

Pelle Snickars

I de mörkaste datahallar

UNDER ANDRA HÄLFTEN av nittiotalet lär två mer eller mindre suspekta personer strukit omkring i närheten av Stanforduniversitetets godscentral. Febrilt och aningen desperat letade de efter hårdvara. När en större leverans med nya skinande datorer anlände var de två efterhängsna unga männen raskt framme och försökte vigma en överbliven maskin för sin forskning. Inte sällan fick dessa doktorander i datalogi tag i en restnoterad burk, som de ivrigt sällade till sin eskalerande maskinpark uppe på campus. Den forskningsuppgift de satt sig för att undersöka – huruvida en sökmotor som analyserade relationerna mellan olika webbsidor kunde producera mer kvalitativa resultat än befintlig teknik – krävde nämligen en helt enorm datakraft. Hårddisk staplades på hårddisk, och enligt uppgift lär de spenderat mer än 15 000 dollar ur egen ficka på sitt databestånd. Att organisera all information som fanns online medförde inte bara ett antal övertrasserade kreditkort, i takt med att webben expanderade föreföll uppgiften också mer eller mindre omöjlig. Men Sergey Brin och Larry Page var lika uthålliga som målmedvetna i skapandet av det som snart skulle bli internets mest hajpade och framgångsrika företag: Google.

1997 hade de lyckats samla ihop data-maskiner som sammantagna hade ungefär en terabytes lagringsutrymme. Men det

räckte ändå bara för att indexera 26 miljoner webbsidor, långt mindre än vad etablerade sökmotorer baserade sina eftersökningar på. Googles träffbild var dock utmärkt och det rena gränssnittet tilltalande, även om de cirka tiotusen användare som då varje dag surfade in på <http://google.stanford.edu> fick ha tålamod. En sökning tog ofta tio sekunder, och därtill knyckte Brins och Pages applikation halva universitetets bandbredd. Det var förstås inte acceptabelt i längden, och beslut togs därför att lämna universitets trygga famn och ge sig ut på den kommersiella marknaden i jakt på riskkapital. I begynnelsen av dot.com-boomen visade det sig nu inte vara speciellt svårt, och Googles popularitet steg i takt med att investerare sköt till pengar. De tiotusen dagliga användarna var snart uppe i hundratusen.

Den stora utmaningen låg emellertid i att bygga ett system med förmågan att anpassa sig till den explosiva webbutvecklingen under slutet av nittiotalet. Tidigare hade ledande sökmotorer som AltaVista och Yahoo förlitat sig på ett slags semi-manuell indexering av webbsidor. Det hade fungerat ganska bra så länge utbudet av sajter inte var alltför omfattande, men det geniala hos Google låg i insikten – och tajmingen – om att endast ett automatiskt mjukvarusystem, deras så kallade PageRank-system, skulle

ha förmågan att hänga med i den vansinnigt snabba tillväxten av world wide web. Mänskliga redaktörer ersattes följaktligen av matematiska logaritmer. Hur sökningar presenterades för användaren var enbart en fråga om diskret binära kalkyler.

Kring millennieskiftet hade knappast någon riktig koll på hur många sajter som det egentligen fanns. Glasklart var emellertid Googles behov av mer eller mindre obegränsad datakapacitet för att indexera allt detta innehåll. Principen var att när någon gjorde en sökning i Google så letade sökmotorn inte direkt fram resultatet på webben, snarare matchades söktermen mot ett index baserat

på kopior av webbsidor som Google tidigare samlat in och lagrat på sina servrar. På så vis kunde Google på mindre än sekund presentera adekvata resultat, och målet som Brin och Page då satte upp var att man skulle ha prestanda att indexera åtminstone en miljard webbsajter. Till skillnad från sina konkurrenter satsade man på att bygga upp företagets datakraft genom egenkonstruerade datarack med enkel och standardiserad PC-hårdvara. Det var kostnadseffektivt, och dessutom öppnade Google långt före alla andra våren 2000 ytterligare en datacentral på östkusten i USA. Även om elektroner förflyttade sig blixtnsnabbt över internet påverkade geografiskt avstånd tidsintervallet för sökresultat. Dessutom blev själva sajten google.com mindre sårbar genom denna binära redundans och långt mer driftsäker.

Framgångssagan kring Google – ett företag som idag har mer än 20 000 anställda omsätter hundratals miljarder dollar – kan

på många sätt reduceras till historien om hur ett skalbart hårdvarusystem byggdes upp. Uppgifter gör gällande att Google idag har uppemot en miljon datorer och servrar till sitt förfogande – och möjligen är de ännu fler. Dessa maskiner är spridda över kontinenter och utlokaliserade i regionala datacentra, men ändå sammankopplade till världens största superdator. Det exakta antalet datorer förefaller vara en affärshemlighet och Eric

Elaka tungor påstår till exempel att företagets enorma inskanningsprojekt av världslitteratur inte främst är avsett för oss användare. Den primäre läsaren är Googles eget datasystem.

Schmidt, som tillsammans med Brin och Page leder Google, har på en direkt förfrågan svarat att han faktiskt inte har en aning om vare sig hur många datorer, datahallar eller datacenter som Google egentligen äger.

Ändå är det i dessa dunkla och nedsläckta datahallar, där bara det surrande ljudet från fläktar hörs, som framtidens informationslandskap just nu håller på att ta form. Om avstånd, positioner och platser alltmer förlorat sin betydelse i den globala tidsåldern så befinner sig likafullt »det virtuella» helt lokalt i dessa datahallar. På många sätt var just Google det första företag som på allvar insåg det som idag är uppenbart, nämligen att dator och nätverk konvergerar. Under de senaste åren har det växt fram en ny infrastruktur av personifierad information, vilken inte längre befinner sig lokalt på den egna hårddisken utan online i internets nätverk. Google har varit ledande i att flytta datanyttjandet från ens »desktop» till ett slags »webtop», inte minst genom lanseringen av Google Apps där all ens personliga information kan lagras online. De flesta inom IT-branschen menar att detta globala nätverk håller på att ta över det mesta, ofta sammanfattat i det som brukar kallas »cloud computing». Till

och med Microsoft har insett att en betydande förändring inträffat, i en intervju i Wired för en tid sedan framhöll företagets mjukvaruchef Ray Ozzie att nya versioner av Office knappast kommer att generera några betydande inkomsterna framöver. Istället måste Microsoft (som alla andra) ut på webben, och för tillfället jobbas det intensivt med Windows Azure, ett online-baserat operativsystem. Andra har varit betydligt snabbare med att ställa om; Adobe lanserade för en tid sedan Photoshop.com där var och en numera kan fixa till sina bilder online, och på Amazons servrar hyr sedan flera år tillbaka hundratusentals applikationer och utvecklare in sig.

För branschens aktörer är det förstas nya marknader och affärsmöjligheter som lockar, men informationsteoretiskt innebär skiftet också en fundamental förändring kring hur man ser på exempelvis kategorier som »dator» eller »webb». Flera mindre aktörer bygger idag sina applikationer på och inte för internet; nätet i sig är snarare det nya operativsystemet. Samtidigt menar kritiker, bland annat företrädare för rörelsen kring Öppen källkod, att det digitala molnet är en marknadshajp där användare skall luras att ladda upp personlig information i privata moln som ägs och sköts av företag. Naturligtvis har Google tillgång till den information som man lagrar på deras servrar – även om de givetvis hävdar att de aldrig skulle utnyttja den. Från ett sådant perspektiv framstår molnet som betydligt mörkare. Ändå hävdar en rad IT-gurus som exempelvis Kevin Kelly att den här utvecklingen är ofrånkomlig. På en konferens nyligen i San Francisco, »Web 2.0 Summit», menade han att eftersom våra medier konvergerar har vi snart bara en enda gemensam medial plattform vare sig det gäller tv, press, radio eller film; allt befinner sig online och drivs av samma slags webbaserade maskin. Kelly menar att tre

övergripande rörelser förmodligen kommer att präglade webben framöver; en rörelse upp i det digitala molnet, en rörelse ner i gigantiska databaser och datahallar – och en rörelse mot ett slags generellt delande. Den information existerar nämligen inte, enligt honom, som inte är en del av molnet och som alla kan nå. Den senare kategorin utgör just grunden för de webbaserade applikationer, som till exempel Google Maps, vilka genom »mash-up»-teknologi kombinerar molnets data eller tjänster med varandra. Det finns faktiskt de som hävdar att om Gutenbergs rörliga bokstavstyper var de moduler som boktryckarkonsten vilade på, så kommer ett slags snarlika webbaserade programmoduler att utgöra grunden för framtidens informationslandskap.

Den här förskjutningen, där våra datorer i allt mindre utsträckning förstås som enskilda och separata enheter, är en förändring som inte bara gäller traditionella PC-maskiner. Nästa generation av mobiler, kameror, musikspelare eller konsoler kommer också att ha »online» som defaultvärde, så kallade mids (Mobile Internet Devices). Det innebär bland annat att det inte finns någon anledning att spara ett fotografi lokalt i digitalkameran, när bilden redan automatiskt lagrats uppe i molnet så fort den tagits. Samtidigt är framtiden på flera sätt redan här. Med den lilla applikationen Dropbox kan man exempelvis på ett bagatellartat sätt (genom att lägga sin fil i en mapp) använda webben båda som lagringsplats och fildelningsserver mellan sina datorer. Den populära svenska sajten Spotify utgör ett annat exempel. Förvisso är det inte en webbaserad tjänst, utan en klient som installeras lokalt och till vilken musik streamas från företagets servrar. Ändå är principen densamma; om all musik – Spotify lär ha mer än två miljoner låtar – finns tillgänglig så fort man slår på datorn, så finns det egentligen

inte någon anledning att vare sig äga eller lagra dessa filer lokalt.

Icke desto mindre är grunden för »cloud computing» långt ifrån ett moln på drift i en upplyst stratosfär. Snarare återfinns det binära molnet i de mörkaste av datahallar. Visserligen har det på sistone svävat över i ett slags mobil variant; Sun microsystems har nämligen utvecklat rörliga datacentra i form av containers med hårdvara som enkelt kan kopplas ihop och staplas på varandra. Men i regel är molnets datacentra byggda på geografiskt avskilda platser där markpriset är överkomligt, gärna invid kraftstationer för billig eldistribution. I dessa kyliga och närmast posthumant kliniska datahallar står rack efter rack med grå hårdvara, och endast kabelhärvor och blinkade dioder avslöjar att burkarna kommunicerar med varandra och omvärlden. Kanske är det här som maskinerna kommer att vakna framöver, men än så länge vakar de mest över nätets innehåll och virtuella verkligheter. I egenskap av digitala moln håller maskinerna i dessa datahallar reda på tre grundläggande informationslager. Det första gäller själva infrastrukturen, det andra de olika centraliserade applikationer som datorerna driver, och det tredje lagret utgörs av ett slags periferi vilken interagerar med omvärlden.

I dessa informationsfabriker är det mjukvara och inte människor som utför själva arbetet. Här läggs därför även grunden till det slags artificiella intelligens som bland andra Google jobbar idogt på att förfina. »More data is better data» lär vara ett av företagets ledord, liksom devisen att man inte skall vara elak: »Don't be evil». I form av ett slags iGod menar dock some- lliga kritiker att Google är tämligen sken- heliga, detta eftersom företaget oupphör- ligen samlar på sig mer och mer data om vårt beteende online. Googles superda- tor; detta nätverk av servrar, datarack och

trafikswitchar, har ju åstadkommit inget mindre än en gigantisk databas av mänsk- liga intentioner. Eftersom träffbilden i sökmotorn baserar sig på en algoritm vars resultat beror på tidigare sökningar blir systemet alltmer förfinat. Sofistikerad användarinformation förs genom feed- back-loopar tillbaka in i superdatorn, med syfte att göra denna så artificiellt intelligent som möjligt. Elaka tungor påstår till exem- pel att företagets enorma inskanningspro- jekt av världslitteratur inte främst är avsett för oss användare. Den primäre läsaren är Googles eget datasystem.

Egentligen har utvecklingen av sökmotorer alltid varit kopplad till ett slags syntetiskt tänkande. Brin och Page är ju dataloger – inte affärsmän. Numerisk analys och beräkningsvetenskap är lika centrala för Google som företagsnamnets ursprung. I matematiken är talet »googol» lika med 10100 , det vill säga en etta följd av hundra nollor. Det nuvarande huvudkvarteret i Kalifornien går dessutom under namnet Googleplex – även detta en referens till det ofattbart stora talet gogoolplex, tio upphöjt i tio upphöjt i hundra. I en presentation på Stanfordinstitutet sommaren 2002 påpekade Larry Page »att den ultimata sök- motorn är lika smart som människor – eller smartare. Att arbeta med sökteknologi är att arbeta med artificiell intelligens.» I andra sammanhang har han liknat Google vid superdatorn HAL i Stanley Kubricks film 2001, ett rymdäventyr (1968) – en dator som ju alltid låg ett snäpp före IBM (bokstaven »I» kommer före »H» etcetera) – men utan den bugg som gjorde att HAL tog kål på sin besättning.

Att den fiktiva datorn HAL, med sitt pulserande röda kameröga fabulerades fram under sextioalet är symptomatiskt efter- som det var då som forskning kring ska- pad eller artificiell intelligens (AI) tog fart.

En förgrundsgestalt var datalogen John McCarthy, som också myntade själva termen. Men egentligen är portalfiguren den lika skarpsinnige som tragiske matematikern Alan Turing. Om Googles sökmotor idag framstår som ett slags embryonal AI-maskin, vilken på basis av tidigare sökmönster på förhand (snart) kommer att veta vad en användare letar efter redan innan hon ställt en fråga, undrade Turing redan 1950 huruvida digitala maskiner egentligen kunde tänka. I artikeln »Computing Machinery and Intelligence» lanserade han idén om det så kallade Turingtestet. I korthet innebar det att om en person var oförmögen att skilja ett datasystems handlande från det beteende som en människa visade upp, så måste detta system anses vara intelligent. Dessa reflexioner grep tillbaka på den essä Turing skrivit som 24-åring i mitten av trettioalet, »On Computable Numbers», där han fullständigt genialt formulerat den teoretiska grunden för all modern datateknik. Turings bidrag till matematikens teori var en kombination av mekanik och talteori; något som för samtidens Cambridgematematiker framstod som »shockingly industrial». Turing tänkte sig en helt abstrakt maskin – Turingmaskinen – som till skillnad från andra gamla maskiner inte blivit omodern. Turingmaskinens låda kunde läsa och skriva på en tänkt remsa indelad i rutor. I varje ruta skrevs antingen 1 eller 0, eller så kunde rutan förbli blank. Turingmaskinen var alltså helt rudimentär, men utgjorde likafullt prototypen för den moderna datorn.

Värt att notera är att Turing var övertygad om att det gick att jämföra det mänskliga tänkandet med hans maskin; den kunde ju i princip utföra vilken mental

process som helst. Likheterna med hur skaparna av Google resonerat är påfallande; bara genom att låta maskinerna arbeta ostört kunde deras system optimeras. Liksom Turing har Brin och Page i många sammanhang uttryckt en betydande tilltro till den artificiella datateknikens möjligheter och utsikter; potentialen är enorm – »den ultimata sökmotorn skulle förstå allting i världen», som Page påpekade för ett år sedan. Föga förvånande sponsrar just Google sedan en tid tillbaka Turingpriset, datavärldens finaste utmärkelse.

Avslutningsvis är det likväl en vedertagen sanning inom matematisk kunskapsöverföring att det rigorösa och exakta ofta blir en belastning. Precisa detaljer riskerar att skymma helheten, och vaghet i kommunikationen kan ibland vara en fördel för att göra något begripligt. Det gäller även sökmotorer och Googles själlösa algoritm har mycket riktigt krockat rejält med den sociala webbens öppenhet. Företaget har kanske alltför länge varit sig själv nog, och alls inte haft behov av de kolloberativa processer och användarorienterade applikationer som utgjort basen för webb 2.0 under de senaste två åren. Men här måste de tveklöst tänka om, annars kommer de att drabbas av en backlash i ljuset av den typ av »social search» som blir alltmer omskriven. Mahalo.com är till exempel en sökmotor som inte är algoritm-baserad, utan istället drar nytta av den sociala webbens struktur där användare rekommenderar länkar till sådant man söker efter – det vill säga ett slags redigerat, snarare än automatiserat informationsflöde. Och till skillnad från nittioalet är det inte längre ett fåtal redaktörer som väljer, utan webbens miljoner användare.■