

LUNDS UNIVERSITET  
*Informatik*  
Ole Römers väg 6, 223 63 LUND  
Hämtställe 38  
AÖlerup / Ö Steen

06-01-30

# Effektiv informationsökning på Internet utifrån användarnas perspektiv

Magisteruppsats, 10 poäng, i informatik

*Framlagd:* januari 2006

*Författare:* Arina Zech

*Handledare:* Konrad Tollmar

# Förord

Jag vill tacka alla de som hjälpt mig genomföra detta examensarbete.  
Lund, november 2005  
Arina Zech

LUNDS UNIVERSITET  
*Informatik*  
Ole Römers väg 6, 223 63 LUND  
Hämtställe 38  
AÖlerup / O Steen

06-01-30

LUNDS UNIVERSITET  
Informatik

## Effektiv informationsökning på Internet utifrån användarnas perspektiv

© Arina Zech

Magisteruppsats framlagd januari 2006  
Omfång: 80 sidor, avrundat till närmaste tiotal, inkl. bilagor o. dyl.  
Handledare: Konrad Tollmar

### **Resume**

Storleken på informationsmängderna i vad som finns tillgänglig på Internet idag gör att söktjänster närmast är en nödvändighet för informationsökning. Ibland kan det upplevas att det är svårt att hitta begärd information hos söktjänster utan att få fram en mängd oönskade sidor, så att man måste hitta relevant information bland dem och filtrera bort irrelevant information. Ett problem för många söktjänster är hur de ska kunna ta hänsyn till användarnas individuella informationsbehov och därigenom skapa förutsättningar för effektivare sökprocesser.

Denna magisteruppsats handlar om användarnas upplevelser av informationssökning på Internet. Syftet med uppsatsen är att undersöka den problematik som finns när man söker information med hjälp av söktjänster på webben, användarnas syn på olika söktjänster och att även få fram förbättringsförslag.

Genom litteraturstudier, enkäter och intervjuer med Internetanvändare har jag försökt besvara hur söktjänsterna på webben ser ut, vilka söktjänster som oftast väljs, vilka problem användarna upplevt med informationsökning och vilka möjligheter de önskar/saknar.

Studien visade att Google var den mest använda söktjänsten. Slutligen har jag fått fram förbättringsförslag för söktjänster som skulle kunna underlätta sökprocessen för användarna. Uppsatsen avslutas med förslag till vidare forskning.

### **Nyckelord**

Informationssökning, Internet, webb, söktjänst, sökverktyg, sökmotor, sökprocess

# Innehållsförteckning

Förord .....	i
Resume .....	iii
Innehållsförteckning .....	iv
1 Inledning .....	1
2 Syfte .....	2
2.1 Frågeställningar .....	2
2.2 Uppsatsens målgrupp .....	2
2.3 Varför ämnet är intressant för mig som systemvetare: .....	2
3 Metod .....	3
3.1 Ansats .....	3
3.2 Metodval .....	4
3.3 Valt angreppssätt .....	4
3.4 Val av deltagare .....	6
3.5 Avgränsningar .....	6
3.6 Etik .....	7
3.7 Validitet .....	7
3.8 Reliabilitet .....	7
3.9 Disposition .....	8
4 Teori .....	9
4.1 Information på Internet/webb .....	9
4.2 Informationssökning med sökmotorer på Internet .....	9
4.2.1 Typer av informationssökning .....	10
4.2.2 Sökprocess .....	10
4.2.3 Modeller för informationssökning .....	14
4.2.3.1 Ellis - 'beteende-modell' för informationssökning .....	14
4.2.3.2 ISP-modell av Kuhlthau .....	16
4.2.3.3 Limbergs A, B och C kategorier .....	17
4.3 Information Retrieval (IR) - modeller .....	19
4.4 Olika typer av söktekniker .....	20
4.4.1 Sekventiell strängsökning .....	20
4.4.2 Strängsökning med index .....	20
4.4.3 Sökning av information som är inte textbaserad .....	21
4.4.4 Semantisk webbt teknik .....	22
4.5 Söktjänster på Internet .....	24
4.5.1 Generellt om söktjänster .....	25
4.5.2 Ranking .....	27
4.5.3 Databaser .....	28
4.5.4 Användbarhet .....	29
4.5.5 Olika typer av söktjänster .....	29
4.5.5.1 Sökmotorer .....	29
4.5.5.2 Metasöktjänster .....	30
4.5.5.3 Ämneskataloger eller ämnesportaler .....	30
4.5.5.4 Jämförande shopping-söktjänster (comparison shopping engines) .....	31
4.5.5.5 Söktjänster som svarar på frågor (QA – Question Answering Systems) .....	31
4.5.5.6 Osynlig webb .....	31

4.5.6 Hur man använder söktjänster.....	32
4.5.7 Några av största söktjänster.....	33
4.5.7.1 Alta Vista ( <a href="http://www.altavista.com">http://www.altavista.com</a> ) .....	33
4.5.7.2 Yahoo! ( <a href="http://www.yahoo.com/">http://www.yahoo.com/</a> ) .....	34
4.5.7.3 Google ( <a href="http://www.google.com">http://www.google.com</a> ) .....	35
4.5.7.3 Google ( <a href="http://www.google.com">http://www.google.com</a> ) .....	35
4.5.7.4 Ask Jeeves ( <a href="http://www.ask.com/">http://www.ask.com/</a> ).....	36
4.5.7.5 Open Directory ( <a href="http://dmoz.org/">http://dmoz.org/</a> ) .....	37
4.5.7.6 Wondir ( <a href="http://www.wondir.com">http://www.wondir.com</a> ).....	38
4.7 Sökprocess - strategisk informationsökning.....	38
5 Undersökning och resultat.....	41
5.1 Observation .....	41
5.2 Enkäter .....	42
5.3 Intervju .....	44
6 Diskussion .....	46
7 Slutsatser .....	46
7.1 Svar på frågeställningar.....	46
7.2 Avslutande sammanfattning.....	48
7.3 Förslag på vidare forskning.....	49
8 Terminologi.....	50
8.1 Förkortningar.....	50
9 Litteratur & övrigt material .....	52
10 Bilagor.....	56
Bilaga 1 .....	56
Bilaga 2 .....	58
Bilaga 3 .....	61
Bilaga 4 .....	64
Bilaga 5 .....	65
Bilaga 6 .....	68
Bilaga 7 .....	69
Bilaga 8 .....	69
Bilaga 9 .....	70
Bilaga 10 .....	71
Bilaga 11 .....	72

## Förteckning över tabeller och figurer

<b>Figur 1.</b> The conceptual framework - the domain of the method schema. Ur Pharom fl (2002) .....	11
<b>Figur 2.</b> Kuhlthaus ISP-modell (Kuhlthau 1993, Figure 3-1 s 43, genom Båge m fl, 2003).....	16
<b>Figur 3.</b> Systemorganisation ur Funkhouser m fl (2002) .....	20
<b>Figur 4.</b> Söktjänst för 3D modeller ur Funkhouser m fl (2002) .....	21
<b>Figur 5.</b> Söktjänsters relationer .....	23
<b>Figur 6.</b> Diagram över sökningar av US Internetanvändare under juli 2005 (Sullivan, 2005).....	24
<b>Figur 7.</b> AltaVistas enkel sökning .....	32
<b>Figur 8.</b> AltaVistas avancerad sökning .....	32
<b>Figur 9.</b> Yahoo! enkel sökning .....	33
<b>Figur 10.</b> Yahoo! avancerad sökning .....	33
<b>Figur 11.</b> Google enkel sökning .....	34
<b>Figur 12.</b> Google avancerad sökning .....	34
<b>Figur 13.</b> Ask Jeeves enkel sökning .....	35
<b>Figur 14.</b> Ask Jeeves avancerad sökning .....	35
<b>Figur 15.</b> Open Directory Project enkel sökning .....	36
<b>Figur 16.</b> Open Directory Project avancerad sökning .....	36
<b>Figur 17.</b> Wondirs sökinterface .....	37

# 1 Inledning

Informationen är en av de viktigaste resurserna i vår tid. Mängden av information ökar hela tiden. Internet ger möjlighet att lätt komma åt information rent tekniskt men samtidigt har det blivit svårare för vanliga användare att hitta det man behöver bland all information som finns i detta nätverk. Grundare till Google, Brin och Page (1998) säger att antalet webbdokument som är tillgängliga för sökning ökar, men inte användarnas möjligheter att gå igenom dem.

” Att finna det man söker på Internet kan liknas vid att leta efter en nål i en höstack, eller kanske snarare vid att vaska efter guld. För att lyckas finna guld krävs tålamod och kunskap men naturligtvis också att man har bra och ändamålsenliga verktyg. På Internet krävs framför allt att man har tillgång till bra sökverktyg för att kunna vaska fram det sökta, som alltför ofta ligger gömt bland en massa virtuella stenar och morän i denna virvlande flod av informationen.” Bjurbäck (1998)

Med den mängd information som finns tillgänglig är söktjänster en nödvändighet. Den perfekta söktjänsten skall kunna begripa användarens informationsbehov och returnera mest relevanta dokument till även vagt formulerade frågor. Det finns många kommersiella söktjänster, alla med egna indexerings- och sökfunktioner, t ex vissa sökmotorer kan leta efter söktermer i hela dokument, andra söker endast genom de första tio raderna, medan en del söker genom de termer som upphovsmannen medvetet har valt och kodat in för att beskriva sitt dokument. Att välja rätt söktjänst är ett viktigt moment i informationssökning på webben därför att olika sökmotorer återfinner olika dokument trots att man har använt samma sökord. Det finns ett stort antal av dem söktjänster med vissa likheter och skillnader: WWW sökmotorer, webbkataloger, metasöktjänster, specialiserade söktjänster, frågetjänster och faktasöktjänster. (Hock, 2001). Några är bättre på att söka viss information än andra och vid en sökning kan det produceras så många resultatsidor att man inte kan leta igenom alla, eller så få resultatsidor att det man söker inte kommer med. Ingen söktjänst känner till allt som finns på Internet. Det är också väldigt viktigt med en bra designad UI för sökningen.

De populära söktjänsterna får miljoner sökfrågor dagligen och användare önskar få ett snabbt och relevant svar. I detta arbete studerade jag hur man kan söka information på Internet effektivt, samtidigt om hur vanliga användarna gör detta och vilka önskemål de har angående söktjänst, vilket med andra ord är hur en perfekt söktjänst borde kunna vara enligt användarna.

## **2 Syfte**

Syftet var att undersöka hur vanliga användare hittar information på Internet samt hur effektivt användarna upplever sökprocessen, om användarna saknar någonting i dem moderna söktjänsterna. Utifrån det framtofs eventuella rekommendationer för framtida utveckling av söktjänster samt besvarades frågeställningar.

### **2.1 Frågeställningar:**

- 1) Hur är tjänster för informationsökning på Internet designade (uppbyggda) och hur fungerar de?
- 2) Vilken erfarenhet har användarna av sökprocessen och vad är deras önskemål dvs hur borde en perfekt söktjänst vara?
- 3) Vad vet användarna om olika söktjänster och vad är användarnas syn på dem?

### **2.2 Uppsatsens målgrupp**

I första hand är det studenter i informatik/systemvetenskap, men även alla andra på akademisk nivå som har allmänt eller professionellt intresse för informationssökning på Internet och som anser att det idag är alldeles för svårt att hitta exakt det de vill på Internet. Det har förutsetts att läsaren har baskunskaper om vad Internet och www är och hur man använder det.

### **2.3 Varför ämnet är intressant för mig som systemvetare:**

Frågan är intressant eftersom söktjänster på Internet är den typ av informationssystem som används så ofta av så många. Alla människor vill kunna använda Internet på ett meningsfullt och effektivt sätt. Söktjänsterna utvecklas och förbättras hela tiden, berör många Internetanvändare och IT-professionella som både söker information eller utvecklar teknik för informationssökning. Som andra nätbaserade applikationer utvecklas söktjänsterna av tekniker som kan få idéer till vidareutveckling.



## 3 Metod

Enligt Göran Wallén (1996) behöver vetenskapligt problem inte betyda att något är direkt problematiskt utan att det är någonting man söker kunskap om.

### 3.1 Ansats

Vetenskapsteori behandlar hur vetenskaplig kunskap bildas, hur man går tillväga för att prova den och kunskapens roll i samhället (Wallén, 1996). Två vetenskapliga synsätt som kommer att användas i denna studie är positivism och hermeneutik.

Hermeneutik kan översättas som tolkningslära. Den handlar framförallt om tolkning av innebörden i bland annat texter, symboler, handlingar och upplevelser. Det är viktigt att man tar hänsyn till den kontext det tolkade materialet har uppkommit i. Tolkning innebär att man visar innebörder eller sammanhang som "ligger bakom" det som vid första anblicken kanske verkar vara något helt annat. (Wallén, 1996)

Inom det positivistiska synsättet anser man att det som inte är empiriskt prövbart, som till exempel känslor och värderingar, inte kan räknas till vetenskaplig kunskap. Förklaringar anges i form av orsak - verkan. Positivism baseras på stark tilltro till vetenskaplig rationalitet och att kunskap ska kunna testas empiriskt. (Wallén, 1996) Jag utgår från ett hermeneutiskt synsätt eftersom jag gör tolkningar av verkligheten. Resultatet bygger till stor del på människors erfarenheter och upplevelser, insamlade via en enkätundersökning och ett mindre antal intervjuer.

Wallén (1996) skiljer mellan följande ansatser: explorativa, deskriptiva, förklarande studier och normativa studier. När det finns luckor i kunskapen om det aktuella ämnet kommer undersökningen att vara explorativ. Syftet med en sådan undersökning är att inhämta så mycket ny kunskap som möjligt om ett bestämt problemområde: när, var, hur, relevanta variabler och begrepp. När det redan finns en viss mängd kunskap inom området, och när man börjar systematisera denna i former av modeller, är undersökningen deskriptiv. När man studerar "varför-frågor" är det förklarande studier. När undersökningen resulteras i norm eller handlingsförslag är det normativa studier. Denna undersökning var av explorativ karaktär med syftet att inhämta så mycket ny kunskap som möjligt om valt problemområde.

Val av forskningsansats beror på uppfattningar och relationer mellan teori och empiri. Det finns två olika övergripande metodansatser: induktion och deduktion. Induktion utgår från en datainsamling och försöker sedan dra generella och teoretiska slutsatser. Nackdelen med denna ansats är att resultatet bygger på ett fåtal enskilda fall det vill säga bara en del av verkligheten. Deduktiv ansats utgår från en generell teori och förklarar ett visst enskilt fall. Nackdelen med denna ansats är att den teorin ibland inte stämmer överens med den verklighet som forskare studerar. Jag valde att använda mig av en kombination av induktion och deduktion. Studien började med en teoristudie och vissa frågeställningar fastställdes utifrån teoristudien. När jag har gjort den empiriska studien jämförde jag resultatet med den teori som jag tidigare har studerat.

## **3.2 Metodval**

Metoderna kan vara kvalitativa, kvantitativa eller kombination av dem, och kan utföras på fältet eller i en laboriemiljö. Kvalitativa metoder har målet att fastställa fenomen, generellt beskriva processer och miljön. Kvalitativa metoder handlar om att karakterisera och är användbara för sådant som vagt och mångtydligt, när man är intresserad av att hitta grundläggande eller speciella egenskaper i en viss miljö. Kvalitativa metoder beskriver vad något är och inte hur ofta något är. Kvalitativa metoder är bra att använda när det är viktigt att förstå människors tolkningar av fenomen eller händelser, för att skapa kunskap om mänskliga beteendet i olika situationer (Bryman, 1997). Wallén (1996) rekommenderar kvalitativa undersökningsformer för att kunna studera komplexa problem som är svåra att mäta med något mätinstrument.

När man tittar efter kvantitet eller frekvensen av något är en kvantitativ metod mer lämplig att använda. Analys av kvalitativt insamlade data kan även ge statistik. När man analyserar insamlade data är det svårt att undvika uttryck som ”de flesta användarna upplever...” och så vidare. (Wallén, 1996)

Denna undersökning var en kombination av kvalitativ och kvantitativ metod främst för att, förutom ett resultat ville jag få en förståelse för helheten i valt ämne och fånga in respondenternas upplevelser gällande informationssökning via Internet. Data som producerades var av både kvalitativ och kvantitativ art. Kvantitativa data kan enligt Miles och Huberman (1994) användas för att fördjupa analysen och ge större detaljrikedom. Begränsningen i detta var att jag bara skulle se hur effektiva, enligt användarna, dagens verktyg är, men tanken med det här arbetet var ju att mer teoretiskt försöka se möjligheter.

Vid användandet av en kvantitativ metod utmärks ofta resultatet av att det kan presenteras i siffror (statistik). Min problemformulering var inte utformad på ett sådant sätt att ett tillfredställande svar kan ges i siffror. Jag var inte heller intresserad av enbart ett sådant mätbart resultat. Dessutom kan det vara mer tidsödande att göra en enbart kvantitativ undersökning på detta område. Undersökningsmetoden avgjorde vilken information jag fick fram i denna uppsats. Informationen som jag eftersträvade att få fram skulle vara så heltäckande och förklarande som möjligt, vilket också talade för en kombination av kvalitativ och kvantitativ undersökning.

## **3.3 Valt angreppssätt**

- Arbetssteg 1 - förarbete genom litteraturstudie
- Arbetssteg 2 - förarbete via observation av två användare
- Arbetssteg 3 – datainsamling via enkäter
- Arbetssteg 4 – datainsamling via intervjuer
- Arbetssteg 5 – bearbetning av data
- Arbetssteg 6 – resultat utvärdering

Arbetet med uppsatsen var indelat i sex steg: förarbete genom litteraturstudie, förarbete via observation av 2 användare, datainsamling via enkäter, datainsamling via intervjuer, bearbetning av data, resultatutvärdering.

Förarbetet inleddes med en **litteraturstudie**. Syftet med litteraturstudien var att få en uppfattning om området i allmänhet och en förståelse för de begrepp, teorier och formuleringar som finns inom området och utifrån dem göra egna tolkningar och dra slutsatser. Litteraturstudierna bestod delvis av litteratur som rekommenderades av min handledare, delvis av artiklar och uppsatser som jag själv hittade på Internet. Den valda litteraturen belyste aspekter om informationssökningsprocessen på Internet, söktjänster och dess jämförelse, rekommendationer av val av verktyg. Det fanns väldigt mycket inom det området vilket gjorde svårt att följa forskningen och att skaffa sig en överblick. Det som var nackdelen med litteraturstudien är att man lätt kan bli styrd av andra författares attityder. Tillförlitligheten av olika typer av litteratur kan variera. Detta gäller särskilt det material som finns på Internet, där det ofta är svårt att veta varifrån informationen kommer från början.

Den här rapporten innehåller mycket information som hämtas från Internet på grund av att den mest väsentliga information på det området finns tillgänglig där. Böcker blir lätt inaktuella innan de kommer ut i tryck, däremot aktuell information på Internet är lättillgänglig och snabb att tillgå. Jag tog ställning för varje informationskälla på Internet till hur tillförlitlig den var, vem som var källa till materialet och om det fanns risk för att källan kunde vara opålitlig. För att bedöma vad som är tillförlitlig information på Internet tittade jag först på upphovet som till exempel informationen från universitet, forskningsorganisation, välansedda tidskrifter och så vidare bedömde jag som objektiv och korrekt. Det fanns också en risk att informationskällor på Internet innehåller partisk information. Det är en naturlig följd av att alla kan skriva om vad som helst där men ändå få det att se helt seriöst ut. En bok eller tidningsartikel är tillförlitligare ur den synpunkten att inte vem som helst kan publicera material inom ett visst ämne i dessa medier.

Förarbetet fortsatte med **observation** av två användare som söker information på Internet. Vill forskaren veta något om verkligheten så måste han observera den (Backman, 1998). Fördelen med observationer, enligt Preece (1994) är att de ger forskaren en direkt tillgång till observerade processer på ett sätt som intervjuer bara kan ge en indirekt kunskap om. Människor vill eller kan inte alltid beskriva situation, problem, krav, önskemål osv. Observationer är både tid- och resurskrävande och som observatör måste forskaren arbeta inom ett ganska avgränsat område. Det är viktigt att tänka på, vid analys av observationer, att forskareffekter kan uppstå (Preece, 1994). Människor som vet om att de blir observerade kan ibland bete sig annorlunda än om de inte vet om att de blir observerade. Även forskaren som person kan påverka observationen.

Observationsmetoden som användes i detta arbete var öppen då deltagarna var medvetna om att de observerades. Öppen observation kräver att man måste bli accepterad. Jag använde ostrukturerad (ad hoc) observation (där jag försökte uppfatta vad som hände) med ”think aloud method”. ”Think aloud method” är när användare under observationer berättar om sina tankar under genomförandet. Jag ställde frågor för att ta reda på användarens tolkning och uppfattning av händelser som jag observerade. Observationerna spelades in på video (Preece, 1994).

Observationer och litteraturstudier skapade grunden för enkätfrågor

Empirisk undersökning var viktig dels eftersom saker i praktiken kan vara annorlunda än man trodde. En av de viktigaste metoder jag använde för att samla in data och skaffa mig en övergripande bild av användarnas synpunkter och attityder om informationssökning på Internet var en **enkätundersökning**. Ett antal deltagare svarade på enkätfrågor om vilka

sökverktyg som används, på vilket sätt, hur hela sökprocessen upplevs och vilka önskemål användare har.

Sedan utförde jag **intervjuer**, som har fördelen att kunna ge en bredare bas av respondenter från olika sociala miljöer (Bryman, 1997). Intervjuer använde jag för att lyfta fram erfarenheter, uppfattningar och upplevelser som varje respondent har om informationssökning via Internet. Jag genomförde semistrukturerade intervjuer. Att arbeta med semistrukturerade intervjuer valdes eftersom det anses kunna ge en mer avslappnad intervjuform som till stor del kan liknas vid ett samtal (Bryman, 1997). Detta gav mig flexibilitet. När nya frågor dök upp under en intervju kunde de ställas utan att orsaka problem. Detsamma gällde om frågor behövde ändras på något sätt. Denna typ av intervju gav även möjligheter för samtalen att gå sin egen väg och komma fram till saker som inte hade framkommit vid en strukturerad intervju. För att lyckas med en ostrukturerad intervju behöver forskare ha en större erfarenhet som intervjuare (Andersson, 2001), och eftersom jag hade inte en sådan praktisk erfarenhet valde jag att inte arbeta på detta sätt.

### **3.4 Val av deltagare**

Till observation tillfrågade jag två personer som båda hade vana att söka information på Internet men inte arbetade med t ex utveckling av söktjänster. Jag delade ut 30 enkäter genom att först har tillfrågat 10 personer ifrån min vankrets och bitt dem att ta med 2 extra enkätexemplar till deras bekanta. På sådant sätt fick jag större variation av respondenter på enkäter än om jag skulle fråga endast folk som jag känner själv samtidigt som jag fick människor att verkligen svara på enkäter. Respondenterna, som fyllde i enkätformulär, var från olika åldersgrupper och yrkeskategorier. Gemensamt för alla de är att de har vana i användning av Internet och de söker information minst någon gång per vecka. Tre av de 30 deltagare har även intervjuats. Se bilaga 3.

### **3.5 Avgränsningar**

Fokus i undersökningen låg på verktyg som finns på www. I andra delar av nätet utöver www (t ex Gopher, ftp, news) söker vissa verktyg efter t ex filer eller diskussioner. Dessa utelämnade jag i undersökningen. I detta arbete låg tyngdpunkten på textbaserad informationssökning på Internet eftersom det användes mest nu och jag var mest intresserad av det. Ljud- eller bildbaserad informationssökning undersökte jag i mindre utsträckning här.

I uppsatsen behandlade jag användningen av Internets söktjänster. En söktjänst kan utvärderas från olika perspektiv, bland annat tekniska och användarnas perspektiv. Eftersom jag valde att studera här användarnas perspektiv gick jag in på områden som bara begränsat berör de tekniska delarna.

### **3.6 Etik**

Det är viktigt att fundera över etiska frågor i ett forskningsarbete. Ett viktigt moment i forskningen är att inte skada eller utlämna en försöksperson. Forskaren ansvarar för inriktningen av sin forskning och tänkbara följder av resultaten. Undersökningen som jag gjorde kommer inte att leda till skada för någon särskild individ. Däremot uppkom en viss kritik mot söktjänster på Internet.

Syftet med undersökningen förmedlades tydligt till deltagare och detta sätt snarare stärkte än försvagade möjligheterna att lyckas med undersökningen. På detta vis kunde en del av misstänksamheten som kunde finnas mot forskare undermineras och gav respondenterna trygghet genom förståelse om att det endast var processen som studerades, inte de individerna.

Forskare har en etisk skyldighet att presentera korrekt information och inte manipulera data för att resultaten som genereras skall bli de som avsetts att generera, eller att dölja känslig information som på något vis är viktigt för studien. (Bryman, 1997)

### **3.7 Validitet**

Validitet betyder giltighet och används för att bedöma om en undersökning mäter det som undersökningen avser att mäta. (Miles m fl, 1994) Jag bedömde att validiteten av mina slutsatser var relativt hög, trots ett mindre antal undersökta personer, genom noggranna urval av respondenter och analysinstrument.

Validitet i de instrument som använts i min kvalitativa forskning betyder bland annat granskning av forskarens färdigheter. Exempel på vad som gör en bra forskare kan vara att hon är bekant med fenomenet, har en multidisciplinär ansats och bra undersökningsfärdigheter (Miles m fl, 1994). För att stärka min validitet som forskare i intervjuerna förberedde jag mig noga genom att läsa litteratur och skapa intervjuguiden med relevanta intervjufrågor. Det är också viktigt att vara lyhörd för sina egna antaganden och samtidigt ha förståelse för deltagarens antaganden och förutfattade meningar eftersom dessa kan påverka de data som samlas in. (Miles m fl, 1994)

Respondentens validitet tryggas genom att jag efter intervjun kontrollerade resultatet med dem som undersöktes. Nackdelen med detta kan vara att de undersökta personerna kunde censurera bort information som var känslig för dem. I mitt arbete valde jag att använda detta sätt eftersom risken för att stöta på känslig information som respondenterna skulle vilja censurera bedömdes som relativt liten. Nyttan av att på ett bra sätt kunna bli försäkrad om att insamlad empiri var korrekt bedömdes vara tillräckligt stor för att motivera detta beslut.

### **3.8 Reliabilitet**

Med reliabilitet menas att man som forskare skall bedöma hur pålitlig den undersökning man genomfört är. Har undersökningen en hög reliabilitet betyder det att undersökningar kan

genomförs upprepade gånger på motsvarande respondenter med ungefär identiska resultat. Med detta brukar menas om undersökningen har utförts på rätt sätt, om det ställts rätt frågor, använts rätt metoder och undersökts rätt personer så att samma resultat skulle ha genererats om undersökningen skulle ha upprepats. (Miles m fl, 1994) I denna studies kontext, om man låter samma respondenter svara på samma frågor, kunde resultatet bli ganska likt det som man fick första gången det vill säga reliabiliteten kunde bedömas som högt.

Att gå tillbaka till respondenterna för att kontrollera resultat användes även för att stärka reliabiliteten. Respondenterna fick chansen att bedöma den information som undersökningen resulterat i.

### **3.9 Disposition**

Uppsatsen är indelad i fyra block: bakgrund, teori, undersökning och slutanalys, fördelade på nio kapitel.

Kapitel 1 - är inledningskapitlet.

Kapitel 2 - här definieras uppsatsens syfte och frågeställningar.

Kapitel 3 - är metodkapitel. Här presenteras det vetenskapliga synsätt och den metod som används.

Kapitel 4 - innehåller teoridel. I detta avsnitt behandlas dels teoretiska resonemang om informationssökning med bl. a informationssökningsprocessen och sökmodeller och dels olika söktekniker och söktjänster.

Kapitel 5 - här jag redogör för mina undersökningar. Först kommer en genomgång av användareobservationer, sedan resultat av enkätundersökningen och slutligen resultat av de intervjuer jag genomfört med användarna.

Kapitel 6 - som avslutning analyseras resultatet av undersökningarna.

Kapitel 7 – uppsatsen summeras i form av en slutsats.

Kapitel 8 – här förklaras termer och förkortningar.

Kapitel 9 - är referenskapitel.

Kapitel 10 - innehåller bilagor med bland annat enkätformulär, sammanställning av inkomna resultaten, intervjuguide och intervjusammanställningar/anteckningar.

## 4 Teori

### 4.1 Information på Internet/webb

Internet och webben kan beskrivas som en fulltextdatabas bestående av flera miljarder dokument där strukturen på informationen och användningen av hyperlänkfunktionen varierar beroende på webbmastern. Det finns ingen formell organisation eller central resurs för informationssökningens skull. Det skapar problem att sortera fram den information som användare behöver bland en mängd av icke-relevant information.

Informationsbitar skiljer sig väldigt mycket i storlek, från hela innehållet i servrar till korta textremsor. Det kan vara svårt att se vad det indexerade dokumentet egentligen innehåller för typ av information (speciellt då det handlar om bilder och liknande). Strukturen på webbdokumenterna är långt ifrån enhetlig (Koch, 1996). Informationen kan vara falsk, ej uppdaterad/inaktuell, felskriven/felstavad osv. (Baeza-Yates m fl, 1999)

För Gopher, WAIS och filarkiv skapas sökfunktioner genom att serveransvarig indexerar dess innehåll och beskrivningar samlas i en databas på en central server. Men för www är vanligaste sättet att navigera är ”surfa”, klicka sig fram genom länkar. I Internet-trafik protokoll finns inga inbyggda funktioner för beskrivning av resursernas innehåll eller stöd för sökning och därför finns ingen central möjlighet att göra all information sökbar. (Hock, 2001)

Skillnaden mellan Internet och bibliografiska databaser är att det inte finns några krav på att dokument på Internet ska gå att återfinna genom användningen av en standardiserad eller kontrollerad vokabulär. Skaparna av HTML-dokumenterna kan också ställa till en del problem med t ex felaktig, okontrollerad HTML-kodning och otillräcklig innehållsbeskrivning, terminologiska fel, felaktiga formuleringar av titlar och rubriker o s v. Detta gör att även informationsspecialister kan ha svårigheter att leta rätt på specifika dokument eller viss information. Webbens kommersiella potential gör att mindre nogräknade webbmästare förser sina webbplatser med irrelevanta söktermer som ett sätt att öka antalet besökare. Det innebär att sökaren kan få ett stort antal träffar som inte har någonting att göra med ämnet han/hon söker om. (Limberg m fl, 2002) Detta s.k. **webb-spam** som enligt Langville m fl (2005) inkluderar all teknik som används för att manipulera relevansen av dokumentet. Webb spammers utvecklar olika scheman för att lura söktjänster att lista deras webbsidor nära top av sökresultat. Internets mix av olika typer av information (listor, menyer, hela böcker, mjukvara, ljud, video, bilder o s v) bidrar till svårigheter med sökningar och indexeringar.

### 4.2 Informationssökning med sökmotorer på Internet

Glover m fl (2000), beskriver processen av användning av sökmotorer på följande sätt:  
*”a user will start with an information need, or concern. This concern will be formalised into a query, and is then submitted to the search engine. Once the engine receives the query, it will construct a list of web pages that it has deemed are matched to the query, and this list will then be sorted into some kind of meaningful order with each matching page summarised, and re-presented to the user. Once the user receives this list back, they will then select which documents they consider meet their information need, hopefully satisfying it.”*

## 4.2.1 Typer av informationssökning

Enligt Ingwersen och Wormell (1989) finns det tre huvudsakliga typer av informationsproblem eller behov som förekommer inom informationssökning: verifierande frågor, fokuserade ämnesfrågor och oklara ämnesfrågor. Författarna talar om etiketteffekten som innebär att användaren anger sitt informationsbehov med ett eller ett par ord som uttrycker det övergripande ämnet utan att specificera närmare i vad eller vilka aspekter inom detta ämne som man är intresserad av.

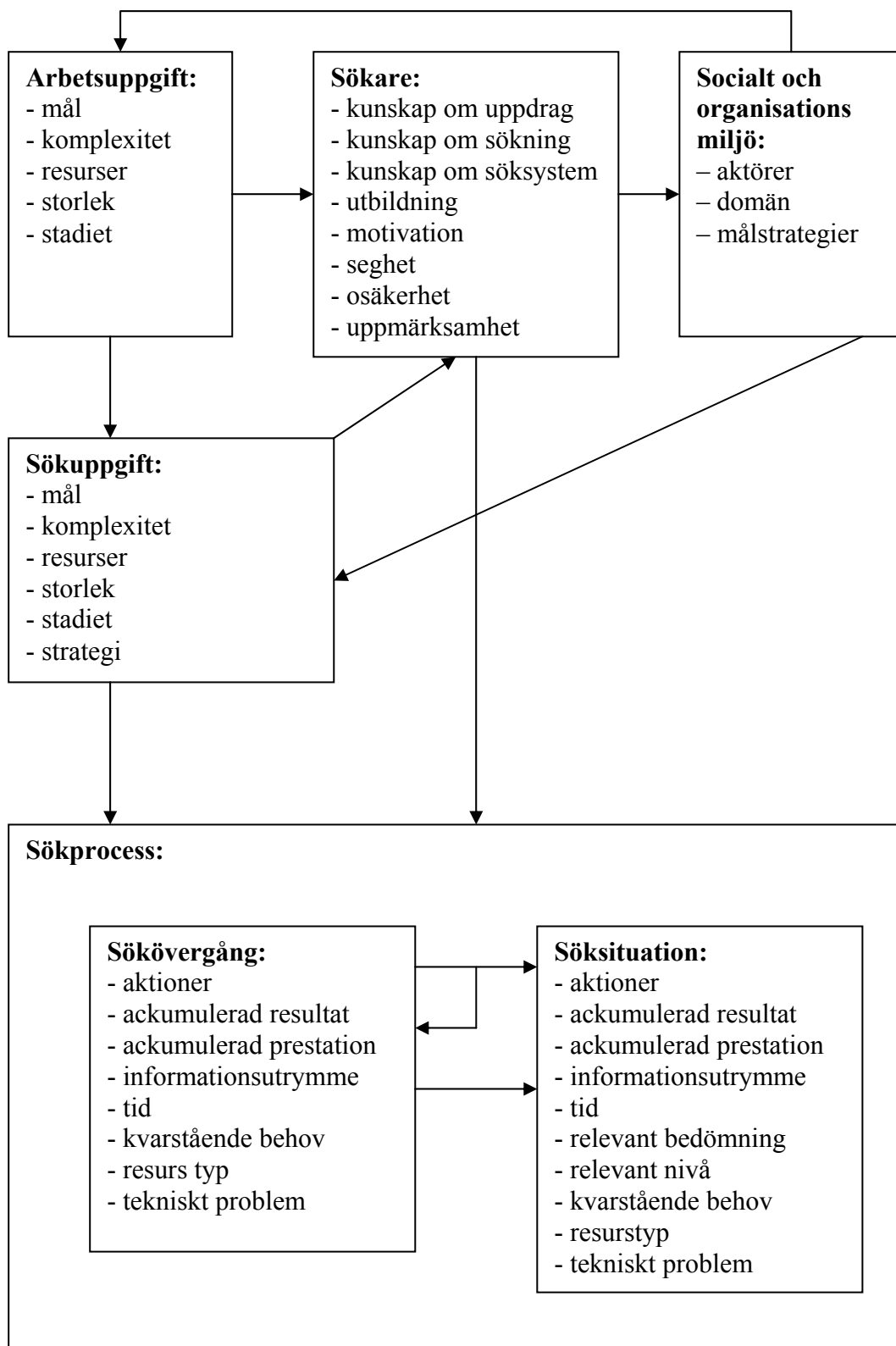
Limberg m fl (2002) beskriver informationssökning som att identifiera ett informationsbehov, att kunna välja sökstrategi, att kombinera söktermer eller välja ämnesord, att formulera söksträngar, att bedöma relevans bland en stor mängd källor, att kunna välja och värdera informationen, att analysera och sammanställa den, att kunna presentera vad man kommit fram till och att kunna utvärdera arbetet.

Enligt Aguilar (1967) finns det 4 kategorier av sökning utifrån användarens motivation: undirected, conditional, informal, formal. Undirected – när användaren inte har några specifika informationsmål och/eller behov. Det är inte riktig sökning utan snarare webbsurfing. Conditional – när användaren har specifika ämnen eller informationstyper i tanke. Informal – när användare söker efter kunskap och förståelse av någon fråga/ämne men utan något tydligt struktur. Formal – när användaren söker efter information på ett strukturerat sätt. Vissa sökningar kan placeras i mer än en kategori medan andra kan byta kategori under söktiden.

## 4.2.2 Sökprocess

Pharom fl (2002) kartlägger situationen kring sökprocessen med fem huvudkategorier: arbetsuppgift, sökare, social och organisationsmiljö, sökuppgift, sökprocess som inkluderar två underkategorier - söksituation och sökövergång. Figur 1 visar kategorier med attributen och relationer mellan dem.





**Figur 1** The conceptual framework - the domain of the method schema. Ur Pharo m fl (2002).

Originalbild finns i bilaga 6

Arbetsuppgift definieras som en serie av aktioner som genomförs för att uppnå ett visst mål som oftast är utanför informationssökning kontext. Sökuppgift är en serie av informationssökningar som genomförs för att hitta mer eller mindre specificerad information. Arbetsuppgift och sökuppgift kan ibland betyda samma process. Kategori sökprocess inkluderar två underkategorier: söksituation och sökövergång. Sökprocessen karakteriseras att sökaren samlar in resultat och väljer till (eller väljer bort) information utifrån dem ackumulerade resultat. Relationen mellan sökuppgift och sökprocess kan vara att man utifrån en sökuppgift planerar t ex vilka resurser man ska använda och vilka aktioner man ska utföra för att uppnå sökuppgiftens mål osv. Relation mellan sökare och sökprocess kan handla om många olika företeelser som t ex sökarens ouppmärksamhet kan hindra honom ifrån att identifiera relevanta resurser eller leda till att göra onödiga aktioner osv. Sökprocessen är dynamisk och sökresultat kan leda till förändringar i arbetsuppgiften. Delarna av sökprocessen (sökövergång och söksituation) har liknande attributen och står i nära relationer till varandra. Sökprocessen i sig är komplex och kan påverkas av många olika faktorer som den övergripliga bilden visar. Jag hänvisar till Pharo m fl (2002) för mer information.

Båge m fl (2003) skriver:

*"Vid utformning av söksystem krävs kunskap om hur människor tänker. Många är överens om att informationssökning är en process och riktningen 'cognitive viewpoint' diskuterar vad som händer när ny information interagerar med redan existerande tankemodeller hos den informationssökande."*

För att kunna förbättra utveckling av söktjänster måste man förstå att sökarens komplexitet, hur sökaren är, hur han fungerar. Dourish (2001) skriver om *embodiment* som är sätt som människor möter den fysiska och den sociala realiteten varje dag. Hans hypotes är att tangible (konkret) och socialt computing är grundat på embodiment. Embodiment innehåller 3 viktiga moment: 1) det är inte bara fysiskt erfarenhet – det är ett sätt att vara och delta; 2) vi är inte bara "embodied" i världen, men världen är formad och formas av aktiviteter av embodied aktioner; 3) "embodied" aktion är en meningskälla.

Dourish (2001) argumenterar att utifrån fenomenologisk tradition kan man skapa position för att förklara, relatera, och utveckla tangible och socialt computing program. Enligt Dourish (2001) borde embodiment vara centralt i design och analys av interaktion. Den riktiga världen kan användas som interaktions metafor för att hjälpa människor att förstå informationssystem och hur de ska användas. "Conversational" datasystem, dvs. den teknik som använder regler för naturligt språk gör det lättare med systeminteraktion.

### ***Lite om teknikfilosofi: Dreyfus och Heidegger.***

Hubert Dreyfus är Professor i filosofi och har skrivit Being-in-the-World, som kommentar av Martin Heideggers verk Being and Time från 1927. Martin Heidegger, existensfilosofen som blev medlem i nazistpartiet har lämnat stora bidrag till teknikfilosofi, men hans texter och åsikter är svåra att förstå.

Dreyfus (2001) skriver att Heidegger inte behandlade traditionella problem som hur vi ska kontrollera teknik. Heideggers fokus är på "varat" (being, eller Dasein på tyska) och att vi som mänskliga individer måste reflektera och tänka om detta "Dasein". Heidegger såg inte teknik som något politiskt utan som något givet i den moderna tiden. Problem som uppstår ligger inte i tekniken i sig, utan är kännetecknande för modern tid. Tekniken mer ses som ett verktyg som kan tjäna som hjälpmedel för arbete. Vad som behövs enligt Heidegger är en ny

realitet som främjar icketekniska aspekter på människors erfarenheter och samhälle. Teknik existerar men formar inte människors existens och förståelse. Heideggers bidrag har legat i en ny syn på hur teknik bättre skall anpassas till människor.

Gustavsson (2001) skriver att Dreyfus kommit med kritik baserat på den filosofiska inriktningen fenomenologi. Kritiken grundar sig i att människans beteende inte kan formaliseras, vilket är nödvändigt för att en dator ska kunna reproducera det. Dreyfus har kritiskt inställning mot teorierna att människor gör sin omvärld meningsfull genom en mental representation av den. Enligt Dreyfus är personers omvärld meningsfull för dem som den är utan att behöva representeras eftersom deras kroppar har socialiserats in i ett sätt att leva, som gör att deras omvärld är meningsfull för dem baserat på den kontext de befinner sig i och deras mål.

### ***Mentala modeller utifrån Jenny Preece (1994).***

Mentala modeller är ett teoretiskt koncept som utvecklades för att beskriva dynamiska aspekter av kognitiv aktivitet. Det är modeller som människor har av sig själv, andra, miljö och saker som de interagerar med. Mentala modeller hjälper människor att beskriva, förklara och skapa förutsägelsen om framtidshändelsen. Människor formar mentala modeller genom erfarenhet, träning och instruktion.

*Strukturala modeller* kan beskrivas som surrogat eftersom de ersätter riktiga saker i människans uppfattning och förklarar hur saker och system fungerar. Det är förenklade modeller som möjliggör för användare att göra förutsägelsen om beteende av den riktiga sak eller system som modellen representerar. Modellerna har begränsad användning för att de inte räknar in användarens agerande. Fördelen med sådana modeller är att genom förklaring av funktioner kan användare förutse vad hon kan göra med saken/systemet och hur hon ska klara flest uppgifter med den. Nackdelen är att det ligger mycket arbete i att skapa struktural modell i hjärnan och lära sig vad man kan göra med modellen.

*Funktionella modeller* (eller task-action mapping modell) beskriver hur man kan använda saken/system. För att hjälpa användaren att lära sig använda en ny sak/system kan befintliga kunskaper om annan sak/system utnyttjas. Modeller är beroende av kontext vilket gör modeller lättare att använda. De flesta människors förståelse om sak/system är funktionellt.

Konceptualisering av användarnas kunskaper med hjälp av mentala modeller kan vara till stor hjälp vid utveckling av användargränssnitt.

### ***Referensbaserad modell***

I en referensbaserad modell samlar och beskriver användarna sina resurser från den egna uppfattningen om resursens innehåll och framtida användning. Den modellen består av komponenter (Andersson, 2001 genom Jakobsson m fl, 2002):

- Användare – person som utnyttjar systemet och bidrar till kunskapsvinning för sig själv och andra användare.
- Referenser – En formaliserad beskrivning av en resurs som finns eller har funnits att observera.
- Egenskaper – Varje referens har en eller flera egenskaper (namn följt av ett värde/digitalt producerbar konstruktion), vilka är kärnan i referensen och hjälper användaren bättre förstå vad referensen beskriver för typ av information.
- Resurser – Ett direkt eller indirekt observerat fenomen som är källan för den kunskap man eftersöker, kan vara ett Word-dokument, en Internetlänk osv.

Referensegenskaperna skapas av användarna när de lägger till en referens för en resurs, för att förklara för användarna om resursen är intressant att undersöka. Egenskapsbeskrivningarna skapade utifrån människors subjektiva uppfattning om resurserna medför att flera användare i systemet därigenom kan skapa skilda referenser till samma resurs

### ***Användarmodellen vid informationssökning***

Genom att ta hänsyn till användarprofilen kan man förbättra sökresultaten. För att få systemet att uppföra sig på ett sätt som överensstämmer med användarens behov kan man göra en användarmodell. Modellen kan vara helt specifik (en modell per användare) eller generaliserad (en modell för ett flertal användare). Specialiserade modeller kräver mycket information om varje användare och det kan ta en lång tid för att modellen ska bli adekvat. Generella modeller innehåller mindre detaljerad information, och är uppbyggd av ett antal fakta som anses vara sanna för alla individer som ingår i modellen (Shapira m fl, 1997, genom Jakobsson m fl, 2002).

Shapira med flera (1997, genom Jakobsson med flera, 2002) skriver om två ansatser till användarmodellering: beteendemässig och matematisk. I den beteendemässiga ansatsen genomgår användarna en övergripande undersökning, som sedan bearbetas och analyseras av beteendevetare. Utifrån den analysen skapas sedan ett antal stereotypa grupper. I den matematiska ansatsen genomför systemet någon form av matematisk beräkning för att skapa grupper, t.ex. med hjälp av klustring. Den matematiska ansatsen kräver dock mycket information om användarna, som tar lång tid att samla in.

## **4.2.3 Modeller för informationssökning**

I sitt arbete redogör Båge med flera (2003) för olika informationssökningsmodeller. De modellerna betonar mer eller mindre informell informationssökning. Här gör jag kort presentation av modellerna utifrån Båge med flera (2003).

### **4.2.3.1 Ellis - 'beteende-modell' för informationssökning**

David Ellis är professor vid institutionen för Information and library studies vid University of Wales i Aberystwyth i Wales. Hans 'behavioral model' (beteende-modell) för informationssökning togs ur en artikel av Ellis i *Journal of documentation* (Ellis, 1989, genom Båge m fl, 2003).

Undersökningen var empiriskt baserad (intervjuer, enkäter och dokumenterade databassökningar) och fokus gjordes på användarperspektivet på informationssökning. Undersökningsgrupperna bestod av forskare och lärare inom samhällsvetenskap på universitetet. Sex mönster i informationssökningen upptäcktes:

1. 'starting'
2. 'chaining'
3. 'browsing'
4. 'differentiating'
5. 'monitoring'
6. 'extracting'

Modellen lägger vikt vid beteenden och varje söksätt (kategori) rymmer en rad möjliga handlingar. Modellen flätar ihop informationssökning och informationsanvändning, eftersom sökbeteenden relaterar till varandra. Modellen anger ingen hierarki eller problemlösningssgång, men kategorierna kan användas för en modell som beskriver individuell informationssökning. Hur den enskildes sökning sedan ser ut är beroende av tid, kontext och individuella sökstrategier vid just det tillfället.

Kategorin *'starting'* innebär att identifiera källor som i sin tur kan peka framåt mot annat material. En startreferens (artikel, bok, person) gör forskaren medveten om centrala arbeten och ger en allmän överblick över forskningsfältet i stort.

Ibland övergår initialstadiet *'starting'* till *'chaining'* nästan omedelbart när forskaren går från den första referensen till material citerat i den. Vid *'chaining'* följs ledtrådar i form av citat och fotnoter, givna av den första referensen. Strategin att följa referenser och fotnoter, s.k. *'backward chaining'*, var det traditionella och huvudsakliga sättet att samla information. *'Chaining'* följer en bevakningsfunktion och kan ge en bibliometrisk överblick, s.k. *'forward chaining'* så att man får reda på hur många referenser det finns till en viss författare, artikel eller monografi i andra artiklar eller verk samt vem som har citerat vem.

*'Browsing'*, eller bläddring, eller halvstrukturerad sökning är den aktivitet där det ingår genomgång av abstrakt och innehållsförteckningar i tidskrifter och böcker, granskning av bibliotekets tidskriftsprenumerationer och att hålla sig uppdaterad genom att titta på bibliotekets eller bokhandelns hyllor. Forskaren har uppsikt över vilka arbeten som görs inom ett område och blir medveten om vad som finns tillgängligt och bekantar sig med olika källor.

*'Differentiating'* innebär att jämföra olika informationskällor och värdera dem efter relevans. Källor kan ha olika status och inom varje vetenskap finns t.ex. en eller flera tidskrifter som är erkänt mer betydelsefulla än andra. Ellis hittade tre kriterier för *'differentiating'*: 1) ämne; 2) perspektiv eller vinkling; 3) nivå, kvalitet eller typ. En tidskrift kan innehålla relevant material för ämnet men på fel nivå. Användningen av booleska operatorer vid Internet- eller databassökning ger stöd vid *'differentiating'*.

*'Monitoring'* innebär löpande bevakning av vissa informationskällor för att hålla sig uppdaterad och även medveten om andra källor än dem som forskaren just nu använder. För denna kategori är informella kontakter, *'invisible colleges'* och *'gatekeepers'*, mycket betydelsefulla. Tidningar, tidskrifter, recensioner, bibliografier och förlagskataloger utgör också källor för *'monitoring'*. I elektronisk miljö kan bokmärken på webben och anpassade egna länksamlingar vara till stöd för *'monitoring'*.

*'Extracting'* innebär ett systematiskt insamlande av material ur specifika källor för senare bearbetning och är en fokuserad och direkt sökaktivitet. Kategorierna *'monitoring'* och *'extracting'* är tätt sammanlänkade – bevakning av en källa kan övergå till insamling av material. Kollegor och handledare är värdefull hjälp under *'extracting'*, men också används skilda typer av index, förlagskataloger, bibliografier och abstrakt för denna aktivitet.

Enligt Ellis, Belkin, Kuhlthau, borde större hänsyn tas till hur användare interagerar med IR-system (information retrieval system) när sådana designas. Ellis ger rekommendationer för utveckling och förbättring av IR-system och menar att hans modell över informationsökningsbeteenden kan tjäna som vägvisare. Det är grundläggande

informationsbehov och sökbeteenden som styr informationssökningen hos Ellis. Känslor finns inte med och informationsanvändning är utanför modellen. Ellis-modellen 1993 kompletterades med två steg, 'verifying' (kontroll och belägg av information), och 'ending' (avslutande informationssökning för komplettering innan arbetet avslutas).

#### 4.2.3.2 ISP-modell av Kuhlthau

I sin bibliotekariegränning stötte Kuhlthau ofta på osäkerhet hos användarna. När Kuhlthau undersökte denna osäkerhetskänsla hos grupper av biblioteksanvändare medan de sökte information, upptäckte hon att den hade mycket bredare implikationer än man kände till. Denna osäkerhet hos användaren ställer Kuhlthau mot den exakthet och precision som informationssystemen bygger på och menar att man måste uppmärksamma de problem som uppstår i mötet mellan dessa två världar. En konstruktionsprocess är när individen går från osäkerhet till klarhet. Genom sin modell för informationssökning vill Kuhlthau komma bort från synen på informationssökning som en färdighet eller teknik och ersätta den med en processororienterad syn.

Kuhlthaus teorin går ut på att användaren ställs inför ett problem som måste lösas, upptäcker en brist i sin kunskapsbas och känner osäkerhet inför sin otillräcklighet. Osäkerheten gör att användaren försöker fylla kunskapsluckan genom att söka ny information. Inhämtning av information går stegvis och följs av lättnad, ny förvirring och slutlig tillfredsställelse, allteftersom informationsbehovet fokuseras och informationen utvinns och integreras med den gamla kunskapsbasen. Det räcker inte med ett enkelt svar på en enkel fråga, utan användaren måste genom en sökprocess förstå betydelsen av den nya informationen och sätta den i relation till den befintliga kunskapen.

Kuhlthaus teori kallas ISP (Information Search Process). Modellen har sex faser och omfattar tre områden: känslor, tankar och handlingar. Varje fas beskrivs med hjälp av uppgiftens natur; tankar och känslor i samband med informationssökningen;

Stadier	Uppgiftens början	Ämnesval	Förberedande utforskning	Fokus formulering	Informations insamling	Sök avslutning	Början på skrivning
<b>Känslor</b>	osäkerhet	optimism	frustration	klarhet	Känsla av tillit	lättnad	Tillfredsställelse eller otillfredsställelse
<b>Tankar</b>		tvetydlighet		Ökat intresse	bestämlighet		
<b>Aktioner</b>	Sökning av relevant information				Sökning av relevant information		

**Figur 2.** Kuhlthaus ISP-modell (Kuhlthau 1993, Figure 3-1 s 43, genom Båge m fl, 2003) Originalbild finns i bilaga 7

I första fasen i informationssökningsprocessen (task initiation) finns mycket osäkerhet och på handlingsplanet en bred sökning efter information. Fas två (topic selection) kännetecknas av optimism och obeslutsamhet och informationssökningen fortsätter. Under den tredje fasen

(prefocus exploration) är man förvirrad och frustrerad och tvivlar på sitt ämnesval. Fjärde fasen (focus formulation) innebär val av fokus, att göra avgränsningar eller välja infallsvinkel, och följs av en lättnadskänsla av och ett stigande intresse för uppgiften. Fas nummer fem (information collection) karakteriseras av självsäkerhet, målmedvetenhet och lättnad och sökningen riktas nu mot relevant information. I den sjätte sökfasen (search closure) sätter användaren punkt för informationssamlandet och förbereder sig för bearbetning och presentation av resultatet.

Informationssökning är en dynamisk process som tar hela personen i anspråk och blir till en holistisk erfarenhet. Kuhlthau säger att ISP-modellen bara är en metafor för en utveckling som i verkligheten är oerhört komplex och individuell. Kuhlthau har genom senare studier konstaterat att processen kan vara mera cyklisk än vad den i grafisk form I figur 2 tyder på.

Kuhlthau belyser en intressant aspekt, den lekfulla sidan av informationssökning. Glädje är en viktig faktor att ta hänsyn till och den står för grundläggande mänskliga behov. Allt lärande är kreativa processer varför lekperspektiv på informationssökning borde vara självklart.

### 4.2.3.3 Limbergs A, B och C kategorier

I en empirisk studie undersökte Limberg samspelet mellan informationssökning och lärande. Via fenomenografisk metod sökte hon fånga variationer i sätt att uppfatta informationssökning. Limberg följde 25 gymnasieelevers fördjupningsarbete rörande Sveriges eventuella medlemskap i EU. Fakta skulle samlas in och analyseras, för- och nackdelar skulle vägas, och eleverna skulle vid arbetets slut själva kunna ta ställning i frågan. Utifrån intervjuer, observationer och skriftliga redogörelser urskildes tre kvalitativt olika uppfattningar:

A informationssökning som att söka fakta

B informationssökning som att väga information för att välja rätt sida

C informationssökning som att granska och analysera

I den första kategorin (A) var målet att hitta rätt svar. Denna syn på informationssökning gav påvert inlärningsresultat och fragmentarisk kunskap eftersom fakta flyttades från källan till elevens redovisning utan att förståelse eller inläring ägt rum.

Nästa kategori (B) hade ett öga på kvantitet istället för kvalitet och samlade in mycket information. Denna grupp fick fördjupade kunskaper men också en s.k. relativistisk kunskapsuppfattning dvs. den åsikt som syns och hörs mest vinner. Gruppen använde många informationskällor.

Kategori C fick bäst inlärningsresultat, eftersom förmåga till analys och förståelse för helheten var viktiga komponenter. Eleverna försökte sätta in informationen i ett större sammanhang bl.a. genom användning av olika källor. Den här kategorin utgör idealbilden för informationssökning och följer arbetsgången i flera modeller.

Limberg jämförde varje kategori med Kuhlthaus ISP-modell och konstaterade att kategorierna B och C stämde rätt väl överens med modellens stadier medan kategori A inte alls följde processen. Hos eleverna i denna kategori ökade varken säkerhet eller intresse. Som en generell process har informationssökning stöd av två kategorier, men den kvalitativa bredden i informationssökning ryms inte i beskrivningen av processen.

Limbergs studie visar att det finns ett starkt förhållande mellan informationssökning och informationens innehåll. Svårigheten med att tillämpa generella modeller för informationssökning blir tydlig när varje individ har sitt eget mönster. Man kan också se på kategorierna som en uppsättning strategier att använda efter behov, konstaterar Limberg. Ur denna mängd av strategier väljer man olika strategier vid olika situationer och fokuserar på det som är väsentligt just då. Limbergs forskning visar att sättet att söka påverkas av synen på vad man ska ha informationen till.



### 4.3 Information Retrieval (IR) - modeller

Traditionella IR-modeller använder index för att hitta dokumenten. Sökfrågor är oftast enkla nyckelord och en rankingalgoritm bestämmer dokumentets relevans så att det mest relevanta kommer högst upp på listan.

Baeza-Yates med flera (1999) klassificerar IR-modeller på följande sätt:

1. Retrieval Adhoc Filtering
  - a. Classic Models
    - i. Boolean (Set Theoretic)
      1. Fuzzy
      2. Extended Boolean
    - ii. Vector (Algebraic)
      1. Generalized Vector
      2. Semantic Index
      3. Neutral Networks
    - iii. Probabilistic
      1. Interference Network
      2. Belief Network
  - b. Structured Modells
2. Browsing
  - a. Flat
  - b. Structured
  - c. Hypertext

Retrieval Adhoc Filtering: Adhoc är en metod där dokumenten är relativt statiska och men nya sökfrågor körs. Däremot är Filtering en metod där dokument förändras ofta medan relativt statiska frågor körs, t ex aktiemarknad eller nyhetsartiklar. (Baeza-Yates m fl, 1999)

I Classic IR beskrivs varje dokument med en grupp av nyckelord som kallas för Index. De ord som inte används i många dokument kan bättre begränsa antal returnerade dokument. Den äldsta av Classic IR modeller är Boolean där indexord relaterar genom tre booleska operatörer: AND OR, NOT. I denna modell finns eller finns inte indexord i dokumentet, vilket betyder att dokument entligen bedöms som relevant eller inte relevant (s.k. binär bedömning). (Baeza-Yates m fl, 1999)

Vektormodellen minskar binära bedömningen genom att beräkna graden av likhet mellan dokument i kollektion vilket möjliggör att få med de dokument som bara delvis matchar sökfrågan. För att beräkna likhetsgrad används clusteringkonceptet. Denna modell är populär att använda nuförtiden. (Baeza-Yates m fl, 1999)

Probabilisticmodellen är grundad på statistisk likhet och letar efter dokument som är exakt relevanta. Efter det att den returnerat resultatet för första gången bestämmer användaren vilka dokument som är relevanta och systemet använder denna information för att förbättra beskrivningen av perfekt resultat. (Baeza-Yates m fl, 1999)

Structured Models kombinerar information om textinnehåll med information om dokumentets struktur. (Baeza-Yates m fl, 1999)

Browsingmodellen hittar dokumenten genom referenser för att användare som surfar (brows) ställer inga specifika sökfrågor till systemet. Surfing har samma mål som sökning vilket är att hitta information. Den interaktiva strukturen av hypertext tillåter grafstrukturen av nodes kopplade genom dirigerade länkar. (Baeza-Yates m fl, 1999)

Jag kommer inte här att beskriva alla modeller i detalj för att det är utanför detta arbetets omfattning, men däremot hänvisar jag för vidare information till Baeza-Yates och Ribeiro-Neto "Modern Information Retrieval" (1999).

## **4.4 Olika typer av söktekniker**

Informationssökning på Internet fungerar inte så att verktyg söker genom nätet efter resurser som matchar sökfrågan. Söktjänsten söker i sin egen databas.

### **4.4.1 Sekventiell strängsökning**

Brute-force algoritm är den enklaste strängsökningsalgoritmen. Det söks genom texttecken för tecken och jämför med angivet mönster på alla positioner i texten. Ett problem med denna algoritm är s.k. backtracking. Exempel: om man söker efter mönstret "internet" och texten är "intranet", algoritmen försöker jämföra från första positionen i texten och upptäcker missmatch vid "r" i texten, men sedan fortsätter den att jämföra från position två ("ntranet") o s v. I värsta fall tidskomplexitet bli  $O(ab)$ , där  $a$  – längden på mönster,  $b$  längden på texten. (Takabayashi, 2004)

*Searching allowing errors* hittar position i texten med mönster som uppkommer med visst antal tillåtna fel, som kan vara av följande typer: återinsättning, insättning, radering. För att ha kontroll över antal fel kan NFA (Non-deterministic finite automata) eller DFA (deterministic finite automata) användas. Algoritmen skapar NFA från mönster som ska uppsökas och NFA tillstånd förändras när den läser tecken efter tecken. (Takabayashi, 2004)

Sökning genom stora texter med den typen av algoritm kan ta lång tid. Istället kan användas annan teknik som omvända filer eller suffix array. (Takabayashi, 2004)

### **4.4.2 Strängsökning med index**

Tekniker som inverted (omvänd) file och suffix array är populära att använda för informationsökning. (Takabayashi, 2004)

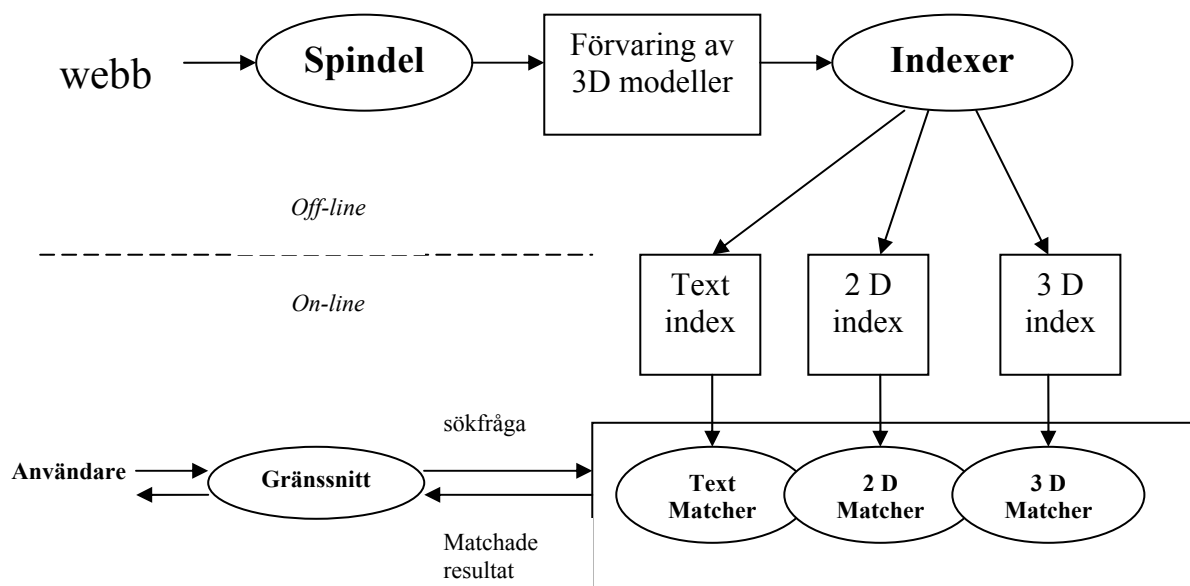
*Inverted (omvänd) file* – är en datastruktur för snabb sökning av strängar. Den består av så kallad vokabulär (alla unika ord i texten) och occurrences (positioner där varje unikt ord finns i texten). Det möjliggör att söka efter dessa nyckelord betydligt snabbare. Algoritmen som skapar inverted file utför 4 steg: 1) läser ord från text, 2) om den inte finns redan i vokabulär skrivs ordet i den med tom occurrence lista, 3) ordpositionen läggs till i slutet av dess occurrence lista, 4) processen upprepas tills texten är slut. (Takabayashi, 2004)

*Suffix array* är en struktur för effektiv sökning av större texter. Den består av en array med indexen till suffixen av lexikografisk sorterad text. Varje suffix är en sträng som börjar på en viss position i texten och avslutas i slut av texten.  
(Takabayashi, 2004)

#### 4.4.3 Sökning av information som är inte textbaserad

Det finns mycket icke textbaserad information ute på webben som t ex ljud- och video- filer, bilder, 2D och 3D objekt. För att hitta dem borde det förutom textbaserad söktechnik, också finnas form-, bild- och ljudbaserad söktechnik, vilket nu enbart i form av utvecklingsprojekt. Ett exempel är system som utvecklas av Funkhouser m fl (2002) för sökning efter 3D modeller på Internet, som är en kombination av 3D skissning, 2D skissning och text. De har utgått ifrån hypotesen att formbaserade sökfrågor skulle underlätta att hitta 3D objekten.

På bilden visas den systemorganisation som Funkhouser med flera (2002) föreslog för sökning efter 3D objekt. Systemet består av 4 steg: krypning, indexering, sökfråga och resultatmatchning.

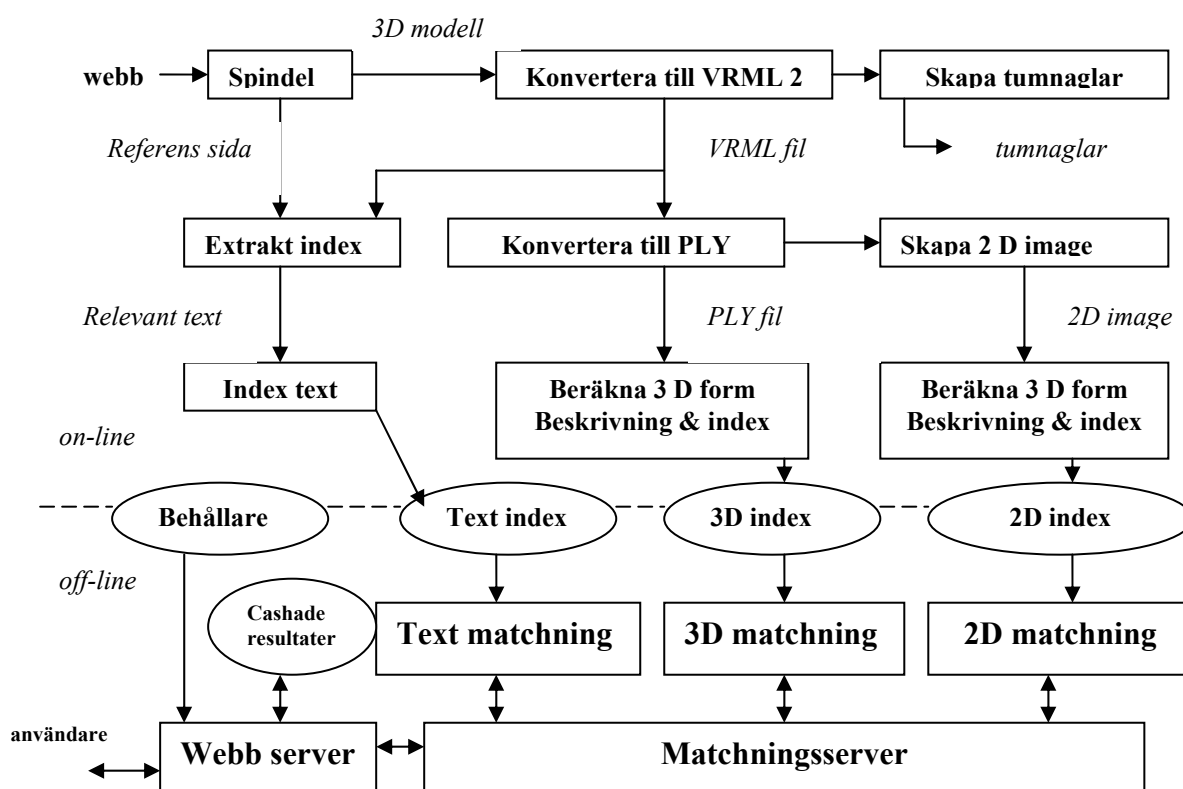


**Figur 3.** Systemorganisation ur Funkhouser med flera (2002)  
Originalbild finns i bilaga 8.

I det första steget *krypning (crawling)* skapades databas av 3D modeller och ranking bestämdes utifrån modellernas kvalitet. I den databasen lagrades 17,834 VRML modeller från webben och 2,873 kommersiella modeller från 3D utvecklare. I nästa steg, *indexering (indexing)*, skapades indexen för sökning efter 3D modeller. Det tredje steget *sökfråga (querying)* är interaktivmetoden för sökning efter 3D modellen med hjälp av nyckelord, 2D sketchning, 3D sketchning. I sista steget *matchning (resultatmatchning)* returneras 16 3D modeller som matchar sökfrågan bäst.

Funkhouser m fl (2002) beskriver 3 möjligheter att söka 3D modeller: formfrågor (shape queries), skissfrågor (sketch queries), textfrågor (text queries). *Formfrågor* utförs genom t ex att användaren anger filnamn från sin dator (“c: \dolphin.wrl”) och system söker 3D modell med liknande form. Sedan kan användaren söka för modeller med form liknande en av dem returnerade modellerna genom att klicka på “Find Similar Shape” länk under vald modell. *Skissfrågor* utförs genom att användaren skissar 2D bild med hjälp pixel paint program och systemet matchar sedan den med 2D projektions av 3D objekten. *Textfrågor* utförs genom att användaren anger nyckelord som matchas med beskrivning av modellen.

Det system eller söktjänst som Funkhouser m fl (2002) utvecklar tillåter att kombinera alla 3 sätt att söka och även köra kombinerade frågor. Denna 3D sökmotor implementerades i C/C++, Java, and Perl. På bilden visas dataflöden i denna söktjänst



**Figur 4.** Söktjänst för 3D modeller ur Funkhouser m fl (2002)  
Originalbild finns i bilaga 9.

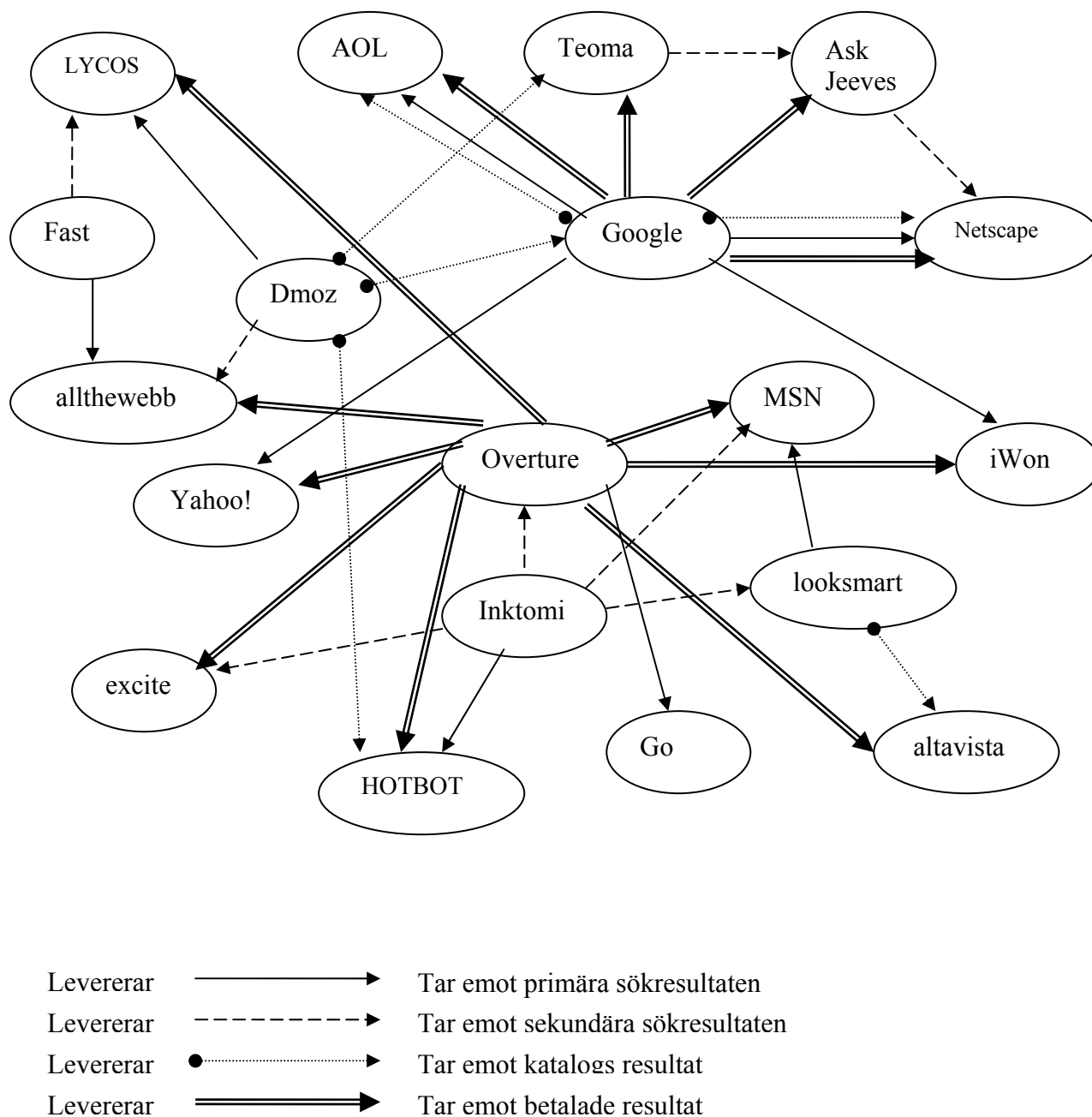
#### 4.4.4 Semantisk webbteknik

Skapare av WWW Berners-Lee (Berners-Lee m fl, 2001) ser den semantiska webben som ett sätt att fånga meningen av innehåll, så att webben mer intuitivt och intelligent kan förstå och tillfredsställa användares behov. Genom att definiera meningen av informationen skapas en struktur av innehållet i den semantiska webben, vilket ska vara till stor hjälp för söktjänster så att de kan bedöma webbsidans innehåll. Berners-Lee föreslog en 5-skiktarkitektur där utvecklare och författare med hjälp av olika teknik möjliggör för program att tolka kontexten och på så sätt förbättra sökningen.

Taibi m fl (2005) skriver att den semantiska webben skulle förbättra webb med semantiskt skikt som möjliggör för program att effektivt behandla information. T ex XML skikt skapar en syntax för att strukturera dokumenten i XML SCHEMA, men skapar inte semantisk struktur. RDF skick skapar en datamodell för objekten och relationer mellan dem i RDF SCHEMA med semantik för generalisering av hierarki. OWL skick eller ontologiskt skick innehåller mer beskrivning av klasser, relationer osv. och representerar allmänt en överenskommelse över betydelsen av data. Logiskt skick möjliggör intelligent resonemang med betydelsefulla data. En sådan struktur på den semantiska webben leder till att dokumenten kan ha relationer av olika typ med varandra till skillnad från den ordinära webben där det bara finns länk-relationstyp.

Enligt Taibi med flera (2005) är semantisk webb-metadata väldigt viktiga i blockbyggnaden i semantik webb och de använder i sitt arbete semantik webb teknik för att utveckla en semantisk söktjänst. Med denna söktjänst skulle studenter kunna köra sökfrågor som skulle kunna bearbetas semantiskt för att hitta konceptuella relationer mellan dokumenten. Författarna föreslår två möjliga arkitekturer: Static Ontology Based Search (SOBS) arkitektur och Dynamics Ontology Based Search (DOBS) arkitektur. I SOBS använder sig söktjänsten av en uppsättning av ontologier för att ordna webbresurser enligt semantiskt grund. DOBS innebär en mer flexibel söktjänst som kan processa sökfrågan med ontologi som användaren definierar själv. För mer information jag hänvisar till Taibi med flera (2005).

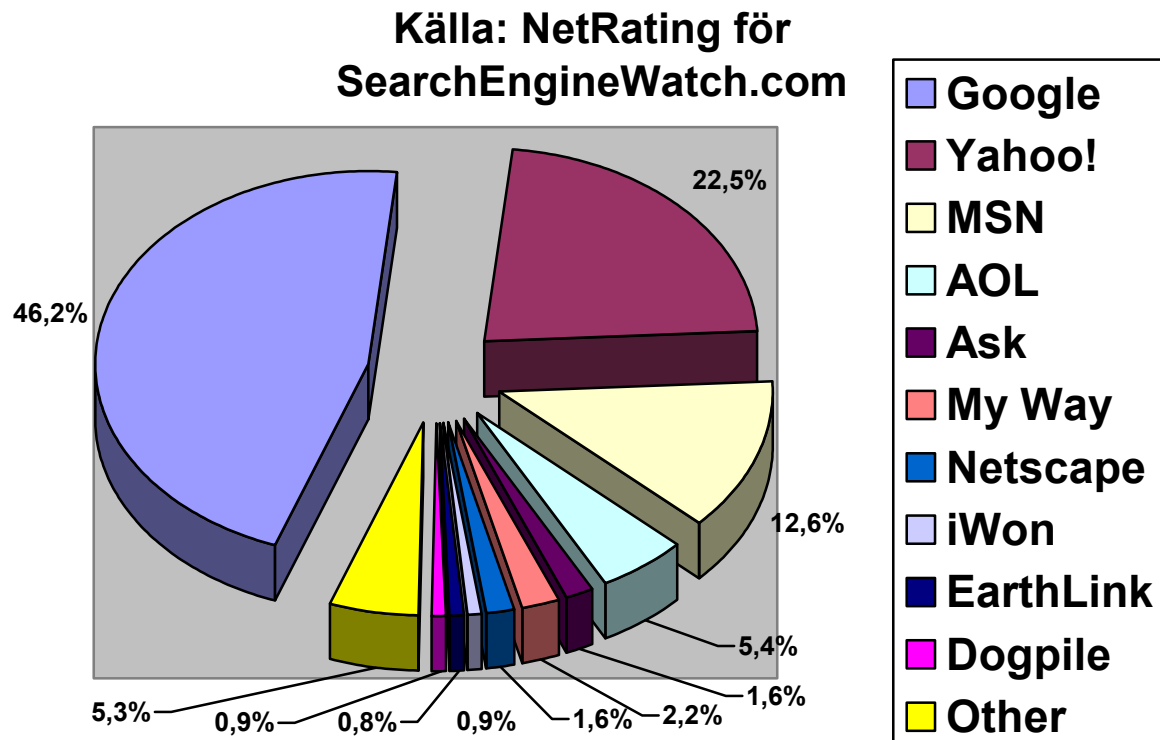
## 4.5 Söktjänster på Internet



**Figur 5.** Söktjänsters relationer ur <http://www.ellipseinc.com/searchengine/images/relationship.jpg> eller <http://www.bruceclay.com/searchenginereationshipchart.htm>  
Originalbild finns i bilaga 10.

På bilden visas befintligt tillstånd av söktjänster och dess relationer. Bland de största söktjänsterna är Google, Ask Jeeves, Yahoo! och MSN Search. De anses vara störst eftersom de antingen är mest omtalade, flitigt använda, eller indexerar flest sidor (Sullivan, 2004).

Välkända och kommersiellt uppbackade söktjänster betyder oftast att de med stor sannolikhet är väl underhållna och uppdaterade, vilket även är attraktivt för användarna som genomför sökningar (Sullivan, 2004).



**Figur 6.** Diagram över sökningar av US Internetanvändare under juli 2005 (Sullivan, 2005) <http://searchenginewatch.com/reports/article.php/2156451>  
Originalbild finns i bilaga 11

Figuren 6 visar i procent vilken söktjänst som användes i 4.5 miljarder sökningar som utfördes av US Internetanvändare både hemma och på arbetet under juli 2005. Enligt den gick de flesta sökningarna via Google – 46,2 %.

#### 4.5.1 Generellt om söktjänster

De flesta söktjänster på Internet är baserade på text för både sökning och presentation för användare, varför alla problem som kan finnas i text är med i sökningen. Synonymer (ex automobil, bil och motorfordon) är flera olika ord som betyder samma sak. Användaren kommer inte åt webbsidan om den inte innehåller den valda synonymen även om dokumentet skulle tillfredsställa informationsbehovet. Om ord har samma stavning men olika betydelse (t ex bok som man läser, bok som en trädsort) kan det också komma fram irrelevanta dokument. (Ahmstedt,2003)

Många söktjänster presenterar för användaren dokumentens sammanfattning som innehåller titel, URL, kort text från dokumentens body (kropp), ofta också storlek och datum. Metoder för att konstruera en sammanfattning kan vara olika: första några rader av text, extrakt av text relaterad till sökfrågan.

Alla sökmotorer lämnar sitt namn i den besökta webbplatsens loggfil. Genom att titta i denna kan alltså den ansvarige avgöra om platsen behöver anmälas manuellt till någon söktjänst. (Bjurbäck, 1998) Vilka sidor en söktjänst känner till beror på var dess ”spindel-program” har varit eller vad användare har anmält för resurser. Olika sökverktyg presenterar resurser i olika typer av strukturer: ämnesbaserad, geografisk (geografisk ordnade efter karta) eller tidsmässig (i tidsordning).

Alla söktjänster inkluderar 4 moment: insamling (Spider), indexering (Indexer), sökning (Query Analyzer), gränssnitt. Olika moment kan implementeras på olika sätt i olika sökverktyg även de som i stora drag är ganska lika varandra.

### ***Spider***

Insamlingsprogram oftast benämnda spindlar, robotar eller krypare (eng. crawlers) skapades eftersom det är omöjligt att manuellt indexera så stora mängder av sidor på Internet. Det är mjukvara som söker webbsidan och kontrollerar alla hyperlänkar. I kontrollmodulen samlas alla URL inifrån returnerade dokument och bestäms vilka länkar som ska besökas av spindeln nästa gång. Websidor som hittas av en spindel lagras temporärt i s.k. Page Repository tills de skickas till indexering. Dokumentinformation sparas sedan i en databas där allt indexeras med någon typ av algoritm. Oftast är hela innehållet i ett HTML-dokument extraherat och indexerat (förutom stoppord). (Baskaran, 2005) Vissa sidor som t ex nyhetssidor uppdateras ofta och då måste spindeln besöka sådana sidor oftare. Spindlar fungerar på olika sätt, vilket kan leda till vissa bias som kan påverka vilken information som lagras i sökverktygens databas. T ex crawlers som uppsöker toppnivådomän eller som undersöker djupare struktur av en toppnivå domän. När sedan en informationssökning utförs, letas sökorden upp i databasen, och i gränssnittet presenteras länkarna till de dokument där sökord förekommer. (Ackerman m fl, 2001)

### ***Index***

Det finns 3 generella typer av indexering (index) för webbsidor: textindex (content), länkindex (struktur), special-purposeindex. Contentindex innehåller samma pressad information för varje webbsida som till exempel lista av alla termer som finns på sidan, nyckelord eller fraser, titel o s v och allt detta lagras i omvänd struktur. I text (content) index ingår så kallas omvänt (inverted) index, som har information om ords placering i kollektionen till exempel sidas identifikator och ords position på sidan osv. Länkindex byggs upp som en modell av webben i grafer där en omvänd webbgraf har information om t ex sidor som innehåller länkar till en viss sida. Sådan information används av algoritmer (PageRank, HITS – Hypertext Induced Topic Search) för att beräkna sidans betydelse. Special-purpose index är index för t ex image, PDF (Langville m fl, 2005).

På vissa tjänster utför människor sortering av information i kategorier (ex Yahoo), vilket enligt Brin och Page (1998) är subjektivt, väldigt dyrt och inte täcker alla ämnen.

### ***Query Analyzer***

Sökning eller Query Analyzer kan fungera på olika sätt till exempel kan vissa sökmotorer leta efter söktermer i hela dokument, andra söker endast genom de första tio raderna, medan en del söker genom de termer som upphovsmannen har valt och kodat in för att beskriva



dokumentet. Att välja rätt söktjänst är en viktig variabel i webbsökning därför att olika sökmotorer finner olika dokument trots att man har använt samma sökord (Limberg m fl, 2002). Vissa verktyg erbjuder möjlighet att i efterhand förbättra en sökning genom att presentera ord som skulle förbättra sökningen (ex Excite, Google)

### **Gränssnitt**

Gränssnitt är den webbsida som användaren ser och därifrån utgår sökningar. Det är ett dilemma för gränssnittsutvecklare att designa webbsidan för användarna: om användargränssnittet är förenklade för att underlätta för den ovane användaren kan de ge sämre möjligheter för experten. Det finns stora skillnader i hur verktyget ser ut för användaren och två generella sätt: nyckelordsökning och bläddring. (Hock, 2001)

Nyckelordsökning betyder att i ett gränssnitt finns en ruta där användaren kan skriva ord som kännetecknar den information man letar efter. Exempel på tjänster som erbjuder nyckelordsökning är Google, AltaVista, MSN search osv. Vissa gränssnitt har bara grundläggande nyckelordsökning. Då bli sättet att utföra en sökning på att skriva en söksträng i ett textfält och klicka på en sök-knapp eller trycka på returtangenten. Men de flesta söktjänster idag erbjuder till användaren två gränssnitt, för enkel resp. avancerad sökning där användaren kan specificera sökningen med bl a booleska operatörer (s.k. Advanced Search på Google, AltaVista osv). Det är ett mer effektivt sätt att söka i dessa stora databaser. Antingen används de booleska operatorerna AND, OR och NOT eller genom att specificera olika tecken såsom plus (+), minus (-) och citationstecken (" ") eller finns det hos en del verktyg möjlighet att kombinera sökord med hjälp av valmenyer vilket ger ett likvärdigt resultat. När användaren formulerar en fråga finns olika begränsningsmöjligheter t ex att ange typ av källa, vilken del av dokumenten man vill söka i, språk, när den är skapad, var den finns (man kan specificera geografiskt område t ex Sverige eller virtuellt område t ex utbildningsinstitutioner.edu). Jokertecken (tecknet \*) fungerar som en ersättning för ett eller flera tecken, t.ex. en ändelse eller en förstavelse. (Koch, 1996)

Bläddring betyder att användaren bläddrar fram i databasen, som presenteras på ett strukturerat sätt och med hierarkiskt struktur på innehållet. Bläddringssätt är vanligast vid manuellt sammanställda databaser, innehållet genomgår någon typ av urval och beskrivningarna är inte automatiskt genererade (Koch m fl, 1996). Exempel på tjänster som erbjuder bläddringssökning är WWW Virtual Library. Koppling mellan sök- och bläddringsstrukturer där om man gör sökning man hamnar i en lämplig del av bläddringsstrukturerna och vice versa är inte så vanligt (InfoSeek).

När en användare frågar efter en eller flera termer, så letas termerna upp i databasen. Webbsidorna där termerna förekommer blir identifierade och därefter rankade beroende på dess relevans till sökfrågan. De rankade sidornas länkar visas sedan upp i klientens gränssnitt i form av en resultatlista med en kort beskrivning taget från de rankade sidorna så att det ger användaren en uppfattning om sidans innehåll. (Baskaran, 2005)

### **4.5.2 Ranking**

När användaren söker alla relevanta resultat som hittas kallas det för "Pertinent Pages" och är sidor som innehåller sökorden. Efter identifieringen skickas dessa sidor till rankingen där de sorteras med hjälp av ett rankingprogram i den ordning som är relevant för användaren och det mest relevanta placeras högst upp (Langville m fl, 2005). Vissa sökmotorer baserar

sorteringen på kommersiella intressen, vilket kan störa sökningen om användaren inte är intresserad av kommersiell service. Dokument kan bedömas mer eller mindre relevant t ex om alla sökta ord finns med i dokumentet, om de sökta orden står nära varandra, om de sökta orden finns i början av dokumentet, om de sökta orden förekommer många gånger i dokumenten, om de sökta orden finns i större eller fet skrivstil. De flesta söktjänster använder variation av boolean och vektormodeller för ranking (Baeza-Yates m fl, 1999). Ranking är väldigt viktigt eftersom användaren vanligtvis bara tittar på första sidan av det returnerade resultatet. Två sätt att mäta om ranking är effektivt är Precision (antal returnerade relevanta sidor i förhållande till totalt antal returnerade sidor) och Recall (antal returnerade relevanta sidor i förhållande till totalt antal relevanta sidor som finns i databasen) (Van Rijsbergen, 1979). Webbssidor som hittas av spindel lagras temporärt i s.k. Page Repository tills de skickas till indexering.

PageRank är en algoritm som utvecklades av Brin och Page och som ligger till grund för Googles globala rankingschema. (Brin, Page, 1998) Goggles framgång visar att PageRank är en effektiv metod för att ranka sidor. Undersökning som gjordes av Baeza-Yates m fl visade att PageRank sätter högre ranking på äldre sidor och de gav förslag på modifiering av algoritmen för att korrigera denna bias.

Tekniken *click popularity* utvecklades av DirectHit ([www.directhit.com](http://www.directhit.com)) och den går ut på att samla information om frågor som användare kör på söktjänster, som sidor besöks efteråt och tiden som spenderas på sidan. Den informationen används för att returnera sidor som flera användare som har samma fråga besöker.

### 4.5.3 Databaser

Alla söktjänster bygger på databaser som innehåller information om resurser som finns på nätet. Databasen fungerar som en stor katalog med kopior av alla besökta och lästa sidor. Databaser sammanställs på olika sätt, skiljer sig i storlek och innehåll. Det finns olika typer av databaser: robotbaserade (ex HotBot, Northern Light), list eller bidragbaserade (Yahoo, Punkt.se).

Robotbaserade databaser byggs genom att en spindel går i nätet genom att följa länkar och samlar in upphittade dokument som indexeras och lagras i databasen. Det är därför mer sannolikt att man via sökmotorer hittar en resurs som många andra har länkat till än en som få eller ingen har länkat till. (Koch m fl, 1996)

List eller bidragbaserade databaser byggs genom att ett antal människor letar efter resurser och lägger till beskrivningar i databasen. Sådana databaser brukar vara mindre i storlek. Manuell katalogisering liknar ett referensbibliotek som är manuellt administrerat. Webbadministratörer skickar in en kort beskrivning på webbplatsen för katalogisering. Vid en sökning sker bara matchning mot beskrivningarna som är insända. Att ändra en webbsida utan att ha fått webbplatsen omregistrerad kommer att resultera i att förändringen inte hittas vid en sökning. (Koch, 1996)

En databas kan vara universell eller begränsad t ex till ett visst ämne eller ett geografiskt område. Databasen består av ett antal organiserade tabeller för att underlätta snabb åtkomst av information. Med ökade antal webbsidor är det vanligt att sökmotorer är kopplade till fler än en databasserver. Eftersom innehållet på webben förändras hela tiden (dokumenten ändras,

raderas, flyttas eller kommer nya) är det omöjligt att skapa en perfekt databas. Sidor som sedan länge tagits bort kan komma upp bland sökresultaten om inte databasen är effektivt uppdaterad. Ämnesstruktur beror på databasens storlek och dess innehåll. Det kan vara olika på olika ställen, olika ämnen kan vara olika bra täckta. Vissa verktyg kan visa på något sätt var i strukturen användaren befinner sig. (Koch, 1996)

#### **4.5.4 Användbarhet**

Begreppet användbarhet är ett nyckelord inom området MDI (människa/dator interaktion) och handlar om att göra system lätta att lära sig och lätta att förstå (Preece, 1994). Preece definition av användbarhet är att det är ett resultat av Relevans, Effektivitet, Attityd, Lärbarhet, Tillförlitlighet och Funktionalitet.

Relevans innebär att en söktjänst ska vara relevant för användarnas behov. En söktjänst som inte kan tillfredsställa användarnas behov förlorar trovärdigheten hos användare. Effektivitet visar hur väl användarna kan utföra sina uppgifter med hjälp av söktjänst, det ska gå lätt och snabbt att använda. Attityd innebär att en söktjänst ska upplevas positivt av användarna vilket kan höja användarnas motivation. Lärbarhet handlar om hur lätt det är att lära sig använda en ny söktjänst. Tillförlitlighet innebär att en söktjänst ska ha en låg felfrekvens så att användarna tror på den. Funktionalitet innebär att en söktjänst innehåller de funktioner som användarna behöver för att utföra sina uppgifter.

#### **4.5.5 Olika typer av söktjänster**

Det finns två generella söktjänstplattformar: directories och crawler-baserade. Directories eller blädderverktyg finns i form av ämneskataloger och ämnesportaler. Crawler-baserade söktjänster inkluderar sökmotorer (indexerade söktjänster) och metasöktjänster (multipla sökmotorer). Det finns även kombinationer av dessa. (Hock, 2001)

Det finns inte söktjänster som täcker alla internetsidor och dokument och som presenterar rätt resultat överst i resultatlistan. Utbytet och variation av söktjänster är ganska stor och det finns ett tiotal stora generella söktjänster och tusentals andra som har olika storlekar, täckning och funktion. Söktjänster använder sig av olika sökmotorer och är ofta beroende av varandra för att få fram sökresultat. De komplexa sambanden mellan bakomliggande sökmotorer kan vara svåra att överblicka. (Hock, 2001)

##### **4.5.5.1 Sökmotorer**

En sökmotor kan man säga är ett verktyg som gör det möjligt att utföra sökningar på Internet (d v s i egen databas) genom att använda sig utav olika fraser eller nyckelord. Sökmotorer är det mest etablerade sättet att söka efter information på Internet. Ranking av returnerade dokument beror på den algoritm som används av söktjänsten. Vissa av sökmotorerna erbjuder mot betalning företag möjligheten att deras webbsidor ska placeras bra i träfflistor. AltaVista var den första crawler-baserade söktjänsten på webben. Den startades 1995 och var ledande i många år tills Google kom. De flesta ämnesportalerna (ex Yahoo) innehåller en sökmotor.

Förutom de mer allmänna finns en mängd specialinriktade sökmotorer:

- nyhetssökmotorer - letar genom nyhetswebbplatser;
- multimediasökmotorer - hittar videofilmer, ljud- och bildfiler;
- svarssökmotorer - hittar svar på sökförfrågan i stället för att hitta en viss webbsida;
- ämnesspecifika sökmotorer (s.k. vertikala sökmotorer) - söker webbplatser som uteslutande handlar om datorer, finans, juridik, medicin, vetenskap, resor etc.;
- barnsökmotorer - filtrerar ut webbplatser som inte är lämpliga för barn;
- sökmotorer som letar igenom den osynliga webben.

De specialinriktade sökmotorer som baseras på informationsfiltrering fokuserar på ett begränsat område, vilket är mer effektivt alternativ till giganter som t ex Google. (Hock, 2001)

#### 4.5.5.2 Metasöktjänster

Metasöktjänster har inga egna databaser och söker istället genom flera andra verktygs databaser. I en metasöktjänst kan man leta i flera separata söktjänster samtidigt genom ett interface. På det sätt går det att i en sökning täcka en större del av webben, än om man söker med vanlig sökmotor. Den innebär samtidigt att man kan få en stor mängd av resurser i en träfflista, vilket kan vara tidskrävande. Även om det ofta finns möjlighet att göra avancerade sökningar är dessa begränsade då det inte går att använda alla de avancerade sökfinesser som de enskilda söktjänsterna har. En metasöktjänst kan vara användbar när man söker inom något ovanligt ämnesområde och kanske inte får tillfredställande svar från endast en söktjänst. Metasöktjänster är mest effektiva när indexen från den använda söktjänsten inte överlappar mycket. (Baeza-Yates m fl, 1999)

Metacrawler är en av de äldsta metasöktjänster och som startades 1995 på Washington universitet och blev uppköpt 1997 av InfoSpace

#### 4.5.5.3 Ämneskataloger eller ämnesportaler

Ämnesportalerna kan kallas för Subject Based Information Gateways (SBIG), Quality Controlled Subject Services eller ”virtuella bibliotek”. En ämnesindelad söktjänst, även kallad katalog, motsvarar bokens register. I en ämnesportal samlas länkar, som är systematiskt organiserade så att man kan bläddra sig fram till önskad kategori. Innehållet i katalogerna är vanligtvis hierarkiskt strukturerat i ämneskategorier. Vissa kataloger ger även omdömen och mer eller mindre utförliga beskrivningar av resurserna. Innehållet i katalogerna är ofta sökbart. Det kan finnas beskrivningar av resurserna och nyckelord. Sidor kategoriseras efter innehåll, vilket sker manuellt och innehållet i katalogerna är utvalt av redaktörer. Ofta kan söktjänstens besökare också själva föreslå egna sidor som ska läggas in. Det finns olika typer av kataloger, de kan vara allmänna eller ämnes/resursspecifika eller vara inriktade mot en speciell målgrupp. Principerna för urval av resurser skiftar. Ämnesinriktade kataloger har vanligtvis mer uttalade kvalitetskrav än allmänna kataloger. Det kan finnas någon form av sökmöjlighet i dessa verktyg. Ämnesportaler är oftast sämre uppdaterade än sökmotorerna. Det görs ingen automatisk uppdatering av databasen, utan redaktörerna måste själva besöka sidor på Internet och lägga in de som de anser vara av allmänt intresse i passande kategori. (Hock, 2001)

Fördelarna med katalogerna är att de ofta erbjuder ett kvalitativt och överblickbart urval av resurser. Med hjälp av den hierarkiska strukturen är det lätt att orientera sig och hitta informationskällor inom ett specifikt ämnesområde. Men även de största katalogerna täcker endast ett litet urval av resurserna på nätet.

Den äldsta ämnesindelade söktjänsten är Yahoo, som lanserades 1994. Yahoo har även ett samarbete med den indexerade söktjänsten Google, vilket innebär att din sökning skickas vidare till dem om inga träffar finns bland Yahoos kategorier. (Sullivan, 2004)

Open Directory Project DMOZ är den största ämneskatalog på webben som upprätthålls av frivilliga redaktörer. 1998 övertogs den av Netscape (AOL). ([www.dmoz.org](http://www.dmoz.org))

Libris (<http://www.libris.kb.se/>), en gemensam katalog för forsknings- och specialbibliotek i hela landet. Libris listar bland annat böcker, rapporter, avhandlingar och tidskrifter. För närvarande finns mer än fyra miljoner titlar i databasen som man kan nå via Internet.

Katalog Bibliotek.se (<http://www.bibliotek.se/>) innehåller hela Libris databas samt databasen BURK (Bibliotekstjänst AB). BURK registrerar folkbibliotekens samlingar av böcker, kartor, skivor, musiktryck, videokassetter, ljudböcker, talböcker, punktskriftsböcker, tidningar, tidskrifter och annat som finns hos biblioteken.

#### **4.5.5.4 Jämförande shopping-söktjänster (comparison shopping engines)**

Hjälper användarna att hitta information om produkter som säljs on-line. Froogle är en exempel på crawler-baserad shopping ([www.froogle.com](http://www.froogle.com)), den letar efter relevanta varor med hjälp av söktermer. Resultatet är baserat på "datafeed" från butiker och varor. (Paulson, 2003) Yahoo!Shopping listar upp varor hos butiker som är med i Yahoo Store.

#### **4.5.5.5 Söktjänster som svarar på frågor (QA – Question Answering Systems)**

Question Answering Systems returnerar svar på ställd fråga och därigenom ger användaren exakt information som den behöver istället för en lång lista av dokument. Direkta svar på faktafrågor som t ex "Vem fick Nobels fredpris 2000?" hämtas i en kunskapsdatabas. (Kwok m fl, 2001)

START ([www.ai.mit.edu/projects/infolab](http://www.ai.mit.edu/projects/infolab)) är en av de första QA systemen med webbinterface. Den utvecklades 1993 av Boris Katz och hans kollegor på InfoLab Group i MIT Computer Science and Artificial Intelligence Laboratory.

Wondir (<http://www.wondir.com>) är icke kommersiell QA söktjänst med möjlighet att söka i databas eller få "live" svar från frivilliga. (<http://www.wondir.com/wondir/jsp/index.jsp>)

#### **4.5.5.6 Osynlig webb**

De delar av webben som inte går att nå genom sökmotorer kallas för osynliga webben. Det finns olika anledningar till att det inte går att nå resurserna genom sökmotorerna som t ex är tekniskt omöjligt för sökmotorerna att indexera en resurs eller att materialet är skyddat bakom lösenord. Mycket av det vetenskapliga materialet går inte att nå via sökmotorer utan tillhör den ”osynliga webben”.

Det finns tjänster för att nå resurser i osynliga webben. Där finns hierarkiskt ordnade webbadresser till databaser i olika ämnen. Det går även att använda sökmotorer för att nå fram till en förstasida för en viss databas. Men det går inte att söka i själva databasen med hjälp av en sökmotor.

Exempel på söktjänster för att nå den osynliga webben: CompletePlanet, ELIN@Lund, Invisible Web Directory.

#### **4.5.6 Hur man använder söktjänster**

En sökmotor kan t ex erbjuda många olika sätt att söka som till exempel att söka på ett eller flera ord, i en eller flera avdelningar på sajten, på dokument inom vissa tidsperioder, på särskilda typer av dokument. Alla dessa metoder kan dessutom kombineras med varandra.

Undersökningar gjorda av företaget Onestat visar att antalet använda sökord fördelar sig enligt följande:

- 30% av alla är två ords sökningar
- 27% är tre ord sökningar
- 17% består av endast ett ord
- 15% består av fyra ord

Sökningar med tre ord ökar, för ett år sedan var det vanligare med sökningar med ett ord än tre ord. (Våge, 2004)

Söktjänsterna erbjuder oftast sökhjälp: Google har beskrivningar av sina funktioner, Yahoo har dokumentation i form av frågor och svar (FAQ), MSN Search har en beskrivande hjälpfunktion som liknar den i Windows, AskJeeves har en MyJeeves Tour i fyra steg för att hjälpa folk komma igång med att spara och organisera träffar, i övrigt tips och FAQ, osv.

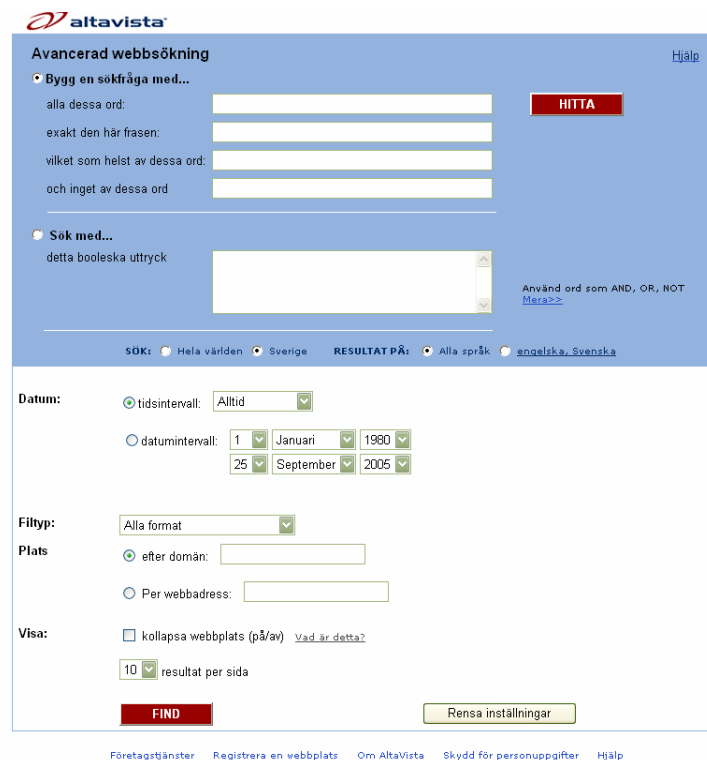
## 4.5.7 Några av största söktjänster

### 4.5.7.1 Alta Vista (<http://www.altavista.com>)

AltaVista är en snabb, kraftfull söktjänst. Den har 2 sökalternativ: Enkel och avancerad sökning (se bilderna):



Figur 7. AltaVistas enkel sökning



Figur 8. AltaVistas avancerad sökning

AltaVista använder en egen sökmotor och var först med att använda en full-text indexering på Internet år 1995. Under 1999 lanserades stöd för sökning av multimedia (ljud/video/bilder). Sökningen är skiftlägeskänslig och tar även hänsyn till diakritiska tecken, såsom é, è, ë, osv, vilket möjliggör en mer nyanserad sökning. Sökresultaten presenteras i första hand med hjälp av webbsidans META-beskrivning och i andra hand med hjälp av dokumentets första två rader. (AltaVista)

## 4.5.7.2 Yahoo! (<http://www.yahoo.com/>)

Yahoo! erbjuder både enkel och avancerad sökning (se bilderna):



Figur 9. Yahoo! enkel sökning



Figur 10. Yahoo! avancerad sökning

Yahoo! är den äldsta och mest populära länkkatalogen som grundades 1994 av Stanfords doktorander David Filo and Jerry Yang. Nuförtiden erbjuder den särskild sökning av bilder, nyheter samt produkter och den nya sökroboten indexerar även pdf, word, excel, powerpoint m.fl. filformat. Yahoo! Inc. har förutom informationssökning annan service Yahoo! Mail, Yahoo! Messenger, Yahoo! Calendar, Yahoo! Chat, Yahoo! Greetings, Yahoo! Clubs and Yahoo! Photos. Under flera år utnyttjade Yahoo Googles databas som ett komplement i sin söktjänst, men den 18 februari 2004 presenterade man en helt ny egenutvecklad sökmaskin och databas som ska förbättra webbsökningen och i stället konkurrera med Google. (Yahoo!)

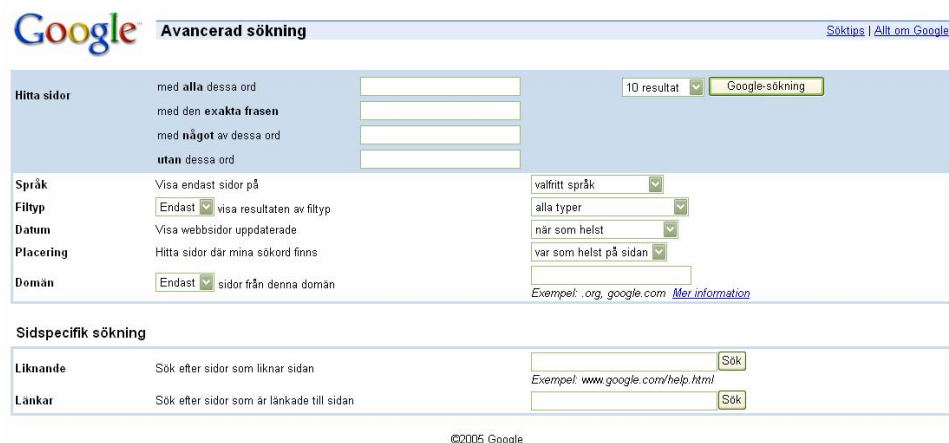


### 4.5.7.3 Google (<http://www.google.com>)

Google har användargränssnitt som är enkelt, rent och erbjuder såväl enkel som avancerad sökning (se bilderna):



Figur 11. Google enkel sökning



Figur 12. Google avancerad sökning

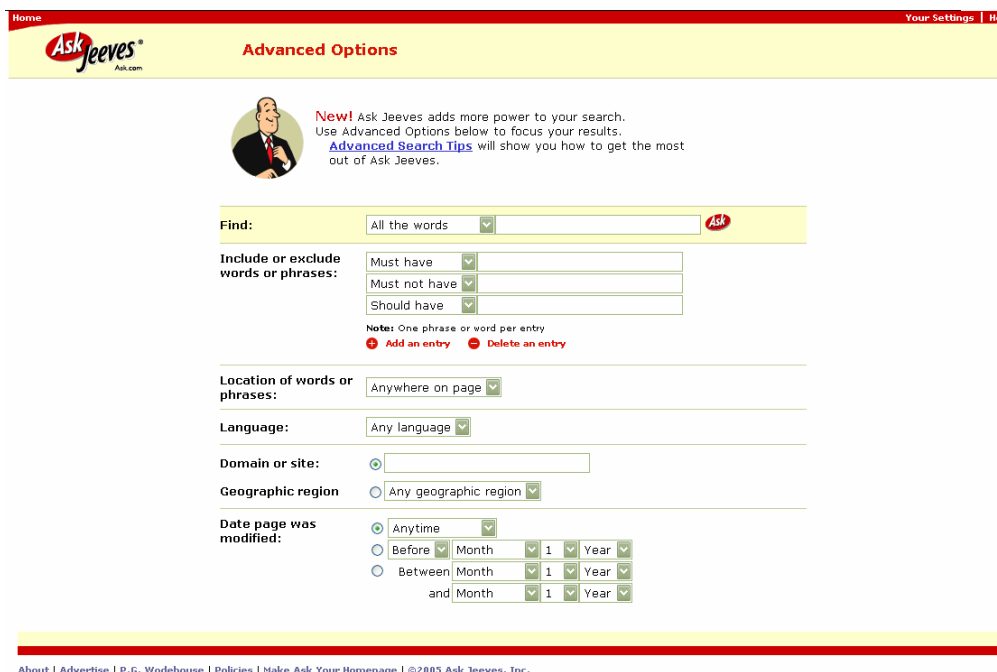
Brin and Page utvecklade en väldigt framgångsrik sökmotor, Google år 1998. När detta arbete skrivs har denna sökmotor i sin katalog ”11,3 miljarder objekt, varav 8,2 miljarder sidor och 2,1 miljarder bilder” (CS1, 2005), besvarar mer än 100 miljoner sökförfrågningar om dagen och har blivit den populäraste sökmotorn. (Om Google) Brin och Page identifierade två huvudmål för deras sökmotor: scalability and relevancy. GoogleBot (Google crawler) går igenom alla webbsidor som den känner till med några veckors mellanrum. Vissa webbsidor (t ex nyheter) kollas dagligen eller även oftare. Programvaran är väldigt snabb och kontrollerar många sidor parallellt (Avi Rapoport, 2002). Googles ranking baseras på PageRank-algoritmen och är en effektiv metod för att ranka sidor (Google).

#### 4.5.7.4 Ask Jeeves (<http://www.ask.com/>)

Ask Jeeves grundades i 1996. Den erbjuder både enkel och avancerad sökning (se bilderna):



Figur 13. Ask Jeeves enkel sökning



Figur 14. Ask Jeeves avancerad sökning

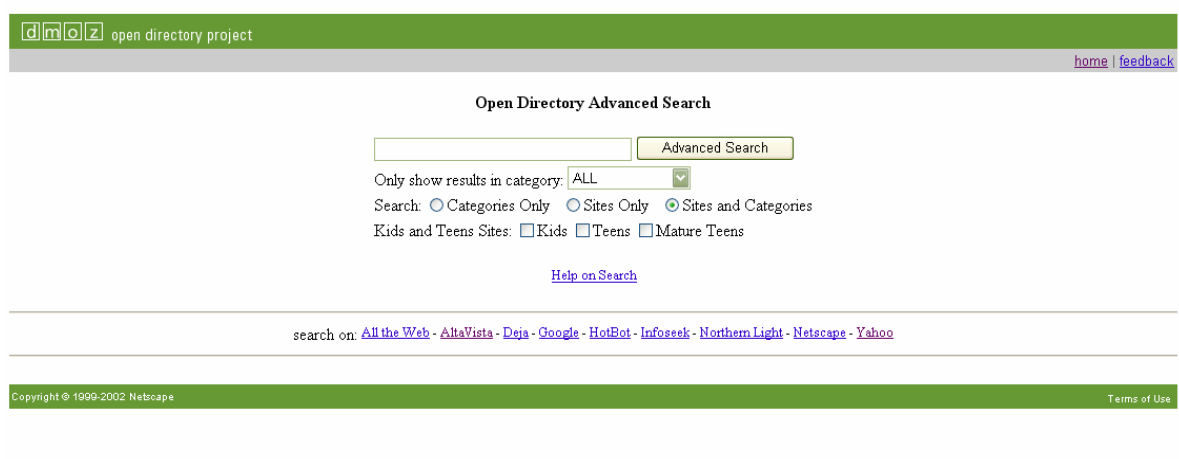
Söktjänsten Ask Jeeves är framför allt specialiserad på s.k. "naturligt språk"- sökningar. Den utnyttjar dels ett manuellt sammanställt index med utvalda svar på 7 miljoner vanliga frågor, dels indexet från sökmaskinen Teoma. Ask Jeeves erbjuder särskild nyhetssökning samt bildsökning från PicSearch. (Ask Jeeves)

#### 4.5.7.5 Open Directory (http://dmoz.org/)

Icke-kommersiell länkkatalog som startade 1998 och som sammanställs av över 66.000 ideellt arbetande redaktörer runt om i världen. Den erbjuder både enkel och avancerad sökning (se bilderna):



Figur 15. Open Directory Project enkel sökning



Figur 16. Open Directory Project avancerad sökning

Open Directory Project (ODP) katalogen omfattar idag runt 4 miljoner länkar och är den största katalog skapad av människor. Eftersom den är fritt tillgänglig (via avtal med Netscape) utnyttjas den som ett komplement till de egna tjänsterna av flera stora söktjänster t.ex. Google (via länken More ovanför sökrutan i den engelska versionen). Sökmöjligheterna är inte så utvecklade men de booleska operatorerna fungerar (Open Directory).

#### 4.5.7.6 Wondir (<http://www.wondir.com>)



Figur 17. Wondirs sökinterface

Wondir är annorlunda söktjänst på så sätt att den för samman människor som behöver information och de (människor och organisationer) som har information. Söktjänsten är gratis. Wondir har kombinationen av människors svar i realtid, metasearch och sökning i osynlig webb. (Wondir)

### 4.7 Sökprocess - strategisk informationssökning

“With so much information now online, it is exceptionally easy to simply dive in and drown.”  
Alfred Glossbrenner.

Limberg m fl skriver att enbart tillgång till information räcker inte för att hitta det man vill. De flesta vanliga användare har inga kunskaper om informationssökning och de utgår från egen erfarenhet och många upplever nuförtiden att det är svårt att hitta önskad information med hjälp av söktjänster utan att få fram en mängd oönskade sidor. Ett flertal författare som bl a Hock (2001), Large m fl (2001), Ackerman m fl (2001) har skrivit om det och hur information kan sökas på det mest effektiva sättet.

Det finns miljoner sidor i sökmotorernas databaser. Det betyder att en sökning måste vara så precis som möjligt för att begränsa resultatet. De olika sökmotorerna har olika informationsresurser indexerade i sin databas och man når fler resurser om man använder flera sökmotorer. Enligt Sherman (2005) gäller för Ask Jeeves, Google, MSN search and Yahoo följande fakta:

- 84.9% of total results are unique to one engine
- 11.4% of total results were shared by any two engines
- 2.6% of total results were shared by any three engines

1.1% of total results were shared by any four engines.

Och:

73.9% of Ask Jeeves first page results were unique to Ask Jeeves

71.2% of Yahoo first page results were unique to Yahoo

70.8% of MSN search first page results were unique to MSN search

66.4% of Google first page results were unique to Google

När man vill söka information på Internet med hjälp av en söktjänst är det viktigt vilken formulering man väljer därför att systemet förstår inte naturligt språk som människor använder. Språk med lingvistisk mångfald har inverkan vid informationssökning. Språkets komplexitet gör att informationssökning är mer komplicerat än att bara ange några nyckelord, och förvänta sig ett sökresultat som ska innehålla allt som finns tillgängligt på det som begärdes. Det kan hända att sökningen misslyckas vid första försöket och i sådana fall behöver man göra korrigeringar för att förbättra sökresultatet.

Enligt Large m fl (2001) finns det två grundläggande principer för informationssökaren. Den första är att minska sökresultatet av ej relevant information utan att förlora alltför mycket av relevant information. Ett sätt för att uppnå detta är att ersätta ett av de sökord som har använts med ett mera preciserat sökord. Den andra är, vid otillräckligt sökresultat, att utvidga sökningen så att mängden relevant information ökar utan att få en mängd av ej relevant information. Detta kan uppnås genom att minska antalet sökord och kan även kombineras med hjälp av operatorm AND. Om de ursprungliga sökvillkoren inte ger önskat resultat, kan man byta ut dem mot synonymer, söka på en fras eller använda andra sökfunktioner.

Hock (2001) skriver om precision och täckning som används i samband med mätning av kapaciteten hos system där information har lagrats. Hock (2001) definierar precision som måttet på i vilken grad en samling dokument motsvarar sökningens önskemål, d v s det är proportion till mängden relevanta hämtade poster och den totala mängden hämtade poster. T ex, om en fråga genererade 10 poster och 4 av dessa är relevanta, då är precisionen 40%. Enligt Hock (2001) mäter täckning den graden i vilken en sökning lyckats hitta alla relevanta dokument i en samling, d v s det är proportion till mängden relevanta hämtade poster och det totala antalet relevanta poster i en databas. Till exempel, om en fråga får fram 5 relevanta poster när det finns 10 relevanta poster i databasen, blir täckningen för denna sökning 50%. Täckning är komplicerat att mäta då det för närvarande är i det närmaste omöjligt att uppskatta det totala antalet relevanta poster.

Om sökning resulterar i många dokument i resultatlistan men där få av dem är relevanta kallas den *låg precision/hög täckning*. Ett resultat med många relevanta men få dokument kallas *hög precision/låg täckning*. Det idealiska resultatet är där alla relevanta dokument finns med vilket kallas *hög precision/hög täckning*. Om täckningen är noll, är precisionen odefinierbar (Ackerman m fl, 2001)

Precision har en tendens att öka och täckning minska när en sökning begränsas för att reducera sökresultatet. Den motsatta tendensen ökar täckningen men minskar precisionen när en sökning breddas för att utöka sökresultatet. (Large m fl, 2001)

Vid sökning i kataloger kan man välja mellan att klicka sig fram bland kategorier och underkategorier eller att skriva in sökord i en textruta, som hos den indexerade söktjänsten. En ämnesindelad söktjänst har inte en lika stor databas som en indexerad, men fördelen med

den är att chansen ökar att få relevanta träffar om man söker inom ett inte alltför ovanligt ämne. Till exempel om man söker på matlagning handlar alla returnerade sidor om just det. Är man däremot ute efter en specifik person som har att göra med matlagning, kan det vara en bättre ide att vända sig till en indexerad söktjänst.

Värdering av information kan ske på olika plan, t ex på ett externt plan då man bedömer en källas äkthet eller ett internt då man bedömer innehållet och hur pass intressant det är för ens syfte. Det är också viktigt hur aktuell den information är som man tänkt använda sig av. Om den baseras på gamla kunskaper så kan den i vissa fall visa sig mindre värdefull. Uppgifter om t ex informationsteknik och Internet ställer mycket höga krav på aktualitet. All information oavsett om den är i tryckt form eller publicerad på Internet behöver värderas både vad gäller sanningshalt och trovärdighet.

Jag avslutar teorikapitlet med högskolebibliotekariens Agneta Lantz färdighetslista för informationskompetens som presenterades vid NVBFs konferens i Linköping 1996, hämtad ur Båge m fl (2003):

1. Utforska och identifiera problemet/behovet av information relaterat till tidigare kunskap
2. Formulera sökfrågan
3. Identifiera nyckelord, begrepp etc
4. Identifiera lämpliga sökhjälpmedel och resurser
5. Utveckla lämpliga sökstrategier för att utnyttja dessa
6. Söka efter och lokalisera relevant information
7. Värdera, välja och organisera den funna informationen
8. Analysera, tolka och dra slutsatser
9. Fastställa hur informationen skall användas, presenteras, kommuniceras, organisera den för tänkt användning, tillämpa informationen
10. Utvärdera resultatet; utvärdera processen

## 5 Undersökning och resultat

### 5.1 Observation

Anledningen till att genomföra undersökningen var att genom observation av användaren under sökprocessen skaffa ett bättre underlag till en enkät. Två frivilliga deltagare deltog vid två separata tillfällen i observationen. Båda deltagare var över 18 år och Internet användare. Deltagare 1 har IT-relaterat yrke, deltagare 2 har inte IT-relaterat yrke. Observationer genomfördes i respektive deltagares hemmiljö, var icke strukturerad och bestod av att se hur deltagarna sökte information på Internet efter eget behov. Observationerna spelades in på video. Under varje observation kommenterade deltagarna högt, så kallat "think aloud method". Deltagarna sökte information efter deras informationsbehov, men vilken exakt information som söktes kommer inte att nämnas och analyseras mer än att förekommande informationsproblem ska klassificeras till någon typ av informationsproblem enligt Ingwersen och Wormell (1998) (se teoriavsnitt 4.2.1).

Observationsbeskrivning finns i bilaga 1.

#### **Resultat**

Under observationerna såg jag skillnader i två sökprocesser. Dessa skillnader kan bero på att det var två olika personer som hade bland annat olika bakgrund, IT vana, informationsöknings-erfarenhet och så vidare. Sökprocessen ser också annorlunda ut beroende på den information som användaren vill ha. Jag noterade att söktjänsten Google var båda deltagarnas första val söktjänst.

Även om de informationssökningar som observerades var korta, kunde faser som finns med i Kuhlthaus ISP-modell identifieras i varje sökning. I deltagare 1:s första sökning var alla sex faserna med. Deltagare 1:s första sökning liknade mest Limbergs kategori C (informationssökning som att granska och analysera), medan de andra informationssökningarna av deltagare 1 och informationssökning av deltagare 2 liknade Limbergs kategori A (informationssökning som att söka fakta). Varje deltagare har i sin egen informationssökning ett mönster och sättet att söka påverkas av synen på vad man ska ha informationen till och det finns ett starkt samband mellan informationssökning och informationens innehåll, vilket gör det svårare enligt Limberg att kategorisera under någon modell/kategori.

Irritationsmomenten var irrelevanta träffar, för många träffar, sidor som är med i träfflista, men inte längre är tillgängliga.

Jag har inte analyserat deltagarnas sökprocess, deras val och vilka faktorer som låg till grund för deras val under sökprocessen. Däremot har observationerna hjälpt till att skapa enkätfrågor.

## 5.2 Enkäter

Genom enkätfrågor ville jag få information om personernas användning av Internets söktjänster samt fånga deras upplevelser om sökprocessen och deras eventuella önskemål. Jag ser resultaten mer som ett indikationsvärde om användarnas upplevelser, problem och önskemål än att enkäten har ett statistiskt värde.

Enkäten innehåller olika frågor: frågor med möjlighet att välja ett eller flera alternativ, frågor som har öppna alternativ där respondenten kan skriva själv och förklara. Enkätfrågorna formulerades så att de inte blev ledande för att få mer spontana svar. Det fanns även ett antal påståenden där respondenten fick ta ställning och välja alternativ.

Ett antal påståenden syftade till att få fram vilka alternativ användaren väljer i sökprocessen och vilka sökmetoder som används samt hur sökresultaten hanteras. Påståendena formulerades med fastställda svarsalternativ för att få mätbara enkätsvar och för att minimera tvetydigheter. I den fråga där jag vill att personerna tar ställning till påståendena har en skala utan medelalternativ avsiktligt använts. I en annan fråga har tre alternativ med medelalternativ använts. Jag använde ordet söktjänster konsekvent genom hela enkäten för att undvika tvetydlighet. Enkätformulär - se bilaga 2.

Jag delade ut 30 enkäter till personer med varierande kunskap om informationssökning på Internet. Av dessa svarade 28 användare, och två svarade inte eftersom dem ansåg sig inte ha tid för det. Jag anser att det föreligger en bra svarsfrekvens som ger tillräckligt underlag för min undersökning.

Enkätsammanställning finns i bilaga 3.

### **Resultat.**

Modern informationssökning på Internet är inte lika lätt för alla. Respondenterna som svarade på enkätfrågor var av olika ålder, olika yrkeskategorier, vilket kan förklara deras olika upplevelser och uppfattningar om informationssökning. Trots att alla respondenter söker information på nätet dagligen eller i vart fall några gånger per vecka upplever lite mindre än hälften av dem informationssökning på Internet som svårt och 11 av 28 uppgav att söktjänster inte underlättar sökningen. Mycket få av respondenterna hade något konkret förbättringsförslag. Hur man lyckas med informationssökning och därefter upplever den kan bero på många faktorer: vad för information man söker, kunskap om olika söktjänster, erfarenhet av hur man söker information på Internet och t ex formulerar och avgränsar sökfrågan.

Alla respondenter använder sökmotorer för informationsökning och hälften använder dessutom länksamling. Ingen använder sig av QA (Question Answering tjänst). Det visar att sökmotorer är det mest använda verktyget för informationssökning på nätet. Den information som mest söks på Internet är först och främst text. Många har någon gång sökt efter bilder med hjälp av sökmotorer, betydligt färre sökte efter ljudfiler och ingen av respondenter sökte efter video.

Alla respondenterna använde Google och de flesta använde denna söktjänst oftast. Eniro väljs också ofta av respondenterna. Andra söktjänster som respondenterna ofta väljer är Yahoo!, MSN Search, ELIN, lite mindre AltaVista och LIBRIS. Söktjänsterna Ask Jeeves, Teoma,



Open Directory Project är antingen helt okända eller väljs aldrig av respondenterna. Söktjänsterna pubmed, hitta, VEGA, Malin användes av någon respondent. Tabellen visar resultat på enkätens fråga om kännedom av olika söktjänster:

	Använder oftast	Använder ibland	Använder sällan	Använder aldrig	Känner inte till	Helhetsbetyg av söktjänsterna (1=mycket dåligt, 5=mycket bra)
AllTheWeb	0	0	1	0	27	3
AltaVista	0	5	10	13	0	2,3,4
AOL Search	0	0	1	14	13	0
Ask Jeeves	0	0	0	4	24	0
Eniro	16	10	2	0	0	3,4,5
Exite	0	0	0	27	1	0
Google	23	4	1	0	0	3,4,5
MSN Search	5	6	7	6	4	3,4,5
Teoma	0	0	0	10	18	0
Yahoo	4	10	7	7	0	1,2,3,4,5
WWW virtual Library	0	3	6	8	11	3
BUBL LINK	0	0	1	10	17	0
Open Directory Project	0	0	0	0	28	
SUNET	2	1	4	11	10	2,3,4
ELIN	4	10	4	5	5	3,4
LIBRIS	0	6	7	10	5	2,3

I sökresultatets träfflista tittar många respondenter främst på de tre-fyra översta träffarna och förutsätter att de som är överst även är relevantast. Tillräckligt många granskar även adresserna för att se om resultatet verkar seriöst, de går ytterst sällan till nästa sida med träffar och ändrar inte sökningen förrän flera resultat granskats. Färre respondenter modifierar oftare sökningen än scollar nedåt i träfflistan. Vid missnöje med sökresultatet valde respondenterna oftast att avgränsa frågan, eller omformulera frågan. Andra alternativ som att ge upp, att byta söktjänst, att maila frågan till en kamrat – kollega valdes inte ofta av respondenterna.

Användarnas förbättringsförslag handlade mest om presentation av resultat som t ex ikon före träff som visar typ av hemsida (universitet, privat osv.), att kunna filtrera vilken typ av information som kommer upp på listan, experter skulle kunna granska och rangordna hemsidor – men man bör kunna välja bort denna funktion, kategorisering av resultat svar enligt användarens önskemål (artiklar, psykologi, osv.).

Resultatet av enkäten hjälpte mig att förbereda mina intervjufrågor om söktjänsterna.

### **5.3 Intervju**

Syftet var att få svar på fördjupade frågor, baserade på svaren i enkätundersökningen, och därigenom få fram användarnas upplevelse av informationssökning på Internet och utförliga förslag för framtida utveckling av söktjänster. Alla respondenter som har ingått i denna studie använder Internet för informationssökning. Varje intervjuperson har svarat på 33 frågor i 4 kategorier: Om söktjänster, Om sökfrågor, Information, Generella/övriga frågor. Ibland ställdes följdfrågor för att förtydliga respondentens svar.

Intervjufrågor - se bilaga 4.

I inledningen av mina intervjuer redogjorde jag för syftet med min studie. Varje intervju tog ungefär 60-80 minuter. Resultaten av intervjuerna presenteras i en sammanställning av svaren.

Intervjusammanställning finns i bilaga 5.

#### **Resultat**

Alla respondenter angav Google som sitt första val av söktjänst, p g a att den är enkel, lättbegriplig, fungerar ofta tillfredställande, har bra rykte som t ex att den är störst, effektiv osv. Men respondenterna varken vet eller ville veta någonting om tekniken bakom Google. Användarna har oftast inga strukturella mentala modeller (Preece, 1994), men de använder funktionella mentala modeller av söktjänster. Det kan vara anledningen till att användarna bl a uppskattar och föredrar enkelt och tydligt gränssnitt där det är lätt att förstå hur den skulle användas. Förutom Google använder respondenterna nischsöktjänster för speciella sökningar, ingen som använder QA. Respondenterna har angett att de även skulle vilja prova en ny söktjänst om de hört att den var bra, men de är inte villiga att betala för en söktjänst annat än i särskilda fall. Respondenterna har olika uppfattning i integritetsfrågor.

Under intervjuerna kom det fram att de flesta inte ansåg att de hade någon plan eller strategi när de började en sökning i söktjänsterna. Efter att ha tänkt ut några sökord provade man vilket resultatet blev och först om man fick massor av träffar eller inga alls tänkte man lite närmare över hur man skulle gå tillväga. En anledning till detta är säkert att vanliga korta frågor inte kräver alltför mycket planering. Alla respondenterna använde avancerad sökning, men inte booleska operatorer. Alla utom en person har sagt att de vanligen fick för många träffar när de sökte. Användarna började ofta med att skriva enkla sökfrågor med ett, två eller tre ord som uttrycker det övergripande ämnets (s k etiketteffekten - Ingwersen och Wormell, 1989), vilket kan resultera i mängd av irrelevanta resultat. Men trots det uppgav respondenterna att de inte vill spendera tid på att lära sig sökspråk som skulle kunna förbättra deras sökning genom t ex bättre begränsning för att få fram relevanta dokument.

Respondenterna hade svårt att beskriva vilka kriterier de använde när de bedömde relevans och trovärdighet men kunde ge några ledtrådar, som domännamn, webbplatsadress, även sidornas layout, textens längd och innehåll. Respondenterna uppgav att de söker även efter information om lokala verksamheter och evenemang och det fungerar generellt sett bra. Negativt var att små verksamheter inte alltid har sin information på Internet eller att den kan vara sämre uppdaterad. Inte alla respondenter var lika intresserade av att få veta hur informationen filtreras och om informationen är sponsrad och inte alla visste om de ville kunna påverka hur resultatet skulle rankas.

Användarna ville ha snabb och lätt tillgång till information. Respondenterna uppgav att informationssökning på Internet är viktigt, men kan kännas både spännande och arbetsamt beroende på hur sökningen går till. De oroade sig för att missa viktigt information när de fick stor resultatlista som man inte kunde gå igenom, om informationens trovärdighet och aktualitet. Stora mängder av resultat som är omöjliga att hantera, dålig uppdatering, fel typ av information irriterar användarna. Som lösningar föreslog respondenterna oftare uppdatering, möjlighet att välja vilken typ av information som önskas (t ex artikel, försäljning), extrahera en del av informationen till informationsdatabas så att man kan läsa sammanställd information utan att besöka många olika sidor, utveckla Internet så att den skulle fungera logiskt likt en mänsklig hjärna (semantikwebb). Respondenterna önskade att kunna sätta filter för olämplig information till barn. Ingen av respondenterna har använt någon hjälpsida till en söktjänst.

## 6 Diskussion

*"Den som letar efter information eller tjänster på en sökmotor ska hitta den hemsida som bäst motsvarar sökningen. Är det alltid så?" Capotondi*

Genom Internet och utvecklandet av webben kan i princip alla nuförtiden få tillgång till den information som finns lagrad där, om man vet hur man ska få fram denna information. Söktjänster är en av de viktigaste komponenterna man har på Internet. Det spelar ingen roll hur många miljoner sidor man har publicerat där om ingen hittar fram till dem. Söktjänstindustrin har stor konkurrens och varje företag försöker höja precision och användbarhet på sin söktjänst för att öka populariteten och trafiken genom den. Komplexitet av söktjänstteknologi gör det svårt för vanliga användare att se vad som saknas i ett sökresultat. Användare vet inte om sidan är den lämpligaste eller om någon annan sida är bättre. Därför fortsätter sannolikt användare att använda den söktjänst som de redan använder. Man har inte heller mycket tid att hålla sig ajour med alla nya finesser och nya söktjänster som utvecklas.

Meningen med hela denna undersökning var att få reda på användares upplevelser och önskemål avseende informationsökning på Internet med hjälp av söktjänster, vilket är ett viktigt steg mot förverkligande av Heideggers teknikfilosofi där teknik bättre skall anpassas till människor. Syfte var att studera vilken sökteknik som används och hur vanliga användare upplever sökprocessen.

## 7 Slutsatser

### 7.1 Svar på frågeställningar:

***Hur är tjänster för informationsökning på Internet designade (uppbyggda) och hur fungerar de?***

Svaret på denna fråga finns i kapitel fyra där jag redogör för olika sökmodeller, sökteknik samtidigt för indexerade, katalogiserade och metasöktjänster på ett generellt plan. Jag lyfter fram deras respektive förtjänster och begränsningar.

***Vilken erfarenhet har användarna av sökprocessen och vad är deras önskemål dvs hur borde en perfekt söktjänst vara?***

De som deltog i min undersökning var personer som söker upp information på Internet regelbundet. De flesta använder Google i första hand och utnyttjar både enkla och avancerade sökfält, skriver in några ord, letar bland de 3 till 20 första länkarna och tittar på sammanfattningstexten för varje länk och får oftast fram exakt det som eftersöktes. Om inte, avgränsar eller omformulerar de sökningen och ibland byter de söktjänst. De anser att söktjänster uppfyller deras förväntningar och behov, trots att de inte alltid var helt nöjda. En av orsakerna till det är att sökmotorn returnerar irrelevant och betydelselös information. Även om man söker med avancerad sökning så kommer man att få en ganska stor mängd irrelevant

information. Informationen kan vara både svåröverskådlig, svår att förstå och av varierande kvalitet. Det krävs att man kan sortera, vilket är oerhört tidskrävande.

Enligt Kuhlthau är osäkerhet en naturlig del av allt informationsinsamlande och därigenom även processen med användandet av söktjänsterna. Min intervjuundersökning har bekräftat att det finns en osäkerhet om hur man använder söktjänsterna, och att man ibland är missnöjd med resultaten av sina sökningar, trots att alla anser sig vara kompetenta att söka information. Under intervjuerna kom det fram att de flesta inte ansåg att de hade någon särskild strategi när de började en informationssökning med söktjänsterna. Efter att ha tänkt ut några sökord prövade man vilket resultatet det blev och först om man fick många träffar eller inga alls tänkte man över hur man lite närmare skulle gå tillväga. En anledning till detta är säkert att vanliga korta frågor inte behöver alltför mycket planering.

Webben är fortfarande under utveckling och dess brister påverkar människors informationssökning. Ett stort problem i informationssökning är att söktjänster inte kan sluta sig till kontexten av human kommunikation på samma sätt som människor kan göra det. Den ideala söktjänsten skulle förstå en fråga även om den var dåligt formulerad och direkt returnera den till en perfekt källa. Moderna teknologier kan inte uppnå detta mål. Utmaningen i söktjänstens design är att konstruera en modell som genererar relevant och hanterligt resultat som svar på välformulerade frågor. Olika användare kan ha olika intentioner med samma handling i ett system, vilket kan bero på användarens bakgrund, erfarenheter och/eller preferenser.

Det är viktigt att användare får intryck av att söktjänsten är lättbegriplig och lätt att använda. En perfekt söktjänst skulle förstå användarens naturliga språk så att den förstår vad frågan handlar om (t ex om användaren söker en viss information om europeiska länder skulle information om t ex Sverige, Danmark, Frankrike komma med). Söktjänsten skulle fungera som flaskhals för att reducera informationsflöden till hanterbar volym. Den ska vara gratis för användarna och vara designad så att de inte skulle behöva lära sig något särskilt sökspråk. Användarna borde ha möjlighet att få veta om länken är sponsrad, att välja hur resultatlistan på sidorna ska sorteras och få returnerad information placerad under olika ämnesmappar.

### ***Vad vet användarna om olika söktjänster och vad är användarnas syn på dem?***

Googles framgång visar att användarna inte vill ha en söktjänst med många förvirrande moment. Många väljer Google eftersom den är lättbegriplig, lätt att använda för de flesta, har stöd för många olika språk och är känd för att ha den största databasen. Googles matriser för att indexera data är mer relevanta och ger snabbare sökning med bättre träffsäkerhet. Förutom Google var respondenterna bekanta med och använde även andra söktjänster (Eniro, Yahoo!, MSN Search, ELIN, lite mindre AltaVista och LIBRIS). Användarna får information om vilken söktjänst som är bäst utifrån media och ifrån bekanta. De väljer och besöker en söktjänst gång efter gång bara om de är nöjda med resultatet.

Det visar sig att respondenterna inte känner till vissa typer av söktjänster som t ex QA (Ask Jeeves). Vissa användare har ett behov av att kunna söka efter bred och grundläggande information i olika ämnen via ämnesportal som t.ex. Yahoo!.

## 7.2 Avslutande sammanfattning

Sökmotortekniken har utvecklats enormt under de senaste åren, vilket man kan se på Internet där Google mer eller mindre har utklassat Alta Vista. Av detta kan man däremot inte dra slutsatsen att Google alltid skulle vara bättre. Om man använder Internet som informationskälla så bör man lära sig hur ett antal sökverktyg fungerar för att kunna få ut mesta möjliga av deras funktioner. Om man söker bred och grundläggande information om ett ämne kan det vara lämpligast att söka via en webbkatalog. Specifik information söks bäst genom en söktjänst.

Ur en söktjänstutvecklarens perspektiv är det svårt att förutsäga användarnas beteenden, men tekniken byggs så att den ska passa så många som möjligt. Jag tycker att det finns mycket stora möjligheter för vidareutveckling av söktjänster och även att skapa nya funktioner.

Då det kan vara svårt för oerfarna användare att formulera en bra sökfråga skulle det vara bra om en söktjänst lätt kunde förstå och översätta användarens önskemål. Genom intuitiva sökmotorer skulle sökresultatet kunna förbättras samt irrelevant och ostrukturerad data bortfiltreras utan att användaren behöver vara medveten om detta. Ett annat sätt kan vara avgränsning av en fråga med hjälp av sökvisard (eller sökguide) Exempel är: fråga om man ska inkludera synonymer, välja bort betydelse av ord (element –fysisk, element värme), inkludera några andra språk, välja typer av dokument som ska vara med och hur de skulle kunna sorteras osv. Den skulle kunna utvecklas som ett "Conversational" datasystem (Dourish, 2001) som är nära det mänskliga språket eftersom användare inte vill investera tid och kraft i att lära sig ett speciellt sökspråk. En söktjänst kan även kombineras med QA-system så att resultatet skulle kunna returneras i form av svar och inte bara stor mängd av dokumentlänkar. Ask Jeeves erbjuder redan möjlighet att ställa frågor, men det antal frågor man kan få svar på är fortfarande väldigt begränsade och frågor kan ställas endast på engelska.

För de som e-handlar skulle det vara bra med en vidare utveckling av comparison shopping söktjänster som skulle kunna jämföra varor och tjänster inte bara till pris utan också andra parameter som t ex rabatt, returrätt och även utveckling av en trådlös shoppingsöktjänst. Framtida söksystem skulle kunna ge användarna möjlighet att definiera olika roller för dem själva och specificera sina intressen för att sökresultatet skall bli maximalt anpassat till varje användare.

Användare och söktjänster har samma mål, att undvika spam och operera i intuitivt interface design. Huvudmål för en kommersiell söktjänst är ekonomisk vinst. De flesta användarna vill inte betala för söktjänstservice och därför måste den söktjänst som vill attrahera många användare vara gratis. Söktjänsterna får betalt när den dirigerar användares uppmärksamhet till aktörer som betalar för trafik som skickas till dem. Användare däremot har som huvudmål att hitta sida/information som är mest relevant till en ställd fråga. Kanske ska staten sponsra en icke kommersiell söktjänst, som huvudsakligen inte fokuserar på vinst, utan på unbiased informationssökning. En sådan tjänst skulle kunna finansieras genom skatter eller publika sponsorer.

Kommersiella söktjänster håller sin programkod i stor sekretess. Öppen kod kan vara ett alternativ i utveckling av söktjänstteknik. Linux framgång visar att iden med öppen kod har livskraft där manipulationer är mindre sannolikt, eftersom rankingteknik skulle kontrolleras av en "community" av utvecklare. En publicerad rankingalgoritm skulle i detta fall minska bl

a informationens asymmetri mot användarna. Reklam kan markeras som sådant och användas för att finansiera en index-infrastruktur så att man kan konkurrera med stora söktjänster.

Själva webben kan bli mer intelligent och intuitivt tillfredställa användares behov (semantisk webb).

Mängden av information på webben leder till att användaren inte kan ta till sig allt potentiellt relevant material och besöka alla potentiellt relevanta webbsidor. Man vill oftast ha tillförlitliga källor därför sortering av informationskollektion med till ex strukturering och informationlabeling (markering) kunde vara bra.

När söktjänsten förlorar användare och trafiken minskar leder det till inkomstminskning för söktjänsten. Å andra sidan om någon söktjänst har vunnit de flesta användarna betyder det inte bara ekonomisk vinst utan också makt för söktjänsten att dirigera användarnas uppmärksamhet. Media skulle bättre övervaka olika söktjänster i deras roll som mäktiga gatekeepers. Användarna är inte intresserade eller förstår inte problemen i koncentration av söktjänstmarknaden, media borde därför försöka öka förståelsen härför.

Jag avslutar mitt arbete med ord från Sellen med flera (2002) om att Internets ”true potential is still yet to be realized”.

### **7.3 Förslag på vidare forskning**

Denna studie började med några frågor men slutade även med fler. Under arbetets gång har det växt fram ett flertal olika frågetecken som man önskade att få svar på, men som det av naturliga skäl inte fanns möjlighet att ägna sig åt under arbetets gång. En frågeställning som inte alls har behandlats i uppsatsen är hur användarna av systemet tar till sig information. Till exempel är det möjligt att urskilja mönster för i vilken ordning som användarna tenderar att besöka olika referenser. Kan man ha nytta av sådana mönster eller borde man bryta ut dem?

Det kan bli intressant med vidare undersökning av sambandet mellan val av sökverktyg och sökresultat. En annan fråga man kan studera mer är vilka sökstrategier är bäst lämpade för informationssökning på Internet.

## 8 Terminologi

**Booleska operatorer** – vanliga operatorer är AND, OR, NOT. De kan kombineras med sökord på olika sätt för att skapa mer komplicerade sökuttryck.

**Hypertext eller hyperlänk** – koppling från en viss plats (en symbol, en bild eller ett markerat ord) på en webbsida till en annan plats på samma eller annan webbsida.

**Internet** - ett globalt nät bestående av en stor mängd lokala datornätverk, där varje nät består av ett varierande antal enskilda datorer.

**Metadata** – betyder data om data, vanligen hur en databas är organiserad. Informationen etiketteras så att det blir lättare att hitta den.

**Söktjänst eller sökverktyg** – engelsk alternativ är Web Search Engines (WSEs), ”search engine”, ”search service”, ”search tool” - tjänst som erbjuds på en webbplats och som är inriktad på att tillhandahålla sökmöjligheter i text på webbsidor och i meddelanden i diskussionsgrupper.  
(Svenska datatermgruppens ordlista).

**Sökmotor** - program för indexering av och sökning i stora textmassor, t.ex. samtliga webbsidor på Internet. Flera olika söktjänster kan använda sig av samma sökmotor.

**Stoppord** – ord som förekommer så ofta i databaser att dem inte meningsfulla som söktermer (ex Internet), eller mycket vanliga ord såsom *och*, *den* etc Man utesluter dem från databas dels för att spara på utrymmet i databasen och dels för att göra själva söktiden kortare.

**Sökfråga** – de sökord man skriver in i sökformuläret i en söktjänst.

**Träfflista** – den lista över träffar, resultat, som kommer fram när man genomfört en sökning i en söktjänst.

**URL** – Universal Resource Locator, användas för att beskriva plats och typ av information på Internet.

**Webben eller www** – (World Wide Web) den gigantiska grafiska informationsdatabasen på Internet.

**Webbplats** – flera webbsidor som är hoplänkade med varandra med en gemensam upphovsman (i form av en person, ett företag osv.) och med en gemensam ingångssida.

**Webbsida** – Sida på World Wide Web. Via länkar är de direkt eller indirekt kopplade till alla andra sidor i www.

### 8.1 Förkortningar

AI – artificial intelligence



API – application program interface  
DMOZ – open directory project  
DSL – digital subscriber line  
HITS – hypertext induced topic search  
HTML – Hyper Text Markup Language.  
HTTP – HyperText Transfer Protocol.  
IA – intermediary agent  
IDF – inverse document frequency  
IE – information extraction  
IPO – initial public offering  
IR – information retrieval  
IPO – initial public offering  
NLP – natural language processing  
OCR – optical character recognition  
QA – question answering  
SUNET – Swedish University Network, gemensamt datornätverk för svenska universiteten och högskolor.  
TF – term frequency  
UI - user interface  
URL – uniform resource locator  
URL – uniform resource locator  
WWW – world wide web

## 9 Litteratur & övrigt material

Ackerman, E, Hartman, K. (2001); The information searcher's guide to searching & researching on the Internet & the world wide web, 2nd edition; Franklin, Beedle & Association Inc., USA,

Aguilar, F. J. (1967); Scanning the Business Environment. New York, NY: Macmillan Co.

Ahmstedt, P. (2003); Sju populära sökmotorer - flest träffar inte alltid bäst; PC För Alla, nr 3, 2003

Andersson, B-E. (2001). Som man frågar får man svar. Prisma, Stockholm

Backman, J. (1998); Rapporter och uppsatser. Studentlitteratur.

Baeza-Yates, R., Ribeiro-Neto, B. (1999); Modern information retrieval, Addison-Wesley

Baskaran, S. (2005); AU-KBC Research Centre; [http://www.au-kbc.org/research\\_areas/nlp/projects/sengine.html](http://www.au-kbc.org/research_areas/nlp/projects/sengine.html), laddat ner 2006.05.10

Berners-Lee, T m fl (2001); The semantic Web;  
<http://www.sciam.com/articlecfm?articleID=00048144-10D2-1C70-84A980EC588EF21>

Bjurbäck, T. (1998); Informationssökning på Internet en handledning;  
<http://www.thn.fhsk.se/thomas/search/default.htm>; laddat ner 2006.06.05

Brin, S.; Page, L. (1998); The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, 30 (1-7), 107 - 117.  
<http://www-db.stanford.edu/pub/papers/google.pdf>; laddat ner 2006.05.10

Bryman, A. (1997); Kvantitet och kvalitet i Samhällsvetenskaplig forskning; Studentlitteratur, Lund.

Båge, S., Ekelund, L. (2003) Informationskompetens – färdighet eller insikt? En jämförelse mellan två modeller över begreppet informationskompetens. BIVIL:s skriftserie 2003:4 ISSN 1401-2375.

CS1, ComputerSweden nr 75, 2005.08.12

Dourish, P. (2001); Where the action is – the foundations of embodied interaction; Massachusetts institute of technology

Dreyfus, H.L. (2001); Being-in-the-world – a commentary on Heidegger's Being and time, Division 1; Massachusetts institute of technology

Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D; A (2002); Search Engine for 3D Models; <http://www.cs.princeton.edu/~funk/tog03.pdf>

Glover, E.J., Lawrence, S., Gordon, M.D., Birmingham, W.P., Lee Giles, C. (2000); Web Search - Your Way. In Communications of the ACM.

Gustavsson, M. (2001); Hubert Dreyfus syn på AI - en kritik utifrån fenomenologisk teori Linköpings universitet Institutionen för datavetenskap HKGGB0, Artificiell intelligens, ht-01, <http://130.236.177.26/~HKGGB0/studentpapper-01/Mattias-Gustavsson.pdf>; laddat ner 2006.05.10

Hock, R. E. (2001); The Extreme Searcher's Guide to Web Search Engines, 2nd edition; CyberAge Books

Ingwersen, P. & Wormell, I. (1989); Modern indexing and retrieval techniques matching different types of information needs, i Information, knowledge, evolution (proceeding of the 44th FID Congress), Amsterdam: Elsevier Science Publishers.

Jakobsson, P., Lindskog, S. (2002); Användaren som resurs vid informationssökning; Report 02076; Växjö University ISSN 1650-2647; SE-351 95 VÄXJÖ ISRN VXU/MSI/IV/E/--02076/--SE

Koch, T, Ardö, A., Brümmer, A., Lundberg, S.(1996); The building and maintenance of robot based internet search services: A review of current indexing and data collection methods, <http://www.lub.lu.se/desire/radar/reports/D3.11/>; laddat ner 2006.05.22

Koch, T. (1996). Internet search services. Rapport, Lund Universitet.  
<http://www.lub.lu.se/tk/demos/DO9603-meng.html>; laddat ner 2006.05.10

Kvale, S (1997), Den kvalitativa forskningsintervjun, Studentlitteratur, Lund

Kwok, C et al (2001), Scaling Question Answering to the web, in: ACM Transaction on information systems

Langville, A. N., Meyer, C. D. (2005); Information Retrieval and Web Search

Large, A., Tedd, L.A., Hartley, R.J. (2001); Information Seeking in the Online Age - Principles and Practice; K.G. Saur Verlag, München,

Limberg, L., Hultgren, F., Jarneving, B. (2002); Informationssökning och lärande – en forskningsöversikt; Form & Tryck: Lenanders tryckeri AB, Kalmar

Miles, M. B., Huberman, A. M. (1994); Qualitative Data Analysis, Sage Publications

Page, L., Brin, S., Motwani, R., Winograd, T. (1998) The Pagerank. Citation Ranking: Bringing Order to the Web. Technical report, Stanford University

Paulson, J. (2003); Fast Facts About Froogle; <http://www.sitepoint.com/print/1060>; laddat ner 2006.07.10

Pharoa, N., Järvelin, K. (2002); The SST method: a tool for analysing Web information search processes; [http://www.info.uta.fi/tutkimus/fire/archive/sst\\_method\\_ipm.pdf](http://www.info.uta.fi/tutkimus/fire/archive/sst_method_ipm.pdf); laddat ner 2006.05.24

Preece, J (1994); Human-computer interaction; addison Wesley

Sellen, A. J., Murphy R., Shaw, K. L.; (2002); How knowledge workers use the web,

Shapira m fl. (1997); Stereotypes in information filtering systems. Information Processing & Management,

Sherman, C. (2005); Search Engine Results Continuing to Diverge; Associate Editor, August 2, 2005; <http://searchenginewatch.com/searchday/article.php/3524411>; laddat ner 2006.05.24

Sullivan, D. (2004, April 28); The Major Search Engines and Directories; Search Engine Watch; <http://www.searchenginewatch.com/links/article.php/2156221>, laddat ner 2006.05.24

Sullivan, D. (2004); <http://searchenginewatch.com/article.php/2156221>; laddat ner 2006.05.24

Takabayashi, S. (2004); Synthetic Assistance for Creation and Communication of Information; Nara Institute of Science and Technology

Taibi, D.; Gentile, M., Seta, L (2005); A Semantic Search Engine for Learning Resources - Recent Research Developments in Learning Technologies

Van Rijsbergen, C.J. (1979). *Information Retrieval*. Butterworths, <http://www.dcs.gla.ac.uk/Keith/Preface.html>; laddat ner 2006.05.29

Våge, L.; Dalianis, H.; Iselid, L. (2003); Informationsökning på Internet; studentlitteratur

Våge, L. (2004); Vad händer med sökmotorerna? Vad blir utrymmet för bibliotekens vägledningstjänster?; Stockholms Universitetsbibliotek; <http://www.sub.su.se/aktuellt/konferenser/sokguider/LV.pdf>; laddat ner 2006.09.11

Wallén, G. (1996). Vetenskapsteori och forskningsmetodik, Studentlitteratur

Alfred Glossbrenner; <http://www.quotationsbook.com/subjects/631/Information>

AltaVista; <http://www.altavista.com/about/>, 2005-07-11

AskJeeves; <http://sp.ask.com/docs/about/index.html>

Google - företagsprofil; <http://www.google.com/intl/sv/profile.html>, 2003-04-11

Google; <http://www.google.com/profile.html> 2005-07-12

Googles specialfunktioner; <http://www.google.com/features.html> 2005-07-12

Yahoo!; <http://help.yahoo.com/>; <http://docs.yahoo.com/info/terms/>

Avi Rappoport; Search Engine Watch; SearchDay - Anatomy of a Search Engine: Inside Google; <http://www.searchenginewatch.com/searchday/article.php/2161091>, 30 October 2002

Search Engine Watch; SearchDay - Anatomy of a Search Engine: Inside Google <http://searchenginewatch.com/searchday/02/sd1030-in-google.html>, 2003-03-18

Search Engine Watch; Searches Per Day <http://www.searchenginewatch.com/reports/perday.html>, 2003-03-15

MSLN Connection, How Search Engines Work <http://www.mstf.org/~usmmsln/work.htm>

Open Directory; <http://dmoz.org/about.html>

<http://www.searchenginecolossus.com/>

Wondir; <http://www.wondir.com/wondir/jsp/index.jsp>

Samuel Johnson; <http://www.u-blog.net/joel>

Capotondi, R. Vad är en bra hemsida enligt MSN search; [http://www.capotondi.se/internetmarknadsforing\\_2.shtml](http://www.capotondi.se/internetmarknadsforing_2.shtml); downloaded 06 juni 2005

## 10 Bilagor

### **Bilaga 1**

#### **Observation 1**

Observation av deltagare 1 tog 20 minuter, två informationssökningar genomfördes, efter observationen av ställde jag några frågor till deltagare 1.

Första informationsökningen började med att deltagare 1 hade ett eget uppdrag att genomföra, men hade en brist i sin kunskapsbas som personen behövde lösa med informationsökning (*Kuhlthau - task initiation*). Informationen som då söktes var text (ingen bild, ljud eller video) och specifikt fokuserad ämnesfråga. För att formulera sökfrågan (*Kuhlthau - topic selection*) valdes först två ord (etiketteffekten - Ingwersen m fl, 1989). Deltagaren började med att välja söktjänsten Google (enkel sökning). Deltagaren fick många returnerade sidor i svar och har besökt de 4 första träffarna där fanns inte den information som deltagare 1 ville ha. Personen läste titlar och beskrivningar och granskade även adresser längre ner i träfflistan, kände sig frustrerad och förtvivlad (*Kuhlthau - prefocus exploration*). Sedan deltagare 1 tänker efter specificerades sökfrågan mer (*Kuhlthau -focus formulation*) med hjälp av Google avancerad sökning. Då fick personen återigen alldeles för stort svar i retur, gick personen till de 2 första av de returnerade svaren, tittade igenom träffarna lägre i träfflistan (läste titlar, beskrivningar, granskade även adresser) och var fortfarande missnöjd med information. Personen gick över till söktjänsten ELIN (*Kuhlthau -focus formulation*), fick många returnerade dokument i svar och redan första träffen innehöll exakt den information som personen letade efter så denne blev nöjd med resultatet, hämtade information (*Kuhlthau - information collection*) och avslutade sökningen (*Kuhlthau - search closure*). Denna informationssökning kan kategoriseras som conditional. (Aguilar, 1967)

Deltagare 1 gjorde ytterligare en informationsökning. Den information som söktes var text och en oklar ämnesfråga (av intresse – vad som finns om ett visst ämne) (*Kuhlthau - task initiation*). Deltagare 1 var inte ute efter konkreta svar, men ville se vad som fanns för information om ett visst intressant ämne. Deltagare 1 började med att välja Google, skrev ett sökord (etiketteffekten - Ingwersen m fl, 1989) (*Kuhlthau - topic selection*), tittade på de 3 första träffarna och klickade vidare på länkar som fanns där och som intresserade deltagare 1 (*Kuhlthau - information collection*), tills denne tröttnade (*Kuhlthau - search closure*). Deltagare 1 verkade vara nöjd med resultaten. Denna informationssökning kan kategoriseras som informal (Aguilar, 1967).

Deltagaren 1 kommenterade sitt val av Google med att den har bra ryckte samt att personen har en egen positiv erfarenhet och vana att använda den.

#### **Observation 2**

Observation av deltagare 2 tog 25 minuter. Efter observationen ställde jag några frågor till deltagaren. (*Kuhlthau - task initiation*) Den information som söktes var text och en specifikt fokuserad ämnesfråga. Deltagare 2 valde att använda Google enkel sökning. Till sökfrågan (*Kuhlthau - topic selection*) valde deltagare 2 från början två ord (etiketteffekten - Ingwersen m fl, 1989). Söksvaret var alldeles för stort och deltagaren gick genom alla träffarna på första sidan, blev inte helt nöjd (*Kuhlthau - prefocus exploration*) och specificerade frågan genom att lägga till ett ord i Google enkel sökning (*Kuhlthau -focus formulation*). Svaret var fortfarande för stort. Deltagaren gick igenom alla träffarna på första sidan och hittade tillfredställande sidor (klickade genom länkar). Sedan ville personen se vad det blev för

alternativ på Yahoo. Personen sökte på Yahoo, skrev samma sökfråga och som svar erhöles många träffar. Personen tittade bara på första sidan och dokumenten var desamma som erhöles genom Google, förutom 3 annorlunda träffar. Deltagaren sade att denne var sådär nöjd med sökningen, det erhöles alternativ som var tillfredställande, men det var osäkert om det var de bästa alternativen och om det skulle finnas bättre alternativ. Irritation skapades om en sida var högt på träfflistan, men inte längre var tillgänglig. Deltagaren läste textsammandrag/sammanfattning men granskade inte adresser och kontrollerade alla träffar på första sidan, oavsett sammanfattningen. Denna informationssökning kan kategoriseras som conditional. (Aguilar, 1967)

## **Bilaga 2**

### **ENKÄT OM INFORMATIONSSÖKNING PÅ INTERNET**

Ruta  betyder att fler alternativ kan kryssas

Ruta  betyder att endast ett alternativ kan kryssas

1) ÅLDER:

20-25  26-30  31-35  36-40  41-45  46-50  över 51

2) KÖN:

man  kvinna

3) UTBILDNINGS NIVÅ:

grundskola  gymnasial  högskola/universitet eller högre

4) YRKE:

---

5) ANVÄNDER INTERNET FÖR INFORMATIONSSÖKNING:

ja  nej

*OM ja BESVARA FÖLJANDE FRÅGOR:*

5) SÖKER INFORMATION PÅ INTERNET:

dagligen  1-3 gånger per vecka  1-3 gånger per månad  sällan

6) VÄLJ RÄTT ALTERNATIV:

Jag upplever att:

a. det är lätt att hitta till hemsidor: ja  nej

b. det är lätt att hitta information på Internet: ja  nej

c. söktjänster underlättar att hitta information: ja  nej

7) VILKEN TYP AV SÖKTJÄNSTER ANVÄNDER DU FÖR ATT SÖKA INFORMATION:

länksamling  sökmotor  QA (Question Answering tjänst)   
annat

---

8) HAR DU SÖKT NÅGON GÅNG EFTER ICKE TEXTBASERAD INFORMATION:



bilder                       ljudfiler                       videofiler   
 annat  \_\_\_\_\_

på vilket sätt sökte du sådan information ? \_\_\_\_\_

9) HUR ANVÄNDER DU SÖKRESULTATET:

- a. Tittar främst på de tre-fyra översta träffarna:  
                                  stämmer helt                       stämmer delvis                       stämmer inte alls
- b. Förutsätter att det som är överst är relevantast:  
                                  stämmer helt                       stämmer delvis                       stämmer inte alls
- c. Läser alltid träffarnas titlar och beskrivningar:  
                                  stämmer: helt                       stämmer delvis                       stämmer inte alls
- d. Granskar även adresserna för att se om det verkar seriöst:  
                                  stämmer helt                       stämmer delvis                       stämmer inte alls
- e. Tittar bara på träffarna längre ner i träfflistan om de inte hittar något i de översta:  
                                  stämmer helt                       stämmer delvis                       stämmer inte alls
- f. Modifierar oftare sökningen än scollar nedåt i träfflistan:  
                                  stämmer helt                       stämmer delvis                       stämmer inte alls
- g. Går ytterst sällan till nästa sida med träffar:  
                                  stämmer helt                       stämmer delvis                       stämmer inte alls
- h. Ändrar inte sökningen förrän flera resultat granskats  
                                  stämmer helt                       stämmer delvis                       stämmer inte alls

10) KÄNNEDOM AV OLIKA SÖKTJÄNSTER:

	Använder oftast	Använder ibland	Använder sällan	Använder aldrig	Känner inte till	Helhetsbetyg av söktjänsterna (1=mycket dåligt, 5=mycket bra)
AllTheWeb						
AltaVista						
AOL Search						
Ask Jeeves						
Eniro						
Exite						
Google						
MSN Search						
Teoma						
Yahoo						
WWW virtual Library						

BUBL LINK						
Open Directory Project						
SUNET						
ELIN						
LIBRIS						

11) ANVÄNDER DU NÅGON ANNAN SÖKTJÄNST ELLER SÖKER INFORMATION PÅ ANNAT SÄTT: \_\_\_\_\_

12) OM DU MISSNÖJD MED SÖKRESULTAT VAD GÖR DU OFTAST:

ger upp                       avgränsar fråga                       omformulerar fråga   
byter söktjänst                       majlar frågan en kompis – kollega   
annat  \_\_\_\_\_

13) BESKRIV HUR DU SKULLE VILJA ATT SÖKTJÄNST PÅ INTERNET SKULLE KUNNA VARA, VILKA MÖJLIGHETER/FUNKTIONER DU SAKNAR; GE FÖRSLAG TILL HUR SÖKTJÄNSTER KAN UTVECKLAS OCH MER SPECIFICERAS

A) GRAFISK PRESENTATION och TYDLIGHET
B) FUNKTIONER
C) PRESENTATION AV RESULTAT
D) ANNAT

## **Bilaga 3**

### ***Sammanställning av enkätsvar.***

Jag inleder med att berätta om de användare som deltagit i enkäten, utifrån den bakgrundsinformation som jag samlade in, och därefter går jag igenom resultatet från de slutna respektive öppna svarsalternativen. Bland respondenterna var 13 män 15 kvinnor, från alla åldersgrupper: (20-25) 1; (26-30) 5, (31-35) 7, (36-40) 6, (40-45) 6, (46-50) 1, (över 51) 2.

Ingen av respondenterna hade utbildning endast på grundskolenivå, gymnasial utbildning hade 3 av respondenterna, högskola/universitet 25, varav 2 har IT-relaterat yrke (systemvetare). Andra yrke var advokat, sekreterare, ekonomibiträde, läkare, sjuksköterska, student, lärare, säljare.

Alla respondenter använder Internet för informationssökning och de sökte så ofta som: dagligen 20, 1-3 gånger per vecka 8, 1-3 gånger per månad 0, sällan 0.

Respondenterna svarade på frågor om sin uppfattning om svårigheten att hitta information:

- a. det är lätt att hitta till hemsidor: ja 25, nej 3
- b. det är lätt att hitta information på Internet: ja 15, nej 13
- c. söktjänster underlättar att hitta information: ja 17, nej 11

För att leta efter information använder respondenterna följande söktjänster: länksamling 14, sökmotor 28, QA (Question Answering tjänst) 0, annat 0

Några av respondenter har erfarenhet av att söka icke textbaserad information: bilder 17, ljudfiler 4, videofiler 0

Icke textbaserad information sökte respondenterna på följande sätt:

lägger till text "picture i sökfrågan, sökmotor, skrev "pictures" och det jag sökte bild av Yahoo, bildsökning, söktext + jpg,

Svar på frågan om hur använder respondenterna sökresultat, fråga med 3 valalternativ:

- a. Tittar främst på de tre-fyra översta träffarna:  
stämmer helt 10, stämmer delvis 14, stämmer inte alls 4
- b. Förutsätter att det som är överst är relevantast:  
stämmer helt 11, stämmer delvis 10, stämmer inte alls 3
- c. Läser alltid träffarnas titlar och beskrivningar:  
stämmer helt 4, stämmer delvis 21, stämmer inte alls 3
- d. Granskar även adresserna för att se om det verkar seriöst:  
stämmer helt 9, stämmer delvis 15, stämmer inte alls 4
- e. Tittar bara på träffarna längre ner i träfflistan om de inte hittar något i de översta:  
stämmer helt 3, stämmer delvis 19, stämmer inte alls 6
- f. Modifierar oftare sökningen än scollar nedåt i träfflistan:  
stämmer helt 0, stämmer delvis 16, inte alls 12
- g. Går ytterst sällan till nästa sida med träffar:  
stämmer helt 9, stämmer delvis 16, stämmer inte alls 3
- h. Ändrar inte sökningen förrän flera resultat granskats  
stämmer: helt 6, stämmer delvis 15, stämmer inte alls 7

Svar på frågan om kännedom av olika söktjänster:

	Använder oftast	Använder ibland	Använder sällan	Använder aldrig	Känner inte till	Helhetsbetyg av söktjänsterna (1=mycket dåligt, 5=mycket bra)
AllTheWeb	0	0	1	0	27	3
AltaVista	0	5	10	13	0	2,3,4
AOL Search	0	0	1	14	13	0
Ask Jeeves	0	0	0	4	24	0
Eniro	16	10	2	0	0	3,4,5
Exite	0	0	0	27	1	0
Google	23	4	1	0	0	3,4,5
MSN Search	5	6	7	6	4	3,4,5
Teoma	0	0	0	10	18	0
Yahoo	4	10	7	7	0	1,2,3,4,5
WWW virtual Library	0	3	6	8	11	3
BUBL LINK	0	0	1	10	17	0
Open Directory Project	0	0	0	0	28	
SUNET	2	1	4	11	10	2,3,4
ELIN	4	10	4	5	5	3,4
LIBRIS	0	6	7	10	5	2,3

Några av respondenterna använder även andra söktjänster för att söka information på Internet: pubmed ([www.pubmed.com](http://www.pubmed.com)), hitta ([www.hitta.se](http://www.hitta.se)), VEGA är Malmö högskolas bibliotekskatalog, Malin - Malmö stadsbiblioteks katalog

Svar på frågan om vad man gör om man är missnöjd med sökresultatet: ger upp 4, avgränsar fråga 26, omformulerar fråga 25, byter söktjänst 7, majlar frågan en kompis – kollega 2, annat: 1 provar igen senare

Respondenternas förslag/önskemål över söktjänsterna på Internet:

A) GRAFISK PRESENTATION och TYDLIGHET

Ikon före texten som visar vilket typ av hemsida ex universitet, privat, försäljning osv  
Viktigt att sidan tydligt visar hur den fungerar, inte för mycket detaljer, bättre enkelt och begripligt

B) FUNKTIONER

Kunna välja kort produkt, försäljningshemsidor

Kunna välja bort annons eller filtrera vilken typ av information som kommer upp i listan (government, educational, företag o s v).

#### C) PRESENTATION AV RESULTAT

Experter skall kunna granska och rangordna hemsidor – men man bör kunna välja bort denna funktion

I ordning och med relevantast resultat först, tydliga fungerande länkar under sidan.

#### D) ANNAT

Ska vara mer kategoriserad. Ex kan man först tala om att jag vill ha artiklar, andra – psykologi, tredje emotionella fjärde, engelska osv.

## **Bilaga 4**

### ***Intervjuguide***

#### Om söktjänster

1. Vilka söktjänster använder du vid informationssökning på Internet?
2. Vilken eller vilka av dem använder du oftast?
3. Varför väljer just den söktjänsten?
4. Vad vet du om ditt första val söktjänst?
5. Vad skulle du vilja veta mer om den söktjänsten och informationsökningen?
6. Vad använder du söktjänster till? (För vilken typ av frågor)
7. Vad tycker du om de söktjänster du använder? (gränssnitt, resultat du få)
8. Vilka möjligheter saknar du?
9. Vad anser du vara speciellt bra resp. dåligt? Vilka problem upplever du finns?
10. Har du något förbättringsförslag?
11. Har du använt QA?
12. Skulle du till söktjänsten lämna personlig information om dig om det kunde förbättra resultaten genom att personifiera sökningen (genom att skapa användarprofilen)?
13. Vad tycker du om teknik som undersöker dina sökmönsters? ex cookies
14. Om du får höra talas om en ny söktjänst fortsätter du då att använda den/de du vanligen använder, eller prövar du den nya? Varför/varför inte?

#### Om sökfrågor

15. Förbereder du din sökning på något sätt? Hur börjar du formulera sökfrågan?
16. Hur många ord brukar dina sökfrågor ha?
17. På vilket sätt specificerar du sökfrågan?
18. Använder du avancerad sökning?
19. Använder du bolesiska operatörer?
20. Har du något särskilt söktips, sökstrategi eller sökmetod?
21. Skulle du vilja lära dig sökspråk för att förbättra sökfrågeutformningen?
22. Om du skulle söka t ex efter en flygresa hur skulle du göra?

#### Information

23. Vilka relevanskriterier och trovärdighetskriterier använde du när du bedömde relevans och trovärdighet på en träff?
24. Har du sökt lokala informationer och hittade du det du ville? Hur upplevde du den?
25. Är du intresserad att veta hur informationen som du får genom söktjänst filtrerad och om informationen är sponsrad?
26. Skulle du vilja ha möjlighet att påverka informationens rankning?

#### Generella/övriga frågor

27. Hur upplever du att söka information på Internet?
28. Vad oroar/irriterar dig när du söker information på nätet?
29. Hade du haft dåligt upplevelse av informationssökning?
30. Hur skulle det vara för att minimera/eliminera den?
31. Vilka önskemål har du när/för dina barn söker information på Internet?
32. Skulle du vilja betala för söktjänstens service om den blev bättre?
33. Har du använt hjälpsida till söktjänsten? Om ja: var du nöjd med den?

## **Bilaga 5**

### ***Sammanställning av intervjusvar.***

#### Deltagare 1

1. Google, Eniro, hitta.se och tidigare Yahoo! AltaVista osv
2. Google, Eniro, men oftast Google.
3. Enkelhet, lättbegriplighet, funkar ofta tillfredställande
4. Alla säger att den är bäst, vilket betyder att den kanske är störst och mest populär. Om hur Google uppbyggd dvs om tekniken bakom vet jag ingenting
5. Vet inte, jag är inte särskild intresserad.
6. Söker allt möjligt information både för mina studier och privata intressen.
7. se svar 3
8. att välja några språk samtidigt kanske
9. Ibland är osäkert om informations trovärdighet, inte alltid lätt att hitta det man behöver, Bra – snabbt, när man skrivit fel kommer med stavning förslag.
10. Ville gärna få meddelande när någon information över det som jag särskild intresserad har inkommit.
11. Aldrig
12. Aldrig gjort men det är tänkbart att skaffa sina personliga inställningar som t ex filtrering av sidor olämpliga för barn, begränsa till dem några språk som jag kan, även att begränsa till ett visst antal returnerade dokumenten.
13. Vet inte
14. Jag kan prova den för att se om den är bättre.
15. Tänker genom vad jag behöver ha för information och sedan väljer ord som avspeglar den mest och se vad det bli för resultat med dem.
16. Oftast ett eller två ord
17. Går till avancerad sökning och utifrån både typ av information jag behöver och resultat jag fick först avgränsa min sökning t ex läger till ord som ska eller inte ska vara med, ange datum osv.
18. Ja
19. nej
20. Inte speciell, att läsa sammanfattningen, lägga sidor man litar på till favoriter
21. Vill inte spendera tid på att lära sig sökspråk.
22. Kolla på sidor av dem flyggbolag som jag känner och bruka flygga med.
23. svårt att säga, kan vara domännamn, webbplatsadress, även sidornas layout och innehåll t o m
24. Sökte lokal info om London och det gick väldigt bra. I Lund är också många men inte alla ställen är med. Affärsverksamheter som t ex restauranger, hotellen, resebyrå är inte särskilt svårt att hitta. Men en gång letade efter alternativ för barnkalas och det var inte lätt hitta olika möjligheter.
25. vet inte
26. vet inte
27. Det är olika, för det mesta få man reda på det man vill även ibland få gå genom många sidor
28. Oro: att missa en del information, oftast få man för mycket information och omöjligt att gå genom allt.
29. När sidan inte finns längre eller där man söker en typ av information och få helt annan t ex försäljningssidor

30. Uppdatera oftare, möjlighet att välja vilket typ av information önskas, t ex artikel, försäljning osv.
31. Filter för våld och pornografisk information
32. Nej, vill inte betala helst söka gratis.
33. Aldrig använd.

### Deltagare 2

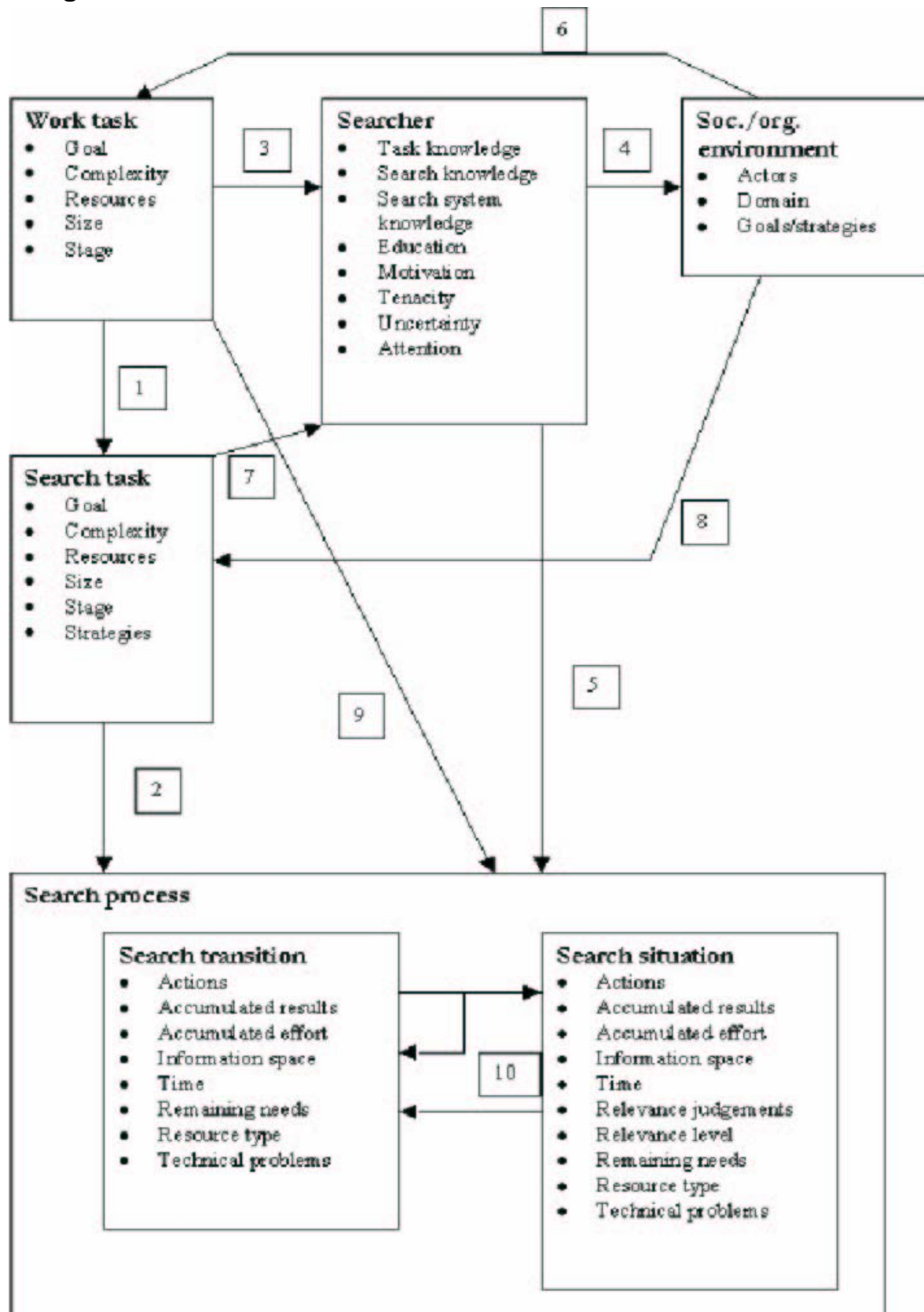
1. Google, Elin, Eniro, Yahoo, Libris och provat ett antal andra någon gång.
2. Oftast använder Google.
3. Hörde och läste att den är större och bättre än konkurrenternas, har dessutom enkelt design som är lätt att begripa och ingen reklam på första sidan, men huvudsaken att det känns att man får bättre information än genom AltaVista som jag använde förut.
4. Att den är störst, har effektivt ranking och inte prioritera betalda sidor.
5. Vet inte.
6. Söker all möjligt information både för arbete och privat.
7. Ganska bra
8. Trådlös shopping söktjänst, röstsöktjänst för mobil
9. Bra: utseende, att reklam inte finns med
10. Skulle vilja att söktjänsten skulle förstå mitt naturliga språk så att jag kan skriva min sökfråga som jag skulle tala
11. Nej
12. Kan tänka sig att lämna ut till söktjänst min personliga information om detta skulle underlätta sökningen, men samtidigt vill vara säker på att informationen inte säljs vidare
13. It ok men vill gärna bli informerad av söktjänst i fall den gör så och vill ha valmöjlighet att välja bort detta.
14. Möjligtvis, om jag hör eller läser mycket positivt om det
15. Väljer ut de viktigaste begreppen/termerna, prova vilken resultat som kommer
16. Ett till tre
17. Läger till ord och fraser som kan hjälpa till att avgränsa resp. utvidga sökningen
18. Ja
19. Jag kan men det behövs inte
20. Inget särskilt
21. Vill inte lära sig något speciellt sökspråk, det ska vara lätt att förstå vad och hur det ska skrivas.
22. Skulle skriva ord ”flyg, flygresa, destinationsnamn”
23. Kollar om adressen tillhör till en pålitlig organisation t ex, läser innehåll
24. Ja, det gick bra med t ex tågtider, biblioteket olika händelser som t ex Malmöfestival eller teaterföreställning.
25. Kanske om det inte för mycket eller för komplicerad.
26. Kanske att välja sorteringsprincip som man kan göra i MS Office
27. Spännande, bra, ibland kan bli jobbigt om t ex ämne är brett och man inte är säker på vilken information man vill ha, så det är svårt att specificera.
28. För mycket reklam och icke trovärdig information, för många svar så att man inte kan kolla alla
29. Ja
30. Utveckla computer och Internet så att den skulle fungera logiskt som en mänsklig hjärna
31. Filter mot pornografisk, rasistisk och annan information som är olämplig för barn
32. Vill helst ha söka gratis, men i visa fall skulle kunna tänka sig att betala för en speciell information.
33. Nej



### Deltagare 3

1. Google, VEGA och har provat MSN Search och (visimo) [www.visimo.com](http://www.visimo.com)
2. Google
3. Väljer Google, den har bra ryckte, är enkelt att använda, lättbegripligt samt har olika moment som tillåter att precisera sökningen, stödjer många språk. Jag vill inte ha en söktjänst med många komplicerade funktioner och förvirrade moment.
4. Att den är mest använd söktjänst på Internet med störst databas
5. Vilka andra söktjänster som är lika goda/effektiva som t ex Google, vad har Google för nackdelar/svaga sidor.
6. Allt möjligt informationssökning.
7. Ok
8. Frågor skulle kunna översättas till andra språk speciellt när man söker efter modeller eller bilder.
9. Bra: lätt begripligt, enkel design, snabbt resultat som är ofta bra, dåligt när man får irrelevant och för mycket information.
10. Funktion där man t ex genom klick på ikon skulle meddela till söktjänst att man anse dokumentet har hög relevans till fråga är kanske inte dåligt. Val av funktion där man kan välja att resultat ska presenteras under kategorier. Kanske sökvizard för komplexa sökningar men sådan att den skulle var enkel att förstå och använda.
11. Nej
12. Skulle inte ge söktjänsten min personliga profil för att underlätta sökningen.
13. Negativt, vill behålla min privata integritet.
14. Jag skulle prova den om jag hör att den är bättre än andra
15. Funderar över vilken information som behövs och vilket sökord som skulle vara mest representativt, provar söka med den och se vilket resultat som blev först.
16. 1-3 men ibland kan vara flera ord bero på den information som sökes.
17. med hjälp av avancerad sökning
18. ja
19. Sällan AND, NOT, ELLER (det behövs inte i Google t ex)
20. Kolla domänadress och sammanfattning i träff innan besök på webbsida
21. Vill helst inte spendera tid på att lära sig sökspråk eller väldigt liten tid.
22. Söka genom en resebyrå eller flyggbolag som redan känns till.
23. Lite svårt att säga, kan vara längden på texten t ex som relevanskriterier, innehåll i texten, datum osv.
24. Ja, det oftast går bra med visa sidor speciellt små affärsverksamheter (ex restauranter, butiker) är inte bra på uppdatering när t ex öppettider förändras osv.
25. Ja
26. Ja
27. Informationsökning är viktig för mitt arbete och privat
28. Bekymmer att missa viktigt information, informationens pålitlighet, man får alldeles för stora resultatlista och kan inte kolla alla svar.
29. Upplever att det är alldeles för mycket information som det är svårt eller rent av omöjligt att hantera.
30. Kanske extrahera en del av informationen till informationsdatabas så att man kan läsa sammanställd information utan att besöka många olika sidor.
31. Att man skulle kunna sätta filter för olämplig information de kan komma åt.
32. Skulle kunna tänka sig att betala ibland till en söktjänst för att få tillgång till information av hög kvalitet, utan reklam och med trovärdigt innehåll och till och med har betalat någon gång för att få tillgång till en viss databas. Men det är bra och nödvändigt med gratis informationsökningservice.

**Bilaga 6**



The conceptual framework - the domain of the method schema. Ur Pharom fl (2002).

## Bilaga 7

Kuhlthaus ISP-modell (Kuhlthau 1993, Figure 3-1 s 43, genom Båge m fl, 2003)

Stages	Task initiation	Topic selection	Prefocus exploration	Focus formulation	Information collection	Search closure	Starting writing
Feelings	uncertainty	optimism	confusion/ frustration doubt	clarity	sense of direction/ confidence	relief	satisfaction or dis- satisfaction
Thoughts		ambiguity	—————→		specificity		
Actions	seeking relevant information		—————→		seeking pertinent information		

Kuhlthaus ISP-modell (Kuhlthau 1993, Figure 3-1 s. 43)

## Bilaga 8

Systemorganisation ur Funkhouser m fl (2002)

6 · Thomas Funkhouser et al.

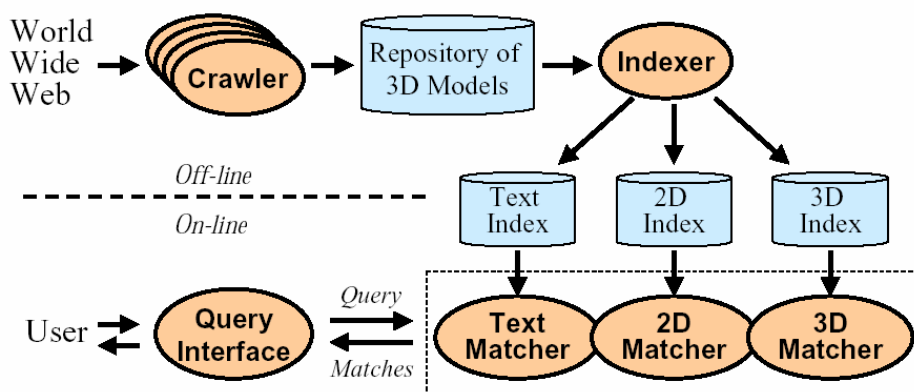
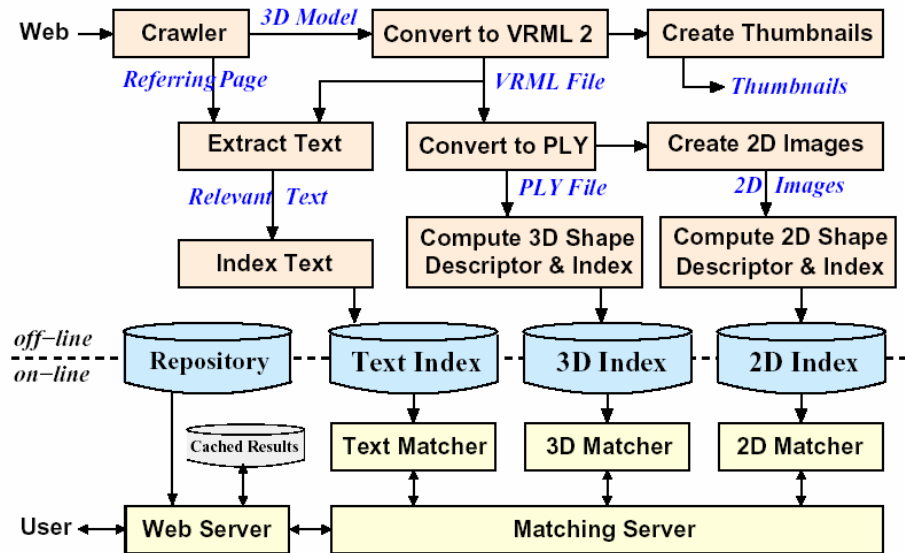


Fig. 2. System organization.

## Bilaga 9

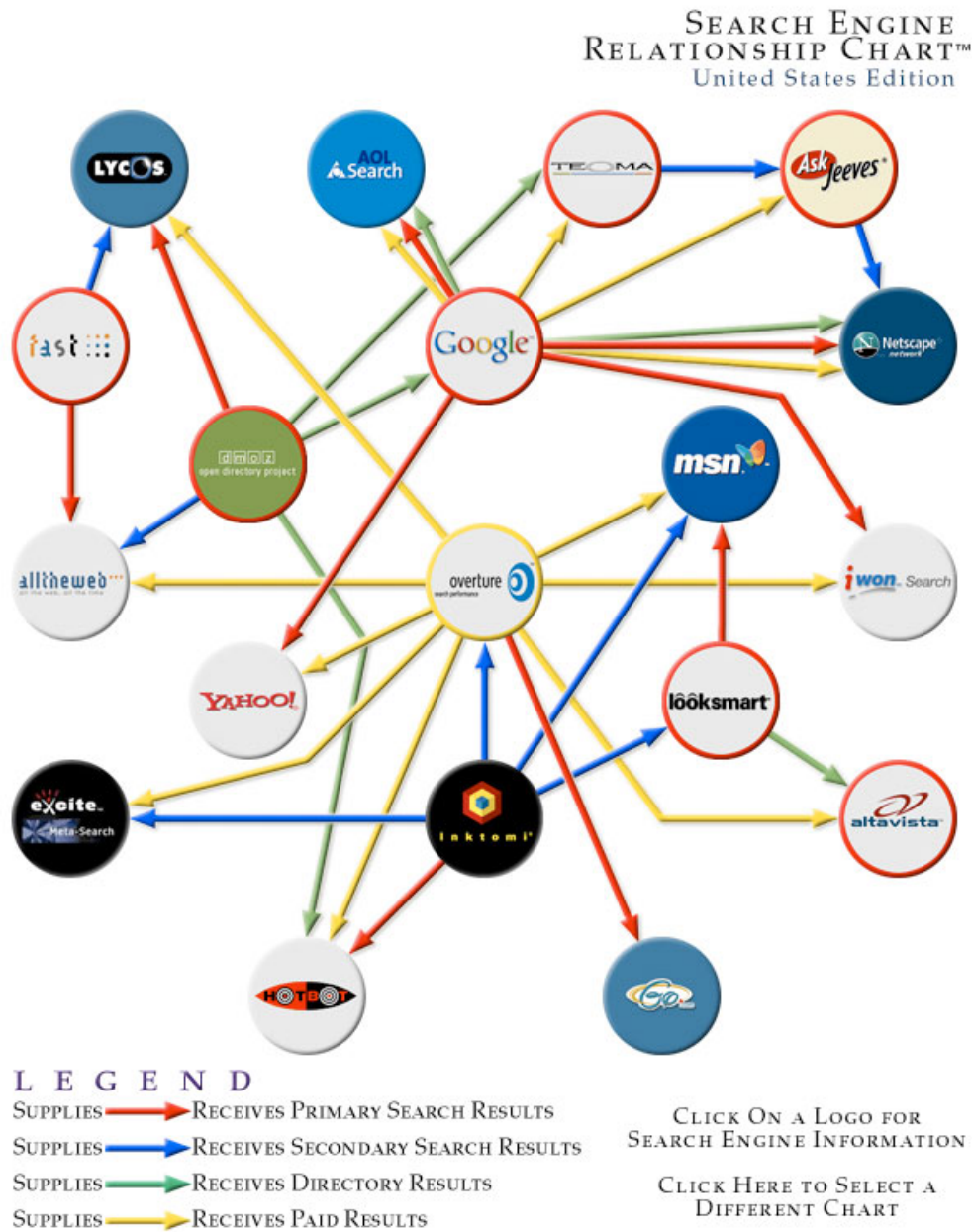
Söktjänst för 3D modeller ur Funkhouser m fl (2002)

A Search Engine for 3D Models · 15



## Bilaga 10

Söktjänsters relationer ur <http://www.ellipseinc.com/searchengine/images/relationship.jpg>  
 eller <http://www.bruceclay.com/searchenginereationshipchart.htm>



ALL LOGOS PROPERTY OF THEIR RESPECTIVE OWNERS  
 COPYRIGHT © 1996-2002 BRUCE CLAY, LLC - ALL RIGHTS RESERVED

0212.1

## Bilaga 11

Diagram över sökningar av US Internetanvändare under juli 2005 (Sullivan, 2005)  
<http://searchenginewatch.com/reports/article.php/2156451>

