

# Human IT

Tidskrift för studier av IT  
ur ett humanvetenskapligt perspektiv

## Kunskapsordning och dynamiska nät

[Larserik Lönn](#)

Institutionen för Data och Affärsvetenskap (IDA), Högskolan i Borås

---

### Abstract

*The article deals with the problem of classification of knowledge for efficient retrieval. One suggestion is the use of a dynamic net, based on the user behaviour in retrieving information from databases. There are two dilemmas pinpointed: the dilemma to establish boundaries, and the dilemma with time. Different traditional methods are reported as well as newer methods as hypertext and World Wide Web methods. The article concludes in the need for cooperation in the intersection area of information science, especially the discipline of classification, and computer science, to be able to manage the problem of efficient retrieval of knowledge and information from the huge flood of information of today.*

---

### Innehåll

[Inledning](#)

[Gränsdragningens dilemma](#)

[Dilemmat med tiden](#)

[Kunskapen om klassifikation](#)

[Informationssystem för kunskapsåtervinning](#)

[Motorvägar och upptrampade stigar](#)

[Det dynamiska nätet](#)

## Inledning

Biblioteken har ända sedan antiken försökt att skapa ordning i den kunskap som vetenskap, kultur och konst har producerat, och då kunskap representerat främst som fysiska dokument. I den här artikeln talar jag om dokument i rollen som bärare av kunskap, oavsett deras form och uppträdande.

Begreppet kunskap är här använt där man skulle kunnat använda ordet information, men då syftet med informationen i dokumenten är att förmedla kunskap och då min ansats i artikeln primärt är kunskapskommunikation från människa till människa, där kunskapen kommuniceras med hjälp av information bärare av dokument, så låter jag orden kunskap och information i detta sammanhang bli nära synonymer till varandra.

Begreppet dokument har förändrats i takt med IT-utvecklingen. Allt oftare är nu dokumenten 'icke materiella', flyktiga eller dynamiska. Mångfalden av kunskapsbärare har ökat vilket gör att begreppet dokument inte så lätt låter sig definieras. Mängden dokument ökar också dramatiskt vilket både ger nya möjligheter och skapar problem. Nyttan av IT har emellanåt ifrågasatts. Clifförd Stoll<sup>1</sup> frågar i sin bok "Silicon Snake Oil" varför t ex on-line service lovar så mycket men ännu ger så lite. On-line kataloger är som restaurangmenyer - de är frestande men tillfredsställer inte.

Det finns ett spännande gränsland mellan ordnandet av kunskap och det som nu åstadkoms med IT. Låt oss först se lite på kunskapsordnandet. I grunden är allt ordnande en slags klassificering, en indelning i grupper av företeelser, objekt, tankar, ja egentligen vad som helst. Redan på 1800-talet började bibliotekarier att intressera sig för klassifikation av kunskap. Man definierade då klassifikation som "the act of grouping like things together, and the members share at least one characteristic which members of other classes do not possess. The things classified may be concrete entities, ideas of such entities or abstractions."<sup>2</sup> Klassificeringsträningen börjar redan under de första barnaåren, till exempel med den traditionella leksaken där olikformade klossar ska stoppas i olika hål som passar klossarna. Vi ägnar oss åt klassificering hela livet igenom, med allt från enkla klossobjekt till komplicerade kunskapsstrukturer.

([Åter](#) till början av artikeln)

## Gränsdragningens dilemma

Vid klassifikation sätter man gränser, mellan kunskapsområden, mellan föreställningar om verklighetens beskaffenhet, mellan tro och vetande, ja listan kan göras oändligt lång. Det finns två fundamentala dilemman med klassifikation. Verkligheten låter sig inte lätt indelas i klasser, speciellt inte den abstrakta verkligheten i form av tankemönster, idéer och intellektuella skapelser. Vi får gränsdragningsproblem. Linné råkade på

problem när han träffade på växten 'pleoria' som inte passade in i den klassindelning han hade skapat. Verkligheten hade större varietet än hans klassifikation kunde svara upp mot. I det uppkomna dilemmat måste man välja, och då vill man naturligtvis välja efter bäst möjliga intellektuella eller pragmatiska grund. Där uppkommer problemet - det som av en kan bedömas som en bra indelningsgrund uppfattas på ett helt annat sätt av någon annan. Effekten blir att det blir svårt att via klassifikationen kunna återfinna det som klassificerades 'fel' enligt någons uppfattning, och det är ju ofta själva återfinnandet som är anledningen till klassifikationen.

Dilemmat med gränsdragningsproblemet har man försökt lösa på många olika sätt. De har t ex gått ut på att det klassificerade fått tillhöra flera klasser. Därigenom kan man klara återfinnandet, men har samtidigt skapat problemet med att det går åt ett hjälpmedel, en strukturerad klassförteckning eller ett utbyggt hänvisningssystem, för att kunna lokalisera t ex ett dokument. Olika metoder för klassifikation har utvecklats. Den indiske klassifikationsteoretikern Ranganathans klassifikationssystem<sup>3</sup> gav möjlighet att tilldela klasser ur fem kategorier på ett dokument, så kallad fasetterad klassifikation. Klasserna var strukturerade på ett generaliserat sätt, vilket minskar behovet av förändring över tiden. Ett annat exempel på klassifikation enligt den fasetterade metoden är den som Roget gjorde inom språkområdet, där resultatet blev Roget's Thesaurus<sup>4</sup>. Även de mer konventionella hierarkiska systemen som idémässigt bygger på att man kan klassindela utan gränsdragningsproblem, har fått funktioner för tilldelning av flera klasser. Ett exempel är Dewey Classification<sup>5</sup>.

Dilemmat med gränsdragningsproblemet har även fått en annan kategori av tekniker som försöker lösa problemet. De brukar sammanfattas med begreppet indexering. Indexering kan ses som en form av klassifikation, varför jag väljer att benämna även dessa tekniker som klassifikation i fortsättningen. Klasserna är här mindre strikt organiserade. En eller flera klassbeteckningar, fritt valda eller valda ur ett normerande hjälpmedel (t ex thesaurus) avgränsar klasserna. Det mest fria sättet att klassificera är när varje ord i ett dokument får utgöra en innehållsbeskrivande klassifikation - en enkel form av automatisk klassifikation.

Det finns också mer eller mindre avancerade metoder för statistisk analys av ett dokument innehåll baserat på språket, och där det statistiska utfallet avgör vilken/vilka klasser det ska tillhöra. Den allra enklaste metoden av det här slaget går ut på att räkna alla ord i ett dokument, sortera bort innehållslösa ord (och, eller, i, på mfl efter en så kallad stoppordslista) och sedan låta de kvarvarande, mest frekventa orden, bli de som innehållsbeskriver dokumentet. Ett lite mer avancerat sätt är att matcha det statistiska ordmönstret mot en mängd 'typmönster' för olika klasser av dokument. Det på det sättet funna typmönstret, avgör dokumentets klassifikation. Metoder där man låter ordens 'tyngd' bli beroende av *var* i dokumentet man finner orden finns också, varefter man låter de 'tyngsta' orden bli innehållsbeskrivande. Klusteranalys med ords relativa närhet (avstånd) till varandra kan också vara en statistisk grund för klassifikation.

Ytterligare ett sätt att se på klasser är när klassbeteckningen sätts in i ett semantiskt sammanhang - man använder klassbeteckningarna som komponenter i en formell innehållsbeskrivande meningsbyggnad, man konstruerar med andra ord en mening baserad på ord i dokumentet, som så bra som möjligt beskriver innehållet i dokumentet. Man konstruerar en ny slags titel.

Således finns det en skala från mycket hårt styrda, föreskrivande klassifikationssystem till helt fria sådana baserade på orden i ett dokumentinnehåll. Exempelen ovan gäller klassifikation av kunskap i allmänhet, men inom olika specifika områden finns det även helt andra klassifikationsgrunder. Två exempel är de Geografiska Informationssystemen (GIS) där geografiska läget är en fundamental indelningsgrund och inom biokemin där geners sammansättning och uppbyggnad kan ligga till grund för en viss klassifikation. Vad som sagts ovan ska därför snarare ses som principexempel än som en uttömmande framställning.

([Åter](#) till början av artikeln)

## Dilemmat med tiden

Det andra dilemmat är att även det dokument som man, vid ett tillfälle, lyckats klassificera utan gränsdragningsproblem, sannolikt, vid något senare tillfälle kommer att generera gränsdragningsproblem. Kunskapsmassan förändras över tiden, våra insikter växer, och det som en gång föreföll uppenbart till vilken klass det hörde, kan nu ifrågasättas. Vem kan bygga ett klassifikationssystem som klarar 'Berlinsmurens fall' eller Jugoslaviens sönderfall? Här får vi in ett tidsperspektiv i klassifikationen. Vi kan inte ändra i klassindelningen, radera bort klassen 'Jugoslavien' och ersätta det med något annat, då naturligtvis 'Jugoslavien' i all framtid kommer att vara intressant ur t ex historisk synpunkt. Ett klassifikationssystem speglar därför endast kunskapens uppbyggnad i ett visst tidsögonblick, och förblir därför alltid otidsenligt sett ur detta perspektiv.

([Åter](#) till början av artikeln)

## Kunskapen om klassifikation

Klassifikationsområdet har en lång historia. &Aring;ringtminstone från antikens tid. Vetenskap, konst och kultur har alla bidragit till uppbyggnaden av olika klassifikationssystem, men arbetet har ofta kanaliserats till biblioteksvetenskapen och dess närliggande områden som t ex informationsvetenskap och systemvetenskap. Detta har naturligtvis vuxit fram av nödvändighet, då biblioteken har haft att svara för att kunskap ackumuleras i dokumentsamlingar och att kunskapen kan återvinnas med ett kunskapsbehov som utgångspunkt.

Inom biblioteksvetenskapen finns således en gedigen kunskapsmassa om hur kunskap växer fram och hur den kan återvinnas. Pragmatiska lösningar, filosofiskt och teoretiskt underbyggda, av dilemmat med gränsdragningen och dilemmat med tiden har vuxit fram, men idag faller dessa ofta p g a volymproblemet. Informationsmängden växer för snabbt och förändringstakten är hög. IT-området har bidragit med vissa kvantitativa lösningar (databaser, elektronisk kommunikation mm), men har samtidigt dels bidragit till att så mycket information kan genereras, dels skapat nya typer av problem med t ex 'flyktiga' dokument och 'dynamiska' dokument, dels också tveksamhet om vad som egentligen är ett dokument. Det är nu dags att IT-området, vilket det nu är (definition och avgränsning är inte möjlig att göra idag), börjar att samarbeta med biblioteks- och informationsvetenskap och nyttiggör den kunskap som redan finns, för att skapa

framtidens informationssystem.

Det mått av stumhet och dövhet för varandra som förekommit inom biblioteks- respektive IT-området behöver brytas, och i stället behöver korsbefruktnings ske och dialog komma tillstånd.

([Åter](#) till början av artikeln)

## Informationssystem för kunskapsåtervinning

De informationssystem för kunskapsåtervinning vi har idag karaktäriseras av att de är en förlängning av de manuella systemen men också av vissa blygsamma nya funktioner som inte praktiskt kunde genomföras manuellt. En katalogs möjlighet att via en klassifikationskod, ett namn, en titel i t ex bokstavsordning, återfinna kunskap hittar vi också i de datorbaserade systemen. Att kombinera olika kataloguppslag, var knappast praktiskt möjligt innan IT, men är idag tillämpat i databaser med setteori och Boolesk logik.

En ny återvinningsstrategi som tillämpas med IT-stöd är ett länkningssystem mellan olika dokument - så kallad hypertext. Med hypertext sätter man upp en eller flera fasta kopplingar mellan ett dokument och andra dokument, eller mellan delar av dokument, eller mellan dokumentsamlingar. Länkarna kan representera ett visst kunskapsområde, eller en viss typ av dokument, en viss upphovsman, en tidsrelation, ett ärende eller vad någon bestämmer sig för. Kunskapen i dokumentet kan sedan återvinnas genom att man följer länksystemet. I praktiken bildas ett nät, där användaren letar sig fram i nätmaskorna. Att leta sig fram i ett större nät kan vara mödosamt, och det blir nästan omöjligt att bedöma de mått på återvinningsresultatet, recall och precision, som man i varje fall vid vetenskapligt arbete bör ha en uppfattning om. World Wide Web bygger på nätprincipen. Dilemmat med tiden har i nätprincipen ett embryo till lösning. Länkarna är visserligen statiska, men då länkarna tillkommer under en tidsperiod får vi i dessa ett slags tidsspeglning. Även om inte nya länkar tillkommer om Östberlin, så finns det eventuellt länkar tillbaka till tidigare Berlin-dokument som i sin tur, då när det var mer aktuellt, länkar till Östberlin. Själva kedjan av länkar representerar tiden. Dilemmat med gränsdragningen är dock inte löst med detta. Det är ju någons uppfattning som styr hur maskan i nätet ska vävas.

En viss automatisering av länkskapandet kan göras med statistiska metoder, precis som vid indexering, eller med hjälp av automatiserad innehållsanalys, skapa semantiska länkar. Fördelen med detta sätt är att vi delvis kommer ifrån den subjektiva och personberoende uppfattningen var gränsen ska dras, men å andra sidan har vi skapat oss stora problem att kunna förstå ett kunskapsinnehåll med automatiserade metoder.

För att möjliggöra/effektivisera återvinningen via nät, så har man skapat mekanismer (datorprogram) som automatiskt letar sig fram via nätet, följer maskorna efter vissa givna riktlinjer och letar efter dokumentinnehåll som uppfyller vissa Booleska eller andra krav som specificerats av användaren. De brukar benämnas sökmotorer eller 'crawlers'. Sökmotorer med semantisk ambition kan man möjligen tänka sig som en utveckling. Vi är alltså tillbaka i ungefär samma stadium som tidigare, vi kan göra en kombinationssökning (Boolesk) i en databas (katalog), men med den förändringen att



vi nu kan göra sökningen automatiserat i väldigt många databaser.

Går vi ännu ett steg längre så kan vi nämna Feigenbaum's<sup>6</sup> expert-system-bibliotek som använder kunskapservers till att automatiserat samla och summera relevant information. Elektroniska läroböcker och kunskapsprocessorer assisterar i problemlösning och med komplext tänkande. Kunskapsstrukturer samlas i kunskapslager och kunskapsbaser underhålls av kunskapsingenjörer.

Det grundläggande problemet hur kunskapen identifieras och återvinns är fortfarande inte löst, även om det finns visioner, det vi gjort är kvantitativa förbättringar i återvinningsledet. Dilemmat med gränsdragningen är knappast ens berört. Vi kan söka effektivt, men vi söker inte efter rätt saker.

([Åter](#) till början av artikeln)

## Motorvägar och upptrampade stigar

Finns det då någon lösning på våra dilemman? Någon gång kring mitten av 80-talet lyssnade jag till ett föredrag av Eugene Garfield, Institute of Scientific Information, ISI, som med hjälp av citeringsindex gjorde en statistisk klusteranalys av hur olika vetenskapsmän citerade varandra. Han kunde urskilja ett mönster hur citeringar "klumpade" sig kring vissa personer och grupper, en framväxt av kunskap kring vissa personer eller grupperingar. Garfield använde denna analys för att förutsäga vilka som skulle få Nobelpris.

På ett liknande sätt skulle man kunna göra en kontinuerlig uppföljning av hur människor använder informationen i en databas (katalog). Alltså inte bara tillhandahålla kunskapen (eller ev kunskap om var kunskapen finns att hämta) utan också ta reda på hur man faktiskt hittade dit. Vilka är sökvägarna? Vilka stigar trampas upp till småvägar, och vilka småvägar växer till kunskapsmotorvägar? Redan etablerade stigar, vägar och motorvägar kan sedan presenteras som en karta i kunskapsvärlden för andra att följa eller avvika från. Ett dynamiskt nät av stigar och vägar som förändras med kunskapens utveckling. Förändras därför att människor söker sig till ny kunskap, nöter in nya vägar, annan kunskap förbleknar kanske för att den är överspelad och inaktuell, vägen blir en stig.

Har vi löst våra dilemman med gränsdragningen och tiden med hjälp av ett dynamiskt nät? Maskorna i nätet som vävs kommer att bli ett mått på samhörigheten mellan olika kunskapsdelar. Många kommer att nöta vägen mellan kloning och DNA-teknik men få mellan kloning och Berlinmurens fall. Vi har fått en kunskapsmotorväg mellan kloning och DNA-teknik, men knappast ens en stig mellan DNA och Berlinmuren eller mellan kloning och Berlinmuren.

Samhörigheten representeras alltså av maskorna (länkarna) i nätet. Nättekniken har vi redan i våra hypertextlänkar där länken har framkommit ur en intellektuell aktivitet, men här har vi en kvantitativ länk som är framkommen ur användningen av kunskapen. En stark länk har många funnit relevant, men en svag länk endast ett fåtal. Vilken slags relevans länken står för vet vi dock inte, då den statistiskt har definierats av användarna tillsammans.

## Det dynamiska nätet

Genom att fortlöpande mäta *hur* olika människor når kunskap kan vi upprätthålla ett länksystem (nät) som förändras med tiden och med länkarnas relativa användning. Nätet blir ett stöd att ta sig fram, samtidigt som själva framfarten påverkar nätet, alltså en återkoppling från användaren. Nätet kommer att spegla en aktuell bild av kunskapens samhörighet. Detta är ett dynamiskt nät. Kevin Kelly<sup>7</sup> har uttryckt om det dynamiska nätet "the symbol of science for the next century is the dynamical net".

Svaret på frågan om vi löst något av våra dilemman kvarstår. Gränsdragningsdilemmat minskar, då det inte längre är lika intressant var någon satt gränsen. &Aring;tervinningen av kunskapen (vilket var den huvudsakliga skälet till gränsdragning och klassifikation) kan i vårt fall ske via dynamiska länkar, som baserat på användning ger definition av vad som hör samman respektive ej hör samman (starka och svaga länkar) - alltså en dynamisk automatisk klassifikation med starka och svaga samband.

Dilemmat med att kunskapen och uppfattningen om kunskapen förändras med tiden har inte lösts med det dynamiska nätet. De dynamiska länkarna bakåt i tiden tenderar att försvagas då nya länkar och områden växer till. Länkarnas relativa styrka är ju kopplade till användningen av dem och inom de flesta områden är nyare kunskap mer använd. Beroende på hur man mäter användningen, kan gamla länkars styrka få avklinga långsamt eller snabbt, och nya länkars genomslag få vara långsamt eller snabbt.

## Kunskaps-IT

Höga hästar är inte gångbara för att lösa morgondagens uppgift att kunna återvinna information ur den störtflod som produceras.

*IT-kunskaparen* (den som besitter kunskap om IT och dess möjligheter) måste inse att information/kunskap inte bara ska produceras utan också användas som en del i kunskapens utveckling och därför behöver lagras för senare återvinning. Samarbete måste skapas med den som har kunskap om hur kunskap organiseras och återvinns. Den kunskapen finns hos kunskapsordnaren.

*Kunskapsordnaren* (den som besitter kunskap om organisation och återvinning av kunskap) måste lära sig att omsätta sina kunskaper om just kunskapsordnande så att de av IT-kunskaparen kan formaliseras i effektiva system för automatisk klassifikation, länkskapande och återvinning.

Det dynamiska nät som beskrivits ovan bidrar som bäst till lösningen av något av de två huvudsakliga dilemman som finns inom kunskapsordnande. Det bör dock snarare uppfattas som ett komplement till det som redan finns. Det räcker inte att bara ha vägarna utsatta på en karta, vägarna finns i ett landskap som också måste beskrivas. De metoder som sedan länge används har en given roll i framtiden, men med en

starkare integrering med möjligheterna inom IT-området. All kreativitet och alla kunskapsresurser behövs, både hos IT-kunskapare, kunskapsordnare och andra, för att nyttiggöra all den kunskap som produceras. Vi behöver utveckla ett nytt kunskapsområde, det gemensamma området IT och kunskapsordnande, kunskaps-IT.

[\(Åter till början av artikeln\)](#)

---

## Referenser

1. **Stoll**, Clifford: Silicon Snake Oil. Second thoughts on the information highway. Doubleday N.Y. 1995
2. **Buchanan**, B: Theory of Library Classification, Bingley London 1979
3. **Ranganathan** S.R.: Impact of growth in the universe of subjects on classification in International Federation for Documentation, Committee on Classification Research, FID/CR Report No 12 (1972), FID Publ. Series no 405.
4. **Roget** P. M.: Roget's Thesaurus. First published 1852
5. **Blomberg** M, Weber H: An Introduction to Classification and Number Building in Dewey. Libraries Unlimited Inc. 1976
6. **Feigenbaum** E., McCorduck, Nii: The Rise of the expert Company: How Visionary Companies Are Using Artificial Intelligence to Achieve Higher Productivity and Profits. N.Y. Random House 1988
7. **Kelly**, Kevin: Out of Control. The new biology of machines. Adison Wesley 1994.

[\(Åter till början av artikeln\)](#)

---

## Författaren

Larserik Lönn är universitetslektor i informatik vid Institutionen för Data och Affärsvetenskap (IDA) vid Högskolan i Borås och doktorand i ämnet 'Informatik med Systemvetenskap' vid Stockholms universitet. LL har tidigare varit aktiv inom näringslivet och som lektor vid Bibliotekshögskolan, med inriktning mot biblioteksautomation och system för informationsåtervinning. Hans nuvarande forskning är inriktad mot systemutveckling och simulering av organisationer som generella system.

---

© Larserik Lönn, 1997

Åter till [Human IT 2/1997](#)