

# Sökteknologi och personlig integritet

Jussi Karlgren

*i Tekniken bakom språket, Rickard Domeij (red) Norstedts. 2008.*

*Tekniken för att söka efter information på nätet blir allt mer avancerad. Det är förstås bra på många sätt. Men det gör också att sökmotorerna kan ta reda på allt mer om oss för att kunna anpassa sökresultat och annonser efter våra intressen. Vill vi det? Och kan vi vara säkra på att uppgifterna inte också används i andra, mer tvivelaktiga syften? Det handlar om vem som ska ha makten över informationen. Jussi Karlgren beskriver situationen, tekniken bakom och de möjligheter vi har att skydda vår integritet på nätet.*

Vårt kommunikativa beteende berättar mycket om oss; med visst fog kan vi säga att det inte finns något viktigare eller bättre sätt förstå vilka vi är än att notera vad vi säger till vem och hur vi gör det.

Det mänskliga språket, vårt kommunikationsverktyg, har genomgått flera teknologiska språng under mänsklighetens senare historia. Vi har lärt oss att vi med hjälp av skrift inte bara kan kommunicera med dem vi har närvarande just nu, inom hör- och synhåll – vi har utvecklat tekniker för att skicka både skrift och tal över tid och rum till motparter bortom horisonten. Vi har en ganska god förståelse för vad språkliga yttranden präntade på papper har för betydelse för oss som skribenter, för mottagaren, vi vet hur texter kan lagras och åter plockas fram vid behov. Vi har sociala konventioner (och i vissa fall till och med tvingande lagregler) för vad som är lämpligt att fästa på papper och vad som bäst sägs ansikte mot ansikte.

Vad vi yttrar, säger, skriver till någon annan är ofta privat och vi vill för det mesta att det ska fortsätta vara så. Men digital teknologi som hjälper oss kommunicera lättare och oftare med nära vänner, avlägsna bekanta och okända likasinnade kan också ha sidoeffekter: den kan röja för andra vilka vi är, vilka vi kommunicerar med, vad vi säger och hur vi tänker. Den kan göra det utan att vi själva märker det och utan att det finns en illasinnad motpart som vi kan identifiera, ansvarsbelägga och lagföra. Det är inte någon annans fientliga motiv vi ska akta oss för, utan den omärkliga glidningen i informationsbalans tekniken i sig orsakar.

## **Gamla vanor gäller inte längre**

Digital teknologi bryter mot och överväldigar en del av de vanor, regler och verktyg vi som språkliga användare har. Digitala dokument uppvisar på många sätt viktiga skillnader gentemot tidigare språkligt material. Enskilda texter som på papper ser olika ut, har olika dignitet, har olika emotiv och social kraft, ser alla likadana ut i sin digitala form. Digitalt lagrade samlingar av texter är svåra att få överblick över utan speciella verktyg. Inspelat tal i digital form är än svårare att hålla reda på än texter. Det finns ingen självskrivna mekanism för att bläddra, sova och gallra i materialet när tillfälliga inköpslistor, klotter, anteckningar skrivna i affekt eller i stundens ingivelse har samma ontologiska status, samma utseende och samma närvaro som vetenskapliga avhandlingar, privata dagböcker, nyckelromaner, kontrakt, lagfarter eller pantbrev.

De intellektuella verktyg vi har att reda oss med när vi försöker förstå digitala kommunikationsyttringar är de vi har från den pappersbaserade verklighet vi hittills befunnit oss i – och en del av de intuitioner vi har håller inte i översättning.

Vi är idag vana vid att skriva brev elektroniskt, men gör det utan att sätta dem i kuvert. Så skulle vi inte bete oss om korrespondensen fortfarande skulle gått på papper och vi hade haft känslan av att varje handpräntat pappersbrev hade gått att följa från oss direkt till vår mottagare genom att brevbäraren känner till vilka vi båda är.

Ännu mindre hade vi gjort det om vi vetat att brevbäraren också är postmästare, handelsman, bibliotekarie och växeltelefonist i byn och att bibliotekarien behåller en kopia på varje brev och varje tidning vi skickat och fått. Och att fjärdingsman, rektor och söndagsskolelärare får glutta i pärmarna vid behov.

Vi skickar våra digitala brev på öppna vykort i förvissning om att några av våra vanligare pappersbaserade intuitioner fortfarande gäller: det som vi inte själva ser på vårt skrivbord ser ingen annan heller, att vi själva är obetydliga och omärkliga i den stora mängden människor vi omges av och att saker som finns i stora osorterade säckar, högar och travar är svåra att bringa reda i. Vi räknar med att vår post försvinner i mängden av annan post och att, liksom vi själva, ingen annan kan hålla reda på allt innehåll. Men det är inte rätt tänkt, längre.

### ***Makten över information***

Informationsåtkomstteknologi har det senaste halvsekle väntat på att få maskiner som är stora nog att kunna hantera stora datamängder och på att få tillgång till datamängder som är stora nog att bli verkligt användbara. Vi är där nu. Och när informationsåtkomstteknologin tillämpas på öppet tillgänglig information vi själva producerar och tillhandahåller, när sökmotorer faktiskt fungerar för att hålla reda på oss och vårt bör vi tänka efter en gång till.

Sökmotorer är nyckeln till makten över information. Utan möjlighet att slå i läggen, att bläddra i pärmarna, behöver vi inte oroa oss för vare sig handelsmannen eller söndagsskoleinkvisitorn. I dag finns sökmotorerna där för att hålla reda på den post vi skickar, de texter vi skriver, de samtal vi gör över telefonnäten. Det är inte bara en osorterad postsäck, en mängd anonyma användare eller ett stökigt skrivbord längre.

### ***Sökmotorerna vill veta mer om oss***

Sökmotorer baserar sig på kunskap om dokumenten de hanterar. För att kunna tillhandahålla användbara resultat måste de analysera dokumentsamlingen på rätt sätt: tillräckligt djupt för att kunna svara på frågor; tillräckligt ytligt för att kunna hantera informationen effektivt. Sökmotorer fungerar så att de läser varje dokument de får tillgång till, tillexempel genom en spindel som följer länkar mellan dokument, de skapar ett sökregister, ett index, med termer skördade ur dokumenten och de tillhandahåller sedan söktjänster och sökfunktioner för att den som så vill skall kunna orientera sig i dokumenten ifråga. Till skillnad från sökregister sammanställda av mänskliga redaktörer och avsedda för direkt mänskligt bruk, sådana som finns i slutet på fackböcker, tillexempel, kan en sökmotor lägga i stort sett varje ord, varje fras, varje uttryck från dokumentsamlingen i sina sökregister – efter att ha gjort en teknisk och statistisk avvägning mellan fullständighet i sökregistret och brus från obetydliga och irrelevanta termer.

När en användare söker efter någonting brukar de ange ett par termer för att beskriva sitt informationsbehov; sökmotorn är byggd för att så snabbt som möjligt – omedelbart, ur användarens synvinkel – leverera en så fullständig lista på förhoppningsvis användbara måldokument som möjligt. I stort sett den enda kunskap de flesta sökmotorer har om sina användare är de två–tre ord som sökbehovet är uttryckt i. De vet att Erik söker efter ”recept rulltårta” och ”hagelbössor kompositmaterial” och att Maria söker efter ”skilsmässoadvokat

äktenskapsförord” och att de använder samma dator. Den fundersamma användaren kan dra slutsatser därur men det är inte alltid lätt att göra det riktigt.

Så det är givet att varje sökmotor arbetar mycket med sitt index – sökregistret måste ha rätt och riktigt innehåll. Men för att kunna förbättra sina resultat räcker inte djupare analys av dokumenten om inte den motsvaras av en mer finkornig analys av sökbehovet och av användaren. Utan mer kunskap om användaren hjälper det inte att läsa dokumenten noggrannare. Och det är den stora flaskhalsen för förbättrade sökmotorer idag. Sökmotorernas kvalitet lider av att de inte känner till mer om sina användare. Med mer kunskap om sina användare, om preferenser i allmänhet, om bevekelsegrunder för innevarande sökning i synnerhet, om egenheter och tidigare erfarenheter skulle söktjänsterna kunna leverera mer skraddarsydda och sannolikt mer tillfredsställande resultat.

Att bättre förstå sina användare är ett viktigt mål för sökmotorer idag. Hittills har de flesta användare dock visat sig kallsinniga till att själva lagra sina favoritsökningar, att logga in till sin egen personliga söksida, eller till att anta andra liknande erbjudanden från de större företagen. Sökmotorföretagen brottas med att förstå varför. Är inte förbättringarna som den mer personliga tjänsten levererar värda besväret? Är användarna oroliga för att avslöja för mycket om sig själva? Är de för bekväma för att orka registrera sig? Är gränssnitten dåliga? Det kommer att komma fler och mer lätthanterliga sätt för systemen att känna igen oss: det läggs mycket arbete på den frågan.

De största och mest ansedda sökmotorföretagen erbjuder kostnadsfria sökningar för allmänheten. Men någonstans får de kostnadstäckning – det är inte enbart av sitt hjärtas godhet de erbjuder kostnadsfria tjänster. Den som tar ett ögonblick för att fundera kan säkert lista ut att affärsidén för sökmotorföretagen ligger i de annonser som också erbjuds läsaren jämte sökträffarna. Ofta nog är annonserna faktiskt ganska intressanta och sökmotorföretagens tjänsteutvecklare ser noga till att de inte tar så mycket plats på skärmen att läsaren inte längre litar på de rena sökträffarna – som de själva säger ska ”användarupplevelsen” förbättras av annonserna, inte försämrats. När sökmotorannonser fungerar väl är de exempel på marknadsföring som den bäst ska utföras: relevanta, icke-störande informativa annonser som erbjuds läsare vid tillfällen när de själva är beredda att ta emot dem.

### ***Inga illvilliga motiv***

Det är inte en slump att alla de största sökmotorföretagen också erbjuder kostnadsfria brevlådor till sina användare: det är en logisk poäng eftersom de ju vill veta mer om sina användare. Med mer data får de en bättre bild av sina användare och kan dels erbjuda skraddarsydd sökning anpassad till de egenheter, de erfarenheter etc etc som användaren visar upp och dels tillhandahålla mer och bättre riktad marknadsföring. Med hjälp av den information de kan samla in om oss som använder systemen kan de avgöra hur de bäst kan förbättra sin informationsanalys och sitt sökregister.

Det är någonstans här vi som användare – som kunder i en affärsförbindelse vi inte medvetet inträtt i – måste skänka en tanke åt vad vi berättar om oss själva. Om vi tillhandahåller information om oss själva för att få mer personliga tjänster av ett informationsföretag – vad används den informationen till, vem vill använda den och på vilket sätt betalar vi för informationen vi får tillgång till? Har vi samma mål som vår affärspartner? Det finns fog för skepticism inför en del av de erbjudanden och en del av de automatiska förfaranden som vi utan att själva veta om initierar när vi använder kostnadsfria tjänster på nätet. En del opinionsbildare försöker också höja vår medvetenhet och pekar på hur information om användare lömskt kan missbrukas eller spridas vidare utan deras tillåtelse

eller vetenskap. Och rent juridiska regelverk kan bara ge oss en del stöd i en stor och gränslös värld.

Det är lätt att antagonister demoniserar varandra. Den som samlar information om just mig kan vilja stjäla min identitet och göra affärer eller pinsamheter i mitt namn; den som följer mina spår kanske är ute efter att skada eller förfölja mig personligen; den som skickar reklam till mig kan tänkas vilja lura mig; den som övervakar säkerheten i ett samhälle kan tänkas vilja urskillningslöst snoka reda på medborgarnas privata hemligheter; den som vill skapa konsumentprofiler för riktad marknadsföring gör bådadera och det på ett dolt och skumt sätt. Sådana brottslingar, illdådare, paranoiker och lurendrejare finns helt visst, lika väl som organisationer vi i förstone litat på kan använda privata data om sina klienter eller sina medborgare för att köpslå med andra så att information om oss kan läcka vidare till parter vi inte ens känner till: andra företag, andra organisationer eller byråkratier hos främmande makt. Skräckscenarierna baserar sig på en illvillig motpart och en omdömeslös användare. Men en tafflig motpart är nästan lika illa. Ibland ser vi i pressen nyheter om hur viktiga eller personliga data råkar bli öppet tillgängliga; oftare får vi inte ens reda på blamagen.

Men bör vi då försöka hindra ansvariga personer genom att röja deras existens och förbjuda deras aktiviteter? Nej, inte i första hand! Det räcker inte. Vi kan försöka begränsa kommersiella parters rätt och förmåga att samla data om sina klienter; vi kan försöka inskränka myndigheters rätt att ha tillgång till våra kommunikationskanaler, men demoner och dummisar, klåpare och klantar behövs inte för att ge oss anledning att vara bekymrade för vår personliga integritet och att måla sådana skräckbilder på väggen riskerar att hindra en vettig dialog. Det är lätt för marknadsförare och statsapparatusens kontrollorgan att bemöta argument om illvilja och ondsinhet där sådant tillskrivs dem eftersom det sällan är det som driver deras handlingar. Det är inte ondska och illvilja vi bör oroa oss för utan enkel, oförfalskad och försvarbar egennytta.

Denna egennytta är inte alltid i samklang med vår egennytta som konsumenter eller medborgare. Det räcker för att vi ska vilja skyla och värja oss. Men trots att det är fråga om en situation där vi vill försvara oss behöver vi inte leta efter mer avancerade motiv hos motparten än så. Vi behöver inte heller motivera våra egna handlingar med mer komplicerade motiv än egennytta: det är i vårt intresse att våra personliga särdrag och egenskaper endast delas ut till andra när vi själva vill och när vi själva får vinning av det.

Ansvar, oavsett lagregler, för att styra åtkomst till den information vi själva delar ut kommer ligga kvar hos den enskilda användaren. Det stora kruxet är att informationstekniken bjuder den ena parten på bättre mekanismer än den andra.

## ***Risk för informationsobalans***

Vi får ofta rådet att vara medvetna om vad vi gör offentligt. Bilder vi publicerar måste vi veta att vi publicerar. Skriver vi arga insändare till lokaltidningen får vi räkna med att grannen ser dem. Skriver vi ett brev till handlaren där vi berömmar eller gnäller på hans varor får vi räkna med att både han och de andra verksamma i handelsboden får reda på det. Men den medvetenheten är svår att upprätthålla om inte tekniken ger oss stöd i det. I dag ser kyrkoherden, handelsmannen och fogden alla vykort vi skickar eftersom vi inte har tillgång till kuvert. Vi har inte möjlighet att själva hålla reda på vad vi berättar om oss själva och att styra den bild vi låter andra få av oss.

Och förlorar vi makten över den bild vi själva skickar ut på nätet så riskerar vi att andra använder den för att komma åt oss – antingen genom att manipulera oss till handling vi inte är intresserade av eller genom att använda informationen mot oss antingen vi vet det eller inte. Vi riskerar råka ut för informationsobalans, där någon motpart vet saker om oss men vi inte

vet det. I en förhandling eller i en diskussion om vad som helst hamnar vi i underläge om motparten kan säga saker som ”men förut tyckte du ju inte så” eller ”men till dina vänner har du tidigare sagt att ...”. Som en marknadschef på ett större konsumentorienterat informationsföretag säger på direkt fråga: ”Vi vill såklart veta betydligt mer om människor än vad människor vill att vi ska veta. ... utgå från att vi vill veta det mesta. Även det som inte verkar relevant kan vara relevant, inte minst då man som marknadsförare lär sig att iakta vad människor faktiskt *gör* och inte vad de *säger*.”

Och världen förändras, men insamlade data består, och det med samma dignitet de hade den dag texten först skrevs, fotot först togs eller inspelningen först gjordes. Precis som på våra individuella hårddiskar det inte finns någon självskrivna överblick som hjälper oss gallra bland dokument så har de digitala datalagren ingen inbyggd utmönstringsmekanism, ingen rost, inga dammlösa och boklösa, inga bränder. Om allt du skrivit och alla samtal du fört är bokförda och tillgängliga för andra – inte nödvändigtvis för allmänheten men för viktiga samtalsparter – är det nog så viktigt att du själv vet vad du säger och vad du sagt.

Så det är tre saker vi bör hålla i minnet som har att göra med att våra egna intuitioner som är baserade på pappersteknologi inte är giltiga längre efter att det finns effektiv informationsteknologi tillgänglig. För det första att det vi gör, även på nätet utan att märka det själva, ger andra en bild av oss som vi kanske vill utforma på ett sätt som överensstämmer med det vi anser vara viktigt. För det andra att den bilden har ett affärsmässigt värde för affärsförbindelser som vi kanske vill eller kanske inte vill delta i och att vi därför inte bör dela med oss av den utan ersättning. För det tredje att även om vi tar bort dokument och själva glömmer bort vad vi gjort så finns informationen kvar för den som en gång samlat på sig den. Om vi inte bryr oss riskerar vi hamna i informationsobalans.

Även utan att tillskriva motparter ondskefulla avsikter kan informationsobalansen te sig obekvämt, obehagligt eller motbjudande. På ett samhälleligt plan aggregeras effekterna till att bli direkt skadliga. Om folk i allmänhet lägger tid på att försöka dölja sina spår, om folk misstror kommunikationssystemen – då finns risker för ett kommunikationsgap: att informationsspecialister, maktinnehavare och andra informationsbesuttna skaffar sig andra, mer slutna, kommunikationssystem och att oprivilegierade användare får klara sig utan privat sfär. Ett sådant kommunikationsgap är inte ett jämviktssystem utan kan leda till att klasser av medborgare motarbetar och misstror varandra och till hastiga och oväntade samhälleliga och ekonomiska omvälvningar. Det är få som är betjänta av den sortens framtidsscenarioer. Och även utan mardrömsscenarioer så riskerar den oro och överdrivna försiktighet i kommunikationsmönster, det val av mindre väl anpassade kommunikationssätt som kan bli följden av misstro mot de offentliga systemen att dämpa den offentliga debatten och i förlängningen lamslå den öppna diskurs som finns på nätet och som idag ger liv åt och karakteriserar mycket av det offentliga samtalet i västvärlden.

## ***Språkteknologi ger kunskap om oss***

Vilken kunskap finns det då om identifierande personliga särdrag i kommunikationsmönster? I synnerhet ur ett språkteknologiskt perspektiv. Vad vi säger till vem och hur vi gör det – vad går det att säga om det? För att hitta och sortera dokument har vi söksystem; för att analysera dem vidare har vi informationsextraktionssystem, författarbestämningsalgoritmer och annan apparatur för informationsförädling. Det finns en hel del teknologi att tillgå för den som vill förstå meddelanden, i synnerhet för textuella data. Det mesta är ganska enkla algoritmer men som med de stora datamängder det rör sig om ändå kan ge en god bild av vad som sägs.

För att sätta dokument i sitt sammanhang går det att analysera sociala mönster och sociala nät, vilket ger en annan sorts analys – modellering av vilka vi kommunicerar med och hur vår kommunikativa omgivning ser ut. Märkligt nog anses inte det vara en del av språkteknologin, trots att det kommunikativa är en nog så viktig del av hur språk ter sig i bruk.

Det behövs inte någon speciellt sofistikerad nästagenerationstjänst för att samla spännande information om individer. Häromåret släppte forskningsavdelningen på America Online – en av USA:s större internetleverantörer – ett ansevärt stycke av sina sökloggar, alltså de sökfrågor deras användare ställt, fritt för forskningsändamål. Det var överraskande lätt för vem som helst med tålmod att tröska igenom dem att hitta underhållande men ur individuell integritetssynvinkel katastrofala berättelser. America Onlines ledning insåg snabbt vad som kunde finnas i loggarna och drog snabbt tillbaka erbjudandet. Men för sent, eftersom filerna spreds ganska snabbt till allmänheten under de första dygnet. Och det var ändå bara sökloggar – inga skrivna texter, inga uttryck för personliga preferenser förutom den som kunde läsas in i sökfrågorna. Har vi tillgång till mer information – till exempel genom att analysera e-postkommunikation – får vi både text och sociala sammanhang att arbeta med.

Söksystem är effektiva sätt att sortera material ämnesmässigt – som också beskrivs i kapitlet *Informationssökning på Internet* i denna bok. Teknologi för informationsextraktion kan plocka ut alla omnämmanden om något visst slags förlopp eller skeende – att ur nyhetstext hitta alla rapporter om flygolyckor eller rapporter om börskursrörelser är exempel ur forskningen. Det gör att ett system kan bevaka en ström av information för att se om någon slags händelse som intresserar uppdragsgivaren omnämns och vad som sägs om det. En sorts specialfall av informationsextraktion är attitydanalys som gör det möjligt att hålla reda på vad någon i skrift ansett om någonting. Attitudinella uttryck är uttryck som ”Jag gillar varm korv” eller ”Bertram har en utmärkt V-botten och tar motsjö fint.” Sådana uttryck har i typiska – men inte alla – fall någon explicit eller lätt inläsbar tyckare i uttrycket som är den som håller attityden, någon tydligt värderande term och någonting som omtalas och avses med attityden. Det går att skriva en språklig analyskomponent för attitudinella uttryck som fortlöpande samlar på sig attityder om något visst och med lite mer omfattande datamängder kan ett attitydanalysystem tränas till att identifiera även uttryck som använder omskrivningar och oväntade ordvändningar. Här finns förstås ett starkt kommersiellt intresse för att tråla efter åsiktsuttryck i text: företag som producerar varor och tjänster vill veta hur de omtalas bland kunderna; konsumenter söker lära sig av andras erfarenheter; media och politiska aktörer vill pejla opinionsläget hos allmänheten. Och då finns en anledning för oss som uttrycker attityder att fundera på huruvida vi vill att de ska lagras, om vi vill stå för dem i eget namn och vilka som ska få ta del av dem.

Rättslingvistiken och författarbestämningen baserar sig på enkel räkning av lexikala eller lättidentifierade strukturella språkliga enheter i texter. I svensk rättshistoria är nog Helanderfallet det mest kända exemplet: år 1952 sökte professorn i teologi vid Uppsala universitet Dick Helander bli vald till biskop i Strängnäs. Inför valet cirkulerade ett hundratal smädesbrev om Helanders motkandidater. Helander vann valet men åtalades och fälldes för av både tingsrätt och hovrätt. Han blev efter att flera år senare lyckats få högsta domstolen behandla ärendet friad från en del av anklagelserna. Rätterna tog hänsyn till många omständigheter i sina överväganden, och bland dem diskuterades beräkning av likheter och skillnader mellan språkliga data tagna ur breven och ur Helanders övriga produktion – vilket efteråt lett till en vetenskaplig diskussion om hurdana statistiska metoder bäst kan användas för de skäligen magra data som fanns att tillgå. I dag har flera rättsfall avgjorts med hänvisning till mindre datamängder än så. 2002 fälldes en person i Storbritannien för mord mot sitt nekande. Mördaren hade tagit mordoffrets mobiltelefon och skickat flera sms med den efter mordet för att fastställa sitt alibi. Efter språklig analys av sms:en ifråga kunde rätten fastställa att de var skrivna i fel stilsort och omöjligt kunde ha skrivits av offret, en tonårig flicka.

Ren råräkning av ordfrekvenser ger inte någon vidare särskiljningsförmåga. Skillnader mellan genrer och av språkbrukaren förutsedd samtalspart eller läsare har större effekt på de flesta språkliga variabler än den enskilde språkanvändarens egenheter. Men språkteknologin

står inte still och vi kan förutse mycket mer kompetenta metoder i takt med att vi får större datamängder och mer effektiva beräkningsmodeller. Vi vet att ordval varierar mellan författare – det är anekdotiskt känt sedan länge och systematiskt och beräkningsmässigt visat och satt i bruk sedan ett drygt hundratal år. Men ordval utgör inte hela eller ens den största delen av den möjliga språkliga variation språkliga data kan uppvisa. Författarbestämning görs dock i första hand på just ordval och enklare statistik på enklare observationer av enklare särdrag. Det fungerar nöjaktigt men är lite som att leta under lyktstolpen: texter, likaväl som tal, kan ju variera på så många andra sätt! Stavning bestäms till största delen av språksamfundet och dess småskolelärare, ordval och syntaktisk komplexitet till ganska stor del av ämne och genre, medan berättelsens och narrationens ämnesmässiga struktur och textualitet visar variation som i mycket högre grad är lämnad åt författarens omdöme.

Språkliga särdrag identifieras oftast genom att i en text eller annan diskurs observera och göra mätningar på något språkligt fenomen varje gång det förekommer, föra in mätningen i en tabell av något slag och sedan beräkna någon slags normaliserat medelvärde för fenomenets förekomstfrekvens. I denna text, tillexempel, kan vi se att författaren använt ”och” ett par hundra gånger och pronomen i första person likaså ett par hundra gånger varav nästan alla i plural. Tabellen ger en del mätningar på några av denna boks kapitel. Det är tydligt att det finns skillnader mellan författarna, redan med så här enkla mått!

Kapitel	semikolon	kommatecken	och	att	vi	man
Sökteknologi och personlig integritet	18	174	206	217	177	3
Nya metoder ger bättre maskinöversättning	8	307	204	154	14	74
Bättre tala än texta – talteknologi nu och i framtiden	2	215	140	224	52	36
Informationssökning på Internet	0	116	151	179	12	82

En sådan beräkning använder bara en del information som skulle kunna ha använts. I ett par experiment har vi tidigare visat hur särdrag kan aggregeras konfigurationellt, i sekvenser av observationer. I stället för att räkna enkla medelvärden (hur många *vi* har författaren använt; hur många *man* har författaren använt) visar vi hur ofta en observation följs av en annan liknande observation (hur ofta använder författaren *vi* i två på varandra följande meningar; hur ofta växlar författaren från *vi* till *man*). Den sortens särdrag, hävdar vi, är mindre bestämd av

språkliga regler och konventioner och mer tillgänglig för individuell variation som är härledbar till författarens preferenser, vanor och val. Poängen här, oavsett just våra specifika hypoteser, är att det här är ett område som just nu genomgår stora omvälvningar. De särdrag vi idag använder för att identifiera vem som skrivit vad kommer att bli mer sofistikerade i takt med att forskningen ger sig tid att analysera problemet med bättre verktyg. Även om vi skriver saker anonymt finns det möjligheter att lista ut vem som är författare, vad vederbörande talar om, vad den säger och hur den gör det.

## ***Hur kan vi värja oss?***

Som redan konstaterats är den viktigaste regeln vi bör agera efter att vara medvetna om vad vi gör, vad vi delar ut, hur vi framstår, vilka gardiner och förlåt vi lättar på och vilka vi håller öppna.

Därnäst är det en vettig princip att i offentliga sammanhang dölja eller modifiera vissa personliga egenheter som vi vet eller märker är speciella för oss själva. Vi bör undvika att publicera namnteckning, personnummer och kontonummer – detsamma gäller för andra egenheter som kan användas för att identifiera oss, som nära relationer till andra.

Och till sist kan vi försöka hålla andra nyfikna ovetande om vilka vi är genom att hålla på vår anonymitet, genom att bara bete oss på vissa sätt i vissa av oss själva valda sammanhang.

Var och en av principerna är naturliga för oss att följa i den fysiska världen, men i den digitala är det svårare. De flesta av oss har bara en eller två interaktionspunkter med den digitala världen och det kan vara svårt för oss att komma ihåg att vi lämnat vårt trygga arbetsbord eller bekväma hemmafåtölj och befinner oss i ett offentligt sammanhang. Vi kan försöka komma ihåg att ha flera e-postadresser – en för politiska diskussioner, en för pryltorget, en för arbetet; vi kan försöka komma ihåg att använda olika läsprogram för att prenumerera på nyheter, skvaller och bloggar, för att sköta bankaffärer, och för att köpa och sälja varor på nätauktioner. Vi kan försöka komma ihåg att använda olika datorer för att kommunicera med arbetskollegor, med statsförvaltningen, med våra vänner på fritiden och med kompisarna vi planerar en statskupp tillsammans med.

Men det är inte lätt för oss att hålla reda på, inte lätt att genomföra praktiskt och det är inte heller tillräckligt kraftfullt för att möta den analys som idag är möjlig och håller på att komma i bruk. Likaväl som analysverktyg för att känna igen och kategorisera beteende går det givetvis att bygga apparatur för att stödja privatlivets helgd. Den bild vi skickar ut om oss själva kan vi själva läsa av, redigera till viss gräns och komplettera; språkliga särdrag går att identifiera, varna för och redigera bort; kontakter, kommunikationsmönster och sociala nät som ser speciella ut eller som öppnar mellan tidigare avgränsade beteendemönster går att identifiera, varna för och undvika.

## ***Koder, chiffer och hemliga språk***

Att dölja det vi skriver genom kryptografi är ett sätt att dölja dessa särdrag. Kryptografi kan utföras på många sätt: tillexempel kan ett meddelande koda genom att viktiga element i det ersätts med andra element på ett sätt om avsändaren och mottagaren överenskommit.

”Bålgetingarna är i flaskan” rapporterar Allan till Rastapopoulos i albumet *Koks i lasten* när han fått Tintin och kapten Haddock i fällan. Ett meddelande kan också chiffreras: innehållet kan algoritmiskt förvanskas så att enbart mottagaren kan konvertera det tillbaka till ett begripligt meddelande. Meddelandet kan också döljas helt, eller begravas i någonting annat genom steganografi: ett textmeddelande kan döljas som en del av bakgrundsbruset i ett musikstycke eller en del av linsdistortionen i ett fotografi – om inte meddelandet syns alls drar det inte uppmärksamhet till sig.

Ett meddelande och dess innehåll går också att göra mindre märkbar genom att dränkas i brus: vi kan automatiskt generera meddelanden om annat som förvirrar ett statistiskt baserat



beräkningsmaskineri. Vi kan generera meddelanden om ingenting alls till skuggidentiteter och skapa ett sammanhang vars enda syfte är att dölja det kommunikationsmönster vi själva verket upprätthåller. Vi kan stycka upp de meddelanden vi i själva verket anser vara viktiga och skicka dem åt olika håll i bitar som sen återsamlas hos mottagaren.

Alla metoderna ovan är i olika grad användbara, beroende på hur sofistikerad kod, chiffer eller maskering vi använder oss av – men framför allt är de besvärliga att hantera för den enskilda användaren. Men uppgifter som är beräkningstunga, rutinmässiga, enahanda och besvärliga för en människa att utföra konsekvent och tillförlitligt är givna kandidater för datorisering. Kryptografi och besläktade funktioner är en självklar tjänst som borde ingå i varje e-postprogramvara och varje Internetleverantörs utbud.

Se till att – precis som ni har för vana att sätta era brev i kuvert – kräva automatiska kryptografiska funktioner i de kommunikationsverktyg ni använder och av de kommunikationstjänstleverantörer ni anlitar!

### ***Digitala identiteter och bilder***

Vilken identitet och vilken bild ger vi av oss själva genom det vi gör tillgängligt? Lika väl som att någon annan analyserar det offentligt eller halvoffentligt tillgängliga materialet vi skickat ut kan vi själva göra det. Vi kan själva använda sökmotorer och databassökningar av olika slag för att kontrollera vad våra digitala spår frammanar för bild. Det finns ingen anledning att låta informationsobalans om oss själva bestå: det finns goda verktyg för oss att använda. Att göra en sådan sökning och sammanställning manuellt tar givetvis en hel del arbete i anspråk. Rent tekniskt är det däremot inte någon större utmaning att samla sådan information automatiskt: beständiga och förutsägbara informationsbehov går att automatisera. Det finns få informationsbehov som är lika förutsägbara som de man har om sig själv!

Men att sedan göra det så att sammanställningen blir begriplig och användbar som grund för att skydda oss som användare är inte trivialt. Vi ska ju inte bara samla utan även kunna bedöma resultatet och sedan agera klokt och klokare än förr på grundval av det.

Anonyma författare brukar inte ha oförvitliga motiv, och pseudonym låter skumt, som om någon har någonting att dölja. Och likväl är de flesta situationer som vi befinner oss i anonyma eller pseudonyma. I butiken vi alltid handlar i, på restaurangen vi ofta besöker, på hotellet vi ibland bor på, i hejaklacken, på konserten – ofta ingår vi i sammanhang där endast en del av vår identitet görs känd för motparten. De sidor av vår personlighet vi utvecklar och prålar med i ett sammanhang undertrycker vi och håller strängt skilda från de vi gör bruk av i andra. Det är inte fusk att ha flera identiteter – det är snarare en naturlig egenskap hos människans anpassliga och socialt bestämda personlighet. Att använda flera identiteter i det digitala umgänget är rimligt och naturligt och har ibland föreslagits som en rättighet för människor att få utnyttja.

Återigen är det helt rimligt att kräva bättre stöd för det i sina kommunikationsmedel. En kommunikationsklient kan varna oss om vi blandar bort korten: Lika väl som en e-postklient kan hantera flera brevlådor ska den kunna hantera flera kommunikationsmönster och markera om vi bryter mot våra tidigare vanor. ”Den där mottagaren har du inte tidigare skrivit till från den här adressen.” ”Det där ämnet brukar du behandla med de där mottagarna och från den där brevlådan.” ”Du brukar skriva på svenska när du använder den identiteten.” är alla varningar som är tekniskt banala att åstadkomma med hygglig om än inte perfekt träffsäkerhet redan med dagens teknik och utan någon vidare betungande beräkningsinsats.

Den stilistiska analys som idag kan användas som skrivstöd eller som grund till författarbestämning kan givetvis köras i andra riktningen också. ”Det du har skrivit nu är väldigt mycket du – vill du få det att verka lite mer som andra?” I kombination med tanken på pseudonyma identiteter kan en sådan analys komma väl till pass. ”Nu har du skrivit mer som

din identitet som motocrossrickard@hoipolloi.org än som din andra identitet som advokat.svensson@juridikkonsult.se”.

### ***Vad måste just jag göra?***

Lita inte på att någon annan gör något åt dig. Det här inte en juridisk fråga, primärt: går saken till rätten är det redan för sent ur integritetssynvinkel. Det är en fråga om kommande, nya konventioner för kommunikation och det är du som användare av digitala medier som formar hur de ska se ut i morgon. Det betyder inte att du behöver sätta dig in i ny teknik, men du måste se till att den du köper informationstjänster av tillhandahåller tjänster som fungerar som du vill att de ska göra.

Så länge den analys som kan göras på våra offentliga data begränsar sig till förekomst- och frekvensanalys av nyckelord och nyckelfraser, och till att läsa kuverteringsinformationen riskerar vi inte att lämna ut oss till särskilt stor insyn. Men mer sofistikerad analys, baserad på analys av textgenre, kommunikationsmönster, individuell variation med finare verktyg än de som sökmotorerna idag använder, ger både oss och våra motparter större möjligheter förstå vilka vi är. Det viktigaste vi som individuella användare måste hålla i minnet är att upprätthålla en ständig medvetenhet om vilken arena vi agerar på. Det är svårt att göra, av anledningar som diskuterats ovan – men det finns goda möjligheter att få tekniskt stöd för att göra det. Sådant stöd kommer inte oss till del om vi inte ber om det. Det bör vi göra!