

4. WWW-SØGEVÆRKTØJER	33
4.1 METODISKE OVERVEJELSER	33
4.1.1 Udgangspunkt i virkeligheden	33
4.1.2 Konfrontation med fænomenet	33
4.1.3 Interaktion med fænomenet	34
4.1.4 Teoretisk perspektiv	34
4.1.5 Dannelse af væsenstyper	35
4.2 SØGESYSTEMERNES BEGRÆNSNINGER	35
4.2.1 Begrænsninger i informationernes kvalitet og tilgængelighed	36
4.2.2 Søgeværktøjerne som kommercielle produkter	37
4.3 FORSTÅ UDFORDRINGEN I WWW-SØGNING	41
4.3.1 Hypertekst og hyperlinks	43
4.3.2 Dokumentkodningsstandard <i>HTML</i>	44
4.3.3 Forudsætninger for at kunne tage udfordringen i <i>WWW-søgning op</i>	46
4.4 FÅ ET OVERBLIK OVER SØGEVÆRKTØJERNES KVALITETER	48
4.4.1 Udgangspunkter	48
4.4.2 Emneindeksene	48
4.4.3 Vejledere	49
4.4.4 Søgnesystemer	49
4.4.5 Bogmærker	50
4.4.6 Fordele og ulemper ved de forskellige typer af søgeværktøjer	51
4.4.7 Vurderingsparametrene når en udbyder af et søgeværktøj skal vælges fremfor en anden	52
4.5 SITUATIONENSBESTEMT ANVENDELSE AF SØGEVÆRKTØJERNE	55
4.5.1 <i>WWW</i> s mangfoldighed - udgangspunkter	55
4.5.2 Specifik information - simpel søgning	55
4.5.3 Information af høj kvalitet - vejledere via emneindeks	59
4.5.4 Samtlige <i>WWW-tilgængelige</i> informationer - avanceret søgning	61
4.6 UNDERSØG SPECIALISEREDE REFERENCESYSTEMER	63
4.7 OVERVÅG NYE INFORMATIONER PÅ <i>WWW</i>	64
4.7.1 Udgangspunkt	64
4.7.2 Emneindeks	64
4.7.3 Søgnesystemer	65
4.7.4 Bogmærker	65
4.7.5 Andre værktøjer	66
4.8 TRANSFORMÉR VIDEN TIL INFORMATION	66

4. WWW-søgeværktøjer

I denne delundersøgelse har vi sat os i informationssøgerens sted, og søgt at oparbejde en erfaring omkring, med hvilke metoder man kan undersøge udbuddet af informationer på WWW, hvorefter det er meningen, at denne erfaring skal bruges på flere måder i forbindelse med udarbejdelsen af dette projekt. De metoder, der findes til informationssøgning på WWW, har vi valgt at kalde WWW-søgeværktøjer.

4.1 Metodiske overvejelser

Inden resultatet af selve delundersøgelsen præsenteres, vil det være nødvendigt med nogle metodiske overvejelser i relation til Eneroth's kundskabstrin, der blev beskrevet i forrige kapitel. Den viden denne delundersøgelse har givet, er primært baseret på refleksion af vores erfaring med at søge efter information på WWW og sekundært på fortolkninger af information vi indhentede fra udbydere af søgeværktøjer - via e-post og produktinformation - og dialog med og information fra andre, der også havde analyseret WWW-søgeværktøjer.

4.1.1 Udgangspunkt i virkeligheden

Med delundersøgelsen ville vi undersøge den virkelighed en virksomhed står overfor, når den skal "på nettet", hvilket bringer os frem til det overordnede mål med denne delundersøgelse, nemlig at svare på vores problemstilling. Således var denne delundersøgelses mål at sætte sig i informationssøgerens sted. Dette var ud fra den tanke, at for en virksomhed kan styrke sin internationaliseringsproces ved at markedsføre sine produkter og serviceydelser på WWW, må dens nuværende og potentielle kunder være i stand til at finde den information, virksomheden har gjort tilgængelig på WWW. Samtidig må virksomheden også være i stand til at finde information om eventuelle konkurrenter, som også søger at styrke deres internationaliseringsproces ved at markedsføre deres produkter og serviceydelser på WWW. Med andre ord, vi beskæftigede os med at undersøge fænomenet "Informationssøgning på WWW", som kunne have mange anvendelsesmuligheder. For at kunne gøre dette måtte vi først beskrive fænomenet og derefter via fortolkninger komme frem til opfyldelsen af de nævnte mål.

4.1.2 Konfrontation med fænomenet

Vi ønskede at skabe en erfaring med så mange sider af fænomenet "Informationssøgning på WWW" som muligt, dog havde vi ikke ubegrænsede ressourcer. Således begrænsede vi undersøgelsen af WWW-søgeværktøjer i retning af, at vi - for det første - kunne bruge den tilegnede viden i forbindelse med vores egen informationssøgning på WWW, således vi effektivt kunne finde information om virksomheder, der er på WWW og som har relevans for vores problemstilling. Med andre ord udgør denne tilegnede viden en del af vores dataindsamlingssteknik. Da vores virkelighed tager udgangspunkt i at betragte WWW som et marked, så vi det nødvendigt at undersøge, hvorledes disse WWW-søgeværktøjer kunne bruges til at skabe kontakt mellem køber og sælger. Derfor skal denne delundersøgelse også ses som udviklingen af en forståelse for, hvorledes en kunde kan finde en virksomhed på WWW, der kan tilbyde løsninger på kundens

problem. Omvendt måtte vi også udvikle en forståelse for, hvorledes virksomheden kunne undersøge om andre (konkurrenter) på WWW, tilbød de samme løsninger. Samlet vil det sige at undersøge, hvorledes køber og sælger kan anvende WWW til informationsøgning.

4.1.3 Interaktion med fænomenet

Vores udgangspunkt for at beskæftiges os med WWW-søgeværktøjer var, at vi havde meget svært ved at finde det vi eftersøgte. Vi begyndte hyppigt at bruge diverse søgeværktøjers hjælpefunktion. Nogen søgeværktøjer var lettere at anvende end andre og nogen mere præcise end andre. Der begyndte at danne sig et mønster. Ved at læse artikler og søgeværktøjernes produktinformation blev vi klar over, at disse søgeværktøjer godt nok var gratis at benytte, men de måtte alligevel betragtes som kommercielle produkter. Endelig havde det også betydning, at vi søgte i en anden form for tekst (hypertekst) end den vante (lineær tekst). Dette stillede nogle nye krav til informationssøgeren. Samt at den information, vi fandt, var af meget forskellig kvalitet. Vi var ved at drukne i disse ustrukturerede erfaringer. Det blev et kaos af erfaringer, som vi ønskede at afklare.

4.1.4 Teoretisk perspektiv

Udfra vores erfaringer var der to teoretiske perspektiver, der på hvert sit niveau kunne hjælpe os til at reflektere over de tilegnede erfaringer og give nye erfaringer og dermed ny viden. Det filosofiske udgangspunkt for denne delundersøgelse var vores forståelse af vidensudviklingsprocessen. Denne iterative proces passede sammen med vores forståelse af det at navigere på WWW med det formål at skabe ny viden omkring et givet emne.

Dette videnskabsteoretiske perspektiv kunne kobles med et teoretisk perspektiv omhandlende informationssøgning på WWW. Vi fandt, at der var nogle teknisk betingede problemer i at navigere på WWW. Dette omtales på udemærket vis i Bryan Pfaffenbergers bog "Web Search Strategies", som udmærker sig ved at være en af de bedste bøger omhandlende WWW-søgning. Denne bog gav os inspiration til at reflektere over vores erfaring. Bryan Pfaffenberger har udarbejdet en 7-trins strategi for informationssøgning på WWW, som han inddeler således:

1. Know what you're up against
2. Find something that's relevant
3. Learn the techniques of deep searching
4. Perform a deep search
5. Explore specialized search resources
6. Keep up with what's new
7. Give something back⁵⁶

Bryan Pfaffenberger er en forfatter med et typisk WWW-fokus - han er amerikaner! I disse amerikanske bøger er der en tendens til, at den internationale dimension er begrænset til sporadiske benævnelser, som *World Wide Web* er *globalt* dækkende. Hvad dette betyder, samt en mere kritisk holdning til det globale i WWW er yderst begrænset. Derfor er Bryan Pfaffenbergers strategi fyldt med amerikanske søgeværktøjer, og derfor forholdte vi os kritisk til Pfaffenbergers 7-trins strategi. Imidlertid indeholder den mange udmærkede betragtninger, vi lod os inspirere af. Dog var det afgørende, at denne 7-trins strategi passede sammen med vores opfattelse af, hvorledes viden udvikles. Vores tolkning af strategien betød, at vi opfattede den som:

⁵⁶ Pfaffenberger: Web Search Strategies, p. 8-10

- Trin 1 indbefatter informationer, der gennemgår begrænsninger og muligheder i informationsøgning på WWW.
- Den viden, der skabes på baggrund af dette, giver et erfaringsgrundlag, som kan anvendes til selv at søge - altså trin 2.
- Erfaringen fra dette giver grobund for at udvikle sin viden om mere avanceret informationssøgning - trin 3.
- Hvor trin 4 giver anvisninger på, hvorledes denne dybere søgning udføres.
- Hvis disse søgninger ikke giver det fornødne resultat, og/eller man har mere specifikke formål med søgningen, kan det være nødvendigt at undersøge mulighederne for en specialiseret søgning, hvilket trin 5 er dedikeret til.
- Trin 6 er blot at undersøge løbende om nye informationer er kommet til siden sidste søgning.
- Trin 7 er en gennemgang af, hvorledes man selv kan oprette informationssøgningssider, så andre kan få tilgang til de informationer, ens viden er baseret på.

Dette lyder jo ganske udemærket, så hvorfor ikke blot opfordre læseren til at skynde sig hen i den nærmeste boghandel og købe bogen eller opfordre til at kigge forbi på bogens hjemmeside¹? For det første vil dette kun kunne lade sig gøre, hvis man opfatter verden som en objektiv og statisk størrelse, hvilket vi i vores videnskabsteoretisk diskussion har argumenteret for vil være forkert. For det andet er "Web Search Strategies" meget amerikansk, og kan derfor ikke stå alene. Endelig vil vores holdning til WWW-søgesystemer og indeks ikke blive beskrevet. Hans håndbogsagtige teori vil være formålsløst at beskrive, da vores holdning ikke vil komme til udtryk.

4.1.5 Dannelse af væsenstyper

De billeder, der havde dannet sig omkring, hvorledes en informationssøgningsproces kunne foretages, blev klarere ved at støtte sig til det teoretiske perspektiv. Vores opfattelse af de væsentlige trin i informationssøgningsprocessen blev som de følgende:

1. Forstå udfordringen i informationsøgning på WWW
2. Få et overblik over søgeværktøjernes kvaliteter
3. Situationsbestemt anvendelse af søgeværktøjerne
4. Undersøg specialiserede referencesystemer
5. Overvåg nye informationer på WWW
6. Transformer viden til information

Disse væsenstyper vil blive beskrevet i det følgende. Dog er der nogle generelle begrænsninger, som først vil blive præsenteret.

4.2 Søgesystemernes begrænsninger

Tilbage i problemformuleringen diskuterede vi forskellige opfattelser af WWW - medie eller marked. Denne delundersøgelses begrænsninger skal også ses i denne sammenhæng. De, som opfatter WWW som et medie, kritiserer WWW for manglende dækning og kvalitet. Dette bliver man nødt til at forholde sig til. Samtidig er WWW også et marked, hvorfor søgeværktøjerne også bør ses i denne sammenhæng. Delundersøgelsen kan dermed karakteriseres som et undersøgelse omhandlende en konkret dataindsamlingssteknik, men også som en markedsundersøgelse på

produktniveau, idet vi sammenligner forskellige søgeværktøjer fra forskellige udbydere. Karakteristisk for disse søgeværktøjer er, at de er gratis at anvende, men alligevel kommercielle produkter. Da søgeværktøjerne er kommercielle produkter, giver det en usikkerhed i informationssøgningens bredde og dybde. Samtidig er den information man finder ikke altid af høj kvalitet. Dette giver således en begrænsning i den fundne informations kvalitet.

Nedenstående begrænsninger bør derfor have i baghovedet, når læseren skaber sin egen viden baseret på den information delundersøgelsen udgør.

4.2.1 Begrænsninger i informationernes kvalitet og tilgængelighed

Når man beskæftiger sig med WWW bliver man let imponeret af de nye, spændende og overvældende muligheder, der ligger i anvendelsen af WWW. Efterhånden ligger der masser af tilgængelig information på WWW. Men ikke al information er tilgængelig via WWW og alt er ikke af ønskelig kvalitet. Clifford Stoll, forfatteren til den kritiske bog "Silicon Snake Oil" udtrykker sin mistillid således:

Researchers naturally save their best work to publish in journals and books, realizing that the review process ensures that better papers make it into print. They're unlikely to post good, original stuff to the network first; somebody might swipe their material. ... Pat Berger, former president of the American Library Association, points out that the Internet was created as a tool for researchers, not as an archive.⁵⁷

Denne kritik er berettiget, dog er det vigtigt at forstå, at Stolls baggrund i den funktionalistiske videnskab "astrofysik" gør, at han forventer at finde "sande" fakta om en objektiv verden. Derfor ser han det som værende negativt, at WWW er en dynamisk størrelse:

I started my own personal catalog of useful Uniform Record Locators; after one year half of them still worked. But the poetry section of my neighborhood library? They don't shuffle books around but on a decade. How should a book's bibliography reference this location? When the URL changes, the printed reference becomes meaningless. Even if the archive stays up, someone may update the files, confusing anyone looking for the original work. Libraries have a nifty system for storing books, it's the same one that I use to look 'em up. I'll bet librarians have an instruction list or some kind of cheat sheet. There's a place for every book. The Internet, however, has not such an organization--files are made available at random locations. To search through this chaos, we need smart tools programs that find resources for us. These tools work after fact: something's posted to an archive, then an automated cataloging tool discovers it and adds it to a catalog. These programs don't do a good job of searching, partly because they pay a lot of attention to titles and ignore files contents. Then, too, computer tools are great at finding strings of characters. It's much harder to search for meaningful information.⁵⁸

For os er det ikke specielt negativt, at WWW er dynamisk, dette afspejler blot vores verdensbillede, idet vi ser verden som dynamisk - WWW afspejler blot den dynamiske verden som helhed. Stoll har dog ret i, at det kan være en ulempe, at WWW er dynamisk, da den information man henviser til ikke længere er tilgængelig. Dette er dog ikke anderledes end med anden information. En bog kan være udsolgt og bliver ikke genoptrykt. En virksomhed kan være gået fallit eller opkøbt, en interviewet person kan have skiftet job eller kan ligefrem være afgået ved døden. Betyder det så, at referencen er ugyldig? Det mener vi ikke. Det hænger sammen med, at vores kvalitetskriterium for god forskning ikke nødvendigvis er, at forskningen kan gentages, men at vi har tydeliggjort vores tolkninger, således disse kan forfølges.

⁵⁷ Stoll: Silicon Snake Oil, p. 39

⁵⁸ Stoll: Silicon Snake Oil, p. 194

Yderligere angriber Stoll, at søgesystemerne kun arbejder med titler og ikke indholdet af de indekserede dokumenter. Som det vil fremgå senere, er dette ikke længere tilfældet, men Stoll har ret i at det er svært at søge efter, hvad han kalder for meningsfuld information. Når information giver mening, er der efter vores opfattelse ikke tale om information men om viden. Viden kan ikke slås op. Med reference til vidensudviklingsprocessen i kapitel 3, er viden et resultat af fortolkninger af information og af vurderingen af erfaring set retrospektivt, hvor både information og erfaring er affødt af handling.

Som det også fremgår af det ovenstående citat, sammenligner Stoll WWW med et traditionelt bibliotek. Dette kan skyldes at hans foretrukne måde at søge information på er at gå på biblioteket. På biblioteker findes kun dokumenter, der er beregnet til at man starter på side 1, følger forfatterens forudbestemte struktur og slutter på side "sidst". Derfor kritiserer han også måden et WWW dokument er opbygget:

These hypertext documents lack unity: one page may be dense with information another may be a vacuum. Its advocates speak of nonlinear text--you don't read a straight line of text, but can jump around as you wish. As a result, the author's logic, structure, and reasoning disappear. You get random facts. The World Wide Web seems well tuned for automated key-word searches. Your computer combs through files, picking out words in titles of text. It feels like you're riding atop of the World's finest index. You're not. Key-word-in-context lists are weak substitutes for a genuine index. Such searches profane the whole idea of research. In their facile expediency, they deliver information too effortlessly to be trustworthy. I know either the depth nor breadth of the search: what databases I did miss? Would I get different results if I'd capitalized the name?⁵⁹

Men at hypertext ikke er lineær er jo hele pointen. Læseren tror at vedkommende hopper rundt efter egen fri vilje, men dette er ikke tilfældet, da forfatteren bestemmer hvilke muligheder der er for at "hoppe rundt". Dette er særdeles interessant ved markedsføring, da sælger kan styre køber, uden køber ligger mærke til det.

Stoll har dog ret i, at det er særdeles vanskeligt at klargøre dybden og bredden af en søgning, men hvor tit kan man være opmærksom på dette. Da verden ændres konstant og da den er dynamisk, vil det aldrig være muligt at få fuld information om et givent emne. Imidlertid kan informationer om det søgeværktøj man benytter, hjælpe én i retning af at forstå bredden af søgningen. Dybden af søgningen kan også tilnærmes ved at kende søgeværktøjets kvaliteter.

4.2.2 Søgeværktøjerne som kommercielle produkter

De søgeværktøjer, der fokuseres på i denne delundersøgelse, er endnu gratis at benytte. Dog skal man stadig betale telefonregningen og have det nødvendige software installeret på den PC, man ønsker at benytte til informationssøgning. På trods af at søgeværktøjerne i traditionel forstand er gratis, er der stadig tale om kommercielle produkter, udbudt af virksomheder, som ønsker at tjene penge på WWW. En interessant artikel "Finans-eventyrer går på Nettet" beskriver markedssituationen således:

Man skal tilbage til forrige århundredes guldfeber i USA for at finde en tilsvarende eufori og magi, som det der i øjeblikket omgiver det famøse Internet. Fascination, nødvendighed og drømmen om en hurtig gevinst tiltrækker en skønsom blanding af fidusmagere, plattenslagere, multinationale koncerner, bladkonger, regeringer, organisationer og privatpersoner - og herhjemme tre af 1980'ernes mest spektakulære finansmænd.⁶⁰

⁵⁹ Stoll: Silicon Snake Oil, p. 197

⁶⁰ Jensen: Finans-eventyrer går på Nettet, Politiken den 28-2-1996

Forbilledet for mange, der søger at skabe en virksomhed i relation til WWW er firmaet Netscape, som har udviklet den mest brugte browser til navigering på WWW. Nedenstående citat udtrykker denne forventning:

I august sidste år [1995], chokerede det halvanden år gamle firma Netscape med formodentlig verdenshistoriens største værditilvækst, da det på den første dag på børsen i New York opnåede en børsværdi på ca. 30 mio. kr. uden på noget tidspunkt have præsenteret én eneste krone i overskud.⁶¹

I Danmark har en af det såkaldte spektakulære finansmænd set den enorme tiltrækning Internettet har. Klaus Riskær Petersen, som nu er direktør for en af de større danske Internet-udbydere "CyberCity" og formand for brancheforeningen af Internet-udbydere i Danmark, citeres for at sige:

..at han ikke opfatter Internet som en genvej til "lette og hurtige penge". Netmediet er et seriøs og kapitalkrævende forretningsområde, som efter hans opfattelse desværre også er truet af mindre pålidelige forretningsfolk, som "forsøger at slå plat på kassen".⁶²

Vi må altså her være opmærksomme på, at vi bevæger os ind på et område, hvor ikke alle oplysninger skal tages for pålydende. Når vi senere vil gennemgå de forskellige søgeværktøjer er forudsætningen, at nogle af udbydere tjener penge på at sælge annoncer på deres WWW-side eller på anden måde bruge tilgængeligheden af søgeværktøjerne i markedsføringen af deres virksomhed. Således ser de oplysninger om søgeværktøjerne som en forretningshemmelighed eller som en markedsføringsmulighed.⁶³ Det kan også betyde, at søgeværktøjet ikke er helt gratis. Som Claus Sølvesteen skriver i kronikken "Overlever avisen Internettet" omkring annoncering på WWW at:

Det bedste sted at annoncere er dér, hvor budskabet bliver set af den rette målgruppe. Kunderne er alfa og omega. Derfor oplever vi også i stigende grad, at man på Internettet skal betale for at få adgang til de interessante steder. Ikke med penge men med information om sig selv.⁶⁴

WWW er dynamisk, derfor ændrer konkurrencesituationen sig konstant, uanset om der betales for produkterne eller ej. Firmaet Cybernet.dk er Internet-udbyder, men udbyder samtidig to af de søgeværktøjer, der senere i denne delundersøgelse vil blive omtalt - søgesystemet Jori og emneindekset Jubii. Cybernet.dk er det danske firma, som er længst i udviklingen mod at tilbyde søgeværktøjer. En artikel om Jori nævnte kort baggrunden for, at Cybernet.dk havde udviklet Jori, idet Jacob Farvang fra Cybernet.dk er citeret for at sige:

Jori blev til på ni dage. Vi arbejdede meget hurtigt, fordi rygterne sagde, at Tele Danmark ville annoncere deres egen søgemaskine.⁶⁵

Man kan altså få på fornemmelsen, at der konkurreres om at være først. Der er ikke så meget tid til udvikling og test af disse søgeværktøjer. Jori er dog også senere blevet programmeret fuldstændig om, da den første version var langsom og fejlbehæftet. Dette bør informationssøgeren have i bagehovedet, når et søgeværktøj anvendes. Således vil et søgeværktøj som hævdes at være det stærkeste, ikke nødvendigvis være det, da det kan være fuld af fejl.

⁶¹ Jensen: Finans-eventyrere går på Nettet, Politiken den 28-2-1996

⁶² Jensen: Finans-eventyrere går på Nettet, Politiken den 28-2-1996

⁶³ Se Lycos og CyberCity Agents svar i afsnittet "Søgeværktøjer" i bilag 3, Dialog via e-post.

⁶⁴ Sølvesteen: Overlever avisen Internettet, kronikken i Politiken den 5-9-1996

⁶⁵ Jørgensen: Dansk stjerne på web, Intenetworld den 23-6-96

Udover der konkurreres på, hvor hurtigt der kommer ny produkter til, kører konkurrencen også på styrke i form af størrelse. Tele Danmark købte i februar 1996 den første og største Internetudbyder i Danmark - DKnet, og blev dermed den største Internetudbyder i Danmark med en markedsandel på over 50%. Dette affødte et andet opkøb, da Telepartner opkøbte Cybernet.dk i juni. Grunden til at Cybenet.dk blev solgt angav en af de tidligere ejere således:

Vi har siden januar fået henvendelser fra over ti firmaer. I starten takkede vi nej, men når Tele Danmark bruger millioner på markedsføring, så smitter det af.⁶⁶

For at forstå dette, må man forstå, at Tele Danmark i løbet af foråret 1996 omlagde deres "ikke-internet-baserede" onlinesystem Diatel til WWW med navnet Opassia. Dette emneindeks er en direkte konkurrent til Cybernet.dk's Jubii emneindeks. Dog er Opassia ikke særlig anvendeligt til informationssøgning, og er direkte rettet mod danske privatbrugere, og vil derfor ikke blive behandlet i denne delundersøgelse. Det essentielle er, at Telepartner og Cybernet.dk havde en fælles konkurrent Tele Danmark, som de ønsker at bekæmpe. Nedenstående citat forklarer, at Telepartner er ejet af endnu en af de 1980'ernes finansyrster:

Aldo Petersen, har siden 1988 drevet firmaet Telepartner...Firmaets aktiv er en database med en halv millioner fax-numre på virksomheder i Skandinavien, som bl.a. udgives som nationale telefax-kataloger i de tre lande. Aldo Petersen dukkede op i avisoverskrifterne i årets [1996] begyndelse, da han havde købt de skandinaviske rettigheder til et program, der gør det muligt at føre telefonsamtaler kloden rundt via Internet til lokaltakst.⁶⁷

Telepartner konkurrerer altså med Tele Danmark på to områder: Telefonbøger og langdistance telefontid, hvor man med Telepartner's software i stedet kan føre en samtale over Internettet og kun til lokaltakst. Ved købet af Cybernet.dk har Telepartner altså endnu et våben mod Tele Danmark.

Ved salget af Cybernet.dk ændredes leverandøren fra det indtrængende svenske teleselskab Telia til Global One. Hvorfor nu det? Ja, dels fordi, Telepartner også konkurrerer med Telia hjemme i Sverige på telefonbøger, og fordi Cybernet.dk nu også skal være Internetudbyder i Sverige. Som beskrevet i artiklen "Telia viser tænder i kampen om Internet", er Telia også stærkt interesseret i Internet som forretningsområde. Deres modspil er, at de har indgået en aftale med det amerikanske computerfirma Digital omkring søgesystemet Alta Vista. I artiklen beskrives aftalen bl.a. således:

Fra 1. oktober i år [1996] skal nordiske og baltiske Internetbrugere derfor ikke længere over dammen for at anvende World Wide Webs mest populære søgemaskine, men kan komme i kontakt med den heftige søgerobot i hjertet af Sverige - vel og mærke på deres eget sprog. Aftalen er en prestigesejr for Telia Telecom AB, fordi Alta Vista-teknologien stiller store krav til nettrafik og underliggende hardware. Således mener Digital, at kun ti til tolv lokaliteter på verdensplan kan tilbyde de nødvendige tekniske faciliteter - og Telia blev valgt som den første og hidtil eneste. For Telia Telecom AB ligner aftalen et væsentligt halvlegsresultat i kampen om fremtidens nordiske Internetkunder...⁶⁸

Dette citat synes som taget ud af Telia's pressemeddelelse. Senere i denne delundersøgelse, vil det vise sig, at Alta Vista ikke er det bedste system. Det der er vigtigt at bemærke her, at aftalen betegnes som værende en prestigesejr for Telia. Tilgangen til Alta Vista via Telia er frit tilgængelig også for Internetbrugere, der ikke er kunder hos Telia. Derfor kan Telia ikke direkte tjene penge på denne aftale. Som det også nævnes i artiklen giver aftalen Telia mulighed "for at spille med

⁶⁶ Marquart: Telepartner køber Cybernet.dk, ComputerWorld den 14-6-96

⁶⁷ Jensen: Finans-eventyrere går på Nettet, Politiken den 28-2-1996

⁶⁸ Steinmark: Telia viser tænder i kampen om Internet, ComputerWorld den 23-8-1996

musklerne og vise styrke". Nu begynder baggrunden for at tilbyde gratis søgeværktøjer at tone frem. Søgeværktøjerne er kommercielle produkter, som skal bruges til at markedsføre udbyderens egentlige produkt, det være sig en Internetopkobling eller et telefonabonnement. Samtidig er det på de WWW-sider, hvor søgeværktøjer befinder sig, mange brugere kommer, da de ellers ikke kan finde rundt på WWW. Det giver derfor mulighed for at annoncere for udbydernes egne produkter, og/eller udbyderen tager sig betalt for, at andre virksomheder kan annoncere på disse sider. For at slå dette fast kan det nævnes, at artiklen "Alta Vista baner vejen" skrevet af den samme journalist, som også skrev artiklen "Telias viser tænder i kampen om Internet", at Alta Vista betragtes af ejeren Digital som en mærkevare på højde med Coca Cola.⁶⁹ Man kan da mistænke Digital for at have kopieret Coca Colas mission om, at "Hele verdens befolkning skulle kunne række hånden ud og få en Coca Cola - uanset, hvor de befinder sig". Derfor skal aftalen med Telia ses som første skridt til at gøre Alta Vista og dermed Digital kendt - uanset, hvor brugeren befinder sig på WWW.

Uddybende kan det nævnes at købet af Cybernet.dk ikke blev til noget, og de to parter Cybernet.dk ApS og Telepartner skændes nu i pressen om, hvem der opsagde aftalen.⁷⁰ Samtidig er Telia voldsomt forsinket med at implementere Alta Vista.⁷¹ Dette viser blot hvor dynamisk WWW-verdenen er, og at visioner på WWW ofte bliver forsinket af tekniske vanskeligheder.

Denne gennemgang har været præget af danske og svenske eksempler. Det må dog antages, som eksemplet med Alta Vista angiver, at uanset hvor i verden et søgeværktøj stammer fra, vil det blive udbudt med kommerciel interesse for øje. Undtaget er de relativt færre og færre privatpersoner, som arbejder af ren og skær interesse. Endnu har disse kommercielle interesser ikke direkte påvirket kvaliteten af søgningen. Imidlertid peger tendenserne i retning af, at virksomheder kan betale for, at en søgning med et bestemt søgeord resulterer i et link til den pågældende virksomhed. F.eks. kunne søgeordet "rejse" resultere i "Spies". For at man som bruger kan undgå dette, planlægger nogle søgesystemer, at brugeren skal betale for adgangen til søgeværktøjet.⁷²

Med andre ord, denne delundersøgelse bør ses som en analyse, der sammenligner konkurrerende produkter indenfor forskellige kategorier, hvor de dog har det tilfælles, at de er rettet mod at hjælpe med til en lettere informationssøgning på WWW.

4.3 Forstå udfordringen i WWW-søgning

Som allerede nævnt, var WWW fra begyndelsen ikke udviklet til at tjene som et globalt informationssystem eller marked (Jævnfør diskussionen i problemformuleringen). WWW startede som et mindre eksperiment for dokumentudveksling. Men meget hurtigt voksede WWW - i udbredelse og anvendelsesmuligheder - i en grad de ikke havde forestillet sig i CERN, hvor WWW blev opfundet. Hvis de kunne have forudset denne udvikling, havde WWW været designet

⁶⁹ Steinmark: Alta Vista baner vejen, ComputerWorld den 23-8-1996

⁷⁰ Marquart og Danscher: Cyber-drengenes mareridt, og Telepartner kommenterer - Cybernet lyver, Communications World den 18-10-1996

⁷¹ Steinmark: Svenskere er pinligt forsinkede, Computer World den 18-10-1996

⁷² Laursen: Reklamens magt på Nettet, Politiken den 17-10-96

anderledes. Formålet med WWW var at udvikle et simpelt brugervenligt informationssystem for forskere med detaljeret kendskab til det emne informationen omhandlede.

Systemet blev opbygget omkring ideen om hypertext. Hvad dette giver af muligheder og begrænsninger for søgning på det WWW, der er en realitet i dag, vil blive gennemgået i dette afsnit.

4.3.1 Hypertekst og hyperlinks

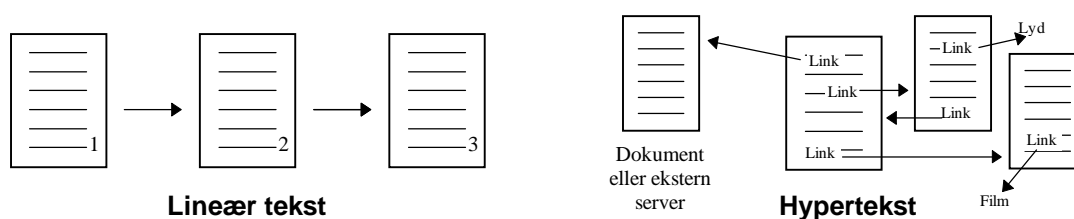
Forskningsministeriet i Danmark har udarbejdet nogle "Retningslinier for offentlige on-line publikationer", som har beskrevet den standard offentlige institutioner i Danmark skal følge ved publicering af dokumenter på en fællesstatslig informationsserver. I denne publikation beskrives hypertekst som:

Den tekst, der ses i WWW-browseren, er normalt en hypertekst. Teksten består af almindelig tekst og ofte også billeder, men i skærbilledet er der markeret særlige felter eller tekststykker, som rummer henvisninger. Disse henvisninger kan pege på andre filer eller dokumenter, og hvis brugeren aktiverer disse henvisninger, vil programmet foretage et hyperspring til det givne sted. En anden kendt form for hypertekst er hjælpefiler til Windows-programmer.⁷³

Disse hyperspring, som også kaldes for hyperlinks eller bare links, er beskrevet i den sammen publikation:

Et hyperspring er betegnelsen for den proces, der sker når en bruger aktiverer en henvisning i *hypertekstdokumentet* i sin WWW-browser fx med et museklik. Henvisningen kan pege på et andet sted i samme dokument eller på et andet dokument, som enten kan ligge på samme fysiske server eller et helt andet sted i verden. WWW-browserprogrammet og Internettet sørger så for, at der sker en opkobling til der, hvor dokumentet ligger, og at det overføres til og vises på brugers maskine. Man kan hyperspringe helt vilkårligt og for det meste følge sin vej tilbage gennem de allerede sete dokumenter for siden at vælge nye forgreninger.⁷⁴

Hypertekst er altså en ikke-lineær metode til at se tekstbaseret information. I et hypertekst baseret dokument, vælger læseren sin egen vej gennem dokumentet ved at klikke på nogle fremhævede links. Denne måde at tilegne sig information på, menes at svare bedre til menneskets måde at lære på, da læseren af en hypertekst, i modsætning til et almindeligt dokument, ikke er tvunget til at følge forfatterens struktur. I stedet bruger læseren intuition og associationer til at bevæge sig igennem teksten. Nedenstående figur viser forskellen på et "almindeligt" dokument og et hypertekst baseret dokument:



Figur 4.1: Lineær tekst versus hypertekst. Kilde: Turlington: *Walking the World Wide Web*, p. 11

Her er det vigtigt at genkalde sig Stolls kritik af hypertekst. Nogle sider kan være fyldt med den information, man leder efter, og andre er, hvad han kalder for et vakuum - uden den information man ledte efter. Søgemetoden i at klikke sig igennem teksten er, hvad Pfaffenberger kalder for browsing:

In a hypertext, a method of information access in which you explore what's available by clicking hyperlinks that look interesting or useful.⁷⁵

⁷³ Forskningsministeriet: Retningslinier for offentlige on-line publikationer, p. 44

⁷⁴ Forskningsministeriet: Retningslinier for offentlige on-line publikationer, p. 44

⁷⁵ Pfaffenberger: *Web Search Strategies*, p. 19

Men husk på, at den dygtige forfatter udnytter læserens intuition og associationer, således læseren tror han selv vælger. Pfaffenberger ser da heller ikke browsing som en særlig optimal form for informationssøgning, idet han kalder afsnittet omkring browsing for "Browse Your Way to Disorientation". Dette var i WWWs unge dage den eneste form for informationssøgning. Browsing er i dag også kendt som surfing. Uanset navnet er der dog en fordel ved browsing/surfing. Læseren opdager nye WWW-sider, som vedkommende ikke havde drømt om at søge på. Dette svarer omtrent til den største fordel ved en avis. Når man skimmer avisen igennem for at komme om til sporten (eller tegneserierne eller noget helt tredje, der er ens primære interesseområde), bliver man måske opmærksom på en ny politisk udvikling, der synes interessant. Hvis man skulle købe avisen i småbidder, havde man ikke købt den politiske sektion.

For at kunne forstå hvorledes man ellers kan søge på WWW vil det først være nødvendigt at forstå, hvorledes en WWW-side er opbygget. Det vil vi tage fat på i det følgende afsnit.

4.3.2 Dokumentkodningsstandarden HTML

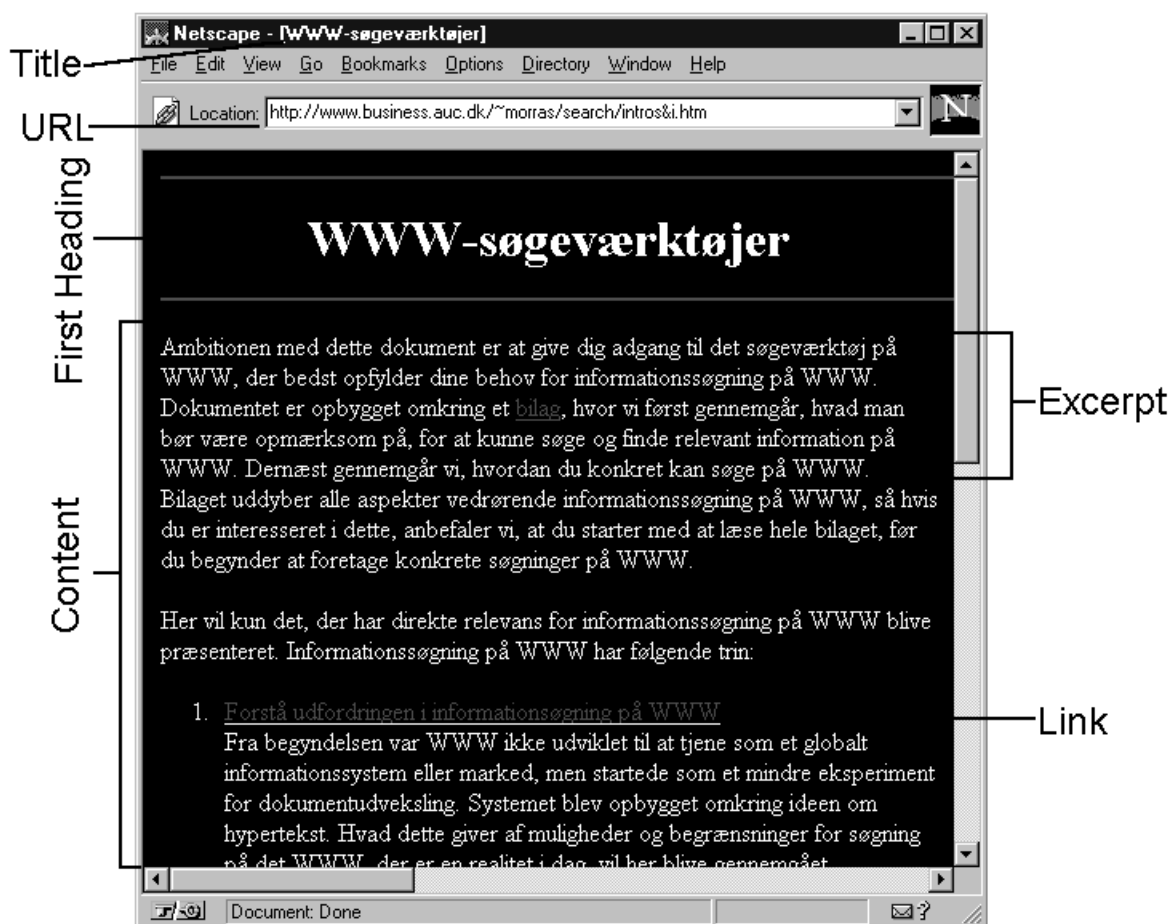
Som nævnt i den videnskabsteoretiske diskussion ligger der nogle klare lovmæssigheder til grundlag for opbygningen af en side på WWW. Disse lovmæssigheder er nedfældet i dokumentkodningsstandarden HTML. Forskningsministeriets "Retningslinier for offentlige on-line publikationer" forklarer HTML således:

HyperText Markup Language (HTML) er det kodeformat, man bruger til at skrive WWW-dokumenter i. Koden består af en række mærker (tags) angivet i mindre-end/større-end tegn, der lægges i en almindelig tekst f.eks. eller . Mærkerne fortæller WWW-browseren, hvad den pågældende tekstfunktion er. Browseren afgør derefter, hvordan det skal fortolkes og vises på skærmen, fx om det er en tekst, der skal fremhæves med fed skrift, om det er et billede, der skal vises i teksten, eller om det er en henvisning til et andet sted på WWW.⁷⁶

Som det fremgår er der tale om nogle meget stringente lovmæssigheder, som ikke altid er lige simple at forstå, men som dog alligevel har en meget klar indre logik. Så selv om der på WWW ofte hersker en stemning af anarki, er forudsætningerne "bagved" den grafiske brugerflade meget entydige og meget funktionalistiske. Denne standard består af mange dele, men det der er essentielt i WWW-søgninger er, hvorledes siden er opbygget samt de overordnede koder (Meta Tags), som kan fortælle et evt. søgesystem, hvad WWW-siden indeholder.

Det er essentielt at forstå, hvorledes en WWW-side er opbygget, da det er forskelligt, hvad de forskellige søgesystemerne tager med i deres søgning, da ikke alle komponenterne bliver indekseret. Nedenstående figur viser en typisk WWW-side, hvor de engelske udtryk er bibeholdt, da er det disse, søgeværktøjerne primært anvender:

⁷⁶ Forskningsministeriet: Retningslinier for offentlige on-line publikationer, p. 44



Figur 4.2: Komponenter i en typisk WWW-side. Kilde: Pfaffenberger: *Web Search Strategies*, p. 145

Figur 4.2 illustrerer komponenterne:

- Uniform Resource Locator (*URL*) er betegnelsen for den unikke adresse, en given ressource har på WWW. Man angiver en URL, når man ønsker, at ens browser skal vise indholdet af en given WWW-side. Henvisninger til steder på WWW gives med URL'er, fx. <http://www.i4.auc.dk/rask/>.
- *Title* er WWW-sidens titel, som optræder mellem <TITLE>...</TITLE> koderne i et HTML dokument. Normalt kan titlen ses i browserens titelbjælke. Problemet ved informationsøgningen er, at ikke alle forfattere vælger en beskrivende og informativ titel.
- *Headings* er overskriften på siden, denne er ofte mere informativ end titlen, da forfatteren opfatter første overskrift som sidens titel.
- *Content* er indholdet af hele sidens tekst inklusiv introduktionen og overskrifterne.
- *Excerpt* er introduktionen til siden, hvilket ofte betyder de første få linier, der beskriver formålet med siden. Men ikke altid, da de første linier kan være forbeholdt links til en overordnet indholdsfortegnelse med WWW-sidens, WWW-stedets og/eller hjemmesidens hovedelementer.
- *Hyperlinks* kan også være inkluderet i søgesystemets database.⁷⁷

⁷⁷ Pfaffenberger: *Web Search Strategies*, p. 144-145

- *Andre komponenter*, kommer hele tiden til, dette kan være billeder (image), programstumper (applets) og meget anden information, der ikke er tekstbaseret.

Foregående komponenter er især vigtige at have i baghovedet, når et givent søgesystem skal anvendes. Dog vil indeksene også kunne anvendes bedre, når man forstår hvorledes en WWW-side er opbygget. Hvordan man konkret anvender søgesystemer og indeks vil vi komme tilbage til senere. Inden dette er vi imidlertid også nødt til at forstå den usynlige del af en WWW-side.

Alle WWW-sider har denne usynlige kode:

```
<!doctype html public "-//IETF//DTD HTML//EN">
<HTML>
<HEAD>
<TITLE>Sidens titel</TITLE>
<META name="navnet på informationen i siden" content="information man vil give videre">
[fx:]
<META name="author" content="Forfatterens navn">
</HEAD>
<BODY>
Her følger sidens synlige tekst m.m.
</BODY>
</HTML>
```

Det vigtige er her at ligge mærke til koden "META". Ved anvendelse af denne kode er det muligt at beskrive alt det ved siden, der er vigtigt for en søgning. Forskningsministeriet har ved at udgive "Retningslinier for offentlige on-line publikationer" prøvet at udarbejde en standard for offentlige WWW-sider, hvor de især har anvendt koden "META" til at konstruere en standard for en elektronisk kolofon⁷⁸. Med denne kode "META" som kolofon, skulle det være nemt at søge på WWW - det er det bare ikke. Koden bliver oftest ikke brugt, der er meget få standarder for, hvorledes kolofonen skal udformes (Forskningsministeriets initiativ er undtagelsen). Dog anvender flere og flere søgesystemer informationerne i en sådan Meta Tag.

Måske vil disse problemer ikke blive løst, da WWW stadig er præget af anarki. Dog vil det være fornuftigt at undersøge søgesystemets evne til at medtage Meta Tag. Dette er især relevant, hvis formålet er markedsføring af virksomheden ved hjælp af eget WWW-sted.

4.3.3 Forudsætninger for at kunne tage udfordringen i WWW-søgning op

Som de ovenstående afsnit angiver, er informationssøgning på WWW ikke identisk med almindelig informationsøgning. Derfor må den *første* forudsætning være, at man gør sig klart, hvad WWW grundlæggende består af. Dette skulle være på plads nu. Den *anden* forudsætning må være en tilpas mængde baggrundsviden om det emne, man ønsker yderligere information om. Ellers kommer man til at lide af "information overload" eller bliver "lost in Cyberspace". *Tredje* forudsætning er, at man lader sin intuition komme frem, associere og tænke i synonymer, når søgningen foretages. *Fjerde* aspekt, er, at man bør være meget kritisk. Ikke alt man finder er brugbart, så en sund skepsis er at foretrække, når informationen fortolkes. *Femte* forudsætning, er at huske at starte forfra igen. Brug erfaringen fra første omgang søgninger til at specificere søgningen og sortere resultaterne.

⁷⁸ En elektronisk kolofon er et sæt koder, der identificerer en række bibliografiske oplysninger om publikationen.

Denne baggrund er vigtig at have, når vi nu går i gang med en gennemgang af de forskellige værktøjer til informationssøgning på WWW.

4.4 Få et overblik over søgeværktøjernes kvaliteter

Der er hundredvis af forskellige værktøjer, der kan anvendes ved søgning på WWW. I dette afsnit vil vi opdele de forskellige værktøjer i fem forskellige typer, samt gennemgå de forskellige typers fordele og ulemper. Desuden vil vi opstille nogle vurderingsparametre, der gør det muligt at vurdere bredden og dybden af et givent søgeværktøj.

Værktøjerne er behandlet i den rækkefølge, det vil være mest optimalt at anvende dem, hvis WWW er helt nyt for én. Disse værktøjer er opdelt efter karakteristika, men fire af dem; *udgangspunkter*, *emneindeks*, *vejledere* og *bogmærker* er alle forskellige former for indeks. Den sidste type af værktøjer *søgesystemer*, hvilket er rene søgesider.

To yderligere værktøjer, som ikke behandles her, er agenter og personlig service. En agent er et computerprogram, som konstant leder efter information man i forvejen har angivet, at man er interesseret i. Disse er stadig i deres vorden og har endnu en begrænset anvendelsesmulighed. Det andet værktøj som ikke er behandlet er personlig service, som i virkeligheden betyder, at man betaler andre personer for at gennemføre søgningen med de værktøjer det følgende vil omhandle.

4.4.1 Udgangspunkter

Udgangspunkter (Starting Points) er ofte et indeks som enten følger med WWW-software, er tilgængelig på elektroniske avisers og blades hjemmeside, eller andre virksomheder eller privat personer, der er interesseret i at vise mangfoldigheden i WWW. Dette kan være et udemærket udgangspunkt for nybegyndere, men man skal ikke forvente at disse indeks dækker alt. I virkeligheden er disse indeks udarbejdet som udgangspunkt for browsing eller surfing, hvilket vi allerede har afskrevet som en solid metode for informationssøgning på WWW. De har dog den fordel, at man hurtigt finder steder med relevans for sin søgning.

4.4.2 Emneindeksene

Emneindeksene (Subject catalogs) er opstået som en WWW pendant til bibliotekernes emneindeks. Således vil man, hvis man søger oplysninger uden at kende forfatter eller titel, slå op under "hovedemne" - "underemne". Emneindeksene er altså et WWW-sted, hvor personer manuelt har kategoriseret nogle WWW-sider ved at bruge emneklassificeringer. Man klikker på links, der fanger ens interesse, indtil man klikker på link, der fører én væk fra emneindekset og til et andet sted på WWW, der kan forsyne én med de oplysninger man søger. Emneindeksene er i modsætning til udgangspunkterne meget mere omfattende, selvom de dog kun dækker en brøkdel af WWW.

Normalt er emneindeksene vedligeholdet af personer, som sidder centralt og finder og klassificerer WWW-siderne. En variant heraf er decentrale emneindeks, hvor man trækker på forskellige (op til flere hundrede) personer med speciel ekspertise på et område.

Uanset om emneindekset er centralt eller decentralt, er ulemperen, at de ser omfattende ud, men ikke dækker hele WWW. Samt at der sidder personer, som udvælger hvilke WWW-sider, der skal med i oversigten. Det vil sige, at andre personer har afgrænset ens muligheder for at finde den

information, man ønsker. Værre bliver det af (især ved de centrale emneindeks), at disse personer ikke har specielt kendskab til det emneområde de vælger at inkludere (eller ikke inkludere) i indekset. Det vil ofte blive deres personlige holdninger og motiver, som afgør valgene. Da ingen indeks beskriver deres normer, værdier, holdninger osv. eksplicit, men da WWW-sider med flot computergrafik er godt repræsenteret, må det vurderes, at det ofte er grafikken på siden og tilfældigheder, der afgør om siden er med i indekset.

4.4.3 Vejledere

Vejledere er oprettet og vedligeholdt af personer med viden omkring et givent meget mere specifikt emneområde som denne form for emneindeks dækker. Vejledere kan derfor være meget værdifulde i en søgningsproces. Vejledere bliver (på engelsk) kaldt meget forskelligt som: collection, compendium, guide, index, indices, meta-page, subject guide, subject index og trailblazer pages. Derved fremgår det, at ulempen ved vejledere er, at de kan være svære at lokalisere. Selvom de personer, der tilbyder en vejleder, kan være meget vidende om deres felt, har vejledere stadig den begrænsning, at andre personers holdninger og interesser begrænser søgningen.

4.4.4 Søgssystemer

Søgssystemer (Search Engines) blev udviklet, da det blev for uoverskueligt at indekser alle WWW-sider under emnegrupper. Søgssystemer bygger på en såkaldt robot (også kaldet spider, wanderer eller worm). Kort sagt, en robot er et software program, som konstant skanner WWW for nye eller opdaterede WWW-sider som den tilføjer i en database. Teoretisk set vokser databasen kontinuerlig med, at WWW vokser i omfang. Selve søgssystemet er et andet program, hvor man kan indtaste nogle nøgleord, som programmet så søger i databasen efter. Søgningens resultat bliver listet efter relevans, således man kan klikke videre til den WWW-side, der kan forsyne én med den information man søger.

Dette er jo udemærket, idet ingen person har bestemt, hvad der skal indekseres, og således er der ingen, som begrænser mangfoldigheden i den information, man har mulighed for at finde. Her bliver problemet dog hurtigt, at man får alt for mange svar tilbage, hvor langt størstedelen ikke har relevans for ens søgning. Med andre ord er de fleste søgssystemer ikke særlig præcise. Nogle søgssystemer søger i en database som ikke indeholder alle dele af WWW-siden. Disse søgssystemer kan derfor melde tilbage uden relevante oplysninger, da søgeordet er gemt langt væk i selve teksten. Da ikke alle søgssystemer indekserer selve teksten betyder det, at en søgning i et system har en anden dybde end et andet søgssystem - søgninger får dermed forskellig dybde. Et andet problem er, at de svar søgssystemet giver kan have forskellig kvalitet, da robotten ikke er i stand til at kvalitetsvurdere WWW-siden.

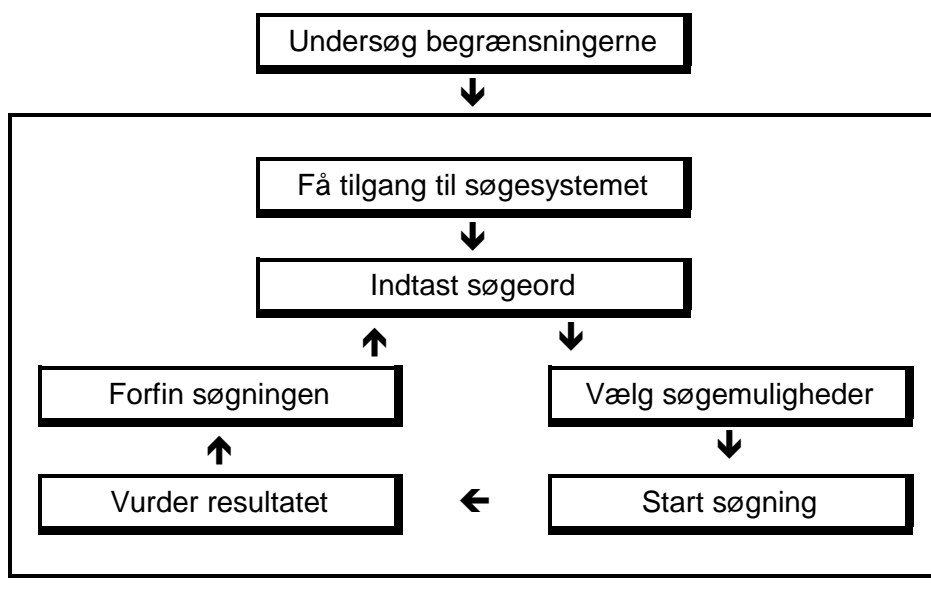
Dog har nogle af søgssystemerne en løsning på ovenstående problemer, da det er muligt at foretage avancerede søgninger, hvor man kan specificere søgningen. Imidlertid er ulempen, at ikke alle søgssystemer er nemme at lave avancerende søgninger i.

Som det fremgår af ovenstående, kan ét søgssystem være meget forskelligt fra et andet med hensyn til at kontrollere bredden og dybden af søgningen. Det er derfor det sikreste at søge i flere søgssystemer. Som løsning på dette problem er der udviklet søgssystemer som søger i mange andre

søgesystemer samtidig. Dog er ulemperne ved disse søg-i-et-systemer, de samme som ved almindelige søgesystemer, og at det sjældent er muligt at kontrollere søgningen. Endelig er der også WWW-sider, der samler en masse forskellige søgesystemers søgefelt på en WWW-side, men igen er det svært at kontrollere søgningen. Dog har det den fordel, at søgesystemer med fokus på emneområde er inddelt under emner efter samme princip som emneindeksene.

Et sidste aspekt man bør være opmærksom på er, at jo bedre søgesystemerne er, desto mere belastede er de, og derfor kan det være svært at få adgang til dem i løbet af arbejdstiden. Da de amerikanske systemer er særdeles anvendt af amerikanere er man tvungen til at undersøge, hvornår amerikanerne arbejder. Således kan "global village"-ideen med at tid og sted forsvinder, nogen gange være svært at få øje på.

Uanset, hvilket søgesystem der benyttes, vil den generelle procedure være følgende:



Figur 4.3: Generel anvendelses procedure for et søgesystem

4.4.5 Bogmærker

Når en given søgning har givet positivt resultat er det vigtigt at gemme disse til evt. senere brug. Browseren har faciliteter, der muliggør dette. Der kan yderligere anskaffes programmer med mere avancerede funktioner. Her skal der dog ikke gås i dybden med faciliteterne i de enkelte software pakker. I stedet vil vi blot gennemgå principperne i denne form for søgeværktøj. Med vidensudviklingsprocessen i baghovedet kan dette søgeværktøj ses som et erfaringsværktøj, forstået på den måde, at når en succesfuld søgning er gennemført, vil en tolkning af den pågældende information give meningsfuld information, altså viden. Denne søgning gemmes som et bogmærke, man nemt kan vende tilbage til. Erfaringen kommer ind, når disse skal organiseres.

En simpel men effektiv plan for organiseringen er, at man grupperer søgeværktøjerne, samt at man samler de WWW-sider, som har relevans i relation til et emneområde. I takt med, at man får erfaring med WWW-søgning og med emneområdet, vil kvaliteten af ens bogmærker stige.

En udemærket metode til at gruppere bogmærkerne under overskrifter, kunne være at følge denne opdeling:

- Udgangspunkter
- Emneindeks
- Søgesystemer
- Emne område #1
 - ⇒ Vejledere
 - ⇒ Overblik
 - ⇒ Rigtige gode informationer
 - ⇒ Anden god information
 - ⇒ Tvivlsom information
- Emne område #2

Fordelen ved bogmærkerne er, at ens erfaring kan anvendes. Ulempen er, at ingen ny information kommer for dagens lys ved at bruge dette søgeværktøj, men dog undgår man at glemme, hvor man fandt gode informationer. Man må dog huske på, at disse bogmærker bliver forældet, og at de registrerede WWW-sider ændres eller forsvinder.

4.4.6 Fordele og ulemper ved de forskellige typer af søgeværktøjer

Som det fremgår af det ovenstående er der fordele og ulemper ved de forskellige typer af søgeværktøjer. Nedenstående figur opsummerer disse, hvor de fire første typer af søgeværktøjer følger en rækkefølge, der er bestemt udfra, hvor dækkende det enkelte værktøj i princippet er. Dette er set i forhold til den information man søger, samt hvor nemt søgeværktøjet er at anvende. Det sidste søgeværktøj - bogmærker - er uden for denne rangering, idet ingen ny information kan søges, men bogmærker kan i stedet betragtes, som en værdifuld opsummering af foretagne søgninger.

Rangering	Værktøjer	Fordele	Ulemper
Mindst dækkende Nemmeste anvendelse ↑	Udgangspunkter	<ul style="list-style-type: none"> • Giver umiddelbart en god ide om WWWs mangfoldighed 	<ul style="list-style-type: none"> • Ikke dækkende overhovedet • Andre personers holdninger og interesser begrænser søgningen
	Emneindeks	<ul style="list-style-type: none"> • Meget omfattende oversigter • Nemme at anvende 	<ul style="list-style-type: none"> • Ikke komplet dækkende • Ikke faglig kompetente personer foretager kvalitetsvurderingen • Andre personers holdninger og interesser begrænser søgningen
	Vejledere	<ul style="list-style-type: none"> • Faglig kompetente personer foretager kvalitetsvurderingen og indekseringen • Kan være guld værd 	<ul style="list-style-type: none"> • Svære at finde frem til • Andre personers holdninger og interesser begrænser søgningen
	Søgesystemer	<ul style="list-style-type: none"> • Ingen personers holdninger og interesser begrænser søgningen • Præcis søgning er mulig, hvis det enkelte systems muligheder og begrænsninger kendes 	<ul style="list-style-type: none"> • Sværere at anvende med et godt resultat • Kan være overbelastede • Ofte nødvendigt at foretage søgninger i flere systemer
↓ Mest dækkende Sværeste anvendelse	Bogmærker	<ul style="list-style-type: none"> • Anvender aktivt erfaring i søgningsprocessen • Undgår at man glemmer gode søgningers resultat 	<ul style="list-style-type: none"> • Ingen ny information findes ved anvendelse af dette værktøj

Figur 4.4: Fordele og ulemper ved de forskellige typer af søgeværktøjer

Fordelene og ulemperne i figur 4.4 skal ses som, hvad der er karakteristisk for den enkelte type af søgeværktøj. Udbydere af et søgeværktøj har forsøgt at udnytte fordelene og begrænse ulemperne. Således vil det være nødvendigt at opstille nogle vurderingsparametre, der kan anvendes ved valget imellem forskellige udbydere af den samme type af søgeværktøjer.

4.4.7 Vurderingsparametrene når en udbyder af et søgeværktøj skal vælges fremfor en anden

De følgende parametre vil ikke være relevant for bogmærker, da dette søgeværktøj ikke anvendes til at finde ny information. Ikke alle parametrene vil være relevante for alle de andre søgeværktøjer. Dette gælder især mulighederne for avanceret søgning, da nogle af søgeværktøjerne kun understøtter simpel søgning og nogle udelukkende baseres på browsing. Et andet aspekt er, at søgeværktøjerne er kommercielle produkter. Således ser udbydere oplysninger om de følgende parametre som en forretningshemmelighed, og/eller ser parameteren som en markedsføringsmulighed. Derfor har det ikke været muligt at få oplysninger om disse parametre, og derfor skal de tilgængelige oplysninger tages med et gran salt.

Oplysninger om *bredden af søgningen* kan tilnærmes, ved at have kendskab til hvilken *type* af sider, der tages med i søgeværktøjet, samt *antallet* af sider, der er tale om. *Typen* kan udover hypertextbaserede WWW-sider være Gopher, FTP og/eller Usenet. Gopher er et menustyret "go for" søgesystem, der tilbyder nogenlunde den samme information som WWW.⁷⁹ Gopher startede i 1991, men er på vej ud, da WWW er endnu mere brugervenligt. File Transfer Protocol (FTP) er en protokol, der beskriver, hvordan filer skal sendes via f.eks. Internet. FTP bruges til overførsel af binære filer, som kan være alt fra computerprogrammer til regnearksfiler.⁸⁰ Usenet (NetNews) er et verdensomspændende kommunikationssystem med tusindvis af diskussionsgrupper.⁸¹ Vi vil her fokusere på WWW, da de andre typer ikke er brugbare for markedsføring af en virksomheds produkter, og da den nuværende og fremtidige udvikling vil koncentrere sig om WWW. Det er svært at sige noget som helst om størrelsen, når det handler om WWW. Antallet af sider udbydere af søgeværktøjerne hævder, at søgeværktøjet har indekseret, kan kun bruges som et fingerpeg om, hvor bred søgningen kan være. På trods af, at WWW (af nogle) bliver opfattet som værende et sted (Cyberspace) uden tid og sted, er nogle søgeværktøjer dedikeret til en bestemt geografisk lokalitet, hvorfra informationen er tilgængelig. Andre søgeværktøjer gør det muligt at begrænse informationssøgningen til bestemte geografisk lokaliteter. Derfor kan et søgeværktøjs *geografisk dækning* anvendes til at kontrollere bredden af søgningen.

Oplysninger om *dybden af søgningen* kan tilnærmes ved at have kendskab til, hvilke *komponenter* af en WWW-side, der tages med i søgningen. De synlige dele af WWW-siden er som før nævnt URL, Title, Headings, Content, Excerpt og Hyperlinks. Den usynlige Meta Tag er også vigtig, da nogle søgesystemer bruger denne som den primære informationskilde til at indeksere WWW-siden. Relateret til dette er *resultat præsentationen* af søgningen. Søgnesystemerne præsenterer ofte en del af WWW-siden som resultat, og ved de forskellige indeks er det relevant om den medtagne

⁷⁹ Heilesen: Helt fra start på Internet og World Wide Web, p. 33

⁸⁰ Forskningsministeriet: Retningslinier for offentlige on-line publikationer, p. 39

⁸¹ Heilesen: På Internet med World Wide Web, p. 35

WWW-side karakteriseres med en beskrivelse og/eller en form for kvalitetskarakter (oftest med stjerner som, ved vurderinger af film). En anden parameter, der kan gøre søgningen nemmere, er mange søgeværktøjers relevanslister som resultaterne rangeres efter.

Følgende parametre er kun anvendelige ved søgesystemer og ved de indeks, hvor det er muligt at foretage en søgning. Vi vil først gennemgå de parametre, som er standardindstillingerne ved en simpel søgning. Heraf vil det fremgå, at en simpel søgning med samme søgeord, men med forskellige søgesystemer eller indeks, kan give et meget forskelligt resultat.

- *Afskæring af søgeord* kan være Automatisk, Præcis eller Stamme. Automatisk (Automatic) er når søgningen giver resultat, hvor søgeordet indgår i et ord i de søgte sider. På den måde vil søgeordet "skibstøm" give resultater, hvor fx. der står "skibstømmer" eller "skibstømmer" indgår. Præcis (Exact word match) giver kun resultat, når søgeordet findes i en side, hvor søgeordet er et selvstændigt ord. Endelig er der Stamme (Stemming), hvor søgningen reducerer ordet til dets ordstamme eller til dets entalsform. Her vil søgningen baseret på et dansk søgeord oftest fejle, når søgningen fortages med et amerikansk søgeværktøj, da det bygger på amerikansk grammatik.
- Således bør man være opmærksom på, hvilket *sprog og grammatik* søgesystemet anvender. Ikke blot ved afskæring af søgeordet til dets stamme, men også om søgesystemet kan forstå tegn såsom det danske "å".
- En anden standardindstilling ved simpel søgning er om *Store bogstaver* genkendes, hvilket man også bør huske på - især ved søgning på egennavne.
- Den sidste parameter, der er standardindstillet ved en simpel søgning, er *Behandlingen af flere søgeord*. Nogle søgeværktøjer sætter automatisk den logiske operator AND ind mellem søgeordene, andre OR og andre igen ser søgeordene som en frase.

Disse standardindstillinger ved en simpel søgning kan med nogle søgeværktøjer ændres. Når dette er muligt taler vi om *avanceret søgning*.

- En parameter, der er god til at kontrollere dybden af søgningen, kaldes for *Begrænset feltsøgning*. En database er opdelt i forskellige felter som repræsenterer forskellige typer af information. Søgeværktøjer, der tillader at begrænse søgningen til et bestemt felt, kan være særdeles anvendelig til fx. kun at søge på titler af WWW-sider.
- En af de mest præcise former for søgninger man kan foretage, er søgninger med flere søgeord som en *frase*. Dermed kontrollerer man, i hvilken rækkefølge ordene skal stå, for at søgningen bør give resultat.
- *Koncept*søgning er en teknologi, som med hensyn til præcision minder om frasesøgning. Derfor kan konceptsøgning være ligeså præcis som frasesøgning.
- En anden måde at kontrollere dybde af søgningen, vil være at bruge *logiske operatører*. Figur 4.5 angiver funktionen af forskellige operatører:

Operator	Betydning	Aliaser	Eksempel	Forklaring
P AND Q	Udtrykket skal forekomme	+, &	ælling AND andersen	Både ælling og Andersen
P OR Q	Udtrykket kan forekomme	-,	ælling OR andersen	Enten eller
P NOT Q	Udtrykket må ikke forekomme	!	ælling NOT andersen	ælling men ikke Andersen
P NEAR Q	Udtrykket skal forekomme inden for en vis afstand	~	ælling NEAR andersen	Ælling tæt på Andersen
P ADJ Q	Udtrykkene skal forekomme lige opad hinanden	Ingen	grimme ADJ ælling	Ælling ved siden af grimme
P*	Kan erstatte et eller flere tegn	Ingen	anders*	Anders med forskellige endelser
"P Q R"	Hele udtrykket skal forekomme	Ingen	"den grimme ælling"	Hele titlen skal forekomme
(P+Q)	Udtrykket skal evalueres samlet	Ingen	(ælling NEAR grimme) AND andersen	Ælling og grimme skal stå tæt ved hinanden i dokumenter med Andersen

Figur 4.5: Logiske operatører. Kilde: Jørgensen, *Web-søgessystemer, Internetworld nr. 1*, p. 29ⁱ

Således vil det i mange tilfælde være muligt at få kendskab til bredden og dybden af den foretagne søgning. Figur 4.6 viser de parametre, man skal være opmærksom på, når man vurderer et søgeværktøj:

	Parametre	Eksempler
Bredden af søgningen	<ul style="list-style-type: none"> • Typer af dokumenter • Antal af WWW-sider • Geografisk dækning 	<ul style="list-style-type: none"> • WWW, Gopher, FTP, Usenet, m.fl.
Dybden af søgningen	<ul style="list-style-type: none"> • Komponenter • Resultat præsentation • Afskæring af søgeord • Sprog og grammatik • Store bogstaver • Behandling af flere søgeord • Begrænset feltsøgning 	<ul style="list-style-type: none"> • De synlige: URL, Title, Headings, Content, Excerpt, Hyperlinks og den usynlige Meta Tag • Komponenter, relevans, kvalitetsbeskrivelse og -stjerner • Automatisk, Præcis eller Stamme • Fx. amerikansk, dansk eller japansk • Altid, Aldrig eller kun ved egennavne • Logiske operatører, frase- og konceptsøgning • URL, Title, Headings, Content, Excerpt, Hyperlinks, image osv.

Figur 4.6: Parametre der angiver bredden og dybden af søgningen

Endelig vil *Bemærkninger* være forskellige andre vurderingsparametre vi vil fremhæve ved et søgeværktøj, og som kan være anvendeligt i specielle søgesituationer.

Nu er så spørgsmålet, i hvilken situation man med fordel kan anvende en type af søgeværktøj frem for en anden, samt hvilke søgeværktøjer af hver type, der er anbefalelsesværdig. Dette vil det følgende afsnit omhandle.

4.5 Situationensbestemt anvendelse af søgeværktøjerne

Inden man går i gang med at analysere, hvad ens informationsbehov er, kan det være en fordel at kende til WWWs mangfoldighed. Som det fremgik af forrige afsnit er netop søgeværktøjet Udgangspunkter fortrinligt til dette formål. Hvis man derimod er ude efter specifik information om et specifikt emne, vil en Simpel søgning i et søgesystem være at foretrække, da den er hurtig at foretage. Formålet kan også være, at man ønsker god information af høj kvalitet. Her vil søgningen i Emneindeksene være at foretrække, således at man finder gode Vejledere, der kan give den ønskede information. Hvis man er meget ambitiøs og ønsker at finde alt WWW-tilgængelig information om et givet emne, bliver man nødt til at ty til Avancerede søgninger i omfattende søgesystemer. Vi har derfor opdelt det følgende i fire underafsnit: 1) WWWs mangfoldighed, 2) Specifik information, 3) Information af høj kvalitet og 4) Samtlige WWW-tilgængelige informationer. I disse underafsnit, vil vi sammenligne konkrete søgeværktøjer i forhold til målet man har med søgningen. Under hvert underafsnit vil vi udvælge de bedste søgeværktøjer samt danske panderter.

4.5.1 WWWs mangfoldighed - udgangspunkter

Den bedste måde at få en ide om WWWs mangfoldighed er at anvende en udgangspunkts-side. Af mange udmærkede eksempler på Udgangspunkter er det amerikanske "*Point*"ⁱⁱⁱⁱ berømt for sin "Top 5% of the Web" priser, der påstås at være de bedste steder på WWW. Det er en emneorienteret oversigt indeholdende WWW-sider af høj kvalitets inddelt efter områderne: "Arts & Humanities, Business, Computers, Education, Entertainment, Government & Politics, Health & Medicine, Internet, Kids, News & Information, Road Less Traveled, Science & Technology, Shopping, Social & Community, Sports og The World."

Af danske eksempler, der i forhold til amerikanske er langt færre eksempler på, er journalisten Steven Snedkers "*Net.sekstanten*"^{iv}, som er resultatet af hans faste artikelserie (Netop nu i Netland^v) i det største computerblad i Danmark "Alt om Data". Endnu bedre er ugebladet Ingeniørens "*Hitlisten*".^{vi}

4.5.2 Specifik information - simpel søgning

Til en hurtig søgning efter specifik information bruges et søgesystem, da det i forhold til de andre søgeværktøjer har den største bredde, hvilket vil sige mange WWW-sider indekseret. Samtidig er det det eneste søgeværktøj, som indekserer alle komponenter i WWW-sider. Yderligere er det en fordel at kunne foretage søgninger af flere ord som fraser, da dette giver de mest nøjagtige søgninger. Endelig spiller hastigheden ind, da det er hurtigere at foretage en søgning end at klikke sig frem i et indeks. Derfor vil valget af søgeværktøj ved søgning efter specifik information være et valg mellem søgesystemer, baseret på antallet af WWW-sider og på, at hele teksten er indekseret, samt at det er muligt at foretage frasesøgning. Således vil det bedste søgesystem ideelt set være det, som indekserer flest WWW-sider, og medtager flest komponenter, samt muliggør frasesøgning. Endelig er det vigtigt at sikre sig, hvorvidt den geografiske dækning er tilfredsstillende i forhold til emnet, man søger belyst ved yderligere information. I de følgende figurer er søgesystemerne rangeret efter disse kriterier, hvor figur 4.7 er store globalt dækkende systemer:

Søgesystem	Antal af WWW-sider	Komponenter	Frasesøgning
HotBot ^{vii}	Ca. 54 mio.	URL, title, headings, all content, hyperlinks, Meta Tag, JavaScript, Java, Audio, Acrobat, Shockwave, VRML, Smiley, File-extensions	Ja (men ikke enkelt bogstavs ord)
Lycos ^{viii}	Ca. 51 mio.	URL, title, headings, all content, hyperlinks	Nej
Excite ^{ix}	Ca. 50 mio.	URL, title, headings, all content, hyperlinks	Ja (koncept søgning)
InfoSeek Ultra ^x	Ca. 50 mio.	URL, title, headings, all content, hyperlinks, Meta Tag	Ja
Alta Vista ^{xi}	^{xii} Ca. 30 mio.	URL, title, headings, all content, hyperlinks, Meta Tag	Ja
OpenText	Ca. 10 mio.	URL, title, headings, all content, hyperlinks	Ja

Figur 4.7: Søgesystemer rangeret efter antal indekserede WWW-sider, komponenter og frasesøgning. Kilder: Search Engine Tutorial^{xiii}, Pfaffenberger: Web Search Strategies, produktinformation fra udbyderne, samt e-post dialog.

Hvis man ved, at den specifikke information er at finde i Danmark, kan der med fordel anvendes et dansk søgesystem. På nuværende tidspunkt er der tre valgmuligheder, hvor Thor ikke er et selvstændigt system, men et filter oven på det amerikanske Alta Vista, således man kan få danske resultater af søgningen. Figur 4.8 illustrerer kvaliteten af de danske søgesystemers bredde:

Søgesystem	Antal af WWW-sider	Komponenter	Frasesøgning
Jori ^{xiv}	Ca. 535.000	URL, title	Ja
CyberCity Agent ^{xv}	Ca. 200.000	URL, title, headings, all content, hyperlinks, Meta Tag	Nej
Thor ^{xvi} (baseret på Alta Vista)	Ca. 100.000	URL, title, headings, all content, hyperlinks, Meta Tag	Nej

Figur 4.8: Danske søgesystemer rangeret efter antal indekserede WWW-sider, komponenter og frasesøgning. Kilde: produktinformation fra udbyderne den 26-11-96.

Som det fremgår af figuren, er Jori det bredeste søgesystem målt på antal indekserede WWW-sider. Dog er CyberCity Agent i stand til at tage alle komponenter med i søgningen, men kan ikke foretage frasesøgning. Derfor anbefales det at starte med Jori og derefter gå nedad i skemaet og afprøve de andre søgesystemer.

Når en simpel søgning foretages i ovenstående amerikanske og danske søgesystemer bør man følge følgende råd:

1. Hvis det er muligt, brug frase søgning
2. Brug din erfaring og intuition til at finde tre ord der flertydigt beskriver den ønskede information
3. Brug din erfaring og intuition til med synonymer at beskrive den ønskede information
4. Husk rigtig stavning
5. Brug ikke WWW-termer som http, WWW og andre almindelige ord som den (the), og (and), da alle søgesystemer på nær InfoSeek Ultra har en liste, der udelader disse ord

6. Brug ikke søgeord i flertal⁸²

⁸² Pfaffenberger: Web Search Strategies, p. 84-85

Det vigtigste er dog de bagvedliggende forudsætninger for en simpel søgning, da disse er forskellige fra søgesystem til søgesystem. Søgesystemerne er rangeret efter, hvor omfattende de er:

Søgesystem	Resultat præsentation	Afskæring af søgeord	Lande, Sprog og grammatik	Store bogstaver	Behandling af flere søgeord	Frase kode
HotBot	Title, Relevans, første 1-5 linier på siden, URL, størrelse, dato	Præcis	Alle lande Amerikansk Alle internationale tegn brugerens browser understøttes	Nej - kun i specielle tilfælde, fx. NeXT	All the words, indsætter AND, any of the words indsætter OR, Frasesøgning, Personsøgning URL	Vælg "the phrase", "-tegn afgrænser
Lycos	Title, URL, Relevans, antal søgeord, sammendrag	Automatisk	Alle lande Amerikansk	Nej	Indsætter OR automatisk	Ikke mulig
Excite	Relevans, Title, URL, automatisk generet beskrivelse, More like this	Ordstamme	Alle lande Amerikansk	Nej	Koncept søgning ved alm. søgning indsættes OR automatisk	Koncept søgning
InfoSeek Ultra	Relevans, Title, første 1-5 linier på siden, URL, størrelse,	Ental	Alle lande Amerikansk	Ja	Indsætter OR automatisk	"-tegn afgrænser
Alta Vista	Title, URL, første 1-3 linier på siden, eller META content=information	Præcis	Alle lande Amerikansk	Ja	Indsætter OR automatisk	"-tegn afgrænser
OpenText	Title, URL, første 5 linier på siden, Relevans	Automatisk	Alle lande Amerikansk	Kun ved egen- navne	Frase søgning	Vælg "this phrase"
Jori	Title, Relevans, første 1-5 linier på siden, URL,	Automatisk	Danmark Dansk	Nej	Indsætter OR automatisk	"-tegn afgrænser
CyberCity Agent	Title, Relevans, første 1-3 linier på siden, URL	Automatisk, dog uden specielle tegn	Danmark Dansk	Nej	Indsætter OR automatisk	Ikke mulig
Thor (baseret på Alta Vista)	Title, URL, første 1-3 linier på siden, eller META content=information	Præcis	Danmark, Finland, Norge, Sverige og alle lande	Ja	Indsætter OR automatisk	Ikke mulig

Figur 4.9: Standardindstillinger ved simpel søgning

Ovenstående kan virke som overvældende at skulle sætte sig ind i - blot for at foretage en simpel søgning. På den anden side kan ovenstående også ses som en forklaring på, hvorfor det umiddelbart er svært at få et godt resultat af ens søgning.

4.5.3 Information af høj kvalitet - vejledere via emneindeks

Når målet med informationssøgningen er at finde 10-20 WWW-sider af høj kvalitet, er bredden af søgningen ikke så vigtig. Samtidig vil søgesystemerne give resultater uden hensyn til kvaliteten af WWW-siden. Her er emneindeks en oplagt mulighed, idet de ikke er så omfattende, og samtidig er indekseret af mennesker af kød og blod. Disse personer foretager en kvalitetsvurdering, hvilket kan være en fordel. Det har dog stadig den ulempe, at der dermed kommer en person ind i mellem personen, der foretager søgningen, og den mulige information, så søgningen allerede er begrænset kraftigt inden man starter søgningen. Den anden ulempe er, at de mennesker som indekserer WWW-siderne ikke nødvendigvis har den fornødne faglige viden om det emne man ønsker at finde information om. Derfor må emneindeksene kun ses som et middel til at finde vejledere varetaget af personer med den nødvendige faglige viden.

De emneindeks vi her vil koncentrere os om, er det altdominerende amerikanske indeks kaldet Yahoo!, samt to danske Jubii og Indeks.dk. Valget er faldet på disse på grund af deres størrelse, hvilket er vigtigt, da emneindeks i sig selv er langt mindre dækkende end søgesystemerne. Med emneindeks er det noget sværere at foretage en vurdering af dybden af ens søgning, idet man blot klikker sig rundt under emner, der har relevans i forhold til ens søgning. Man kan dog vurdere ud fra resultatpræsentationens omfang, samt i hvilket omfang der er foretaget en kvalitetsvurdering i form af kvalitetsbeskrivelser og/eller -stjerner. Endelig har de største af emneindeksene mere eller mindre avancerende søgemuligheder, hvilket også hjælper til en bedre kontrol af søgningens dybde. Dette har også været med til at anbefale de nævnte emneindeks, som er opstillet i figur 4.10:

Emneindeks	Antal WWW-sider	Resultat præsentation	Søgemulighed
Yahoo! ^{xvii}	Ca. 200.000	Titel, Få liners beskrivelse, Beskrivende symboler for "nyhed" og "blandt de bedste"	Som søgesystemernes simpel søgning
Jubii ^{xviii}	Ca. 5.700	Titel, 1 linies beskrivelse sidens ansvarlig selv har indtastet	Meget simpel
Indeks.dk ^{xix}	Ca. 6.500	Titel, URL og evt. få liners beskrivelse	Meget simpel (indsætter AND mellem søgeord)

Figur 4.10: Oversigt over Emneindeksene. Kilde: For Yahoo! er det Search Engine Tutorial^{xx}, for Indeks.dk er det produktinformation fra udbyderen den 26-11-96 og for Jubii er det e-post den 27-11-96.

Vi foreslår Jubii som et bedre valg end Indeks.dk, da Indeks.dk ofte registrerer de enkelte WWW-sider to gange, og dermed er Indeks.dk ikke nær så dækkende, som antallet af WWW-sider umiddelbart giver indtryk af. Indeks.dk muliggør, at den ansvarlige for WWW-siden kan indtaste en beskrivelse af WWW-siden, men Indeks.dk sletter ikke den gamle. Som det fremgår af figuren er deres præsentation af WWW-siderne særdeles begrænset, og derfor må man håbe på, at disse emneindeks også har listet nogle vejledere man kan klikke videre til. Yahoo kalder vejledere for "Indices". Et andet godt sted at starte i emneindeksene for at finde vejledere, er at undersøge de

listede forskningsbiblioteker. Ofte tilbyder eller henviser disse biblioteker til særdeles gode vejledere inden for mange emneområder, og det er ikke, som man ellers kunne tro, kun forskningsrelaterede emner.

4.5.4 Samtlige WWW-tilgængelige informationer - avanceret søgning

Når formålet med søgningen er at finde al information om et givet emne, der er tilgængelig på WWW, må man igen bruge et søgesystem. Igen vil en stor database og flest mulige komponenter være at foretrække, således søgningen kan blive så bred som mulig. Derfor vil de søgesystemer, som blev anbefalet i underafsnittet om "Specifik information" stadig være de bedste. Igen gælder det, at man må være opmærksom på begrænsningerne ved de forskellige søgesystemer. Muligheden for at kontrollere søgningen er dog her endnu mere vigtig, når man ønsker at finde al WWW-tilgængelig information. Derfor er det vigtigt at forstå, hvilke yderligere muligheder man har for at kunne kontrollere søgningen. Disse muligheder kan opnås ved at man vælger "options", "Advanced search" eller lignende søgning. Figur 4.11 gennemgår de forskellige søgesystemers kapacitet i denne retning. Når "Ingen" er angivet, betyder det, at søgesystemets standardindstilling ikke kan ændres.

Søgesystem	Resultat præsentation	Afskæring af søgeord	Begrænset feltsøgning	Behandling af flere søgeord	Bemærkninger
HotBot	Ingen	Ingen	Tekst, Frase, person, URL, Date, Location, JavaScript, Java, Audio, Acrobat, Shockwave, VRML, Smiley, File-extensions	Must eller should eller must not contain words, the phrase, the person, the URL	Specificer regioner og lande, Datosøgning, Mediesøgning, Geografisk placering, kan bruge Meta Tag som beskrivelse
Excite	Kan sortere resultatet ud fra relevans eller URL	Ingen	Ikke mulig	AND, OR, AND NOT og ()	Find lignende sider, forespørgsel-med-eksempel
InfoSeek Ultra (Beta)	Ingen	Ental	Hyperlink, Site, URL, Title, Image og Java	Ingen	Søge forbedring, Genkender egennavne, Søger på alle ord, kan bruge Meta Tag som beskrivelse
Alta Vista	Relevans lister kan bestemmes med at indtaste søgeordet	Præcis ændres til Ordstamme ved brug af * efter søgeordet	Summary, Title, First Heading, Hyperlinks, Java Applets, Host, Images	AND, OR, NOT, NEAR (højst 10 ord imellem søgeord), brug af parenteser	Kan bruge Meta Tag som beskrivelse
Lycos	Formatet af resultatet kan være: Standard, Summary eller Details	Automatisk kan ændres til Præcis med punktum efter søgeordet eller med \$ som det første i søgeordet ændres til Ordstamme	Ikke mulig	AND, OR og "-" tegnet som NOT	Ingen
OpenText	Relevans kan sorteres efter antal gange søgeordet optræder eller om søgeordet forekommer eller ikke forekommer	Automatisk ændres til Præcis med mellemrum efter søgeordet	Summary, Title, First Heading, Hyperlinks	AND, OR, BUT_NOT NEAR, FOLLOWED BY	Søge forbedring, Find lignende sider, Vægtet søgning, Frase søgning
Jori	Ingen	Ingen	Ingen	AND (eller og) og "-" tegnet som NOT, "+" tegnet prioriterer søgeordet højere	Ingen
CyberCity Agent	Opdateret sidst, links til og fra siden, sidens hierakiske placering	Ingen	Ingen	OR (skal-ord), AND (kan-ord), NOT (nej-ord) og NEAR	Se resultat præsentation
Thor (baseret på Alta Vista)	Ingen	Ingen	Ingen	Ingen	Kan bruge Meta Tag som beskrivelse

Figur 4.11: Yderligere kontrolmuligheder ved avanceret søgning

Figur 4.11 viser, at der er mange måder at specificere sin søgning på. Den viser også, at der er stor forskel på, hvorledes dette foretages i de enkelte søgesystemer. Derfor må det anbefales altid at afprøve forskellige søgesystemer, inden man stiller sig tilfreds med søgningen. Hvis man alligevel ikke har fundet den ønskede information, kan det være nødvendigt at anvende specialiserede referencesystemer. Mulighederne for dette vil det følgende afsnit kort skitsere.

4.6 Undersøg specialiserede referencesystemer

Referencesystemer kan være emneindeks og/eller søgesystemer, der er dedikeret et smalt område. For at forstå referencesystemer må vi tilbage til indledningen af dette kapitel. Som skitseringen af vidensudviklingsprocessen viste, er der to "hovedveje" for udvikling af viden - informationsøgning og erfaring. Når man ved hjælp af de gennemgåede søgeværktøjer finder information, der på basis af fortolkninger ikke har givet den ønskede viden, og ens erfaring ikke har kunnet bidrage yderligere, må man ty til de specialiserede referencesystemer. Her kommer ens erfaring og nuværende viden ind. Denne må nødvendigvis bruges til at finde og søge i relevante specialiserede søgesystemer.

For at få en ide om, hvad disse områder kan være, vil det være udemærket at kaste et blik på Pfaffenbergers opdeling i hans bog "Web Search Strategies":

- | | | |
|-------------------------------|----------------------------|--------------------|
| • Reference Trailblazer Pages | • Acronyms | • Authors |
| • Book Publishers | • Calendars | • Census Data |
| • Colleges and Universities | • Dictionaries | • Earthquakes |
| • Encyclopedias | • Geography and Travel | • Literary Works |
| • Maps | • Measurements | • Movies |
| • Patents | • Phone Numbers | • Quotations |
| • Religious Texts | • Scientific Constants | • Tax Forms |
| • Thesauri | • Time Zone Converters | • Translation Aids |
| • Usenet FAQs | • Zip Codes ^{xxi} | |

Ovenstående liste viser, at der er tale om meget forskellige emneområder med forskellige abstraktionsniveauet. Udover disse kilder nævner Pfaffenberger også nogle andre emner der kan være af interesse:

- | | | |
|----------------|--------------|----------------------|
| • News/Weather | • Government | • Stocks/Investments |
| • Periodicals | • Software | • People |
| • Jobs | | |

Disse kilder beskrives og vurderes af Pfaffenberger i hans bog, hvor bogens hjemmeside^{xxii} kan anvendes til at klikke videre til de beskrevne systemer. Problemerne ved sådanne beskrivelser er, at systemerne meget vel kan have ændret karakter, samt at eksemplerne er meget amerikanske og dermed er ubrugelige i en sammenhæng, hvor det ikke er amerikanske forhold eller fænomener ens søgning er rettet i mod. En andet sted at gå hen, hvor de videre links til specifikke referencer kan findes i en opdateret version, samt er mere globalt dækkende, er "The Internet Sleuth", hvor utroligt mange søgemuligheder er samlet på et sted. I "The Internet Sleuth" er søgemulighederne opdelt under følgende kategorier:

- Agriculture
- Bio Sciences
- Chemistry
- Computer Related
- Employment
- Environment
- Genealogy
- Humanities
- Mathematics
- News
- Reference
- Shopping
- Sports
- Travel
- Web Search Engines^{xxiii}
- Arts
- Business
- Commercial Databases
- Economics
- Engineering
- Finance
- Government
- Internet
- Medicine
- Physics
- Regional
- Social Sciences
- Technology
- Usenet News
- Biology
- Business Directories
- Companies
- Education
- Entertainment
- Food & Drink
- Health
- Legal
- Music
- Recreation
- Sciences
- Software
- Trade & Industry
- Veterinary

Problemet er dog stadig, at det for søgninger rettet mod danske (eller andre ikke amerikanske) forhold kan være meget svært at finde hjælp. Man er her nødt til at bruge danske (eller andre nationale) emneindeks og søgesystemer i sine søgninger.

4.7 Overvåg nye informationer på WWW

WWW er en meget dynamisk størrelse, der udvikler sig konstant. Derfor er det nødvendigt, at man hele tiden holder sig opdateret med, hvilke nye informationer indenfor et givent emneområdet, der er kommet til siden sidste søgning.

Der er forskellige værktøjer, der kan hjælpe i denne proces. Disse værktøjer er som de gennemgåede søgeværktøjer forskellige, men fælles gælder det, at de kan hjælpe én, således at man ikke skal klikke sig igennem emneindekset igen eller foretage søgningen på ny.

4.7.1 Udgangspunkt

Som der findes udgangspunkts-sider, der giver én oversigt over WWWs mangfoldighed findes også Udgangspunkts-sider eller vejledere om man vil, dedikeret til at give et overblik over nye tiltag på WWW. Her vil det være passende at starte med "What's New on Yahoo"^{xxiv}. The National Center for Supercomputing Applications (NCSA)s "What's New"^{xxv} kan også anbefales. Af danske kan Jubii's emneindeks nævnes med deres "Nyheder på Jubii"^{xxvi}.

4.7.2 Emneindeks

Da folkene, der varetager indeksering af emneindeksene, ofte er underbemandede, kan det tage tid inden tilmelding af WWW-siden bliver registret (dette kalder vi for den gennemsnitlige tilmeldingstid). Ajourføring af indekset foretages ikke nødvendigvis hver dag. Derfor må man være opmærksom på disse parametre, når man vurderer, hvorvidt det er muligt at finde opdateret information i det pågældende emneindeks. Nogle emneindeks giver én mulighed for at orientere sig under dets nyhedsside, således det ved eventuel ny information er nemt at se om det emne, der har ens interesse, er kommet til. Ligeledes kan man tilmelde sig på en e-post-liste, hvor man løbende

for tilsendt nyhedsbreve om nyheder på det pågældende emneindeks. I figur 4.12 er nyhedsparametrene listet for de før gennemgåede emneindeks:

Emneindeks	Gennemsnitlig tilmeldingstid	Indekset ajourføres	Nyhedsside
Indeks.dk	Ca. 2 sekunder	Hver nat for nye sider, løbende for at fjerne gamle sider	Nej
Jubii	Maksimum 4 dage	Mandag og torsdag	Ja
Yahoo!	Vides ikke	Vides ikke	Ja (de sidste 3 dage)

Figur 4.12: Emneindeksenes nyhedsparametre. Kilder: Northern Webs: "Search Engines Tutorial", <http://www.digital-cafe.com/~webmaster/set01.htm> og dialog med udbyderne. Se bilaget "Dialog via e-post".

4.7.3 Søgnesystemer

Søgnesystemerne har lignende nyhedsmuligheder. Dog findes der ikke en decideret nyhedsside. I stedet kan man vælge at gemme en søgning man ønsker at holde konstant opdateret. I figur 4.13 er nyhedsparametrene listet for de før gennemgåede søgnesystemer:

Søgnesystem	Gennemsnitlig tilmeldingstid	Databasen ajourføres	Kan søgning gemmes?	Nyheds - brev/-side
InfoSeek Ultra (Beta)	Med det samme	Konstant	Ja	Intet
CyberCity Agent	Gennemsnitlig 12 timer	Konstant	Ja	Intet
Alta Vista	24 timer	Konstant	Ja	Side
Thor (baseret på Alta Vista)	24 timer	Konstant	Ja	Intet
OpenText	24 timer	Hver måned eller længere	Nej	Side
Jori	1 uge	Hver uge	Ja	Intet
HotBot	1 uge	Uklart	Ja	Intet
Lycos	3 uger	Hver 2. uge	Ja	Brev/side
Excite	6 uger	Hver uge	Ja	Intet

Figur 4.13: Søgnesystemernes nyhedsparametre. Kilder: Northern Webs: "Search Engines Tutorial", <http://www.digital-cafe.com/~webmaster/set01.htm> og Dialog med udbyderne. Se bilaget "Dialog via e-post".

Det skal dog bemærkes, at søgnesystemernes nyhedsparametre er et meget fejlbehæftet område. Eller sagt på en anden måde, her er det vigtigt at anse de enkelte søgnesystemer som kommercielle produkter, da der er stor forskel på, hvad de siger, og hvad de i realiteten gør.

4.7.4 Bogmærker

Bogmærkefunktionen i ens browser kan anvendes til at gemme gamle søgninger. Når man foretager søgningen igen, vil man automatisk få den nye information for det pågældende WWW-sted med i ens søgning. Samtidig giver bogmærkefunktionen mulighed for at overvåge om bogmærkerne (fx. en konkurrents, et emneindeks underemne, en tidligere søgning eller lignende) har ændret sig, uden at man selv kigger efter. Denne funktion kaldes i Netscape 2.0s bogmærkefunktion for "What's New?". Her kan man vælge overvågning af alle eller udvalgte bogmærker.

4.7.5 Andre værktøjer

Den kommercielle service "What's New Too"^{xxvii} er et simpelt søgesystem, som udelukkede beskæftiger sig med nye informationer på WWW. Dette kan være en sidste mulighed for at finde nye relevante informationer på WWW.

En meget kraftfuldt service, der løser problemet med at interessante WWW-steder opdateres eller ændres, uden at man ved det, er overvågningsservicen The URL-minder.^{xxviii} Her tilmeldes de WWW-sider, der ønskes overvåget. Herefter vil man ved hver ændring få en e-post baseret besked. Herved undgår man selv at skulle checke sine bogmærker.

4.8 Transformér viden til information

Denne gennemgang af forskellige søgeværktøjer skulle gerne hjælpe med til at oparbejde en viden om et givent emne. Denne viden kan ikke oplagres, men informationerne man har fundet værdifulde for ens vidensudviklingsproces, kan også være gavnlige for andre, som ligeledes ønsker at tilegne sig viden om samme emne. Andre kan være ens kollegaer, studiekammerater, venner eller mere idealistisk alle WWW-brugere.

Ofte har man gemt disse informationer som bogmærker, og således kan disse være til hjælp for andre. Disse bogmærker kan gives videre ved at sende dem til de interesserede personer eller ved at gøre dem tilgængelige via WWW. En mere krævende opgave er at skabe sin egen vejleder-side om et givet emne, og gøre den tilgængelig via WWW. Denne opgave er mere krævende, da det kræver en detaljeret beskrivelse af, hvilken information man vil finde på den pågældende URL. En evaluering af kvaliteten af denne information, samt en kort beskrivelse af, hvorledes informationen på dette WWW-sted er organiseret, må også beskrives på vejleder-siden.

For at få en bedre forståelse for informationssøgning på WWW, har vi valgt at udarbejde et WWW-sted om dette ^{xxix}. WWW-stedet er baseret på denne delundersøgelse. Da delundersøgelsen er meget omfattende, har vi ved udarbejdelsen af WWW-stedet prøvet at udnytte hypertexts fordele, og dermed gøre undersøgelsen mere "brugervenlig". Dette WWW-sted vil vi desuden bruge til at foretage egne søgninger på WWW.

I forbindelse med besvarelsen af projektets problemstilling er det også vigtigt at forstå, hvordan virksomheden bliver i stand til at transformere viden til information, da det netop er betingelsen for at kunne skabe kontakt til køberen, hvilket igen er betingelsen for at virksomheden effektivt kan markedsføre sine produkter og serviceydelser på WWW.

Her bør man også overveje, hvem der skal/kan få tilgang til den fundne information og hvilke informationer de pågældende personer skal have adgang til. Her må man f.eks. sørge for, at konkurrenter ikke får adgang til information, der kan styrke deres position.

4.1 Referencer der forfindes på World Wide Web

ⁱ **Web Search Strategies:** <http://watt.seas.virginia.edu/~bp/searchme/welcome.html>

- ii **Internetworld:** <http://www.idg.dk/internetworld/artikler/searchtest/logik.htm>
- iii **Point:** <http://www.pointcom.com>
- iv **Net.sekstanten:** <http://194.19.145.24:80/aod/net-kurs/sekstant/sekstant.htm>
- v **Netop nu i Netland:** <http://inet.uni-c.dk/~snekker/artikler.htm>
- vi **Hitlisten:** <http://www.ingenioeren.dk/redaktion/usite.html>
- vii **HotBot:** <http://www.hotbot.com>
- viii **Lycos:** <http://www.lycos.com>
- ix **Excite:** <http://www.excite.com>
- x **Infoseek Ultra:** <http://ultra.infoseek.com>
- xi **Alta Vista:** <http://altavista.digital.com>
- xii **OpenText:** <http://index.opentext.net>
- xiii **Search Engine Tutorial:** <http://www.digital-cafe.com/~webmaster/set01.html>
- xiv **Jori:** <http://jori.jubii.dk>
- xv **CyberCity Agent:** <http://agent.cybercity.dk>
- xvi **Thor:** <http://www.danielsen.com/thor/>
- xvii **Yahoo!:** <http://www.yahoo.com>
- xviii **Jubii:** <http://www.jubii.dk>
- xix **Indeks.dk:** <http://www.forum.dk/index/>
- xx **Search Engine Tutorial:** <http://www.digital-cafe.com/~webmaster/set01.html>
- xxi **Specialiserede referencesystemer i følge Web Search Strategies:**
<http://watt.seas.virginia.edu/~bp/searchme/refdesk.htm>
- xxii **Flere specialiserede referencesystemer i følge Web Search Strategies:**
<http://watt.seas.virginia.edu/~bp/searchme/welcome.html>
- xxiii **The Internet Sleuth:** <http://www.isleuth.com/>
- xxiv **What's New on Yahoo:** <http://www.yahoo.com/new/>
- xxv **The National Center for Supercomputing Applications (NCSA)s "What's New":**
<http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html>
- xxvi **Nyheder på Jubii:** <http://www.jubii.dk/nyheder/default.htm>
- xxvii **What's New Too:** <http://newtoo.manifest.com/today.html>
- xxviii **The URL-minder:** <http://www.netmind.com/url-minder>
- xxix **Informationssøgning på World Wide Web:**
<http://www.i4.auc.dk/rask/search/cover.htm>