

MAGISTERUPPSATS I BIBLIOTEKS- OCH INFORMATIONSVETENSKAP  
VID INSTITUTIONEN BIBLIOTEKS- OCH INFORMATIONSVETENSKAP/BIBLIOTEKSHÖGSKOLAN  
2007: 126  
ISSN 1654-0247

# Forskningsfronten inom informationsvetenskap

## En bibliometrisk analys

PATRIK SPENDRUP THYNELL  
EVA WESTLIN



HÖGSKOLAN I BORÅS  
VETENSKAP FÖR PROFESSION

© **Författarna**

Mångfaldigande och spridande av innehållet i denna uppsats  
– helt eller delvis – är förbjudet utan medgivande.

Svensk titel: Forskningsfronten inom informationsvetenskap – en bibliometrisk analys

Engelsk titel: The research front of information science – a bibliometric study

Författare: Patrik Spendrup Thynell och Eva Westlin

Kollegium: Kollegium 2

Färdigställt: 2007

Handledare: Per Ahlgren

Abstract: The aim of this study is to analyze and describe the research front within information science. To reach this purpose we ask the following research question: Which research themes can be identified within the subject field of information science by using bibliographic coupling in combination with cluster analysis.

The data for our study consists of articles from the journals Journal of the American Society for Information Science and Technology and Journal of Documentation throughout the years 2004-2006. The bibliographic information from the articles was analyzed with the combined methods of bibliographic coupling and cluster analysis. This gave fifteen clusters within which we identified research themes. When analyzing these results we found four main fields of research within the research front of information science: Information retrieval, bibliometrics, knowledge organization and information science as research discipline. Information retrieval proved to be the largest field of research followed by bibliometrics and thereafter the two significantly smaller fields of research knowledge organization and information science as research discipline. When comparing our results with previous research we found that despite different choice of methods and data, there were some similarities.

Nyckelord: Bibliometri, bibliografisk koppling, citeringsanalys, forskningsfront, informationsvetenskap, klusteranalys

|       |  |    |
|-------|--|----|
| 1     | Inledning.....   | 4  |
| 1.1   | Syfte och forskningsfrågor .....                       | 4  |
| 1.2   | Avgränsningar .....                                    | 5  |
| 1.3   | Definitioner .....                                     | 5  |
| 1.4   | Disposition.....                                       | 6  |
| 2     | Teoretisk bakgrund.....                                | 7  |
| 2.1   | Vetenskaplig kommunikation.....                        | 7  |
| 2.1.1 | Citeringar inom vetenskaplig kommunikation.....        | 7  |
| 2.2   | Bibliometri.....                                       | 8  |
| 2.2.1 | Citeringsanalys .....                                  | 9  |
| 2.3   | Multivariat analys .....                               | 10 |
| 2.3.1 | Klusteranalys .....                                    | 10 |
| 2.3.2 | Multidimensionell skalning.....                        | 11 |
| 3     | Bibliometriska arbets- och datamaterial.....           | 13 |
| 3.1   | Citeringsindex .....                                   | 13 |
| 3.2   | Citeringsmått .....                                    | 13 |
| 4     | Tidigare forskning .....                               | 15 |
| 5     | Metod och material.....                                | 21 |
| 5.1   | Insamling av data.....                                 | 21 |
| 5.2   | Behandling av data .....                               | 21 |
| 5.2.1 | Bibexcel och bibliografisk koppling .....              | 21 |
| 5.2.2 | Normalisering av referenslistors längd .....           | 22 |
| 5.2.3 | Klustring.....   | 23 |
| 6     | Resultat.....  | 25 |
| 6.1   | Identifierade forskningsteman .....                    | 25 |
| 7     | Analys och diskussion .....                            | 37 |
| 7.1   | Forskningsteman och huvudinriktningar.....             | 37 |
| 7.1.1 | Informationsåtervinning .....                          | 38 |
| 7.1.2 | Bibliometri .....                                      | 39 |
| 7.1.3 | Kunskapsorganisation .....                             | 40 |
| 7.1.4 | Informationsvetenskapens filosofi .....                | 40 |
| 7.1.5 | Sammanställning av huvudinriktningar .....             | 41 |
| 7.2   | Forskningsteman och tidigare forskning.....            | 41 |
| 7.3   | Reflektioner över urval samt tillämpning av metod..... | 43 |
| 8     | Sammanfattning .....                                   | 46 |
|       | Källförteckning.....                                   | 47 |

# 1 Inledning

*Informationsvetenskap* är ett brett och svårdefinierat ämnesområde. Ofta diskuteras vad som skiljer begreppen information och kunskap åt eller om de ens går att skilja åt och istället är synonyma begrepp. Chaim Zins (2006) menar att informationsvetenskapen består av tre koncept: data, information och kunskap; där data är råmaterial för information medan information är råmaterial till kunskap

Informationsvetenskapen som den ser ut idag har sina anor från tiden efter andra världskriget, men även tidigare har man studerat information. Efter andra världskriget började forskningen dock mer inrikta sig på databehandling och olika kommunikationstekniker och därmed ökade också informationsvetenskapens forskningsobjekt (Mäkinen 2003, s.31). Främst handlar informationsvetenskap om hur information förmedlas mellan producent och användare, alltså hur man söker och återvinner information. Centrala forskningsområden är bland annat informationssökning, informationsåtervinning och informationsadministration (Vakkari 2003, s. 25ff.)

I denna uppsats gör vi en bibliometrisk analys av forskningsfronten inom informationsvetenskap. Bibliometri är något som vi har kommit i kontakt med i någon större utsträckning under utbildningen men metoden används inom många områden, inte sällan inom informationsvetenskap. Bibliometri kan användas för att kartlägga forskningsområden, studera vetenskaplig litteratur, till exempel i samband med forskningspolitiska frågor, men även för att beskriva och utveckla bibliotekens bestånd och inköp.

Den bibliometriska analysmetoden bibliografisk koppling introducerades i mitten av 1900-talet som en metod för att mäta relationer mellan dokument i form av gemensamma referenser. Ju fler gemensamma referenser dokument emellan desto större likhet har dokumenten innehållsmässigt. Om man tänker sig en karta kan man utifrån ett par dokument gjorda under samma tidsperiod i ett specifikt forskningssammanhang se hur dessa dokument bildar grupper utifrån vilka dokument de refererar till. Med hjälp av denna metod kan man beskriva ett forskningsområdes finstruktur, och därigenom se vad kring vilka områden, och i vilken utsträckning, det bedrivs forskning under en viss tidsperiod – en så kallad forskningsfront. (Persson, 1991, s.64ff)

Då vi studerat informationsvetenskap under utbildningen vid Institutionen biblioteks- och informationsvetenskap tyckte vi att det skulle vara intressant att försöka kartlägga en forskningsfront inom informationsvetenskap genom att använda oss av den bibliometriska analysmetoden bibliografisk koppling. För att få fram en aktuell forskningsfront, vad det forskas kring ”just nu”, använder vi oss av nyligen publicerade artiklar i tidskrifter betydelsefulla för ämnet.

Det har gjorts ett flertal kartläggningar av forskningsområdet informationsvetenskap, denna uppsats är ett bidrag till denna forskning.

## 1.1 Syfte och forskningsfrågor

Syftet med denna uppsats är att undersöka och beskriva forskningsfronten inom ämnesområdet informationsvetenskap. För att uppnå syftet tar vi upp följande forskningsfråga:

- Vilka forskningsteman kan identifieras inom ämnesområdet informationsvetenskap genom användning av bibliografisk koppling i kombination med klusteranalys?

## 1.2 Avgränsningar

Vi har valt att använda artiklar ur tidskrifterna *Journal of Documentation* (JDOC) och *Journal of the American Society for Information Science and Technology* (JASIST) för vår undersökning. Båda dessa tidskrifter behandlar ämnesområdet informationsvetenskap<sup>1</sup>, vi anser dock att de tydligt representerar olika områden inom den aktuella forskningen och därför ger de en bred bas för representation av ämnet. JASIST är mer inriktad på praktiska tillämpningar av informationsforskning (utvärderingar av informationssystem, bibliometriska undersökningar, undersökningar kring elektronisk dokumenthantering etcetera). JDOC representerar en "mjukare" forskning kring informationsvetenskap (informationsteori, -filosofi, -historia etcetera). I valet av tidskrifter ville vi också få en spridning av forskningsområdet ur en geografisk aspekt och valde därför en tidskrift med bas i USA (JASIST) och en i europa (JDOC - Storbritannien).

Vid sökning i Journal citations reports på ISI Web of Knowledge är båda tidskrifterna högt rankade i fråga om citeringar bland informationsvetenskapliga tidskrifter<sup>2</sup>. År 2004 får JDOC 1.54 och JASIST 2.09 i impact factor (plats 7 respektive 5), 2005 är impact factor för JDOC 0.98 och JASIST 1.58 (plats 15 respektive 6) och slutligen år 2006 är impact factor för JDOC 1.439 och för JASIST 1.55 (plats 9 respektive 6).

För att få fram forskningsfronten inom ämnesområdet informationsvetenskap har vi valt att undersöka nyare artiklar i JDOC och JASIST. De årtal som vi har valt att avgränsa oss till är 2004-2006. Antalet år (tre) valde vi utifrån premisen att vi skulle få ett urval som dels gav en större mängd data samt att risken att få ett jämnare material över tid (om man väljer färre antal år ökar "risken" att till exempel temanummer och dylikt får stort genomslag i materialet). I de studier som rör kartläggning av forskningsfronten inom informationsvetenskap som vi studerat har man valt tidsperioder från ett till fem år, utifrån detta valde vi "mellanvärdet" på tre år.

Vi har valt att använda tidskriftsartiklar som dokument för vår undersökning och utesluter därmed brev, bokrecensioner och annat publicerat material som kan återvinnas i citeringsdatabaser.

Vårt val av citeringsindex för uppsatsen är Social Science Citation Index (SSCI) som ingår i ISI Web of Knowledge.

## 1.3 Definitioner

Här tar vi upp aktuella termer och dess betydelse inom uppsatsens ramar.

Forskningsfront: Med forskningsfront menas vad det forskas kring och i vilken utsträckning det görs så utifrån en vald tidsperiod.

---

<sup>1</sup> Se kapitelavsnitt 6.1

<sup>2</sup> Se kapitelavsnitt 4.2

Forskningsområde: i uppsatsen används även termerna ämnesområde samt forskningsdisciplin.

Informationsvetenskap: Med informationsvetenskap avser vi ett brett vetenskapligt område inom vilket man behandlar studier av information. Inom detta område rymms forskning kring informationsbehandling, informationsteori, kommunikationsteori, systemteori, biblioteksvetenskap, dokumentation m.fl.

#### **1.4 Disposition**

I kapitel 2 presenterar vi en teoretisk bakgrund till de metoder som vi använder i uppsatsen. I kapitel 3 tar vi upp ytterligare bakgrund i form av vetenskaplig kommunikation och citeringar. I kapitel 4 redogör vi för bibliometriska datamaterial. I kapitel 5 tar vi upp tidigare forskning. Metod och material presenteras i kapitel 6 där vi redogör för hur vi samlat in och bearbetat den data som undersökningen bygger på. I kapitel 7 redovisar vi undersökningens resultat i form av presentation och kommentarer till de forskningsteman som vi har fått fram. Efter det följer kapitel 8 med analys och diskussion, dels av resultat och dels i förhållande till tidigare forskning. Kapitlet avslutas med kommentarer och förslag på fortsatt forskning. Uppsatsen avslutas med en sammanfattning i kapitel 9.

## 2 Teoretisk bakgrund

Här redovisar vi den teoretiska bakgrunden för vår uppsats: vetenskaplig kommunikation i förhållande till bibliometri, bibliometri och några av de metoder man använder sig av inom bibliometrisk forskning samt multivariat analys och metoder inom dess användningsområde. Främst beskriver vi de metoder vi kommer att använda oss av i vår egen undersökning, bibliografisk koppling och klusteranalys.

### 2.1 Vetenskaplig kommunikation

Vetenskaplig kommunikation kan delas in i formell och informell kommunikation. Formell kommunikation avser officiell informationsöverföring i huvudsak genom publikationer. Informell kommunikation är informationsöverföring mellan personer på exempelvis seminarier och konferenser där en inofficiell diskussion kan föras (Kärki & Kortelainen 1998, s. 5). Betydelsen av den informella kommunikationen i vetenskapliga sammanhang har inte studerats i lika stor utsträckning som den formella kommunikationen där man använt kvantitativa metoder, exempelvis bibliometri, för att studera denna (Kärki 2003, s. 205).

Inom formell kommunikation spelar publikationer en central roll. En tidskriftsartikel med tillhörande referenslista där forskaren presenterar sina undersökningar och resultat är ett universellt accepterat medium för vetenskaplig kommunikation. En av orsakerna till att publicering i vetenskapliga tidskrifter är attraktivt beror på systemet med peer review. (Cronin 1984, s. 11f). Peer review innebär att ett antal granskare, exempelvis sakkunniga forskare, granskar och väljer ut de artiklar som ska publiceras i en tidskrift (Lunds universitets bibliotek 2007).

I vår uppsats studerar vi de referenser till artiklar som publicerats i de tidskrifter som vi valt att undersöka, alltså består vårt material av en formell kommunikation mellan olika forskare. Den kommunikationsaspekt man studerar med bibliografisk koppling är det medvetna syftet en forskare har med att skapa en relation utifrån sin aktuella text till de texter man där refererat till.

#### 2.1.1 Citeringar inom vetenskaplig kommunikation

Det finns olika skäl till att en författare citerar andra författares verk. Citeringsprocessen är subjektiv och det kan vara svårt att veta motivet till varför en författare citerar. Det finns alltid en lucka mellan varför en författare citerar och varför vi tror att författaren citerar. Orsakerna till att en författare citerar kan vara seriösa eller manipulativa (Cronin 1984, s. 30). Weinstock (enligt Egghe & Rousseau, 1990, s. 211) identifierade 15 specifika citeringsfunktioner som anses vara seriösa orsaker till citering:

1. Visa sin vördnad för pionjärer
2. Ge erkännande för relaterade arbeten
3. Identifiera metodik, utrustning etc.
4. Tillgodose bakgrundsläsning
5. Korrigera sitt eget arbete
6. Korrigera arbeten gjorda av andra författare
7. Kritisera arbeten gjorda av andra författare
8. Bekräfta resultat
9. Uppmärksamma forskare på kommande resultat

10. Synliggöra dåligt spridda, dåligt indexerade eller icke-citerade arbeten
11. Bekräfta data och klasser av fysikaliska konstanter, etc.
12. Identifiera ursprungspublikationer där idéer eller koncept diskuterats
13. Identifiera ursprungspublikationer där eponymer<sup>3</sup> eller termer beskrivs, exempelvis Hodgkin's sjukdom
14. Förkasta andra författares arbeten eller idéer
15. Bestrida andra författares prioritetsanspråk

Thorne (enligt Egghe & Rousseau, 1990, s. 212) har gjort en sammanställning på oseriösa eller manipulativa citeringsorsaker som är raka motsatsen till Weinstocks lista. Thorne nämner, som exempel på oseriösa citeringar, bland annat uppdelning av en forskningsrapport i flera olika delar, citeringar med uppenbar giltighet som kan spegla vilken typ av synsätt som helst och självciteringar.

De olika skäl som en författare har för att citera till ett visst verk är som sagt svårt att utröna och det är något som man kan ha i betänkande när man använder citeringsanalys för att exempelvis kartlägga ett forskningsområde eller hur ett visst verk har citerats. Vid bibliografisk koppling används referensförteckningar som undersökningsmaterial och en viss risk finns att de referenser som ingår i referensförteckningen har förts in av olika orsaker vilket gör att alla referenser inte är likvärdiga (Kärki 2003, s. 202). Därmed kan det finnas en viss felmarginal eller problem med exempelvis referensmaterialet vid en studie med metoden bibliografisk koppling. Man kan tänka sig att självciteringar kan få ett stort genomslag i materialet då en författare citerar sig själv och dessa referenser jämförs med andra författare som refererat till samma författare.

## 2.2 Bibliometri

”Bibliometri kan enklast beskrivas som läran om hur man med hjälp av kvantitativa metoder kan beskriva samlingar av dokument” (Persson, 1991, s. 6). Inom bibliometrisk forskning arbetar man efter antagandet att forskningsaktiviteten avspeglas i litteraturen och med hjälp av kvantitativa analyser av densamma vill man beskriva vetenskapens och teknologins utveckling (Kärki & Kortelainen, 1998, s. 1). Bibliometrisk forskning är interdisciplinär och dess metoder innehåller element från matematik, samhällsvetenskap, naturvetenskap och teknik (Glänzel, 2003, s. 5). De variabler man främst arbetar med inom bibliometrisk forskning är publikationer, författare, referenser och citeringar (Kärki & Kortelainen, 1998, s. 6).

Termen *bibliometri* introducerades av Pritchard 1969 och med bibliometri menade han en tillämpning av matematiska och statistiska metoder på böcker och andra kommunikationsmedier (Glänzel 2003, s. 6). Samtidigt började termen *scientometri* användas i Östeuropa och termen är en översättning av ryskans *naukometrija* som användes av Nalimov och Mulscenko, senare grundades tidskriften *Scientometrics* i Ungern 1978 (Kärki och Kortelainen, 1998, s. 4). Med scientometri menades från början kvantitativa studier för att analysera vetenskaplig kommunikation, men enligt Glänzel har gränsen mellan termerna bibliometri och scientometri suddats ut och de används nästan som synonymer idag (Glänzel 2003, s. 6). Enligt Egghe och Rousseau är termen *bibliometri* begränsad till analys av bibliotek och bibliografier, och scientometri används för tillämpning av vetenskaplig policy. Istället använder de termen *informetri*, som förespråkats av Brooks, som enligt dem tar hänsyn till modern teknologi som gett

---

<sup>3</sup> Benämning på exempelvis en sjukdom som innehåller ett personnamn (Nationalencyklopedin 2007)

nya former av kunskapsframställning och kommunikation. Informetri behandlar mätning av alla aspekter på information med hjälp av matematiska teorier och metoder. I vår uppsats har vi valt att använda termen bibliometri eftersom vi tycker att den förekommer mest i det material vi har använt oss av.

Bibliometri har idag åtminstone fyra användningsområden. Dels kan man utvärdera hur vetenskapliga forskningsenheter, exempelvis universitet eller institutioner, presterar i förhållande till varandra (Noyons 1999, s. 3ff). Glänzel kallar detta användningsområde för vetenskapspolicy och menar att det är det mest betydelsefulla området inom bibliometri idag (Glänzel 2003, s 10). Ett annat användningsområde för bibliometri är kartläggning av vetenskap. Genom att analysera forskningens produktion i form av publikationer med hjälp av olika klustrings- och skaleringsmetoder kan strukturen och förändringar av ett vetenskapsområde fås fram. Ett tredje användningsområde för bibliometri är informationsåtervinning. Bibliometri kan här visa på olika ämnes inbördes relationer. Sist kan bibliometri användas i bibliotek för beståndsutveckling av tidskrifter, exempelvis för att se vilka tidskrifter som citeras mest inom olika ämnesområden (Noyons 1999, s. 4f).

### **2.2.1 Citeringsanalys**

Den vetenskapliga traditionen kräver att forskare som skriver vetenskapliga artiklar refererar till tidigare artiklar som relaterar till den aktuella artikelns ämne. Citeringsanalys är det område inom bibliometri där relationen mellan citerade och citerande artiklar studeras (Egghe, 1990, s. 203f). Vilken relation artiklarna emellan man väljer att lägga fokus på kan variera, det kan gälla allt från dokumenten i sig till deras författare, tidskrifterna de förekommer i, den institution eller det land de produceras i etc. (Diodato, 1994, s. 33).

#### ***Bibliografisk koppling***

Kessler introducerade bibliografisk koppling genom ett antal artiklar i slutet av -50 och början av 1960-talet (Kessler 1963). Ursprungligen beskrevs det som en metod för att gruppera artiklar med tekniskt eller vetenskapligt innehåll för att underlätta anskaffandet av vetenskaplig information och dokumentåtervinning (Jarneving, 2005a, s. 247). Grupperandet av artiklarna görs på basis av antal gemensamma referenser. Om två eller flera publikationer refererar till samma källor bör, enligt tanken med bibliografisk koppling, även dess innehåll vara gemensamt; ju fler gemensamma referenser desto större likheter innehållsmässigt (Persson, s. 64f). En fördel med att använda sig av bibliografisk koppling vid kartläggning av ett forskningsområde är det att man kan utföra en studie på ett dokument så snart det är publicerat. Användandet av nyligen publicerat material i tidskrifter betydelsefulla för ämnet är en värdefull möjlighet om man vill studera för tillfället aktuella skeenden, så kallade ”snapshots”, eller vad vi i denna uppsats kallar forskningsfronten, inom ett forskningsområde (Glänzel, 2003, s. 84f). Hicks uttryckte 1987 fördelarna med att använda sig av bibliografisk koppling gentemot cociteringsanalys på följande vis (enligt Glänzel, 2003, s. 86):

- ...just published papers that are closely related by bibliographic coupling links can provide snapshots of early stages of a specialty's evolution. By contrast, it may take time before a 'critical mass' of papers on a new research topic is created that is needed to produce the highly cited publications on which the cocitation mapping is based.

Glänzel gör följande tillägg till ovanstående kommentar:

- From the viewpoint of information retrieval, co-citation clustering results in the restriction to frequently cited papers, whereas bibliographic coupling extends to practically all publications.

### ***Cociteringsanalys***

Citeringsanalysmetoden cocitering introducerades av Small (1973). Small ville visa att genom att studera frekvensen av samciterade dokument och utifrån detta bilda grupper/kluster kan man få fram forskningens intellektuella bas. Den intellektuella basen är, till skillnad från den mer tidslåsta forskningsfront man får fram genom bibliografisk koppling, rörlig. Rörligheten beror på att användandet av de citerade dokumenten kan minska eller öka över tiden, emedan de citerade dokumenten man studerar i bibliografisk koppling blir statistiska så fort de publiceras (Jarneving, 2005a, s. 248). När man gör en cociteringsanalys arbetar man stegvis. Första steget är att göra en rangordnad frekvenslista över citerade dokument. Sedan går man vidare genom att räkna fram cociteringarna mellan varje par. För att kontrollera för variation ifråga om enskilda citeringsfrekvenser konstrueras en normaliserad cociteringsfrekvens genom att cociteringsfrekvensen divideras med roten ur produkten av de båda citerade dokumentens enskilda citeringsfrekvenser (Persson 1991, s. 58f). En variant av cocitering är författarcocitering (author cocitation analysis, ACA) där man studerar cociteringar mellan författare istället för dokument (Diodato 1994, s. 43).

## **2.3 Multivariat analys**

Statistiska analyser benämns och behandlas olika beroende på hur många variabler som studeras. Utgår man från en variabel talar man om *univariat* analys, när två variabler studeras samtidigt benämns det som *bivariat* analys. När man, som i vårt fall, simultant studerar tre eller flera variabler kallas det *multivariat* analys. Det finns ett flertal analysmetoder att utgå från vid multivariat analys. Vanligt förekommande är *multipl regressionanalys* – då variationen i en beroende variabel, så kallad responsvariabel analyseras med hjälp av flera förklarande variabler. En annan vanligt förekommande multivariat analys är *faktoranalys* – här undersöker man möjligheterna att komprimera den information man arbetar med; vilka faktorer mäter vi egentligen med vårt datamaterial och är det möjligt att minska antal variabler? (Körner & Wahlgren 2005, kap. 9). Vi kommer i vår undersökning att använda oss av den multivariata analysmetoden *klusteranalys*.

### **2.3.1 Klusteranalys**

Klusteranalys är en av de mest populära teknikerna för multivariat analys (Egghe, 1990, s. 112). Syftet vid klusteranalys är att samla objekt som är så lika varandra som möjligt i naturliga grupper, så kallade kluster. Vid användning av klusteranalys vill man alltså upptäcka grupper av objekt i data man använder sig utav – inte föra in objekt i redan kända klassificerade grupper. Vad själva likheten mellan objekten skall representera väljer man själv utifrån sin undersökning; det kan gälla tid, geografiskt avstånd, kostnader etc. I vårt fall mäter vi likhet i form utav gemensamma bibliografiska referenser. Eftersom man har flera variabler går det att räkna ut likhet till exempel med hjälp av matematiska formler. Ett annat alternativ är att normera mätvärdena genom att

alla variabler standardiseras eller genom att transformera variablerna så att värdena hamnar mellan 0 och 1 (Körner & Wahlgren 2005, s. 175ff).

### ***Hierarkisk klusteranalys***

”Hierarkisk klusteranalys innebär att observationerna stegvis förs samman till allt större kluster till dess att samtliga observationer bildar ett enda kluster.” (Körner & Wahlgren 2005, s. 178). Hierarkisk klusteranalys är oåterkallelig på så vis att när en sammanslagning utav två objekt till ett kluster väl har skett kan man inte på nytt dela upp objekten var för sig igen. När man använder sig av hierarkisk klusteranalys är det upp till den som gör studien att själv att avgöra när man har det optimala antalet kluster innan det slutligen blir ett enda stort sådant, i normalfallet är det ju resultaten någonstans dessförinnan man är ute efter (Everitt, Landau & Leese 2001, s. 55). Det finns även en del utarbetade metoder för att avgöra när man nått det optimala antalet kluster, bland annat kan man utgå från en visuell bild av klustringsprocessens olika steg; ett två-dimensionellt diagram kallat *dendrogram*. Men, då någon generell metod för detta inte existerar så är det ofta, såsom Baxter (1994) påpekar ”informella och subjektiva kriterier, baserat på områdesexpertis, som troligast är den vanligaste avgörande faktorn” (Everitt et al. 2001, s. 76f).

Det finns ett antal olika metoder för hierarkisk klusteranalys, som exempel kan nämnas *Complete link clustering*, *Group average link clustering*, *Centroid clusterin* samt *Single link clustering* (SLINK), även kallad *nearest-neighbour technique* eller *närmsta-grannemetoden*. Den enklaste metoden för hierarkisk klusteranalys är sistnämnda SLINK (Aldenderfer & Blashfield 1984, s. 35). Vid klustring med SLINK förs klustren samman när två objekt inom två olika kluster befins vara lika (Romesburg 2004, s. 317). Objekten ställs upp i en matris och klustringsprocessen inleder med att söka upp de objekt som är mest lika. Klustringsprocessen fortskrider därefter att länka samman objekt i kluster och därpå kluster med andra kluster efter regeln att nya ”klustermedlemskap” beviljas för ett kluster eller enskilt objekt på basis av högsta ”likhetsnivå” med vilket objekt – ”medlem” som helst i ett redan existerande kluster. Det krävs alltså endast en enda länk (single link) inom två existerande kluster för att en sammanslagning skall ske (Aldenderfer & Blashfield 1984, s. 36).

### **2.3.2 Multidimensionell skalning**

Multidimensionell skalning (MDS) är ett antal tekniker som använder sig av närhetsvärden av indata för att visualisera dessa. Närhetsvärde är ett tal som visar på hur lika eller olika två objekt är. Med MDS gestaltas objekten i två- eller flerdimensionella kartor som visar på strukturen mellan de data som använts. Kartorna består av punkter som representerar objekten och här kan man se deras inbördes närhet. Två objekt som är lika varandra hamnar nära varandra i kartan, medan två objekt som är olika varandra får ett större avstånd (Kruskal & Wish 1978, s. 5).

De matematiska uträkningarna för MDS är komplexa och kräver hjälp av datorprogram. Matematiskt är det möjligt att utföra MDS i olika dimensioner. Vanligtvis använder man två- eller tredimensionella kartor för att visualisera data men det kan även utföras i fler dimensioner. Om värdet för dimension är  $R$  så kan dimensionerna vara:  $R=1, 2, 3, 4$  osv. Problemet med större dimensioner är att ju högre värde på  $R$  som används desto mer otillräckliga blir de olika grafiska gestaltningarna (Kruskal & Wish 1978, s. 15). En tvådimensionell karta visar punkterna i ett koordinatsystem och den kan enkelt återges

på papper. Den tredimensionella kartan däremot är inte lika enkel att återge utan visas bäst på en datorskärm eftersom den tredje dimensionen gestaltas bättre där jämfört med en utskrift (Kärki & Kortelainen 1998, sid. 35ff).

Data som ska visualiseras genom MDS är en samling av objekt, som exempelvis kan vara länder, politiska kandidater och färger mm. Objekten indexeras först och främst med bokstaven  $i$  och i andra hand med  $j$ . Härmed ges att  $i$  och  $j$  går från 1 till  $n$  om det finns  $n$  objekt. Detta ger att närhetsvärdet för objekt  $i$  med objekt  $j$  kan formuleras som  $d_{ij}$ . Därefter ordnar man  $d_{ij}$  i en matris som sedan kan användas som indata för MDS. Varje objekt är representerat av en punkt och avståndet mellan punkterna fås av  $d_{ij}$  (Kruskal & Wish 1978, s. 15ff).

## 3 Bibliometriska arbets- och datamaterial

I detta kapitel presenterar vi bibliometriska arbets- och datamaterial och då främst de vi använder oss av i vår undersökning.

### 3.1 Citeringsindex

Då man vill göra en bibliometrisk undersökning baserat på citeringar av större omfattning använder man sig med fördel utav en citeringsdatabas, ett så kallat citeringsindex. Enligt Hjerppe är citeringsindex den totalt dominerande källan för utvinning av data till bibliometriska undersökningar (1978, s. 10). Ett citeringsindex är en förteckning över var tidskriftsartiklar, rapporter, böcker etc. har citerats. Möjliga sökningar är (Lunds universitets bibliotek 2007):

- Vem/vilka har citerat ett visst arbete?
- Vilka forskare eller vilka dokument samciteras ofta?
- Vilka arbeten citerar gemensamma referenser?
- Vilka tidskrifter citeras mest?
- Vilka institutioner citeras mest?
- Vilka ämnesområden citeras mest?
- Vilka individuella forskare citeras mest?

Exempel på citeringsindex är: Arts & Humanities Citation Index, Science Citation Index och Social Sciences Index vilka alla ingår i plattformen ISI Web of knowledge.

### 3.2 Citeringsmått

Om man vill göra en bibliometrisk studie över ett forskningsområde utifrån dess vetenskapliga tidskrifter och inte har tillgång till obegränsade resurser kan det vara lämpligt att göra någon form av urval. Ett bra mätinstrument på för ämnet representativa tidskrifter är att se på hur en tidskrifts artiklar används genom att på olika sätt studera hur dessa citeras i andra vetenskapliga tidskrifter inom ämnesområdet. För att mäta och jämföra hur mycket en tidskrift används som citeringskälla finns det olika metoder att tillgå. Journal Citation Reports (JCR) är en databas tillgänglig via ISI Web of Knowledge där man presenterar citeringsstatistik över 7500 vetenskapliga tidskrifter som täcker en rad varierande ämnesområden (ISI Web of Knowledge 2007b).

De tre mått man på JCR använder sig av för att visa på tidskrifters citeringsfrekvens är *Immediacy Index*, *Cited Half-life* och *Journal Impact Factor*.

I "Glossary of Thomson Scientific<sup>4</sup> terminology" (2007b) kan man läsa att *Immediacy Index* är ett mått på hur snabbt en specifik tidskrift citeras i övrig publicerad litteratur, på så sätt kan man se hur pass "på framkant" en tidskrift håller sig inom sitt forskningsområde. Man får fram en tidskrifts *Immediacy Index* genom att dividera det totala antalet mottagna citeringar under ett år med det totala antalet publicerade artiklar under samma år. Ett exempel från tabell 1: Tidskriften JDOC innehöll 33 artiklar 2006. Dessa artiklar citerades under samma år 11 gånger, dvs. varje artikel citerades i genomsnitt 0,333 gånger (=Immediacy Index)

---

<sup>4</sup> Thomson Scientific tillhör The Thomson corporation och tillhandahåller bl.a ISI Web of knowledge, Journal Citation Reports och ett flertal andra informationsökningstjänster.

Cited Half-life använder man, enligt Thomson Scientific's ordlista (2007b), när man vill ha ett mått på en tidskrifts långsiktiga värde som källa. Cited Half-life är det antal år om man går tillbaka från det innevarande året som står för 50 % av citeringarna, som mottagits av den citerade tidskriften under det innevarande året. Exempel från tabell 1 på uträkning av en tidskrifts Cited Half-lifemått: Tidskriften JDOC får 2006 värdet 8,8, dvs. hälften av alla citeringar under år 2006 av tidskriften Cell gäller artiklar, som publicerats i denna tidskrift under de sista 8,8 åren. Övriga hälften av citeringarna gäller då artiklar i JDOC, som är äldre än 8,8 år.

Impact factor är det mått man får för ett givet år genom att dividera antalet citeringar som de artiklar, vilka publicerats i tidskriften under de två föregående åren, mottagit året ifråga med antalet artiklar som publicerats i tidskriften de två föregående åren (Thomson Scientific 2007b). Exempel: I tabell 1 har tidskriften JDOC för år 2006 impact factor 1,439, dvs. en artikel, som publicerades i årgångarna 2004 och 2005 av JDOC, citerades under år 2006 i genomsnitt drygt 1 gång

| Rank | Abbreviated Journal Title | ISSN      | Total Cites | Impact Factor | Immediacy Index | Articles | Cited Half-life |
|------|---------------------------|-----------|-------------|---------------|-----------------|----------|-----------------|
| 1    | MIS QUART                 | 0276-7783 | 3186        | 4.731         | 0.610           | 41       | 9.5             |
| 2    | J AM MED INFORM ASSN      | 1067-5027 | 2143        | 3.979         | 0.587           | 80       | 4.7             |
| 3    | INFORM SYST RES           | 1047-7047 | 1508        | 2.537         | 0.304           | 23       | 6.6             |
| 4    | INFORM MANAGE-AMSTER      | 0378-7206 | 1466        | 2.119         | 0.215           | 79       | 5.4             |
| 5    | J MANAGE INFORM SYST      | 0742-1222 | 1523        | 1.818         | 0.195           | 41       | 6.9             |
| 6    | J AM SOC INF SCI TEC      | 1532-2882 | 2761        | 1.555         | 0.366           | 142      | 7.3             |
| 7    | INFORM PROCESS MANAG      | 0306-4573 | 1347        | 1.546         | 0.475           | 101      | 7.5             |
| 8    | INFORM SYST J             | 1350-1917 | 290         | 1.543         | 0.500           | 14       | 5.6             |
| 9    | J DOC                     | 0022-0418 | 700         | 1.439         | 0.333           | 33       | 8.8             |
| 10   | J HEALTH COMMUN           | 1081-0730 | 550         | 1.387         | 0.161           | 62       | 4.4             |

**Tabell 1** Tabell hämtad från JCR innehållande de, efter Impact factor, tio högst rankade tidskrifterna inom Biblioteks- och informationsvetenskap år 2006. (ISI Web of Knowledge, 2007a)

## 4 Tidigare forskning

De studier som vi har valt ut att jämföra vårt resultat med i den kommande analysen är alla olika bibliometriska undersökningar gjorda i syfte att kartlägga informationsvetenskap utifrån varierande faktorer och metoder. Dessa kriterier, bibliometriska undersökningar av informationsvetenskap är den inledande anledningen till att vi specifikt valt just de undersökningar som ges en närmre presentation i detta kapitel.

Ytterst ville vi helst finna studier som använt sig av mer exakt samma metoder som oss vid kartläggning av informationsvetenskap för att så tydligt som möjligt kunna jämföra resultat. Det visade sig dock att vi endast kunde finna en tidigare undersökning som med de kombinerade metoderna bibliografisk koppling och klusteranalys kartlagt forskningsområdet informationsvetenskap i syfte att få fram en forskningsfront: *The Intellectual Base and Research Fronts of JASIS 1986-1990*. I denna studie redogör Olle Persson (1994) för sin undersökning av artiklar publicerade i JASIS under åren 1986-1990. I undersökningen gjordes dels en cociteringsanalys för att få fram den intellektuella basen och dels en analys av forskningsfronten med hjälp av bibliografisk koppling.

Utifrån cociteringsanalysen av materialet från JASIS skapade Persson en MDS-karta för att kunna visa på rådande förhållanden sinsemellan de författare som genererats av analysen. Med utgångspunkt i MDS-analysen kunde Persson urskilja två huvudämnesområden inom informationsvetenskap: en grupp med bibliometri och en med IR, där IR-gruppen är ungefär dubbelt så stor som bibliometrigruppen. IR-gruppen delades in i två undergrupper: en som Persson kallar hård som behandlar tekniska aspekter på IR som algoritmer och automatisk indexering samt en undergrupp som benämns mjuk som tar upp exempelvis utvärdering av IR-system, användarvänlighet och teoretiska aspekter på IR. Bibliometrigruppen kan delas in två undergrupper där en del behandlar citeringsanalys och en del behandlar bibliometriska fördelningar. Cociteringsanalysen av artiklarna som är gjord speglar enligt Persson den intellektuella basen av informationsvetenskap som forskningsområde. Persson hänvisar sitt resultat till White och Griffith 1981 (se nedan) och menar att namnen och positionerna av författare i respektive undersökning, trots olika metoder, är ganska likartade. De skillnader som finns beror enligt Persson på att han endast använt en tidskrift som underlag samt att tidsaspekten visar på att det kan ha skett en förändring där de klassiska studierna av vetenskaplig kommunikation har mindre inflytande än vid tidpunkten för White och Griffiths undersökning.

Forskningsfronten definieras enligt Persson av artiklar som citerar samma litteratur, artiklarna bör behandla samma frågor och problem som de citerande artiklarna. För att få fram en sådan forskningsfront över forskningsområdet informationsvetenskap gjordes utifrån materialet från JASIS en analys med en kombination av metoderna bibliografisk koppling och SLINK. De kluster som utvanns i analysen kartlades utifrån antal gemensamma citerade författare som indata och ord från titlarna på artiklarna användes för att benämna klustret på kartan. På kartan fick Persson fram en struktur som liknar den med cociterade författare, en grupp med artiklar om bibliometri och en grupp om IR som går att dela in i en hård och en mjuk grupp. Därmed menar han att forskningsfronten speglar den intellektuella basen utifrån sin undersökning. Persson

menar också att det går att kartlägga strukturer på ett ämnesområde som är giltiga trots att det sker på en mikronivå med endast en tidskrift som underlag.

Bibliografisk koppling har inte använts lika mycket som metod jämfört med cocitering när det gäller utvärdering av forskning (Glänzel 2003, s. 86). Vid sökningar på respektive metod i databaser som exempelvis SSCI, där olika ämnesområden är representerade i urvalet, så visar resultatet att det finns fler artiklar om cocitering än om bibliografisk koppling. De undersökningar som har gjorts på informationsvetenskap har använt sig av olika metoder, men trots att de har använt andra metoder än bibliografisk koppling så tar vi upp dem här för att kunna jämföra med vår kartläggning av forskningsfronten inom ämnesområdet informationsvetenskap.

Två betydande, som ofta refereras i övriga studier, bibliometriska undersökningar av forskningsområdet informationsvetenskap baserade på författarcociteringsanalys är: *Author Cocitation: A Literature Measure of Intellectual Structure* av Howard D. White och Belver C. Griffith (1981) samt *Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972–1995* av Howard D. White och Katherine W. McCain (1998). I den förra är syftet med undersökningen att visa på metoder för att kartlägga intellektuella strukturer i vetenskap och andra områden och i den senare vill man visa på förändringar som har skett inom biblioteks- och informationsvetenskap under en tidsperiod från 1972 till 1995.

I båda dessa undersökningar framhåller man fördelarna med ACA gentemot cociteringsanalys och dess styrka när det gäller att få fram intellektuella strukturer inom ett forskningsområde. Att denna åsikt framförs i båda undersökningarna är väl inte heller särskilt anmärkningsvärt då biblioteks- och informationsforskaren Howard D. White är medförfattare till bägge. Fördelar med ACA gentemot cociteringsanalys som tas upp i *Author Cocitation: A Literature Measure of Intellectual Structure* är att det behövs enbart författarnamnen för att analysen ska kunna genomföras, vilket gör att sökning av cociteringsdata går snabbt att genomföra och därmed blir mindre kostsamt (denna kostnads- och tidsaspekt bör man se i det sammanhang undersökningen gjordes, det fanns inte 1981 samma förutsättningar att bearbeta data som det finns idag – uppsatsförfattarnas anmärkning). Vid cocitering av dokument så behövs enligt White och Griffith en större kunskap om området som ska undersökas jämfört med ACA. Det blir svårare att kategorisera ett område utifrån dokument jämfört med författare eftersom författare täcker in ett bredare område.

I undersökningen som presenteras i *Author Cocitation: A Literature Measure of Intellectual Structure* gjordes en ACA på 39 olika författare som enligt White och Griffith väl representerade ämnesområdet. Cociteringsdatan fördes in i en MDS-karta. De fyra stora författargrupper som urskiljdes på MDS-kartan var följande (i tur och ordning från vänster till höger i kartan): vetenskaplig kommunikation, bibliometri, generalister samt IR. En mindre grupp, benämnd föregångare, med två författare placerades utanför de stora författargrupperna och det beror på att de har mindre antal cociteringar med de andra författarna. Placeringen av författargrupperna visar på att bibliometrigruppen är den mest centrala och det förklaras med att författarna i denna cociteras både med författare i grupperna vetenskaplig kommunikation och IR. Generalisterna har starka band till både bibliometri- och IR-gruppen, men inte med vetenskaplig kommunikation.

En faktoranalys<sup>5</sup> av författarna utfördes också och det resulterade i sju olika faktorer: automatisk återvinning och dokumentanalys (1), nätverk av dokument och citeringsstudier (2), bibliometri (3), vetenskaplig kommunikation (4), utvärdering av återvinning och indexering (5), äldre teoretiska modeller (6) samt utvärdering av informationssystem (7). Samtliga 39 författare placerades i åtminstone en faktor, 14 författare placerades i två faktorer och en författare placerades i tre faktorer. Genom en grafisk framställning av faktorerna kunde man se vilka faktorer som hade samband med varandra. Faktor 1, 5 och 7 bildade en grupp, faktor 3 och 6 en grupp samt 2 och 4 den tredje gruppen. Faktor 3, bibliometri, hade i olika grad samband med alla andra faktorer och det överensstämmer med gruppen bibliometris centrala placering på MDS-kartan.

I den omfattande undersökningen som redogörs i artikeln *Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972–1995* görs en ACA av de 120 mest citerade författarna från tolv olika betydande tidskrifter inom informationsvetenskap under tre åttaårsperioder. En faktoranalys av de 120 författarna gjordes på hela tidsperioden för att se hur strukturen för ämnesområdet såg ut. Faktoranalysen gav tolv faktorer som representerar informationsvetenskap: experimentell återvinning, citeringsanalys, onlineåtervinning, bibliometri, generella bibliotekssystem, vetenskaplig kommunikation, användarteorier, OPAC (Online Public Access Catalog), importerade idéer, indexeringsteori, citeringsteori och kommunikationsteori.

De två största specialiteterna inom informationsvetenskap, utifrån resultatet av faktorerna, är experimentell återvinning, som behandlar design och utvärdering av system för dokumentåtervinning och citeringsanalys, där fokus ligger på samband mellan vetenskaplig litteratur. Nästan hälften av författarna i undersökningen tillhör något av de två största faktorerna som nämnts ovan. Den tredje största faktorn är onlineåtervinning, som beskrivs som praktisk återvinning, där diskussionerna rör verkliga databaser till skillnad från experimentell återvinning som mer tar upp teorier om återvinning. Bibliometri är den fjärde faktorn och författarna inom denna faktor tar upp matematiska modeller avseende text- och bibliografiska statistiska fördelningar. Faktorn bibliometri skiljer sig från faktorn citeringsanalys då bibliometri tar upp aspekter som är tidlösa, generella och abstrakta. Den femte faktorn är generella bibliotekssystem och här tas ämnen som exempelvis biblioteksautomatisering, biblioteks- och informationsservicepolicier och utvärdering av återvinningssystem upp. Vetenskaplig kommunikation är den sjätte faktorn och White och McCain anser att detta område har kopplingar till ämnen utanför informationsvetenskapen som exempelvis sociologi. Den sjunde faktorn är användarteorier och här tas artiklar som handlar om den kognitiva aspekten av informationsvetenskap. OPAC är den åttonde faktorn och här behandlas datoriserade bibliotekskataloger och då särskilt de som grundar sig på sökning genom ämnesord. I den nionde faktorn, importerade idéer, finns författare som tillhör andra discipliner som exempelvis kognitiv vetenskap, informationsteori och datorvetenskap. De tre följande faktorerna, indexeringsteori, citeringsteori och kommunikationsteori är mindre ämnesfält eller innehåller författare som har förekommit främst i andra faktorer.

White och McCain anser att de åtta första faktorerna representerar specialiteter inom informationsvetenskap. Utifrån dessa åtta faktorer går det att urskilja två större

---

<sup>5</sup> Se kapitelavsnitt 2.3

huvudkategorier varav den ena är analytiska studier av forskares publikationer och forskarnas sociala kontext, som består av citeringsanalys och citeringsteori, bibliometri och vetenskaplig kommunikation. Den andra huvudkategorin benämns som studier av människa – dator – litteratur samspel och består av experimentell och praktisk återvinning, generella bibliotekssystem, användarteori, OPAC och indexeringsteori.

En MDS-karta skapades för varje åttårsperiod. MDS-kartorna visade att de författare som formade de olika faktorerna även hamnade nära varandra i kartorna. När kartorna från de tre tidsperioderna jämfördes visade dessa på ungefär samma resultat, inga större förändringar har skett inom informationsvetenskap förutom att vissa ämnen har blivit större och andra mindre. De författare som skriver om informationsåtervinning skapar ett stort kluster som placeras till höger i kartorna och de författare som behandlar vetenskaplig kommunikation hamnar till vänster tillsammans med ämnen som bibliometri och citeringsanalys. Till höger i kartorna finns även författare som behandlar generella bibliotekssystem. Det finns inget kluster av författare i mitten av MDS-kartorna. White och McCain tolkar detta som att det inom informationsvetenskap saknas betydelsefulla centrala författare, som används av de andra ämnesområdena, vilket visar på att strukturen för informationsvetenskap består av flera specialiteter runt en svag mitt. Ur kartorna kunde utläsas två huvudinriktningar: informationsåtervinning till höger i kartan och bibliometri till vänster. White och McCain väljer dock att använda termen *domänanalys*, som Hjörland och Albrechtsen lanserade (1995), istället för bibliometri. De anser att domänanalys bättre beskriver inriktningen med ämnen som exempelvis vetenskaplig kommunikation och vetenskaps- och kunskaps sociologi.

Då ovanstående undersökningar ligger ett par år bakåt i tiden ville vi även kunna jämföra våra resultat med undersökningar som gjorts nära vår egen tidsmässigt. En undersökning av biblioteks- och informationsvetenskap genom fulltextanalys och bibliometriska metoder gjordes 2006 av Janssens, Leta, Glänzel & De Moor vilken presenteras i artikeln *Towards mapping library and information science*. De forskningsfrågor som ställs är bland annat vilka forskningsområden som är de huvudsakliga inom informationsvetenskap idag, om nya områden har skapat en egen terminologi samt hur ämnen och underdiscipliner är representerade av tidskrifter inom området. Materialet bestod av artiklar och anteckningar från tidskrifterna *Information Processing & Management*, *JASIST*, *JDOC*, *Journal of Information Science* och *Scientometrics* under åren 2002-2004.

Utifrån insamlat och bearbetat material fick man fram sex kluster som grupperade sig i tre större och tre mindre kluster. De tre större klustren är Bibliometri1 som innehåller artiklar om citeringsanalys, utvärdering av vetenskap och ämnesområden, kluster IR som innehåller artiklar om både teoretisk och anpassad IR samt kluster Social där artiklar som behandlar exempelvis kunskapsorganisation, användarstudier och informationssökning återfinns. De tre mindre klustren är följande: Bibliometri2 som till skillnad från Bibliometri1 är mer inriktad på teori och metod inom bibliometri, Webometri innehållande artiklar om exempelvis länkning av information på webben och det minsta klustret Patent där artiklarna behandlar tekniska studier och patent.

De olika klustren placerades i en MDS-karta för att visa hur forskningsområdena inom biblioteks- och informationsvetenskap förhåller sig till varandra och till de tidskrifter som användes i undersökningen. Det visade sig att Patent klustret var det som skiljde ut sig mest från de övriga. Bibliometri1 var närmre klustren IR och Social än vad

Bibliometri<sup>2</sup> var. Webometri låg dock närmre Bibliometri<sup>2</sup>. När tidskrifterna togs med i resultatet så visade det på att artiklarna i klustren IR, Social och Webometri alla var representerade i tidskrifterna JIS, IPM, JASIST och JDOC, medan de båda bibliometri- och patentklustrens artiklar var hemmahörande i Scientometrics.

Ännu en studie där man kartlagt ämnesområdet informationsvetenskap som gjorts på senare tid, men med ett historiskt jämförande perspektiv är Peter van den Besselaar och Gaston Heimeriks från 2006: *Mapping research topics using word-reference co-occurrences: A method and exploratory case study*. Syftet med denna studie var att introducera en metod för att kartlägga forskningsområden genom förekomster av ord i artiklarnas titlar och de citerade referenser som används till just dessa artiklar. De citerade referenserna används för att sätta in titelorden i en större kontext. Orden i artiklarnas titlar används för att klustra artiklar och enligt författarna bildar artiklarna från tidskrifterna en forskningsfront. För att undersöka hur forskningsfronten har förändrats över tid har man valt ut följande år för studien: 1986, 1992, 1996, 2000 och 2002. Materialet för analysen är artiklar och recensioner i åtta tidskrifter som enligt författarna behandlar forskningsområdet informationsvetenskap: JDOC, *Information Processing & Management*, *Annual Review of Information Science and Technology*, *Proceedings Asist*, *Canadian Journal of Information Science*, *Journal of Information Science*, *Scientometrics* och JASIST.

De resultat som man fick fram för 1986 med avseende på forskningsfronten inom informationsvetenskap bestod av två huvudgrupper av artiklar: informationsåtervinning och scientometri. Ett mindre forskningsfält var biblioteksstudier. Informationsåtervinning bestod av följande undergrupper: generell informationsåtervinning, informationssökning, kataloger och index samt sociala och politiska aspekter på informationssamhället. Scientometriklustret var indelat i artiklar om indikatorer och kartläggning, scientometriska fördelningar, utvärdering av vetenskap och metoder för kvantitativa analyser.

Forskningsfronten 1992 liknade resultaten för 1986 med informationsåtervinning och scientometri som de största forskningsområdena. Scientometri var dock bättre representerat 1992 än det var 1986. Resultaten från 1996 visade igen att informationsåtervinning och scientometri var de två huvudgrupperna av forskningsområden. Inom scientometri var ämnesområdena impact factor, vetenskapligt samarbete och citeringsanalys de mest betydelsefulla. Även år 2000 var informationsåtervinning och scientometri de två stora forskningsområdena men ett nytt forskningsområde kallat webbstudier framträder. Detta forskningsområde behandlar bland annat webbanvändning och vetenskaplig kommunikation på webben som exempelvis e-publicering. På kartan placeras och länkas webbstudier mellan informationsåtervinning och scientometri. Det sista året för undersökningen, 2002, har forskningsområdet webbstudier växt sig större än tidigare och behandlar webbanvändning och sökning samt digitala bibliotek. Under informationsåtervinning finns undergrupperna informationssökning, informationsåtervinning och metoder för informationsåtervinning. Forskningsområdet scientometri bestod av impact analys, citeringsanalys, domänanalys, vetenskaplig kommunikation och metoder. År 2002 kunde även följande forskningsområden urskiljas: teorier om information, biblioteksforskning och kunskapshantering (knowledge management).

Sammanfattningsvis anser författarna att metoden med samförekomster av ordreferenser är användbar för att kartlägga forskningsområden. Det som gör den användbar är att titelorden reflekterar ämnet inom forskningsområdet och referenserna visar på ämnets position inom forskningsområdet. Med avseende på kartläggningen av forskningsfronten inom informationsvetenskap fick författarna fram att informationsåtervinning och scientometri var de två stora forskningsområdena och att webbstudier har kommit fram som ett betydelsefullt forskningsområde under senare tid. Författarna anser att detta visar på att informationsvetenskap har gått från att ha en struktur med två poler, informationsåtervinning och scientometri, till en struktur med tre poler, med webbstudier som den tredje polen, under tiden för undersökningen.

## 5 Metod och material

Här redovisas varifrån vi fått vårt datamaterial, hur insamlingen av densamma gått till samt med vilka metoder vi har bearbetat den för vår analys.

### 5.1 Insamling av data

Vi använder oss i denna undersökning utav citeringsdatabasen SSCI som produceras av The Institute for Scientific Information (ISI). SSCI innehåller bibliografiska uppgifter från artiklar publicerade i ca 1725 vetenskapliga tidskrifter som täcker mer än 50 olika discipliner (Thomson Scientific 2007a).

För att få fram de poster som användes i vår undersökning sökte vi i ISI Web of Science. Vi använde "Advanced search" och angav där vilka tidskrifter som sökningen skulle utföras i samt tidsperioden år 2004-2006. Posterna som återvanns med sökningen hänvisar till olika typer av dokument men vi har valt att avgränsa materialet för undersökningen till att enbart omfatta artiklar. Med denna avgränsning så fick vi fram totalt 455 artiklar, JASIST 363 och JDOC 92 artiklar, för åren 2004-2006. Dessa 455 artiklar ligger till grund för vår undersökning.

JDOC är en av de äldsta tidskrifterna inom informationsvetenskap och utkom första gången 1945, fulltextartiklar på nätet finns tillgängliga från och med år 1997. Tidskriften som kommer 6 gånger per år ges ut av Emerald Group i London. De ämnen som tas upp i JDOC är bland annat: informationsvetenskap, biblioteksvetenskap och andra discipliner, informations- och kunskapshantering, informations- och kunskapsorganisation och informationssökning och återvinning (Journal of Documentation 2007).

JASIST (2001-) är den nya titeln på Journal of the American Society for Information Science (JASIS, 1970-2000). JASIST behandlar ny forskning inom teorier om informationsvetenskap som exempelvis informationsteori, bibliometri, IR, utvärdering och mätning av information. Även aspekter som kommunikation, management, ekonomi och marknadsföring av information tas upp. Tillämpad informationsvetenskap som exempelvis design av informationssystem, informationsteknologi och elektronisk dokumenthantering är ett annat område som JASIST behandlar (Journal of the American Society for Information Science and Technology 2007).

De tidskrifter, JDOC och JASIST, som vi har valt för vår undersökning behandlar därmed olika aspekter på informationsvetenskap som forskningsområde.

### 5.2 Behandling av data

För att kunna behandla och analysera den data som vi har samlat in använder vi programmet Bibexcel. Vi tar här upp de olika moment som genomfördes för att bearbeta datamaterialet.

#### 5.2.1 Bibexcel och bibliografisk koppling

För att bearbeta den data som består av referenser använde vi programmet Bibexcel som har skapats av Olle Persson på Inforsk vid Umeå universitet. Bibexcel är ett free-ware program och är därmed gratis att använda i akademiskt syfte. Bibexcel kan användas för flera olika bibliometriska analyser, exempelvis co-citeringsanalys, klusteranalys och

analys avseende bibliografisk koppling. Med Bibexcel får man fram datafiler med tabbad data som kan importeras till Excel eller liknande program för att bearbetas eller analyseras ytterligare. I vårt fall använder vi Bibexcel för att få fram de bibliografiska kopplingarna som finns i tidskriftsartiklarna utifrån vår sökning i SSCI.

Av de 455 artiklarna som utvanns ifrån tidskrifterna togs 158 artiklar bort, en avgörande anledning till detta bortfall var att vi ej gjort någon standardisering av referenssträngarna<sup>6</sup>, samt att vissa av artiklarna helt saknade referenser eller gemensamma referenser med de övriga artiklarna. Kvar återstod då 297 artiklar. Data från dessa 297 artiklar bearbetades i Bibexcel med syfte att göra de bibliografiska kopplingarna. Vi valde att sätta ett värde på minst två gemensamma referenser mellan artiklarna för att inte få för många par, vilket skulle leda till ett svårarbetat material. Den bibliografiska kopplingsfunktionen i Bibexcel gav utifrån ovanstående material slutligen 918 kopplingar.

### 5.2.2 Normalisering av referenslistors längd

Olika artiklar har olika antal referenser, vilket är något man måste ta hänsyn till vid uträkningar av olika dokumenters kopplingsstyrka, dvs. antal gemensamma referenser. För att komma till rätta med referenslistornas olika längder bör man göra en så kallad normalisering av dessa. Det finns olika metoder för normalisering, en av dem är cosinusmättet, vilket vi har använt oss utav. Det definieras enligt följande (Jarneving 2005b, s. 58):

$$NCS_{ij} = \frac{r_{ij}}{\sqrt{(n_i * n_j)}}$$

$NCS_{ij}$  är den normaliserade kopplingsstyrkan mellan artiklarna i och j;

$r_{ij}$  är antalet gemensamma referenser för artiklarna i och j;

$n_i$  är antal referenser i referenslistan i artikel i;

$n_j$  är antal referenser i referenslistan i artikel j.

När listan med kopplingsstyrka för varje dokument genererats i Bibexcel beräknade vi kopplingsstyrkan utifrån cosinusmättet. Detta arbete utfördes i Excel enligt följande tabell:

---

<sup>6</sup> Se kapitelavsnitt 7.3

| Kolumn1                     | Kolumn 2 | Kolumn 3 | Kolumn 4                       | Kolumn 5                       | Kolumn 6     |
|-----------------------------|----------|----------|--------------------------------|--------------------------------|--------------|
| antal gemensamma referenser | doknr    | doknr    | antal refs. för dok i kolumn 2 | antal refs. för dok i kolumn 3 | Cosinusvärde |
| 2                           | 1        | 30       | 30                             | 24                             | 0,0745       |
| 2                           | 1        | 74       | 30                             | 28                             | 0,0690       |
| 2                           | 1        | 86       | 30                             | 41                             | 0,0570       |
| 2                           | 1        | 255      | 30                             | 49                             | 0,0522       |
| 2                           | 1        | 374      | 30                             | 20                             | 0,0816       |
| 2                           | 1        | 398      | 30                             | 52                             | 0,0506       |
| 2                           | 4        | 74       | 31                             | 28                             | 0,0679       |
| 2                           | 4        | 109      | 31                             | 42                             | 0,0554       |
| 2                           | 4        | 164      | 31                             | 51                             | 0,0503       |
| 2                           | 4        | 180      | 31                             | 77                             | 0,0409       |
| 2                           | 4        | 185      | 31                             | 35                             | 0,0607       |
| 2                           | 4        | 204      | 31                             | 47                             | 0,0524       |
| 2                           | 4        | 255      | 31                             | 49                             | 0,0513       |
| 2                           | 4        | 298      | 31                             | 38                             | 0,0583       |

**Tabell 2.** Tabellen redovisar antal gemensamma referenser (kolumn 1) för ett urval kopplade dokumentpar (kolumn 2 och 3) samt de enskilda dokumentens totala antal referenser (kolumn 4 och 5). I kolumn 6 redovisas det uträknade cosinusvärdet, avrundat till fyra decimaler, för de enskilda dokumentparen

När normaliseringen är genomförd kan man med hjälp av de värden man fått fram i nästa steg gruppera dokumenten med hjälp av olika klustringsmetoder, i vårt fall metoden Perssons Party clustering som finns i programmet Bibexcel.

### 5.2.3 Klustring

Vid klustringen av vårt material använde vi oss av klustringsrutinen som ingår i programmet Bibexcel, även kallad Persson's party clustering. Persson's party clustering finns beskrivet i Bibexcels hjälpfil på följande vis [fritt översatt från engelska]:

“Föreställ dig att du har följande lista med par (i första kolumnen antal gemensamma citeringar; 10 A B = A och B har 10 gemensamma citeringar):

|    |   |   |
|----|---|---|
| 10 | A | B |
| 9  | D | F |
| 8  | B | C |
| 7  | A | C |
| 6  | F | G |
| 5  | H | I |
| 4  | A | H |

Klustringsrutinen kommer att ha följande händelseförlopp:  
(paren är inbjudna till ett party):

A-B kommer först, måste vänta i hallen

D-F kommer därefter, måste vänta i hallen.

B-C kommer sen, bildar ett kluster-bord tillsammans med A-B inne i rummet

A-C kommer sen, blir raderade eftersom A-C redan är inne i rummet

F-G kommer sen, hittar ingen vän i rummet, går till hallen och hittar D-F

och D-F-G bildar ett kluster inne i rummet  
H-I kommer sen, måste vänta i hallen  
A-H kommer sen, H klustrar med A-B-C, letar sedan i hallen och hittar H-I  
I klustrar med A-B-C-H

Kluster 1 innehåller: A-B-C-H-I

Kluster 2 innehåller: D-F-G”

När man utför klusteranalys med Bibexcels klustringsrutin Persson's party clustering kan inte objekten förekomma i fler än ett kluster samtidigt och objekten behöver bara visa likhet med ett av objekten i ett visst kluster för att få ingå i detta, alltså en form av SLINK<sup>7</sup>. Som exempel på detta kan vi närmare studera objektet I. I ingår endast i ett par, par H I med fem gemensamma citeringar. H ingår däremot även i par med A. Då A även ingår i par med B och C bildar A-B-C-H ett kluster som sedan ytterligare inkluderar I på basis av att I har sin enda samhörighet med H.

De objekt som användes vid klustringen var de kopplade dokumenten med tillhörande cosinusvärde. Vid klustringen bildades det 15 kluster av 277 artiklar. Klustringsrutinen sorterade ut 20 artiklar och dessa ingår därför inte i något kluster.

---

<sup>7</sup> Se kapitelavsnitt 2.3.1

## 6 Resultat

I detta kapitel redovisar vi de kluster med tillhörande identifierat forskningstema som vi har fått fram. Klustren presenteras i den ordning som de utföll i klustringsprocessen.

De olika artiklarna i varje kluster redovisas med artiklarnas författare, titel samt ämnesord (i de fall de finns). Då en del av klustren är väldigt stora redovisas av utrymmesskäl endast femton artiklar. I dessa fall har artiklarna valts ut för att visa på ämnesmässigt typiska samt ämnesmässigt avvikande artiklar. Ordningen som artiklarna inom de olika forskningsteman presenteras i har ingen betydelse i någon form av rangordning.

### 6.1 Identifierade forskningsteman

Vid klustringen av artiklarna i Bibexcel skapades totalt femton kluster. Storleken på klustren varierar mycket, allt från kluster innehållande endast tre artiklar (kluster 5, 10 och 15) till det klart största klustret (kluster 1) som innehöll så mycket som 111 artiklar. Då vi undersökte ämnesinnehållet för varje kluster, för att på så vis få fram ett forskningstema, studerades artiklarnas titel samt ämnesord. I de fall det saknats ämnesord eller då det har varit svårt att definiera tema utifrån enbart dessa samt titel, har vi även studerat artiklarnas abstract. Vi har, i den mån det för oss varit möjligt, valt att benämna de forskningsteman som vi har fått fram med termer inom informationsvetenskap, företrädesvis svenska. När vi inte har hittat någon lämplig term, utifrån ovan nämnda kriterier, för ett forskningstema har vi valt att benämna dessa med egna ord. Nedan presenteras de olika dokumenten i varje kluster med författarnamn (**FF**), titel (**TI**) samt ämnesord (**ÄO**).

#### Kluster 1. (111 artiklar) Forskningstema ”Informationsåtervinning”

**1. FF:** Wacholder N; Liu L **TI:** User preference: A measure of query-term quality **ÄO:** INFORMATION-RETRIEVAL; SEARCHERS SELECTION; VOCABULARY; SEEKING; KEYS

**2. FF:** Tudhope D; Binding C; Blocks D; Cunliffe D **TI:** Query expansion via conceptual distance in thesaurus indexed collections **ÄO:** INFORMATION-RETRIEVAL; DESIGN; NAVIGATION; SUPPORT; MODEL

**3. FF:** Yi K; Beheshti J; Cole C; Leide JE; Large A **TI:** User search behavior of domain-specific information retrieval systems: An analysis of the query logs from PsycINFO and ABC-Clio's Historical Abstracts/America: History and life **ÄO:** HUMANITIES SCHOLARS; RESEARCH PROPOSAL; SUBJECT ACCESS; WEB USERS; ONLINE; DESIGN; VOCABULARY; LIBRARIES; KNOWLEDGE; STUDENTS

**4. FF:** Wu YFB; Li QZ; Bot RS; Chen X **TI:** Finding nuggets in documents: A machine learning approach **ÄO:** DIGITAL LIBRARIES

**5. FF:** Hong T **TI:** The influence of structural and message features on Web site credibility **ÄO:** WORLD-WIDE-WEB; HEALTH INFORMATION; COGNITIVE AUTHORITY; MEDIA CREDIBILITY; DECISION-MAKING; INTERNET; QUALITY; TRUST; JUDGMENT; ATTITUDE

**6. FF:** Shiri A; Revie C **TI:** Usability and user perceptions of a thesaurus-enhanced search interface **ÄO:** INTERACTIVE QUERY EXPANSION; INFORMATION-RETRIEVAL; BIBLIOGRAPHIC DATABASES; SUBJECT KNOWLEDGE; SELECTION; DESIGN; ENVIRONMENT; TACTICS; ACCESS; SYSTEM

**7. FF:** Jorgensen C; Jorgensen P **TI:** Image querying by image professionals **ÄO:** INFORMATION; WEB

**8. FF:** Bystrom K; Hansen P **TI:** Conceptual framework for tasks in information studies **ÄO:** RETRIEVAL SYSTEMS; RELEVANCE; SEEKING; DESIGN; INTERFACE; USER; COMPLEXITY; ENGINEERS; SELECTION; BEHAVIOR

**9. FF:** Spink A **TI:** Multitasking information behavior and information task switching: an exploratory study **ÄO:** SEEKING BEHAVIOR; LIBRARY USE; USERS; RELEVANCE; SYSTEMS; DESIGN; MODELS; LIFE

**10. FF:** Hjørland B **TI:** Empiricism rationalism and positivism in library and information science **ÄO:** DOMAIN-ANALYSIS; KNOWLEDGE; RELEVANCE; EXAMPLE; WORK

**11. FF:** Crudge SE; Johnson FC **TI:** Using the information seeker to elicit construct models for search engine evaluation **ÄO:** RETRIEVAL-SYSTEMS; QUERY REFORMULATION; SEEKING STRATEGIES; DESIGN; SATISFACTION; INTERFACE; RELEVANCE; USERS

**12. FF:** Sihvonen A; Vakkari P **TI:** Subject knowledge improves interactive query expansion assisted by a thesaurus **ÄO:** INFORMATION-RETRIEVAL; SEARCH TACTICS

**13. FF:** Koshman S **TI:** Comparing usability between a visualization and text-based system for information retrieval **ÄO:** saknas

**14. FF:** Muresan G; Harper DJ **TI:** Topic modeling for mediated access to very large document collections **ÄO:** INFORMATION-RETRIEVAL; DESIGN

**15. FF:** Bodoff D; Robertson S **TI:** A new unified Probabilistic model **ÄO:** INFORMATION-RETRIEVAL; DOCUMENT-RETRIEVAL; MAXIMUM-LIKELIHOOD; RELEVANCE; LOGIC

Detta tema är undersökningens i särklass största bestående av 111 artiklar och vi har valt att benämna det ”Informationsåtervinning”. En stor del av artiklarna har orden ”Information retrieval”, eller ”Retrieval” som ämnesord. En del av artiklarna tar även upp informationssökning, men detta ämne och dess koppling till informationsåtervinning återkommer och presenteras längre fram i våra resultat.

I detta tema tar man i artiklarna upp informationsåtervinning i olika databaser och på Internet och webben. Olika undersökningar angående arbete med thesaurus behandlas samt så kallad ”query-expansion” – utvidgning av sökfrågor. Majoriteten av artiklarna behandlar textåtervinning, men exempelvis artikel sju tar upp bildåtervinning.

Det finns dock en del artiklar där innehållet avviker något eller helt från det presenterade temat. I urvalet ovan är artiklarna fem och tio exempel på sådana artiklar med avvikande innehåll. I artikel fem tar man upp trovärdighet på hälsorelaterade webbsidor. I artikel tio undersöker man hur viktiga och inflytelserika epistemologierna<sup>8</sup> ”empirism”<sup>9</sup>, ”rationalism”<sup>10</sup> och ”positivism”<sup>11</sup> varit för biblioteks- och informationsvetenskap.

## **Kluster 2. (35 artiklar) Forskningstema “Webometri”**

---

<sup>8</sup> Epistemologi, även kallat kunskapsteori, är en huvudgren av filosofin som studerar frågor om kunskapens natur, objekt och källor, exempelvis: vad är kunskap? (Nationalencyklopedin 2007).

<sup>9</sup> Filosofisk inriktning som betonar erfarenhet som bas för vår kunskap (ibid.).

<sup>10</sup> Filosofisk inriktning som menar att det är möjligt att uppnå kunskap genom att använda förnuftet (ibid.).

<sup>11</sup> Filosofisk inriktning där de två centrala ståndpunkterna är att kunskap om faktiska förhållanden vilar på sinneserfarenhet och kunskap som inte rör det faktiska rör förhållandet mellan idéer (ibid.).

1. **FF:** Bar-Ilan J **TI:** Web links and search engine ranking: The case of Google and the query "jew" **ÄO:** ACADEMIC-INSTITUTIONS; INFORMATION; PAGES
2. **FF:** Thelwall M; Stuart D **TI:** Web crawling ethics revisited: Cost, privacy, and denial of service **ÄO:** INFORMATION; COMMUNICATION; PLACE; SITES
3. **FF:** Leydesdorff L; Bensman S **TI:** Classification and powerlaws: The logarithmic transformation **ÄO:** INFORMATION-SCIENCE; BRADFORD LAW; QUANTITATIVE ASPECTS; INTELLECTUAL SPACE; AUTHOR COCITATION; CO-WORDS; INDICATORS; FREQUENCY; COLLABORATION; DISTRIBUTIONS
4. **FF:** Brody T; Harnad S; Carr L **TI:** Earlier web usage statistics as predictors of later citation impact **ÄO:** ONLINE
5. **FF:** Goodall AH **TI:** Should top universities be led by top researchers and are they? A citations analysis **ÄO:** ASSESSMENT EXERCISE RATINGS; COUNTS; IMPACT; ARCHAEOLOGY; LIBRARY
6. **FF:** Herrera-Viedma E; Pasi G; Lopez-Herrera AG; Porcel C **TI:** Evaluating the information quality of Web sites: A methodology based on fuzzy computing with words **ÄO:** GROUP DECISION-MAKING; WORLD-WIDE-WEB; AGGREGATION OPERATORS; OWA OPERATORS; RETRIEVAL; FRAMEWORK; DESIGN; MODEL
7. **FF:** Kuperman V **TI:** Productivity in the Internet mailing lists: A bibliometric analysis **ÄO:** SUCCESS-BREEDS-SUCCESS; LOTKAS LAW; INFORMETRIC DISTRIBUTIONS; SCIENTIFIC PRODUCTIVITY; CUMULATIVE ADVANTAGE; ZIPFS LAW; NEWSGROUPS; SCIENCE; MODEL
8. **FF:** Price L; Thelwall M **TI:** The clustering power of low frequency words in academic webs **ÄO:** SCIENCE; COMMUNICATION; ALGORITHM; SITES
9. **FF:** Thelwall M **TI:** Interpreting social science link analysis research: A theoretical framework **ÄO:** WEB IMPACT FACTORS; WORLD-WIDE-WEB; CITATION ANALYSIS; SCHOLARLY COMMUNICATION; WEBOMETRICS; SITES; BIBLIOMETRICS; HYPERLINK; NETWORK
10. **FF:** Cothey V **TI:** Web-crawling reliability **ÄO:** WORLD-WIDE-WEB; SEARCH ENGINES; BIBLIOMETRICS; INFORMATION; COLLECTION; INTERNET
11. **FF:** Robertson S **TI:** Understanding inverse document frequency: on theoretical arguments for IDF **ÄO:** PROBABILISTIC MODELS; INFORMATION; SPECIFICITY; RETRIEVAL
12. **FF:** Cyr D; Trevor-Smith H **TI:** Localization of Web design: An empirical comparison of German Japanese, and United States Web site characteristics **ÄO:** saknas
13. **FF:** Fan WG; Fox EA; Pathak P; Wu H **TI:** The effects of fitness functions on genetic programming-based ranking discovery for web search **ÄO:** RELEVANCE FEEDBACK; INFORMATION-RETRIEVAL; TEXT RETRIEVAL; ALGORITHMS; QUERY
14. **FF:** Feitelson DG; Yovel U **TI:** Predictive ranking of computer scientists using CiteSeer data **ÄO:** CITATION ANALYSIS; MODEL
15. **FF:** Björneborn L; Ingwersen P **TI:** Toward a basic framework for webometrics **ÄO:** WORLD-WIDE-WEB; COMPLEX NETWORKS; INTERNET; UNIVERSITY; COMMUNICATION; BIBLIOMETRICS; INFORMATION; DOCUMENTATION; CONNECTIVITY; IMPACT

Vårt andra tema består till största del av artiklar där författarna på ett eller annat sätt beskriver, eller utför egna, studier av webben så kallad *Webometri*. Termen webometri introducerades av Almind och Ingwersen 1997. Det finns olika definitioner på webometri, men en som vi har valt att följa är Björneborn och Ingwersen (2004) vilka

menar att webometrisk forskning består av fyra huvudområden; innehållsanalys av webbsidor, analys av länkstruktur på webben, analys av webbanvändande samt webbt teknisk analys. Sett till ämnesorden i forskningstemat förekommer många termer som används inom bibliometri tillsammans med Internet och Webben, exempel på sådana artiklar är artikel fyra, artikel sju och artikel åtta. Ordet *webometri* förekommer i sig inte särskilt ofta i varken ämnesord eller titel, men ser man till ovanstående definition av Björneborn och Ingwersen så stämmer flertalet av artiklarna överens med forskningstemat ”Webometri”. Exempel på artiklar i urvalet ovan som inte kan sägas handla om webometri är artikel fem och artikel elva. I artikel fem har man med citeringsanalys undersökt hur olika universitets främsta forskare och ledare finns representerade i aktuell forskning, undersökningen inbegriper inget webbmaterial. I artikel elva har man studerat olika försök att teoretisera inverse document frequency (IDF).

### **Kluster 3. (39 artiklar) Forskningstema “Bibliometri”**

- 1. FF:** Leydesdorff L; **L:** Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment **ÄO:** AUTHOR COCITATION ANALYSIS; PEARSONS-R
- 2. FF:** Zuccala A **TI:** Author cocitation analysis is to intellectual structure as web colink analysis is to ... ? **ÄO:** SITES; LINKS; INFORMATION; SIMILARITY; SCIENCE; WEBOMETRICS; CITATION; SPACE
- 3. FF:** Leydesdorff T **TI:** Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? **ÄO:** SCIENCE; SPECIALTIES; MAPPINGS; MATRICES
- 4. FF:** Freitas HR; Ribeiro-Neto B; Vale RF; Laender AHF; Lima LRS **TI:** Categorization-driven cross-language retrieval of medical information **ÄO:** PROBABILISTIC INFERENCE
- 5. FF:** Chen CM **TI:** CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature **ÄO:** TRIASSIC MASS EXTINCTION; DOMAIN VISUALIZATION; TERRORIST ATTACKS; SCIENCE; PARADIGMS; KNOWLEDGE; NETWORKS; CITY; SEPTEMBER-11; TECHNOLOGY
- 6. FF:** van Raan AFJ **TI:** Statistical properties of Bibliometric indicators: Research group indicator distributions and correlations **ÄO:** CITATION DISTRIBUTIONS; SCIENTIFIC PUBLICATION; JOURNAL IMPACT; SCIENCE; NETWORKS; COMPETITION
- 7. FF:** Jimenez-Contreras E; De La Moneda M; de Osmá ER; Bailon-Moreno R; Ruiz-Banos R **TI:** A bibliometric model for journal discarding policy at academic libraries **ÄO:** saknas
- 8. FF:** Morris SA **TI:** Manifestation of emerging specialties in journal literature: A growth model of papers, references, exemplars, bibliographic coupling, cocitation, and clustering coefficient distribution **ÄO:** SCIENTIFIC PAPERS; COMPLEX NETWORKS; SCIENCE; VISUALIZATION; PUBLICATION; DOCUMENTS
- 9. FF:** Egghe L **TI:** Zipfian and Lotkaian continuous concentration theory **ÄO:** SQUARE-ROOT LAW; SCIENTIFIC PRODUCTIVITY; GINI-INDEX; PRICE LAW; DISTRIBUTIONS; CURVE; CONSEQUENCES; CONSTRUCTION; AMBIGUITY; NETWORKS
- 10. FF:** Frandsen TF **TI:** Journal interaction - A bibliometric analysis of economics journals **ÄO:** EUROPEAN ECONOMICS; RESEARCH OUTPUT; CORE JOURNALS; NEW-MODEL; SCIENCE; CITATION; PUBLICATIONS; ALLOCATION; AMERICAN; RANKINGS
- 11. FF:** Coleman A **TI:** Instruments of cognition: Use of citations and web links in Online teaching materials **ÄO:** RESEARCH-PROJECT; DOCUMENT USE; MODEL

**12. FF:** Bensman SJ **TI:** Urquhart and probability: The transition from librarianship to library and information science **ÄO:** NATIONAL-LENDING-LIBRARY; AMERICAN MEN; SERIALS; TECHNOLOGY; PERIODICALS; LISTS

**13. FF:** Meadows J **TI:** The immediacy effect - then and now **ÄO:** HALF-LIFE; IMPACT FACTOR; OBSOLESCENCE; CITATION; INDEX; JOURNALS; SCIENCE; AGE

**14. FF:** Schloegl C; Stock WG **TI:** Impact and relevance of LIS journals: A scientometric analysis of international and German-language LIS journals - Citation analysis versus reader survey **ÄO:** FOR-INFORMATION-SCIENCE; LIBRARY; RANKINGS; JASIS

**15. FF:** Rousseau R; Zuccala A **TI:** A classification of author co-citations: Definitions and search strategies **ÄO:** INTELLECTUAL SPACE; CITATION ANALYSIS; SCIENCE; COMMUNICATION; PERFORMANCE; ISSUES

Inom detta tema beskrivs i de olika artiklarna olika aspekter på bibliometri. I en del av artiklarna beskrivs olika teorier om bibliometri, i andra artiklar har man utfört olika bibliometriska undersökningar. I ytterligare andra artiklar har man studerat vissa specifika analysmetoder man använder sig av inom bibliometrisk forskning.

En del av artiklarna i klustret har inte bibliometri som huvudtema. Som exempel kan nämnas artikel nummer tolv där författaren presenterar biblioteks- och informationsforskaren Donald J. Urquhart och dennes gärning. Urquhart hade viss betydelse för den bibliometriska forskningen vilket beskrivs i artikeln, men i övrigt så handlar inte artikeln främst om bibliometri. Ett annat exempel på en artikel som ej stämmer överens med detta tema är artikel nummer fyra. I artikel nummer fyra tar man upp en metod för att kunna söka efter medicinsk information i databaser med artiklar på olika språk, något som snarast kan sägas tillhöra forskningstema "Informationsåtervinning".

#### **Kluster 4. (37 artiklar) Forskningstema "Informationssökning"**

**1. FF:** Kelly D **TI:** Measuring online information seeking context, part 1: Background and method **ÄO:** TASK COMPLEXITY; RETRIEVAL; KNOWLEDGE; USER; RELEVANCE; SERVICES; SYSTEMS; DOMAIN; MODEL; LIFE

**2. FF:** Heinström J **TI:** Broad exploration or precise specificity: Two basic information seeking patterns among students **ÄO:** BEHAVIOR; PERSONALITY; SCIENTISTS; EDUCATION; INTERNET; CONTEXT

**3. FF:** Xu YJ; Chen ZW **TI:** Relevance judgment: What do information users consider beyond topicality? **ÄO:** RESEARCH-PROJECT; COGNITIVE MODEL; DOCUMENT USE; CRITERIA; RETRIEVAL; DIMENSIONS; SCIENCE; PERSPECTIVES; HISTORY; VIEW

**4. FF:** Kalbach J **TI:** "I'm feeling lucky": The role of emotions in seeking information on the Web **ÄO:** SEARCH PROCESS; UNCERTAINTY; BEHAVIOR; SYSTEM

**5. FF:** Shiri A; Revie C **TI:** Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation **ÄO:** INFORMATION-RETRIEVAL INTERFACES; CONTROLLED VOCABULARY; FREE-TEXT; NATURAL-LANGUAGE; RELEVANCE FEEDBACK; SEEKING STRATEGIES; DIGITAL LIBRARIES; COMMUNITY SYSTEM; SUBJECT ACCESS; DESIGN

**6. FF:** Spink A; Cole C **TI:** Human information behavior: Integrating diverse approaches and information use **ÄO:** SEEKING RESEARCH; RETRIEVAL; EVOLUTION; SCIENCE; LIFE; CONTEXT; NEEDS

- 7. FF: Xie H TI:** Understanding human-work domain interaction: Implications for the design of a corporate digital library **ÄO:** INFORMATION-SEEKING; TASK COMPLEXITY; RETRIEVAL; USER; EXPLORATION; INTENTIONS; INTERFACES; FRAMEWORK; SYSTEMS
- 8. FF: Budd JM TI:** Phenomenology and information studies **ÄO:** SCIENCE
- 9. FF: Vilar P; Zumer M TI:** Comparison and evaluation of the user interfaces of e-journals **ÄO:** COMPUTER INTERACTION; CD-ROM; WEB; EASE
- 10. FF: Talja S TI:** The social and discursive construction of computing skills **ÄO:** INFORMATION TECHNOLOGY; LITERACY; ATTITUDES
- 11. FF: Hepworth M TI:** A framework for understanding user requirements for an information service: Defining the needs of informal carers **ÄO:** SEEKING; RETRIEVAL; BEHAVIOR; DESIGN
- 12. FF: Rieh SY TI:** On the web at home: Information seeking and web searching in the home environment **ÄO:** SPECIAL ISSUE; CONTEXT; BEHAVIOR; INTERNET; LIFE; RETRIEVAL; SERVICES; USER
- 13. FF: Spink A; Sollenberger TI:** Elicitation purposes and tasks during mediated information search **ÄO:** RETRIEVAL; SEEKING; DESIGN; USERS; BEHAVIOR; DATABASE; IR
- 14. FF: Macpherson K TI:** An information processing model of undergraduate electronic database information retrieval **ÄO:** ONLINE CATALOGS; SEARCH INTERFACE; SYSTEMS-DESIGN; LITERACY; PERSPECTIVES; INSTRUCTION; FEEDBACK; SEEKING; HARD
- 15. FF: Kabel S; de Hoog R; Wielinga BJ; Anjewierden A TI:** The added value of task and ontology-based markup for information retrieval **ÄO:** COMPLEXITY; KNOWLEDGE; IR

Forskningstema "informationssökning" består till huvuddel av artiklar som behandlar hur människor söker information, och olika sökstrategier. Flera av studierna är inriktade på olika kategorier av yrkesgrupper. I artiklarna undersöks också i vilka olika miljöer människor söker information och om detta påverkar deras sökstrategier. Som exempel på en artikel representativ för temat kan nämnas undersökningen i artikel elva som går ut på att utarbeta ett ramverk som ska kunna hjälpa personer vilka obetalt vårdar anhöriga i hemmet att söka information. Ett annat exempel är artikel tolv där man studerat om och i så fall hur hemmiljön påverkar informationssökning och sökningar på webben.

Det finns en del artiklar som inte riktigt faller inom ramen för temat, vi har i ovanstående urval av temats artiklar valt att redovisa fyra sådana; artikel sju, åtta, nio samt artikel elva. Dessa artiklar handlar i viss mån, ur ett större perspektiv betraktat, även om informationssökning, men inriktningarna på artiklarna skiljer sig från de övriga. I artikel sju och nio tar man istället upp användargränssnitt. I artikel åtta beskrivs hur människor som arbetar med information professionellt med hjälp av fenomenologi<sup>12</sup> kan interagera med personer som söker information.

I en del av artiklarna tar man upp informationsåtervinning, vilket vi tidigare tagit upp i vårt första forskningstema. Informationssökning och informationsåtervinning är besläktade ämnen och studeras tillsammans i många artiklar, både i detta tema och i tema "Informationsåtervinning". Här har vi ändå funnit att i majoriteten av artiklarna ligger huvudfokus på informationssökning.

---

<sup>12</sup> Lärnan om det som visar sig för medvetandet (Nationalencyklopedin 2007)

## **Kluster 5. Forskningstema “Vetenskaplig kommunikation”**

- 1. FF:** Birnholtz JP **TI:** What does it mean to be an author? The intersection of credit, contribution, and collaboration in science **ÄO:** SCIENTIFIC COMMUNICATION; HYPERAUTHORSHIP; PROPOSAL
- 2. FF:** Cronin B **TI:** Bowling alone together: Academic writing as distributed cognition **ÄO:** COLLABORATION; HYPERAUTHORSHIP
- 3. FF:** White HD; Wellman B; Nazer N **TI:** Does citation reflect social structure? Longitudinal evidence from the "Globenet" interdisciplinary research group **ÄO:** HYBRID PROBLEM AREA; SCHOLARLY COMMUNICATION; SCIENTIFIC LITERATURE; ORGANIZATION; DYSLEXIA; AUTHORS; MODEL; TIES

Vetenskaplig kommunikation är som tidigare nämnts i uppsatsen<sup>13</sup> något som kan studeras med citeringsanalys. I artikel tre i detta forskningstema undersöker man om man kan utläsa även sociala kontaktnät med hjälp av citeringsanalys. I artikel ett och två beskrivs vetenskaplig kommunikation och speciellt undersöker man forskningskollektiv.

## **Kluster 6. Forskningstema “Informationsvetenskapens filosofi”**

- 1. FF:** Karamuftuoglu M **TI:** Information arts and information science: Time to unite? **ÄO:** MEDICAL LITERATURES; RETRIEVAL; ORGANIZATION; KNOWLEDGE
- 2. FF:** Swanson DR; Smalheiser NR; Torvik VI **TI:** Ranking indirect connections in literature-based discovery: The role of medical subject headings **ÄO:** COMPLEMENTARY LITERATURES; GENERATING HYPOTHESES; SCIENTIFIC DISCOVERY; ALZHEIMERS-DISEASE; FISH OIL; KNOWLEDGE; IMPLICIT; TEXT; MAGNESIUM; MEDLINE
- 3. FF:** Zins C **TI:** Redefining information science: from "information science" to "knowledge science" **ÄO:** FOUNDATIONS; PHILOSOPHY; METATHEORY
- 4. FF:** Holland GA **TI:** Associating social constructionism and extended cognition in information studies **ÄO:** SCIENCE
- 5. FF:** van der Eijk CC; van Mulligen EM; Kors JA; Mons B; van den Berg J **TI:** Constructing an associative concept space for literature-based discovery **ÄO:** INFORMATION-RETRIEVAL; SCIENCE; SYSTEM

Artiklarna i detta tema behandlar informationsvetenskap utifrån ett mer teoretiskt perspektiv där man diskuterar ämnets uppbyggnad. Vissa av författarna försöker utarbeta olika ramverk för hur man kan studera ämnet tillsammans med andra forskningsområden. Exempelvis tar man i artikel ett upp hur information arts och informationsvetenskap skulle kunna sammanföras under en gemensam ram, som domänanalys. I artikel fyra menar författarna att kognitiv vetenskap och informationsvetenskap skulle kunna studeras gemensamt vilket skulle vara positivt för båda ämnesområdena. Artikel två och fem faller lite utanför de övriga tre då man i dessa båda främst behandlar textanalys.

## **Kluster 7. Forskningstema “Utvärdering av vetenskapliga tidskrifter”**

---

<sup>13</sup> Se kapitelavsnitt 3

1. **FF:** Walters GD **TI:** Measuring the utility of journals in the crime-psychology field: Beyond the impact factor **ÄO:** Saknas
2. **FF:** Knothe G **TI:** Comparative citation analysis of duplicate or highly related publications **ÄO:** ENANTIOSELECTIVE TOTAL-SYNTHESIS; 1ST TOTAL-SYNTHESIS; DOUBLE-BOND; STEREOCONTROLLED SYNTHESIS; MASS-SPECTROMETRY; PINENE PATH; TAXOL; ACID; POSITION; IMPACT
3. **FF:** Wang JX; Guan JC **TI:** The analysis and evaluation of knowledge efficiency in research groups **ÄO:** DATA ENVELOPMENT ANALYSIS; IMPACT; SCIENCE; PERFORMANCE; MANAGEMENT; JOURNALS; CREATION; MODEL; DEA
4. **FF:** Sombatsompop N; Markpin T **TI:** Making an equality of ISI impact factors for different subject fields **ÄO:** Saknas
5. **FF:** Clyde LA **TI:** Evaluating the quality of research publications: A pilot study of school librarianship **ÄO:** LIBRARY INFORMATION-SCIENCE; JOURNAL LITERATURE; IMPACT FACTOR; PROFESSIONAL JOURNALS; PERCEIVED PRESTIGE; CITATION
6. **FF:** Vinkler P **TI:** Characterization of the impact of sets of scientific papers: The Garfield (Impact) Factor **ÄO:** INDICATORS; CITATION; SCIENTOMETRICS; MODEL

Detta forskningstema har vi valt att kalla ”Utvärdering av vetenskapliga tidskrifter” - något som kan göras på olika sätt. De flesta av artiklarna handlar om citeringsanalys. I tre av artiklarna tar man upp Impact factor<sup>14</sup>, som mäter vilken betydelse ett objekt (till exempel tidskrifter) har.

### **Kluster 8. Forskningstema “Textkategorisering”**

1. **FF:** Koppel M; Akiva N; Dagan I **TI:** Feature instability as a criterion for selecting potential style markers **ÄO:** TEXT CATEGORIZATION
2. **FF:** Zheng R; Li JX; Chen HC; Huang Z **TI:** A framework for authorship identification of Online messages: Writing-style features and classification techniques **ÄO:** SUPPORT VECTOR MACHINES; NEURAL-NETWORKS; STYLOMETRIC ANALYSIS; ATTRIBUTION; FEDERALIST; DESIGN
3. **FF:** Sun A; Lim EP **TI:** Web unit-based mining of homepage relationships **ÄO:** saknas
4. **FF:** Yoon Y; Lee C; Lee GG **TI:** An effective procedure for constructing a hierarchical text classification system **ÄO:** CATEGORIZATION
5. **FF:** Debole F; Sebastiani F **TI:** An analysis of the relative hardness of Reuters-21578 subsets **ÄO:** AUTOMATED TEXT CATEGORIZATION
6. **FF:** Diaz I; Ranilla J; Montanes E; Fernandez J; Combarro EF **TI:** Improving performance of text categorization by combining filtering and support vector machines
7. **FF:** Melucci M **TI:** Making digital libraries effective: Automatic generation of links for similarity search across hyper-textbooks **ÄO:** DESIGN; CONSTRUCTION; TOOL; WEB
8. **FF:** Frank E; Paynter GW **TI:** Predicting library of Library of Congress Classifications from congress subject headings

Med forskningstema “Textkategorisering” menar vi här analyser av text i syfte att kunna dela in antingen textens författare eller textens olika byggstenar i kategorier. Artikel ett handlar till exempel om att kunna identifiera och utarbeta synonymer i en text. I artikel

---

<sup>14</sup> Se kapitelavsnitt 3.2

två är det istället författare som ska kunna identifieras utifrån text i meddelanden som skrivits online. Ett flertal av artiklarna behandlar eller utarbetar system för automatisk textkategorisering.

### **Kluster 9. Forskningstema “Användandet av e-tidskrifter ”**

- 1. FF:** Huntington P; Nicholas D; Jamali HR **TI:** Article decay in the digital environment: An analysis of usage of OhioLINK by date of publication, employing deep log methods **ÄO:** CONTEMPORARY ACADEMIC RESEARCHER; INFORMATION NEEDS; JOURNAL USE; OBSOLESCENCE; BEHAVIOR; PROGRESS; INSIGHTS
- 2. FF:** Nicholas D; Huntington P; Jamali HR **TI:** What deep log analysis tells us about the impact of big deals: case study OhioLINK **ÄO:** saknas
- 3. FF:** Davis PM; Price JS **TI:** eJournal interface can influence usage statistics: Implications for libraries, publishers, and project COUNTER **ÄO:** ELECTRONIC JOURNALS
- 4. FF:** Nicholas D; Huntington P; Watkinson **TI:** Scholarly journal uagee: The results of deep log analysis **ÄO:** ELECTRONIC JOURNALS; INFORMATION GATEWAYS; DIGITAL LIBRARIES; BEHAVIOR; UNIVERSITY

Artiklarna i kluster 9 handlar om användandet av elektroniska tidskrifter, e-tidskrifter. I tre av artiklarna, artikel ett, två och fyra, har man studerat användarstatistik med hjälp av metoden deep log analysis, P. Huntington och D. Nicholas är medförfattare till dessa tre. I artikel tre var syftet med att studera användandet av e-tidskrifter att se hur gränssnitt kan påverka användarstatistik och därmed i förlängningen vara till hjälp för bland annat bibliotekarier och publicister.

### **Kluster 10. Forskningstema “Kunskapshantering”**

- 1. FF:** Haythornthwaite C **TI:** Learning and knowledge networks in interdisciplinary collaborations **ÄO:** ORGANIZATIONAL KNOWLEDGE; MEDIA
- 2. FF:** Desouza KC; Awazu Y; Wan Y **TI:** Factors governing the consumption of explicit knowledge **ÄO:** ORGANIZATIONAL MEMORY; MANAGEMENT-SYSTEMS; USER ACCEPTANCE; INFORMATION; TECHNOLOGY; TRUST; MODEL; FOUNDATIONS; FIRM
- 3. FF:** Fleischmann KR **TI:** Do-it-yourself information technology: Role hybridization and the design-use interface **ÄO:** RETRIEVAL; SYSTEMS; SCIENCE

Forskningsstema ”Kunskapshantering” är ett av de mindre klustren. Artikel ett och två tar upp olika aspekter på hantering av kunskap inom organisationer, medan man i artikel tre har undersökt en grupp av IT-designers samt användare av deras system och vilka förmågor de har att ta till sig och hantera kunskap inom vitt skilda ämnen. Artikel ett och tre tar bägge upp interdisciplinärt samarbete och utbyte av kunskap. I abstract till artikel ett och två kan man läsa att ”Knowledge management” används som ett övergripande tema, men termen finns inte med i titel eller ämnesord.

### **Kluster 11. Forskningstema “Indexering”**

- 1. FF:** Newman DJ; Block S **TI:** Probabilistic topic decomposition of an eighteenth-century American newspaper **ÄO:** LATENT SEMANTIC ANALYSIS; ALGORITHM
- 2. FF:** Chen L; Zeng J; Tokuda N **TI:** A "stereo" document representation for textual information retrieval **ÄO:** LATENT SEMANTIC ANALYSIS; KNOWLEDGE; SPACE

3. **FF:** Garces PJ; Olivas JA; Romero FP **TI:** Concept-matching IR systems versus word-matching information retrieval systems: Considering fuzzy interrelations for indexing Web pages **ÄO:** MODEL
4. **FF:** Efron M **TI:** Eigenvalue-based model selection during latent semantic indexing **ÄO:** NUMBER; RULES; COMPONENTS; RETAIN
5. **FF:** Ding CHQ **TI:** A probabilistic model for Latent Semantic Indexing **ÄO:** INFORMATION-RETRIEVAL; TEXT

Forskningsstema "Indexering" innehåller fem artiklar varav fyra tar upp indexeringsmetoden Latent Semantic Indexing (LSI). I artikel tre presenteras en indexeringsmodell för webben baserad på Fuzzy Interrelations (FIS). Artiklarna i temat "Indexering" presenterar olika perspektiv på ämnet. I artikel ett tar man upp en historisk textanalys medan andra artiklar behandlar webben och informationsåtervinning med hjälp av indexering.

### **Kluster 12. Forskningsstema "Kunskapsorganisation"**

1. **FF:** Pan S; Pan G; Hsieh MH **TI:** A dual-level analysis of the capability development process: A case study of TT&T **ÄO:** RESOURCE-BASED VIEW; DYNAMIC CAPABILITIES; COMPETITIVE ADVANTAGE; CALL CENTERS; STRATEGY; FIRM; ORGANIZATIONS; MANAGEMENT; EVOLUTION
2. **FF:** Nunes MB; Annansingh F; Eaglestone B; Wakefield R **TI:** Knowledge management issues in knowledge-intensive SMEs **ÄO:** TACIT KNOWLEDGE; INNOVATION; CAPABILITY; FIRM
3. **FF:** Cheung CMK; Lee MKO **TI:** Understanding consumer trust in Internet shopping: A multidisciplinary approach **ÄO:** ELECTRONIC COMMERCE; ONLINE; MODEL; PRIVACY; TRUSTWORTHINESS; TECHNOLOGY; ACCEPTANCE; VARIABLES; SECURITY; IMPACT
4. **FF:** Teo HH; Wang XW; Wei KK; Sia CL; Lee MKO **TI:** Organizational learning capacity and attitude toward complex technological innovations: An empirical study **ÄO:** CONFIRMATORY FACTOR-ANALYSIS; INFORMATION-SYSTEMS; KNOWLEDGE INTEGRATION; MARKET ORIENTATION; PERSPECTIVE; BUSINESS; ADOPTION; FIRM; ENVIRONMENTS; ASSIMILATION
5. **FF:** Kankanhalli A; Tan BCY; Wei KK **TI:** Understanding seeking from electronic knowledge repositories: An empirical study **ÄO:** INFORMATION-TECHNOLOGY; MANAGEMENT-SYSTEMS; USER PARTICIPATION; COMMUNITIES; MODEL; FIELD; DETERMINANTS; SATISFACTION; PERFORMANCE; ACCEPTANCE
6. **FF:** Desouza KC; Awazu Y **TI:** Maintaining knowledge management systems: A strategic imperative **ÄO:** saknas
7. **FF:** Awazu Y; Desouza KC **TI:** Open knowledge management: Lessons from the open source revolution **ÄO:** OPEN SOURCE SOFTWARE; INNOVATION; SCIENCE

"Kunskapsorganisation" är forskningsstemat för kluster 12. Majoriteten av artiklarna innehåller olika infallsvinklar på ämnet. På de flesta artiklar är det tydligt, sett till titel och ämnesord, att de handlar om organisation av kunskap. Detta tema har likheter med forskningsstema "Kunskapshantering" (kluster 10), i bägge teman finns "Knowledge management" representerat i flera artiklar. En del artiklar i temat "Kunskapsorganisation" behandlar organisation av kunskap inom företag medan andra är mer inriktade på temat ur en teknisk aspekt. Artikel tre faller dock utanför vår kategorisering. Efter att ha studerat artikel tres abstract kan vi konstatera att artikeln handlar om hur man kan undersöka vilket förtroende kunden har för försäljare på Internet efter att ha gjort inköp via webben.

### **Kluster 13. Forskningstema “Hermeneutiska perspektiv på biblioteks- och informationsvetenskap”**

1. **FF:** Andersen J **TI:** The public sphere and discursive activities: information literacy as sociopolitical skills **ÄO:** RETRIEVAL
2. **FF:** Hjørland B; Pedersen KN **TI:** A substantive theory of classification for information retrieval **ÄO:** SCIENCE; PERSPECTIVE; SEMANTICS
3. **FF:** Sundin O; Johannisson J **TI:** Pragmatism neo-pragmatism and sociocultural theory - Communicative participation as a perspective in LIS **ÄO:** INFORMATION-SCIENCE; PROFESSIONALISM; LIBRARY
4. **FF:** Hansson J **TI:** Hermeneutics as a bridge between the modern and the postmodern in library and information science **ÄO:** TEXT; LIS
5. **FF:** Fonseca FT; Martin JE **TI:** Toward an alternative notion of information systems ontologies: Information engineering as a hermeneutic enterprise **ÄO:** SCIENCE
6. **FF:** McKenzie PJ **TI:** Positioning theory and the negotiation of information needs in a clinical midwifery setting **ÄO:** MIDWIVES; COMMUNICATION; SEEKING; PATIENT; MODEL; CARE; EXPERIENCES; PERSPECTIVE; STRATEGIES; CHILDBIRTH

Dessa artiklar tar upp olika aspekter av informationsvetenskap utifrån ett hermeneutiskt perspektiv<sup>15</sup>. Man har bland annat studerat hur viktigt det är att i dagens samhälle ha kunskap om hur man söker information. Att behärska informationssökning är lika viktigt som att kunna läsa och skriva enligt artikel etts författare. Klustret innehåller även artiklar om hur man tar till sig information utifrån de socio- kulturella förutsättningar man har. Artikel sex har studerat en särskild yrkesgrupp, barnmorskor, och deras förmåga att söka och förmedla information inom sin yrkesroll.

### **Kluster 14. Forskningstema “Forskares informationssökningsbeteende”**

1. **FF:** Blake C; Pratt W **TI:** Collaborative information synthesis I: A model of information behaviors of scientists in medicine and public health **ÄO:** RANDOMIZED-TRIALS; CONSORT STATEMENT; SEEKING PATTERNS; CLINICAL-TRIALS; METAANALYSIS; RETRIEVAL; INCLUSION; QUALITY
2. **FF:** Palmer CL **TI:** Scholarly work and the shaping of digital access **ÄO:** INFORMATION-SEEKING BEHAVIOR; UNDISCOVERED PUBLIC KNOWLEDGE; COMMUNICATION; HUMANITIES; PATTERNS; NEUROSCIENCE; MAGNESIUM; DISEASE; BRAIN
3. **FF:** Tenopir C; King DW; Boyce P; Grayson M; Paulson KL **TI:** Relying on electronic journals: Reading patterns of astronomers **ÄO:** SCHOLARLY COMMUNICATION; PHYSICISTS; CONTINUUM; BEHAVIOR
4. **FF:** Borgman CL; Smart LJ; Millwood KA; Finley JR; Champeny L; Gilliland AJ; Leazer GH **TI:** Comparing faculty information seeking in teaching and research: Implications for the design of digital libraries **ÄO:** EARTH PROTOTYPE; SEARCH PROCESS; BIODIVERSITY; UNIVERSITY; MODEL
5. **FF:** Selden L **TI:** On grounded theory - with some malice **ÄO:** saknas

---

<sup>15</sup> Läran om texttolkning och mer utvidgat en vetenskapsteoretisk riktning där betydelsen av förståelse och inlevelse framhävs i human- och samhällsvetenskaperna (Nationalencyklopedin 2007)

**6. FF:** Yitzhaki M; Hammershlag G **TI:** Accessibility and use of information sources among computer scientists and software engineers in Israel: Academy versus industry **ÄO:** D LABORATORIES; COMMUNICATION PATTERNS; SEEKING BEHAVIOR; NETWORKS; TASK

**7. FF:** Kling R; Spector LB; Fortuna J **TI:** The real stakes of virtual publishing: The transformation of E-biomed into PubMed central **ÄO:** PREPRINT ARCHIVE; COMMUNICATION; PLAN; NIH; JOURNALS; VARMUS; RESEARCHERS; PROPOSAL; INTERNET; SCIENCE

Artiklarna i kluster 14 handlar om hur forskare inom olika vetenskaper söker och tillgodogör sig information inom sitt område. Ett flertal av artikelförfattarna tar upp digital information. En av artiklarna jämför tillgänglighet och användandet av informationskällor mellan forskare inom datorvetenskap och utvecklare av mjukvara i Israel. Detta för att se skillnader mellan den akademiska världen jämfört med den industriella.

### **Kluster 15. Forskningstema ”Användares attityder gentemot ny informationsteknik”**

**1. FF:** Vishwanath A; Chen H **TI:** Technology clusters: Using multidimensional scaling to evaluate and structure technology clusters **ÄO:** COMMUNICATION; ADOPTION

**2. FF:** Kwon N; Onwuegbuzie AJ **TI:** Modeling the factors affecting individuals' use of community networks: A theoretical explanation of community-based information and communication technology use **ÄO:** PLANNED BEHAVIOR; PERCEIVED USEFULNESS; SELF-EFFICACY; ACCEPTANCE; USAGE; EASE; WEB; INTERNET; SITES

**3. FF:** Vishwanath A **TI:** Impact of personality on technology adoption: An empirical IVIodel **ÄO:** INNOVATIVENESS

Vårt sista forskningstema handlar om hur ny teknik och informationssystem accepteras i relation till användares attityder och förkunskaper. Artiklarnas författare angriper detta utifrån olika infallsvinklar. I artikel ett undersöker man två olika användargrupper, så kallade adopters och non-adopters. I artikel två utgår man ifrån en beteendevetenskaplig modell för att undersöka en homogen grupp människor i ett och samma samhälle. I artikel tre, som har gemensam författare med artikel ett har man undersökt hur personligheten spelar in vid tillämpning av ny teknik.

## 7 Analys och diskussion

Under analys och diskussion resonerar vi kring de resultat som vi har fått fram samt belyser vår undersöknings resultat relaterat till tidigare forskning. I texten gör vi vissa antaganden om att de olika forskningsteman är besläktade med varandra. Dessa antaganden bygger vi på artiklarnas titlar och ämnesord samt vår kunskap om ämnena.

Vi redovisar en tabell med de olika forskningsteman och dess samhörighet indelat i fyra olika huvudinriktningar, för att på så sätt visualisera den forskningsfront inom informationsvetenskap vi genom vår undersökning kommit fram till. Varje huvudinriktning presenteras sedan för sig. I presentationerna för varje huvudinriktning redovisar vi en tabell med antal artiklar för varje forskningstema samt totalt antal artiklar för huvudinriktningen. Syftet med att redovisa antal artiklar inom varje huvudinriktning är att kunna se vilken inriktning som är störst sett till antal artiklar. Vi anser att storleken på huvudinriktningarna kan visa på deras betydelse inom forskningsfronten.

### 7.1 Forskningsteman och huvudinriktningar

I föregående kapitel (7) presenterades resultatet av vår undersökning. Vi fick med hjälp av metoderna bibliografisk koppling i kombination med klusteranalys fram femton forskningsteman som tar upp olika aspekter inom forskningsområdet informationsvetenskap. De forskningsteman vi fick fram var följande (antal artiklar för varje tema angivna inom parentes):

- Informationsåtervinning (111)
- Webometri (35)
- Bibliometri (39)
- Informationssökning (37)
- Vetenskaplig kommunikation (3)
- Informationsvetenskapens filosofi (5)
- Utvärdering av vetenskapliga tidskrifter (6)
- Textkategorisering (8)
- Användandet av elektroniska tidskrifter (4)
- Kunskapshantering (3)
- Indexering (5)
- Kunskapsorganisation (7)
- Hermeneutiska perspektiv på biblioteks- och informationsvetenskap (6)
- Forskares informationssökningsbeteende (7)
- Användares attityder gentemot ny informationsteknik (3)

Ovanstående forskningsteman representerar forskningsfronten inom informationsvetenskap utifrån underlaget för vår undersökning. Storleken på de olika forskningsteman varierar med avseende på antal artiklar.

Vi tycker oss se fyra huvudinriktningar bland de olika forskningsteman som vi har fått fram i vår undersökning. Dessa fyra huvudinriktningar är **Informationsåtervinning**, **Bibliometri**, **Kunskapsorganisation** och **Informationsvetenskapens filosofi**. Vi har återanvänt begreppen från de forskningsteman inom varje huvudinriktning som innehöll flest antal artiklar. Samtidigt är det de termer som vi uppfattar som mest övergripande och bäst beskriver inriktningarna. Undantag från ovanstående gäller huvudinriktningen

**Informationsvetenskapens filosofi** där vi tyckte att det minsta temat var det bäst beskrivande för huvudinriktningen i stort. Vi har valt att presentera huvudinriktningarna och dess tillhörande forskningstema i tabell 3:

| <b>Informationsåtervinning</b>   | <b>Bibliometri</b>   | <b>Kunskapsorganisation</b>   | <b>Informationsvetenskapens filosofi</b>   |
|--|--|---|--|
| <i>"Informationsåtervinning"</i><br><i>"Informationssökning"</i><br><i>"Textkategorisering"</i><br><i>"Forskares informationssökningsbeteende"</i><br><i>"Indexering"</i><br><i>"Användandet av e-tidskrifter"</i> | <i>"Bibliometri"</i><br><i>"Webometri"</i><br><i>"Utvärdering av vetenskapliga tidskrifter"</i><br><i>"Vetenskaplig kommunikation"</i> | <i>"Kunskapsorganisation"</i><br><i>"Kunskapshantering"</i><br><i>"Användares attityder gentemot ny informationsteknik"</i> | <i>"Hermeneutiska perspektiv på biblioteks- och informationsvetenskap"</i><br><i>"Informationsvetenskapens filosofi"</i> |

**Tabell 3.** Tabell över forskningsfrontens fyra huvudinriktningar med tillhörande forskningsteman.

Våra antaganden att de olika forskningsteman är besläktade bygger på artiklarnas titlar och ämnesord samt vår kunskap om ämnena. Sett till ett större perspektiv är det egentligen inga skarpt tydliga gränser emellan de fyra huvudinriktningarna, vi har ändå, för att tydliggöra vår resultatanalys, valt att göra denna indelning.

### 7.1.1 Informationsåtervinning

| <b>Forskningstema</b>                    | <b>Antal artiklar</b> |
|--|-----------------------|
| "Informationsåtervinning"                | 111                   |
| "Informationssökning"                    | 37                    |
| "Textkategorisering"                     | 8                     |
| "Forskares informationssökningsbeteende" | 7                     |
| "Indexering"                             | 5                     |
| "Användandet av e-tidskrifter"           | 4                     |
| <i>Totalt antal artiklar</i>             | <b>172</b>            |

**Tabell 4.** Antal artiklar inom huvudinriktning **Informationsåtervinning**

**Informationsåtervinning** är en av de huvudinriktningar vi fått fram som en beståndsdel av forskningsfronten inom informationsvetenskap. Hur vi kommit fram till denna huvudinriktning samt dess tillhörande forskningsteman redogör vi för i detta avsnitt. Forskningstemat "Informationsåtervinning" var det klart största forskningstemat sett till antal artiklar som genererades av klustringsprocessen. Ett närbesläktat forskningstema till "Informationsåtervinning" är "Informationssökning", även detta ett av de större temana, dock bara en tredjedel så stort jämfört med forskningstema "Informationsåtervinning". Ser man till artiklarnas ämnesord inom de båda forskningsteman förekommer ofta samma. Vid närmare studie av titlar, ämnesord och i vissa fall abstract tycker vi oss se att vissa av artiklarna skulle kunna passa lika bra, eller i vissa fall bättre, inom det andra forskningstemat och vice versa.

Forsknings temat vi valt att benämna ”Forskarens informationssökningsbeteende” innehåller även dessa artiklar med liknande ämnesord som förekommer i de båda tidigare nämnda forskningsteman. Exempel på sådana ämnesord är retrieval samt information-seeking behavior. ”Indexering” och ”Textkategorisering” är forskningsteman man arbetar med inom inriktningen **Informationsåtervinning**. Att indexera och/eller kategorisera text görs bland annat för att bättre kunna återvinna information vid sökningar. Forsknings temat ”Användandet av elektroniska tidskrifter” hade vi något svårt att kategorisera samt benämna. I tre av artiklarna medverkar samma författare och dessa behandlar främst statistik över användning av e-tidskrifter. I den återstående fjärde artikeln tar man istället upp gränssnittets betydelse för användandet av e-tidskrifter. Det vi kunde se som gemensamt fokus för temats artiklar var användandet av e-tidskrifter. Vi har valt att placera forsknings temat ”Användandet av e-tidskrifter” inom inriktningen **Informationsåtervinning** för att temats artiklar behandlar ämnen som gränssnitt och informationsbehov.

### 7.1.2 Bibliometri

| Forsknings tema                            | Antal artiklar |
|--|----------------|
| ”Bibliometri”                              | 39             |
| ”Webometri”                                | 35             |
| ”Utvärdering av vetenskapliga tidskrifter” | 6              |
| ”Vetenskaplig kommunikation”               | 3              |
| <i>Totalt antal artiklar</i>               | <b>83</b>      |

**Tabell 5.** Antal artiklar inom huvudinriktning **Bibliometri**

Inom huvudinriktning **Bibliometri** finner vi två av de största forskningsteman; ”Webometri” och ”Bibliometri”. Dessa båda teman är besläktade på så vis att man i vissa fall använder sig av samma mät- och analysmetoder. Skillnaden, förenklat sett, mellan webometri och bibliometri är att webometri enbart är inriktad på webben medan bibliometri behandlar alla olika sorters publikationer. Sett till ämnesord i artiklarna inom de båda sist nämnda forskningsteman förekommer exempelvis citeringsanalys, impact factor och olika matematiska lagar i bägge. Anmärkningsvärt är att det bland ämnesorden av vårt urval av artiklar till forsknings temat ”Webometri” endast finns en enda artikel med webometrics som ämnesord. Att det ändå är webometri som blir den samlande termen för att beskriva forsknings temat beror på att flertalet artiklar, vid närmre studier av titlar samt abstract, i någon form tar upp något av de fyra forskningsområden som enligt Björneborn och Ingwersen (2004) definierar webometrisk forskning<sup>16</sup>. Andra gemensamma nämnare för temana ”Bibliometri” samt ”Webometri” får vi när vi närmre studerar författarfältet. Vissa författare förekommer som upphovsmän till artiklar inom bägge forskningsteman, som exempel från vårt urval kan nämnas Leydesdorff.

I forsknings temat ”Utvärdering av vetenskapliga tidskrifter” används i många av artiklarna bibliometriska metoder som citeringsanalys för att utvärdera vetenskapliga tidskrifter; dess betydelse samt produktivitet. Därför har vi valt att placera forsknings temat inom huvudinriktningen **Bibliometri**. Det återstående forsknings temat ”Vetenskaplig kommunikation” är mer svårplacerat. Detta forsknings tema är litet och består endast av tre artiklar. Två av artiklarna beskriver vetenskaplig kommunikation ur

<sup>16</sup> Se kapitelavsnitt 6.1 s. 27

ett teoretiskt perspektiv medan i den tredje artikeln har man undersökt vetenskaplig kommunikation ur ett socialt perspektiv med hjälp av citeringsanalys. Kanske borde vetenskaplig kommunikation ha blivit en egen huvudinriktning, frågan är om det är starkt förankrat inom informationsvetenskaplig forskning. Till exempel påpekar White och McCain (1998, s. 337) att författare som skriver inom området vetenskaplig kommunikation har starka identiteter inom discipliner utanför informationsvetenskap. Vi har ändå valt att placera forskningstema "Vetenskaplig kommunikation" inom huvudinriktningen **Bibliometri** eftersom det undersöks med bibliometriska metoder.

### 7.1.3 Kunskapsorganisation

| Forskningstema  | Antal artiklar |
|---|----------------|
| "Kunskapsorganisation"                                | 7              |
| "Kunskapshantering"                                   | 3              |
| "Användares attityder gentemot ny informationsteknik" | 3              |
| <i>Totalt antal artiklar</i>                          | <b>13</b>      |

**Tabell 6.** Antal artiklar inom huvudinriktning **Kunskapsorganisation**

Inom huvudinriktning **Kunskapsorganisation** finner vi de båda närbesläktade forskningsteman "Kunskapsorganisation" och "Kunskapshantering". I flera av artiklarna inom dessa två forskningsteman används termen *knowledge management*. Det förekommer också artiklar med samma upphovsmän i de båda forskningsteman. Forskningstemat "Användares attityder gentemot ny informationsteknik" har vi placerat under denna huvudinriktning då artiklarna i forskningstemat beskriver hur man, utifrån den kunskap man har, accepterar ny teknik och informationssystem.

### 7.1.4 Informationsvetenskapens filosofi

| Forskningstema  | Antal artiklar |
|---|----------------|
| "Hermeneutiska perspektiv på biblioteks- och informationsvetenskap" | 6              |
| "Informationsvetenskapens filosofi"                                 | 5              |
| <i>Totalt antal artiklar</i>  | <b>11</b>      |

**Tabell 7.** Antal artiklar inom huvudinriktning **Informationsvetenskapens filosofi**

Bägge forskningsteman inom huvudinriktningen **Informationsvetenskapens filosofi** har artiklar som behandlar informationsvetenskap (i det ena fallet biblioteks- och informationsvetenskap) ur ett mer teoretiskt perspektiv. Forskningstemat "Informationsvetenskap som forskningsområde" sätter informationsvetenskap i relation till andra discipliner. Forskningstemat "Hermeneutiska perspektiv på biblioteks- och informationsvetenskap" innehåller artiklar där man teoretiserar kring tolkningen av fenomen inom biblioteks- och informationsvetenskap. Inom huvudinriktning **Informationsvetenskapens filosofi** har artiklarna inom de båda forskningsteman ibland samma ämnesord, exempelvis science, text och retrieval.

### 7.1.5 Sammanställning av huvudinriktningar

| Huvudinriktning       | Informations-<br>återvinning | Bibliometri | Kunskaps-<br>organisation | Informationsvetenskapens<br>filosofi |
|-----------------------|------------------------------|-------------|---------------------------|--------------------------------------|
| <i>Antal artiklar</i> | 172                          | 83          | 13                        | 11                                   |

Tabell 8. Antal artiklar per huvudinriktning.

Vi har delat in forskningsfronten inom informationsvetenskap i fyra huvudinriktningar baserade på vår undersöknings resultat. Utgår vi från antal artiklar inom varje huvudinriktning så ser man att huvudinriktningen **Informationsåtervinning** är den ojämförligt största. Efter **Informationsåtervinning** följer i storleksordning **Bibliometri**, **Kunskapsorganisation** samt **Informationsvetenskapens filosofi**. Det är olika antal forskningsteman inom de olika huvudinriktningarna och inom huvudinriktningarna varierar det i storlek baserat på antal artiklar mellan de olika forskningsteman. Oavsett detta så är **Informationsåtervinning** den klart största huvudinriktningen beroende på att forskningstema "Informationsåtervinning" är det ojämförligt största forskningstemat innehållande 111 artiklar. Dessutom ingår i huvudinriktningen även forskningstema "Informationssökning" som är det tredje största forskningstemat med sina 37 artiklar. Inom huvudinriktning **Bibliometri** finner vi två andra stora forskningsteman; "Bibliometri"(näst störst, 39 artiklar) samt "Webometri" (35 artiklar). Dessa två forskningsteman utgör majoriteten av huvudinriktningen, sett till antal artiklar. Huvudinriktningarna **Kunskapsorganisation** samt **Informationsvetenskapens filosofi** innehåller tre respektive två forskningsteman. Ingen av forskningsteman inom dessa två huvudinriktningar är särskilt stora, de största består av sju artiklar ("Kunskapsorganisation" samt "Hermeneutiska perspektiv på biblioteks- och informationsvetenskap").

Detta resultat ger slutsatsen att den huvudinriktning som innehåller flest artiklar; **Informationsåtervinning**, har haft störst del av forskningsfronten inom forskningsområdet informationsvetenskap under de aktuella åren som undersökts (2004-2006). **Bibliometri** har också en stor del, men ändå knappt hälften så stor som **Informationsåtervinning**. **Kunskapsorganisation** och **Informationsvetenskapens filosofi** har jämfört med de två förstnämnda en relativt liten del.

### 7.2 Forskningsteman och tidigare forskning

Här kommer vi att diskutera våra resultat i relation till tidigare studier som tagits upp i avsnittet med tidigare forskning. Då vår metod skiljer sig från de flesta av de studier som tagits upp kan det vara svårt att jämföra och dra slutsatser utifrån de tidigare studierna jämfört med vårt resultat. Materialet för vår undersökning skiljer sig också jämfört med tidigare studier avseende exempelvis antal artiklar, tidskrifter och tidsperioder som använts. Fokus på vårt material jämfört med tidigare forskning kommer därmed att vara vilka forskningsområden som har framkommit i de olika undersökningarna.

De studier vars resultat mest liknar vårt eget är Perssons *The Intellectual Base and Research Fronts of JASIS 1986-1990* från 1994, Van den Besselaars och Heimeriks *Mapping research topics using word-reference co-occurrences: A method and exploratory case study* från 2006 samt White och McCains resultat i artikeln *Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972–1995* från 1998.

De likheter vi kan se jämfört med Perssons studie är att de huvudinriktningar inom forskningsfronten inom informationsvetenskap som vi fått fram har likheter med Perssons huvudområden. Där vi fick fram **Informationsåtervinning**, **Bibliometri**, **Kunskapsorganisation** och **Informationsvetenskapens filosofi**, varav informationsåtervinning och bibliometri var de klart största huvudinriktningarna, fick Persson fram två huvudämnesområden, IR och bibliometri. Liksom i våra resultat var IR-gruppen dubbelt så stor som bibliometrigruppen.

Van den Besselaars och Heimeriks kom fram till två huvudgrupper i sin undersökning av forskningsfronten inom informationsvetenskap: informationsåtervinning och scientometri. Jämfört med vår undersökning så fick även vi fram **Informationsåtervinning** och **Bibliometri**, som är en annan term för scientometri, som två huvudinriktningar. Även de undergrupper som ingår i de två huvudgrupperna i Van den Besselaars och Heimericks studie stämmer ganska bra överens med våra resultat. Inom huvudgruppen informationsåtervinning i Van den Besselaars och Heimeriks studie ingår generell informationsåtervinning, informationssökning, kataloger och index samt sociala och politiska aspekter på informationssamhället. Huvudgruppen scientometri bestod av artiklar om indikatorer och kartläggning, scientometriska fördelningar, utvärdering av vetenskap och metoder för kvantitativa analyser. I vår huvudinriktning **Informationsåtervinning** ingår bland annat forskningsteman "Informationsåtervinning", "Informationssökning" och "Indexering". Inom vår huvudinriktning **Bibliometri** ingår forskningsteman "Bibliometri", "Webometri", "Utvärdering av vetenskapliga tidskrifter" och "Vetenskaplig kommunikation".

Även i White och McCains undersökning var de två stora ämnesinriktningarna: domänanalys, innehållande bibliometri och vetenskaplig kommunikation, och informationsåtervinning. Vad gäller de faktorer som framkom i deras undersökning så finns vissa gemensamma beröringspunkter med våra forskningsteman, till exempel så har även vi valt benämningar som "Bibliometri" samt "Vetenskaplig kommunikation".

De undersökningar vars resultat minst liknar vårt är White och Griffiths *Author Cocitation: A Literature Measure of Intellectual Structure* från 1981 samt Janssens et al. *Towards mapping library and information science* från 2006. Även dessa två studier får fram liknande resultat med IR och bibliometri som framträdande forskningsområden men den struktur på informationsvetenskap som de får fram stämmer inte överens med vår.

En orsak till att vi får liknande resultat som Persson kan vara att vi har använt samma metod som honom, bibliografisk koppling och klusteranalys, för att kartlägga forskningsfronten inom informationsvetenskap. Vi har även använt samma tidskrift som Persson: JASIS(T) men dock inte enbart den tidskriften. Trots att Van den Besselaars och Heimericks använder en annan metod för att studera förhållandet mellan dokument och citerade dokument än oss så överensstämmer resultaten bra. En orsak till detta kan

vara att deras studie tidsmässigt ligger nära vår studie och att JASIST och JDOC ingår i deras undersökningsmaterial, de har dock ytterligare sex tidskrifter med i sin undersökning.

White och McCain har använt sig av cociteringsanalys och de undersöker informationsvetenskapens intellektuella bas till skillnad från vår forskningsfront. Det som är anmärkningsvärt är att vid jämförelse så liknar vårt resultat deras trots skillnader vad gäller metod och inriktning på forskningsfront respektive intellektuell bas. Persson menar att hans forskningsfront speglar den intellektuella basen som han får fram i sin studie och hänvisar även till likheter med White och McCains studie. Därmed kan det vara så att i grova drag så stämmer forskningsfronten inom informationsvetenskap överens med den intellektuella basen inom densamma. Förändringarna kanske sker inom mindre ämnesområden, där nya tillkommer och andra försvinner, över tid men de dominerande områdena som exempelvis IR och bibliometri består.

Orsakerna till att våra resultat skiljer sig något mer ifrån de resultat som White och Griffith presenterar kan bero på skillnader i urval av undersökningsmaterial. White och Griffith gjorde i sin undersökning inte ett urval från tidskrifter, vilket man gjort i alla de övriga presenterade studierna, utan valde istället ut specifika författare de ansåg väl representerade forskningsområdet informationsvetenskap. Deras undersökning är också den äldsta, från 1981, av de undersökningar vi studerat närmre och det kan också utgöra en faktor till det avvikande resultatet.

Det är svårt att jämföra vår undersökning med den undersökning Janssens et al. gjort. Visserligen finns det likheter som att två av tidskrifterna som används i deras studie är JDOC och JASIST, vilka även vi använder. Det är stor skillnad på val av metod mellan undersökningarna då man i Janssens et al. studie inte alls använder bibliografiska uppgifter eller citeringar utan istället grupperar dokumenten utifrån textinnehåll, detta kan vara en bidragande faktor till de annorlunda resultaten i jämförelse med våra egna.

### **7.3 Reflektioner över urval samt tillämpning av metod**

Vi har nu undersökt och beskrivit forskningsfronten inom informationsvetenskap utifrån forskningsteman som vi identifierat genom användning av bibliografisk koppling i kombination med klusteranalys.

Vårt material från tidskrifterna JDOC och JASIST under åren 2004-2006 var omfattande och de forskningsteman som vi fick fram varierade markant i storlek samt var emellanåt svårtolkade på grund av varierande innehåll. Dessa problem hade kunnat reduceras om vi valt att sätta ett högre minimivärde, till exempel fyra, för antal gemensamma referenser vid den bibliografiska kopplingen. Vi satte värdet till minst två gemensamma referenser.

Stundtals har det varit svårt att definiera forskningsområdet informationsvetenskap då en del av artiklarna i resultatet även handlat om biblioteks- och informationsvetenskap. Möjligtvis hade detta kunnat undvikas om vi valt andra tidskrifter. I tidskrifterna JDOC och JASIST behandlar man främst informationsvetenskap, men i JDOC tar man även upp biblioteksvetenskap. I en av undersökningarna, Janssens et al. 2006, som tas upp i kapitlet om tidigare forskning studerar man vilka forskningsfrågor som är aktuella inom just biblioteks- och informationsvetenskap. Även White och McCain (1998, s. 328) diskuterar informationsvetenskap i förhållande till biblioteks- och

informationsvetenskap. Det kan vara en svår gräns att dra mellan båda dessa ämnen. Vi valde att använda oss av benämningen informationsvetenskap som en bredare term vilken inkluderar biblioteks- och informationsvetenskap.

Angående tidsperioden hade vårt resultat kanske blivit annorlunda om vi använt ett längre tidsintervall. Jämfört med tidigare studier varierar de tidsperioder som använts i studierna av just forskningsfronten inom informationsvetenskap med vårt tidsintervall. Tidsperioden i Perssons (1994) undersökning av forskningsfronten var 1986-1990 medan i Van den Besselaar och Heimeriks (2006) studie var tidsperioden ett år. Vi valde att avgränsa vår undersökning till tre år eftersom vi inte skulle få ett för stort underlag men ändå en längre tidsperiod samt försöka få ett tillräckligt jämnt material över tid.

Det resultat vi har fått speglar givetvis de tidskrifter vi valt. Resultatet hade möjligen kunnat se annorlunda ut om vi valt andra tidskrifter. Att Informationsåtervinning samt Bibliometri i vår undersöknings resultat tar upp stora delar av forskningsfronten inom informationsvetenskap kan bero på att det är områden som är dominerande i JASIST. JASIST är sett till antal artiklar större än JDOC, 363 jämfört med 92 i vårt undersökningsunderlag, vilket kan bidra till att forskningsfrontens tyngdpunkt främst förlagts till de områden som det skrivs om mest i JASIST. Dock visar det sig vid en närmare studie av kluster ett ("Informationsåtervinning") ändå att klustret består av en relativt jämn fördelning av artiklar både från JASIST och JDOC, vilket tyder på att vi ändå valt tidskrifter som kompletterar varandra sett till att täcka in forskningsområdet informationsvetenskap.

Vad gäller tillämpning av metod är det viktigt att vara medveten om och ordentligt inläst på vilken inverkan olika metoder har på slutresultatet innan man genomför själva undersökningen. Vi valde klustringsmetoden SLINK vilket genererar stora samt något "brusiga" kluster, vilket fungerar bra om man som vi studerar en forskningsfront, vad det forskas kring "just nu". I ett annat fall då man med klustringsmetoder vill studera ett område mer ingående för att fånga alla dess beståndsdelar, även mindre, kan till exempel complete link clustering vara lämpligare. Complete link clustering genererar till skillnad från SLINK små och väl sammansatta kluster.

När vi gjorde de bibliografiska kopplingarna i Bibexcel försvann ett stort antal (158) poster. Den främsta anledningen till att så många poster försvann var att vi olyckligtvis missat att standardisera referenssträngarna. Ett problem som finns med referensförteckningar i ett citeringsindex som SSCI är att referenssträngarna inte alltid är likvärdiga då det gäller författarnamn och tidskriftstitlar. Exempelvis kan en författares hela förnamn skrivas ut, men i andra poster kanske bara initialerna finns med, författare med flera efternamn kan skrivas på olika sätt i olika strängar. En standardisering av referenssträngarna skall därmed göras för att inte poster ska falla bort "i onödan" vid den bibliografiska kopplingen. För vår del innebar missen att standardisera referenssträngarna att närmre en tredjedel av alla våra poster föll bort vilket säkert haft en stor inverkan på vårt slutgiltiga resultat. Bland annat hade förmodligen klustrens storlek samt innehåll påverkats vilket i sin tur skulle kunna ha gett de olika huvudinriktningarna annan betydelse jämfört med de resultat vi fick fram.

En fortsättning på vår studie skulle kunna vara att rådfråga en eller flera experter om hur vårt resultat speglar aktuell forskning inom ämnet informationsvetenskap. Vidare kan

det vara intressant att se hur vårt resultat eventuellt tillsammans med övriga liknande studier speglar de utbildningar som finns inom informationsvetenskap. Något som vi reflekterat över när vi har sett resultatet är att bibliometri upptar en stor del av forskningsfronten men då vi har studerat informationsvetenskap under utbildningen var detta ämne sparsamt representerat.

## 8 Sammanfattning

Syftet med denna uppsats var att undersöka och beskriva forskningsfronten inom ämnesområdet informationsvetenskap. För att uppnå syftet tog vi upp följande forskningsfråga:

- Vilka forskningsteman kan identifieras inom ämnesområdet informationsvetenskap genom användning av bibliografisk koppling i kombination med klusteranalys?

Materialet till undersökningen begränsades till artiklar i tidskrifterna JDOC och JASIST från åren 2004-2006.

Den teoretiska bakgrunden består av bibliometri och multivariat analys samt en kortfattad beskrivning av vetenskaplig kommunikation och citeringar. Vi beskriver bibliometriska arbets- och datamaterial som citeringsindex samt olika citeringsmått.

I avsnittet om tidigare forskning tar vi upp studier som beskriver ämnesområdet informationsvetenskap med hjälp av bibliometriska metoder.

Den data vi behövde för att besvara vår forskningsfråga samlades in från citeringsdatabasen SCIE. Vi fick då fram 455 artiklar från JDOC och JASIST. Vid bearbetning av dessa artiklars bibliografiska poster i programmet Bibexcel fick vi fram 297 artiklar med gemensamma referenser som bildade 918 kopplingar. Detta material utgjorde basen för undersökningen. Efter en normalisering av dokumentens kopplingsstyrka klustrades materialet i Bibexcel. Vid klustringsprocessen bildades femton kluster av varierande storlek. Vi identifierade olika forskningstema för varje kluster genom att studera ämnesord, titlar och abstracts till varje artikel i de olika klustren.

Under analysen av våra resultat fick vi fram fyra huvudinriktningar: **Informationsåtervinning**, **Bibliometri**, **Kunskapsorganisation** och **Informationsvetenskapens filosofi**, inom forskningsfronten inom informationsvetenskap. Varje huvudinriktning innehöll de olika forskningsteman som vi tidigare identifierat. Huvudinriktningen **Informationsåtervinning** visade sig vara klart störst av dessa då det innehöll flest forskningsteman samt flest antal artiklar. **Bibliometri** var näst störst, följt av de två markant mindre huvudinriktningarna **Kunskapsorganisation** och **Informationsvetenskapens filosofi**.

Vi jämförde vårt resultat med den tidigare forskning som presenterats i uppsatsen. Trots att valet av metod och material skiljde sig åt fanns vissa likheter med tidigare forskning. De största likheterna var att huvudinriktningarna **Informationsåtervinning** och **Bibliometri**, i liknande former och termer, även förekom som stora forskningsområden i vissa av de tidigare studierna.

Uppsatsen avslutas med kommentarer där vi redogör för problem som uppkommit under arbetets gång samt ger förslag till fortsatta studier.

## Källförteckning

Alderderfer, Mark S. & Blashfield, Roger K. (1984), *Cluster analysis*. Beverly Hills, California: SAGE publications.

Baxter, M.J (1994), *Exploratory multivariate analysis in archaeology*. Edinburgh: Edinburgh University Press.

Björneborn, Lennart & Ingwersen, Peter (2004), Toward a basic framework for webometrics. *Journal of the American society for information science and technology*, 55 (14), s. 1216-1227. Även tillgänglig:  
[http://vip.db.dk/lb/papers/bjorneborn\\_&\\_ingwersen\\_2004\\_toward\\_a\\_basic\\_framework.pdf](http://vip.db.dk/lb/papers/bjorneborn_&_ingwersen_2004_toward_a_basic_framework.pdf) [2007-07-20]

Cronin, Blaise (1984). *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham.

Diodato, Virgil (1994). *Dictionary of Bibliometrics*. New York: The Haworth Press.

Egghe, L. & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Amsterdam: Elsevir.

Everitt, B. S., Landau, S. & Leese, M. (2001). *Cluster analysis*. Fourth edition. London: Arnold.

Glänzel, Wolfgang (2003). *Bibliometrics as a research field: A course on theory and application of bibliometric indicators*.  
[http://www.norslis.net/2004/Bib\\_Module\\_KUL.pdf](http://www.norslis.net/2004/Bib_Module_KUL.pdf) [2006-10-02]

Hicks, D. (1987), Limitations of co-citation analysis as a tool for science policy. *Social Studies of Science*, 17, s. 295-316.

Hjerppe, Roland (1978). *An outline of bibliometrics and citation analysis*. Stockholm: Royal Insitute of Technology Library.

Hjörland, B. och Albrechtsen, H. (1995). Toward a new horizon in information-science: Domain-analysis. *Journal of the American Society for Information Science*, 46(6), s. 400-425.

ISI Web of knowledge (2007a), *Journal Citation Reports*.  
<http://portal.isiknowledge.com.lib.costello.pub.hb.se/portal.cgi?DestApp=JCR&Func=Frame> [2007-07-20]

ISI Web of knowledge (2007b), *Web of Knowledge Help*.  
[http://portal.isiknowledge.com.lib.costello.pub.hb.se/help/h\\_info.htm](http://portal.isiknowledge.com.lib.costello.pub.hb.se/help/h_info.htm) [2007-07-29]

Janssens, F., Leta, J., Glänzel, W & De Moor, B. (2006). Towards mapping library and information science. *Information processing & management*, 42(6), s. 1614-1642.

Jarneving, Bo (2005a). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics*, vol. 65, no. 2, s. 245-263.

Jarneving, Bo (2005b) *The combined application of bibliographic coupling and the complete link cluster method in bibliometric science mapping*. Borås: Valfrid.

Journal of Documentation.

<http://www.emeraldinsight.com/info/journals/jd/jourinfo.jsp> [2007-07-29]

Journal of the American Society for Information Science and Technology.

<http://www3.interscience.wiley.com/cgi-bin/jabout/76501873/ProductInformation.html?CRETRY=1&SRETRY=0> [2007-07-29]

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), s. 10-25.

Kruskal, Joseph B. & Wish, Myron (1978). *Multidimensional scaling*. Beverly Hills and London: Sage Publications (Quantitative Applications in the Social Sciences, vol. 11).

Kärki, Riita (2003). Vetenskaplig kommunikation. Ingår i Mäkinen, Ilkka och Sandeqvist, Katja, red. *Introduktion till informationsvetenskapen*. Tammerfors, Tampere University press. S. 189-218.

Kärki, Riitta & Kortelainen, Terttu (1998). *Introduktion till bibliometri*. Helsingfors: Nordinfo.

Körner, Svante & Wahlgren, Lars (2005). *Statistiska metoder*. Lund: Studentlitteratur.

Lunds universitets bibliotek (2002-11).

<http://www.lub.lu.se/ub/distans/infokompetens/vetkomm/vetkommc.html> [2006-10-13]

Mäkinen, Ilkka (2003). Det gick att leva förr också: Informationshistoria. Ingår i Mäkinen, Ilkka och Sandeqvist, Katja, red. *Introduktion till informationsvetenskapen*. Tammerfors, Tampere University press. S. 30-68.

Nationalencyklopedin. [Online]

<http://www.ne.se> [2007-07-21]

Noyons, C. M. (1999). *Bibliometric mapping as a science policy and research management toll*. Leiden: DSWO Press.

Persson, Olle (1991). *Forskning i bibliometrisk belysning*. Umeå: Inum.

Persson, Olle (1994). The Intellectual Base and Research Fronts of JASIS 1986-1990. *Journal of the American society for information science*, 45(1), s. 31-38.

Romesburg, Charles, H. (2004). *Cluster analysis for researchers*. North Carolina: Lulu press.

Small, Henry (1973). Co-citation in the scientific literature: e new measure of the relationship between two documents. *Journal of the American Society for Information Science*. 24(4) s. 265-269. Även tillgänglig:  
<http://www.garfield.library.upenn.edu/essays/v2p028y1974-76.pdf> [2007-08-23]

Thomson Scientific (2007a), *Social Science Citation Index*.  
<http://scientific.thomson.com/cgi-bin/productinfo.pl?product= Social Science Citation Index> [2007-03-04]

Thomson Scientific (2007b), *Glossary of Thomson Scientific terminology*.  
<http://scientific.thomson.com/support/patents/patinf/terms/#C> [2007-07-29]  
<http://scientific.thomson.com/support/patents/patinf/terms/#I> [2007-07-29]

Vakkari, Pertti (2003). Informationsförsörjningen och informationsvetenskapen. Ingår i Mäkinen, Ilkka och Sandeqvist, Katja, red. *Introduktion till informationsvetenskapen*. Tammerfors, Tampere University press. S. 9-29.

Van den Besselaar, Peter & Heimeriks, Gaston (2006). Mapping research topics using word-reference co-occurrences: A method and exploratory case study. *Scientometrics*, vol. 68 no. 3, s. 377-393.

White, Howard D. & Griffith, Belver C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American society for information science*, 32(3), s. 163-171.

White, Howard D. & McCain, Katherine W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American society for information science*, 49(4), s. 327-355.

Zins, Chaim (2006). Redefining information science: from "information science" to "knowledge science". *Journal of Documentation*, 64(4), s. 447-461.