

# INTERNATIONAL BUSINESS ECONOMICS

## Study Materials Series



### Informationssøgning på World Wide Web

by

Morten Rask og Kenneth Skærm

No.: 10

1997



**Aalborg University**

Centre for International Studies  
Fibigerstraede 2, 9220 Aalborg, Denmark  
Tel. 45 96 27 00 00 Fax 45 96 15 00 50

# **Informationssøgning på World Wide Web**

**Morten Rask og Kenneth Skræm**

**International Virksomhedsøkonomi  
Aalborg Universitet  
Februar 1997**

© Morten Rask og Kenneth Skræm.

Dette paper forefindes på World Wide Web via adressen:

<http://www.morten-rask.dk>

---

## Indhold

---

<b>INDHOLD .....</b>	<b>2</b>
<b>INTRODUKTION .....</b>	<b>4</b>
BAGGRUND .....	4
OPDELING OG LÆSEVEJLEDNING .....	5
<b>FORSTÅ UDFORDRINGEN I INFORMATIONSSØGNING PÅ WWW.....</b>	<b>8</b>
SØGEVÆRKTØJERNES BEGRÆNSNINGER.....	8
HYPERTEKST OG HYPERLINKS .....	9
DOKUMENTKODNINGSSTANDARDEN HTML.....	10
FORUDSÆTNINGER FOR AT KUNNE TAGES UDFORDRINGEN I WWW-SØGNING OP .....	13
<b>FÅ ET OVERBLIK OVER SØGEVÆRKTØJERNES KVALITETER.....</b>	<b>14</b>
UDGANGSPUNKTER.....	14
EMNEINDEKSENE .....	15
VEJLEDERE .....	16
SØGESYSTEMER.....	16
BOGMÆRKER.....	18
FORDELE OG ULEMPER VED DE FORSKELLIGE TYPER AF SØGEVÆRKTØJER .....	18
VURDERINGSPARAMETRENE NÅR EN UDBYDER AF ET SØGEVÆRKTØJ SKAL VÆLGES FREMFOR EN ANDEN.....	19
<i>Bredden af søgningen .....</i>	<i>19</i>
<i>Dybden af søgningen - generelt .....</i>	<i>20</i>
<i>Dybden af søgningen - søgesystemer og emneindeks med søgemuligheder.....</i>	<i>20</i>
<i>Opsamling - Parametre der angiver bredden og dybden i søgningen .....</i>	<i>22</i>
<b>SITUATIONSBESTEMT ANVENDELSE AF SØGEVÆRKTØJERNE .....</b>	<b>23</b>
WWW'S MANGFOLDIGHED - UDGANGSPUNKTER.....	23
SPECIFIK INFORMATION - SIMPEL SØGNING .....	24
<i>Danske søgesystemer .....</i>	<i>25</i>
<i>Globale søgesystemer.....</i>	<i>26</i>
<i>Opsamling - simpel søgning.....</i>	<i>27</i>
INFORMATION AF HØJ KVALITET - VEJLEDERE VIA EMNEINDEKS .....	28
SAMTLIGE WWW-TILGÆNGELIGE INFORMATIONER - AVANCERET SØGNING .....	30
<i>Danske søgesystemers yderligere kontrolmuligheder ved avanceret søgning .....</i>	<i>30</i>
<i>Globale søgesystemers yderligere kontrolmuligheder ved avanceret søgning .....</i>	<i>31</i>
<i>Opsamling - avanceret søgning .....</i>	<i>32</i>
<b>UNDERSØG SPECIALISEREDE REFERENCESYSTEMER .....</b>	<b>32</b>
<b>OVERVÅG NYE INFORMATIONER PÅ WWW.....</b>	<b>33</b>
NYE INFORMATIONER PÅ WWW .....	34

SØGEVÆRKTØJERNES AJOURFØRING.....	34
<i>Emneindeks</i> .....	35
<i>Søgesystemer</i> .....	35
NY INFORMATION PÅ KENDTE WWW-SIDER .....	36
<i>Bogmærker</i> .....	36
<i>The URL-minder</i> .....	37
OPSAMLING - OVERVÅGNING AF NYE INFORMATIONER PÅ WWW .....	37
<b>TRANSFORMÉR VIDEN TIL INFORMATION.....</b>	<b>38</b>
<b>ORDLISTE.....</b>	<b>39</b>
<b>REFERENCER.....</b>	<b>42</b>
<b>STIKORDSREGISTER .....</b>	<b>43</b>
<b>FIGURER:</b>	
FIGUR 1: VIDENSUDVIKLINGSPROCESSEN .....	5
FIGUR 2: NETSCAPE NAVIGATORS "GO TO"-BOKS .....	8
FIGUR 3: LINEÆR TEKST VERSUS HYPERTEKST .....	10
FIGUR 4: KOMPONENTER I EN TYPISK WWW-SIDE.....	12
FIGUR 5: EKSEMPEL PÅ ET UDGANGSPUNKT .....	14
FIGUR 6: EKSEMPEL PÅ ET EMNEINDEKS.....	15
FIGUR 7: EKSEMPEL PÅ EN VEJLEDER .....	16
FIGUR 8: EKSEMPEL PÅ ET SØGESYSTEM .....	16
FIGUR 9: GENEREL ANVENDELSESPROCEDURE FOR ET SØGESYSTEM.....	17
FIGUR 10: EKSEMPEL PÅ BRUG AF BOGMÆRKER.....	18
FIGUR 11: WHAT'S NEW? I NETSCAPE BROWSERENS BOGMÆRKEFUNKTION .....	36
FIGUR 12: NETSCAPE BROWSEREN KAN UNDERSØGE OM BOGMÆRKEDE WWW-SIDER HAR ÆNDRET SIG .....	36
<b>TABELLER:</b>	
TABEL 1: FORDELE OG ULEMPE VED DE FORSKELLIGE TYPER AF SØGEVÆRKTØJER.....	19
TABEL 2: LOGISKE OPERATORER .....	22
TABEL 3: PARAMETRE DER ANGIVER BREDDEN OG DYBDEN AF SØGNINGEN.....	22
TABEL 4: DANSKE SØGESYSTEMER RANGERET EFTER ANTAL INDEKSEREDE WWW-SIDER, KOMPONENTER OG FRASESØGNING.....	25
TABEL 5: STANDARDINDSTILLINGER VED SIMPEL SØGNING I DANSKE SØGESYSTEMER .....	25
TABEL 6: GLOBALE SØGESYSTEMER RANGERET EFTER ANTAL INDEKSEREDE WWW-SIDER, KOMPONENTER OG FRASESØGNING.....	26
TABEL 7: STANDARDINDSTILLINGER VED SIMPEL SØGNING I GLOBALE SØGESYSTEMER .....	27
TABEL 8: OVERSIGT OVER EMNEINDEKSENE .....	29
TABEL 9: YDERLIGERE KONTROLMULIGHEDER VED AVANCERET SØGNING I DANSKE SØGESYSTEMER .....	30
TABEL 10: YDERLIGERE KONTROLMULIGHEDER VED AVANCERET SØGNING I GLOBALE SØGESYSTEMER .....	31
TABEL 11: EMNEINDEKSENE NYHEDSPARAMETRE.....	35
TABEL 12: SØGESYSTEMERNES NYHEDSPARAMETRE .....	35

---

## Introduktion

---

Dette paper henvender sig til personer som i arbejdet med en given problemstilling ønsker at finde brugbare oplysninger på den grafiske del af Internettet kaldet World Wide Web (WWW). Efterhånden som WWW er vokset i kvantitet og kvalitet, kan WWW betragtes som en meget stor informationsressource. Derfor vil virksomhedsfolk, ansatte i offentlige organisationer, studerende, forskere og andre personer have stor gavn af at søge efter information på WWW. Informationer som af mange grunde ellers ville være svære at få adgang til.

Formålet med dette paper er at gøre disse mennesker i stand til at forstå og beherske forskellige metoder til at søge efter information på WWW. Disse metoder er samlet under begrebet WWW-søgeværktøjer.

### Baggrund

Dette paper tager sit udgangspunkt i vores speciale med titlen "International Markedsføring på World Wide Web", som vi udarbejdede i perioden maj 1996 til januar 1997.

I dette speciale gennemførte vi bl.a. en undersøgelse af WWW-søgeværktøjer. Målet med denne undersøgelse var at sætte sig i informationssøgerens sted. Dette var ud fra den tanke, at for at 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 måtte 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.

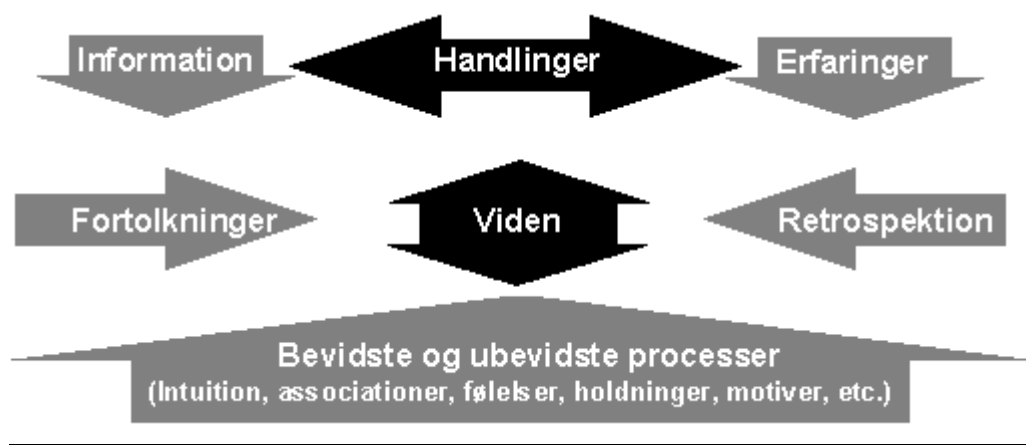
Vores motivation for at beskæftiges os med informationssøgning på WWW i specialet var, at vi havde meget svært ved at finde det, vi eftersøgte. Vi begyndte hyppigt at bruge diverse søgeværktøjers hjælpefunktion. Nogen søgeværktøjer var lettere at anvende end andre og nogen mere præcise end andre. Der begyndte at danne sig et mønster. Ved at læse artikler om søgeværktøjernes produktinformation blev vi klar over, at disse søgeværktøjer godt nok var gratis at benytte, men alligevel måtte 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 andre krav til informationssøgeren, end der var gældende ved andre former for informationsindsamling.

Den viden denne undersøgelse har skabt, er altså 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 - samt dialog med og information fra andre, der også har analyseret WWW-søgeværktøjer.

Dette skal forstås i sammenhæng med den måde vi opfatter viden på. Viden kan ikke lagres (information kan lagres) og er ikke en lineær proces, da mennesker ikke tænker lineært. Viden er en proces, som aldrig stopper. Forståelsen er knyttet til fortolkning af information. Igennem denne tolkning er det muligt at opbygge en forståelse af det beskrevne - at skabe ny viden. Såvel som information er et udtryk for handlinger, der er foretaget, gælder det også, at erfaringen kommer fra handling. Disse erfaringer kan kun forstås retrospektivt og igennem denne refleksionsproces skabes ligeledes ny viden. Således vil den opnåede viden ikke være statisk, det er en proces, eller som Fast udtrykker det:

Knowledge can thereby be understood as moving pictures of reality: experiences and information are produced through actions and are transformed by the processes of interpretation and retrospection to the knowledge that the actor experiences as useful and relevant.<sup>1</sup>

Når der udvikles viden, er der altså en stærk sammenhæng mellem handling og viden. Denne sammenhæng kan illustreres således:



Figur 1: Vidensudviklingsprocessen.

Kilde: Fast: Videnskabsteori og metodologi i studier i livsverden, p.243

Denne vidensudviklingsproces vil vi referere til løbende, da den repræsenterer vores grundlæggende opfattelse af, hvorledes viden udvikles - også i forbindelse med søgning efter information på WWW.

## Opdeling og læsevejledning

Mønstret der dannede sig af, hvorledes informationssøgning på WWW foregår, indeholder nogle væsentlige forhold. Vores opfattelse af disse 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

<sup>1</sup> Fast: Internationalization as a social construction, p. 3

## 6. Transformér viden til information

Når formålet er at få et solidt kendskab til informationssøgning på WWW kan disse trin ses som en fortløbende proces, hvor man starter med trin 1 og arbejder sig videre til og med trin 6. Når man har et kendskab til informationssøgning på WWW, vil man ofte fokusere på det trin, der er relevant i forhold til ens behov for oplysninger.

Disse trin vil blive beskrevet i det følgende og danner derfor strukturen for dette paper. Dog bør det også bemærkes, at det behandlede område indeholder mange nye termer, som alle ikke nødvendigvis kender, og da disse termer opfattes forskelligt, har vi udarbejdet en ordliste, således misforståelser kan undgås. Denne ordliste er sammen med referencelisten placeret sidst i dette paper.

Vi anbefaler at alle læsere - uanset baggrund - læser dette paper første gang i den rækkefølge afsnittene bliver præsenteret, for derefter at anvende dette paper som en håndbog, hvor relevante oplysninger kan slås op. Det hænger sammen med, at dette paper skal opfattes som en rejsebog baseret på vores subjektive oplevelser af en rejse ind gennem et nyt og ukendt land, som vi nu prøver at give videre både på et højt abstraktionsniveau, men også med konkrete overvejelser om, hvilket søgeværktøj der bedst anvendes i en bestemt situation, samt hvordan dette søgeværktøj konkret anvendes.

Men det spændende og samtidig problematiske, når omdrejningspunktet er WWW, er dog den meget dynamiske udvikling på området. Således tales der om, at den udvikling området gennemgår på tre måneder, svarer til et "almindeligt" år. Derfor kan vi citere en del af indledningen fra de kendte rygsæk-rejse-bøger "Travel Survival Kit":

### **Warning and Request**

Things change - prices go up, schedules change, good places go bad and bad places go bankrupt - nothing stays the same. So if you find things better or worse, recently opened or long since closed, please write and tell us and help make the next edition better.

Citatet er taget med, da det netop fanger det besværlige i at skrive rejsebøger såvel som papers om WWW, der er "up-to-date". Virkeligheden er en anden så snart sidste ord er skrevet. Således kan informationssøgningsværktøjerne have ændret karakteristika. Konsekvensen af denne turbulens kunne derfor være, at vi hævdede os over det specifikke og skrev om emnet på et højt abstraktionsniveau. Dette er da også tilfældet i dette paper, men som rejsebøger ville det ikke være meget værd, hvis det sker på bekostning af deltagerne. Forestil dig en rejsebog med en generel introduktion til landet, uden angivelser af, hvor man kan sove, spise osv. Det ville ikke være en rejsebog, men en geografibog.

Således vil første del af dette paper repræsentere *forberedelsen til rejsen*, hvor afsnittet **Forstå udfordringen i informationssøgning på WWW** "klæder informationssøgeren på" mht. til at forstå søgeværktøjernes begrænsninger i form af informationernes tilgængelighed og kvalitet, og i form af at søgeværktøjerne bør ses som kommercielle produkter. Endelig vil dette afsnit også diskutere muligheder og begrænsninger for søgning på WWW, når der er tale om hypertekst - kodet efter

dokumentkodningsstandarden HTML. Dette afsnit opsummerer disse diskussioner i en "mental pakkelse", der gør dig i stand til at "tage af sted".

Men hvor skal vi hen. Afsnittet **Få et overblik over søgeværktøjernes kvaliteter** ser på dette. Det hænger sammen med, at der er hundredvis af forskellige værktøjer, der kan anvendes ved søgning på WWW. Her opdeler vi de forskellige værktøjer i fem forskellige typer, samt gennemgår de forskellige typers fordele og ulemper. Desuden opstiller vi nogle vurderingsparametre, der gør det muligt at vurdere bredden og dybden af et givent søgeværktøj.

Nu er vi klar, *rejsen kan begynde*, idet dette papers anden del indeholder vores undersøgelse af, hvorledes de forskellige søgeværktøjer anvendes. Hvorvidt man vil springe de foregående afsnit over og blot starte her afhænger af ens temperament - det samme gælder ved rejser - ikke alle for læst rejseguiden. Men de foregående afsnit repræsenterer baggrunden for de følgende afsnit. Derfor anbefales det at læse de foregående afsnit. Med en analogi til rejsebøger svarer det til, at det vil være fornuftigt at kende lidt til baggrunden for islam for at forstå, hvilket tøj der anbefales i det pågældende land. Samtidig betyder det også, at man kan anvende de følgende afsnit som en slags opslagsbog, når man på ny foretager søgninger efter information på WWW.

I afsnittet **Situationsbestemt anvendelse af søgeværktøjerne** gennemgår vi i hvilken situation man med fordel kan anvende en bestemt type af søgeværktøj frem for en anden, samt hvilke søgeværktøjer af hver type, der er anbefalelsesværdig.

Har du stadig ikke fundet den ønskede information, kan det være nødvendigt at anvende specialiserede referencesystemer. Det diskuteres i afsnittet **Undersøg specialiserede referencesystemer**.

WWW er meget dynamisk og ændrer sig derfor konstant. Derfor er afsnittet **Overvåg nye informationer på WWW** til dig, der er interesseret i af holde dig ajour med, hvilke nye informationer der er kommet indenfor et bestemt emneområde siden sidste søgning. Vi gennemgår derfor her de værktøjer, der kan være en hjælp i denne proces.

*Rejsen er slut*, du er hjemme igen, men mon andre er interesseret i, hvad du fandt? I afsnittet **Transformer viden til information** diskuterer vi, hvordan man transformerer sin oparbejdede viden til information, så andre med samme interesseområde har nemmere ved at finde værdifulde oplysninger.

En sidste bemærkning er, at dette paper henvender sig til personer, som har anvendt en WWW-browser<sup>2</sup> før. Det skal dog bemærkes, at det anbefales stærkt at anvende enten Netscape Navigator version 2.0 eller højere - eller Microsoft Explorer version 3.0 eller højere, da disse anses for at være standarden.

---

<sup>2</sup> Browseren er det program, som brugeren anvender til at se på WWW-sider med. Programmet læser siderne med eventuelle tilhørende billeder via nettet og fortolker koderne for derefter at opbygge skærbilledet på brugerens maskine.



Når der i dette paper bliver henvist til en bestemt ressource på WWW angives en URL, som f.eks.: <http://www.i4.auc.dk/ivo/default.htm>. Denne lange række af bogstaver og tal indsættes i WWW-browsersens "Go to"-boks:



Figur 2: Netscape Navigators "Go to"-boks.

Således vil man ved at indsætte URL'en her få tilgang til det givne ressource.

---

## Forstå udfordringen i informationsøgning på WWW

---

WWW var fra begyndelsen ikke udviklet til at tjene som et globalt informationssystem. 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<sup>3</sup>, hvor WWW blev opfundet. Hvis de kunne have forudset denne udvikling, havde WWW været designet anderledes. Formålet med WWW var at udvikle et simpelt brugervenligt informationssystem for forskere med detaljeret kendskab til det emne informationen omhandlede, og ikke et medie for oplysninger om alt mellem himmel og jord eller et marked for alverdens produkter og serviceydelser.

### Søgeværktøjernes begrænsninger

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 ikke alt er af ønskelig kvalitet. Dette er begrænsninger, som vil blive uddybet i det følgende.

Der er forskellige opfattelser af WWW - medie eller marked. De, som opfatter WWW som et medie, kritiserer WWW for manglende dækning og kvalitet. Dette bliver man nødt til at forholde sig til. Det kan være særdeles vanskeligt at klargøre dybden og bredden af en søgning, men hvor tit kan man være opmærksom på dette? Da verden ændres konstant, og dermed 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.

---

<sup>3</sup> Conseil Européen pour la Recherche Nucléaire - Det Europæiske Råd for Kerneforskning. Hovedkvarteret er placeret i Geneve, Schweiz

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.<sup>4</sup> Men de søgeværktøjer, der fokuseres på her, 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 informationsøgning. 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. Derfor bør WWW også ses som et marked, hvorfor søgeværktøjerne også bør ses i denne sammenhæng<sup>5</sup>. Da søgeværktøjerne er kommercielle produkter, giver det en usikkerhed i informationsøgningens bredde og dybde.

Dette paper kan dermed karakteriseres som en undersøgelse omhandlende en konkret dataindsamlingsteknik, men også som en markedsundersøgelse på produktniveau, idet vi sammenligner forskellige søgeværktøjer fra forskellige udbydere. Med andre ord, bør dette paper 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 informationsøgning på WWW.

## Hypertekst og hyperlinks

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

Forskningsministeriet i Danmark udarbejdede i 1995-96 nogle "Retningslinier for offentlige on-line publikationer", som beskrev den standard offentlige institutioner i Danmark skal følge ved publicering af dokumenter på en fællesstatslig informationsserver. Denne standard er dog allerede blevet erstattet af "Standard for statslig elektronisk publicering - Version 18". Desværre udelades ordlisten fra den gamle version, der havde nogle gode begrebsafklaringer. Her blev hypertekst beskrevet som:

Den tekst, der ses i WWW-browseren, er normalt en hypertekst. Teksten består af almindelig tekst og ofte også billeder, men i skærmbilledet 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.<sup>6</sup>

Disse hyperspring, som også kaldes for hyperlinks eller bare links, her beskrevet samme sted:

Et hyperspring er betegnelsen for den proces, der sker når en bruger aktiverer en henvisning i *hypertekstdokumentet* i sin WWW-browser f.eks. med et museklik. Henvisningen kan pege på et andet sted i

---

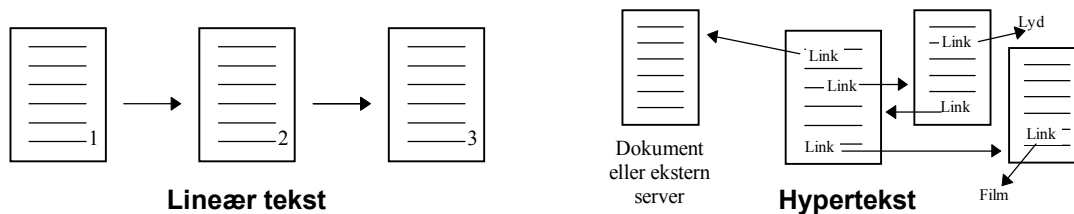
<sup>4</sup> Laursen: Reklamens magt på nettet, Politiken den 17-10-96

<sup>5</sup> Yderligere information om strategiske alliancer, diverse søgesystemer indgår i, kan findes på Calafia Consulting: A Webmaster's Guide to Search Engines, Strategic Alliances: <http://calafia.com/webmasters/compare.htm>

<sup>6</sup> Forskningsministeriet: Retningslinier for offentlige on-line publikationer, Ordliste

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 brugerens 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.<sup>7</sup>

Hypertekst er altså en ikke-lineær metode til at se og læse 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, svarer 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. Figur 3 viser forskellen på et "almindeligt" dokument og et hypertekst baseret dokument:



Figur 3: Lineær tekst versus hypertekst.

Søgemetoden i at klikke sig igennem teksten ved at klikke på de såkaldte hyperlink er, hvad Bryan Pfaffenberger i sin bog "Web Search Strategies" - som udmærker sig ved at være en af de bedste bøger omhandlende WWW-søgning - 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.<sup>8</sup>

Men husk på, at den dygtige hypertekstforfatter udnytter læserens intuition og associationer, således at læseren tror, vedkommende 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.

## Dokumentkodningsstandard HTML

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.

<sup>7</sup> Forskningsministeriet: Retningslinier for offentlige on-line publikationer, Ordliste

<sup>8</sup> Pfaffenberger: Web Search Strategies, p. 19

Der ligger nogle klare lovmæssigheder til grundlag for opbygningen af en side på WWW. Disse lovmæssigheder er nedfældet i dokumentkodningsstandarden HTML. Ordlisten i Forskningsministeriets "Retningslinier for offentlige on-line publikationer" forklarede HTML således:

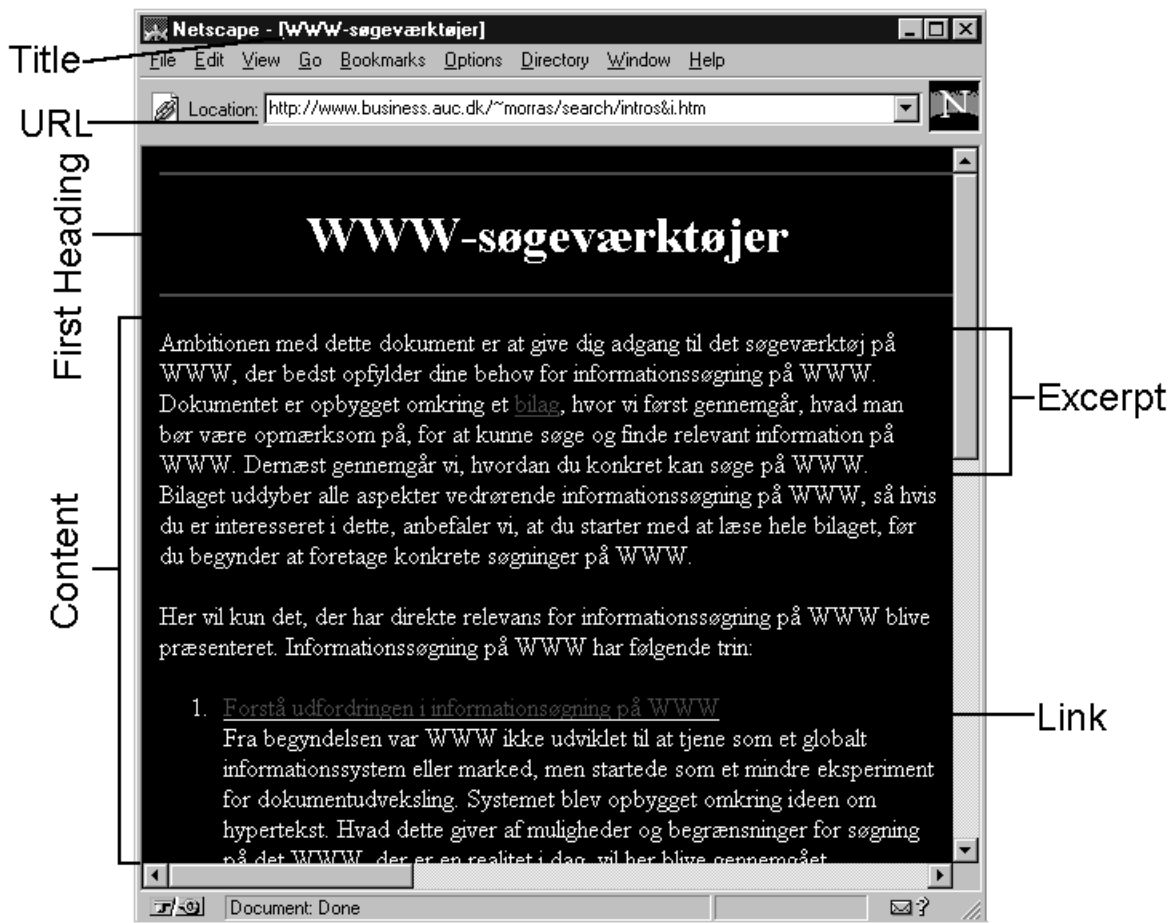
HyperText Markup Language (HTML) er det kodeformat, man bruger til at skrive WWW-dokumenter i. Koden består af en række mærker (tags) angivet i mindre-end/større-end tegn, der lægges i en almindelig tekst f.eks. <STRONG> eller <A HREF="2">. Mærkerne fortæller WWW-browseren, hvad den pågældende tekstfunktion er. Browseren afgør derefter, hvordan det skal fortolkes og vises på skærmen, f.eks. 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.<sup>9</sup>

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 komponenter bliver indekseret. Figur 4 viser en typisk WWW-side, hvor de engelske udtryk er bibeholdt, da det er disse, søgeværktøjerne primært anvender:

---

<sup>9</sup> Forskningsministeriet: Retningslinier for offentlige on-line publikationer, Ordliste



Figur 4: Komponenter i en typisk WWW-side.

Inspireret af Pfaffenberger: *Web Search Strategies*, p. 145

Figur 4 illustrerer komponenterne:

- Uniform Resource Locator (*URL*) er betegnelsen for den unikke adresse, en given WWW-side 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, f.eks. <http://www.i4.auc.dk/ivo/default.htm>
- *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 ofte opfatter første overskrift som sidens titel. I Figur 4 er Title og First Heading den samme, hvilket godt kan være forvirrende i forhold til de ovenstående bemærkninger, men ideelt set bør Title og First Heading altid være ens.
- *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. Dette er ikke altid tilfældet, 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.<sup>10</sup>
- *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>
  [f.eks.:]
  <META name="navnet på informationen i siden" content="information man vil give videre">
  <META name="author" content="Forfatterens navn">
</HEAD>
<BODY>
Her følger sidens synlige tekst m.m.
</BODY>
</HTML>
```

Det vigtige er her at lægge mærke til koden "META". Ved anvendelse af denne kode er det muligt at beskrive alt det ved WWW-siden, der er vigtigt for en søgning. Forskningsministeriet har ved at udgive "Retningslinier for offentlige on-line publikationer" og den efterfølgende "Standard for statslig elektronisk publicering - Version 18" prøvet at udarbejde en standard for offentlige WWW-sider, hvor de især har anvendt koden "META" til at konstruere en standard for en elektronisk kolofon<sup>11</sup>. 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" kode - kaldet 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.

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

Som de foregående afsnit angiver, er informationsøgning 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*

---

<sup>10</sup> Pfaffenberger: Web Search Strategies, p. 144-145

<sup>11</sup> En elektronisk kolofon er et sæt koder, der identificerer en række bibliografiske oplysninger om publikationen.

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.

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

---

## Få et overblik over søgeværktøjernes kvaliteter

---

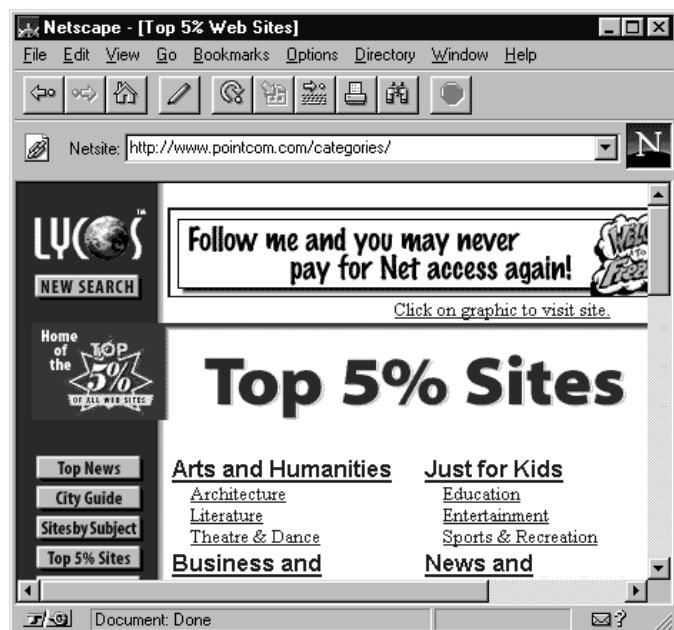
Der er hundredvis af forskellige værktøjer, der kan anvendes ved søgning på WWW. I dette afsnit vil vi opdele de forskellige værktøjer i fem forskellige typer, samt gennemgå de forskellige typers fordele og ulemper. Desuden vil vi opstille nogle vurderingsparametre, der gør det muligt at vurdere bredden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

### Udgangspunkter

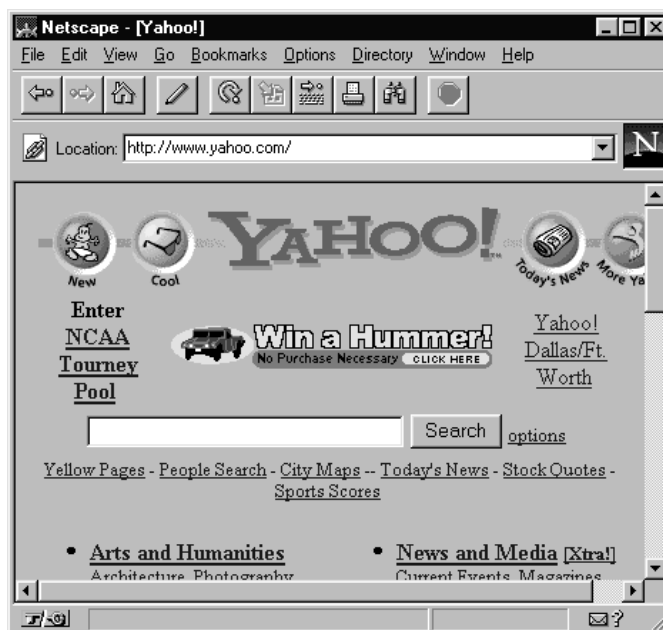
Udgangspunkter (Starting Points) er ofte et indeks som enten følger med WWW-software, eller er tilgængelig på elektroniske avisers og blades hjemmeside, eller andre virksomheder eller privatpersoner, der er interesseret i at vise mangfoldigheden i WWW. Dette kan være et udemærket mulighed for nybegyndere, 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.



Figur 5: Eksempel på et udgangspunkt.

## 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.



Figur 6: Eksempel på et emneindeks

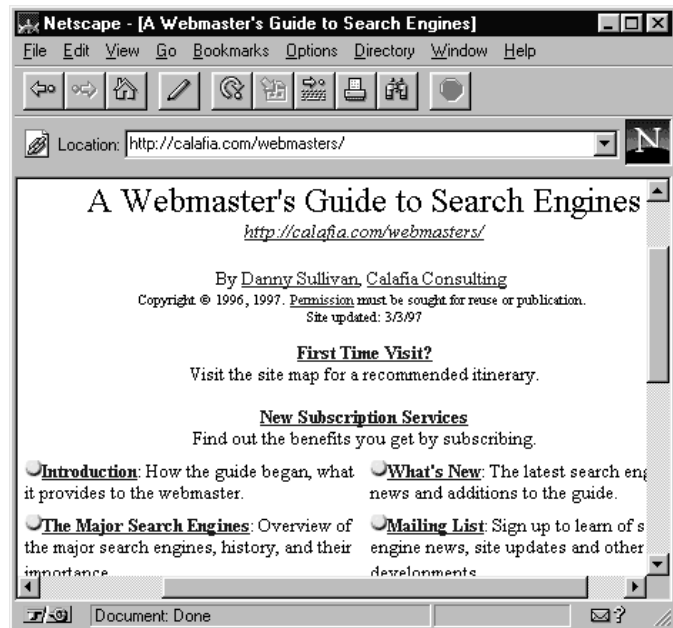
Normalt er emneindeksene vedligeholdt af personer, som sidder centralt og finder og klassificerer WWW-siderne. En variant heraf er decentrale emneindeks, hvor man trækker på forskellige (op til flere hundrede) personer med speciel ekspertise på et område.

Uanset om emneindekset er centralt eller decentralt, er ulempen, 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. tydeligt, og da WWW-sider med flot computergrafik er godt repræsenteret, må det vurderes, at det ofte er form og tilfældigheder, der afgør om siden er med i indekset. Yderligere må det antages at disse personers nationalitet og sprog spiller ind på udvælgelsen hvilke WWW-sider, der skal med i emneindekset.



## Vejledere

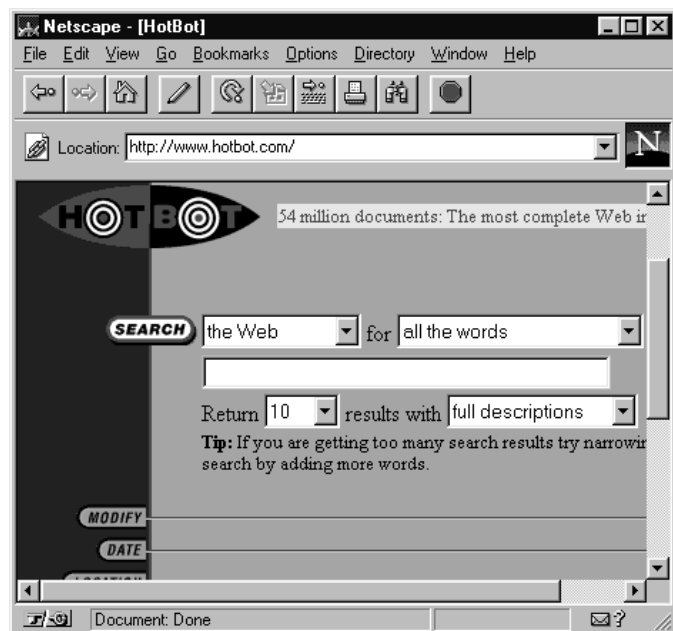
Vejledere er oprettet og vedligeholdt af personer med viden omkring et givent meget mere specifikt emneområde som denne form for emneindeks dækker. Vejledere kan derfor være meget værdifulde i en søgningsproces. Vejledere bliver (på engelsk) kaldt meget forskelligt som: collection, compendium, guide, 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.



Figur 7: Eksempel på en vejleder.

## Søgesystemer

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



Figur 8: Eksempel på et søgesystem.

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 søgesystemer ikke særlig præcise. Nogle søgesystemer søger i en database som ikke indeholder alle dele af WWW-siden. Disse søgesystemer kan derfor melde tilbage uden relevante oplysninger, da søgeordet er gemt langt væk

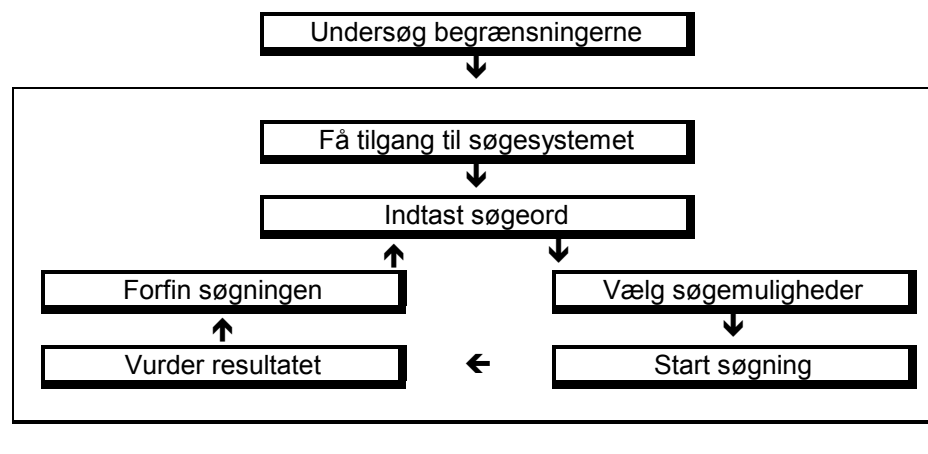
i selve teksten. Da ikke alle søgesystemer indekserer selve teksten betyder det, at en søgning i et system har en anden dybde end et andet søgesystem - søgninger får dermed forskellig dybde. Et andet problem er, at de svar søgesystemet giver kan have forskellig kvalitet, da robotten ikke er i stand til at kvalitetsvurdere WWW-siden.

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

Herved fremgår det, at ét søgesystem kan 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 søgesystemer. Som løsning på dette problem er der udviklet søgesystemer som søger i mange andre søgesystemer samtidig. Dog er ulemperne ved disse søg-i-et-systemer de samme som ved almindelige søgesystemer, og det er sjældent 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 tilgang 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 man kan anvende søgesystemet på andre tidspunkter. Derfor kan "global village"-ideen med at tid og sted forsvinder, nogen gange være svært at få øje på.

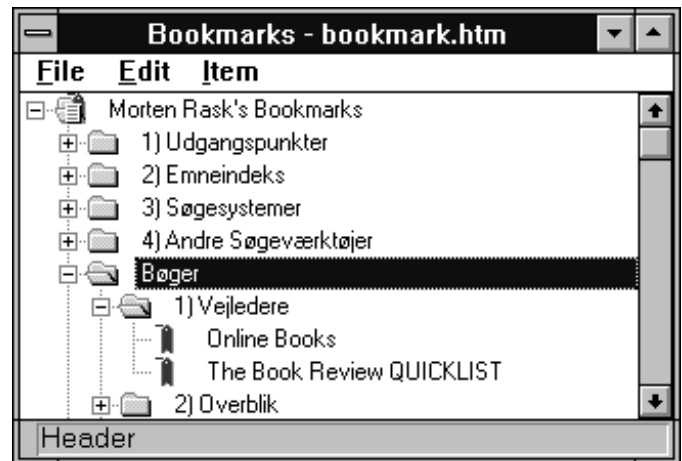
Uanset, hvilket søgesystem der benyttes, vil den generelle procedure være følgende:



Figur 9: Generel anvendelsesprocedure for et søgesystem

## 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 softwarepakker. 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.



Figur 10: Eksempel på brug af bogmærker.

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
- Søgnesystemer
- Emneområde #1
  - ⇒ Vejledere
  - ⇒ Overblik
  - ⇒ Rigtige gode informationer
  - ⇒ Anden god information
  - ⇒ Tvivlsom information
- Emneområde #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 man undgår at glemme, hvor man fandt gode informationer. Man må dog huske på, at disse bogmærker bliver forældede, og at de registrerede WWW-sider kan ændres eller slettes.

## Fordele og ulemper ved de forskellige typer af søgeværktøjer

Som det fremgår af det forrige er der fordele og ulemper ved de forskellige typer af søgeværktøjer. Tabel 1 opsummerer disse, hvor de fire første typer af søgeværktøjer følger en rækkefølge, der er bestemt ud fra, 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"> <li>Giver umiddelbart en god ide om WWWs mangfoldighed</li> </ul>	<ul style="list-style-type: none"> <li>Ikke dækkende overhovedet</li> <li>Andre personers holdninger og interesser begrænser søgningen</li> </ul>
	Emneindeks	<ul style="list-style-type: none"> <li>Meget omfattende oversigter</li> <li>Nemme at anvende</li> </ul>	<ul style="list-style-type: none"> <li>Ikke komplet dækkende</li> <li>Ikke faglig kompetente personer foretager kvalitetsvurderingen</li> <li>Andre personers holdninger og interesser begrænser søgningen</li> </ul>
	Vejledere	<ul style="list-style-type: none"> <li>Faglig kompetente personer foretager kvalitetsvurderingen og indekseringen</li> <li>Kan være guld værd</li> </ul>	<ul style="list-style-type: none"> <li>Svære at finde frem til</li> <li>Andre personers holdninger og interesser begrænser søgningen</li> </ul>
Mest dækkende Sværeste anvendelse ↓	Søgesystemer	<ul style="list-style-type: none"> <li>Ingen personers holdninger og interesser begrænser søgningen</li> <li>Præcis søgning er mulig, hvis det enkelte systems muligheder og begrænsninger kendes</li> </ul>	<ul style="list-style-type: none"> <li>Sværere at anvende med et godt resultat</li> <li>Kan være overbelastede</li> <li>Ofte nødvendigt at foretage søgninger i flere systemer</li> </ul>
	Bogmærker	<ul style="list-style-type: none"> <li>Anvender aktivt erfaring i søgningsprocessen</li> <li>Undgår at man glemmer gode søgningers resultat</li> </ul>	<ul style="list-style-type: none"> <li>Ingen ny information findes ved anvendelse af dette værktøj</li> </ul>

Tabel 1: Fordele og ulemper ved de forskellige typer af søgeværktøjer

Fordelene og ulemperne i Tabel 1 skal ses som, hvad der er karakteristisk for den enkelte type af søgeværktøj. Udbyderne 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.

## 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 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 udbyderne 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.

### Bredden af søgningen

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.<sup>12</sup> 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.<sup>13</sup> Usenet (NetNews) er et verdensomspændende kommunikationssystem med tusindvis af diskussionsgrupper.<sup>14</sup> Vi vil her fokusere på WWW, 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* udbyderne 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 område (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 *landedækning* anvendes til at kontrollere bredden af søgningen.

### Dybden af søgningen - generelt

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 indekse WWW-siden.

Relateret til dette er *resultat præsentationen* af søgningen. Søgesystemerne præsenterer ofte en del af WWW-siden som resultat efter relevans i forhold til søgeordet, og ved de forskellige indeks karakteriseres ofte den medtagne WWW-side med en beskrivelse og/eller en form for kvalitetskarakter (oftest med stjerner som ved vurderinger af film). Man bør her være opmærksom på, at søgesystemernes relevansvurdering er baseret på statistiske beregninger, hvor indeksenes relevansvurdering er baseret på personernes subjektive vurderinger.

### Dybden af søgningen - søgesystemer og emneindeks med søgemuligheder

Følgende parametre er kun anvendelige ved søgesystemer og ved de emneindeks, hvor det er muligt at angive søgeord.

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.

---

<sup>12</sup> Heilesen: Helt fra start på Internet og World Wide Web, p. 33

<sup>13</sup> Forskningsministeriet: Retningslinier for offentlige on-line publikationer, Ordliste

<sup>14</sup> Heilesen: Helt fra start på Internet og World Wide Web, p. 35

- *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 der f.eks. "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 kan søgningen baseret på et dansk søgeord mislykkes, 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 de danske "æ", "ø" og "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 - altså en sætning, der angiver et bestemt udtryk.

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 f.eks. 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, men hvor frasesøgning søger på ord i en bestemt rækkefølge, anvender konceptsøgning en statistisk baseret metode til at definere en bestemt sammenhæng.
- En anden måde at kontrollere dybden af søgningen, vil være at bruge *logiske operatoren*. Tabel 2 angiver funktionen af forskellige operatoren:

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

Tabel 2: Logiske operatører.

Kilde: Jørgensen, Web-søgessystemer, internetworld den 23-8-96, p. 29 eller <http://www.idg.dk/internetworld/artikler/searchtest/logik.htm>

Operatørene eller det alternative alias indsættes mellem søgeordene i søgeværktøjets søgefelt. Således kan en større kontrol med søgningens dybde opnås.

### Opsamling - Parametre der angiver bredden og dybden i søgningen

Således vil det i mange tilfælde være muligt at få kendskab til bredden og dybden af den foretagne søgning. Tabel 3 viser de parametre, man skal være opmærksom på, når man vurderer et søgeværktøj:

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

Tabel 3: Parametre der angiver bredden og dybden af søgningen

Endelig vil en sidste parameter være *Bemærkninger*, som er 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.

---

## Situationsbestemt 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* fortrinlige 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 efter formålet man kan have med søgningen:

1. WWWs mangfoldighed - Udgangspunkter
2. Specifik information - Simpel søgning
3. Information af høj kvalitet - Vejledere
4. Samtlige WWW-tilgængelige informationer - Avanceret søgning

I disse underafsnit, vil vi sammenligne konkrete søgeværktøjer i forhold til målet med søgningen. Under hvert underafsnit vil vi udvælge de bedste søgeværktøjer samt danske pendanter. Det kommer sig af, at de bedste søgeværktøjer vi kender til er af amerikansk oprindelse. Hvis vi var svenskere og havde henvendt os til andre svenskere, havde vi formodentlig valgt at tage svenske søgeværktøjer med i analysen.

### 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 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 kvalitet, der er inddelt efter emner som: "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 af, er ugebladet Ingeniørens HITLISTEN.



Således vil vi anbefale, at informationssøgeren undersøger WWWs mangfoldighed ved at besøge følgende i nævnte rækkefølge:

1. **Point** <http://www.pointcom.com/categories/>
2. **Hitlisten** <http://www.ingenioeren.dk/redaktion/usite.html>

Rækkefølgen er bestemt på baggrund af, at POINT giver den bedste idé om WWWs mangfoldighed, hvor HITLISTEN er det bedste danske alternativ.

## 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. Det er samtidigt det eneste søgeværktøj, som kan indekserer alle komponenter i WWW-siden. 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 i et søgesystem end at klikke sig frem i et emneindeks.

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.

I de følgende tabeller er søgesystemerne rangeret efter disse kriterier. En vigtig faktor er dog, at nogle søgesystemer som f.eks. ALTAVISTA hævder at fokusere på at indekserer så mange *forskellige* WWW-steder, som muligt, hvor andres antal af indekserede WWW-sider er mere et udtryk for, at søgesystemet har indekseret *alle* WWW-sider på *hvert* WWW-sted. Dette er dog ikke muligt at konstatere på andre måder, end hvis en søgnings resultat er præget af WWW-sider fra samme WWW-sted, må det antages, at pågældende søgesystem fokuserer på at indekserer i dybden.

Endelig er det vigtigt at sikre sig, hvorvidt den geografiske dækning er tilfredsstillende i forhold til emnet, man søger belyst ved yderligere information. 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 seriøse valgmuligheder - CYBERCITY AGENT, JORI og THOR, hvor THOR ikke er et selvstændigt system, men et filter oven på det amerikanske ALTAVISTA, der betyder, at man kun får danske resultater af søgningen.

Fordelen består primært af to ting. For det første behøver man ikke at sætte sig ind i, hvorledes de store globalt dækkende søgesystemer instrueres om kun at søge efter danske informationer. Det er altså hurtigere at gå direkte til et dansk søgesystem. For det andet giver de to største danske søgesystemer den bredeste søgning, hvis der er tale om danske oplysninger. Vores argumentation baseres på at Thor har en bredde på 200.000 WWW-sider, hvilket blot er et udtryk for, at 200.000 ud af ALTAVISTA'S 32 millioner WWW-sider er danske. Når de to andre søgesystemer er ca. dobbelt så brede som THOR, er det kun LYCOS, der i teorien kunne give en bredere søgning efter

danske oplysninger. Men, som det senere vil fremgå, har LYCOS ikke mulighed for at begrænse søgningen til kun danske oplysninger, og er derfor ikke relevant i denne sammenhæng.

Vi har derfor valgt at skelne mellem danske og globale søgesystemer, når der skal foretages en simpel søgning efter specifik information på WWW.

### Danske søgesystemer

Desværre er der ingen af de danske søgesystemer, der muliggør frasesøgning. Derfor rangeres de danske søgesystemer i Tabel 4 kun efter bredde i form af antal WWW-sider og komponenter, der medtages ved søgningen:

Søgesystem	Antal WWW-sider	Komponenter	Frasesøgning
<b>1. CyberCity Agent</b>	Ca. 420.000	URL, title, headings, all content <sup>15</sup> , hyperlinks, Meta Tag (kun relevant ved tilmelding)	Nej
<b>2. Jori</b>	Ca. 365.000	URL, title, all content, hyperlinks	Nej
<b>3. Thor (baseret på AltaVista)</b>	Ca. 200.000	URL, title, all content, hyperlinks, Meta Tag som beskrivelse og søgeord	Nej

Tabel 4: Danske søgesystemer rangeret efter antal indekserede WWW-sider, komponenter og frasesøgning.

Kilde: Bekræftet af service-udbyderne midtfebruar 1997.

Som det fremgår af Tabel 4, er CYBERCITY AGENT det bredeste søgesystem målt på antal indekserede WWW-sider og er i stand til at tage alle komponenter med i søgningen. Derfor anbefales det at starte med CYBERCITY AGENT og derefter gå nedad i skemaet og afprøve de andre søgesystemer, hvis man ikke finder relevante informationer.

Inden en søgning foretages er det også vigtigt at forstå, hvad man kan forvente af søgningen, og hvorledes søgesystemet behandler ens søgning:

Søgesystem	Resultat præsentation	Afskæring af søgeord	Lande / Sprog / grammatik	Store bogstaver	Behandling af flere søgeord
<b>1. CyberCity Agent</b>	Title, Relevans, første 1-3 linier på siden, URL	Automatisk, dog uden specielle tegn	Danmark / Dansk / Prøver at bestemme sproget	Aldrig	Indsætter OR automatisk
<b>2. Jori</b>	Title, Relevans, første 1-5 linier på siden, URL	Automatisk	Danmark / Dansk / Dansk	Aldrig	Indsætter AND automatisk
<b>3. Thor</b>	Title, URL, første 1-3 linier på siden, eller Meta Tag som beskrivelse	Præcis	Danmark, Finland, Norge, Sverige og alle lande / Vest Europæisk / Ikke afhængig af grammatik	Altid	Indsætter OR automatisk

Tabel 5: Standardindstillinger ved simpel søgning i danske søgesystemer.

Kilde: Bekræftet af service-udbyderne midtfebruar 1997.

<sup>15</sup> ALL CONTENT: er en sandhed med modifikationer, de indekserer det første 2048 (2K) af dokumentet, efter det er befriet for html-koder, specialtegn, og mellemrum - dette svarer meget ofte til hele dokumentet, men ikke altid.

Tabel 5 viser, at resultatet præsenteres på omtrent den samme måde. Men meget vigtigt er, at oplysningerne i Tabel 5 angiver, hvorfor den samme søgning i de forskellige søgesystemer behandles forskelligt. Hvis man ikke er opmærksom på dette, vil det blive svært at få succes med informationsøgningen.

### Globale søgesystemer

Hvis man ikke kunne finde, hvad man søgte, eller man ikke er særlig interesseret i danske oplysninger, er det nødvendigt at anvende større globale søgesystemer. Dette kan også være tilfældet, hvis det er helt nye oplysninger man søger, da de to danske søgesystemer for tiden ikke opdateres ofte. Dette vil blive uddybet i afsnittet "Overvåg nye informationer på WWW".

Vi har udvalgt de seks største søgesystemer, som i Tabel 6 også er rangeret efter den rækkefølge vi ville anvende dem for hurtigt at foretage en præcis og bred søgning:

Søgesystem	Antal af WWW-sider	Komponenter	Frasesøgning
<b>1. HotBot</b>	Ca. 54 mio.	URL, title, headings, all content, hyperlinks, Image JavaScript, Java, Audio, Acrobat, Shockwave, VRML, Smiley, File-extensions, Meta Tag beskrivelse og søgeord	Ja (men ikke enkelt bogstavs ord)
<b>2. InfoSeek</b>	Ca. 50 mio.	URL, title, headings, all content, hyperlinks, Meta Tag beskrivelse og søgeord	Ja
<b>3. Excite</b>	Ca. 50 mio.	URL, title, headings, all content, hyperlinks	Nej men koncept søgning
<b>4. AltaVista</b>	Ca. 32 mio.	URL, title, headings, all content, hyperlinks, ActiveX, Java, Image, Meta Tag beskrivelse og søgeord	Ja
<b>5. Lycos</b>	Ca. 66 mio.	URL, title, headings, excerpt, hyperlinks	Nej
<b>6. OpenText</b>	Ca. 20 mio.	URL, title, headings, all content, hyperlinks	Ja

Tabel 6: Globale søgesystemer rangeret efter antal indekserede WWW-sider, komponenter og frasesøgning.

Kilde: Bekræftet af service-udbydere midtfebruar 1997.

Disse globale søgesystemer er ikke rangeret på den samme logiske måde som ved de danske søgesystemer. Men stadig anbefales det at starte med den øverste - her HOTBOT og derefter gå nedad i skemaet og afprøve de andre søgesystemer, hvis man ikke finder relevante informationer.

Grunden til at INFOSEEK rangerer højere end EXCITE er, at INFOSEEK medtager flere komponenter i søgningen. LYCOS er det næstlaveste rangerede, selvom den har indekseret flest WWW-sider. Det skyldes, at LYCOS ikke kan foretage frasesøgning, og derfor er upræcis.

Igen er det vigtigt at forstå forudsætningerne bag søgesystemets behandling af søgningen. Dette er endnu vigtigere ved globale søgesystemer, især da nogle er baseret på amerikansk sprog, grammatik og tegnsætning.

Tabel 7 opsummerer standardindstillingerne ved simpel søgning i globale søgesystemer:

Søgesystem	Resultat præsentation	Afskæring af søgeord	Lande / Sprog / grammatik	Store bogstaver	Behandling af flere søgeord	Frasekode
<b>1. HotBot</b>	Title, Relevans, første 1-5 linier på siden eller Meta Tag som beskrivelse, URL, størrelse, dato	Præcis	Alle lande / Amerikansk / Ikke afhængig af grammatik	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
<b>2. InfoSeek</b>	Relevans, Title, første 1-5 linier på siden eller Meta Tag som beskrivelse, URL, størrelse	Ental	Alle lande / Amerikansk / Specielle tegn som ø, æ og å medtages ikke	Altid	Indsætter OR automatisk	"-tegn afgrænser
<b>3. Excite</b>	Relevans, Title, URL, automatisk generet beskrivelse	Ordstamme	Alle lande / Amerikansk / ?	Altid	Koncept søgning, ellers indsættes OR automatisk	Ikke mulig - bruger i stedet koncept søgning
<b>4. AltaVista</b>	Title, URL, første 1-3 linier på siden eller Meta Tag som beskrivelse	Præcis	Alle lande / Vest Europæisk / Ikke afhængig af grammatik	Altid, men små fanger alt	Indsætter OR automatisk	"-tegn afgrænser og andre tegn sammenbinder frasen
<b>5. Lycos</b>	Title, URL, Relevans, antal søgeord, første 3-8 linier på siden	Automatisk	Alle lande / Amerikansk / ?	Aldrig	Indsætter OR automatisk	Ikke mulig
<b>6. OpenText</b>	Title, URL, første 1-5 linier på siden, Relevans	Automatisk	Alle lande / Amerikansk / ?	Egen- navne	Indsætter OR automatisk	Vælg "this phrase"

Tabel 7: Standardindstillinger ved simpel søgning i globale søgesystemer.

Kilde: Bekræftet af service-udbyderne midtfebruar 1997.

Igen er der stor forskel på, hvorledes de forskellige søgesystemer vil behandle den samme søgning. Yderligere er det vigtigt at være opmærksom på, at mange af søgesystemerne kun forstår amerikansk, samtidig med at de afskærer søgeordet. Dette kan give nogle meget underlige søgeresultater.

### Opsamling - simpel søgning

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

Generelt er det en fordel at huske følgende, når en simpel søgning foretages:

1. Hvis det er muligt, brug frasesøgning
2. Brug et søgesystem med bred dækning
3. Brug din erfaring og intuition til at finde tre ord der flertydigt beskriver den ønskede information
4. Brug din erfaring og intuition til med synonymer at beskrive den ønskede information
5. Husk rigtig stavning
6. Brug ikke WWW-termer som http, WWW og andre almindelige ord som den (the), og (and), da alle søgesystemer på nær INFOSEEK har en liste, der udelader disse ord
7. Brug ikke søgeord i flertal, med mindre du bruger frasesøgning

Det vigtigste er dog de bagvedliggende forudsætninger for en simpel søgning, da disse er forskellige fra søgesystem til søgesystem. Samtidigt er det vigtigt at huske, at søgesystemerne er kommercielle produkter, således kan der ikke stoles hundrede procent på den oplyste funktionalitet. Med dette i baghovedet vil vi anbefale anvendelse af søgesystemerne i denne rækkefølge:

Hvis man kun er interesseret i danske informationer:

1. **CyberCity Agent** <http://agent.cybercity.dk/ormen/simple.idc>
2. **Jori** <http://jori.jubii.dk>
3. **Thor (baseret på AltaVista):** <http://www.danielsen.com/thor/>

I alle andre tilfælde:

1. **HotBot** <http://www.hotbot.com>
2. **InfoSeek** [http://www.infoseek.com/Home?pg=ultra\\_home.html](http://www.infoseek.com/Home?pg=ultra_home.html)
3. **Excite** <http://www.excite.com>
4. **AltaVista** <http://altavista.digital.com>
5. **Lycos** <http://www.lycos.com>
6. **OpenText** <http://index.opentext.net>

Hvis du har foretaget søgninger efter foregående anvisninger, har du sikkert hurtigt fundet en mængde information. Imidlertid er det ikke dermed sagt, at al den fundne information er af lige god kvalitet. Hvis du ønsker at finde information af god kvalitet, må du nok ty til at finde nogle gode vejledere via søgning i emneindeks. Dette handler det næste afsnit om.

## Information af høj kvalitet - vejledere via emneindeks

Når målet med informationsøgningen er at finde 10-20 WWW-sider af høj kvalitet, er bredden af søgningen ikke så vigtig. Med hensyn til kvaliteten vil søgesystemerne ikke være oplagte, da de giver 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 de to danske INDEKS.DK og JUBII. Valget er faldet på disse på grund af deres størrelse, hvilket stadig er vigtigt, da emneindeks i sig selv er langt mindre dækkende end søgesystemerne. Med hensyn til at finde information af høj kvalitet er det ikke længere så vigtigt, hvilket geografisk område et emneindeks dækker, da man ikke på forhånd kan vurdere, om den ønskede information er af dansk eller af anden national oprindelse. Imidlertid har vi set os nødsaget til at tage danske eksempler med, da vores erfaringer viser, at YAHOO - som ellers er at foretrække - har en modvilje mod at optage WWW-sider, der er skrevet på andre sprog end engelsk.

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 Tabel 8:

Emneindeks	Antal WWW-sider	Resultat præsentation	Søgemulighed
<b>1. Yahoo!</b>	Ca. 200.000	Titel, Få liners beskrivelse, Beskrivende symboler for "nyhed" og "blandt de bedste"	Som søgesystemernes simpel søgning
<b>2. Indeks.dk</b>	Ca. 8.900	Titel, URL og evt. få liners beskrivelse sidens ansvarlig selv har indtastet	Meget simpel (indsætter AND mellem søgeord)
<b>3. Jubii</b>	Ca. 8.000	Titel, 1 linies beskrivelse sidens ansvarlig selv har indtastet	Meget simpel

Tabel 8: Oversigt over Emneindeksene.

Kilde: Bekræftet af service-udbyderne midtfebruar 1997.

Vi foreslår altså Yahoo! som første prioritet - YAHOO kalder vejledere for "Indices". Hvis man har på fornemmelsen, at informationen foreligger på dansk, vil vi foreslå INDEKS.DK og derefter prøve JUBII. Rækkefølgen de to danske emneindeks bør benyttes, er dog i praksis underordnet, da Indeks.dk kan have dubletter, og derfor ikke er helt så omfattende som antallet af WWW-side angiver.

Et andet godt sted at starte i emneindeksene for at finde vejledere, er at finde forskningsbibliotekernes WWW-steder. 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.

Men som start på søgningen efter gode vejledere i emneindeks er rangeringen følgende:

1. **Yahoo!** <http://www.yahoo.com>
2. **Indeks.dk** <http://www.forum.dk/index/>
3. **Jubii** <http://www.jubii.dk>

Forudsat at man stadig ikke er tilfreds med resultatet af ens søgning, eller formålet med søgningen ganske enkelt er at finde samtlige WWW-tilgængelige informationer - så er det en god idé at anvende avanceret søgning. Dette diskuteres i næste afsnit.

### 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 at søgningen kan blive så bred som mulig. Derfor vil de søgesystemer, som blev anbefalet i afsnittet om "Specifik information - simpel søgning" stadig være relevante. Igen gælder det, at man må være opmærksom på begrænsningerne ved de forskellige søgesystemer.

Men muligheden for at kontrollere søgningen er her endnu vigtigere, 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", "Avanceret søgning" eller lignende.

Således vil de to følgende tabeller gennemgå de forskellige søgesystemers kapacitet i denne retning. Når "Ingen" er angivet, betyder det, at søgesystemets standardindstilling ikke kan ændres. Rangeringen er ikke nødvendigvis helt den samme som i afsnittet "Specifik information - simpel søgning", da det er de yderligere muligheder man har for at kunne kontrollere søgningen, der i realiteten afgør, hvilke søgesystemer - sammen med bredden af søgesystemet - der er bedst til at finde al WWW-tilgængelig information.

### Danske søgesystemers yderligere kontrolmuligheder ved avanceret søgning

Vi har igen opdelt søgesystemerne i danske og globale søgesystemer, da danske søgesystemer kan være relevante hvis al WWW-tilgængelig information på dansk ønskes. Tabel 9 viser, at de danske søgesystemer ikke kan byde på mange yderligere kontrolmuligheder:

Søgesystem	Begrænset feltsøgning	Behandling af flere søgeord	Bemærkninger	Afskæring af søgeord	Resultat præsentation
1. CyberCity Agent	Kan begrænse søgning til URL eller et givent WWW-sted	OR (skal-ord), AND (kan-ord), NOT (nej-ord)	Kan vise: Opdateret sidst, links til og fra siden, sidens hierarkiske placering	Ingen	Ingen
2. Jori	Ingen	AND (eller og) og "-" tegnet som NOT, "+" tegnet prioriterer søgeordet højere	Ingen	Ingen	Ingen
3. Thor	Ingen	Ingen	Ingen	Ingen	Ingen

Tabel 9: Yderligere kontrolmuligheder ved avanceret søgning i danske søgesystemer.

Kilde: Bekræftet af service-udbyderne midtfebruar 1997.

Grunden til at CYBERCITY AGENT anses for at være de bedste søgesystem i denne sammenhæng er, at det samtidig med at være det mest dækkende, også giver informationsøgeren de bedste muligheder for at kontrollere sin søgning.

### Globale søgesystemers yderligere kontrolmuligheder ved avanceret søgning

I de fleste tilfælde, hvor al WWW-tilgængelig information ønskes, må man ty til avanceret søgning i de globale søgesystemer. Dette skyldes, at alle de undersøgte globale søgesystemer har en langt større bredde end de danske søgesystemer, og at alle de globale søgesystemer giver meget bedre muligheder for at kontrollere søgningen. Sidstnævnte faktor ses tydeligt i Tabel 10:

Søgesystem	Begrænset feltsøgning	Behandling af flere søgeord	Bemærkninger	Afskæring af søgeord	Resultat præsentation
<b>1. HotBot</b>	Tekst, Frase, Person, URL, Date, Location, Image, JavaScript, Java, Audio, Acrobat, Shockwave, VRML, Smiley, File-extensions	Must eller should eller must not contain words	Specificer regioner og lande	Ingen	Ingen
<b>2. AltaVista</b>	Summary, Title, First Heading, Hyperlinks, Java Applets, Host, URL, Images, Dates	AND, OR, NOT, NEAR (højst 10 ord imellem søgeord), brug af parenteser	Søge forbedring,, LiveTopics er en kommende service som repræsenterer resultaterne grafisk som et tænketræ eller i tabelformat a la emneindeks.	Præcis ændres til Ordstamme ved brug af * efter søgeordet	Resultat præsentationstyper kan bestemmes.
<b>3. InfoSeek</b>	Hyperlink, Site, URL, Title, Image og Java	Ingen	Søge forbedring, Genkender egennavne	Ingen	Ingen
<b>4. OpenText</b>	Summary, Title, First Heading, Hyperlinks	AND, OR, BUT_NOT, NEAR, FOLLOWED BY	Søge forbedring, Find lignende sider, Vægtet søgning	Automatisk ændres til Præcis med mellemrum efter søgeordet	Relevans kan sorteres efter antal gange søgeordet optræder eller om søgeordet forekommer eller ikke forekommer
<b>5. Excite</b>	Egennavne	AND, OR, AND NOT og ()	Find lignende sider, Vægtet søgning, Forespørgsel-med-eksempel	Ingen	Kan sortere resultatet ud fra relevans eller URL
<b>6. Lycos</b>	Ikke mulig	AND, OR og "-" tegnet som NOT	Ingen	Automatisk kan ændres til Præcis med punktum efter søgeordet eller med \$ som det første i søgeordet ændres til Ordstamme	Formatet af resultatet kan være: Standard, Summary eller Details

Tabel 10: Yderligere kontrolmuligheder ved avanceret søgning i globale søgesystemer.

Kilde: Bekræftet af service-udbyderne midtfebruar 1997.

Ved rangeringen af søgesystemer, når de skal anvendes til avanceret søgning, har vi fokuseret på informationsøgerens muligheder for at kontrollere søgningen. Tabel 10 er opstillet således, at jo



længere til højre kontrolmulighederne er beskrevet, desto mindre vigtige er de for kontrollen af søgningen. Således er rangeringen foretaget på baggrund af, i hvor høj grad det er muligt at kontrollere søgningen sammenholdt med hvor dækkende søgesystemet er.

### Opsamling - avanceret søgning

Det foregående viser, at der er mange måder at specificere sin søgning på. Det 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. Vi vil anbefale anvendelse af søgesystemerne i denne rækkefølge:

Hvis man kun er interesseret i danske informationer:

1. **CyberCity Agent** <http://agent.cybercity.dk/ormen/advanced.idc>
2. **Jori** <http://jori.jubii.dk>
3. **Thor** <http://www.danielsen.com/thor/>

I alle andre tilfælde:

1. **HotBot** <http://www.hotbot.com>
2. **AltaVista** <http://altavista.digital.com/cgi-bin/query?pg=aq>
3. **InfoSeek** <http://www.infoseek.com/Home?pg=Home.html>
4. **OpenText** <http://index.opentext.net/search/powersearch.html>
5. **Excite** <http://www.excite.com/>
6. **Lycos** <http://www.lycos.com/customsearch.html>

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.

---

## 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":

- |   |   |  |  |
|---|---|--|--|
| <ul style="list-style-type: none"> <li>• Reference Trailblazer Pages</li> <li>• Acronyms</li> <li>• Authors</li> <li>• Book Publishers</li> <li>• Calendars</li> <li>• Census Data</li> </ul> | <ul style="list-style-type: none"> <li>• Colleges and Universities</li> <li>• Dictionaries</li> <li>• Earthquakes</li> <li>• Encyclopedias</li> <li>• Geography and Travel</li> </ul> | <ul style="list-style-type: none"> <li>• Literary Works</li> <li>• Maps</li> <li>• Measurements</li> <li>• Movies</li> <li>• Patents</li> <li>• Phone Numbers</li> <li>• Quotations</li> </ul> | <ul style="list-style-type: none"> <li>• Religious Texts</li> <li>• Scientific Constants</li> <li>• Tax Forms</li> <li>• Thesauri</li> <li>• Time Zone Converters</li> </ul> |
|---|---|--|--|

- Translation Aids
- Usenet FAQs
- Zip Codes

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

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

Disse kilder beskrives og vurderes af Pfaffenberger i hans bog, hvor bogens hjemmeside 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 mange søgemuligheder er samlet på et sted. I THE INTERNET SLEUTH er søgemulighederne opdelt under følgende kategorier:

- |  |   |   |  |
|--|---|---|--|
| <ul style="list-style-type: none"> <li>• Agriculture</li> <li>• Arts</li> <li>• Biology</li> <li>• Bio Sciences</li> <li>• Business</li> <li>• Business Directories</li> <li>• Chemistry</li> <li>• Commercial Databases</li> <li>• Companies</li> <li>• Computer Related</li> </ul> | <ul style="list-style-type: none"> <li>• Economics</li> <li>• Education</li> <li>• Employment</li> <li>• Engineering</li> <li>• Entertainment</li> <li>• Environment</li> <li>• Finance</li> <li>• Food &amp; Drink</li> <li>• Genealogy</li> <li>• Government</li> <li>• Health</li> <li>• Humanities</li> </ul> | <ul style="list-style-type: none"> <li>• Internet</li> <li>• Legal</li> <li>• Mathematics</li> <li>• Medicine</li> <li>• Music</li> <li>• News</li> <li>• Physics</li> <li>• Recreation</li> <li>• Reference</li> <li>• Regional</li> <li>• Sciences</li> <li>• Shopping</li> </ul> | <ul style="list-style-type: none"> <li>• Social Sciences</li> <li>• Software</li> <li>• Sports</li> <li>• Technology</li> <li>• Trade &amp; Industry</li> <li>• Travel</li> <li>• Usenet News</li> <li>• Veterinary</li> <li>• Web Search Engines</li> </ul> |
|--|---|---|--|

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.

Men som start kan vi anbefale at besøge følgende adresser i håb om at finde specialiserede referencesystemer:

- **The Reference Desk** <http://watt.seas.virginia.edu/~bp/searchme/refdesk.htm>
- **Web Search Strategies** <http://watt.seas.virginia.edu/~bp/searchme/welcome.html>
- **The Internet Sleuth** <http://www.isleuth.com/>

---

## 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, samt hvor nye informationerne, man har fundet ved hjælp af de gennemgåede søgeværktøjer, kan forventes at være. Der kan med andre ord være tale om

forskellige situationer, hvor man ønsker at være opdateret med hensyn til at overvåge nye informationer på WWW:

- Er nye informationer blevet tilgængelige på WWW?
- Hvor nye kan de fundne informationer forventes at være?
- Har en bestemt WWW-side ændret sig?

Der er forskellige værktøjer, der kan hjælpe med til at svare på disse spørgsmål. Disse værktøjer er som de gennemgåede søgeværktøjer forskellige, men fælles for dem gælder det, at de kan hjælpe én, således at man ikke skal klikke sig igennem emneindekset igen eller foretage søgningen på ny.

## Nye informationer på WWW

Når der er tale om at blive opmærksom på, at nye informationer er blevet tilgængelige på WWW, kan man ikke søge efter dem, da man jo ikke ved de eksisterer. Der findes heldigvis søgeværktøjer som kun beskæftiger sig med nylig tilgængelig informationer på WWW. Således kan man anvende dem til at orientere sig om nye tiltag på WWW.

Som der findes udgangspunkt-sider, der giver et billede af WWWs mangfoldighed, findes der også udgangspunkt-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 THE SCOUT REPORT, som kan sendes via e-post eller læses i ens browser. Af danske eksempler kan DOM@IN nævnes, der varetages af den danske ugeavis ComputerWorld. DOM@IN er kun tilgængelige via ens browser.

Endelig er der den kommercielle service WHAT'S NEW TOO, som er et simpelt søgesystem, som udelukkede beskæftiger sig med nye informationer på WWW. Dette kan også være en mulighed for at finde nye relevante informationer på WWW.

## Søgeværktøjernes ajourføring

I den tidligere beskrivelse af forskellige søgeværktøjer har vi været lidt inde på at alle søgeværktøjer ikke er lige ajourført. Med andre ord man kan ikke være sikker på, at det man har fundet er særligt opdateret. Hvor ny information man kan forvente at finde i emneindeks og søgesystemer, vil vi nu diskutere.

De personer, der varetager indekseringen af emneindeksene, er ofte overbelastede, og de maskiner der varetager indekseringen af søgesystemernes databaser, kunne være hurtigere. Det betyder, at det kan tage tid inden tilmelding af WWW-siden bliver registreret - dette kalder vi for "Forsinkelse på indekseringen". Endvidere foretages ajourføring af emneindekset/databasen ikke nødvendigvis hver dag. Derfor må man være opmærksom på disse parametre, når man vurderer, hvor nye informationer man kan forvente at finde med det pågældende søgeværktøj.

Nogle søgeværktøjer giver én mulighed for at orientere sig under dets nyhedsside, således det vil være muligt at se om informationer, der har ens interesse, er kommet til. Ligeledes kan man

tilmelde sig en e-post-liste, så man løbende får tilsendt nyhedsbreve om nye informationer, der er blevet tilgængelig via det pågældende søgeværktøj.

### Emneindeks

I Tabel 11 er nyhedsparametrene listet for de før gennemgåede emneindeks:

Emneindeks	Forsinkelse på indekseringen	Indekset ajourføres	Nyhedsbrev/-side
1. Indeks.dk	Ingen	Hver nat for nye, hver måned for fjernelse af gamle	Brev/Side
2. Jubii	1-4 dage	Mandag og torsdag	Intet/Side
3. Yahoo!	Vides ikke	Vides ikke	Intet/Side

Tabel 11: Emneindeksenes nyhedsparametre.

Kilde: Bekræftet af service-udbyderne midtfebruar 1997.

Tabel 11 er rangeret efter emneindeksets nyhedsgrad. Yderligere kan man se, at alle emneindeks har en nyhedsside og INDEKS.DK ligeledes giver mulighed for at få et e-post brev tilsendt.

### Søgesystemer

Med undtagelse af nyhedsbrev og -side har søgesystemerne lignende nyhedsmuligheder. Udover dette kan man vælge at gemme en søgning man ønsker at holde konstant opdateret. I Tabel 12 er nyhedsparametrene for søgesystemerne igen rangeret efter nyhedsgrad:

Søgesystem	Forsinkelse på indekseringen	Databasen ajourføres	Kan søgning gemmes?
1. AltaVista	Fra ingen forsinkelse til 4 dage (1 dag)	1-2 dage (1 dag)	Ja
2. OpenText	1 dag	Hver dag	Nej
3. InfoSeek	1 dag (Ingen)	1-2 dage	Ja
4. HotBot	1 uge	Hver uge	Ja
5. Excite	2 uger (Ca. 2 uger)	Hver uge (ca. hver uge)	Ja
6. Lycos	3 uger	2-4 uger	Ja
<b>Danske søgesystemer</b>			
1. Thor	Fra ingen forsinkelse til 4 dage (1 dag)	1-2 dage (1 dag)	Ja
2. CyberCity Agent	Måneder (Gennemsnitlig 12 timer)	Flere måneder i mellem (Konstant)	Ja
3. Jori	Måneder (1 uge)	Måneder (Hver uge)	Ja

Tabel 12: Søgesystemernes nyhedsparametre.

Kilder: Oplysningerne i parentes er service-udbydernes oplysninger hvis disse afviger fra vores egne erfaringer eller oplysninger fra Search Engine Guidens "Search Engine Features"<sup>16</sup>.

Denne rangering viser, at når det gælder søgesystemernes nyhedsgrad, er der ikke tale om samme rangering, som vi kom frem til under de andre trin. Forskellen er dog ikke ret stor (med undtagelse af CYBERCITY AGENT og JORI). Det skal dog også bemærkes, at søgesystemernes nyhedsparametre er et meget fejlbehæftet område. Eller sagt på en anden måde, her er det vigtigt at anse de enkelte

<sup>16</sup> Search Engine Guidens "Search Engine Features": <http://calafia.com/webmasters/chart.htm>

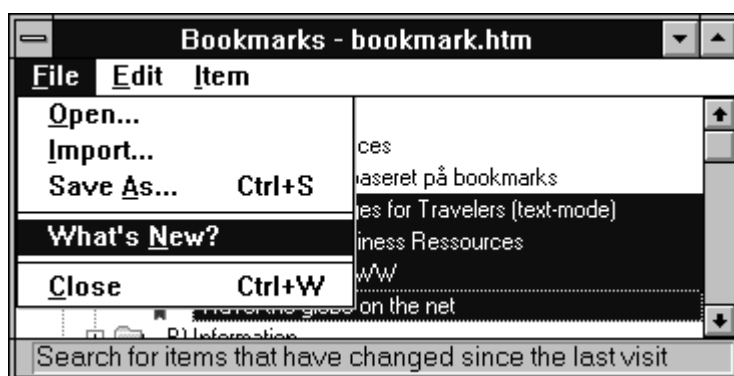
søgesystemer som kommercielle produkter, da der er stor forskel på, hvad de siger, og hvad de i realiteten gør. Søgningen kan ofte gemmes som bogmærke. Bogmærkefunktionen har nogle yderligere nyhedsmuligheder. Dette beskrives i det følgende afsnit.

## Ny information på kendte WWW-sider

Hvis man allerede kender URL'en på en WWW-side og blot vil være sikker på at blive gjort opmærksom på, at informationerne her har ændret sig, skal bogmærkefunktionen og andre lignende værktøjer anvendes.

### 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 (f.eks. 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?:



Figur 11: What's New? i Netscape browserens bogmærkefunktion

Her kan man vælge overvågning af alle eller udvalgte bogmærker:



Figur 12: Netscape browseren kan undersøge om bogmærkede WWW-sider har ændret sig

Dermed er denne WHAT'S NEW?-funktion meget anvendelig til at overvåge, om fundne WWW-sider har ændret sig. Det kræver dog, at man selv giver programmet besked på at undersøge dette.

### The URL-minder

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. Her tilmeldes de WWW-sider, der ønskes overvåget. Denne service kan selvfølgelig også overvåge søgeværktøjers nyhedssider. Herefter vil man ved hver ændring få en e-post baseret besked. Herved undgår man selv at skulle checke sine bogmærker. Dog kan den ikke håndtere gemte søgninger - her må man anvende WHAT'S NEW? -funktionen. Brug derfor:

- **The URL-minder:** <http://www.netmind.com/url-minder/>

### Opsamling - overvågning af nye informationer på WWW

Ovenstående gennemgang af de muligheder, informationssøgeren har for at være opdaterede med hensyn til ny information på WWW, kan forekomme besværlige at håndtere, da man konstant skal undersøge så mange WWW-sider. Men hvis man kombinerer de forskellige værktøjer fornuftigt, kan der blive tale om en forholdsvis nem og effektiv overvågning.

Dog bør man altid være opmærksom på, at søgeværktøjerne ikke er helt opdaterede, og at man kan gemme ens søgninger som bogmærker. Udover dette opfordres informationssøgeren til at bruge bogmærkefunktionens WHAT'S NEW?-funktion og THE URL-MINDER til at undersøge om enkelte centrale WWW-sider (som f.eks. konkurrentens produktside) har ændret sig.

Når det gælder oplysninger fra søgeværktøjerne er det kun for et enkelt søgeværktøjs vedkommende muligt få et nyhedsbrev tilsendt via e-post. Dette gælder for:

- **The Scout Report** <http://wwwscout.cs.wisc.edu/scout/report/index.html>

Dette er klart den bedste måde at holde sig opdateret på, da man blot en gang behøver at tilmelde sig på ovenstående URL'er og derefter løbende vil få tilsendt nyhedsbreve. Andre søgeværktøjer er man tvunget til at besøge. Følgende lister disse søgeværktøjer og de aktuelle URL'er:

- **The Scout Report** <http://wwwscout.cs.wisc.edu/scout/report/current/index.html>
- **Dom@in** <http://www.idg.dk/cw/domain.htm>
- **What's New Too** <http://newtoo.manifest.com/today.html>
- **Indeks.dk** <http://www.forum.dk/index/nyheder/>
- **Jubii** <http://www.jubii.dk/nyheder/default.htm>
- **Yahoo!** <http://www.yahoo.com/new/>

Men ved at bruge bogmærkefunktionens WHAT'S NEW?-funktion eller THE URL-MINDER kan man holde sig opdateret med ændringer på nyhedssiderne og kun besøge ovenstående søgeværktøjer, når ændringer har fundet sted.

Det er imidlertid langt fra alle søgeværktøjer kan oplyse om nye informationer, der er blevet gjort tilgængelige via det pågældende søgeværktøj. Derfor er det vigtigt, at den enkelte informationssøger videregiver sin viden til andre personer, der kan være interesseret i den fundne information. Dette er det sidste trin i søgningsprocessen, som vi nu kort vil skitsere indholdet af.

---

## **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. Det kan være ens kollegaer, studiekammerater, venner eller mere idealistisk alle WWW-brugere.

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

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. Udover dette kræver det kendskab til design af WWW-sider, hvilket kan være ret så tidskrævende.

Vi har oprettet et WWW-sted baseret på dette paper, således at alle de ændringer, der vil komme, vil blive opdateret her. Således har vi dels opnået, at det er altid er muligt at have et opdateret paper og samtidig har vi prøvet at give vores viden videre til andre, som ønsker at forstå informationssøgning på WWW. Denne viden har vi transformeret til nogle WWW-sider, som gør det nemt og enkelt, at foretage informationssøgning på WWW uden at skulle have dette paper ved siden af computeren. Adressen er <http://www.i4.auc.dk/rask/infoswww/infoguid.htm>

Endelig, hvis du har spørgsmål eller kommentarer vedrørende informationssøgning på WWW, kan du kontakte os på følgende e-post adresse: [rask@i4.auc.dk](mailto:rask@i4.auc.dk)

---

## Ordliste

---

Term	Beskrivelse
Adresse	Adressen er det unikke navn, som en ressource på Internet kan have. Adressen kan f.eks. henvise til personer ( <i>e-postadresser</i> ), dokumenter ( <i>URL'er</i> ), binære filer ( <i>FTP-adresser</i> ) eller andet. Adressen er systematisk opbygget. Ved e-postadresser står personens e-postnavn først efterfulgt af serverens navn, f.eks.: <code>en_person@fsk.dk</code> , mens andre ressourcer ofte angives med protokollen først efterfulgt af serverens navn og dokumentets eller filens beliggenhed og navn f.eks.: <code>http://www.auc.dk/html-filer/et_dokument.html</code> .
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 søge ordet "skibstøm" give resultater, hvor f.eks. 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.
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 f.eks. kun at søge på titler af WWW-sider.
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.
Browser	Browseren, WWW-browseren eller Web-browseren, er det program, som brugeren anvender til at se på WWW-sider med. Programmet læser siderne med eventuelle tilhørende billeder via nettet og fortolker koderne for derefter at opbygge skærbilledet på brugerens maskine. Det er derfor browseren og evt. brugerens opsætning af den, der afgør, hvordan et givet dokument ser ud hos brugeren.
E-post	E-post eller e-mail står for elektronisk post og er betegnelsen for den slags postbeskeder, der kan sendes via et computernet. I dag er e-post næsten blevet synonymt med Internet-post, altså elektronisk post, der kan sendes via Internet. Mange brugere kan sende e-post ad mange forskellige veje fra deres daglige postprogram; f.eks. intern post, X-400-post og post via Internet.
Emneindeks	Emneindeksene (Subject catalogs) er et WWW-sted, hvor personer manuelt har kategoriseret nogle WWW-sider ved at bruge emne klassificeringer. Man klikker på links der fanger ens interesse indtil man klikker på et 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.
Frasesøgning	Frasesøgning er søgninger med flere søgeord som en frase. Dermed kontrollerer man, i hvilken rækkefølge ordene skal stå, for at søgningen bør give resultat.
Homepage	Homepage eller hjemmeside er den side, som brugeren ser som den 'øverste' i hierarkiet på en <i>WWW-sted</i> . Mange virksomheder, myndigheder eller enkeltpersoner har en hjemmeside som indgang til de informationer, de gerne vil stille til rådighed via <i>WWW</i> . Der kan sagtens ligge mere end én hjemmeside på en server, ligesom begrebet er så åbent, at en enkelt instans også sagtens kan have flere hjemmesider.



HTML	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. <STRONG> eller <A HREF="2">. Mærkerne fortæller WWW-browseren, hvad den pågældende tekstfunktion er. Browseren afgør derefter, hvordan det skal fortolkes og vises på skærmen, f.eks. 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.
Hyperspring	Et hyperspring er betegnelsen for den proces, der sker når en bruger aktiverer en henvisning i <i>hypertekstdokumentet</i> i sin WWW-browser f.eks. 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 brugerens 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.
Hypertekst	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.
Internet	Internet er en verdensomspændende sammenkobling af mange computernet. Disse computernet kan alle kommunikere med hinanden, og brugerne på de individuelle net kan komme i kontakt med brugere og ressourcer på de andre net. Den enkelte bruger behøver ikke at vide noget om nettet, men bruger blot Internets system af <i>adresser</i> , så sørger nettet for resten. Internet er udsprunget af et amerikansk, militært forskningsnet, men er i dag en blanding af et stort antal kommercielle og ikke-kommercielle computernet. Der er ingen central styring eller overordnet myndighed.
Konceptøgning	Er en teknologi, som med hensyn til præcision minder om frasesøgning. Derfor kan konceptøgning være ligeså præcis som frasesøgning.
Landedækning	De lande, hvor søgeværktøjet er dedikeret til, er den bestemte geografisk lokalitet, hvorfra informationen er tilgængelig.
Meta Tag	I forbindelse med søgesystemer kan et Meta Tag være et kort resumé af indholdet i ens dokument eller nøgleord fra dokumentet. Et Meta Tag skrives i starten af siden (i <HEAD>-delen), som er usynlig for brugeren. Laver man et resumé skal det have følgende form: <META NAME="description" CONTENT="Her skrives resuméet"> Et Meta Tag kan som sagt også være beskrivende nøgleord for dokumentet, og skrives på følgende måde: <META NAME="keywords" CONTENT="Her skrives nøgleordene">
Resultat præsentationen	Søgesystemerne præsenterer ofte en del af WWW-siden som resultat, og ved de forskellige indeks er det relevant om det medtagede WWW-side karakteriseres med en beskrivelse og/eller en form for kvalitetskarakter. En anden parameter, der kan gøre søgningen nemmere, er mange søgesystemernes relevanslister som resultaterne rangeres efter.
Sprog og grammatik	Søgningen baseret på et dansk søgeord kan mislykkes, 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 "æ, ø og å".
Udgangspunkter	Udgangspunkter (Starting Points) er et indeks, der har formålet at vise mangfoldigheden i WWW. Man skal ikke forvente at disse udgangspunkter er dækkende. Disse indeks er udarbejdet som udgangspunkt for browsing eller surfing. De har den fordel, at man hurtigt finder et WWW-sted med relevans for sin søgning.

URL	Uniform Resource Locator (URL) er betegnelsen for den unikke adresse, en given ressource har på World Wide Web. Man angiver en URL, når man ønsker, at ens browser skal vise indholdet af en given Web-side eller hente en given fil. En URL kan være en WWW-adresse (http://), en FTP-adresse (FTP://) eller f.eks. en nyhedsgruppe (news:). Henvisninger til steder på WWW gives med URL'er, f.eks. http://www.fsk.dk
Vejledere	Vejledere er oprettet og vedligeholdt af personer med viden omkring et givent meget mere specifikke emneområde, denne form for emneindeks dækker. Vejledere kan derfor være meget værdifulde i en søgningsproces. Vejlederes bliver (på engelsk) kaldt meget forskelligt som: collection, compendium, guide, index, indices, meta-page, subject guide, subject index og trailblazer pages.
World Wide Web	World Wide Web kaldes ofte blot for Web, WWW eller Spindet, og er den betegnelse, der bliver brugt om den enorme mængde af forbundne HTML-sider, der ligger på Internet. Spindmetaforen hidrører fra det faktum, at mange af dokumenterne er forbundet på kryds og på tværs med hyperspringhenvisninger. World Wide Web er den hurtigst voksende del af Internet.
WWW-side	Eller en Web-side er et enkelt HTML-dokument. Et WWW-sted består normalt af mange forbundne WWW-sider. WWW-sider kan rumme både tekst, billeder og henvisninger, foruden lyd, video eller data i andre formater, der kan kræve særlig hardware eller programmer for at kunne ses.
WWW-søgesystem	Search Engine, Søgemaskine, Indeks-database, Internet-robot, Søgerobot eller bare Søgeselement. Søgeselementer 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 søgesystemet er et andet program, hvor man kan indtaste nogle nøgleord, som programmet så søger i databasen efter.
WWW-søgeværktøj	Er en samling af forskellige metoder til at søge efter information på WWW. Disse metoder er materialiserede i forskellige værktøjer som "Udgangspunkter", "Emneindeks", "Vejledere", "Søgeselementer", "Bogmærker", "Agenter" og "Personlig Service". Det fælles formål med disse værktøjer er at effektivisere og præcisere en søgning på WWW.
WWW-sted	Eller et Web-sted (Web-site) er et begreb, der bruges om det samlede antal sider, der findes i hierarkiet under en given hjemmeside. WWW-stedet er ligesom hjemmeside et noget upræcist begreb, da siderne på et Web-sted i princippet kan være spredt over et meget stort geografisk område. Brugeren vil normalt ikke kunne mærke dette. I de fleste tilfælde er alle HTML-siderne på et WWW-sted dog samlet på den samme server, som også sagtens kan rumme flere Web-steder.

---

## Referencer

---

Forfatter (e)	Titel, Forlag/Tidsskrift, Udgivelsestidspunkt
Calafia Consulting	A Webmaster's Guide to Search Engines, 31-1-1997, <a href="http://calafia.com/webmasters/">http://calafia.com/webmasters/</a>
Fast, Michael	<i>Internationalization as a Social Construction</i> , Paper presented at the 9th IMP annual conference - Bath UK, Aalborg University, 1993
Fast, Michael	<i>Videnskabsteori og metodologi i studier af livsverden</i> , Aalborg Universitet, Upubliceret, 1996
Forskningsministeriet, Danmark	<i>Retningslinier for offentlige on-line publikationer</i> , Forskningsministeriet, juli 1996
Forskningsministeriet, Danmark	Standard for statslig elektronisk publicering - Version 18, Forskningsministeriet, januar 1997, <a href="http://www.fsk.dk/fsk/publ/online-pub/">http://www.fsk.dk/fsk/publ/online-pub/</a>
Heilesen, Simon	<i>Helt fra start på Internet og World Wide Web</i> , Borgens Forlag, 1995
Jørgensen, Martin	<i>Web-søgesystemer</i> , internetworld 23-8-1996, <a href="http://www.idg.dk/internetworld/artikler/searchtest/logik.htm">http://www.idg.dk/internetworld/artikler/searchtest/logik.htm</a>
Laursen, Jesper Vissing	<i>Reklamens magt på nettet</i> , Politiken, 17-10-1996
Northern Webs	<i>Search Engines Tutorial</i> , 26-9-1996, <a href="http://www.digital-cafe.com/~webmaster/set01.html">http://www.digital-cafe.com/~webmaster/set01.html</a>
Pfaffenberger, Bryan	<i>Web Search Strategies</i> , MIS Press, 1996
Rask, Morten og Skræm, Kenneth	International Markedsføring på World Wide Web, Aalborg Universitet, 1997, <a href="http://www.i4.auc.dk/rask/p10/front.htm">http://www.i4.auc.dk/rask/p10/front.htm</a>

---

## Stikordsregister

---



---

### A

Adresse .....	39
Afskæring af søgeord .....	21;39
Ajourføring af emneindekset/databasen .....	34
ALTA VISTA .....	24;25;26;27;28;31;32;35
AND .....	22
Antal WWW-sider .....	20
Automatisk .....	21
Avanceret søgning .....	19;21;23;30;31;32
definition .....	21

---

### B

Begrænset feltsøgning .....	21;39
Behandlingen af flere søgeord .....	21;39
Bemærkninger .....	23
Bogmærker .....	18;19;36;37;38
definition .....	18
Bredden af søgningen .....	8;14;22;28
definition .....	19
Browser .....	39
Browsing .....	10;14;19
definition .....	10

---

### C

Content .....	12
CyberCity Agent .....	25;28;30;31;32;35

---

### D

Dokumentkodningsstandard HTML .....	10
Dom@in .....	34;37
Dybden af søgningen .....	8;14;20;22
definition .....	20

---

### E

Emneindeks .....	15;19;20;23;28;29;35;39
definition .....	15
Emneindeks, decentralt .....	15
Entalsform .....	21
Excerpt .....	12
Excite .....	26;27;28;31;32;35

---

### F

Få et overblik over søgeværktøjernes kvaliteter .....	5;14
Fordele og ulemper ved de forskellige typer af søgeværktøjer .....	18;19
Forsinkelse på indekseringen .....	34
Forstå udfordringen i informationsøgning på WWW .....	5;8
Forudsætninger for WWW-søgning .....	13
Frase .....	21
Frasesøgning .....	21;39

---

### H

Headings .....	12
<b>Hitlisten</b> .....	23;24
HotBot .....	26;27;28;31;32;35
HTML .....	10;11;40
Hyperlinks .....	9;13
definition .....	9
Hyperspring .....	9;40
Hypertekst .....	9;10;40
definition .....	9
HyperText Markup Language .....	11;40

---

### I

Indeks.dk .....	29;35;37
Information af høj kvalitet .....	28
Informationssøgningsprocessen, trin i .....	5
InfoSeek .....	26;27;28;31;32;35

Internet.....	40
Internet Sleuth.....	33

**J**

Jori.....	25;28;30;32;35
Jubii.....	29;35;37

**K**

Komponenter i en typisk WWW-side.....	20;25;26;30
definition.....	12
Komponenter, andre.....	13
Koncept søgning.....	40

**L**

Landedækning.....	20;40
Lineær tekst versus hypertext.....	10
Links.....	9
Logiske operatører.....	21
Lycos.....	24;26;27;28;31;32;35

**M**

META.....	13;40
Meta Tag.....	13;40

**N**

NEAR.....	22
NOT.....	22
Nye tilgængelige informationer på WWW.....	34
Nyhedsbrev.....	35
Nyhedsside.....	34

**O**

OpenText.....	26;27;28;31;32;35
OR.....	22
Ordstamme.....	21
Overvåg nye informationer på WWW.....	5;33

**P**

Point.....	23;24
Præcis.....	21

**R**

Reference Desk, the.....	33
Resultat præsentation.....	20;40
Robot.....	16

**S**

Samtlige WWW-tilgængelige informationer.....	23;30
Scout Report, the.....	34;37
Search Engines.....	16;33
Simpel søgning.....	19;20;23;24;25;27
Situationsbestemt anvendelse af søgeværktøjerne.....	5;23
Søgesystem.....	9;16;19;20;23;24;28;35;41
danske.....	25;30
definition.....	16
generel anvendelsesprocedure.....	17
globale.....	26;27;31
Søgeværktøjer	
ajourføring.....	34
begrænsninger.....	8
fordele og ulemper ved de forskellige typer.....	19
vurderingsparametrene.....	19
Søg-i-et-systemer.....	17
Specifik information.....	23;24;30
Spider.....	16
Sprog og grammatik.....	21;40
Stamme.....	21
Standardindstillingerne.....	20;27
Starting Points.....	14
Store bogstaver.....	21
Subject Catalogs.....	15
Surfing.....	10;14

**T**

THOR.....	24;25;28;30;32;35
Title.....	12
Transformér viden til information.....	6;38

---

**U**

Udfordringen i informationsøgning på WWW .....	5;8
Udgangspunkter .....	14;19;23;40
definition .....	14
Undersøg specialiserede referencesystemer .....	5;32
Uniform Resource Locator .....	12;41
URL .....	12;41
URL-minder, the .....	37
Usynlig del af en WWW-side .....	13

---

**V**

Vejledere .....	16;19;23;28;38;41
definition .....	16
Viden, definition .....	5
Vidensudviklingsprocessen .....	5;18;32
Vurderingsparametrene for et søgeværktøj .....	19

---

**W**

Wanderer .....	16
What's New Too .....	34;37
What's New? .....	36;37
World Wide Web .....	41
Worm .....	16
WWWs mangfoldighed .....	19;23;24;34
WWW-side .....	41
WWW-Søgesystem .....	41
WWW-søgeværktøj .....	41
WWW-sted .....	41

---

**Y**

YAHOO! .....	29;35;37
--------------	----------

## PAPERS IN STUDY MATERIALS SERIES

- No. 1: *Poul Weihrauch (1993)* International Dimensions of Branding and Promotion.
- No. 2: *Thomas S.M.T. Nielsen (1993)* Contrary Market Research: The Japanese Way of Conducting Market Research.
- No. 3: *Stig H. Christiansen (1993)* Towards a New Approach to Strategic Industry Analysis - A Preliminary Discussion.
- No. 4: *Michael Fast (1993)*: Internationalisering og organisationen som en social konstruktion.
- No. 5: *Olav Jull Sørensen (1996)*: Approaches to Analyzing International Industries.
- No. 6: *Olav Jull Sørensen (1996)*: Competitive Advantage of Firms and Nations. The Porter Approach.
- No. 7: *Olav Jull Sørensen (1996)*: The International Dimension of Industrial Economics.
- No. 8: *Olav Jull Sørensen (1996)*: The Network Theory. An Introduction to its Conceptual World.
- No. 9: *Olav Jull Sørensen (1996)*: Alternative Ways of Choosing International Market Entry Modes.
- No. 10: *Morten Rask og Kenneth Skærm (1997)*: Informationssøgning på World Wide Web.

---

Copies of publications in this series can be purchased from:  
**Department of Development and Planning, Fibigerstraede 11, 9220 Aalborg, Denmark,**  
**att.: Dorte Johansson, Phone: +45 96 35 83 52.**