Netið

Netið er sannkallaður ævintýraheimur þar sem finna má hvaðeina, hvort sem það eru uppskriftir að indverskum grænmetismat, leiðbeiningar um samsetningu á tölvum og hátæknibúnaði, upplýsingar um frægt fólk, helstu bókmenntaverk heimsins eða ferðaupplýsingar um framandi staði. Svo er komið að allt sem nöfnum tjáir að nefna er að finna á Netinu, en sá er hængurinn að í takt við aukið streymi af efni inn á Netið verður æ erfiðara að finna það sem leitað er að.

Netnotendur

Netnotendur eru eitthvað í kringum 300 milljónir, en tölvurnar sem tengdar eru Netinu ríflega 75 milljónir. Til þess að leita á öllum þessum milljónum tölva í milljarði vefsíðna þarf öflug tæki og í þeim efnum hefur orðið mikil og hröð þróun á síðustu mánuðum eftir því sem samkeppni hefur harðnað á leitarsviðinu. Síður á vefnum fóru yfir milljarðinn fyrir nokkru sem samsvarar tæpum 20 terabætum af gögnum. Á hverjum degi bætist við hálf önnur milljón síðna, en flestir spá því að þegar hægja tekur á stækkun vefjarins eftir 10–15 ár verði síðurnar komnar vel á tíunda milljarð.

Leitarvélar eða síðusöfn

Ýmist nota menn leitarvélar eða síðusöfn til að leita sér upplýsinga á vefnum, en mörgum hættir til að rugla þessu saman. Þannig er Yahoo iðulega kallað leitarvél, en er í raun síðusafn, því ekki er þar skráð nema lítið brot af þeim síðum sem finna má á vefnum.

Starfsmenn Yahoo sjá um að skrá síðurnar og yfirfara þær, en fjölmörg áþekk síðusöfn eru til.

Þannig má t.d. nefna www.ebig.com, sem er vefsetur tengt alfræðibókinni Encyclopædia Britannica og about.com sem hét The Mining-Company í eina tíð. Sambærilegt safn við Yahoo er rekið af sjálfboðaliðum og kallast ýmist www.whoohoo.com, www.newwhoo.com eða dmoz.org.

Hinar eiginlegu leitarvélar leita ekki á öllum vefnum, það tæki gríðarlegan tíma að leita á milljarði vefsíðna á 75 milljón tölvum, og því beita leitarvélarnar fyrir sig róbótum sem eru á sífelldum þeysingi um vefinn að leita uppi síður og skrá innihald þeirra í gagnagrunn viðkomandi leitarvélar.

Lycos

Í árdaga Netsins var helsta leitarvélin Lycos-köngurlóin og Lycos er reyndar enn með öflugustu leitarvélum, þó hlutverk þess og starfsemi sé orðin mun umfangsmeiri.

Lycos hefur verið fljótt að henda á lofti nýjungar, jafnvel umdeildar, og þannig keypti fyrirtækið norska ftp-leitarvél og setti upp á setri sínu þrátt fyrir deilur um að hún stuðlaði að útbreiðslu á ólöglegum hugbúnaði.

Aðrar stórar leitarvélar eru t.d. Excite, www.excite.com, og HotBot, www.hotbot.com, sem báðar eru allöflugar. Sú síðarnefnda er þó talin betri.

Lengi vel var stærsta leitarvélin AltaVista. Hún var sett var af stað til að auglýsa geysiöflugan örgjörva sem kallast Alpha. Til að kynna hann setti fyrirtækið upp leitarvél sem það kallaði AltaVista, www.altavista.com.

AltaVista varð snemma öflugasta leitarvélin og dugði því vel til að sýna fram á burði Alpha örgjörvans.

Leit á Netinu

Hvernig á að leita?

Þegar leitað er í gagnagrunni sem inniheldur mörg hundruð milljón síður er eins gott að skilgreina nákvæmlega hvað það er sem leitað er að. Vilji menn til að mynda finna upplýsingar um Jaguar-bifreiðar verða þeir að gæta að því að einnig er til dýr af kattaætt sem heitir þessu nafni, gítar frá Fender og leikjatölva frá Atari sem var vinsæl í eina tíð, svo dæmi séu tekin. Sé aðeins slegin inn jaguar í altavista.com koma upp 337.910 síður. Við slíka leit má reyndar sjá að AltaVista er með grófflokkun á niðurstöðum, því efst á síðunni eru helstu niðurstöður flokkaðar saman, bifreiðar, leikjatölvur og gítar. Með því að slá í leitarreitinn jaguar -cat-fender -atari fækkar síðunum jniður í 301.440, sem er vissulega ágætis árangur, en alls ekki nógu gott. Með því að bæta við +car fækkar síðunum í 194.525. Síðan má enn fínstilla leitarskilyrðin þar til kominn er hægilegur skammtur af síðum.

Ef finna á orð sem standa saman eða setningarhluta eru orðin sett innan enskra gæsalappa, þ.e. gæsalappir uppi fyrir framan og aftan orðið: "jaguar xj". Þannig leit skilar 3.184 síðum, en sé leitað að xj og jaguar án gæsalappa koma 362.210 síður þar sem hvort orðið fyrir sig kemur fyrir og orðin bæði að auki. Vitanlega má einnig nota + og - í leit þar sem gæsalappir eru notaðar og þannig útiloka ákveðnar setningar.

Hástafir og lágstafir

Ef orð eru skrifuð með lágstöfum gengur leitarvélin út frá því að sama sé hvort þau séu með lágstöfum eða hástöfum. En ef hástafir eru aftur á móti notaðir finnur hún aðeins þau tilvik þar sem orðið er skrifað á þann hátt. Þannig skilar xj og jaguar 362.210 síðum eins og getið er en sé slegið inn XL og Jaguar verða síðurnar 297.055.

Einnig er hægt að leita að stöku orði á einu vefsetri með því að skrifa orðið, t.d. með + fyrir framan, og síðan vefsetrið með + og host: fyrir framan: +leitarorð+host:vefsetur.lén.

Eins má útiloka vefsetur með því að nota +host:vefsetur.land. Ef útiloka á t.d. heilt land, er notað -domain:land (-domain.is útilokar allar vefsíður með íslenskum lénum).

Hægt er að leita eftir heiti síðna til að þrengja leitina enn frekar með því að skrifa title: "heiti síðu" því gera má ráð fyrir því að höfundar vefsíðna setji ekki lykilorð í heiti síðnanna nema á þeim sé efni um viðkomandi.

Hjá AltaVista er einnig hægt að leita að myndum og hljóðum. Til þess eru sérstakir flipar en einnig má slá í venjulegan leitarglugga: media: image: eða audio: á undan leitarorðinu. Ef skrifað er inn image:cartman skilar leitin mynd af Eric Cartman, audio:cartman skilar hljóðum sem hann hefur gefið frá sér og media:cartman, hreyfimyndum o.s.frv.

Að lokum er svo hægt að leita að slóðum og tenglum í síður. Með því að setja url: fyrir framan leitarorð koma síður sem eru með leitarorðið í slóðinni. Sé aftur á móti sett link: og slóð þar fyrir aftan koma upp síður sem eru með tengla í viðkomandi slóð. T.d. skilar url:.is þeirri niðurstöðu að 484.849 síður séu til í safni AltaVista með slóð sem endar á .is. Leit með link:www.mbl.is skilar 766 síðum sem bendir til þess að 766 síður séu með tengil á Fréttavef Morgunblaðsins á sínum síðum.