærmere "computerforståelse", medfører det en problematisk handling i formaliseringsprocessen, der er indsat mellem "menneskelig viden" og "vidensbase" i vores model. Når man bygger vidensbaser, er det da virkelig "viden", man repræsenterer? Ofte tager repræsentationen en omvej af natursproget. I ontologier for eksempel er subtyperelationen måske nok en repræsentation af menneskelig viden om typer, men typerne defineres, uanset om man benytter naturlige sprog eller formelle sprog til at notere definitionerne, gennem begreber fra naturlige sprog. For eksempel er konceptuelle grafer et formelt sprog, men er baseret på det

naturlige sprog (Sowa, 2000, p. 476), da det er et metasprog herover. Dermed bliver *Conceptual Graphs* en repræsentation af den repræsentation, som natursprog jo også er. Men er denne "metaformalisering" tilstrækkelig. Den er jo ikke en formalisering af kognitionsprocessen. En diskussion af, hvorvidt det er nødvendigt at formalisere denne kognitionsproces (hvis funktion ingen vel kan påstå at have fuldstændigt kendskab til) lægger sig tæt op ad diskussionen af Searles kinesiske rum. Man kan så finde denne diskussion interessant eller ej. Et andet interessant problem er, at så længe de formelle sprog, der bruges til vidensbaserne, er funderede i natursprogene, må det forholde sig sådan, at hvis der er grænser for hvilken viden eller mening, der kan udtrykkes i natursprog, gør disse grænser sig også gældende for de formelle sprogs vedkommende.

Man kan hævde, at ovenstående ikke er et problem i forhold til semantisk web, og at viden, der ikke kan udtrykkes i normalsprog, ikke er relevant, når man søger på tekster, der netop er skrevet i natursprog. Men også natursprog kan afkræve viden af læseren, der ikke umiddelbart lader sig udtrykke. Hvis man søger på "ungdomskultur i firserne", skal der megen baggrundsviden til for at finde følgende:

”Vi er forkomne englebørn
med vinger af fremtidssang,
med barnet i blodet og smøgen i kæften.
Vi har hud af den sarteste drøm
og hjerter der lyser mere end neon.
Vi er kvæstede af dagens skarpe lyde,
blødende lyserød sne,
spiddet af avisoverskrifter.
Vi er en del af Natmaskinen
transformerer angst til venskab.
Vi bærer vore hjerner med stolthed
bytter drømme og cigaretter,
fylder os med rus og musik
bytter køn og masker...”

(Strunge 1981)

Selvfølgelig er det muligt, at der selv af en tekst, der er så pakket med metaforer og analogi, kan tolkes et enkelt emneord eller en betydning ud fra regler og ontologi, men problemet er, at det kan være svært selv for mennesker at forstå betydningen af f.eks. lyrik. Og det, at forskellige mennesker udleder forskellige meninger, kan være særligt udpræget, netop når det kommer til litteratur. Spørgsmålet er så også, om man overhovedet ønsker, at en maskine skal kunne fortolke tekster af denne karakter. En meget væsentlig del af det at læse lyrik er jo netop at skabe sin egen tolkning.

Desuden er der fremført den kritik (Glasgow, 1993 og Sowa 2000, p. 399f), at menneskeligt ræsonnement ikke kun baserer sig på sprog og logik, men også på billeder og symboler. Som eksempel giver Birch spørgsmålet: ”Do German shepards have triangular-shaped ears?”. Dette spørgsmål kan umiddelbart besvares med et ”Ja”, men Birch har empirisk konstateret, at de fleste mennesker hævder, at de når dette svar ved at se på en intern, billedlig repræsentation af dyret. Disse repræsentationer har grund i erfaring; typisk noget konkret fysisk og altså repræsenterbart, og ovenstående eksempel kan da også besvares ved, at en ontologi definerer en schäfer, som

havende trekantede ører. Men tanken om, at dele af menneskeligt ræsonnement ikke er funderet i sproget, men er repræsenteret ved et "billede i hovedet" i stedet for et ord, åbner mulighed for, at der kan eksistere menneskelig viden, der ikke kan repræsenteres efter traditionelle principper.

Vi har i ovenstående kredset om de to begreber, "betydning" og "forståelse", uden klart at skelne imellem dem. Vores problem ligger i at skille de to begreber fra hinanden. Giver det mening at tale om, at computeren kan skelne mellem forskellige betydninger, hvis den ikke har nogen forståelse, og omvendt *hvis* computeren er i stand til at afgøre en sætnings betydning, *medfører* dette så forståelse? Man kan hævde, at denne diskussion er ligeegyldig. Hvis computeren ved hjælp af ontologier og syntaktiske regler bliver så "dygtig", at den kan give korrekte tilbagemeldinger på enhver forespørgsel og afgøre den korrekte betydning af sætninger i lige så høj grad som mennesker, så er der vel ingen grund til at diskutere, hvorvidt computeren har forståelse eller ej. Grunden til, at vi alligevel beskæftiger os med problemstillingen om forståelse, er at vi tidligere har påpeget det som et problem, at formaliseringen fra "menneskelig viden" til "vidensbaser" tager en omvej ad natursprogene, der lægger nogle begrænsninger i forhold til, hvad der lader sig formalisere. Det, vi her har villet påpege, er de mange problemer, der vil være forbundet med at komme ud over dette problem; det vil sige realisere den formaliseringsproces, der i modellen er placeret mellem menneskelig viden og vidensbaserne; en form for menneskelig kognitionsproces eller det, der "skaber semantik".

Selvom alt menneskelig viden bliver formaliseret, vil der opstå et problem med at komme fra vidensbasen videre til mening i det semantiske web, fordi denne behandling af viden er underlagt en computers logiske behandling af data. Ifølge Gödels Incompleteness Theorem, som siger: "*The implication is that all logical system of any complexity are, by definition, incomplete theory; each of them contains, at any given time, more true statements than it can possibly prove according to its own defining set of rules.*", vil der i ethvert logisk system være sande påstande, som ikke kan bevises ud fra systemets egne regler (Internet 9). Hvis den formaliserede viden skal repræsentere alt viden også logik, vil der altså være nogen sandheder, der ikke kan udledes af denne vidensbase. Ud fra denne betragtning ser vi også et problem med det logiske lag i lagkagemodellen over det semantiske web, da dette lag skal udlede logiske

sammenhænge ud fra ontologien i laget under. Berners-Lee er selv inde på denne problematik:

”The logic must be powerful enough to describe complex properties of objects but not so powerful that agents can be tricked by being asked to consider a paradox.”(Berners-Lee, 2001, p. 38).

I vores analyse af Ubot har vi påpeget, at Ubot har visse problemer med at udlede simple relationer af en tekst, men selvom disse fejl bliver rettet, viser Gödels bevis, at Ubot aldrig vil være i stand til at udlede alle sandheder af en tekst, uanset hvor avanceret dens ontologi er.

For at det semantiske web skal kunne opføre sig meningsfuldt, er det altså ikke nok med formelle sprog, syntaktiske regler og ontologier. Der mangler en eller anden form for forståelse. Og så er vi tilbage ved spørgsmålet om hvornår eller hvorvidt, man kan hævde, at en computer forstår.

Forståelse i interaktionen

I diskussionen tidligere kom vi ind på, at computeren ikke kan opnå forståelse, men kun simulere forståelse. For det, den gør, er bare at behandle data, som den ikke kender indholdet af, men kun formen, det repræsenteres på. Ud fra en brugers perspektiv er det derimod ikke så relevant, om computeren reelt ”forstår”. Hvis den kan give de rigtige svar med lige så stor nøjagtighed som en person, der reelt forstår, så er det vel ikke vigtigt, om den har forståelse i menneskelig forstand. Det vigtige er, om interaktionen mellem menneske og computer er lige så forståelseskabende, som interaktionen mellem to mennesker. Hvis man anskuer menneske-computer-interaktionen som et menneske, der udtrykker et behov, og en computer, der kommer med et løsningsforslag, og hvis vi har den indstilling, at en computer kun er et system, som gør, hvad den har fået besked på, så er denne situation vel ikke udtryk for mere interaktion, end når man benytter det system, der sørger for, at toilettet skyller ud. Spørgsmålet er, om der opstår betydning i interaktion mellem computer og menneske. Vi vil i et forsøg på at besvare dette spørgsmål igen ende ved, at der kun kan opstå forståelse i kognitiv forstand inde i mennesket, så evt. betydning og meningsdannelse vil opstå i det menneske, der bearbejder outputtet fra computeren og ikke udspringe af selve interaktionen. Alligevel vil mennesket,

hvis det semantisk web fungerer optimalt, få de samme informationer, som i interaktion med et andet menneske. Hvis denne situation opnås vil det derfor ikke være nødvendigt at prøve at få computeren til at forstå i en kognitions-mæssig forstand. Spørgsmålet er så bare, hvorvidt man kan nå dertil uden at finde frem til, hvordan forståelse opstår. Altså hvordan kan vi få computeren til, i bedste fald, at simulere forståelse uden at vide, hvordan forståelse opstår, og videre hvordan man formaliserer denne forståelsesproces. Desuden er der dem, der vil hævde, at forståelse i mennesket ikke adskiller sig fra den måde, en computer "forstår". Altså at menneskers forståelse i bund og grund også blot er udtryk for avancerede regneprocesser (Johansen, 2003, p. 43).

Værktøj eller interaktionspartner

Hele diskussionen, om hvorvidt computeren forstår eller ej, kan være lidt misvisende i forhold til det semantiske web, som ifølge Berners-Lee (2001) ikke er et forsøg på at udvikle stærk kunstig intelligens: "it will "know" all this without needing artificial intelligence on the scale of 2001's Hal or Star Wars's C-3PO."

I bund og grund ville det semantiske web være et enestående værktøj til hurtigt at overskue meget store mængder af information, hvilket er et stort problem med de søgeredskaber, vi har i dag.

Netop spørgsmålet om, hvorvidt det semantiske web skal være i stand til at efterligne de menneskelige forståelsesprocesser eller om det blot et værktøj, har bevirket, at vores diskussioner i dette projekt ofte kommer til at køre i ring. Ud fra ovenstående kan man udlede to argumenter, der ikke udelukker hinanden, dog giver de samtidig ikke noget endegyldigt svar. De to argumenter er:

- En computer vil aldrig få nogen forståelse. Hvis den giver udtryk for forståelse, så er dette kun et udtryk for velstrukturerede taksonomier og logik. Computeren er altså kun et værktøj uden forståelse.
- Hvis Semantisk web fungerer på en måde, så det kan simulere en interaktionspartner, som giver svar, der er fuldt ud lige så meningsfulde som en menneskelig interaktionspartner, er det så ikke lige meget, om den forstår?

Diskussion ud fra disse to standpunkter ender ofte i en cirkulær diskussion(argumentation) i stil med følgende:

Hvis en computer giver udtryk for at den har opnået forståelse, er det så simuleret forståelse eller er det rigtig forståelse (hvad er egentlig rigtig forståelse og kan man formalisere noget man ikke forstår fuldt ud?). Er det nødvendigt at diskutere hvorvidt computeren forstår, hvis bare det fungerer, altså at den giver fyldestgørende svar. Spørgsmålet er så hvorvidt det kan lade sig gøre, at give fyldestgørende svar på en brugers søgning uden rigtig forståelse.

Kan man hævde at computeren blot er et værktøj uden forståelse eller intelligens, hvis den er i stand til at give et meningsfyldt svar på en forespørgsel og derved fremstå som værende intelligent? Eller kunne det tænkes, at forståelsen opstår i interaktionen, så computeren i stedet for selv at have forståelse blot bidrager til brugerens forståelse gennem en interaktion? Og så er man tilbage ved spørgsmålet om, hvorvidt computeren kan bidrage som interaktionspartner og derved bidrage til forståelse, hvis den ikke selv har forståelse eller intelligens.

Konklusion

Dette projekt har taget afsæt i en analyse af søgesituationen i dag, hvor vi har belyst nogle af de problemer, som eksisterer med informationssøgning. Internettets enorme vækst har medført, at brugeren kan have svært ved at finde relevante informationer, specielt uden at der kommer en stor mængde irrelevant information med i søgeresultaterne. En af grundene til disse problemer er, at de søgeredskaber, der findes i dag, søger på form og ikke på indhold.

En mulig løsning på dette problem kan være det semantiske web. Vi har beskrevet en del af den teknologi, som ligger til grund for, at det semantiske web kan fungere. I en analyse af et konkret værktøj (AeroSWARM) har vi beskrevet, hvor langt udviklingen af denne del af teknologien er kommet. I beskrivelsen af det semantiske web fremgår det, at der yderligere er en del teknologier, der skal udvikles, før alle kriterierne for Berners-Lees idé om det semantiske web er opfyldt. Dette understøttes i analysen af eksisterende ontologier på Internettet, da vi her redegør for kvaliteten af det output, AeroSWARM producerer. Vi finder frem til, at AeroSWARM er i stand til at danne bestemte betydningssammenhænge mellem koncepter, samt inddele koncepter og instanser i en relation til abstrakte koncepter, der udgør fundamentet for meningstilskrivelse. Ydermere finder vi frem til, at der er problemer med at præcisere disse relationer, såvel som at der er problemer med at lave helt korrekte inddelinger. Med udgangspunkt i denne analyse har vi diskuteret nogle af de problemer, der kan være forbundet med udviklingen af det semantiske web. Blandt

disse problemer er problemet med flertydige ord, som kræver en større kontekstforståelse, for at disse kan kategoriseres korrekt.

Hvis det semantiske web bliver en realitet, er der sandsynligvis udsigter til noget bedre end det traditionelle web bl.a. fordi det semantiske web vil have en vidensbase at trække på, som vil hjælpe med at skabe den kontekst, som er afgørende for den semantiske betydning. Uanset om computeren forstår i menneskelig forstand, så vil en realisering af det semantiske web potentielt skabe væsentligt bedre forhold for brugeren. Af de mulige forhold, der kan løses for brugerne, er blandt andet problemerne med flertydighed og for meget irrelevant materiale.

Desuden forestiller vi os, at nogle af de overvejelser, man går igennem i en søgesituation illustreret ved ottetrinsmodellen, vil blive varetaget af søgemaskinen. Dette vil muligvis ikke ske fuldt automatisk, men gennem interaktion, hvor søgemaskinen spørger efter uddybning ved f.eks. flertydighed i en forespørgsel.

I en diskussion af de resultater vi har fundet frem til i henhold til analysen af AeroSWARM, behandles spørgsmålet om, hvad der skal til, før det semantiske web bliver en realitet, eller om det i hele taget er mulig at lave et reelt semantisk web i den forstand, at maskinen kan blive i stand til at "forstå" mening. Her argumenterer vi for, at det højeste krav, vi kan stille til maskinerne, som vi kender dem i dag, er, at de i en informationssøgningskontekst vil blive i stand til at give en meningsfuld tilbagemelding på en forespørgsel. En sådan meningsfuld tilbagemelding vil i praksis implicere, at man på en eller anden måde må formalisere menneskelig viden eller *common sense*. I diskussionen påpeger vi en række problemer med denne formalisering:

I vores model har vi illustreret et syn på, at meningsberigelse af semantisk web i henhold til menneskelige forståelsesprocesser kan ses som en parallel proces. I diskussionen af denne model kommer vi ind på nødvendigheden af at gøre det klart, at det i øjeblikket ikke er mening, man repræsenterer direkte. Ontologier og vidensbaser er derimod en metarepræsentation i form af en formalisering af den formalisering, som også natursproget er. De grænser, der gælder for natursprog, gælder altså også for vidensbasen, så længe den bruger formelle sprog, der er funderet i de naturlige sprog. Der er altså mulighed for, at der er dele af menneskelig kognition, der

ikke kan repræsenteres af den nuværende vej, det vil sige gennem en sproglig formalisering.

Vil det være muligt at formalisere den menneskelige viden (common sense), der er nødvendig for at realisere det semantiske web, så længe formaliseringsprocessen skal en omvej ad sproget?

Med hensyn til hvorvidt det semantiske web kan blive semantisk, finder vi i diskussionen frem til, at begrebet semantik dækker over det at udlede mening fra natur-sprog. Analysen af AeroSWARM viser, at man allerede har taget de første skridt mod en realisering af denne meningsudledning. Det er et spørgsmål, om det bliver et problem, at det ikke umiddelbart synes muligt at formalisere al menneskelig kognition ad den omvej, som sproget er? Og følger man tanken om at dette er muligt, så peger teoretikere på at kognition er en individuel proces. Forskellige mennesker vil udlede forskellige meninger af den samme tekst, hvorimod den centrale vidensbase, som for eksempel Cyc, vil nå en og kun en mening, hvis der er konsistens i basen.

Dermed kan man stille spørgsmålet, om hvorvidt det er en konsistent database, der skal til, for at vi har et semantisk web, som vil løse vores søgningsproblem?

Yderligere kan vi stille spørgsmålet om hvorvidt, "Det semantiske web" er en passende titel at give den teknologi, der er under udvikling?

Temarammen

En del af kravet til temarammeredegørelsen er at forholde projektet og dets problemstilling til de på semesteret afholdte kurser. Følgende afsnit vil kort forklare, hvordan problemstillingen forholder sig til temarammen og vil senere tage form af uddybning af forholdet mellem det enkelte kursusforløb og projektet. Det, at denne del af temarammeredegørelsen er henlagt til slutningen af rapporten, skal ikke tages som udtryk for, at vi tillægger den mindre betydning end det øvrige indhold, men blot for, at den ikke umiddelbart falder naturligt ind i andre sammenhænge. Afsnittet vil reflektere over projektets sammenhæng med alle kurser, der afløses ved projektsamen, således også kurser, hvor undervisningen og kursuslitteraturen ikke har fundet direkte anvendelse i dette projekt, men som alligevel indgår, som en del af den helhed semestret har skabt i form af en forståelsesramme, der uundgåeligt har præget projektet. Desuden vil der være overvejelser om, hvordan kurserne eventuelt kunne have bidraget med andre vinkler eller perspektiver end dem, vi har valgt.

Vores projekt tager udgangspunkt i temarammen for 5. semester humanistisk datalogi: Edb-systemer; brug og betydning. På semesteret arbejdes der ifølge beskrivelsen af temarammen med: "interaktionen mellem mennesker og informations- og kommunikationsteknologi (IKT) i en brugs kontekst." (http://www.huminf.aau.dk/vejledning/grund_ba.html#38) Brugsiden er vigtig for dette projekt, fordi vi tager vores udgangspunkt i den måde informationsøgning foregår på i dag set fra brugerens synsvinkel. Vi tager således fat om nogle konkrete

problemer, der eksisterer med brugen af Internettet i dag. En del af projektets sigte er at give et bud på, hvordan Semantisk Web kan hjælpe med til at løse de problemer, vi konstaterer i den forbindelse. Vi undersøger et mere fremtidsorienteret brugsperspektiv i vores analyse af den automatiske markup-generator Ubot, som producenter af websider vil kunne benytte til at lave den metatekst, som er essentiel for det semantiske web. På betydningssiden diskuteres mulighederne i det semantiske web i en mere filosofisk kontekst. Vores interesse i den forbindelse er rettet mod de bagvedliggende ontologier og vidensbaser, som af mange, herunder udviklerne af RDF og OWL, betragtes som en forudsætning for at Semantisk Web vil kunne fungere. Betydningsdelen vil bevæge sig ud over brugen. Vi diskuterer hvilken betydning, man kan tillægge den meningstilskrivelse, som Semantisk Web overfor brugeren giver indtryk af. Både brug og betydning spiller en væsentlig rolle, når vi i projektet diskuterer om og hvordan, menneskelige processer automatiseres eller assisteres af "Semantisk Web"-teknologien. Det kan betragtes som en forbedring af brugervenligheden, hvis computeren bliver i stand til simulere eller forstå brugers søgeprocesser og dermed assistere i informationssøgningen. Samtidig kræver det overvejelser af betydningsmæssig karakter, når det kommer til spørgsmålet om, hvorvidt dette overhovedet kan lade sig gøre. Derfor mener vi, at brug og betydning i dette projekt er to uadskillelige sider af samme sag.

Kurserne

Da vi har valgt at skrive om Semantiske Web, så er de kurser, der mest direkte har haft relevans for os, "Kunstig Intelligens" og "Programmeringsmetoder".

Kunstig Intelligens (KI): Mange, blandt andre Sowa (2000) placerer vidensrepræsentation indenfor forskningsfeltet KI, som dette kursus giver en introduktion til. Af alle kurser har dette været projektets væsentligste inspirationskilde, idet det dels har givet os begyndelsen til en forståelsesramme for projektet, dels har bidraget med en teoretisk baggrund. Vi har benyttet kursuslitteraturen herfra, f.eks. Searles eksempel med "det kinesiske rum". Der står i kursusbeskrivelsen, at der lægges stor

vægt på "et nærmere studium af formaliseringens muligheder og begrænsninger", der jo er et af dette projekts centrale fokuspunkter. En enkelt forelæsning beskæftigede sig direkte med vidensrepræsentation, og var, selv om den ikke tog udgangspunkt i Sowa, selvsagt af stor relevans for vores projektemne. I øvrigt har kurset hovedsagelig omhandlet KI's historiske, filosofiske og etiske perspektiver, der udgør en del af den kontekst, som vores projekt befinder sig i på trods af, at vi ikke forholder os til historiske og etiske problemstillinger. Den filosofiske side er meget central i både undervisning og kursuslitteratur, ligesom i dette projekt. Den viden vi har opnået, blandt andet ved læsning af Copeland (1993), har fungeret som en type oversigt over især den filosofisk orienterede teoridannelse indenfor KI som konkret teori og har derforuden været en vigtig inspiration i vores valg af problemområde. En del af kurset har taget form af "de opfattelser inden for AI-forskningen, som tager udgangspunkt i en sammenligning mellem den menneskelige hjerne og computerens potentiale." (kursusbeskrivelse), der jo også har været et vigtigt interesseområde hele vejen igennem vores projekt. De refleksioner, vi har haft i forbindelse med denne del af projektet, har for en stor dels vedkommende været en fortsættelse af de tankerækker, der blev sat i gang af dette kursus.

Kursets historiske vinkel kunne også inspirere mange spændende projekter. En mulighed kunne være at indplacere Semantisk Web i den kontekst eller diskurs, som er forskningen i kunstig intelligens, eller med udgangspunkt i forskningsfeltets historiske udvikling at vurdere om visionen om Semantisk Web er realistisk.

Det samme gælder for kursets etiske vinkel. Dette projekts overvejelser omkring forståelse lægger sig til debatten om stærk kontra svag KI, som kunne have dannet baggrund for en diskussion af det etiske i at tilføje en computer "forståelse". Diskussionen om centrale kontra distribuerede vidensbaser kan også ses ud fra en etisk vinkel. En central vidensbase vil for eksempel altid være udtryk for et bestemt perspektiv, det vil sige, at andre vil kunne have ganske stor indflydelse på, hvilke søgeresultater brugeren får. En af de egenskaber, der ofte fremhæves ved World Wide Web som medie, er jo netop det demokratiske i, at alle dokumenter har samme status. Der er mulighed for, at dette vil kunne ændre sig, hvis dokumenternes relevans i forhold til en bestemt mening fastlægges centralt. Desuden kunne man frygte, at teknologien åbner mulighed for misbrug fra for eksempel virksomheder,

der ønsker flere hits, regeringer, der ønsker at visse meninger ikke kommer til udtryk eller hackere, der af den ene eller anden grund har et ønske om at sabotere systemerne.

Udover en historisk, en etisk og en filosofisk side har kurset en datalogisk side, hvor fokus ligger på konceptuelle grafer, som vi jo kort inddrager, og logikprogrammering. Vi har valgt et andet fokus end logikprogrammering i projektet, men det kunne have stor relevans, idet vidensrepræsentation netop er et af de områder, hvor logikprogrammering kan applikeres. Desuden har vores kendskab til programmering i forhold til KI, givet os en stærkere baggrund for at vurdere de problemer, der er forbundet med programmering af vidensbaser. Kendskab til programmeringsmæssig syntaks har desuden forbedret vores forudsætninger for at kunne forstå XML, RDF, etc.

Programmeringsmetoder: Kursets indhold har haft stor relevans for dette projekt; først og fremmest har vi arbejdet direkte med ontologi på AMINE-plattformen og i forbindelse med logikprogrammering i PROLOG+CG. En enkelt kursusgang havde direkte form af en indføring i vidensbaser og ontologier. Vi blev introduceret til OntoClean, en metode til evaluering af ontologier, som også har givet inspiration til dette projekts problemstilling, idet vi flere gange har overvejet på det ene eller anden måde at inddrage OntoClean. Selv om det ikke er sket, har vi altså gjort os tanker omkring metoden, der har været påvirket af idèen om OntoClean. Også den del af kurset, der beskæftiger sig med objektorienteret programmering, har haft indhold, der ligner ontologibygning. Grundbegreberne *objekt*, *klasse* og *struktur* blev forholdt til den klassiske begrebsanalyse. Dermed bidrog også dette kursus til forståelsen af ontologi.

En væsentlig del af kursets indhold var en indføring i logikken. Dette grundlæggende kendskab viste sig at være meget værdifuldt for at kunne forstå Sowa og anden litteratur om vidensrepræsentation, hvor logik inddrages. På den måde sparede denne introduktion til den grundlæggende logik os formentlig for mange timers læsning.

Det udbytte, vi har haft af kurserne i "Programmeringsmetoder" og "KI", og som vi i høj grad har brugt direkte eller implicit i projektet, lader sig ikke uden videre adskille. Sammenhængen mellem de to kurser har været tæt, idet de har haft en henholdsvis praktisk og teoretisk tilgang til det samme emne og som helhed har dannet den forståelse hos os, som er fundamentet for dette projekt.

Foruden disse to kursusgange, der har haft stor indflydelse på vores projekt, har vi også haft følgende tre kursusforløb.

Menneske Maskine Interaktion (MMI): Diskussioner omhandlende interaktion mellem mennesker og maskiner har været meget centrale for dette projekt. Diskussionerne har imidlertid antaget en overvejende filosofisk karakter og adskiller sig dermed noget fra de "kognitive, kulturhistoriske og kommunikationsteoretiske tilgange", der lægges vægt på på MMI-kurset. Der lægges i forbindelse med kurset vægt på interfacebegrebet, noget der har været uden for fokus i projektet. Vidensrepræsentation, ontologi og sprog er imidlertid kun en del af SW-teknologien. En indførelse af en ny måde at søge på vil sandsynligvis have implikationer for de interfaces søgefunktionerne tilgås igennem. I sin artikel fremlægger Berners-Lee også visioner for Semantisk Webs interface, f. eks. den håndholdte webbrowser. Særdeles relevant og interessant kunne det i den forbindelse være at forholde søgningens udvikling til den udvikling, der er indenfor brugerflader. Kursuslitteraturen har ikke fundet anvendelse i forbindelse med dette projekt.

MM-produktion: Sigtet for dette kursus var at give os et indblik i og en forståelse af en praktisk systemudviklingsproces. Meget af kurset bestod i at skrive pseudo-kode, som så senere blev ændret til Lingo. I dette projekt har vi ikke udviklet noget konkret produkt, ligesom vi heller ikke på anden måde inddrager konkret systemudvikling. Den viden om systemudvikling, som vi har opnået på dette kursus, kunne have været inddraget ved at udvikle et konkret produkt. Det kunne være en applikation der på den ene eller anden måde har forbindelse til semantisk web, eller, mere realistisk, en lille ontologi eller vidensbase. Dette kunne have været en meget interessant tilgang til emnet, om end en sådan vinkel formentlig ville adskille sig en del fra vores. Således har vi ikke inddraget noget af den litteratur, som har været anvendt i dette kursus.

Formidling af EDB: Denne kursusgang er efter vores forståelse uafhængig af projektføløbet, da denne bliver afløst af et undervisningsforløb og bestået ved aflevering af en evaluerings rapport. Forholder man alligevel dette kursusforløb til projektet, så må vi igen erkende, at dette kursusforløb ikke har nogen direkte forbindelse med vores projekts problemstilling.

Procesbeskrivelse

Gruppen mødtes på grundlag af en fælles interesse for det meget brede område "kunstig intelligens". Inspireret af semesterets kursus i dette emne opnåede vi senere konsensus omkring en fokusering af denne interesse på den del af KI, der handler om logik og sprogformalisering. Efter en del læsning om forskningen i dette område konkretiserede vi yderligere denne interesse til emnet "Semantisk Web". Den konkrete problemformulering udspringer således af en meget bred interesse, og det har påvirket de mål, vi har haft med projektet. Vores hovedmål har hele vejen igennem været at danne os et overblik over forskningen og teorierne indenfor dette område, som flere i gruppen har ambitioner om senere at arbejde videre med, uden fra begyndelsen dog at have erkendt, hvor omfattende blot det at danne sig et overblik over dette meget store område er. I processen har vi naturligvis hele tiden haft fokus på vores produkt, det vil sige projektet og rapporten, men har hele tiden også arbejdet ud fra målet om at sætte os ind i området. Derfor har vores læsning af teorier været bred, og vores diskussioner har bevæget sig i mange forskellige områder inden for KI – heriblandt nogle som lå ret langt fra sprogformaliseringen, der jo er det fokus, vi har valgt i problemformuleringen. Vi har haft en oplevelse af, at det har krævet stor disciplin at fastholde dette fokus, idet der hele tiden har været mange forskellige mulige veje at følge, og problemområdet går på tværs af mange andre, meget forskellige, forskningsområder, herunder filosofi, psykologi, programmering, etik og logik. Alle emner virkede med vores nyfundne interesse for KI tillokkende og interessante, og vi har set alle disse som mulige vinkler på vores problem. Vi har simpelthen haft svært ved at vælge.

Dette har naturligvis også sat sit præg på projektet. Diskussionerne bevæger sig i en del forskellige områder, og for eksempel diskussionen svag kontra stærk KI har fået stor opmærksomhed både i processen og rapporten, selv om den måske ikke er den mest oplagte at tage i forbindelse med den brugskontekst, vi og temarammen har fastlagt.

Selvom disse forhold til tider har virket særdeles frustrerende, har det samtidig været meget berigende for os. Vi kan nok ikke påstå endnu at have overblik over hverken teorier eller forskning indenfor sprogformalisering, endsige KI, men vi mener, at projektprocessen har givet megen viden i forhold til området, skærpet vores interesse og givet os bedre forudsætninger for at arbejde videre indenfor området.

Møder

Generelt var der i starten en del hjemmearbejde, hvor opgaverne primært bestod af at indsamle og læse litteratur om det semantiske web og de værktøjer, der benyttes til at udvikle det semantiske web. Desuden skulle der selvfølgelig skrives lidt om det, der blev læst. Gruppen blev inddelt i mindre grupper af tre gange to personer, som sammen varetog opgaver med at indsamle viden om forskellige områder indenfor det semantiske web. På denne måde mente vi, at vi nemmest kunne skaffe os viden om et emne, som på mange områder var helt nyt for os. I denne tid mødtes vi et par gange om ugen, når vi fandt det relevant og havde nogle problemer eller problemstillinger, vi ønskede at drøfte med gruppen. Egentlig faste møder, med en på forhånd kendt mødestruktur og med en referent, har der ikke været tale om, men mere en løs snak/diskussion.

Denne arbejdsgang blev der ændret på i slutningen af november og starten af december, hvor vi begyndte at mødes fast hver dag kl. 9 i hum.dat-rummet. De mindre grupper blev stadig bevaret til en vis grad, så der ikke var én, der havde hele ansvaret for et rapportafsnit. På trods af aftale om fast mødetid har der været nogle forsinkelser på ca. 10-15 minutter, men der har ikke været ytret mening om, at det har været kilde til irritation. På disse møder har der både været diskuteret projektrelevante ting, samt skrevet på rapporten. På dette tidspunkt var denne arbejdsgang

meget brugbar, da man relativt hurtigt kunne få svar på evt. tvivlsspørgsmål til et rapportafsnit. På denne måde kunne man altså hurtigere komme videre, hvis man var gået i stå.

Indbyrdes kontakt, styring af dokumenter

Vi blev i starten enige om, at diskussioner udenfor gruppemøderne var lettest at holde enten i de små grupper eller som et gruppemøde på Messenger. Indkaldelser til gruppemøder kunne dog forekomme via mail eller sms.

Af erfaring vidste vi, at det er meget vigtigt at holde styr på hvilke dokumenter, der er de nyeste, således at man ikke kommer til at ændre i et gammelt dokument. Dette gælder især, når der er flere mennesker, der skriver på de samme dokumenter. Derfor var der fra projektstart enighed om, at dokumentnavn blev efterfulgt af versionsnummer f.eks. "indledning-v1". For at undgå at "spamme" hinanden med nye versionsnumre af de samme dokumenter oprettede vi hurtigt et grupperum på QuickPlace. Dette havde desuden den fordel, at det blev nemmere at styre dokumenterne. Der blev f.eks. oprettet en mappe, hvori alt materiale til det kommende vejledermøde blev lagt, så det var nemmere at samle de dokumenter, der skulle sendes til vejleder. Ligeledes kunne andre hente nyeste version ned af et dokument, hvis denne f.eks. skulle skrive videre på dokumentet og efterfølgende uploade den nye version som en revision. På denne måde havde vi altid minimum to udgaver af et dokument, således at det var muligt at gå tilbage til en tidligere version, hvis dette var ønskeligt.

Tidsplan og planlægning

Vi har ikke arbejdet ud fra en egentlig tidsplan andet end "deadline 21. december". Dog har vi løbende haft del-deadlines, hvor vi fra det ene møde til det andet aftalte, at der som minimum skulle være et udkast til et rapportafsnit, som enten forfatteren selv kunne fortsætte med eller overlade til en anden.

Vejleder

Vores vejleder på dette semester har været Henrik Schärfe. Vores egentlige kontakt til Henrik blev etableret relativt hurtigt, især fordi vi i starten var lidt usikre på, hvad vi havde af muligheder for projekt indenfor semantisk web. Kontakten til Henrik har fungeret udmærket, og vi har fået hurtig respons på vores henvendelser. Han har gennem flere vejledermøder vist interesse for projektet og været villig til at indgå i diskussioner om forskellige emner, som har været relevante for projektet. Henriks rettelser til rapporten har generelt været overordnede i den forstand, at vi har diskuteret meget med ham, men han har ikke gået ned i små detaljer, medmindre det var virkelig vigtigt. I de første uger af projektet blev der indgået den aftale, at hvis vi havde noget tekst, som vi ville have respons på til et vejledermøde, skulle det mailes 48 timer forinden, for at give Henrik mulighed for at læse og ræsonnere over stoffet. Denne deadline blev ikke altid overholdt, på trods af dette formåede Henrik alligevel at læse det fremsendte materiale. Den sidste uge op imod deadline for aflevering af det færdige projekt blev det muligt at sende materiale til Henrik mere løbende og få hurtigere respons, hvis dette var ønsket.

Empiriindsamling

Vi har i dette projekt ikke været ude og indsamle empiri i traditionel forstand med et optageværktøj (båndoptager eller videokamera). Da det semantiske web stadig er på udviklingsstadiet, har det ikke været muligt at observere, hvordan folk benytter det. Vi har derfor indsamlet information om, hvordan det er tiltænkt at skulle fungere, og med udgangspunkt i en filosofisk og teoretisk diskussion funderet over, hvorvidt semantisk web er muligt, samt perspektiver i det. Derudover har vi inddraget mere generel litteratur om f.eks. den nuværende internetsøgesituation, logik og mening.

Ansvarsområder

Ansvaret for rapportens enkelte afsnit har for det meste været delt ud på flere af gruppemedlemmerne, som har skrevet på dem. Alle har dog været med inde over de fleste afsnit og kommet med input hertil i forbindelse med diverse diskussioner. Herunder er der derfor listet dem, der har hovedansvaret for de enkelte afsnit.

	Helle	Janus	Magnus	Morten	Peter	Rasmus
INDLEDNING	X	X			X	
Problemformulering	X	X	X	X	X	X
Analyse af problemformulering	X					
INFORMATIONSSØGNING I DAG	X				X	
VIDENSREPRÆSENTATION:						
Mening			X			
Ontologibygning		X				
Sandhed		X	X			
DET SEMANTISKE WEB						X
UBOTS ONTOLOGIER				X		
AEROSWARM -ANALYSE			X	X		
SEMANTISK WEB I BRUGSKONTEKST	X				X	X
DISKUSSION		X				X
Derivationsmodellen						X
Semantik		X				
Mening og betydning		X				X
Kan computeren forstå?		X			X	X
Grænser for formalisering		X			X	X
Forståelse i interaktionen					X	X
Værktøj eller interaktionspartner					X	X
KONKLUSION	X	X	X	X	X	X
TEMARAMMEN		X	X			
PROCESBESKRIVELSE				X		

Litteratur

BAEZA-YATES, R., 1998. Searching the web: challenges and partial solutions. *Lecture Notes in Artificial Intelligence. 6th Iberoamerican Conference on Artificial Intelligence oktober 1998 Lissabon*. Springer Verlag, 39-51.

BANKO, M. et al., 2002. AskMSR: Question Answering Using the Worldwide Web. *In: S. M. HARABAGUI & V. CHAUDHRI. Mining Answers from Texts and Knowledge Bases. Papers from the 2002 AAAI Spring Symposium. Technical Report SS-02-06. 25-27 marts 2002 Stanford University*. Menlo Park, CA: AAAI Press.

BERNERS-LEE, T., 2001. The Semantic Web. *Scientific American*. Maj 2001. 28-37.

BIRCH, A., *Introduction to the contemporary imagery debate*. [Online]. Tilgængelig fra: <http://www.gis.net/~tbirch/mi11.htm>.

BODEN, M. A. (ed.), *Artificial Intelligence. Handbook of Perception and Cognition*. San Diego, CA: Academic Press, Inc.

BRICKLEY, D. & GUHA, R. V., 2000. *Resource Description Framework (RDF) Schema Specification 1.0 - W3C Candidate Recommendation*. W3C. Tilgængelig fra: <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>. [Tilgået 19-12-2004].

BRIGGS, J. W. *The Red Baron: Rittmeister Manfred Albrecht Freiherr von Richthofen* [Online]. Tilgængelig fra: <http://www.briggsenterprises.com/bluemax/>. [Tilgået 20-12-2004].

CARLSEN, M., 2003. *Semantic Web – Ontologies*. [Online]. Aalborg Universitet. Tilgængelig fra: http://it.bt.aau.dk/it/education/slides/semantic_web_ontologies.html. [Tilgået 19-12-2004].

CHOWDHURY, C. G., 1999. *Introduction to modern information retrieval*. London: Library Association Publishing.

COLAPIETRO, V. M. & OLSHEWSK, T. M. (eds.), 1996. *Peirce's Doctrine of Signs, Theory, Applications and Connections*. Berlin: Walter de Gruyter & Co.

COPELAND, J., 1993. *Artificial Intelligence: A Philosophical Introduction*. Oxford, UK: Blackwell.

DROTNER, K. et al., 1995. *Medier og Kultur*. København: Borgen.

FØLLESDAL, D. et al., 1999. *Politikens bog om Moderne Videnskabsteori*. København: Politikens Forlag.

GLASGOW, J. I., 1993. The imagery debate revisited: A computational perspective. *Computational Intelligence*. 9. 303-333.

GLYNGØ, A., 2001. *Pragmatik*. [Online]. Århus Universitet. Tilgængelig fra: http://www.hum.au.dk/romansk/romhn/Sprog_og_Kom/Pragmatik.htm. [Tilgået 19-12-2004].

HANKS, S. & McDERMOTT, D., 1986. Default Reasoning, Nonmonotonic Logics and the Frame Problem. In: T. KEHLER & S. ROSENSCHEIN. *Proceedings of the Fifth National Conference on Artificial Intelligence. 11-15 august 1986 Philadelphia, PA*. Menlo Park, CA: AAAI Press.

HEFLIN, J. et al., 2002. *Requirements for a Web ontology language*. [Online]. W3C. Tilgængelig fra: <http://www.w3.org/TR/2002/WD-webont-req-20020307/>. [Tilgået 19-12-2004].

HENDLER J., 2001. Agents and the Semantic Web. *Intelligent Systems, IEEE*. 16, 30-37.

JOHANSEN, M. W., 2003. *Kunstig Intelligens*. København: Fremad.

KRISTENSEN, T., 2003. *Semantisk web – bedre samarbejde mellem computere og mennesker?* [Online]. Kommunikationsforum. Tilgængelig fra: <http://www.kommunikationsforum.dk/artikler.asp?articleid=10917>. [Tilgået 19-12-2004].

LI, W., 2002. Intelligent Information Agent with Ontology on the Semantic Web. In: Y. XI et al. *Intelligent Control and Automation. Proceedings of the 4th World Congress on Intelligent Control and Automation. 10-14 juni 2002 Shanghai*. Shanghai: Press of East China University of Science and Technology.

MAUTNER, T. (ed.), 2000. *The Penguin dictionary of Philosophy*. London: Penguin.

NAVARETTA, C., 2004. *Det Semantiske Web, Ontologier, dialoger kontra tekster*. [Online]. København: IT-Universitetet. Tilgængelig fra: <http://www.itu.dk/courses/DXML/E2004/del11/del11.html>. [Tilgået 13-12-2004].

PARK, J. & HUNTING, S. (eds.), 2002. *XML Topic Maps. Creating and Using Topic Maps for the Web*. Boston, MA: Addison-Wesley.

PEIRCE, C. S., 1992. *Past Masters. The Collected Papers of Charles Sanders Peirce, I-VIII 1931-1966*. Cambridge: Harvard University Press.

PETERSEN, U. et al., 2003. *Conceptual Graphs*. [Online]. Aalborg: Aalborg Universitet. Tilgængelig fra: <http://www.huminf.aau.dk/cg/index.html>. [Tilgået 19-12-2004].

SAEED, J. I., 2002. Semantics, In: N. J. SMELSER & P. B. BALTES (eds.). *International Encyclopedia of the social and behavioral Sciences*. [Online]. Cambridge, MA: Harvard College.

SEARLE, J., 1980. Minds, Brains and Programs. *The Behavioral and Brain Sciences*, 3. Cambridge, UK: Cambridge University Press.

SINGH, P., 2002. *The Open Mind Common Sense Project*. [Online]. KurzweilAI.net. Tilgængelig fra : <http://www.kurzweilai.net/articles/art0371.html?printable=1>. [Tilgået 19-12-2004].

SOWA, J. F., 1990. Crystallizing theories out of knowledge soup. In: Z. W. RAS & M. ZEMANKOVA (eds.), *Intelligent Systems: State of the Art and Future Directions*. London: Ellis Horwood Ltd. 456-487.

SOWA, J. F., 2000. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove, CA: Brooks/Cole.

STRUNGE, M., 1991. Natmaskinen. In: M. STRUNGE. *Vi folder drømmens faner ud*. Valby: Borgen.

SPINK, A. et al., 2002. From E-sex to e-commerce: Web search changes. In: *IEEE Computer*, 35 (3) 107-109.

WALKER, G. & JANES, J., 1999. *Online Retrieval. A Dialogue og Theory and Practice*. Englewood, CO: Libraries Unlimited, Inc.

Internet 1: <http://www.internetworldstats.com/emarketing.htm>

Internet 2: http://www.w3schools.com/xml/cd_catalog.xml

Internet 3: http://www.w3schools.com/rdf/rdf_example.asp

Internet 4: http://www.w3schools.com/rdf/rdf_schema.asp

Internet 5:

http://searchwebservices.techtarget.com/sDefinition/0,,sid26_gci212955,00.html

Internet 6: http://www.google.dk/intl/da/why_use.html

Internet 7: http://help.lycos.com/hotbot/hb_q_12_2.asp

Internet 8: <http://ubot.lockheedmartin.com/ubot/hotdaml/ontologyguide.html>

Internet 9: <http://www.miskatonic.org/godel.html>

Bilag 1

Herunder en sammenligning af hvad de tre forskellige ontologier finder

ENTITET	AEROSWARM	SUMO	CYC
Personer	Persons	Human	Person
III	Also know as: III	III (defineres i en same as-relation)	III
Roy Brown	Also know as: Roy Brown Same as: Brown	#RoyBrown Same as: #Brown	Roy Brown Same as: Brown
Brown	Also know as: Brown	#Brown	Brown (defineres i en same as-relation)
ThorstenPietsch	Står i forfatterlisten til sidst	#ThorstenPietsch	Thorsten Pietsch
LeopoldI.	Also know as: Lepold I.	#LeopoldI.	Lepold I.
Hermann	Also know as: Hermann	#Hermann	Hermann
Camel	Also know as: Camel	#Camel	Camel
OswaldBoelcke	Also know as: Oswald Boelcke	#OswaldBoelcke	Oswald Boelcke
Corbie	Also know as: Corbie	#Corbie	Corbie
StevenLongstreet	Also know as: Steven Longstreet, forfatter	#StevenLongstreet	Steven Longstreet
Ilse	Also know as: Ilse	#Ilse	Ilse
CedricPolkin	Forfatter	#CedricPolkin	Cedric Polkin

ShaneSimmons	Forfatter	#ShaneSimmons	Shane Simmons
NormanFranks	Also know as: Norman Franks Personal Contact: HalGiblin, Begger er forfattere, men til samme værk	#NormanFranks	Norman Franks
HalGiblin	Also know as: Hal Giblin. Hvor- for kender han ikke NF, når NF kender HG?	#HalGiblin	Hal Giblin
ManfredVon	Also know as: Manfred Von	#ManfredVon	Manfred Von
GeorgeZeumer	Also know as: Georg Zeumer	#GeorgeZeumer	George Zeumer
AlbatrosD.	Also know as: Albatros D.	#AlbatrosD.	Albatros D.
RichardPlatt	Forfatter	#RichardPlatt	Richard Platt
AlbrechtVonRichthofen	Also know as: Albrecht Von Richthofen	#AlbrechtVonRichthofen	Albrecht Von Richthofen
Lothar	Also know as: Lothar	Lothar (defineres i en same as- relation)	Lothar
PeterKilduff	Forfatter	#PeterKilduff	Peter Kilduff
Richthofen	Also know as: Richthofen Same as: Lothar	#Richthofen Same as: #Lothar	Richthofen Same as: Lothar
vonRichthofen	Also know as: von Richthofen Associates with: Kunigunde Same as: AlbrechtvonRichthofen	vonRichthofen Same as: #AlbrechtvonRichtho- fen	vonRichthofen Same as: AlbrechtvonRichthofen
Kunigunde	Also know as: Kunigunde	#Kunigunde	Kunigunde
Arthur	Also know as: Arthur	#Arthur	Arthur

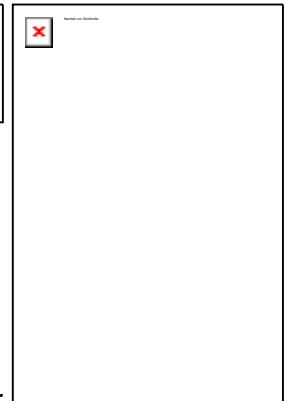
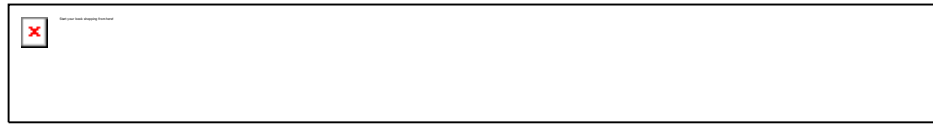
RichardTownshendBickers	Forfatter	#RichardTownshendBickers	Richard Townshend Bickers
Manfred	Also know as: Manfred	#Manfred	Manfred
RomanEmperor	Also know as: Roman Emperor	#RomanEmperor	Roman Emperor
Roy	Also know as: Roy Same as: Brown	#Roy Same as: #Brown	Roy Same as: Brown
LotharVonRichthofen	Also know as: Lothar Von Richthofen Same as: Lothar	#Lothar Von Richthofen Same as: #Lothar	Lothar Von Richthofen Same as: Lothar
VonRichthofen	Also know as: Von Richthofen Same as: AlbrechtvonRichthofen	#Von Richthofen Same as: #AlbrechtvonRichthofen	Von Richthofen Same as: AlbrechtvonRichthofen
ManfredvonRichthofen	Also know as: Manfred von Richthofen Same as: Manfred	#Manfred von Richthofen Same as: #Manfred	Manfred von Richthofen Same as: Manfred
KaiserAlexanderIII	Also know as: Kaiser Alexander III Same as: III	#Kaiser Alexander III Same as: #III	Kaiser Alexander III
Organisationer	Organizations	Organization	Organization
Amazon	Also know as: Amazon	#Amazon	Amazon
JohnWhiley	Also know as: John Whiley	#JohnWhiley	John Whiley
PBS	Also know as: PBS	#PBS	PBS
SocietyofAustralia	Also know as: Society of Australia	#SocietyofAustralia	Society of Australia

RoyalMilitaryAcademy	Also know as: Royal Military Academy Located in: Lichterfelde	#RoyalMilitaryAcademy	Royal Military Academy
AOL	Also know as: AOL	#AOL	AOL
Luftwaffe	Also know as: Luftwaffe	#Luftwaffe	Luftwaffe
CareerAir	Also know as: Career Air	#CareerAir	Career Air
Events			CalenderDay
GreatWar	Also know as: GreatWar		Great War (event)
Measurements			UnitOfMeasure
1289lbs	Also know as: 1289 lbs		1289 lbs
125miles	Also know as: 125 miles		125 miles
110hp	Also know as: 110 hp		110 hp
14000ft	Also know as: 14000 ft		14000ft
24hours	Also know as: 24 hours		24 hours
Time Values	Time Values	Day	CalenderDay
Twentiethcentury	Also know as: Twentieth century (century)	#Twentiethcentury (TmeDuration)	twentieth century (CalenderCentury)
1917to1918	Also know as: 1917 to 1918	#1917to1918	1917 to 1918
APR	Also know as: APR	#APR	APR
July1995	Also know as: July 1995	#July1995	July 1995
Obtober101915	Also know as: Obtober 10, 1915	#Obtober101915	Obtober 10, 1915
JUL	Also know as: JUL	#JUL	JUL
NOV	Also know as: NOV	#NOV	NOV

MAY	Also know as: MAY	#MAY	MAY
SEP	Also know as: SEP	#SEP	SEP
August1914	Also know as: August, 1914	#August1914	August, 1914
AUG	Also know as: AUG	#AUG	AUG
FEB	Also know as: FEB	#FEB	FEB
JUN	Also know as: JUN	#JUN	JUN
August13th1918	Also know as: August 13th, 1918	#August13th1918	August 13th, 1918
JAN	Also know as: JAN	#JAN	JAN
April1911	Also know as: April, 1911	#April1911	April, 1911
OCT	Also know as: OCT	#OCT	OCT
1912	Also know as: 1912	#1912	1912
May	Also know as: May	#May	May
MAR	Also know as: MAR	#MAR	MAR
DEC	Also know as: DEC	#DEC	DEC
April21	Also know as: April 21	#April21	April 21
1915	Also know as: 1915	#1915	1915
Nationaliteter	Nationalities	Citizen	Nationality
Australian	Also know as: Australian	#Australian	Australian
British	Also know as: British	#British	British
Canadian	Also know as: Canadian	#Canadian	Canadian
German	Also know as: German	#German	German
Titler	Titles		Titles
Sergeant	Also know as: Sergeant		Sergeant

Major	Also know as: Major		Major
Køretøjer	Vehicles	TransportationDevice	LandTransportation-Device
Legend	Also know as: Legend	#Legend	Legend
Weapons	Weapons	Weapon	Weapon
Machinegun	Also know as: Machine gun	#Machinegun	Machine gun
Guns	Also know as: Guns	#Guns	Guns
Guns	Also know as: guns	#guns	guns
Lokalitet	Locations	City	City
Wahlstatt	Also know as: Wahlstatt (city)	#Wahlstatt	Wahlstatt
Lichterfelde	Also know as: Lichterfelde (city)	#Lichterfelde	Lichterfelde
			Description
			11

Bilag 2



The Red Baron:

Rittmeister Manfred Albrecht

Freiherr von Richthofen

*Der rote
Kampfflieger:
Rttm. Manfred Von
Richthofen wearing
the Pour le Mérite*

Manfred von Richthofen was born on May 2, 1892 in Breslau, Germany (now Wroclaw Poland). He died on April 21, 1918, age 25, in the skies over Vaux sur Somme, France. His people called him *der rote Kampfflieger* (The Red Battle-Flyer), The French called him *le petit rouge*, and he is known in the English speaking world as *the Red Baron*. In a time of wooden and fabric aircraft, when twenty air victories insured a pilot legendary status and the coveted *Pour Le Mérite* (the famous "Blue Max") , Richthofen had eighty victories, and is regarded to this day as the ace of aces.

This page was initially intended to provide only a brief outline of Richthofen's career. However, due to feedback from visitors, it's been expanded to provide some biographical details of Richthofen's life.

Manfred von Richthofen was born the son of Major Albrecht von Richthofen, a Prussian nobleman, and his wife, Kunigunde. (The name *Richthofen* means "court of judgement" and was bestowed by the Holy Roman Emperor Leopold I.) Manfred was the eldest of three sons (the eldest child was his sister, Ilse). He was enrolled at age 11 at the military school at Wahlstatt, and then attended the Royal Military Academy at Lichterfelde. Manfred was a far better athlete than he was a scholar, and applied his horseback riding skills to become a cavalry officer. He was commissined in April, 1911 in the 1st Regiment of Uhlans Kaiser Alexander III. He was promoted to Leutnant in 1912.

Unfortunately, twentieth century warfare had little use for mounted cavalry. The invention of the machine gun had led to the need for combat operations to be carried out from the relative safety of trenches dug into the countryside. When war broke out in August, 1914, Richthofen looked to the air service for a new challenge. He initially joined the Fliegertruppe (air service) in 1915 as an observer because the training course was shorter and would get him to combat faster. After meeting Oswald Boelcke, who would remain his hero and idol, Manfred was committed to becoming a pilot. After only 24 hours of flight training from his friend, *Oberleutnant* Georg Zeumer, he made his first solo flight on October 10, 1915. (He crashed his plane attempting to land.)

1915

- September - In his second air combat (still as an observer), Richthofen exchanges fire with a French plane. The Allied plane dropped away and crashed, but Richthofen was not credited with the victory as the enemy plane crashed behind French lines (and could therefore not be confirmed). Later in his career, Richthofen would not be held to this restriction: he was taken at his word for his kills.

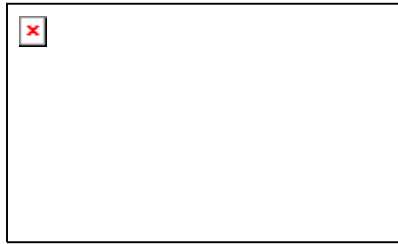


Richthofen in 1916.

1916

- Apr 24 - Richthofen shoots down his first plane as a pilot. The craft, a French Nieuport, crashed behind enemy lines, and again the victory was not officially credited to Richthofen.
- Sep 1 - At Oswald Boelcke's invitation, Richthofen reports for duty on the Western Front. He begins his career with Jagdstaffel 2 in an Albatros D.II biplane. Although it was the Fokker Dr.I triplane with which Richthofen is remembered, he spend the vast majority of his time flying biplanes like the Albatros D.II and D.III.
- Sep 17 - Richthofen scores his first confirmed air victory.
- October: After 40 victories, Oswald Boelcke is killed in a mid-air collision during combat. Some accounts blame Richthofen's enthusiasm for the collision which caused pilot Erwin Böhme's undercarriage to collide with Boelcke's upper wing. Others place the blame on Böhme, or call it simply an accident not attributable to anyone's mistake.

- Nov 23 -- Richthofen, with the help of a superior aircraft, makes British ace Major Lanoe Hawker his eleventh victim.



**Richthofen in his red Albatros D.III with Jasta 11.
Seated on the ground is Manfred's younger brother, Lothar.**

1917

- Jan 4 -- Richthofen scores his 16th air victory, making him the top living German ace. On January 12th, Richthofen receives the *Orden Pour le Mérite* (aka the "Blue Max"). He is given command of Jasta 11. Richthofen decides to paint parts of his aircraft red, in part to identify himself easily to his allies on the ground (whom he feared would otherwise shoot at him). It has been suggested that he chose red because it was the color of his old Uhlan cavalry regiment. This
- begins a series of traditions: each plane in his squadron begins to display some red coloring to show solidarity. Later in the war, some British planes bore red noses, announcing their intention to hunt down the Red Baron.
- Jan 24 -- The lower wing of Richthofen's Albatros D.III breaks off in flight (a common problem for that type).
- Mar 9 -- Richthofen is shot down over Oppy, but was flying again the same day.
- Apr 7 -- Richthofen is promoted to Rittmeister (Cavalry Captain)
- Apr 8 -- The lower wing of the Albatros D.III flown by another member of Jasta III breaks off in flight. Richthofen writes an angry letter to Berlin, and is visited by aircraft designer Anthony Fokker, who literally goes to the trenches to observe his aircraft in action. The result of this visit (in which he was able to observe the British Sopwith Triplane) led to Fokker's development of the Dr.I triplane.
- Apr 29 -- Richthofen shoots down four planes in one day, a personal best. April, 1917 is known as "Bloody April". Britain lost 912 pilots and observers during the month, while Richthofen scored an incredible 21 victories during the same period. After his 41st victory, Richthofen was ordered on leave. Turning command over to his brother



Albatros D.III

Lothar, Manfred left the front on May 1st and did not return until early June. He spent his vacation hunting in his home town, on propoganda tours, and meeting with Kaiser Wilhelm.

- Jun 24 -- Jagdgeschwader 1 (Fighter Wing 1) is formed, with Manfred von Richthofen in command. The wing would later be renamed "Jagdgeschwader Frieherr von Richthofen" in his honor.

-

- July: Richthofen is shot down by Capt. Douglas Cunnell and 2nd Lt. Albert Woodbridge. Though he lands safely, he has suffered a serious bullet wound to the head. Manfred would suffer from terrible headaches until the end of his life, but there were other ramifications. German command, recognizing the propaganda value to the enemy of Richthofen's loss, begins to pressure him to retire from air combat, going so far as to forbid him from flying *unless absolutely necessary* (a loophole Richthofen exploited

Fokker Dr.I Dreidecker

Service Dates: 1917 to 1918 (320 built)

18'-11"L 23'-7"W 9'-8"H

Weight: 1,289 lbs (893 empty)

Maximum Speed: 103 mph

Maximum Range: 125 miles

Maximum Altitude: 14,000 ft

Engine: Oberursel UR II (110 hp)



at every opportunity).

Richthofen began to spend more and more of his time in more administrative and public relations roles. Eventually he returned to the front full-time when Germany's leaders realized that no matter the risks, they could not afford not to let Richthofen do what he did better than anyone else.

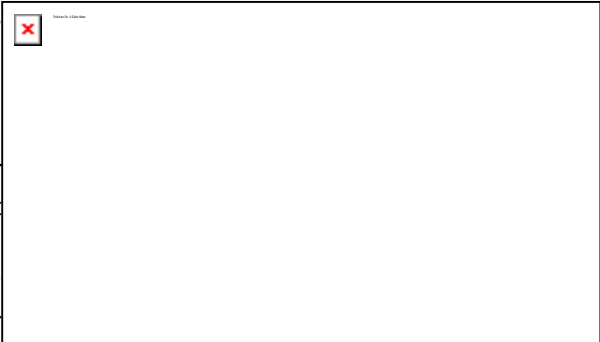
- August: The first Fokker triplanes are delivered to Jagdgeschwader 1.
- September 1 -- Richthofen scores his 60th victory, his first in the Dr.I triplane. On September 6th, he took a leave of convalescence, and when he returned in the next month he went back to flying the Albatros D.V.

1918

- April: Richthofen achieves two victories flying Fokker Dr.1 triplane (number 425/17). Though he flew biplanes for nearly all of his career, and most of these were only partly painted red, it is the Dr.1 triplane, blood-red from cowl to tail, which is commonly


associated with the Red Baron.

On April 21, Richthofen followed the Sopwith Camel of Wilfred May far into British territory. The end of the war was only months off by this time, and the German air command faced both ever-improving British airplanes and their own dwindling numbers. The thrill of the hunt was all but gone for Baron von Richthofen, as most of his peers had already been killed and his own wounds agonized him. Though the German air doctrine he himself wrote stated that "one should never obstinately stay with an opponent which, through bad shooting or skillful turning, he has been unable to shoot down while the battle lasts until it is far on the other side", he chased his British quarry far deeper into enemy territory and far lower to the ground than his own doctrine permitted. May later said that it was only his erratic, untrained piloting which saved him. Richthofen followed the erratic path of the novice pilot until a single bullet, shot from behind him, passed diagonally through his chest. The shot is commonly believed to have come from Australian gunners on the ground, but might have also come from the guns of Canadian flier Arthur "Roy" Brown who was coming to May's aid. Manfred von Richthofen crashed into a field alongside the road from Corbie to Bray. His body was recovered by British forces, and he was buried with full military honors.



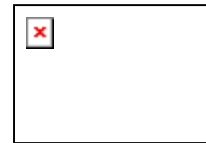
Manfred's brother, Lothar (also a *Pour le Mérite* recipient) was himself recovering from being shot down when his older brother was killed in combat. He returned to Jagdgeschwader 1 and carried on the Richthofen tradition of fearlessness in combat in a blood-red fighter. Lothar was shot down again on August 13th, 1918, and forced into retirement with 40 kills. Manfred's eventual successor was Hermann Göring (who would later become the head of the Luftwaffe and a particularly infamous Nazi), who chose to paint his aircraft completely white, ending the reign of the blood-red German fighters.

<h2>Career Air Victories by Month</h2>	1916					
	JUL	AUG	SEP	OCT	NOV	DEC
			3	3	5	4
	1917					
	JAN	FEB	MAR	APR	MAY	JUN
	3	3	10	21	0	4
JUL	AUG	SEP	OCT	NOV	DEC	
1	2	2	0	0	0	

	1918					
	JAN	FEB	MAR	APR	MAY	JUN
	0	0	11	6		

Books on Manfred von Richthofen and WWI Aviation available from Amazon.com
(Please, if you plan to purchase any of these books, please do so by following these links to Amazon's site. Support from Amazon.com is the only source of revenue for this page.)

- [Richthofen : Beyond the Legend of the Red Baron](#) ©1993, Peter Kilduff. Published by John Wiley & Sons, Inc.
- [The Canvas Falcons](#) by Steven Longstreet
- [Von Richthofen : The Legend Evaluated](#) by Richard Townshend Bickers
- [Under the Guns of the German Aces : Immelmann, Voss, Goring, Lothar Von Richthofen : The Complete Record of Their Victories and Victims](#) by Norman Franks and Hal Gibling
- [Illustrated Book of Great Adventures](#) retold by Richard Platt. (This is an illustrated children's book with a chapter featuring Richthofen.)



Manfred von Richthofen on the Web ([e-mail me](#) if you have any links to add)

- [Thumbnail Biography of Richthofen](#)
- [Der rote Kampfflieger](#) by Manfred von Richthofen (German)
- [Der rote Kampfflieger](#) by Manfred von Richthofen (English)
- **Related Sites**
 - [PBS's "The Great War"](#)
 - [The Aerodrome](#)
 - Though Canadian flier Roy Brown is credited with downing the Red Baron, [The Death of Manfred von Richthofen: Who fired the fatal shot?](#), published by *The Journal and Proceedings of the Military History Society of Australia* presents a detailed argument that the fatal shot was most likely fired by Sergeant Cedric Popkin of the Australian ground forces.

Partial**Bibliography**

I have received many requests for a bibliography (students don't go to the library anymore, it seems). Information for this page has come from numerous sources, most of which I cannot recall. In addition to the books above, the following also provided information for this page:

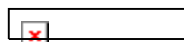
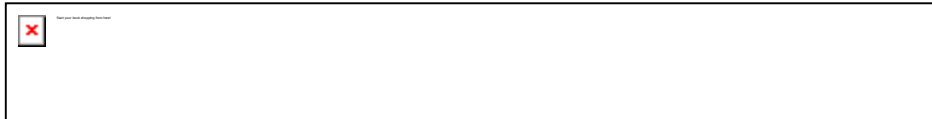
- *Red Baron: An Ace for the Ages* by Shane Simmons, ©1995 Aviation History (magazine, July 1995 issue).
- *Ace of Aces* episode of A&E Biography.

NOTE: All attempts are made to ensure that the information on this page is accurate, but errors may remain. Thanks to RedBa7210 at AOL and Thorsten Pietsch either for correcting errors or helping me clarify information on this page.

Before**you****write:**

I appreciate your feedback. I especially appreciate mail which will help enhance or clarify the information on this page. If presenting such information, please let me know what your source is for corroboration. Also, please note:

- This page is intended to present a capsule biography of Manfred von Richthofen. It is not intended to suggest that no other pilot has equaled or surpassed his success in combat. The fact that he is called the "Ace of Aces" is simply a matter of historical record.
- Historians have argued for eight decades whether Richthofen was shot down from the air or ground. The issue will not be settled here.
- Many people e-mail me asking for more links with information on Richthofen.... all the ones I know of are listed right here.



[LE FastCounter](#)

Bilag 3

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE owl [
  <!ENTITY aeroswarm
"http://ubot.lockheedmartin.com/ubot/2004/04/aeroswarmOntology.owl#">
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
]>
<rdf:RDF
  xmlns:owl      = "http://www.w3.org/2002/07/owl#"
  xmlns:rdf      = "http://www.w3.org/1999/02/22-rdf-syntax-
ns#"
  xmlns:rdfs     = "http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd      = "http://www.w3.org/2001/XMLSchema#"
  <owl:Ontology rdf:about="">
    <owl:versionInfo>v 1.00 2004/04/11 12:56:51</owl:versionInfo>
    <rdfs:label>AeroSwarmExtractionOntology</rdfs:label>
    <rdfs:comment>AeroSwarm OWL ontology</rdfs:comment>
  </owl:Ontology>
  <!-- Entities -->
  <owl:Class rdf:ID="Person">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasStringName"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasAlias"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#LifeForm"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasIdentity"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasNationality"/>
        <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasSpouse"/>
        <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
```

```

    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasBirthPlace"/>
        <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
        </owl:Restriction>
      </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasBirthYear"/>
        <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
        </owl:Restriction>
      </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#transOccurred"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
      </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasPersonalContact"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
      </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasEmployRole"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
      </rdfs:subClassOf>
    </owl:Class>

    <owl:Class rdf:ID="Organization">
      <rdfs:subClassOf>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#hasStringName"/>
          <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
          </owl:Restriction>
        </rdfs:subClassOf>
      <rdfs:subClassOf>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#hasAlias"/>
          <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
          </owl:Restriction>
        </rdfs:subClassOf>
      <rdfs:subClassOf rdf:resource="#Entity"/>
      <rdfs:subClassOf>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#hasIdentity"/>
          <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
          </owl:Restriction>
        </rdfs:subClassOf>
      <rdfs:subClassOf>

```

```

    <owl:Restriction>
      <owl:onProperty rdf:resource="#transOccurred"/>
      <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <!-- Org should be able to have an employ Role because an Org can
work
      For another Org or belong to another Group-->
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasEmployRole"/>
      <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:ID="LifeForm">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#Entity"/>
</owl:Class>
<owl:Class rdf:ID="NamedEvent">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTextPosition"/>
      <owl:allValuesFrom rdf:resource="#TextPosition"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="Abstraction">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasTextPosition"/>
      <owl:allValuesFrom rdf:resource="#TextPosition"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="Measure">
  <rdfs:subClassOf>
    <owl:Restriction>

```

```

        <owl:onProperty rdf:resource="#hasStringName"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="Email">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="Percent">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="Time">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="PhoneNumber">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="Nationality">
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:cardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="Title">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>

```

```

        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="URL">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="IDNum">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="LatLong">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Abstraction"/>
</owl:Class>
<owl:Class rdf:ID="Ago">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="Century">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="Season">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>

```

```

        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="Decade">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="RelativeUnit">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="AbsoluteDate">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="DayOfWeek">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="Quarter">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="TimeOfDay">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>

```



```

        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="Holiday">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<!-- I was not able to get TimeRange to extract as its own entity
    It was extracted as a property of other Dates and Times.-->
<owl:Class rdf:ID="TimeRange">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Time"/>
</owl:Class>
<owl:Class rdf:ID="Entity">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasTextPosition"/>
            <owl:allValuesFrom rdf:resource="#TextPosition"/>
        </owl:Restriction>
    </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="PhysicalObject">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Entity"/>
</owl:Class>
<owl:Class rdf:ID="Vehicle">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>

```

```

    <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
  </owl:Class>
  <owl:Class rdf:ID="Money">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasStringName"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
        </owl:Restriction>
      </rdfs:subClassOf>
      <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
    </owl:Class>
    <owl:Class rdf:ID="Weapon">
      <rdfs:subClassOf>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#hasStringName"/>
          <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
          </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
      </owl:Class>
      <owl:Class rdf:ID="Drug">
        <rdfs:subClassOf>
          <owl:Restriction>
            <owl:onProperty rdf:resource="#hasStringName"/>
            <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
          </rdfs:subClassOf>
          <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
        </owl:Class>
        <owl:Class rdf:ID="Facility">
          <rdfs:subClassOf>
            <owl:Restriction>
              <owl:onProperty rdf:resource="#hasStringName"/>
              <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
              </owl:Restriction>
            </rdfs:subClassOf>
            <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
          </owl:Class>
          <owl:Class rdf:ID="Aircraft">
            <rdfs:subClassOf>
              <owl:Restriction>
                <owl:onProperty rdf:resource="#hasStringName"/>
                <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
                </owl:Restriction>
              </rdfs:subClassOf>
              <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
            </owl:Class>
            <owl:Class rdf:ID="Chemical">
              <rdfs:subClassOf>
                <owl:Restriction>
                  <owl:onProperty rdf:resource="#hasStringName"/>
                  <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
                  </owl:Restriction>
                </rdfs:subClassOf>
                <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
              </owl:Class>

```

```

</owl:Class>
<owl:Class rdf:ID="Location">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#PhysicalObject"/>
  </owl:Class>
<owl:Class rdf:ID="Place">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Location"/>
  </owl:Class>
<owl:Class rdf:ID="Nation">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Place"/>
  </owl:Class>
<owl:Class rdf:ID="County">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Place"/>
  </owl:Class>
<owl:Class rdf:ID="Region">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Place"/>
  </owl:Class>
<owl:Class rdf:ID="Province">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#Place"/>
  </owl:Class>

```

```

<owl:Class rdf:ID="City">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#Place"/>
</owl:Class>
<owl:Class rdf:ID="Water">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#Place"/>
</owl:Class>
<owl:Class rdf:ID="LandForm">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#Place"/>
</owl:Class>
<owl:Class rdf:ID="Continent">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStringName"/>
      <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#Place"/>
</owl:Class>

<!-- LINKS -->

<!-- pers/orgToPhone replaced by hasPhone and
pers/org to url/email/aka/idnum added -->

<owl:ObjectProperty rdf:ID="hasIdentity">
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Entity"/>
        <owl:Class rdf:about="#Abstraction"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:range>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasPhone">
  <rdfs:subPropertyOf rdf:resource="#hasIdentity" />
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#PhoneNumber" />

```

```

</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasEmail">
  <rdfs:subPropertyOf rdf:resource="#hasIdentity" />
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#Email" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasURL">
  <rdfs:subPropertyOf rdf:resource="#hasIdentity" />
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#URL" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasIDNum">
  <rdfs:subPropertyOf rdf:resource="#hasIdentity" />
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#IDNum" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasBirthPlace">
  <rdfs:subPropertyOf rdf:resource="#hasIdentity" />
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="#Location" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasNationality">
  <rdfs:domain rdf:resource="#Person"/>
  <rdfs:range rdf:resource="#Nationality"/>
</owl:ObjectProperty>

<!-- hasBirthYear cannot be a subpropOf hasIdentity because
hasBirthYear is a datatype Property
and hasIdentity is an ObjectProperty-->

<owl:DatatypeProperty rdf:ID="hasBirthYear">
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="&xsd;nonNegativeInteger" />
</owl:DatatypeProperty>

<!-- pers/org to Location replaced by hasLocation -->

<owl:ObjectProperty rdf:ID="hasLocation">
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#Location" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasTravelLocation">
  <rdfs:subPropertyOf rdf:resource="#hasLocation" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasAddress">
  <rdfs:subPropertyOf rdf:resource="#hasLocation" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasPosition">
  <rdfs:subPropertyOf rdf:resource="#hasLocation" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="isLocated">
  <rdfs:subPropertyOf rdf:resource="#hasPosition" />
</owl:ObjectProperty>

<!-- (org/pers) to (org/pers) replaced by TransactionOccurred and
hasPersonalContact.
TransactionOccurred has 5 sub types (Bus_Dealings,
Provides_Services, Sub_Org, Associate, Targets),
but Targets does not extract well so not included here. Also,
I left associate out because

```

Personal_Contacts has the same subtype so I thought this was confusing. -->

```
<owl:ObjectProperty rdf:ID="transOccurred">
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#Entity" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="busTransOccurred">
  <rdfs:subPropertyOf rdf:resource="#transOccurred" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="providedServices">
  <rdfs:subPropertyOf rdf:resource="#transOccurred" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasSubOrg">
  <rdfs:subPropertyOf rdf:resource="#transOccurred" />
</owl:ObjectProperty>
```

<!-- Employee Roles (works for, belongs To) are for Person(or Org) to Org Contact-->

```
<owl:ObjectProperty rdf:ID="hasEmployRole">
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#Organization" />
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="belongsTo">
  <rdfs:subPropertyOf rdf:resource="#hasEmployRole" />
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="worksFor">
  <rdfs:subPropertyOf rdf:resource="#hasEmployRole" />
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="#Organization" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="isClient">
  <rdfs:subPropertyOf rdf:resource="#worksFor" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="isManagement">
  <rdfs:subPropertyOf rdf:resource="#worksFor" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="isGeneralStaff">
  <rdfs:subPropertyOf rdf:resource="#worksFor" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="owns">
  <rdfs:subPropertyOf rdf:resource="#worksFor" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="isAffiliatePartner">
  <rdfs:subPropertyOf rdf:resource="#worksFor" />
</owl:ObjectProperty>
```

<!-- Personal Contacts are only for Person to Person Contact. They are symmetric.

-->

```
<owl:ObjectProperty rdf:ID="hasPersonalContact">
  <rdfs:type rdf:resource="#owl:SymmetricProperty"/>
  <rdfs:domain rdf:resource="#Person" />
  <rdfs:range rdf:resource="#Person" />
</owl:ObjectProperty>
```

```

<owl:ObjectProperty rdf:ID="associatesWith">
  <rdfs:subPropertyOf rdf:resource="#hasPersonalContact" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasAcquaintance">
  <rdfs:subPropertyOf rdf:resource="#hasPersonalContact" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="isRelated">
  <rdfs:subPropertyOf rdf:resource="#hasPersonalContact" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasParent">
  <rdfs:subPropertyOf rdf:resource="#isRelated" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasSibling">
  <rdfs:subPropertyOf rdf:resource="#isRelated" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasSpouse">
  <rdfs:subPropertyOf rdf:resource="#isRelated" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasGrandparent">
  <rdfs:subPropertyOf rdf:resource="#isRelated" />
</owl:ObjectProperty>

```

<!-- locationInText denotes the start and end position, by character count from the head of the file, of the found entity.

```

-->
<owl:Class rdf:ID="TextPosition">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasStartPosition"/>
      <owl:cardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasEndPosition"/>
        <owl:cardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
        </owl:Restriction>
      </rdfs:subClassOf>
    </owl:Class>
  <owl:DatatypeProperty rdf:ID="hasStartPosition">
    <rdfs:domain rdf:resource="#TextPosition" />
    <rdfs:range rdf:resource="&xsd;nonNegativeInteger" />
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="hasEndPosition">
    <rdfs:domain rdf:resource="#TextPosition" />
    <rdfs:range rdf:resource="&xsd;nonNegativeInteger" />
  </owl:DatatypeProperty>
  <owl:ObjectProperty rdf:ID="hasTextPosition">
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#Entity"/>
          <owl:Class rdf:about="#Abstraction"/>
          <owl:Class rdf:about="#NamedEvent"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:range rdf:resource="#TextPosition" />

```

```

</owl:ObjectProperty>
<owl:DatatypeProperty rdf:ID="hasSourceURI">
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="&xsd:anyURI" />
</owl:DatatypeProperty>

<owl:ObjectProperty rdf:ID="isConsistentWith">
  <rdf:type rdf:resource="&owl;SymmetricProperty"/>
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#Entity" />
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="isInconsistentWith">
  <rdfs:subPropertyOf rdf:resource="&owl;differentFrom" />
  <rdf:type rdf:resource="&owl;SymmetricProperty"/>
  <rdfs:domain rdf:resource="#Entity" />
  <rdfs:range rdf:resource="#Entity" />
</owl:ObjectProperty>

<owl:DatatypeProperty rdf:ID="hasStringName">
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Entity"/>
        <owl:Class rdf:about="#Abstraction"/>
        <owl:Class rdf:about="#NamedEvent"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:domain>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="hasAlias">
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Person"/>
        <owl:Class rdf:about="#Organization"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:domain>
  <rdfs:range rdf:resource="&xsd:string"/>
</owl:DatatypeProperty>

</rdf:RDF>

```


Bilag 4



SPACE SYSTEMS > MANAGEMENT & DATA SYSTEMS

DAML UBOT Project

[Introduction](#)

[Team Members](#)

[Details](#)

[Lessons Learned](#)

[Papers](#)

AeroSWARM Extraction Explained

The purpose of the table below is to help users map their ontology to the AeroSWARM ontology. It provides a description of what AeroSWARM extracts from a document. The following classes are used to structure the ontology but you cannot map to them: ENTITY, PHYSICALOBJECT, LIFEFORM, LOCATION, ABSTRACTION, TIME, and GROUP. DAML does not currently have a way to indicate that a class is abstract (i.e., it has no instance).

| Extraction Class | Description | Examples |
|------------------|---|--|
| ENTITIES | | |
| PERSON | A human, individual | John Smith, Jane, Osama bin Laden |
| ORGANIZATION | Generic organization | University of Texas, EuroBankers Inc., Taliban |
| Time | | |
| ABSOLUTEDATE | A specific day | May 31, 2002, Thu Jul 18 |
| AGO | A time earlier than the present time | 2 years ago, this morning |
| HOLIDAY | All US holidays | New Year, Christmas, Labor Day |
| CENTURY | A period of 100 years | 18th century, first century |
| DECADE | A period of 10 years | 1920's, 1920s, 20s |
| SEASON | Season of the year | Spring, Summer, Fall, Winter |
| DAYOFWEEK | Day of week | Monday, Tuesday, Sunday |
| TIMEOFDAY | Time of a day | HH:MM:SS, HH:MM (am pm) |
| QUARTER | Fiscal quarter | first quarter, |
| RELATIVEDATE | Describes a point in time | the end of the year |
| RELATIVEUNIT | Time, relative to another point in time | in the past 27 years, last week |

| | | |
|-------------------------|---|--|
| TIMERANGE | A period of time that spans over two instances in time | between 1992-2000 |
| Locations | | |
| PLACE | Generic physical location | When the AeroSWARM engine does not have enough to subtype, it will default a physical location to 'Place'. |
| CONTINENT | One of the great divisions of land on the globe | Europe, Asia, North America |
| LANDFORM | A natural feature of a land surface | Death Valley Basin, Rocky Mountain |
| WATER | A body of water | Amazon River, Pacific Ocean |
| REGION | Geopolitical regions | Middle East, Pacific Rim |
| NATION | A political state | USA, Great Britain, England |
| PROVINCE | An administrative district or division of a country | Pennsylvania, New York State |
| DISTRICT | A territorial division (as for administrative or electoral purposes) | Chicago Park District |
| COUNTY | The largest local administrative unit in various countries | Madison county |
| CITY | An urban population center | Washington, Miami, Kuala Lumpur |
| Physical Objects | | |
| VEHICLE | A land vehicle, automobile | Ford Explorer, Chevrolet |
| AIRCRAFT | An airplane | F-16, Boeing 747 |
| VESSEL | A watercraft | USS Enterprise, Kidd class destroyer |
| DRUG | An illegal substance that causes addiction | Heroin, LSD, Cocaine |
| FACILITY | A Building, an installation, something that is built, installed, or established to serve a particular purpose | White House, Children's Hospital |
| WEAPON | A device used to injure, defeat, or destroy | gun, knife, nuke |

| | | |
|-------------------|--|---|
| MONEY | A denomination of money | \$40, 30 pound [euro, yen, ringgit, ...] |
| CHEMICAL | A chemical reagent | Sulfur, Helium |
| Relations | | |
| hasPhone | Associates a phone number to an entity | John can be reached at 555-5555 |
| hasEmail | Associates an email address to an entity | John Doe, j.doe@email.com |
| hasURL | Associates a URL to an entity | The White House, www.WhiteHouse.com |
| hasLocation | Associates an entity to a location | Senator John Smith of Washington |
| hasTravelLocation | Associates an entity to a travel location | Jane studied at the University of Texas at Austin; John moved to Houston, Texas |
| hasAddress | Associates an entity to an address | Hello Inc., P.O. Box 1000, Philadelphia, PA |
| isLocated | Associates an entity to a location | John Doe of Newbury Port, MA; Southern Methodist University in Dallas, Texas |
| hasPosition | Associates an entity to a Position | Senator John Smith of Washington |
| hasBirthPlace | Associates a person to his/her BirthPlace | John was born in Milton, New York |
| hasNationality | Associates a person to his/her Nationality | Jane Doe, a polish immigrant |
| hasBirthYear | Associates a person to his/her BirthYear | John was born on July 15, 2004; The following extracts BirthYear of 1968: June 16, 2004 - John, age 36. |
| transOccurred | Associates an entity to an entity; signals that a transaction has occurred between the entities | ABC Inc. arranged a deal for CDE Inc; ABC Inc. transferred \$800 to CDE Inc |
| busTransOccurred | Associates an entity to an entity; signals that a business transaction has occurred between the entities | ABC Inc. sold xyz to CDE Inc. |
| providedServices | Associates an entity to an entity; signals that one entity provided | The US sells weapons to France; The bank gave John a loan; Jack used First Union |

| | | |
|--------------------|--|--|
| | services to the other | |
| hasSubOrg | Associates an organization to a suborganization | Big Inc. owns Small Inc. |
| hasEmployRole | Associates an entity to an organization; signals that the entity is an employee of the organization | John Smith, who heads Operations; Jane Doe, Co-Founder of ABC Inc. |
| belongsTo | Associates an entity to an organization; signals that the entity is a member of the organization | John joined Hello Inc.; John is on the board of Directors |
| worksFor | Associates a person to an organization; signals that the person works for the organization | John Smith of IBM; John's company, ABC Inc. |
| isClient | Associates a person to an organization; signals that the person is a client Of the organization | A client of John, Jane; |
| isManagement | Associates a person to an organization; signals that the person is a management at the organization | John became a leader of Phillip's Academy; John ran for the Senate |
| isGeneralStaff | Associates a person to an organization; signals that the person is general staff at the organization | Laura worked at Dawson Elementary School |
| owns | Associates a person to an organization; signals that the person owns the organization | Mr. Jones, owner of the account; Mr Jones is the owner of the club |
| isAffiliatePartner | Associates a person to an organization; signals that the person is a partner of the organization | Batman and his sidekick RObin; John's assistant Anne |
| hasPersonalContact | Associates a person to a person; signals that the persons are personal contacts | George and Laura Bush |
| associatesWith | Associates a person to a person; signals that the persons associate with | John, Jane, and Julie; John, with Senator Jane; John's associate, Jill |

| | | |
|----------------------|--|---|
| | one another | |
| hasAcquaintance | Associates a person to a person; signals that the persons are acquaintances | John told Jane; They (John and Jane) |
| isRelated | Associates a person to a person; signals that the persons are related | John Smith, related to Jill SMith; |
| hasParent | Associates a person to a person; signals that one person is a parent of the other | Jane's parents are Jack and Jill; |
| hasSpouse | Associates a person to a person; signals that one person is a spouse of the other | George married Laura; George's wife Laura |
| hasSibling | Associates a person to a person; signals that one person is a sibling of the other | Ralph, the brother of Jane and Peter; Jane's brother, Peter |
| hasGrandparent | Associates a person to a person; signals that one person is a grandparent of the other | John's grandfather Jack; Mary's grandmom Kate |
| Miscellaneous | | |
| TITLE | An appellation of dignity, honor, distinction, or preeminence attached to a person | President, Vice Admiral, Mr. |
| EMAIL | An email address | John.dole@xyz.com |
| PHONENUMBER | A telephone/fax number | 212-555-5555 |
| NATIONALITY | Membership in a particular nation | American, Russian, Arab |
| NAMEDEVENT | An event | World War II, American Lung Association Conference |
| URL | A web site URL | www.abc.com |
| PERCENT | Percent of something | 10%, 25.3% |
| MEASURE | Some kind of measurement | 10 miles, 100 tons |

