

Träffa rätt på Internet

**– Tio tips för att förbättra dina
sökningar på Internet**

Att kunna hitta rätt information på Internet är mer eller mindre ett krav för många arbeten. Många känner sig hemma på nätet och upplever inte att de har några problem med att hitta rätt information, medan andra har på känn att det finns mycket mer att hämta än det de hittar. Det spelar ingen roll vilken av dessa grupper du tillhör – det finns alldeles säkert saker du kan göra för att förbättra resultaten av dina sökningar.

Är du nöjd med dina sökningar på Internet? Hittar du nålen i höstacken? Jämt? Oftast? Ibland eller kanske aldrig?

NPD New Media Services har varje kvartal sedan våren 1997 gjort undersökningar hos de största söktjänsterna för att mäta hur besökarna använder respektive tjänst.

Enligt deras rapport från i somras visar det sig att besökarna blivit mer nöjda med sina sökningar. En bottennotering nåddes hösten 1999, men under vintern har besökarna blivit mer nöjda och nu visar det sig att 81% av besökarna är nöjda med sina svar. 20% är dessutom alltid nöjda med sina svar.

Oavsett om du är nöjd med dina sökningar eller inte tänkte vi presentera tio tips för att förbättra dina sökningar.

Tips ett

Använd flera söktjänster, i stället för att vara trogen en enda!

I en undersökning gjord av NPD New Media Services förra vintern visade det sig att bara en femtedel av dem som inte hittar vad de söker byter till en annan söktjänst. 75% av dem som misslyckas omformulerar sökningen i stället.

Eftersom dagens söktjänster täcker en mycket liten del av Internet och många söktjänster har unika dokument i sitt index är det viktigt om man vill göra en någorlunda uttömmande sökning att man söker med flera söktjänster.

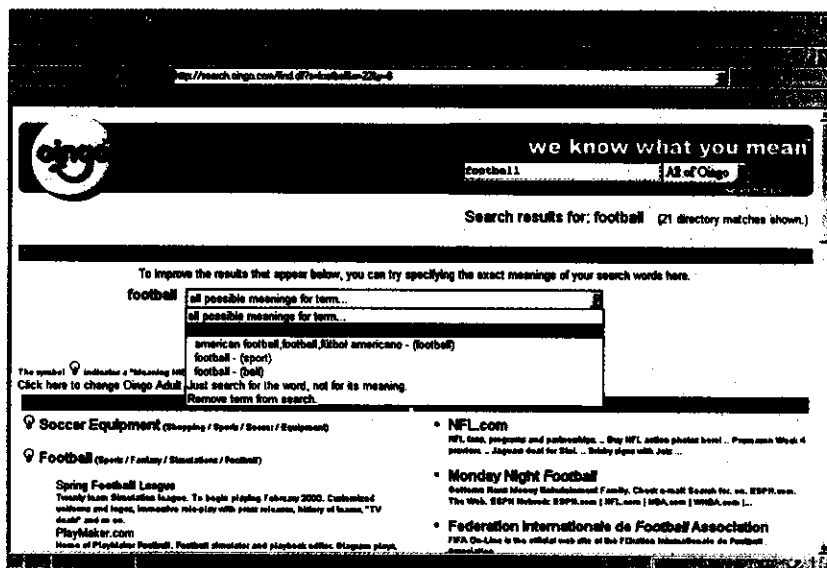
I Searchengineshowdowns test av unika dokument från februari i år visade det sig att av 110 unika dokument funna hos 14 stora söktjänster fanns 40 hos norska FAST, 23 hos AltaVista och 18 hos Excite.

Kontentan är alltså att man inte bör hålla sig till samma söktjänst hela tiden. Lär dig några söktjänster ingående men använd även andra ibland.

Tips två

Användandet av booleska operatörer slår ut relevansrankningen (och viktiga dokument kan hamna långt ner i träfflistan)!

Elektroniska databaser har funnits långt före Internet, ända sedan sextiotalet, och booleska operatörer har flitigt använts som ett effektivt sätt att precisera sin sökning. I och med Internets intåg blev HTML en standard för att presentera webbsidor. Tyvärr är HTML mer anpassat för att styra utseendet på webbsidor än för att presentera information. HTML är bristfälligt vad gäller



Om man slår in ordet "football" i söktjänsten Dingo får man upp en listlista med olika begrepp att sortera efter. Man kan till exempel välja att söka på "american football" eller "association football". Men man kan också välja att söka på "football" som substantiv och då får man upp information om "football equipment".

tydligt definierade fält där beskrivande information om webbsidan kan läggas in. Visa metadata, det vill säga information om webbsidans innehåll, går att lägga in i så kallade metataggar i HTML-koden. Men fler och fler sökmotorer bryr sig inte om denna meta-information eftersom så många webbmakare missbrukar metataggarna.

Av denna orsak, plus att det rör sig om så många miljoner webbsidor i de stora söktjänsternas index, skapar det problem om man söker booleskt. Man slår nämligen ut relevansrankningen, det vill säga i vilken ordning träffarna presenteras. Booleska operatörer säger nämligen inget om ordens placering, bara om ordens förekomst.

Det innebär att man slår ut vanliga kriterier för relevansrankning: närheten mellan orden, var orden förekommer och om de förekommer på flera ställen. Om man exempelvis söker på 'sverige' AND 'fotboll', så kan det betyda att ordet 'sverige' kan förekomma högt upp i ett dokument och där användas i sammanhanget brottnig i sverige. Ordet 'fotboll' kan i sin tur förekomma långt nere i dokumentet och användas i sammanhanget fotboll i Norge. Se följande exempel:

sverige AND fotboll

Sverige har äntligen fått OS-guld i brottnig igen och...

I Norge är det populärare än nånsin med fotboll...

Dokument där 'sverige' och 'fotboll' förekommer i samma stycke eller kanske i samma rad kan med andra ord hamna långt ner på träfflistan.

Särskilt om man har vana från att söka i elektroniska databaser med booleska operatörer ska man inte ta för självklart att det är bästa sättet att söka avancerat i sökmotorer på Internet.

För en förklaring av booleska operatörer (se faktaruta 1).

Tips tre

Använd närhetsoperatören NEAR!

Förutom booleska operatörer finns det många andra operatörer som brukar användas vid sökning i elektroniska databaser, men som tyvärr inte går att använda i många söktjänster. Om man tittar i bibliotekens svenska webguide Svesök, producerad av Kungliga Biblioteket, klickar på Hjälp-knappen och längre ned på sidan väljer länken Söktips, så kan man se en uppräknig av olika operatörer. Ett flertal av dessa är så kallade närhetsoperatörer. Till exempel NEAR, som med fördel går att använda i AltaVista avancerad sökning och i den gamle trotjänaren Webcrawler. Ett varningens finger måste höjas för Webcrawler, som har dålig webbtäckning.

NEAR betyder att sökorden skall stå bredvid varandra men i godtycklig ordning. I Altavistas fall inom tio ords avstånd från varandra. Närhet används i de flesta söktjänster som ett viktigt kriterium vid rankning. NEAR är ofta mer användbart än booleska operato-

rer när man söker med en söktjänst på Internet. Se följande exempel och jämför med det tidigare exemplet då operatör AND användes:

sverige NEAR fotboll

I Sverige är det fler utövare av fotboll än...

Webcrawler har också en närhetsoperator som heter ADJ och som innebär att orden måste stå intill varandra och i angiven ordning. Det blir detsamma som att söka på en fras med citations-tecken ” ”.

Tips fyra

Tänk vertikalt, inte bara portalt!

Många människor tycker det är praktiskt med en ingång till Internet.

Företrädesvis en söktjänst och gärna en som har gjorts till portal, det vill säga erbjuder tjänster som gratis webbplats, gratis e-post, väderinformation, kalender och annat. Exempelvis Yahoo eller AltaVista. Eller så en portal med integrerad söktjänst. Exempelvis Spray eller American Online.

Vanligt är också att söktjänsten eller portalen används som startsida i webbläsaren. Portaler är utformade just för människor som tycker att det är bekvämt att ha så mycket som möjligt på samma ställe.

Men vill man göra en effektiv och mer uttömmande sökning på Internet innebär portaltänkandet att man missar väldigt mycket information.

Av denna anledning bör man också tänka vertikalt. Portalerna och de stora söktjänsterna har ambitionen att ge ett

bra svar oavsett vilket ämnesområde det rör sig om, vilket många gånger visar sig vara väldigt svårt. Det beror naturligtvis inte enbart på själva söktjänsterna utan även på vad som egentligen finns på Internet inom olika ämnesområden.

Därför har det växt fram en allt mer vildvuxen flora av så kallade specialiserade söktjänster med ambitionen att täcka ett visst ämnesområde på Internet. Många av dessa kallar sig missvisande portaler (se faktaruta 2). Vad det egentligen handlar om är vertikala portaler. På engelska brukar det kort kallas vortals. På svenska går det bra med vertikaler, även om det är uttrycket portaler som fått mest genomslag. Andra ord är Subject gateways och Specialized searchengines.

Vertikaler riktar in sig på ett specifikt ämnesområde, och ofta har de skapats av människor som själva besitter specialkunskaper inom ämnet. Det finns en mängd vertikaler av skiftande kvalitet. Tjusningen med vertikaler, som kan vara rena webbkataloger och/eller sökmotorer, är att det inte krävs samma stora apparat för att starta en vertikal som för en portal, där du ska täcka alla ämnesområden och erbjuda gratistjänster för att generera besökare.

Ett svenskt exempel på en vertikal är Crossnet.se. De kallar sig den kristna portalen. Men det handlar alltså inte om en portal i ordets rätta bemärkelse, utan det är en vertikaliserad portal. Du kan söka inom ämnet kristendom i ett index som innehåller andra webbsidor.

Hur hittar man vertikaler? Vanligt är att vertikaler döper sig till det ämne det handlar om till exempel Business.com eller Golf.com.

Man måste också göra skillnad mellan en vertikal som erbjuder en att söka i ett index, en webbkatalog som hänvisar till andra webbsidor på nätet och webbtjänster som bara söker på innehållet i den egna webbplatsen, till exempel Intranetica.com, som inte är en vertikal i ordets rätta bemärkelse men ändå kan ge nog så bra svar.

Ett sätt att hitta vertikaler är att använda sökfunktionen "page depth" i HotBot och bara söka på toppsidor. Man slår in ett ord för det ämne man är intresserad av till exempel "music". En av träffarna längre ned i listan blev vertikalen Musicsearch.com.

Det finns även tjänster som förtecknar olika vertikaler: Searchengineguide.com. Man kan även gissa webbadresser. Om man är intresserad av

Complete Planet är i dagsläget en av de allra bästa webbplatserna att gå till för att söka information.

Booleska operatörer – vad är det?

Matematikern George Boole (1815–64) räknas som upphovsman till så kallad boolesk algebra. De tre booleska operatörerna, som är AND, OR och NOT, ger möjlighet att inkludera och exkludera sökord.

Så här fungerar operatörerna:

fotboll AND sverige

ger dokument som innehåller båda termerna.

fotboll OR rugby

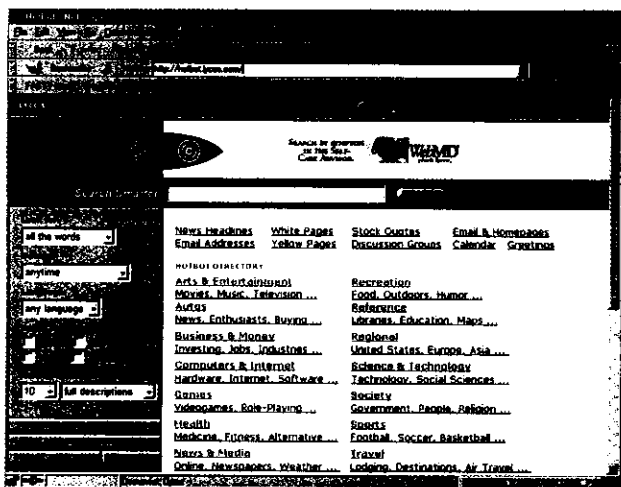
ger dokument som innehåller åtminstone ett av orden.

fotboll NOT amerikansk

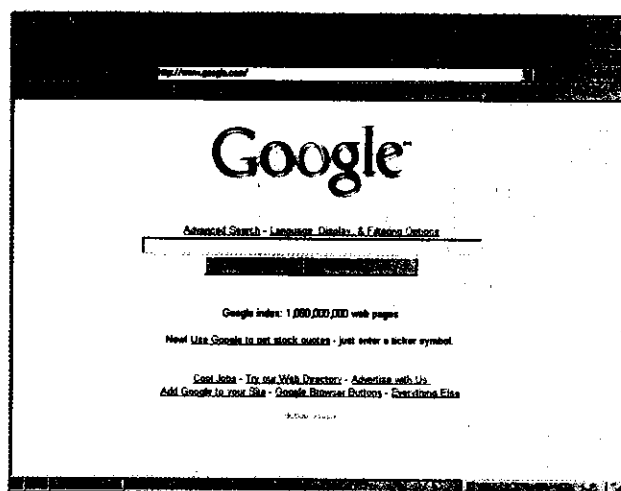
ger dokument som innehåller det första sökordet men inte det andra.

Ha alltid för vana att slå in operatörerna med stora bokstäver, eftersom vissa söktjänster annars inte kan förstå om det är en operatör eller ett sökord.

Faktaruta 1.



I HotBot kan man använda färitsökningen 'depth:' genom en siffra anger man hur djupt ner i underkatalogerna sökningen ska göras.



Google är en söktjänst som märkligt nog inte har någon möjlighet till trunkering.

konst: Art.com, Artnet.com, Artguide.com, Artdirectory.com etc.

Tips fem

Det är sökord, inte kontrollerade ord, man söker på i en sökmotor!

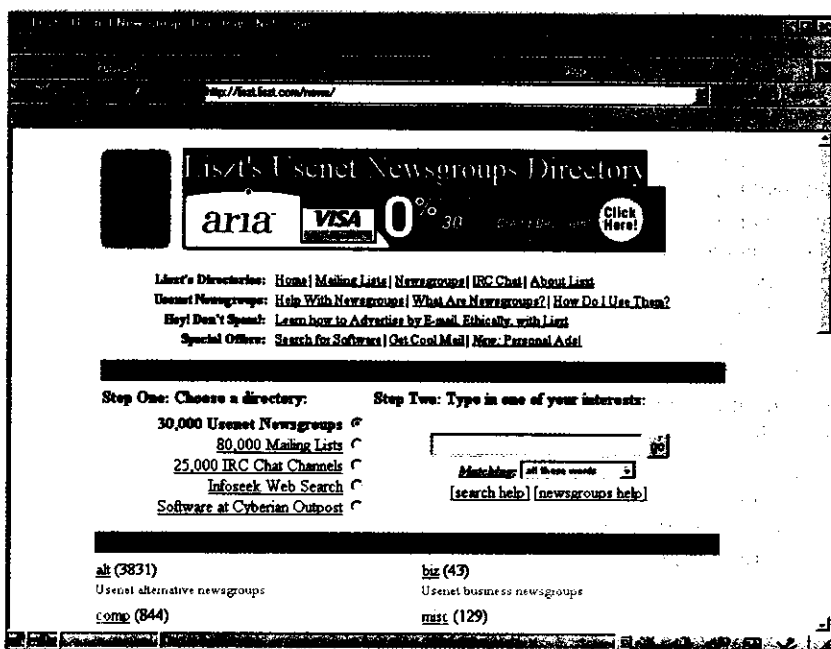
Det kan låta självklart, men glöm inte bort att det är vanliga ord du söker på som den som gjort webbsidan själv har definierat. Används ordet "blad" så kanske det är blad i en bok som avses, du menade blad på en växt när du skrev in ordet i en söktjänst. Ord som blad, som har dubbla betydelser, brukar med en språkvetenskaplig term kallas polysema. Det finns en uppsjö av polysema ord. Kort, bila, hinna, tunna och så vidare.

Synonymer är ett annat problem. Om jag vill söka efter poesi måste jag söka även på dikter, lyrik, poem, verser, skaldekunst och så vidare, eftersom det inte finns någon överenskommelse om vilket ord som bör användas. Orden är inte kontrollerade.

Sökmotorn Excite använder sig av ett söksystem som brukar kallas "concept-based". Den försöker förstå begrepp i stället för exakta ordmatchningar, med hjälp av bland annat synonymlistor och ämnesordlistor.

Ett typexempel på problem med ord är om man söker efter 'football', vilket resulterar i sidor om amerikansk fotboll också. Det kan man försöka lösa genom att utesluta ett ord med hjälp av ett minustecken: 'football -american'. De flesta söktjänster förstår sökspråket '-' och '+'.

Men en sådan sökning gör att man ändå får upp sidor om amerikansk fotboll. Ordet football betyder i USA just amerikansk fotboll och alla amerikan-



Lizt's Usenet Newsgroups Directory är ett bra ställe att söka nyhetsgrupper och e-postlistor på.

ska sidor som handlar om american football innehåller därför inte alltid ordet american. Syftar en amerikan på fotboll spelad med huvudsakligen fötterna så använder han ordet soccer.

Men om man då följaktligen söker enbart på sidor om soccer missar man ändå många sidor om fotboll. Ett olösligt problem?

Oingo är en metasöktjänst som har försökt lösa detta begreppsmässiga problem. Om man slår in ordet football i Oingo får man upp en rullista med olika begrepp att sortera efter (se bild 1).

Man kan välja att söka på 'american football' eller 'association football'. Men man kan också att välja på att söka på 'football' som substantiv och då får man upp information om 'football

equipment'. Vad man bör vara medveten om är alltså att det inte handlar om kontrollerade ord, som är fallet i många elektroniska databaser, utan det är helt upp till den som skapat informationen på webbsidan att välja sina ord. I en databas med kontrollerade ord kan man bestämma sig för att använda ordet 'football' som ämnesord i meningen 'amerikansk fotboll' och 'soccer' som ämnesord i meningen 'association football'. Någon sådan bredare lösning för kontroll av ord och begrepp finns inte idag på Internet.

Tips sex

Slå in fler än ett ord!

Det finns undersökningar som visar att de flesta slår bara in ett eller två ord

i en söktjänst. Vad gör de sen? Ger upp, nöjer sig eller går in i avancerad sökning? Oavsett vad du gör så är ett råd att du slår in fler än två ord innan du ändrar sökstrategi. Ett exempel kan vara om du är intresserad av vin och särskilt då vitt vin. Du slår in 'vitt vin' men får för många svar. Försök precisera frågan med ytterligare ett ord. Exempelvis 'Riesling', som är en speciell druva. Du kan också precisera ytterligare med att det ska vara från Moseldalen genom att slå in 'Mosel'.

Enligt undersökningar från NPJ New Media Services har benägenheten hos besökarna att slå in flera ord ökat

markant för första gången sedan de började sina undersökningar sommaren 1997. Varför? Kanske beror det på att söktjänsterna har blivit duktigare på att tala om för besökarna att använda flera sökord för att precisera sin sökning och att söktjänsterna själva har förbättrat sina rankningsmetoder så att besökarna samtidigt blivit mer nöjda med sina svar. Vem vet?

Tips sju

Sök på ytan av webben och borra dig ner i djupet med "page depth"-funktionen i HotBot!

Många webbplatser är väldigt stora

och innehåller många underliggande dokument långt ner i underliggande kataloger. En sökning på ordet Internet i Google ger bara från webbplatsen Internet.com drygt 800 träffar. Därför brukar bland annat Google, Northern Light, Infoseek och HotBot sortera träfflistan genom klustring. Det innebär att träffar från samma webbplats presenteras som en träff. Om man klickar på en länk under träffen visas alla träffar från just den webbplatsen i en egen träfflista.

När man söker i en sökmotor som inte klustras presenteras många gånger oviktiga träffar på webbsidor långt ner i

Vad då portal?

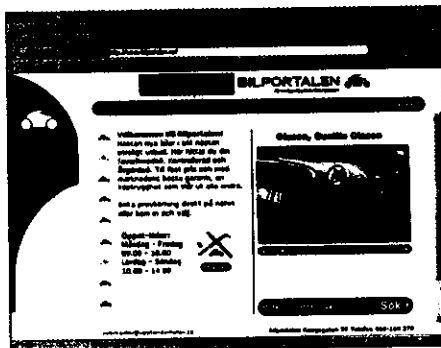
Om man söker på Internet på ordet "portalen" hittar man en uppsjö av olika webbplatser som kallar sig för portaler. Juridikportalen, Bilportalen, Crossnet – den kristna portalen, Danmarksportalen, Lantbruk.com – portalen för lantbrukaren. Många är de på Internet som försöker slå mynt av ordet portal, och dess ursprungliga betydelse har urvattnats mer och mer. Med en portal menar man en dörr ut till Internet, med ambition att samla flera ingångar till Internet på en och samma webbplats för att locka till sig så många surfare som möjligt. Man kan säga att det finns två grundkrav för att en portal ska kunna kallas portal:

1. Att den innehåller "kanaler" liknande en webbkatalog där man kategoriserat ingångarna efter ämnesgrupper.
2. Att det finns en sökfunktion.

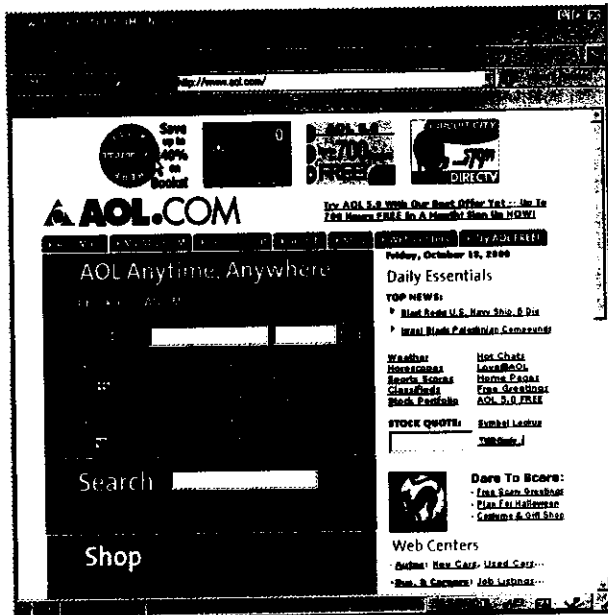
Men en portal med stort P erbjuder också kringtjänster som gratis e-post, gratis hemsida och olika former av anpassade inställningar som till exempel egen almanacka.

Därför måste man göra skillnad på de stora allmänportalerna och de vertikala portalerna. De stora allmänportalerna i Sverige är Spray, Passagen och Torget. I USA är det America Online som dominerar. Man brukar också prata om portaliseringen av webben. Det vill säga att fler och fler webbplatser, speciellt söktjänster, lägger till fler kringtjänster för att dra till sig fler surfare. AltaVista och Yahoo är typiska exempel på detta.

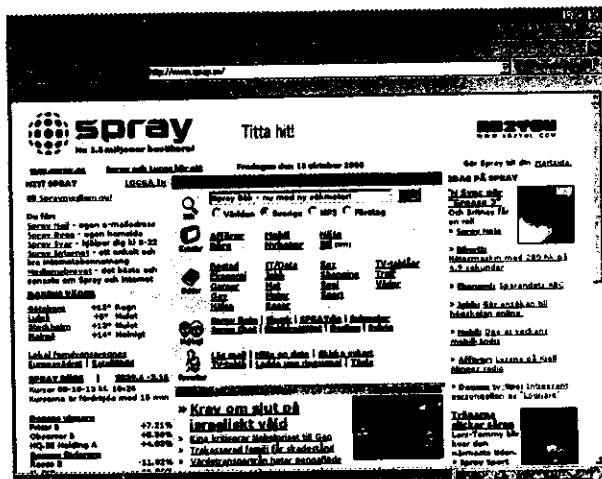
De vertikala portalerna då? Det vill säga ämnesinriktade portaler. Vertikaler, vertikaler eller vertikala portaler är begrepp som i princip är okända på den svenska webben och många som kallar sig "portal" eller "portalen" har varken något med allmänportalering eller vertikalisering att göra. Bilportalen, till exempel, är en ren köp och sälj-sajt.



Bilportalen är en webbplats som i strikt mening knappast kan definieras som en portal. Vågarna ut mot resten av Internet är begränsade och sökfunktionen inskränker sig till att söka bilar som är till salu.



America Online är den dominerande portalen i USA. Den uppfyller gott och väl kraven på en portal, då den har både en sökfunktion och kategoriserade länkar.



Sprays portal är en av de mer välbesökta i Sverige. Ingångssidan är nära nog överbelastad med innehåll och funktioner.

man kan söka på olika stavningar. Så till exempel ger följande sökning:
Lar?on

bland annat följande ord:
Larson, Larzon, Larmon.

En del söktjänster erbjuder också "stemming", som det kallas på engelska. Det betyder att söktjänsten till exempel söker både på pluralis och singularis eller olika tidsformer av ett verb.

Vanligast är dock sökning på pluralis och singularis. Detta använder Northern Light. Yahoo autotrunkerar, vilket ger annat än bara pluralis och singularis. För att få en sammanställning över hur de olika söktjänsterna hanterar trunkering se Searchengineshowdowns sammanställning över respektive söktjänsts sökspråk.

Anmärkningsvärt är att Google inte ger möjlighet till trunkering eller erbjuder någon form av stemming.

Tips nio

Var inte blind för den osynliga webben!

Jag har skrivit om det tidigare och det tål att upprepas: dagens stora söktjänster täcker en mycket liten del av all information på Internet. Enligt Lawrence och Giles studie från februari 1999 täckte den största söktjänsten Northern Light bara 16 procent av den indexeringsbara webben, som då uppskattades till 800 miljoner sidor. Den indexeringsbara webben är den del av webben som sökmotorerna med dagens tekniker skulle kunna klara av att indexera.

Det finns en stor del av webben som inte sökmotorerna klarar av att indexera av tekniska skäl. Bland annat göms mycket information i databaser.

Det finns även information som sökmotorerna inte indexerar av utrymmesskäl och faktiskt även av "etiska" skäl.

Vill man göra en uttömmande sökning krävs det därför att man också söker på den osynliga webben.

Bästa tjänsterna för närvarande är: *Invisibleweb.com*, Direct search (<http://gwis2.circ.gwu.edu/~gprice/direct.htm>) och *Completeplanet.com*.

Tips tio

Glöm inte heller bort diskussionsgrupper och frågetjänster!

Hittar man inte svaret på en fråga kan det bero på att frågan är så specialiserad att svaret kanske inte finns pub-

Egroups är ett bra ställe att gå till för att hitta e-post-listor i olika ämnen.

licerat på Internet, eller är mycket svårt att identifiera. Men kanske finns det i någon annan människas huvud. Ett utmärkt sätt att lite mer informellt nå andra människors kloka kunskaper är diskussionsgrupper och frågetjänster.

E-postlistor är diskussionsgrupper som man prenumererar på via e-post. Hittar man bara rätt grupp så kan det vara en guldgruva. Om det är rätt grupp avgörs av ämnet som diskuteras, vilken kunskap de som diskuterar besitter och deras vilja att svara. Här kan man ställa frågor via e-post till flera människor samtidigt. Har man tur får man ett exakt svar på nolltid och har man otur får man ingen kommentar alls, för att det kanske visar sig vara en inaktiv e-postlista med ett fåtal prenumeranter.

Det handlar främst om att hitta "rätt" listor. Vissa e-postlistor arkiverar också alla e-postbrev, som det går att söka och bläddra bland innan man beslutar sig för att bli prenumerant och ställa en fråga.

En del e-postlistor hittar du på *egroups.com*.

Nyhetsgrupper fungerar som elektroniska anslagstavlor. Man går in i en nyhetsgrupp, bläddrar och söker bland tidigare meddelanden och ställer egna frågor som förhoppningsvis någon annan svarar på.

En del nyhetsgrupper hittar du på www.liszt.com/news

Frågetjänster finns det lite olika varianter av. Här kan du ställa frågor till kunniga personer som i vissa tjänster tar betalt och i vissa tjänster svarar dig helt gratis. I Askme.com kan man anmäla sig som redaktör i ett ämne man anser sig behärska. Ju snabbare och ju bättre svar du ger desto högre poäng får du och kan hamna på en tio-i-topp-lista. Det är andra deltagare som genom att ranka dina svar avgör om du ger bra svar. Du kan naturligtvis ställa egna frågor också.

Eller varför inte helt enkelt fråga biblioteket på nätet? "Fråga biblioteket" heter en frågetjänst där 19 folkbibliotek i samarbete svarar på frågor. Dem kan man hitta här: <http://www.dds.se/bibl/fraga.htm>

Om man trots allt, oavsett vilka sökmetoder man använder inte hittar ett tillfredställande svar, kan man aldrig vara helt säker på att svaret inte finns därute någonstans på Internet. Att förbättra sina sökningar kan många gånger helt enkelt handla om att bli mera säker på att svaret inte finns därute någonstans. För myten om att allting finns på Internet måste en gång för alla avlivas.

Så glöm inte uppslagsböckerna i din bokhylla, faktaböckerna på biblioteket eller kanske din granne som påstås kunna allt och lite till om fanerogamer och kryptogamer.

Lars Iselid