MAGISTERUPPSATS I BIBLIOTEKS- OCH INFORMATIONSVETENSKAP
VID BIBLIOTEKS- OCH INFORMATIONSVETENSKAP/BIBLIOTEKSHÖGSKOLAN
2004:84

# Query expansion med semantiskt relaterade termer

Sofia Höglund

**© Sofia Höglund**
Mångfaldigande och spridande av innehållet i denna uppsats
– helt eller delvis – är förbjudet utan medgivande av författaren/författarna.

| | |
|---|---|
| Svensk titel: | Query expansion med semantiskt relaterade termer |
| Engelsk titel: | Query expansion with semantically related words |
| Författare: | Sofia Höglund |
| Kollegium: | 2 |
| Färdigställt: | 2004 |
| Handledare: | Anders Stenström |

Abstract: The aim of this master's thesis is to examine query expansion. Query expansion is the process of adding new terms to a query to improve the retrieval effectiveness. In this study the baseline query was expanded in five different modes. The first expansion strategy was formulated by the inflected terms (as the system does not allow truncation, I have inflected the terms, e. g. airplane, airplanes, the airplanes, etc.) in the baseline strategy. The second expansion strategy consisted of synonyms from a dictionary of synonyms. For the terms in the remaining three expansion strategies a general thesaurus was used which contained hierarchical and associative term relationships in order to find broader, narrower and related expansion terms. Relative recall and average precision was measured. The average results show that the expansion with inflected terms from the baseline query gave the most effective retrieval, both concerning increase in relative recall and average precision.

Nyckelord: Query expansion, tesaurus, synonymordbok, precision, recall

# Innehåll

1. Inledning....4

2. Syfte och frågeställningar....5

3. Bakgrund....6
3.1 Query expansion....6
3.1.1 Automatisk och interaktiv query expansion....7
3.1.2 Manuell query expansion....8
3.2 Effektivitetsmått....9

4. Tidigare forskning....11
4.1 Sammanfattning och reflektion kring den tidigare forskningen....15

5. Metod....17
5.1 Testmiljö....17
5.2 Topics och relevansbedömningar....18
5.3 Expansionsstrategier....19
5.4 Expansionsverktyg....20
5.4.1 Tesaurus....20
5.4.2 Ordbok och synonymlexikon....20
5.5 Evalueringsmetod....21
5.6 Tillvägagångssätt....21

6. Resultat....24
6.1 Relativ recall....24
6.2 Genomsnittlig precision....25

7. Analys....26
7.1 Konklusion och sammanfattning av analysen....31

8. Sammanfattning....33

9. Referenslitteratur....35

Appendix 1: Topics och sökfrågor....37

Appendix 2: Bortfall....52

Appendix 3: Borttagna termer....55

# 1. Inledning

Denna uppsats undersöker query expansion med semantiskt relaterade termer. Begreppet query expansion (QE) kan översättas med ”expansion av en sökfråga” och är ett delområde inom Information Retrieval (IR). IR handlar om representation, lagring, organisering och åtkomst till information. Syftet med QE är att förbättra återvinningseffektiviteten genom att utöka en initialt formulerad sökfråga på ett sådant sätt att fler relevanta dokument återvinns.

I föreliggande uppsats studeras hur återvinningseffektiviteten påverkas när termer i en initialt ställd sökfråga utökas med semantiskt relaterade termer i fem olika expansionsstrategier. Dessa fem strategier innehåller böjningsformer, synonymer, relaterade termer, snävare termer samt bredare termer i förhållande till termerna i grundstrategin, dvs. den initialt ställda sökfrågan. I denna studie hämtas expansionstermerna från två kollektionsoberoende källor: dels en allmän tesaurus och dels en synonymordbok.

Att göra en mindre undersökning beträffande QE ingick i kursen ”*Information Retrieval som forskningsområde*” vilken genomfördes på c-nivå. Trots att den studien var relativt liten gav den en inblick i hur QE kan utföras vilket var både givande och intressant. Att fortsätta på samma linje men mer utförligt kunna undersöka QE var därför lockande och även passande för en magisteruppsats. Även intresset för att göra en empirisk studie var ytterligare ett skäl för detta ämnesval.

Uppsatsen är upplagd enligt följande: närmast efter denna inledning beskrivs studiens syfte och frågeställningar. Därefter följer ett bakgrundsavsnitt med en presentation av QE och effektivitetsmåtten precision och recall. Syftet med detta är att presentera QE i ett bredare sammanhang och visa de mått som behövs för att kunna mäta och utvärdera återvinningseffektiviteten hos olika strategier för QE. På detta följer några utdrag ur den tidigare forskning som arbetat med liknande frågeställningar. Kapitlen bakgrund och tidigare forskning gör inga anspråk på att vara heltäckande utan avser snarare ge en bild av och förståelse för olika typer av QE samt tidigare undersökningar inom området. Därefter beskrivs uppsatsens metod och även det praktiska tillvägagångssättet. Att presentera uppsatsen och dess delar enligt ovanstående syftar till att ge läsaren en relativt god bakgrundsbild för att kunna förstå resultat- och diskussionskapitlet. Uppsatsen avrundas med en sammanfattning och sist återfinns appendix med samtliga sökfrågor och bortfall.

# 2. Syfte och frågeställningar

Syftet med denna studie är att utvärdera återvinningseffektiviteten hos fem strategier för automatisk QE. Studiens syfte kan uttryckas i följande frågeställningar:

- Hur presterar de fem strategierna avseende relativ recall i förhållande till grundstrategin?
- Hur presterar de fem strategierna avseende genomsnittlig precision i förhållande till grundstrategin?
- Finns det skillnader i återvinningseffektivitet mellan grundstrategin och de fem expansionsstrategierna beträffande genomsnittlig precision och relativ recall?
- Om skillnader finns, vilken strategi presterar bäst och vad kan det bero på?

# 3. Bakgrund

Utgångspunkten för QE är ett antagande att den initiala sökfråga som ställs till systemet fungerar som en ingång, en start, och att denna sökfråga oftast inte till fullo uttrycker det informationsbehov som användaren har. En sökfråga eller en sökformulering är en formell representation av det som eftersöks uttryckt med ett IR-systems språk. Svårigheterna med att uttrycka ett informationsbehov ligger inte bara i att konkretisera sina önskemål utan även i att hitta passande termer. Termerna skall hjälpa till att återvinna relevanta dokument men utelämna irrelevanta sådana vilket i praktiken kan innebära problem. Språkets skiftningar med synonymer, homonymer och så vidare gör att de termer som används i ett dokument eller i en databas kanske inte finns i informationssökarens ordförråd. En kontrollerad vokabulär kan avhjälpa detta men även här finns risk att de termer som tillskrivs ett dokument inte beskriver dess innehåll på ett bra sätt.

Många gånger är det även så att informationssökarens initialt formulerade sökfråga endast innehåller några få termer eller kanske bara en enstaka term.[1] Utifrån en första sökfråga kan således förändringar göras exempelvis genom att man lägger till nya termer, viktar om eller förändrar formen på de termer som ingår. Att på olika sätt expandera sökfrågan gör att chansen att återvinna fler relevanta dokument ökar. Förutsättningen för expansionen är därmed också att användaren önskar fler relevanta dokument. Resultaten för QE styrs av den återvinningsmodell som finns i det system som används. Föreliggande undersökning är utförd i ett probabilistiskt system vilket beskrivs i avsnitt 5.1.

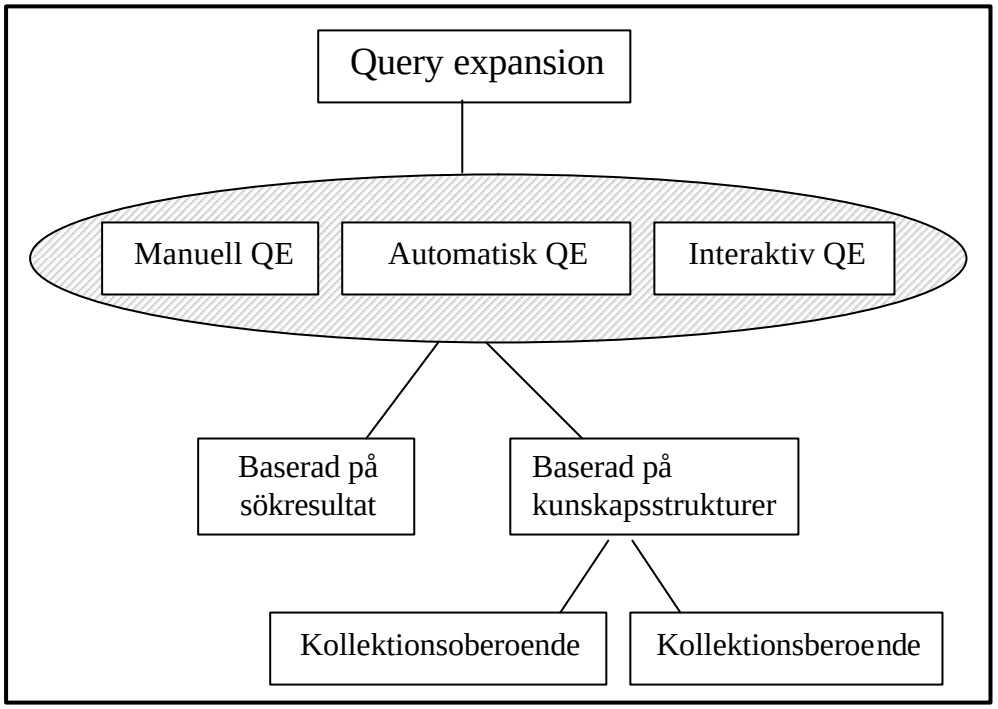
## 3.1 Query expansion

De olika varianterna av QE kan delas in i tre grupper baserat på varifrån expansionstermerna hämtas, vilken metod som används för att välja expansionstermer samt vilken metod som används för att lägga till expansionstermerna (Efthimiadis 1996, s.124). Detta ger en indelning i manuell, automatisk och interaktiv expansion (Fig.1).

Den fortsatta indelningen kan dels baseras på sökresultat och dels på kunskapsstruktur. Den typ av QE som är baserad på kunskapsstrukturer kan delas upp i kollektionsberoende eller kollektionsoberoende kunskapsstrukturer. En kollektionsberoende kunskapsstruktur är bunden till en viss kollektion eller samling. Det kan exempelvis vara en automatiskt skapad tesaurus eller termkluster. En kollektionsoberoende kunskapsstruktur är inte bunden till en viss kollektion utan kan t ex vara en allmän tesaurus eller en synonymordbok.

I den här studien undersöks automatisk[2] QE där expansionstermerna hämtas från kollektionsoberoende källor. För att ge en helhetsbild och en bättre uppfattning av QE kommer även de andra typerna av QE att förklaras kortfattat nedan.

[1] För undersökningar som studerat antal söktermer som användare nyttjat se exempelvis Jansen, Spink, Saracevic 2000 eller Spink et al, 2000.
[2] Expansionen kommer dock att utföras manuellt, för närmare förklaring se avsnitt 5.6




*Figur 1. Metoder och källor för query expansion. (Efter Efthimiadis 1996)*

### 3.1.1 Automatisk och interaktiv query expansion

Automatisk QE innebär att systemet utser expansionstermerna utan att användaren deltar i detta val. Olika typer av algoritmer används som utifrån användarens initiala sökfråga gör en eller flera sökningar för att återvinna fler relevanta dokument. Automatisk QE kan göras enligt metoder baserade på information från de initialt återvunna dokumenten (kallad lokal analys), eller metoder baserade på information från hela kollektionen (kallad global analys) (Baeza-Yates & Ribeiro-Neto 1999, s.117ff). Dessa analyser bygger främst på klustring.

Genom att bilda kluster av termer som samförekommer utgår man från antagandet att termer som återfinns i samma dokument ett flertal gånger, och då har en relativt hög frekvens i dessa dokument, liknar varandra och är bra som expansionstermer.[3] En automatisk lokal analys sker genom att den första sökningen rankas och för varje originalterm skapas ett lokalt associationskluster baserat på samförekomst i dokumenten. Sökfrågan expanderas sedan med de termer som uppvisar högst grad av likhet utifrån dessa kluster (Baeza-Yates & Ribeiro-Neto 1999, s.123ff). Detta förutsätter då att den första sökningen ger ett bra och relevant resultat.

En automatisk global analys använder globalt härledd information, dvs. information från hela kollektionen, för att expandera sökfrågan. En sådan analys kan baseras på en likhetstesaurus eller på en statistisk tesaurus.[4] Ytterligare ett sätt att automatisk expandera är genom användningen av en stamningsalgoritm som bryter ned termerna till ordstammar genom eliminering av suffix och prefix. Algoritmen används för termer både i sökfrågor och dokument. Avsikten är att öka antalet återvunna relevanta dokument genom att även de dokument som innehåller varianter av sökfrågetermen återvinns.

[3] Observera att klustring inte måste vara baserat på samförekomst av termer.
[4] Angående förklaring av likhetstesaurus och statistisk tesaurus se exempelvis: Baeza-Yates & Ribeiro-Neto 1999, s.131-137

Interaktiv QE innebär att användaren interagerar med systemet och får hjälp att utse expansionstermer. Interaktionen består i att systemet tar fram presumtiva expansionstermer vilka användaren tar ställning till. Det kan även vara så att systemet tar fram potentiellt relevanta dokument vilka användaren relevansbedömer varpå en ny sökning görs baserat på denna bedömning. Metoder baserade på relevansfeedback från användaren går ut på att användaren efter en första sökning pekar ut de mest relevanta dokumenten. Systemet ökar då vikten, dvs. betydelsen, för de förekommande termerna och använder de i nästa sökning, som är själva expansionen (Baeza-Yates & Ribeiro-Neto 1999, s.118ff).

### 3.1.2 Manuell query expansion

Som namnet antyder utförs manuell QE av användaren utan någon hjälp av det system som sökningen utförs i. Det är följaktligen användaren som själv väljer expansionstermer. Expansionstermerna kan exempelvis hämtas från användarens eget ordförråd, från återvunna dokument eller från hjälpmedel som tesaurus, index och ordlistor. Återvänder man till Efthimiadis indelning ovan (Fig.1) gäller även här kollektionsberoende och kollektionsoberoende källor för expansionstermerna om de hämtas från en kunskapsstruktur och inte från de återvunna dokumenten, dvs. sökresultatet. De nya termerna läggs sedan till sökfrågan och en ny expanderad sökning görs.

Manuell QE har även ibland kallats för intellektuell QE och enligt Efthimiadis kräver denna expansionstyp en stor intellektuell insats av användaren för att utfallet skall bli återvinningseffektivt (Efthimiadis 1996, s.126). Användaren behöver då bland annat ha kunskap om matchningstekniken i det system som används samt en klar bild av vad som eftersöks. En sökstrategi bör även utvecklas där användaren beslutar vilka termer som skall ingå i sökfrågan och hur sökningen ska utformas. Det finns flera olika sökstrategier och några av de vanligast förekommande är *building blocks*, *citation pearl growing*, *most specific fact*, *fractionalizing* och *exakt matchning* (Efthimiadis 1996, s.127). Samma strategi kan figurera med flera benämningar. Några av de engelska namnen på dessa begrepp används här men förklaras i nedanstående stycke.

Strategin *building blocks* innebär att användaren söker på ett begrepp i taget, till en början kan flera synonymtermer testas genom den booleska operatorn OR. Sedan sätts resultatet av dessa termblock samman med operatorn AND, och kan även begränsas genom operatorn NOT. *Citation pearl growing* utgår från ett för användaren redan känt och bra dokument, en ”pärla”, och använder det som källa för att gå vidare och hitta t ex söktermer, expansionstermer, författarnamn och så vidare. Genom *most specific fact* startas sökningen med en smal term och expanderas sedan med bredare och bredare termer. Dess motsats är *fractionalizing* som utgår från en bred term för att sedan kombinera med snävare termer. Den sistnämnda och enklaste av dessa är *exakt matchning* där användaren känner till något med säkerhet, det kan vara en titel eller ett författarnamn som kan användas i sökningen.

## 3.2 Effektivitetsmått

För att utvärdera hur en sökfråga presterar i fråga om återvinning, och kunna jämföra detta resultat med andra, krävs att man för situationen och ändamålet har fungerande effektivitetsmått. De vanligast förekommande och mest använda måtten inom IR är precision och recall (Baeza-Yates & Ribiero-Neto 1999, s.74), vilka även i något modifierad form kommer att användas i denna studie. Precision är ett mått på den del av de återvunna dokumenten som är relevanta:

$$\frac{\text{återvunna relevanta dokument}}{\text{samtliga återvunna dokument}}$$

Om samtliga återvunna dokument är relevanta blir precisionen 100 %. Recall är ett mått på förhållandet mellan antalet återvunna relevanta dokument och det totala antalet relevanta dokument i samlingen:

$$\frac{\text{återvunna relevanta dokument}}{\text{samtliga relevanta dokument i samlingen}}$$

Har alla relevanta dokument i samlingen återvunnits är recall 100 %. Beräkningen av recall förutsätter att det totala antalet relevanta dokument i samlingen är känt, vilket inte alltid är fallet. Istället för absolut recall kan då relativ recall beräknas. Här behöver man således inte känna till samtliga relevanta dokument utan beräknar antalet relevanta dokument för varje informationsbehov. Relativ recall har beräknats inom exempelvis *Text REtrieval Conference* (TREC), och metoden kallas då *pooling*. Metoden går ut på att antalet relevanta dokument uppskattas för ett givet topic. En relevansbedömning görs då först för ett antal dokument och dessa ingår sedan i en *pool* som kan användas för beräkning av (relativ) recall (Chowdhury 1999, s.433f). Eftersom denna metod bygger på en approximation kan det dock finnas relevanta dokument utanför poolen.

Sett ur användarens perspektiv varierar precision och recall allt vartefter de återvunna dokumenten i rankningslistan undersöks, förutsatt att sökningen görs i ett system med relevansrankning. För att då få en bild av de båda måtten kan de visas i diagramform och man talar om precision/recallkurvor. Ett sådant diagram visar hur recall och precision förhåller sig till varandra vid olika nivåer. Precision och recall kan också mätas inom ett visst intervall, kallat *Document Cut-off Value* (DCV). Då kan beräkningen göras när exempelvis 5, 10, 15 eller 20 träffar i listan av återvunna dokument har undersökts (Baeza-Yates & Riberio-Neto 1999, s.78f). För en viss sökfråga kan även genomsnittlig precision vid genererade recallnivåer uträknas, då beräknas precisionen vid varje observerat relevant dokument och efter det ett genomsnittsvärde.

Genomsnittlig precision är intressant eftersom man får fram ett tal som tar hänsyn till rankningseffektiviteten. Den genomsnittliga precisionen beräknas på följande vis: låt säga att DCV = 10 då endast de 10 först placerade träffarna i rakningslistan undersöks samt att 2 av dessa är relevanta. Det första observerade relevanta dokumentet ligger på plats 2 i träfflistan och det andra ligger på plats 4. Således är ett av två (1/2) relevanta när de två översta träffarna är undersökta och när fyra träffar är undersökta är två av dem (2/4) relevanta. För att få fram genomsnittet summeras dessa två tal (1/2 + 2/4 = 1) och delas med antalet träffar som är två (1/2 = 0.5). Resultatet ger därmed en genomsnittlig precision på 50 %. Detta skiljer sig markant från beräkning av ordinär precision vilket i detta fall hade blivit 2/10; en precision på 20 %. Den siffran säger dock inget om rankningseffektiviteten eftersom samma resultat erhållits oavsett var dessa träffar hade varit placerade inom intervallet.

Precision och recall har en lång tradition och är vedertagna effektivitetsmått inom IR. De har använts i flertalet av de tongivande experimentella evalueringsstudierna inom IR-forskningen. En av dessa, som redan nämnts ovan, är TREC, men även andra som Cranfield, STAIRS med flera har involverat recall och precision. STAIRS, som står för *STorage And Information Retrieval System*, var ett projekt som presenterades år 1985, med huvudsakligt syfte att undersöka hur bra ett system kunde återvinna enbart relevanta dokument givet en viss förfrågan, och för denna evaluering användes recall och precision. Cranfieldtesterna I och II utfördes under 1950 och 60-talen och kom att bli stilbildande inom IR och evalueringsforskningen. Även här användes främst precision och recall. TREC kontrasterar mot tidigare kollektioner eftersom man där söker efterlikna verkliga förhållanden genom storskalighet och dokument i fulltext. Projektet startades i början på 1990-talet och fortgår årligen. Här används standardiserad evaluering, och tillsammans med ovan nämnda poolingmetod för beräkning av relativ recall används även precisionsmåttet i utvärderingarna (Chowdhury 1999, s.215ff).

De båda måtten har således varit flitigt använda, men har även fått utstå hård kritik. En reflektion kring precision och recall är att dessa bör ses som mått på två olika aspekter av återvinningseffektivitet. Det råder ett visst motsatsförhållande dem emellan och ökar recall tenderar precisionen att minska och vice versa (Chowdhury 1999, s.205). Att dock mäta eller presentera precision men inte recall eller vice versa är inte praktiskt, eftersom det är relativt enkelt att uppnå hög recall genom att återvinna samtliga dokument i en databas, eller att tvärtom att få hög precision på bekostnad av låg recall (Harter & Hert 1997, s.10). Den framförda kritiken har därför bland annat rört denna problematik, samt gällande beräkning av recall och precision i större samlingar. Vidare har TREC och Cranfieldexperimenten kritiserats för att de standardmått som precision och recall som använts inte är informativa, och man menar istället att evaluering av individuella sökfrågor ger mer vägledande resultat (Harter & Hert 1997, s.27). Trots kritiken mot dessa båda mått används de dock alltjämt och det finns inga riktigt vassa konkurrerande alternativ.

Avslutningsvis bör understrykas att inom undersökningar där QE är involverat eftersträvas främst en ökad grad av recall, varför precisionen kan ha en benägenhet att minska. I nästa kapitel om tidigare forskning tas ett antal studier inom QE och deras resultat gällande precision och recall upp.

# 4. Tidigare forskning

Forskningen kring QE har bedrivits utifrån flertalet olika angreppssätt och syften. Redan på 1960-talet föreslogs att en sökfråga kunde expanderas med näraliggande termer (Maron & Kuhns 1960) och sedan dess har man inom detta område fortsatt att undersöka QE på olika sätt. Urvalet av tidigare forskning har till viss del varit problematisk. Det beror på att området query expansion rymmer en mängd specialiserade delområden. Som en följd av detta är en stor del av forskningen alltför specifik och avancerad för att kunna bidra med något till denna studie.

Detta kapitel kommer att fokusera på några tidigare utförda undersökningar som i relativt hög grad liknar föreliggande studie gällande metod och tillvägagångssätt. Dessa undersökningar är här tänkta att dels fungera som jämförelsematerial och dels som exempel för att visa hur forskningen gällande query expansion med semantiskt relaterade termer kan se ut.

- *J. Kristensen & K. Järvelin*

Jaana Kristensen (senare Kekäläinen) och Kalervo Järvelin påbörjade en serie undersökningar i samma riktning genom en pilotstudie (1990). I denna pilotstudie konstaterade de att den tesaurus som användes för att expandera den initiala sökfrågan med synonymer och relaterade termer förbättrade relativ recall i samtliga försök. Den initiala sökfrågan expanderades först med synonymer och sedan med relaterade termer. Undersökningen utfördes i en boolesk testmiljö med hjälp av IR-systemet BASIS och en kollektion bestående av 34 000 finska tidningsartiklar i fulltext.

En testtesaurus konstruerades med termer inom ekonomiområdet som hämtades från tidningsartiklar inom detta ämne. Tesauren innehöll synonymer, bredare, snävare och relaterade termer samt bestod av 328 olika termer. I undersökningen var man inte intresserad av att veta om samtliga relevanta dokument hittades eller inte utan syftet var just att jämföra vilken av strategierna som presterade bäst.

Fem journalister verksamma inom ekonomi- och handelsområdet ombads att formulera informationsbehov och av 30 formulerade sådana kunde 18 utföras komplett med samtliga två expansionsstrategier. Samma journalister utförde relevansbedömningar av de återvunna dokumenten där tre grader av relevans var möjliga: irrelevant, kanske relevant och relevant. Dock räknades de kanske relevanta samman med de irrelevanta vilket efter summeringen gav endast två relevansmöjligheter. I evalueringen mättes relativ recall, precision samt net effect.[5]

Resultaten gav att relativ recall förbättrades i båda expansionsstrategierna jämfört med den initiala sökningen. Eftersom recall ökade blev följden att precisionen minskade; från 48,5% i den första, icke expanderade sökningen, till expansionen med synonymer sjönk precisionen till 40,2% och i expansionen med relaterade termer landade precisionen på 35,0%. Relativ recall ökade i samma fall som ovanstående från 44,4% i första sökningen, till 76,8% i synonymsökningen och till 100% i den relaterade termsökningen.

[5] Förklaring angående *net effect* se Kristensen & Järvelin 1990, s. 80

Vad som även sågs i denna undersökning var att expansionen med synonymer gav en något mer tydlig ökning gällande relativ recall än expansionen med relaterade termer. De procenttal från deras undersökning som jag valt att redovisa ovan gäller de 18 informationsbehov med tillhörande sökfrågor som kunde fullföljas i samtliga strategier. Utifrån dessa resultat drar författarna slutsatsen att en tesaurus som används vid söktillfället är ett effektivt redskap för att förbättra sökresultaten i en fulltextmiljö.

- *J. Kristensen*

I nästa studie (Kristensen 1993) visade återvinningsresultatet för relativ recall på förbättring i de sökfrågor som expanderades jämfört med grundstrategin, men nu med hjälp av en större tesaurus. Utöver expansion med synonymer och relaterade termer testade Kristensen här även snävare termer samt en strategi med kombination av synonymer, relaterade termer och snävare termer. Var för sig kombinerades varje expansionsstrategi med grundstrategin till skillnad från föregående studie. Samma IR-system användes som i föregående studie, dvs. BASIS, men kollektionen bestod här av 227 000 st finska tidningsartiklar i fulltext i en operationell databas. Ämnet ekonomi utökades till att även inkludera miljöområdet varefter tidningsartiklar som överensstämde med dessa båda ämnen användes som källor för att skapa en testtesaurus. Även den inverterade filen i databasen användes som källa för att finna tesaurustermer.

Tesauren för denna studie bestod av 1573 antal termer. Precis som i förra studien engagerades journalister för att formulera informationsbehov och utifrån detta ställdes sökfrågor. Även en nyhetsservice nyttjades för detta ändamål. Journalisterna, samt anställda vid den nämnda servicearbetsplatsen, bedömde relevansen för de återvunna dokumenten binärt, dvs. dokumenten var antingen relevanta eller inte. Antalet sökfrågor som fullföljdes i samtliga expansionsstrategier var 14 stycken. I undersökningen testades även huruvida eventuella skillnader kunde ses i återvinningsresultat mellan användning av den booleska operatorn AND och användning av närhetsoperatorer.

Resultaten visade inga statistiskt signifikanta skillnader i återvinningseffektivitet mellan de tre expansionsstrategierna. Samtliga expansionsstrategier i kombination gav den högsta ökningen av relativ recall; från 47,2% i den icke expanderade sökningen mot 100 % i kombinationen med grundstrategi, synonymer, relaterade samt snävare termer. Den genomsnittliga precisionen minskade här i samma fall som ovan från 62.5% till 51.2%. Totalt sett var användningen av en tesaurus i denna studie återvinningseffektiv. Dubbelt så många relevanta dokument återvanns genom kombinationen där samtliga termrelationer från tesauren användes jämfört med dem som återvanns med den icke expanderade sökfrågan. Kristensen understryker att användningen av en tesaurus vid sökning är inriktad på ökning av recall, och menar vidare att resultaten överensstämmer med den tidigare nämnda pilotstudien.

- *J. Greenberg*

En undersökning av Jane Greenberg (2001a) syftar till att undersöka huruvida semantiskt relaterade termer till den initiala sökfrågan förbättrar återvinningseffektiviteten genom expansion. De semantiska relationerna som testades var synonymer, snävare termer, relaterade termer samt bredare termer. Greenbergs studie utfördes i en operationell miljö genom databasen ABI/Inform via DIALOG. Studenter vid Katz Graduate School of Business vid universitetet i Pittsburgh engagerades för att formulera informationsbehov och relevansbedöma dokumenten. Det var således verkliga användare och reella

informationsbehov i denna studie. Expansionstermerna hämtades från tesauren ProQuest Thesaurus. Användarna formulerade informationsbehov ovetande om tesauren och de frågor som matchade med minst en term gentemot tesauren togs med i undersökningen. Greenberg använde fyra hypoteser för att bekräfta eller bestrida att synonymer, snävare termer, relaterade termer och bredare termer förbättrar relativ recall samt på olika sätt, signifikant eller inte, minskar precisionen. Studien testade en expansionsstrategi åt gången. Sökningarna gjordes i en boolesk miljö. Om 15 eller färre än 15 dokument återvanns granskades dessa. Om fler än 15 återvanns rankades dessa baserat på Dice likhetskoefficient. Dokumenten relevansbedömdes sedan enligt en tregradig skala; irrelevant, delvis relevant samt relevanta. Dock lades delvis relevanta och relevanta dokument samman i den slutliga bedömningen. Sammanlagt undersöktes 42 stycken informationsbehov inom handels/affärsområdet.

Resultaten visade att relativ recall ökade för samtliga expansionsstrategier, men mest för expansionen med bredare och relaterade termer. Precisionen var bäst gällande expansionsstrategin med synonymer och därefter i sjunkande ordning snävare, bredare och relaterade termer.

Greenberg presenterade ytterligare en studie samma år (2001b) vilken syftade till att se vilken metod eller process som lämpar sig bäst och är optimal för tesaurusbaserad QE. I undersökningen var det QE med semantiska relationer som studerades, och frågeställningarna fokuserade på vilken metod som är bäst för detta: automatisk eller interaktiv QE? Greenberg var intresserad av att veta om alla semantiska relationer (synonymer, snävare termer, bredare termer och relaterade termer) är bra för QE eller om någon av dem är bättre. Hon ställde här liksom i den förra undersökningen upp fyra hypoteser med antaganden om att t ex relaterade termer är bättre för interaktiv QE än för automatisk eftersom de ökar relativ recall och så vidare. Greenbergs frågeställning här ligger långt från föreliggande studie men resultaten visar intressanta indikationer på hur QE med semantiska relationer presterar. IR-system, kollektion och tesaurus var samma som i föregående studie och även här engagerades studenter från ovan nämnda utbildning för att formulera informationsbehov. Både automatisk och interaktiv QE testades i samma system, något som Greenberg påpekar är sällan förekommande i den tidigare forskningen.

Hennes resultat visade att både synonymer och snävare termer generellt sett är bra för både interaktiv och automatisk QE. Dock valde användarna synonymer i markant högre grad framför snävare termer för QE. Studien visade även att bredare och relaterade termer är effektiva för automatisk QE om det är hög recall som eftersträvas. Är det istället hög precision som användaren vill ha lämpar sig interaktiv QE bättre gällande expansion med bredare och relaterade termer. Denna studie var relativt liten och det som Greenberg konstaterar utifrån resultaten är att mer forskning behövs med fokus på vilka typer av semantiskt relaterade termer som är bäst för olika QE-metoder.

- *E. M. Voorhees*

Ellen M. Voorhees studerade i en undersökning (1994a) resultaten av QE med semantiskt relaterade termer. Sökningarna gjordes i testkollektionen TREC och tesauren WordNet nyttjades för att ta fram fyra expansionsstrategier. WordNet är en allmän tesaurus som har konstruerats manuellt av forskare på Cognitive Science Laboratory vid universitetet i Princeton. Expansionsstrategierna, som följde WordNets uppbyggnad, bestod av expansion med synonymer, med alla hierarkiskt under- och överliggande liggande termer samt en

expansion med de närmast liggande termerna i tesauren.[6] Samtliga expansionstermer handplockades. Återvinningen baserades på vektormodellen. Resultaten visade att expansionen inte gav någon markant ökning i återvinningseffektivitet. Istället var det tydligt att QE i detta fall inte var fördelaktigt för att återvinna fler relevanta dokument jämfört med icke expanderade sökfrågor. Voorhees menar att detta troligen beror på TREC-kollektionens topics som är komplexa och formulerade så att de innehåller många termer. Detta gör att trots att nya termer läggs till uppträder de termer med högst vikter, dvs. de viktigaste termerna, redan i den initiala sökfrågan. Det gav negativ effekt på expansionen, speciellt när dessa viktiga termer inte hade några näraliggande expansionstermer i tesauren.

En annan orsak som Voorhees menar kan bidra till resultatet är att WordNet inte är utformad för att användas på det här sättet. Trots det tror hon dock inte att någon annan bred kollektionsoberoende kunskapsstruktur skulle prestera bättre. Hon konstaterar även att det inte tycks vara några skillnader i vilka typer av termer som är bra eller bättre för expansion. Detta stämmer väl överens med ytterligare resultat av Voorhees (1994b). Resultaten i denna andra studie, utförd på samma kollektion och med samma tesaurus, visade att QE med hjälp av en generell tesaurus var till liten hjälp för att öka återvinningseffektiviteten för en sökfråga om denna redan från början var välformulerad och detaljerad. Det visade sig nämligen att i de fall där den initiala sökfrågan var illa formulerad och innehöll få eller dåliga termer förbättrades återvinningsresultaten vid expansion. Var däremot den initiala sökfrågan från början detaljerad och välformulerad gavs inte någon förbättring.

Voorhees konstaterar därmed att korta sökfrågor är potentiellt bra för expansion men att resultatet beror på vilka termer som läggs till, samt att om dåliga termer läggs till kan det istället försämra resultatet. Vidare menar Voorhees att QE med semantiska relationer har potentialen att förbättra återvinningsresultaten eftersom användare vanligen inte formulerar detaljerade sökfrågor med många termer.

- *R. Lehtokangas & K. Järvelin*

Den sista studie som kommer att presenteras i detta kapitel är gjord av Raija Lehtokangas & Kalervo Järvelin (2001). K. Järvelin finns med som författare i en av de ovan nämnda studierna, men denna studie är dock av en något annan typ och att den tas med här är att tillföra ytterligare en dimension till den tidigare forskningen i detta kapitel. Artikeln argumenterar för användning av semantiskt relaterade expansionstermer vid QE. Antagandet grundar sig på resultaten i en undersökning där de studerar hur konsekvent, eller icke konsekvent, användandet av termer är vid nyhetsrapportering i dagspressen. Syftet med deras studie är för QE mer indirekt och den går från andra hållet jämfört med traditionella QE-studier. Istället för att undersöka och testa QE går de till själva källorna och texterna för att se hur språket används när det gäller dessa fenomen med bland annat synonymer och ambiguitet, samt för att se hur stor lingvistisk variation som förekommer.

De tre största finska dagstidningarna och deras sätt att rapportera, dvs. vilka termer som de använde i rapporteringen, om samma händelser studerades. Tidningarna liknar varandra ifråga om språk, storlek och typ samt även gällande typ av läsare och storleken på publiken/läsekretsen. Undersökningen fokuserade på ord och fraser för att ligga på samma nivå som de flesta IR-system, vilka också opererar med ord och fraser. Från de tre tidningarna hämtades 60 nyhetshändelser vilka sedan grupperades beroende på artikeltyp. Termer och

[6] För närmare förklaring angående termrelationerna och WordNets uppbyggnad se Voorhees 1994a, s. 225.

fraser handplockades ur dessa. Analysen visade att det var klara skillnader mellan val av termer och fraser i korta jämfört med långa artiklar. Korta artiklar innehöll maximalt 100 ord och långa minst 200 ord. Konsekvensen var över lag ca 65 % och varierade som mest mellan de olika artikelkategorierna. Kortare artiklar var mer konsekventa med ett genomsnitt runt 83 % medan långa artiklar hade ett genomsnitt på 46 %. Detta tycks även helt naturligt med tanke på att i längre artiklar kan en reporter breda ut texten och har möjlighet att formulera sig på andra sätt jämfört med kortare artiklar. De typer av inkonsekvens i användning av termer och fraser som hittades kunde sorteras in i semantiska relationer som ekvivalens, hierarkiska och associativa sådana.

Författarna menar att eftersom automatisk indexering blir alltmer gällande är det viktigt att belysa dessa skillnader i språket och ta hänsyn till dem vid informationssökning. Detta kan då ske genom att bland annat semantiska relationer mellan termer kan erbjudas informationssökare genom QE. Enligt författarna kan IR-forskningen dra nytta av dessa resultat genom att rikta uppmärksamheten till de dokument som figurerar i återvinningssystemen.

## 4.1 Sammanfattning och reflektion kring den tidigare forskningen

En summering av ovanstående studier ger att samtliga, förutom Lehtokangas & Järvelin (2001) vilka platsar i en annan kategori, presenterar resultat som visar att QE med semantiskt relaterade termer förbättrar återvinningseffektiviteten när det gäller recall eller relativ recall. Resultaten har dock vissa undantag varav Voorhees (1994a, 1994b) visar att QE inte är lika effektivt när den initialt ställda sökfrågan är detaljerad och välformulerad. Gällande skillnader mellan de olika expansionstermernas inverkan på recall är resultaten inte entydiga; Kristensen & Järvelin (1990) visar att synonymer ger bättre ökning än relaterade termer, Kristensen (1993) får inga markanta skillnader mellan expansionsstrategierna och Greenberg (2001a) ser mest ökning för bredare och relaterade termer. I samtliga undersökningar minskar dock precisionen vid expansion. Att det även finns skillnader mellan automatisk och interaktiv QE gällande vilka termrelationer som är mest effektiva för respektive metod visar Greenberg i sin studie (Greenberg 2001b).

Kristensen & Järvelin (1990) drar slutsatsen att en tesaurus som används vid söktillfället är ett effektivt redskap för att förbättra resultaten för fritextsökning i en fulltextmiljö, dock kan viss kritik riktas mot denna generella slutsats, dels på grund av den relativt småskaliga tesauren och dels på grund av de fåtal informationsbehov som verkligen kunde ge sökfrågor för samtliga expansionsstrategier. Dock indikerar den att tesaurusbaserad QE är återvinningseffektivt vilket Kristensen även söker belägga genom fortsatt forskning (1993).

Voorhees (1994a, 1994b) resultat är tänkvärda och visar att forskningen kring QE inte är fix och färdig med entydiga resultat. Att en från början välformulerad och detaljerad sökfråga inte har någon hjälp av expansion ter sig ganska självklart, dock visar undersökningar, t ex de tidigare nämnda i denna uppsats, att det är vanligare med få termer i sökfrågan och dåliga sådana än bra sökfrågor med många termer varför fortsatt forskning är intressant och relevant.

Trots att Lehtokangas & Järvelin (2001) inte gör någon regelrätt studie av QE visar resultaten i deras undersökning ändå att QE med semantiskt relaterade termer borde vara effektivt eftersom det förekommer stora variationer i användningen av termer gällande skriftliga rapporter, dvs. tidningsartiklar, från flera källor angående samma händelser. Detta är

användbart och intressant att ha i åtanke för föreliggande studie eftersom den kollektion som har använts här består just av tidningsartiklar i fulltext från två olika dagstidningar.

Tillsammans ger detta en bild av att forskningen gällande QE med semantiskt relaterade termer inte är helt samstämmig men att detta ändå tycks vara en relativt effektiv metod att finna fler relevanta dokument vid sökningar. Flertalet frågor står obesvarade och motiverar således fortsatt forskning inom området. Någon tidigare utförd studie som ur liknande eller samma perspektiv undersöker återvinningseffektiviteten i ett probabilistiskt system där expansionstermerna hämtas från kollektionsoberoende källor har inte kunnat hittas. Det tillsammans med ovanstående kan ses som ytterligare motivering till genomförandet av föreliggande studie.

# 5. Metod

I detta kapitel beskrivs undersökningens metod. I avsnitt 5.1 redogörs för testmiljön, vilket innefattar IR-systemet, kollektionen samt verktyget QPA. I avsnitt 5.2 beskrivs urvalet av topics (även kallat informationsbehov) samt hur dokumenten i kollektionen har relevansbedömts. Avsnitt 5.3 tar upp de olika expansionsstrategierna och i avsnitt 5.4 med underavdelningar finns redovisning av de hjälpmedel som kommer att användas för expansionen. Slutligen presenteras evalueringsmetoden i avsnitt 5.5, och allt detta sätts sedan samman i en mer konkret beskrivning av tillvägagångssättet i avsnitt 5.6.

## 5.1 Testmiljö

Det IR-system som har använts i denna studie är *InQuery* version 3.1, utvecklat vid Center for Intelligent Information Retrieval, Department of Computer Science vid University of Massachusetts i USA. Här ges en kortfattad beskrivning av InQuery där samtliga fakta är hämtade från Kekäläinen (1999, s. 25-29) och Ahlgren & Eklund (2003, s. 1-4). InQuery är ett probabilistiskt system vilket gör partiell matchning möjlig. Då partiell matchning är möjlig behöver sökfrågevillkoren endast uppfyllas delvis. Systemet är baserat på bayesianska nätverk där sannolikhetslära används.[7] Mycket förenklat kan man säga att när en sökfråga ställts till systemet beräknas sannolikheten att ett informationsbehov är mött; denna sannolikhet beräknas för varje dokument i samlingen och på detta baseras rankningen av de återvunna dokumenten. I återvinningen matchas också representationerna för dokumenten mot den sökfråga som ställts till systemet och graden av likhet mellan dessa mäts.

Genom InQuerys metod för termviktning ges varje term en vikt, som motsvarar dess betydelse eller *belief value*, beliefvärde, i dokumentet. Ett beliefvärde beräknas även för sökfrågan och dess termer. För att en term ska få en stor vikt i ett dokument måste det ha en hög frekvens, dvs. förekomma många gånger. Vidare bör dokumentet vara relativt kort jämfört med de andra dokumenten i samlingen, och den aktuella termen bör även förekomma i endast ett fåtal av dokumenten i samlingen. Det ideala för en term med hänseende på dess beliefvärde är således att förekomma många gånger i ett fåtal, relativt korta dokument i samlingen.

Som nämnts ovan beräknas även ett beliefvärde för sökfrågan och dess termer. För att formulera en sökfråga erbjuder InQuery ett flertal olika operatorer som alla föregås av tecknet # och begränsas genom parenteser. De operatorer som har använts i denna studie är #syn, #sum och #od. Vid användning av #syn-operatorn placeras termer inom denna som är synonymer, eller korrekt uttryckt termer som är instanser av samma begrepp. Det kan t ex röra sig om olika stavning av samma ord, böjningsformer eller olika termer för samma begrepp. Frekvensen av alla de termer som finns inom operatorn #syn summeras. Det ger att för storleken på beliefvärdet där operatorn #syn används är det oväsentligt vilka av sökfrågans termer som är närvarande i dokumentet, så länge summan av dem är konstant. Det betyder exempelvis att om en sökning görs på följande termer: #syn(förväxling sammanblandning) blir beliefvärdet lika stort om termerna förekommer var sin gång i ett dokument jämfört med om en av termerna förekommer två gånger och den andra ingen.

[7] Angående förklaring av bayesianska nätverk, se Baeza-Yates & Ribeiro-Neto 1999, s.48 eller Kekäläinen 1999, s.25ff.

Operatorn #sum ger att samtliga termer inom denna summeras och alla har lika stor betydelse för återvinningsresultatet. Inom denna operator kan både termer och uttryck skapade med andra operatorer, exempelvis inom operatorn #syn, användas. Beliefvärdet för en sökfråga med operatorn #sum ges genom en beräkning av medelvärdet av samtliga beliefvärden för de termer som ingår.

Den sista av de nämnda operatorerna är #od vilket står för *ordered distance* och är en närhetsoperator. Denna anges tillsammans med en siffra, t ex #od1, som då binder samman två eller fler komponenter. Termer som anges inom denna skall förekomma i given ordning. I ovanstående anger siffran 1 att mindre än ett ord får förekomma mellan komponenterna.

Resultatet av sökningarna i InQuery har i denna studie bearbetas med hjälp av verktyget *Query Performance Analyser* (QPA). InQuery är inbyggt i QPA. QPA är ett redskap och hjälpmedel som möjliggör analysering, tolkning, jämförelse och visualisering av återvinningseffektiviteten för olika sökfrågor. Utvecklingen av QPA har skett vid Institutionen för Informationsstudier vid Tammerfors universitet i Finland. Samtliga sökningar utförs i kollektionen ”GP_HDINF”, vilken består av dagstidningsartiklar i fulltext. Artiklarna är på svenska och hämtade från Helsingborgs Dagblad och Göteborgs Posten från år 1994. Artiklarna varierar i längd från endast ett par meningar upp till ett par tre sidor. Kollektionen består av 161 336 st dokument.

## 5.2 Topics och relevansbedömningar

Ett topic kan förstås som ett informationsbehov. Ett exempel på ett topic som har använts i föreliggande undersökning är ”*Vilka faktorer undergräver den europeiska industrins konkurrenskraft på världsmarknaden*”? För kollektionen ”GP_HDINF” finns sammanlagt 51 informationsbehov formulerade. Utifrån dessa har ett urval gjorts där de 26 topics som uppvisar flest antal relevanta dokument har valts ut. Detta har gjorts dels på grund av att studien måste avgränsas för att inte bli alltför omfångsrik och dels på grund av att de topics som enbart har ett fåtal relevanta dokument är relativt ointressanta ur expansionssynpunkt. För att ett topic ska tas med i undersökningen har det behövt uppvisa minst 9 relevanta dokument.

Relevansbedömningarna är gjorda sedan tidigare och har utförts av fyra f d studenter inom ämnet biblioteks- och informationsvetenskap. Den relevansskala som användes är fyrgradig och går från 0 till 3. De olika nivåerna kan beskrivas enligt följande (Sormunen 2002, s.325):

(0) **Irrelevant**. Dokumentet innehåller ingen relevant information för gällande topic.

(1) **Marginellt relevant**. Dokumentet innehåller ingen mer information än vad som framgår i gällande topic. Vanligen innehållande en mening eller några få uppgifter som är relevanta.

(2) **Relevant**. Dokumentet innehåller mer information än gällande topic men är inte uttömmande. Vanligen innehållande ett relevant stycke eller 2-3 relevanta meningar.

(3) **Mycket relevant**. Dokumentet innehåller relevant och uttömmande information för gällande topic. Vanligen innehållande flera relevanta stycken eller mer än 4 relevanta meningar.

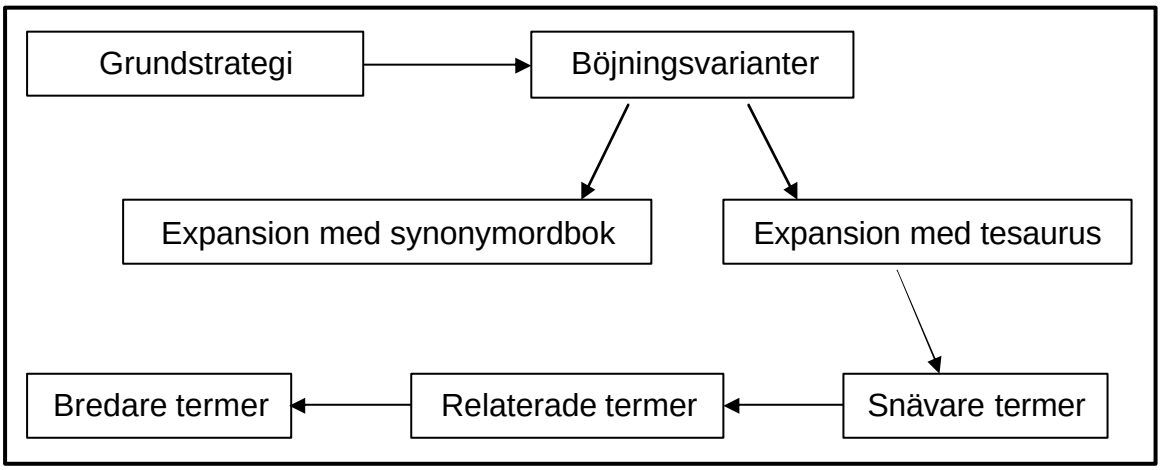
I denna undersökning har dock tolkningen av relevans i själva verket varit binär. Oavsett ovanstående relevansgrad har nivåerna 1-3 räknats samman som relevanta medan endast 0 bedömts som irrelevant.

## 5.3 Expansionsstrategier

För varje informationsbehov har sex sökfrågor formulerats. Det innebär en grundstrategi samt fem expansionsstrategier. Grundstrategin motsvarar termer för varje identifierad aspekt av ett givet topic. De olika strategierna är följande:

- Grundstrategi
- Expansion med böjningsvarianter av termer i grundstrategin
- Expansion med böjningsvarianter och med synonymer med böjningsvarianter
- Expansion med böjningsvarianter och snävare termer med böjningsvarianter
- Expansion med böjningsvarianter, snävare termer och relaterade termer med böjningsvarianter
- Expansion med böjningsvarianter, snävare termer, relaterade termer samt bredare termer med böjningsvarianter

I nedanstående figur (Fig.2) ges en schematisk bild av de olika strategierna och hur de förhåller sig till varandra. Mer detaljerat finns tillvägagångssättet beskrivet i avsnitt 5.6.




*Figur 2. Schematisk bild av de olika expansionsstrategierna.*

## 5.4 Expansionsverktyg

### 5.4.1 Tesaurus

För att kunna expandera med hierarkiskt under- och överordnade termer samt även med relaterade termer i förhållande till grundstrategin har tesauren Allärs använts. Namnet Allärs står för *allmän tesaurus på svenska* och är en ämnesoberoende tesaur med ca 16 000 ämnesord och ungefär 3500 hänvisningstermer. Den publicerades i tryckt form år 1996 och finns sedan 1999 på webben. Allärs är en översättning av den finska tesauren YSA och skapades på initiativ av Finlands svenska biblioteksförening. Helsingfors universitetsbibliotek tillsammans med Åbo akademis bibliotek ansvarar för dess underhåll och utveckling. Ämnesorden är substantiv och finns angivna som enkla begrepp eller sammansatta ord och uttryck. De relationer mellan termer som visas är snävare, bredare och relaterade (Allärs 2004). Allärs webbversion uppdateras kontinuerligt och nya termer läggs till. Eftersom tesauren är en översättning av YSA sker samma uppdateringar gällande nya termer i båda, och de nya termer som skall infogas i Allärs måste då först översättas till svenska. Enligt uppgift är det främst bibliotek, museer och arkiv som sänder förslag på nya termer och dessa diskuteras sedan av ansvariga i en speciell redaktion. Flera gånger årligen hålls större möten där man beslutar om nya termer ska få vara med eller inte. Det är ungefär 300-400 potentiella termer om året som diskuteras vid dessa sammankomster (Kärki 2004).

### 5.4.2 Ordbok och synonymlexikon

Ytterligare expansionsverktyg som har använts är en ordlista (Svenska akademiens ordlista över svenska språket 1998) för att få fram böjningsvarianter av expansionstermerna i korrekt form. Bonniers synonymordbok (Walter 2003) har nyttjats för att ta fram synonymer. Syftet med användningen av dessa båda hjälpmedel har dels varit att hämta expansionstermer från lämpliga källor och dels att få genomgående likartade expansionsstrategier över alla topics. I valet av termer för grundstrategin användes enbart termer som redan fanns presenterade i varje topic. Dock har vissa ändelser i dessa termer omformuleras eftersom deras form i varje topic är utfört för att passa in i de formulerade frågorna. I vissa fall har även sammansatta termer brutits ned, ett sådant exempel är *dopinganvändning* som blivit *doping*. Detta har gjorts på grund av möjligheten att hitta synonymer och andra semantiskt relaterade termer vilket annars skulle vara omöjligt för vissa sammansatta ord. För samtliga termer i grundstrategin gäller dock att inga nya ord utöver de som finns i varje topic har tagits med.

I expansionsstrategin med böjningsformer har grundstrategins termer utökats respektive böjningsvarianter. De olika böjningsformerna är obestämd form singular, bestämd form singular, obestämd form plural, bestämd form plural. I de fall där egennamn, samt geografiska namn på t ex länder och städer förekommit, har även genitivformen tagits med i expansionen, t ex Sydafrika och Sydafrikas. Dock inte sydafrikansk/sydafrikanska och så vidare. Att ta med genitivformerna endast här, och inte på samtliga termer, har skett för att detta var den enda typ av expansion som kunde utföras på dessa termer. Eftersom inga ingångar finns i synonymordboken eller tesauren för länder, egennamn etc. har dessa expanderats enbart genom tillägg av genitivändelser och t ex utskrivna eller sammanskrivna förkortningar. Ett exempel är *EU* som i synonymexpansionen även skrivits som *europeiska unionen*.

Tre större svenska synonymordböcker undersöktes i syfte att passa som hjälpmedel för denna studie: Svensk synonymordbok (1992), Nya stora synonymordboken (Strömberg 1998) samt Bonniers synonymordbok (Walter 2003). I samtliga av dessa visas utöver strikt ekvivalensrelaterade termer även termer som är hierarkiskt under- och överordnade. Ingen av

dem har således enbart synonymer i ordets korrekta bemärkelse. Detta medförde att expansionsstrategin med synonymer därför kunde komma att innehålla både bredare eller snävare termer i förhållande till termerna i grundstrategin. För att undvika sammanblandning och överlappning gjordes därför en särskiljning mellan expansionen med synonymer och de expansionsstrategier som skedde med hjälp av tesauren.

Eftersom varken innehåll eller uppbyggnad tycktes skilja synonymordböckerna åt valdes den som utkommit senast. I expansionsstrategin med synonymer har samtliga synonymer tagits med som presenterades i den aktuella synonymordboken. Detta för att undvika ett alltför subjektivt urval av passande termer.

Efthimiadis skriver om antalet termer som tas med vid QE och menar att detta antal beror på syftet med studien. Antalet termer i olika undersökningar är kraftigt varierande från ett fåtal upp till flera hundra (Efthimiadis 1996, s.135). I denna undersökning har samtliga termer som finns i de källor som används tagits med. Det betyder att inga urval har skett utan för att vara så konsekvent som möjligt har samtliga synonymer för en viss term tagits med och samma sak har gällt för tesauren och termerna där. Eventuella undantag från ovanstående beskrivs i resultatkapitlet.

## 5.5 Evalueringsmetod

Resultaten för varje sökfråga över samtliga topics har utvärderats genom relativ recall och genomsnittlig precision. Även medelvärden för varje strategi över samtliga topics har beräknats. För beräkningen av relativ recall är DCV = 200 vilket betyder att de 200 första träffarna har undersökts. Därmed kan det vara så att relevanta dokument för ett visst topic hamnar utanför den givna gränsen. För den genomsnittliga precisionen är DCV = 10 vilket betyder att genomsnittlig precision har beräknats endast med avseende på de tio första placerade dokumenten i rankningslistan.

## 5.6 Tillvägagångssätt

Varje expansionsstrategi för samtliga topics har tagits fram manuellt med hjälp av de ovan beskrivna hjälpmedlen. Som nämnts i de inledande kapitlen är det dock inte manuell utan automatisk QE som undersöks i denna uppsats. Det är automatisk QE i betydelsen att på ett så konsekvent och likartat sätt som möjligt, med ordbok, synonymordbok och tesaurus formulera sökfrågor. Tanken är att man i ett system skulle kunna implementera synonymordbok och/eller tesaurus och låta utföra QE med termer från dessa källor. Således finns en klar skillnad gentemot manuell QE. Målet har varit att och göra på samma sätt vara konsekvent genom hela studien. På följande sätt har strategierna utformats:

1. Grundstrategin (s1) representeras av termer för varje identifierad aspekt av ett givet topic. Enbart termer uttryckta i respektive topics har använts och dessa har valts manuellt. Dock har, som ovan nämnts, i vissa fall ändelser bytts ut och sammansatta ord brutits ned.

2. Strategi 2 (s2) är en expansion av grundstrategin med böjningsvarianter av samtliga förekommande termer. Böjningsformerna har tagits fram med hjälp av ordlista.

3. Strategi 3 (s3) bygger på s2 och är en expansion med synonymer. Även dessa synonymers böjningsvarianter har tagits med. Synonymerna togs fram med hjälp av synonymordbok. Samtliga synonymer som presenterades för den aktuella termen i synonymordboken togs med i expansionen.

4. Strategi 4 (s4) bygger på s2 och är en expansion med snävare termer. Även dessa snävare termers böjningsvarianter togs med. Dessa termer togs fram med hjälp av tesauren Allärs. Samtliga termer som presenterades för den aktuella termen togs med.

5. Strategi 5 (s5) bygger på s4 och är en expansion med relaterade termer. Även dessa relaterade termers böjningsvarianter tas med. Dessa termer togs fram med hjälp av tesauren Allärs. Samtliga termer som presenterades för den aktuella termen togs med.

6. Strategi 5 (s6) bygger på s5 och är en expansion med bredare termer. Även dessa bredare termers böjningsvarianter togs med. Dessa termer togs fram med hjälp av tesauren Allärs. Samtliga termer som presenterades för den aktuella termen togs med.

Strategierna kördes sedan mot databasen. För att det skulle vara meningsfullt att utvärdera återvinningseffektiviteten sattes följande kriterier upp: för att ett topic med tillhörande strategier skulle tas med behövde minst 3 av 6 strategier gå att genomföra med någon av de termer som presenterats i grundstrategin. För att kunna svara på frågeställningarna beräknades relativ recall samt genomsnittlig precision för varje sökfråga. Relativ recall gavs av InQuery i form av cirkeldiagram medan genomsnittlig precision beräknades fram manuellt. Dessa tal har sedan placerats i tabeller vilka återfinns i kapitel 6. För samtliga strategier och för de båda effektivitetsmåtten har även genomsnittlig prestation tagits fram och visas i tabellerna.

Följande exempel visar ett informationsbehov hämtat från kollektionen samt de sökfrågor som formulerats. Samma numrering som ovan gäller för de olika strategierna. De nya termer som har lagts till i varje ny sökfråga har markerats med fetstil.

**Nr 263. Topic 047:** *Vilka skäl finns det för den ryska militära interventionen i Tjetjenien?*

1. #sum(rysk intervention Tjetjenien)

2. #sum(#syn(rysk **ryska**) #syn(intervention **interventionen interventioner interventionerna**) #syn(Tjetjenien **Tjetjeniens**))

3. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna **inskridande inskridandet inskridanden inskridandena ingripande ingripandet ingripanden ingripandena inblandning inblandningen inblandningar inblandningarna medling medlingen**) #syn(Tjetjenien Tjetjeniens))

4. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna #od1(**humanitär intervention**) #od1(**humanitära interventionen**) #od1(**humanitära interventioner**) #od1(**humanitära interventionerna**)) #syn(Tjetjenien Tjetjeniens))

5. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna #od1(humanitär intervention) #od1(humanitära interventionen) #od1(humanitära interventioner) #od1(humanitära interventionerna) **erövring erövringen erövringar erövringarna fredsframtvingande fredsframtvingandet fredsframtvinganden fredsframtvingandena krigshandlingar krigshandlingarna krigshandling krigshandlingen ockupation ockupationen ockupationer ockupationerna** ) #syn(Tjetjenien Tjetjeniens))

6. Inga bredare termer gavs av Allärs.

# 6. Resultat

I detta kapitel visas resultaten av sökningarna i form av två tabeller. Avsnitt 6.1 visar relativ recall och avsnitt 6.2 genomsnittlig precision. Samtliga beräkningar anges som procenttal med två decimaler. Fem topics föll bort och finns redovisade i Appendix 2. Bortfallen berodde i tre fall på att för få expansionstermer hittades för att motsvara uppsatta kriterier med minst 3 av 6 genomförbara strategier. I de andra två fallen var det i expansionen med snävare termer inte givande att genomföra någon sökning och därmed togs båda dessa topics med samtliga strategier bort. Att inga sökningar gjordes här beror på att Allärs erbjöd ett mycket stort antal snävare termer till ”företag” och ”sjukdomar”, 35 respektive 58 st. En diskussion kring detta återfinns i kapitel 7. Samtliga av dess borttagna termer är placerade i appendix 3. Resultaten visar därmed sökfrågorna 1-6 för 21 st topics. Det har inte varit möjligt att finna expansionstermer för samtliga strategier för något topic, dvs. om det t ex funnits snävare termer har bredare termer saknats, eller om det funnits synonymer har det inte funnits relaterade termer. Detta har gjort att för inget topic har undersökningen kunnat genomföras komplett. Där expansionstermer saknas finns dessa markerade i appendix 1 i anslutning till respektive sökfråga och syns därmed inte i tabellerna.

## 6.1 Relativ recall

Tabell 1 visar procenttalen av relativ recall för samtliga sökfrågor. Längst ner i tabellen finns ett genomsnitt för hur varje strategi presterat. I tabellen är de sex strategierna beskrivna som s1-6.

| | s1 | s2 | s3 | s4 | s5 | s6 |
|---|---|---|---|---|---|---|
| Topic 005 | 53,57 | 59,52 | 59,52 | 59,52 | 54,76 | 54,76 |
| Topic 007 | 57,69 | 61,54 | 61,54 | 61,54 | 65,38 | 65,38 |
| Topic 011 | 33,33 | 55,56 | 22,22 | 0,00 | 0,00 | 0,00 |
| Topic 012 | 100,00 | 100,00 | 100,00 | 100,00 | 72,73 | 72,73 |
| Topic 013 | 68,09 | 59,57 | 40,43 | 59,57 | 59,57 | 59,57 |
| Topic 015 | 7,41 | 44,44 | 37,04 | 44,44 | 48,15 | 48,15 |
| Topic 017 | 21,05 | 47,37 | 47,37 | 47,37 | 47,37 | 57,89 |
| Topic 022 | 14,29 | 42,86 | 78,57 | 42,86 | 42,86 | 42,86 |
| Topic 026 | 67,57 | 70,27 | 70,27 | 70,27 | 70,27 | 48,65 |
| Topic 032 | 6,06 | 15,15 | 27,27 | 15,15 | 9,09 | 9,09 |
| Topic 042 | 75,93 | 87,04 | 48,15 | 87,04 | 87,04 | 87,04 |
| Topic 045 | 0,00 | 0,00 | 11,11 | 0,00 | 0,00 | 0,00 |
| Topic 046 | 0,00 | 0,00 | 0,00 | 0,00 | 10,00 | 10,00 |
| Topic 047 | 9,76 | 56,10 | 51,22 | 56,10 | 56,10 | 56,10 |
| Topic 048 | 25,00 | 28,57 | 28,57 | 28,57 | 21,43 | 21,43 |
| Topic 050 | 39,58 | 62,50 | 47,92 | 62,50 | 62,50 | 62,50 |
| Topic 051 | 33,33 | 77,78 | 77,78 | 77,78 | 77,78 | 77,78 |
| Topic 066 | 66,67 | 85,71 | 76,19 | 85,71 | 85,71 | 85,71 |
| Topic 080 | 42,86 | 47,62 | 47,62 | 47,62 | 47,62 | 47,62 |
| Topic 082 | 35,71 | 50,00 | 50,00 | 50,00 | 42,86 | 35,71 |
| Topic 086 | 34,29 | 34,29 | 45,71 | 48,57 | 48,57 | 48,57 |
| Genomsnitt | 37,72 | 51,71 | 48,98 | 49,74 | 48,09 | 47,22 |

*Tabell 1. Relativ recall i procent.*

## 6.2 Genomsnittlig precision

Tabell 2 visar procenttalen av genomsnittlig precision för samtliga sökfrågor. Längst ner i tabellen finns ett genomsnitt för hur varje strategi presterat. I tabellen är de sex strategierna beskrivna som s1-6.

| | s1 | s2 | s3 | s4 | s5 | s6 |
|---|---|---|---|---|---|---|
| Topic 005 | 53,33 | 53,33 | 53,33 | 53,33 | 42,66 | 42,66 |
| Topic 007 | 36,67 | 98,21 | 98,21 | 98,21 | 96,27 | 96,27 |
| Topic 011 | 100,00 | 39,29 | 0,00 | 0,00 | 0,00 | 0,00 |
| Topic 012 | 89,29 | 98,21 | 96,17 | 98,21 | 100,00 | 100,00 |
| Topic 013 | 100,00 | 26,67 | 40,28 | 26,67 | 26,67 | 26,67 |
| Topic 015 | 0,00 | 10,00 | 0,00 | 10,00 | 0,00 | 0,00 |
| Topic 017 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Topic 022 | 25,00 | 45,00 | 0,00 | 36,67 | 33,33 | 33,33 |
| Topic 026 | 22,50 | 41,55 | 41,55 | 41,55 | 41,55 | 32,50 |
| Topic 032 | 25,00 | 100,00 | 100,00 | 100,00 | 75,00 | 75,00 |
| Topic 042 | 88,02 | 92,83 | 34,29 | 92,83 | 92,83 | 92,83 |
| Topic 045 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Topic 046 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Topic 047 | 0,00 | 83,33 | 33,33 | 83,33 | 22,50 | 22,50 |
| Topic 048 | 0,00 | 0,00 | 0,00 | 0,00 | 14,29 | 14,29 |
| Topic 050 | 69,84 | 97,50 | 83,89 | 98,61 | 98,61 | 98,61 |
| Topic 051 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Topic 066 | 70,00 | 44,38 | 50,99 | 44,38 | 42,06 | 42,06 |
| Topic 080 | 57,87 | 68,57 | 67,46 | 68,57 | 68,57 | 68,57 |
| Topic 082 | 0,00 | 0,00 | 0,00 | 0,00 | 41,67 | 22,62 |
| Topic 086 | 47,62 | 50,00 | 58,89 | 58,45 | 62,95 | 62,95 |
| Genomsnitt | 37,39 | 45,18 | 36,11 | 43,37 | 40,90 | 39,56 |

*Tabell 2. Genomsnittlig precision i procent.*

# 7. Analys

Detta kapitel syftar till att diskutera och analysera ovanstående resultat. Innehållet är upplagt efter de fyra frågeställningarna och diskussionen knyter även an till den tidigare forskning som presenterades i kapitel 3. Slutligen avrundas kapitlet genom avsnitt 7.1 där en konklusion samt en sammanfattande analysbild av resultaten ges.

- Hur presterar de fem strategierna avseende relativ recall i förhållande till grundstrategin?

Grundstrategin gav ett genomsnitt på 37,72% vilket var klart lägst av samtliga strategier. Av de fem expansionsstrategierna presterade strategi 2 (s2) bäst med 51,71% i genomsnitt för samtliga topics. Näst därefter följde s4 (49,74%) och sedan s3 (48,98%), s5 (48,09%) och sist s6 (47,22%). Att grundstrategin visade det lägsta genomsnittet av strategierna var väntat eftersom det är här som syftet med QE kommer in i bilden. Siffrorna visar tydligt att ingen av expansionsstrategierna lyckades återvinna fler relevanta dokument än s2 där endast böjningsformer utnyttjades. Därmed visar genomsnitten att ingen expansionsstrategi lyckades bättre gällande relativ recall än expansionen med böjningsformer. Efter s2 med böjningsformer presterade strategierna i fallande ordning: snävare termer, synonymer, snävare + relaterade och sist snävare + relaterade + bredare termer.

De genomsnittliga talen är dock inte direkt rättvisande eftersom flera topics skiljer ut sig med högsta respektive lägsta värden som påverkar genomsnittet. Detta gäller bland andra topic 012 som återvann samtliga relevanta dokument, dvs. 100%, från grundstrategin fram till och med s4, för att sedan sjunka till en recallnivå på knappt 73% (se tabell 3 nedan). Detta topic är formulerat *"Sök dokument om Soltempelorden och mordet eller självmordet på dess medlemmar"*. Den initiala sökfrågan innehöll termerna *soltempelorden* och *självmord*, och med de termerna rankades tre relevanta dokument på position 1-3. Ytterligare ett relevant dokument återfanns på position 7.

| Topic 012 | s1 | s2 | s3 | s4 | s5 | s6 |
|---|---|---|---|---|---|---|
| Relativ recall % | 100,00 | 100,00 | 100,00 | 100,00 | 72,73 | 72,73 |
| Genomsnittlig precision % | 89,29 | 98,21 | 96,17 | 98,21 | 100,00 | 100,00 |

*Tabell 3. Medelvärden för topic 012 avseende relativ recall och genomsnittlig precision.*

Att samtliga relevanta dokument återvanns redan från grundstrategin i detta topic kan bero på att detta egennamn var en bra sökfrågeterm. Särskilt intressant här är relationen mellan precision och recall eftersom båda visar förhållandevis höga värden. Detta tycks bero på att de två termer som finns med redan i den initiala sökfrågan bra söktermer för detta topic. Det som sedan händer i s5 och s6 är att precis som brukligt är minskar recall när precisionen ökar. Allärs erbjöd här varken snävare eller bredare termer till självmord varför sökfrågorna i s5 och s6 är formulerade exakt likadant. Vad är det då som sker mellan s4 och s5? De tre termer som läggs till är *mord*, *självdestruktivitet* samt *självmordsförsök*. Med dessa termer återvinns endast 8 av 11 relevanta dokument men eftersom de placerar sig på position 1-8 ger det en genomsnittlig precision på 100%.

Vidare lyckades inte topic 045 och 046 återvinna några relevanta dokument alls i s1-2. Topic 046 presterade klart sämst här och lyckas inte förrän i s5 att öka recall från 0 till 10%. Det tycks i detta fall handla om endast en relaterad term som dyker upp i s5 och påverkar resultatet. Gällande topic 045 var det endast i expansionen med synonymer som recall ökade; från 0 till drygt 11%. I övriga följande strategier sjönk recall här tillbaka till 0%. För detta topic fanns inga snävare eller bredare termer, och endast en relaterad term hittades. Således var synonymordboken överlägsen tesauren i detta fall. Ännu ett topic som fått 0% är 011 som visar detta resultat för s4-6. Här är s6 samma som s5 eftersom inga bredare termer gavs av Allärs, men varför s4 och s5 presterar så dåligt är svårt att svara på. De expansionstermer som Allärs gav här har troligen helt enkelt fått för låga beliefvärden vid sökningen och är i detta fall dåliga ur återvinningssynpunkt.

Resultaten gällande relativ recall i denna studie visar att varken tesaurus eller synonymordbok lyckades bättre än strategin där enbart böjningsformer utgjorde expansionen. Expansionen med böjningsformer kan ses som ungefär motsvarande trunkering. Resultaten visar också att expansion med snävare termer var marginellt bättre än expansion med synonymer. Detta till skillnad från resultaten i undersökningen av Kristensen & Järvelin (1990) där expansionen med synonymer var den bäst presterande gällande relativ recall. Kristensen & Järvelin drog också slutsatsen att en tesaurus som används vid söktillfället är effektiv för återvinningsresultatet, något som delvis kan bekräftas i föreliggande studie om man gör jämförelsen enbart mellan grundstrategin och s4-s6 där tesauren Allärs användes. Kristensens undersökning (1993) visade att kombinationen av expansion med synonymer, snävare och relaterade termer gav det bästa resultatet för relativ recall, medan detta inte är något som kan ses i föreliggande studie. Istället är det här expansionen med synonymer som gav bäst resultat jämfört med snävare, relaterade och bredare termer i kombination.

Att synonymerna från den synonymordbok som använts i denna studie presterade så pass dåligt kan även helt enkelt bero på att den synonymordbok som använts inte var bra eller passande i återvinningssyfte i förhållande till termerna i grundstrategin. En orsak kan vara att det rör sig om en del specialtermer och termer som används mest i skrift vilka inte har några uppslag eller motsvarigheter i synonymordboken. Mot detta kan ställas att det dock heller inte finns synonymer till mer vanligt förekommande termer som t ex slutspel, landning eller skogsbrand.

En möjlig orsak här kan även vara att det, oavsett val av källa, inte förekommer synonymer för vissa termer. Synonymordbokens syfte är att erbjuda ett brett spektra av termer för flertalet olika sammanhang och användare, och det kan här konstateras att användningen av den inte gav bra resultat i denna studie. Detta kan även ses som att termerna i den initiala sökfrågan var bra återvinningsmässigt men alltför specifika och därmed inte bra gällande att finna synonymer eller andra semantiskt relaterade expansionstermer. Resultaten överensstämmer därmed delvis med Voorhees (1994a; 1994b) som i sina studier visade att QE inte var fördelaktigt, i hennes studie kanske främst pga. att termerna i den initiala sökfrågan var alltför specifika och för att hela de initiala sökfrågorna var jämförelsevis komplexa. Voorhees kunde också konstatera att dessa initiala termer som var bra ur återvinningssynpunkt inte hade några motsvarigheter eller näraliggande termer i hennes tesaurus, vilket även stämmer med föreliggande studie gällande tesaurus och synonymordbok.

- Hur presterar de fem strategierna avseende genomsnittlig precision i förhållande till grundstrategin?

Grundstrategin gav ett genomsnitt på 37,39%. I jämförelse visar detta att grundstrategin ändå, genomsnittlig sett, var relativt bra på att placera relevanta dokument bland de tio först visade. Bland expansionsstrategierna var det även här s2 som presterade bäst med 45,18%. I fallande ordning kom sedan s4 (43,37%), s5 (40,90%), s6 (39,56%) och s3 (36,11%). För fyra topics blev resultaten 0% från s1-6, detta gäller topic 017, 045, 046 samt 051. Problematiken kring topic 045 och 046 har berörts i ovanstående diskussion, medan 017 och 051 tycks ha andra orsaker. De termer som ingår i 017 är egentligen bara en eftersom den andra termen är ett egennamn. Inga synonymer eller snävare termer fanns här och att inga relevanta dokument har lyckats hamna bland de 10 första kan bero på att den använda söktermen *skogsbrand* kan förekomma i många sammanhang i dokumenten och därmed inte är bra på att skilja ut relevanta dokument från icke relevanta sådana. För söktermerna i topic 051 hittades inga relaterade termer, och för synonymer samt snävare och bredare termer saknades alternativ för alla termer i sökfrågan. De expansionstermer som finns med är relativt breda och håller sig inte inom ämnet vilket har påverkat precisionen negativt.

Även här bör man vara uppmärksam på att de genomsnittliga talen inte helt rättvist återspeglar resultaten. För s1 är så många som 8 av 21 resultat 0% vilket drar ned genomsnittet. För 2 av 21 topics är dock den genomsnittliga precisionen 100% redan i s1. Här gäller samma som ovanstående att termerna i den initialt ställda sökfrågan är relativt specifika och snäva vilket gör att de jämförelsevis är bäst ur återvinningssynpunkt. Det gör att även här lyser svårigheterna med att hitta expansionstermer igenom i resultaten. Om man antar att resultaten delvis beror på att termerna i grundstrategin är relativt specifika borde ändå s6 med bredare termer prestera bättre än näst sämst av samtliga. Även s5 med relaterade termer borde ha presterat bättre. Att s6 presterar så pass dåligt kan bero på kombinationen, dvs. att de snävare och relaterade termerna hänger med även här och därför drar ner resultatet. Hur förklarar man då att även s5 visar dåliga resultat? Även här kan det vara som tidigare beskrevs att problematiken med att finna expansionstermer till termerna i grundstrategin kvarstår, samt att de källor som expansionstermerna hämtades från inte var passande för denna studie.

Tvärtemot Greenberg (2001a) där precisionen var bäst gällande expansionsstrategin med synonymer är det i föreliggande studie sämst precision för synonymexpansionen. Detta kan ha sina grunder i, som även tidigare påpekats, att det här återfinns expansionstermer som hamnar utanför ämnet, som är mycket ovanliga eller äldre termer som knappast används idag. Att det är så har sin naturliga förklaring i synonymordbokens upplägg och syfte. Ett synnerligen anmärkningsvärt exempel på synonymordbokens bredd är i topic 032 där synonymer till *kvinnor* återfinns. Knappast några av dessa termer känns igen från dagspressen och även om detta exempel är väl ytterligt återkommer liknande mönster i flera andra topics vilket därmed troligen påverkat resultaten och dragit ner precisionen.

Inte heller Allärs var särskilt användbar här, genomsnittstalen visar dock att s4-6 presterade bättre än s3 och s1, samt att s4 med snävare termer fått högre precision än s5 och s6. Att snäva expansionstermer som i s4 fick hög precision var väntat, men inte ens expansionen med snävare termer lyckas bättre än böjningsformerna vilket kan tyda på att de initiala sökfrågetermerna redan är specifika och relativt snäva.

- Finns det skillnader i återvinningseffektivitet mellan grundstrategin och de fem expansionsstrategierna beträffande genomsnittlig precision och relativ recall?

Gentemot grundstrategin skiljer sig resultaten för relativ recall markant med nästan 14 procentenheter. Grundstrategin visar genomsnittligt sett lägst resultat av samtliga och till nästföljande strategi (s6) gällande effektivitetsmåttet är det nästan 10 procentenheter. Här kan således konstateras att alla expansionsstrategierna presterade bättre än grundstrategin och att både synonymordbok och tesaurus sett ur detta perspektiv var effektiva för återvinningen. Dock måste det genomsnittliga bästa resultatet ses som jämförelsevis lågt med dryga 50%. I föreliggande studie uppnås inte i mer än ett enda topic relativ recall på 100%. Detta kan bero på avsaknaden av expansionstermer som gjort att inget topic kunnat expanderas komplett, samt att söktermerna fått för låga beliefvärden och därmed hamnat utanför det givna intervallet. För genomsnittlig precision är skillnaderna dock inte lika stora: resultatet för grundstrategin är 37,39% vilket därmed är bättre än den sämsta strategins resultat som är s3 med 36,11%. Att s3 presterar sämre än grundstrategin kan bero på flertalet orsaker; dels det som tidigare nämnts att termerna i grundstrategin redan är så specifika och snäva att precisionen blir relativt hög här och dels kan det bero på att expansionen med synonymerna i s3 inte är bra på att placera de återvunna relevanta inom de 10 först visade. Skillnaden är inte heller så stor mellan dessa utan endast drygt en procentenhet.

- Om skillnader finns, vilken strategi presterar bäst och vad kan det bero på?

För både relativ recall och genomsnittlig precision är det strategi 2 med böjningsformer som visar bäst resultat sett till genomsnitten. Det mest vanligt förekommande mönstret att precisionen sjunker när recall ökar stämmer inte om man ser till genomsnittstalen. Istället ökar både precision och recall från grundstrategin s1 till s2. Att det är just s2 med böjningsformer som ger bäst resultat och vad detta kan bero på har redan diskuterats och en möjlig orsak är att synonymordboken och tesauren är inte lämpade för detta ändamål. Om man ser till synonymordboken är denna uppbyggd på ett sätt så att de termer som visas inte nödvändigtvis måste vara synonymer i ordets mest strikta betydelse, utan dessa termer kan vara både bredare, snävare och relaterade. Det borde ge bra resultat återvinningsmässigt, vilket det dock inte gjort. De termer som finns med i synonymordboken är förutom de allra mest självklara synonymer även en hel del ovanliga former av ord och vissa mycket sällan använda termer, och framförallt sällan använda i tidningsartiklar som ovan konstaterats. Ser man t ex till synonymexpansionen av termen *olycka* i topic 022, som handlar om flygplansolyckor, återfinns termer som *kalamitet*, *ofärd*, *malör* och *otrevlig händelse.* Det är mindre troligt att en flygplansolycka i en dagstidningsartikel beskrivs med någon av dessa termer. Detta gör då att de dokument som trots allt är relevanta, men inte innehåller några av dessa termer, petas ner och får låga beliefvärden. Det gör i sin tur att de kanske hamnar utanför satta intervall. Att synonymordboken presenterar för detta sammanhang relativt opassande termer kan även ses i flera topics, t ex topic 032 och 053. I hela 13 fall av 21, nästan 62% saknas synonymer till en eller flera av termerna från grundstrategin. Även detta visar på att synonymordboken inte varit behjälplig för QE i denna studie.

Gällande tesauren lyckades inte heller expansionsstrategierna s4-6 med termer från denna att överstiga återvinningseffektiviteten i s2. Allärs är en allmän tesaurus varför man inte heller kan vänta sig att finna uppslagstermer inom mycket specifika ämnen, och än mindre då snävare, bredare eller relaterade termer. Tesauren är avsedd att användas för indexering och sökning av böcker, artiklar och andra typer av material men det kan konstateras att den i detta

sammanhang gav dåliga resultat. Detta också trots att den jämfört med den tidigare forskningens använda tesaurer är relativt stor med sina 16 000 ämnesord. I 18 fall av 21 saknas snävare termer till en eller flera av termerna i grundstrategin, samma siffra gäller för bredare termer. I 10 fall av 21 saknas också relaterade termer till en eller flera av termerna i grundstrategin.

I två fall gjordes inga sökningar på grund av att Allärs erbjöd ett mycket stort antal snävare termer till ”företag” och ”sjukdomar”, 35 respektive 58 st. Bland snävare termer till företag räknar Allärs upp en mängd olika typer av företag (se Appendix 3 för samtliga av dessa termer). Samma fenomen dyker upp när det gäller snävare termer till sjukdomar där 58 st olika typer av sjukdomar räknas upp. Dessa två fall är de enda i undersökningen där Allärs uppvisar dessa typer av snävare termer, och de har därför inte tagits med. Vad detta beror på och varför det endast uppträder vid ett fåtal termer i Allärs är inte en fråga för denna uppsats, utan här kan bara konstateras att detta sätt att ange snävare termer inte var passande för denna studie. På Allärs webbplats ges heller ingen närmare förklaring eller beskrivning utan endast att de hierarkiskt ordnade termerna skall vara *”… aspekter på ett och samma ämne”* (Allärs 2004).

Ytterligare aspekter som påverkat resultaten rör testkollektionen och dess utformning. Indexeringen möjliggör inte trunkering varför varje böjningsform måste skrivas ut för att man skall kunna söka på den. I och med det finns risken att man går miste om relevanta dokument som eventuellt hade kunnat hittas vid trunkering. Detta dels eftersom genitiv uteslutits ur böjningsformerna (förutom gällande egennamn och geografiska namn) och dels genom att sammansatta termer skulle kunna förekomma där söktermerna ingår. Kollektionen består också av kortare dokument i form av tidningsartiklar, och som konstaterades av Lehtokangas & Järvelin (2001) verkar kortare artiklar vara mer konsekventa gällande val av termer vilket naturligtvis kan påverka återvinningsresultatet om man misslyckas med att pricka in dessa termer. Här spelar även orsaker som dokumentets längd in eftersom ett dokument som är kort och innehåller en term många gånger ses som bra ur återvinningsperspektiv. Kollektionen består som nämnts av tidningsartiklar och även det kan vara en trolig orsak till resultatet; det gäller vad man skulle kunna kalla ett tidningsspråk eller skillnader i språket mellan tal och text. Klara skillnader finns här och det kan mycket väl vara så att termer används i dagspressen som inte vanligtvis nyttjas i vardagsspråk eller för den delen inte heller används i synonymordböcker eller tesaurus. Detta har diskuterats tidigare. Uppmärksamhet bör även riktas mot att skillnader skulle kunna förekomma mellan en dagsaktuell tidning och en synonymordbok med några år på nacken vilken inte kan uppdateras löpande eller ta in aktuella termer. Dock bör detta inte ha varit något problem i denna studie eftersom tidningsartiklarna är från år 1994 och synonymordboken från år 2000 vilket bör ha gjort att nyare termer hunnit komma med.

Hur är det då med reliabilitet och validitet gällande de erhållna resultaten? Kan resultat och metod såsom de använts i denna studie ge underlag för några generella slutsatser i ett större sammanhang? Svaret på denna fråga måste bli nej. Istället bör man se det som att för just *denna studie* av QE var varken synonymordbok eller tesaurus behjälplig för att finna bra expansionstermer.

Gällande antalet informationsbehov som användes i denna studie var de 21 st, och det verkar troligt att även ett större antal skulle ge liknande resultat om övriga faktorer bibehölls. Det som sannolikt inträffar om antalet informationsbehov ökas är att medelvärdet inte påverkas lika mycket av enskilda resultat med ytterliga värden. Syftet från början var att ta de 26

informationsbehov med flest antal relevanta dokument, detta tal sjönk dock sedan till 21 st. Att inte fler än dessa valdes ut beror på, som beskrevs i metodkapitlet, att de resterande informationsbehoven skapade för denna kollektion uppvisade mycket låga antal relevanta dokument. Detta gör då att QE blir relativt ointressant om det för ett informationsbehov endast finns två eller tre relevanta dokument. Här kan även framhållas att i den tidigare forskningen har antalen informationsbehov varit lägre: Kristensen & Järvelin (1990) lyckades fullfölja samtliga expansionsstrategier för 18 informationsbehov och för Kristensen (1993) fullföljdes endast 14 st. Dock bör här understrykas att av de 21 informationsbehov som användes här kunde inte en enda fullföljas för samtliga termer och expansionsstrategier vilket skiljer sig från den tidigare forskningen. Här kommer även längden på de dokument som ingår i testkollektionen in, de är som tidigare beskrivits av varierande längd från några meningar upp till ett par tre sidor. Trots att inga exakta uppgifter finns här står det ändå klart att samtliga dokument är artiklar av kortare art vilket, med Lehtokangas & Järvelins (2001) tankegångar, skulle kunna tyda på att variationen bland termerna inte är så stor.

Skulle resultaten då vara desamma om kollektionens storlek ändrades? Om samtliga omkringliggande faktorer var konstanta skulle resultaten med största sannolikhet bli likartade, detta eftersom det redan rör sig om en så pass stor kollektion. Antalet dokument i samlingen, 161 336 st, står sig också väl i sammanhanget jämfört med tidigare forskning.

Att återvinningstekniken påverkar resultaten framhölls i början av denna uppsats. Matchningstekniken i IR-systemet har också beskrivits kortfattat i metodavsnittet. Vad som kan sägas om detta är att de operatorer som använts visserligen är specifika för InQuery men att deras principer dock gäller generellt. Därför borde även resultaten kunna sägas vara tillförlitliga i denna fråga och liknande resultat bör kunna uppnås under samma premisser.

## 7.1 Konklusion och sammanfattning av analysen

De genomsnittliga resultaten finns sammanfattade i nedanstående tabell (tabell 4). Här syns att strategi 2 presterade bäst vad gäller både relativ recall och genomsnittlig precision.

| | s1 | s2 | s3 | s4 | s5 | s6 |
|---|---|---|---|---|---|---|
| Recall | 37,72 | 51,71 | 48,98 | 49,74 | 48,09 | 47,22 |
| Precision | 37,39 | 45,18 | 36,11 | 43,37 | 40,90 | 39,56 |

*Tabell 4. Medelvärden i procent för relativ recall och genomsnittlig precision.*

Därmed kan konstateras att gentemot grundstrategin lyckades strategierna 3, 4, 5 och 6 bra där termerna hämtades från synonymordboken och tesauren. Hjälpmedlen kan ur detta perspektiv ses som effektiva ur återvinningssynpunkt. Ett undantag från detta är dock s3 som presterade sämre än grundstrategin gällande genomsnittlig precision. Jämfört med s2 där endast böjningsformer av termerna i s1 användes som expansion står sig dock inte synonymordboken och tesauren som effektiva hjälpmedel för QE. Orsakerna till dessa resultat kan vara följande:

- termerna i grundstrategin, som hämtades ur de sedan tidigare formulerade informationsbehoven, är från början relativt specifika och snäva.

- som en följd av ovanstående skapar detta problem för expansionsstrategierna s3-6 eftersom specifika och snäva termer dels inte har några motsvarigheter och dels inte finns som uppslagstermer i synonymordboken eller tesauren.

- synonymordboken och tesauren är kollektionsoberoende och därmed inte anpassade för att användas i detta sammanhang, detta blir tydligt eftersom det inte för något informationsbehov har gått att skapa kompletta strategier från 1-6.

- eftersom trunkering inte är möjlig på grund av testkollektionens utformning kan eventuellt relevanta dokument ha missats. Detta genom att endast obestämd och bestämd form samt singular och plural tagits med som böjningsformer och inte några genitivformer förutom vid egennamn och geografiska platser.

Slutsatsen blir därmed att varken synonymordboken eller tesauren som användes här var bra som termkällor och därmed var de inte heller effektiva ur återvinningssynpunkt. Det vore intressant med fortsatt forskning inom detta område, under liknande premisser, för att se om snarlika resultat uppnås där trunkering är möjlig. Även att se huruvida andra termkällor står sig i jämförelse med dem som använts i denna studie vore intressant. Området med semantiskt relaterade termer inom QE visar delvis motstridiga resultat vilket också det gör att fortsatt forskning vore givande.

# 8. Sammanfattning

Query expansion (QE) är ett forskningsområde inom information retrieval (IR), där en sökfråga utökas med nya termer i syfte att öka graden av recall och återvinna fler relevanta dokument. Syftet med denna uppsats har varit att utvärdera återvinningseffektiviteten hos fem strategier för automatisk query expansion. Studiens syfte uttrycktes i följande frågeställningar:

- Hur presterar de fem strategierna avseende relativ recall i förhållande till grundstrategin?

- Hur presterar de fem strategierna avseende genomsnittlig precision i förhållande till grundstrategin?

- Finns det skillnader i återvinningseffektivitet mellan grundstrategin och de fem expansionsstrategierna beträffande genomsnittlig precision och relativ recall?

- Om skillnader finns, vilken strategi presterar bäst och vad kan det bero på?

Query expansion brukar indelas in i tre typer: manuell, interaktiv och automatisk expansion. Inom manuell QE agerar användaren själv utan assistans från systemet. Interaktiv QE sker genom att system och användare på något sätt interagerar, t ex genom att systemet visar potentiella expansionstermer. I automatisk QE låter man systemet utföra expansionen utan inverkan av användaren. Denna uppsats har studerat automatisk QE där dock utformningen av expansionsstrategierna skett manuellt. Strategierna för expansionen är alltigenom likartad och genomförd på samma grunder genom hela studien. Inom QE berörs frågor som varifrån expansionstermerna skall hämtas och hur de skall utses. I denna uppsats har expansionstermerna hämtats från kollektionsoberoende källor där samtliga presenterade termer har använts i expansionen. Den tidigare forskningens metoder och resultat inom detta område är inte helt samstämmiga men merparten visar att QE med semantiskt relaterade termer är återvinningseffektivt och ökar recall, dock inte i fall där den initiala sökfrågan är komplex och välformulerad.

Studien utfördes i det probabilistiska IR-systemet InQuery. Testkollektionen bestod av 161 336 fulltextdokument i form av tidningsartiklar från Göteborgs Posten och Helsingborgs Dagblad från år 1994. För att beräkna relativ recall samt genomsnittlig precision användes verktyget Query Performance Analyser (QPA). En grundstrategi samt fem expansionsstrategier formulerades för att testa återvinningseffektiviteten.

- Grundstrategin (s1) skapades enbart med termer från de redan formulerade informationsbehoven tillhörande kollektionen.
- Strategi två (s2) utgjordes av böjningsformer av dessa termer genom singular och plural i bestämd och obestämd form. För geografiska namn samt namn på personer lades enbart genitivformen till här.
- Strategi tre (s3) formades av synonymer hämtade från en synonymordbok,
- strategi fyra (s4) av snävare termer från en allmän tesaurus,
- strategi fem (s5) av snävare samt relaterade termer från samma tesaurus samt slutligen
- strategi sex (s6) med kombinationen snävare, relaterade och bredare termer från tesauren.

Genom dessa strategier formulerades sedan sökfrågor för 21 informationsbehov. Resultaten, här presenterade i form av genomsnittliga procentsatser (tabell 5), visade att både avseende relativ recall och genomsnittlig precision gav expansionen med böjningsformer i strategi 2 bäst resultat. Jämfört med grundstrategin presterade s6 sämst gällande relativ recall och s3 var t o m sämre än grundstrategin avseende genomsnittlig precision.

| | s1 | s2 | s3 | s4 | s5 | s6 |
|---|---|---|---|---|---|---|
| Recall | 37,72 | 51,71 | 48,98 | 49,74 | 48,09 | 47,22 |
| Precision | 37,39 | 45,18 | 36,11 | 43,37 | 40,90 | 39,56 |

*Tabell 5. Medelvärden i procent för relativ recall och genomsnittlig precision.*

Sett till dessa genomsnittliga resultat kan sägas att QE i denna studie var återvinningseffektivt med ett undantag där s3 presterade sämre än grundstrategin för genomsnittlig precision. Att s2 med böjningsformerna presterade bäst av samtliga gör dock att varken synonymordbok eller tesaurus var övertygande effektiva som hjälpmedel för QE i denna undersökning. Vad detta kan bero på sammanfattas i ett antal punkter innehållande möjliga och troliga orsaker:

- termerna i grundstrategin, som hämtades ur de sedan tidigare formulerade informationsbehoven, var från början relativt specifika och snäva.

- som en följd av ovanstående skapade det problem för expansionsstrategierna s3-6 eftersom specifika och snäva termer dels inte har några motsvarigheter och dels inte fanns som uppslagstermer i de använda hjälpmedlen synonymordbok och tesaurus.

- synonymordboken och tesauren är kollektionsoberoende och därmed inte anpassade för att användas i detta sammanhang, detta blev tydligt eftersom det inte för något informationsbehov har gick att skapa kompletta strategier från s1-6 med samtliga termer.

- även testkollektionens utformning inverkade på resultaten: eftersom trunkering inte var möjlig kan eventuellt relevanta dokument ha missats. Detta genom att endast obestämd och bestämd form samt singular och plural tagits med som böjningsformer och inte några genitivformer förutom vid egennamn och geografiska platser. Att dokumenten i samlingen utgjordes av tidningsartiklar i kortare form kan även det ha inverkat på resultaten.

Resultaten stämmer delvis överens med den tidigare forskningen även om de genomsnittstal som här räknats fram för denna studie inte är helt rättvisa. Fortsatt forskning vore intressant gällande området inom QE med semantiskt relaterade termer, detta eftersom många av de undersökningar som gjorts visar motstridiga resultat.

# 9. Referenslitteratur


Ahlgren, Per & Eklund, Johan (2003). *Manual för Query Performance Analyser*. Kursmaterial för kursen “Information Retrieval”, 5p. Borås: Bibliotekshögskolan.

Allärs: Allmän tesaurus på svenska
http://vesa.lib.helsinki.fi/allars/index.html [Hämtad 2004-03-03]

Baeza-Yates, Ricardo & Ribeiro-Neto, Berthier (1999). *Modern information retrieval*. Harlow: Addison-Wesley.

Chowdhury, G. G. (1999). *Introduction to modern information retrieval*. London: Library Association Publishing.

Efthimiadis, Efthimis, N. (1996). Query expansion. Ingår i: Williams, Martha E., ed. (1996) *Annual Review of Information Science and Technology (ARIST)*, vol. 31, Medford, NJ: Information Today. s. 121-187.

Greenberg, Jane (2001a). Automatic query expansion via lexical-semantic relationships. *Journal of the American Society for Information Science and Technology (JASIST)*, vol. 52, no. 5, s. 402-415.

Greenberg, Jane (2001b). Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology. *Journal of the American Society for Information Science and Technology (JASIST)*, vol. 52, no. 6, s. 487-498.

Harter, Stephen P. & Hert, Carol A. (1997). Evaluation of information retrieval systems: approaches, issues and methods. Ingår i: Williams, Martha E., ed. (1997) *Annual review of information science and technology (ARIST)*, vol. 32, Medford, NJ: Information Today, s. 3-94.

Jansen, Bernard, J., Spink, Amanda, & Saracevic, Tefko (2000). Real life, real users and real needs: a study and analysis of user queries on the web. *Information Processing and Management*, vol. 36, no. 2, March 2000, s. 207-227.

Kekäläinen, Jaana (1999). *The effects of query complexity, expansion, and structure on retrieval performance in probabilistic text retrieval*, Acta Universitatis Temperensis 678, Diss. Department of Information Studies, University of Tampere.

Kristensen, Jaana (1993). Expanding end-users' query statements for free text searching with a search-aid thesaurus, *Information Processing and Management*, vol. 29, no. 6, s. 733-744.

Kristensen, Jaana & Järvelin, Kalervo (1990). The effectiveness of a searching thesaurus in free-text searching in a full-text database. *International Classification*, vol. 17, no. 2, s. 77-84.

Kärki, Eeva (2004). Kontaktperson för YSA och Allärs. E-postmeddelande 2004-03-15.

Lehtokangas, Raija & Järvelin, Kalervo (2001). Concistency of textual expression in newspaper articles: an argument for semantically based query expansion, *Journal of Documentation*, vol. 57, no. 4, July 2001, s. 535-548.

Maron, M. E. & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *Association for Computing Machinery*, 7(3), s. 216-244.

Sormunen, Eero (2002). Liberal relevance criteria of TREC - counting on negligible documents? Ingår i: *Proceedings of the 25th annual ACM-SIGIR International Conference on Research and Development in Information Retrieval*. M. Beaulieu, R. Baeza-Yates, S. H. Myaeng, K. Järvelin, eds. ACM: New York. s. 324-330.

Svenska akademiens ordlista över svenska språket (1998). Stockholm: Norstedts.

Spink, Amanda, et. al. (2000). Searching the web, the public and their queries. Wolfram, Dietmar., Jansen, Bernard, J., & Saracevic, Tefko. *Journal of the American Society of Information Science and Technology (JASIST)*, vol. 52, no. 3, s. 227-234

Strömberg, Alva (1998). *Nya stora synonymordboken: 2000-talets synonymordbok*. Andra helt rev. uppl. Stockholm: Strömbergs.

*Svensk synonymordbok* (1992). Stockholm: Norstedts.

Walter, Göran (2003). *Bonniers synonymordbok*. 3. rev. uppl. Stockholm: Bonniers.

Voorhees, Ellen M. (1994a). On expanding query vectors with lexically related words. Ingår i: *Proceedings of the Second Text REtrieval Conference (TREC-2)*, Donna K. Harman ed., NIST Special Publication 500-207, March 1994. s. 223-231.

Voorhees, Ellen M. (1994b). Query expansion using lexical-semantic relations. Ingår i: *Sigir 94: Proceedings of the seventeenth annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University, 3-6 July 1994, Dublin, Ireland*. Ed. by Croft, Bruce W. & van Rijsbergen, C.J. London: Springer.

# Appendix 1: Topics och sökfrågor

Nedan återfinns samtliga topics med tillhörande sökfrågor för de olika strategierna. I de fall där expansionstermer saknas är detta markerat med fetstil i anslutning till den aktuella sökfrågan.

**Nr 238. Topic 005**: *Undersök attityderna hos icke-medlemsländer beträffande framtida medlemskap i Europeiska Gemenskapen eller Europeiska Unionen.*

1. #sum(medlemskap #od1(europeiska unionen))

2. #sum(#syn(medlemskap medlemskapet medlemskapen) #syn(#od1(europeiska unionen) #od1(europeiska unionens)))

3. #sum(#syn(medlemskap medlemskapet medlemskapen) #syn(#od1(europeiska unionen) #od1(europeiska unionens) EU))
**Inga synonymer till medlemskap**

4. #sum(#syn(medlemskap medlemskapet medlemskapen) #syn(#od1(europeiska unionen) #od1(europeiska unionens)))
**Inga snävare**

5. #sum(#syn(medlemskap medlemskapet medlemskapen) #syn(#od1(europeiska unionen) #od1(europeiska unionens)) #syn(medlemskår medlemskåren medlemskårer medlemskårerna medlemsstater medlemsstaterna medlemsstat medlemsstaten))

6. #sum(#syn(medlemskap medlemskapet medlemskapen) #syn(#od1(europeiska unionen) #od1(europeiska unionens)) #syn(medlemskår medlemskåren medlemskårer medlemskårerna medlemsstater medlemsstaterna medlemsstat medlemsstaten))
**Inga bredare**

**Nr 239. Topic 007:** *Sök dokument om dopinganvändning inom fotbollen.*

1. #sum(doping fotboll)

2. #sum(#syn(doping dopingen) #syn(fotboll fotbollen))

3. #sum(#syn(doping dopingen dopning dopningen #od1(användning av stimulerande medel) #od1(användningen av stimulerande medel) #od1(bruk av stimulantia) #od1(bruket av stimulantia) #od1(otillåten stimulering) #od1(otillåtna stimuleringen) #od1(otillåtna stimuleringar) #od1(otillåtna stimuleringarna) droganvändning droganvändningen dopande dopandet) #syn(fotboll fotbollen bollsport bollsporten bollsporter bollsporterna lagspel lagspelet lagspelen fotbollsmatch fotbollsmatchen fotbollsmatcher fotbollsmatcherna))

4. #sum(#syn(doping dopingen) #syn(fotboll fotbollen innefotboll innefotbollen))
**Inga snävare till doping**

5. #sum(#syn(doping dopingen #od1(anabola steroider) #od1(anabola steroiderna) #od1 (anabol steroid) #od1(anabola steroiden) antidopning antidopningen) #syn(fotboll fotbollen innefotboll innefotbollen))
**Inga relaterade till fotboll**

6. #sum(#syn(doping dopingen #od1(anabola steroider) #od1(anabola steroiderna) #od1 (anabol steroid) #od1(anabola steroiden) antidopning antidopningen) #syn(fotboll fotbollen innefotboll innefotbollen bollsport bollsporten bollsporter bollsporterna lagspel lagspelet lagspelen))
**Inga bredare till doping**

**Nr 241. Topic 011**: *Sök dokument som behandlar den nya författningen i Sydafrika och dess slutgiltiga struktur.*

1. #sum(författningen Sydafrika)

2. #sum(#syn(författningen författning författningar författningarna) #syn(Sydafrika Sydafrikas))

3. #sum(#syn(författningen författning författningar författningarna grundlag grundlagen grundlagar grundlagarna konstitution konstitutionen konstitutioner konstitutionerna regeringsform regeringsformen regeringsformer regeringsformerna styrelsesätt styrelsesättet styrelsesätten) #syn(Sydafrika Sydafrikas))

4. #sum(#syn(författningen författning författningar författningarna förordning förordningen förordningar förordningarna lag lagen lagar lagarna) #syn(Sydafrika Sydafrikas))

5. #sum(#syn(författningen författning författningar författningarna förordning förordningen förordningar förordningarna lag lagen lagar lagarna direktiv direktivet direktiven #od1(rättslig reglering) #od1(rättsliga regleringen) #od1(rättsliga regleringar) #od1(rättsliga regleringarna) rättsnorm rättsnormen rättsnormer rättsnormerna) #syn(Sydafrika Sydafrikas))

6. #sum(#syn(författningen författning författningar författningarna förordning förordningen förordningar förordningarna lag lagen lagar lagarna direktiv direktivet direktiven #od1(rättslig reglering) #od1(rättsliga regleringen) #od1(rättsliga regleringar) #od1(rättsliga regleringarna) rättsnorm rättsnormen rättsnormer rättsnormerna) #syn(Sydafrika Sydafrikas))
**Inga bredare**

**Nr 242. Topic 012:** *Sök dokument om Soltempelorden och mordet eller självmordet på dess medlemmar.*

1. #sum(soltempelorden självmord)

2. #sum(#syn(soltempelorden soltempelordens) #syn(självmord självmordet självmorden))

3. #sum(#syn(soltempelorden soltempelordens) #syn(självmord självmordet självmorden #od1(död för egen hand) #od1(dog för egen hand) självspillan suicidum suicid suicidal suicidalt suicidala harakiri harakirit harakirin))

4. #sum(#syn(soltempelorden soltempelordens) #syn(självmord självmordet självmorden))
**Inga snävare till självmord**

5. #sum(#syn(soltempelorden soltempelordens) #syn(självmord självmordet självmorden mord mordet morden självdestruktivitet självdestruktiviteten självmordsförsök självmordsförsöket självmordsförsöken))

6. #sum(#syn(soltempelorden soltempelordens) #syn(självmord självmordet självmorden mord mordet morden självdestruktivitet självdestruktiviteten självmordsförsök självmordsförsöket självmordsförsöken))
**Inga bredare till självmord**

**Nr 243. Topic 013:** *Vilka diskussioner fördes och vilka resolutioner antogs vid befolkningskonferensen om familjeplanering i Kairo?*

1. #sum(resolutioner befolkningskonferens Kairo)

2. #sum(#syn(resolutioner resolutionerna resolution resolutionen) #syn(befolkningskonferens befolkningskonferensen befolkningskonferenser befolkningskonferenserna) #syn(Kairo Kairos))

3. #sum(#syn(resolutioner resolutionerna resolution resolutionen #od1(samfällt uttalande) #od1(samfällda uttalandet) #od1(samfällda uttalanden) #od1(samfällda uttalandena) #od1(gemensamt uttalande) #od1(gemensamma uttalandet) #od1(gemensamma uttalanden) #od1(gemensamma uttalandena) uttalande uttalandet uttalanden uttalandena) #syn(befolkningskonferens befolkningskonferensen befolkningskonferenser befolkningskonferenserna) #syn(Kairo Kairos))
**Inga synonymer till befolkningskonferens**

4. #sum(#syn(resolutioner resolutionerna resolution resolutionen) #syn(befolkningskonferens befolkningskonferensen befolkningskonferenser befolkningskonferenserna) #syn(Kairo Kairos))
**Inga snävare**

5. #sum(#syn(resolutioner resolutionerna resolution resolutionen) #syn(befolkningskonferens befolkningskonferensen befolkningskonferenser befolkningskonferenserna) #syn(Kairo Kairos))
**Inga relaterade**

6. #sum(#syn(resolutioner resolutionerna resolution resolutionen) #syn(befolkningskonferens befolkningskonferensen befolkningskonferenser befolkningskonferenserna) #syn(Kairo Kairos))
**Inga bredare**

**Nr 244. Topic 015:** *Vilka faktorer undergräver den europeiska industrins konkurrenskraft på världsmarknaden?*

1. #sum(Europa industri konkurrens världsmarknad)

2. #sum(#syn(Europa Europas) #syn(industri industrin industrier industrierna) #syn(konkurrens konkurrensen) #syn(världsmarknad världsmarknaden världsmarknader världsmarknaderna))

3. #sum(#syn(Europa Europas) #syn(industri industrin industrier industrierna näringsliv näringslivet företagsamhet företagsamheten råvaruförädling råvaruförädlingen masstillverkning masstillverkningen stordrift stordriften stordrifter stordrifterna industriföretag industriföretaget industriföretagen fabrik fabriken fabriker fabrikerna bruk bruket bruken verk verket verken tillverkningsföretag tillverkningsföretaget tillverkningsföretagen) #syn(konkurrens konkurrensen tävlan tävling tävlingen tävlingar tävlingarna kamp kampen kamper kamperna rivalitet rivaliteten rivaliteter rivaliteterna #od1(fri handel) #od1(fria handeln) #od1(fri företagsamhet) #od1(fria företagsamheten) #od1(kamp om kunder) #od1(kamp om kunderna) #od1(kampen om kunderna)) #syn(världsmarknad världsmarknaden världsmarknader världsmarknaderna))
**Inga synonymer till världsmarknad**

4. #sum(#syn(Europa Europas) #syn(industri industrin industrier industrierna) #syn(konkurrens konkurrensen #od1(illojal konkurrens) #od1(illojala konkurrensen) offerttävling offerttävlingen offerttävlingar offerttävlingarna priskonkurrens priskonkurrensen) #syn(världsmarknad världsmarknaden världsmarknader världsmarknaderna))
**Inga snävare till industri, världsmarknad**

5. #sum(#syn(Europa Europas) #syn(industri industrin industrier industrierna #od1(medelstor industri) #od1(medelstora industrin) #od1(medelstora industrier) #od1(medelstora industrierna) #od1(liten industri) småindustri småindustrin småindustrier småindustrierna storindustri storindustrin storindustrier storindustrierna) #syn(konkurrens konkurrensen #od1(illojal konkurrens) #od1(illojala konkurrensen) offerttävling offerttävlingen offerttävlingar offerttävlingarna priskonkurrens priskonkurrensen konkurrensfördel konkurrensfördelen konkurrensfördelar konkurrensfördelarna konkurrenskraft konkurrenskraften konkurrenskrafter konkurrenskrafterna konkurrenskrafters konkurrensutsättning konkurrensutsättningen konkurrensutsättningar konkurrensutsättningarna marknaden marknad marknader marknaderna marknadsdominans marknadsdominansen monopol monopolet oligopol oligopolet) #syn(världsmarknad världsmarknaden världsmarknader världsmarknaderna #od1(internationell handel) #od1(internationella handeln) utrikeshandel utrikeshandeln))

6. #sum(#syn(Europa Europas) #syn(industri industrin industrier industrierna #od1(medelstor industri) #od1(medelstora industrin) #od1(medelstora industrier) #od1(medelstora industrierna) #od1(liten industri) småindustri småindustrin småindustrier småindustrierna storindustri storindustrin storindustrier storindustrierna) #syn(konkurrens konkurrensen konkurrensens #od1(illojal konkurrens) #od1(illojala konkurrensen) offerttävling offerttävlingen offerttävlingar offerttävlingarna priskonkurrens priskonkurrensen konkurrensfördel konkurrensfördelen konkurrensfördelar konkurrensfördelarna konkurrenskraft konkurrenskraften konkurrenskrafter konkurrenskrafterna

konkurrensutsättning konkurrensutsättningen konkurrensutsättningar konkurrensutsättningarna marknaden marknad marknader marknaderna marknadsdominans marknadsdominansen monopol monopolet oligopol oligopolet) #syn(världsmarknad världsmarknaden världsmarknader världsmarknaderna #od1(internationell handel) #od1(internationella handeln) utrikeshandel utrikeshandeln))
**Inga bredare till industri, konkurrens**

**Obs! Termen ”marknaden” ges som bredare term till världsmarknad men finns redan med som relaterad till konkurrens. Denna har därför inte tagits med två gånger utan finns endast som relaterad till konkurrens.**

**Nr 245. Topic 017**: *Vilken omfattning hade skogsbränderna utanför Sydney och hur allvarligt var läget?*

1. #sum(skogsbränder Sydney)

2. #sum(#syn(skogsbränder skogsbränderna skogsbrand skogsbranden #syn(Sydney Sydneys)))

3. #sum(#syn(skogsbränder skogsbränderna skogsbrand skogsbranden #syn(Sydney Sydneys))
**Ingen synonym**

4. #sum(#syn(skogsbränder skogsbränderna skogsbrand skogsbranden #syn(Sydney Sydneys))
**Inga snävare**

5. #sum(#syn(skogsbränder skogsbränderna skogsbrand skogsbranden skogsförstörelse skogsförstörelsen skogsförstörelser skogsförstörelserna) #syn(Sydney Sydneys))

6. #sum(#syn(skogsbränder skogsbränderna skogsbrand skogsbranden skogsförstörelse skogsförstörelsen skogsförstörelser skogsförstörelserna eldsvådor eldsvådorna eldsvåda eldsvådan markbränder markbränderna markbrand markbranden) #syn(Sydney Sydneys))

**Nr 249. Topic 022:** *Sök exempel på flygplansolyckor som inträffat på start- eller landningsbanan under pågående flygning.*

1. #sum(flygplan olyckor start landning)

2. #sum(#syn(flygplan flygplanet flygplanen) #syn(olyckor olyckorna olycka olyckan) #syn(start starten starter starterna) #syn(landning landningen landningar landningarna))

3. #sum(#syn(flygplan flygplanet flygplanen flygmaskin flygmaskinen flygmaskiner flygmaskinerna flyg flyget flygen maskin maskinen maskiner maskinerna kärra kärran kärror

kärrorna plan planet planen aeroplan aeroplanet aeroplanen flygfarkost flygfarkosten flygfarkoster flygfarkosterna luftfarkost luftfarkosten luftfarkoster luftfarkosterna luftfartyg luftfartyget luftfartygen) #syn(olyckor olyckorna olycka olyckan olyckshändelse olyckshändelsen olyckshändelser olyckshändelserna olycksfall olycksfallet olycksfallen ofall missöde missödet missöden missödena kalamitet kalamiteten kalamiteter kalamiteterna våda vådan vådor vådorna tragedi tragedin tragedier tragedierna haveri haveriet haverier haverierna krasch kraschen krascher krascherna katastrof katastrofen katastrofer katastroferna fördärv fördärvet fördärven ofärd ofärden förstörelse förstörelsen förstörelser förstörelserna elände eländet eländen eländena misär misären misärer misärerna otur oturen motgång motgången motgångar motgångarna #od1(otrevlig händelse) #od1(otrevliga händelsen) #od1(otrevliga händelser) #od1(otrevliga händelserna) malör malören malörer malörerna tillbud tillbudet tillbuden) #syn(start starten starter starterna igångsättning igångsättningen igångsättningar igångsättningarna början begynnelse begynnelsen begynnelser begynnelserna avstamp avstampet grundande grundandet grundanden grundandena inkörsport inkörsporten inkörsportar inkörsportarna utgångspunkt utgångspunkten utgångspunkter utgångspunkterna avfärd avfärden avfärder avfärderna avresa avresan avresor avresorna avgång avgången avgångar avgångarna startplats startplatsen startplatser startplatserna) #syn(landning landningen landningar landningarna))
**Inga synonymer till landning**

4. #sum(#syn(flygplan flygplanet flygplanen militärflygplan militärflygplanet militärflygplanen segelflygplan segelflygplanet segelflygplanen) #syn(olyckor olyckorna olycka olyckan drunkning drunkningen drunkningar drunkningarna eldsvådor eldsvådorna eldsvåda eldsvådan explosionsolyckor explosionsolyckorna explosionsolycka explosionsolyckan gruvolyckor gruvolyckorna gruvolycka gruvolyckan katastrofer katastroferna katastrof katastrofen kemikalieolyckor kemikalieolyckorna kemikalieolycka kemikalieolyckan miljöolyckor miljöolyckorna miljöolycka miljöolyckan naturkatastrofer naturkatastroferna naturkatastrof naturkatastrofen skeppsbrott skeppsbrottet skeppsbrotten storolyckor storolyckorna storolycka storolyckan trafikolyckor trafikolyckorna trafikolycka trafikolyckan) #syn(start starten starter starterna) #syn(landning landningen landningar landningarna))
**Inga snävare till start, landning**

5. #sum(#syn(flygplan flygplanet flygplanen militärflygplan militärflygplanet militärflygplanen segelflygplan segelflygplanet segelflygplanen) #syn(olyckor olyckorna olycka olyckan drunkning drunkningen drunkningar drunkningarna eldsvådor eldsvådorna eldsvåda eldsvådan explosionsolyckor explosionsolyckorna explosionsolycka explosionsolyckan gruvolyckor gruvolyckorna gruvolycka gruvolyckan katastrofer katastroferna katastrof katastrofen kemikalieolyckor kemikalieolyckorna kemikalieolycka kemikalieolyckan miljöolyckor miljöolyckorna miljöolycka miljöolyckan naturkatastrofer naturkatastroferna naturkatastrof naturkatastrofen skeppsbrott skeppsbrottet skeppsbrotten storolyckor storolyckorna storolycka storolyckan trafikolyckor trafikolyckorna trafikolycka trafikolyckan olycksfall olycksfallet olycksfallen) #syn(start starten starter starterna) #syn(landning landningen landningar landningarna))
**Inga relaterade till flygplan, start, landning**

6. #sum(#syn(flygplan flygplanet flygplanen militärflygplan militärflygplanet militärflygplanen segelflygplan segelflygplanet segelflygplanen) #syn(olyckor olyckorna olycka olyckan drunkning drunkningen drunkningar drunkningarna eldsvådor eldsvådorna eldsvåda eldsvådan explosionsolyckor explosionsolyckorna explosionsolycka

explosionsolyckan gruvolyckor gruvolyckorna gruvolycka gruvolyckan katastrofer katastroferna katastrof katastrofen kemikalieolyckor kemikalieolyckorna kemikalieolycka kemikalieolyckan miljöolyckor miljöolyckorna miljöolycka miljöolyckan naturkatastrofer naturkatastroferna naturkatastrof naturkatastrofen skeppsbrott skeppsbrottet skeppsbrotten storolyckor storolyckorna storolycka storolyckan trafikolyckor trafikolyckorna trafikolycka trafikolyckan olycksfall olycksfallet olycksfallen) #syn(start starten starter starterna) #syn(landning landningen landningar landningarna))
**Inga bredare**

**Nr 252. Topic 026:** *Ge exempel på användning av vindkraft.*

1. vindkraft

2. #sum(#syn(vindkraft vindkraften vindkrafter vindkrafterna))

3. #sum(#syn(vindkraft vindkraften vindkrafter vindkrafterna))
**Inga synonymer**

4. #sum(#syn(vindkraft vindkraften vindkrafter vindkrafterna))
**Inga snävare**

5. #sum(#syn(vindkraft vindkraften vindkrafter vindkrafterna ekoenergi ekoenergin))

6. #sum(#syn(vindkraft vindkraften vindkrafter vindkrafterna ekoenergi ekoenergin energi energin energier energierna))

**Nr 254. Topic 032:** *Sök dokument om prästvigning av kvinnor inom de europeiska kyrkorna.*

1. #sum(prästvigning kvinnor Europa)

2. #sum(#syn(prästvigning prästvigningar prästvigningarna prästvigningen) #syn(kvinnor kvinnorna kvinna kvinnan) #syn(Europa Europas))

3. #sum(#syn(prästvigning prästvigningar prästvigningarna prästvigningen) #syn(kvinnor kvinnorna kvinna kvinnan dam damen damer damerna fruntimmer fruntimret kvinnfolk kvinnfolket kvinnfolken kvinnsperson kvinnspersonen kvinnspersoner kvinnspersonerna kvinns kvinnset lady ladyn madam madammen madammer madammerna tant tanten tanter tanterna kjoltyg kjoltyget kjoltygen tjej tjejen tjejer tjejerna donna donnan donnor donnorna brud bruden brudar brudarna brutta bruttan bruttor bruttorna krok kroken krokar krokarna tjack tjacket eva evan evor evorna femininum femininet femininer kärring kärringen kärringar kärringarna kärringarnas matrona matronan matronor matronorna kona konan konor konorna) #syn(Europa Europas))
**Inga synonymer till prästvigning**

4. #sum(#syn(prästvigning prästvigningar prästvigningarna prästvigningen) #syn(kvinnor kvinnorna kvinna kvinnan) #syn(Europa Europas))
**Inga snävare**

5. #sum(#syn(prästvigning prästvigningar prästvigningarna prästvigningen vigsel vigseln vigslar vigslarna) #syn(kvinnor kvinnorna kvinna kvinnan flickor flickorna flicka flickan) #syn(Europa Europas))

6. #sum(#syn(prästvigning prästvigningar prästvigningarna prästvigningen vigsel vigseln vigslar vigslarna) #syn(kvinnor kvinnorna kvinna kvinnan flickor flickorna flicka flickan) #syn(Europa Europas))
**Inga bredare**

**Nr 259. Topic 042:** *Sök dokument om FN-/USA-truppers invasion på Haiti.*

1. #sum(FN invasion Haiti)

2. #sum(FN #syn(invasion invasionen invasioner invasionerna) #syn(Haiti Haitis))

3. #sum(#syn(FN #od1(Förenta Nationerna) #od1(Förenta Nationernas)) #syn(invasion invasionen invasioner invasionerna #od1(fientlig inryckning) #od1(fientliga inryckningen) #od1(fientliga inryckningar) #od1(fientliga inryckningarna) #od1(fientligt infall) #od1(fientliga infallet) #od1(fientliga infallen) inträngande inträngandet #od1(plötsligt överfall) #od1(plötsliga överfallet) #od1(plötsliga överfall) #od1(plötsliga överfallen) ockupation ockupationen ockupationer ockupationerna besättande besättandet besättanden besättandena landstigning landstigningen landstigningar landstigningarna massinvandring massinvandringen massinvandringar massinvandringarna ström strömmen strömmar strömmarna #od1(ankomst av stora skaror) översvämning översvämningen översvämningar översvämningarna) #syn(Haiti Haitis))

4. #sum(FN #syn(invasion invasionen invasioner invasionerna) #syn(Haiti Haitis))
**Inga snävare**

5. #sum(FN #syn(invasion invasionen invasioner invasionerna) #syn(Haiti Haitis))
**Inga relaterade**

6. #sum(FN #syn(invasion invasionen invasioner invasionerna) #syn(Haiti Haitis))
**Inga bredare**

**Nr 261. Topic 045:** *Sök rapporter som innehåller namnen på huvudförhandlarna till det fredsfördrag som undertecknats i Mellanöstern mellan Israel och Jordanien samt dessutom dokument som ger detaljerad information om fördraget.*

1. #sum(förhandla fred fördrag Israel Jordanien)

2. #sum(#syn(förhandla förhandlat förhandlade) #syn(fred freden freder frederna) #syn(fördrag fördraget fördragen) #syn(Israel Israels) #syn(Jordanien Jordaniens))

3. #sum(#syn(förhandla förhandlat förhandlade underhandla underhandlade underhandlat diskutera diskuterat diskuterade överlägga överlade överlagt #od1(träffa överenskommelse) #od1(träffade överenskommelse) #od1(träffat överenskommelse) parlamentera parlamenterat parlamenterade dryfta dryftade dryftat rådslå rådslog rådslagit konferera konfererade konfererat) #syn(fred freden freder frederna fredstillstånd fredstillståndet fredstillstånden pax fredstid fredstiden fredstider fredstiderna #od1(tillstånd utan krig) #od1(tillståndet utan krig) #od1(tillstånden utan krig) vapenstillestånd vapenstilleståndet vapenstillestånden fredsslut fredsslutet fredssluten fredsfördrag fredsfördraget fredsfördragen sämja sämjan försoning försoningen försoningar försoningarna endräkt endräkten enighet enigheten) #syn(fördrag fördraget fördragen avtal avtalet avtalen traktat traktaten traktater traktaterna konvention konventionen konventioner konventionerna överenskommelse överenskommelsen överenskommelser överenskommelserna pakt pakten pakter pakterna) #syn(Israel Israels) #syn(Jordanien Jordaniens))

4. #sum(#syn(förhandla förhandlat förhandlade) #syn(fred freden freder frederna) #syn(fördrag fördraget fördragen) #syn(Israel Israels) #syn(Jordanien Jordaniens))
**Inga snävare**

5. #sum(#syn(förhandla förhandlat förhandlade) #syn(fred freden freder frederna fredsfördrag fredsfördraget fredsfördragen krig kriget krigen krigföring krigföringen vapenvila vapenvilan vapenvilor vapenvilorna) #syn(fördrag fördraget fördragen) #syn(Israel Israels) #syn(Jordanien Jordaniens))
**Inga relaterade till förhandla, fördrag**

6. #sum(#syn(förhandla förhandlat förhandlade) #syn(fred freden freder frederna fredsfördrag fredsfördraget fredsfördragen krig kriget krigen krigföring krigföringen vapenvila vapenvilan vapenvilor vapenvilorna) #syn(fördrag fördraget fördragen) #syn(Israel Israels) #syn(Jordanien Jordaniens))
**Inga bredare**

**Nr 262. Topic 046:** *Vilka effekter har FN:s handelsblockad haft på det irakiska folkets tillvaro?*

1. #sum(FN handelsblockad Irak)

2. #sum(FN #syn(handelsblockad handelsblockaden handelsblockader handelsblockaderna) #syn(Irak Iraks))

3. #sum(#syn(FN #od1(Förenta Nationerna) #od1(Förenta Nationernas) #syn(handelsblockad handelsblockaden handelsblockader handelsblockaderna)) #syn(Irak Iraks))
**Inga synonymer till handelsblockad**

4. #sum(FN #syn(handelsblockad handelsblockaden handelsblockader handelsblockaderna) #syn(Irak Iraks))
**Inga snävare**

5. #sum(FN #syn(handelsblockad handelsblockaden handelsblockader handelsblockaderna #od1(ekonomiska sanktioner) #od1(ekonomiska sanktionerna) #od1(ekonomisk sanktion) #od1(ekonomiska sanktionen)) #syn(Irak Iraks))

6. #sum(FN #syn(handelsblockad handelsblockaden handelsblockader handelsblockaderna #od1(ekonomiska sanktioner) #od1(ekonomiska sanktionerna) #od1(ekonomisk sanktion) #od1(ekonomiska sanktionen) blockad blockaden blockader blockaderna) #syn(Irak Iraks))

**Nr 263. Topic 047:** *Vilka skäl finns det för den ryska militära interventionen i Tjetjenien?*

1. #sum(rysk intervention Tjetjenien)

2. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna) #syn(Tjetjenien Tjetjeniens))

3. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna inskridande inskridandet inskridanden inskridandena ingripande ingripandet ingripanden ingripandena inblandning inblandningen inblandningar inblandningarna medling medlingen) #syn(Tjetjenien Tjetjeniens))

4. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna #od1(humanitär intervention) #od1(humanitära interventionen) #od1(humanitära interventioner) #od1(humanitära interventionerna)) #syn(Tjetjenien Tjetjeniens))

5. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna #od1(humanitär intervention) #od1(humanitära interventionen) #od1(humanitära interventioner) #od1(humanitära interventionerna) erövring erövringen erövringar erövringarna fredsframtvingande fredsframtvingandet fredsframtvinganden fredsframtvingandena krigshandlingar krigshandlingarna krigshandling krigshandlingen ockupation ockupationen ockupationer ockupationerna) #syn(Tjetjenien Tjetjeniens))

6. #sum(#syn(rysk ryska) #syn(intervention interventionen interventioner interventionerna #od1(humanitär intervention) #od1(humanitära interventionen) #od1(humanitära interventioner) #od1(humanitära interventionerna) erövring erövringen erövringar erövringarna fredsframtvingande fredsframtvingandet fredsframtvinganden fredsframtvingandena krigshandlingar krigshandlingarna krigshandling krigshandlingen ockupation ockupationen ockupationer ockupationerna) #syn(Tjetjenien Tjetjeniens))
**Inga bredare**

**Nr 264. Topic 048:** *Skäl för tillbakadragandet av FN:s fredsbevarande styrkor ur Bosnien.*

1. #sum(FN tillbakadragande Bosnien)

2. #sum(FN #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena) #syn(Bosnien Bosniens))

3. #sum(#syn(FN #od1(Förenta Nationerna) #od1(Förenta Nationernas) #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena) #syn(Bosnien Bosniens))
**Inga synonymer till tillbakadragande**

4. #sum(FN #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena) #syn(Bosnien Bosniens))
**Inga snävare**

5. #sum(FN #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena reträtt reträtten reträtter reträtterna) #syn(Bosnien Bosniens))

6. #sum(FN #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena reträtt reträtten reträtter reträtterna) #syn(Bosnien Bosniens))
**Inga bredare**

**Nr 265. Topic 050:** *Sök dokument om indianernas uppror i Chiapas (Mexico).*

1. #sum(indianer uppror Chiapas)

2. #sum(#syn(indianer indianerna indian indianen) #syn(uppror upproret upproren) Chiapas))

3. #sum(#syn(indianer indianerna indian indianen #od1(amerikansk urinvånare) #od1(amerikanska urinvånaren) #od1(amerikanska urinvånare) #od1(amerikanska urinvånarna) rödskinn rödskinnet rödskinnen) #syn(uppror upproret upproren revolt revolten revolter revolterna revolution revolutionen revolutioner revolutionerna resning resningen resningar resningarna uppresning uppresningen uppresningar uppresningarna upplopp upploppet upploppen bråk bråket bråken tumult tumultet myteri myteriet myterier myterierna) Chiapas)

4. #sum(#syn(indianer indianerna indian indianen azteker aztekerna aztek azteken inkaindianer inkaindianerna inkaindian inkaindianen mayafolk mayafolket mayafolken tolteker toltekerna toltek tolteken) #syn(uppror upproret upproren bondeuppror bondeupproret bondeupproren) Chiapas)

5. #sum(#syn(indianer indianerna indian indianen azteker aztekerna aztek azteken inkaindianer inkaindianerna inkaindian inkaindianen mayafolk mayafolket mayafolken tolteker toltekerna toltek tolteken) #syn(uppror upproret upproren bondeuppror bondeupproret bondeupproren upprorsrörelser upprorsrörelserna upprorsrörelsen upprorsrörelse) Chiapas)
**Inga relaterade till indianer**

6. #sum(#syn(indianer indianerna indian indianen azteker aztekerna aztek azteken inkaindianer inkaindianerna inkaindian inkaindianen mayafolk mayafolket mayafolken tolteker toltekerna toltek tolteken) #syn(uppror upproret upproren bondeuppror bondeupproret bondeupproren upprorsrörelser upprorsrörelserna upprorsrörelsen upprorsrörelse) Chiapas)
**Inga bredare**

**Nr 266. Topic 051:** *Sök dokument om slutspelet i 1994 års världsmästerskap i fotboll.*

1. #sum(slutspel världsmästerskap fotboll 1994)

2. #sum(#syn(slutspel slutspelet slutspelen) #syn(världsmästerskap världsmästerskapet världsmästerskapen) #syn(fotboll fotbollen) 1994)

3. #sum(#syn(slutspel slutspelet slutspelen) #syn(världsmästerskap världsmästerskapet världsmästerskapen VM) #syn(fotboll fotbollen bollsport bollsporten bollsporter bollsporterna lagspel lagspelet lagspelen fotbollsmatch fotbollsmatchen fotbollsmatcher fotbollsmatcherna) 1994)
**Ingen synonym till slutspel, världsmästerskap**

4. #sum(#syn(slutspel slutspelet slutspelen) #syn(världsmästerskap världsmästerskapet världsmästerskapen) #syn(fotboll fotbollen innefotboll innefotbollen) 1994)
**Inga snävare till slutspel, världsmästerskap**

5. #sum(#syn(slutspel slutspelet slutspelen) #syn(världsmästerskap världsmästerskapet världsmästerskapen) #syn(fotboll fotbollen innefotboll innefotbollen) 1994)
**Inga relaterade**

6. #sum(#syn(slutspel slutspelet slutspelen) #syn(världsmästerskap världsmästerskapet världsmästerskapen) #syn(fotboll fotbollen innefotboll innefotbollen bollsport bollsporten bollsporter bollsporterna lagspel lagspelet lagspelen) 1994)
**Inga bredare till slutspel, världsmästerskap**

**Nr 274. Topic 066:** *Sök rapporter och diskussioner om ryska truppers tillbakadragande från Lettland.*

1. #sum(ryska trupper tillbakadragande Lettland)

2. #sum(#syn(ryska rysk) #syn(trupper trupperna trupp truppen) #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena) #syn(Lettland Lettlands))

3. #sum(#syn(ryska rysk) #syn(trupper trupperna trupp truppen militärstyrka militärstyrkan militärstyrkor militärstyrkorna förband förbandet förbanden armé armén arméer arméerna tropp troppen troppar tropparna) #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena) #syn(Lettland Lettlands))

**Inga syn till tillbakadragande**

4. #sum(#syn(ryska rysk) #syn(trupper trupperna trupp truppen) #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena) #syn(Lettland Lettlands))
**Inga snävare**

5. #sum(#syn(ryska rysk) #syn(trupper trupperna trupp truppen) #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena reträtt reträtten reträtter reträtterna) #syn(Lettland Lettlands))
**Inga relaterade till trupp**

6. #sum(#syn(ryska rysk) #syn(trupper trupperna trupp truppen) #syn(tillbakadragande tillbakadragandet tillbakadraganden tillbakadragandena reträtt reträtten reträtter reträtterna) #syn(Lettland Lettlands))
**Inga bredare**

**Nr 283. Topic 080:** *Dokument som innehåller information om hungerstrejksaktioner som satts i gång för att dra uppmärksamheten till en sak.*

1. #sum(hungerstrejk aktioner)

2. #sum(#syn(hungerstrejk hungerstrejken hungerstrejker hungerstrejkerna) #syn(aktioner aktionerna aktion aktionen))

3. #sum(#syn(hungerstrejk hungerstrejken hungerstrejker hungerstrejkerna) #syn(aktioner aktionerna aktion aktionen #od1(aktivt ingripande) #od1(aktiva ingripandet) #od1(aktiva ingripanden) #od1(aktiva ingripandena) ingripande ingripandet ingripanden ingripandena handling handlingen handlingar handlingarna verksamhet verksamheten verksamheter verksamheterna))
**Inga syn till hungerstrejk**

4. #sum(#syn(hungerstrejk hungerstrejken hungerstrejker hungerstrejkerna) #syn(aktioner aktionerna aktion aktionen))
**Inga snävare**

5. #sum(#syn(hungerstrejk hungerstrejken hungerstrejker hungerstrejkerna #od1(passivt motstånd) #od1(passiva motståndet)) #syn(aktioner aktionerna aktion aktionen))
**Inga relaterade till aktioner**

6. #sum(#syn(hungerstrejk hungerstrejken hungerstrejker hungerstrejkerna #od1(passivt motstånd) #od1(passiva motståndet)) #syn(aktioner aktionerna aktion aktionen))
**Inga bredare**

**Nr 284. Topic 082:** *Sök dokument som beskriver terroristaktioner utförda av den irländska republikanska armén (IRA) på europeiska flygplatser.*

1. #sum(terrorism #od1(irländska republikanska armén) flygplatser)

2. #sum(#syn(terrorism terrorismen) #syn(#od1(irländska republikanska armén) #od1(irländska republikanska arméns)) #syn(flygplatser flygplatserna flygplats flygplatsen))

3. #sum(#syn(terrorism terrorismen skräckvälde skräckväldet skräckvälden skräckväldena våldsmetoder våldsmetoderna våldsmetod våldsmetoden förföljelse förföljelsen förföljelser förföljelserna förtryck förtrycket tyranni tyranniet) #syn(IRA #od1(irländska republikanska armén) #od1(irländska republikanska arméns)) #syn(flygplatser flygplatserna flygplats flygplatsen flygfält flygfältet flygfälten flyghamn flyghamnen flyghamnar flyghamnarna aerodrom aerodromen aerodromer aerodromerna landningsplats landningsplatsen landningsplatser landningsplatserna lufthamn lufthamnen lufthamnar lufthamnarna))

4. #sum(#syn(terrorism terrorismen bioterrorism bioterrorismen) #syn(#od1(irländska republikanska armén) #od1(irländska republikanska arméns)) #syn(flygplatser flygplatserna flygplats flygplatsen))
**Inga snävare till flygplatser**

5. #sum(#syn(terrorism terrorismen bioterrorism bioterrorismen attentat attentatet attentaten flygkapningar flygkapningarna flygkapning flygkapningen gisslan kidnappningar kidnappningarna kidnappning kidnappningen #od1(politiska brott) #od1(politiska brottet) #od1(politiska brotten) #od1(politiskt brott) #od1(politiskt mord) #od1(politiska mordet) #od1(politiska mord) #od1(politiska morden) #od1(politiskt våld) #od1(politiska våldet) sabotage sabotaget terrorattacker terrorattackerna terrorattack terrorattacken) #syn(#od1(irländska republikanska armén) #od1(irländska republikanska arméns)) #syn(flygplatser flygplatserna flygplats flygplatsen flygfält flygfältet flygfälten))

6. #sum(#syn(terrorism terrorismen bioterrorism bioterrorismen attentat attentatet attentaten flygkapningar flygkapningarna flygkapning flygkapningen gisslan kidnappningar kidnappningarna kidnappning kidnappningen #od1(politiska brott) #od1(politiska brottet) #od1(politiska brotten) #od1(politiskt brott) #od1(politiskt mord) #od1(politiska mordet) #od1(politiska mord) #od1(politiska morden) #od1(politiskt våld) #od1(politiska våldet) sabotage sabotaget terrorattacker terrorattackerna terrorattack terrorattacken) #syn(#od1(irländska republikanska armén) #od1(irländska republikanska arméns)) #syn(flygplatser flygplatserna flygplats flygplatsen flygfält flygfältet flygfälten stationer stationerna station stationen))
**Inga bredare till terrorism**

**Nr 286. Topic 086:** *Sök dokument som beskriver användningen av eller inställningen till "grön" energi, d.v.s. energi från förnyelsebara energikällor.*

1. #sum(energi förnyelsebar grön)

2. #sum(#syn(energi energin energier energierna) #syn(förnyelsebar förnyelsebart förnyelsebara) #syn(grön gröna))

3. #sum(#syn(energi energin energier energierna elenergi elenergin elkraft elkraften bränsle bränslen bränslena bränslet drivmedel drivmedlet drivmedlen) #syn(förnyelsebar förnyelsebart förnyelsebara ) #syn(grön gröna miljövänlig miljövänliga miljövänligt miljömedveten miljömedvetna miljömedvetet))
**Ingen syn till förnyelsebar**

4. #sum(#syn(energi energin energier energierna bioenergi bioenergin ekoenergi ekoenergin elenergi elenergin #od1(geotermisk energi) #od1(geotermiska energin) jordvärme jordvärmen #od1(kinetisk energi) #od1(kinetiska energin) kärnenergi kärnenergin #od1(potentiell energi) #od1(potentiella energin) solenergi solenergin strålningsenergi strålningsenergin tidvattenenergi tidvattenenergin vattenkraft vattenkraften vindkraft vindkraften värme värmen vågenergi vågenergin ångkraft ångkraften) #syn(förnyelsebar förnyelsebart förnyelsebara) #syn(grön gröna))
**Inga snävare till förnyelsebar, grön**

5. #sum(#syn(energi energin energier energierna bioenergi bioenergin ekoenergi ekoenergin elenergi elenergin #od1(geotermisk energi) #od1(geotermiska energin) jordvärme jordvärmen #od1(kinetisk energi) #od1(kinetiska energin) kärnenergi kärnenergin #od1(potentiell energi) #od1(potentiella energin) solenergi solenergin strålningsenergi strålningsenergin tidvattenenergi tidvattenenergin vattenkraft vattenkraften vindkraft vindkraften värme värmen vågenergi vågenergin ångkraft ångkraften energikällor energikällorna energikälla energikällan) #syn(förnyelsebar förnyelsebart förnyelsebara) #syn(grön gröna))
**Inga relaterade till förnyelsebar, grön**

6. #sum(#syn(energi energin energier energierna bioenergi bioenergin ekoenergi ekoenergin elenergi elenergin #od1(geotermisk energi) #od1(geotermiska energin) jordvärme jordvärmen #od1(kinetisk energi) #od1(kinetiska energin) kärnenergi kärnenergin #od1(potentiell energi) #od1(potentiella energin) solenergi solenergin strålningsenergi strålningsenergin tidvattenenergi tidvattenenergin vattenkraft vattenkraften vindkraft vindkraften värme värmen vågenergi vågenergin ångkraft ångkraften energikällor energikällorna energikälla energikällan) #syn(förnyelsebar förnyelsebart förnyelsebara) #syn(grön gröna))
**Inga bredare**

# Appendix 2: Bortfall

**Nr 248. Topic 021:** Sök dokument om Europeiska ekonomiska samarbetsområdet (EES)

1. EES

**Nr 276. Topic 070**: *Sök dokument som ger biografisk information om Kim Il Sung, Nordkoreas president som dog 1994.*

1. #sum(#od1(Kim Il Sung) Nordkorea)

**Nr 285. Topic 085:** *Sök detaljerad information om operation "Turquoise", den franska humanitära aktionen i Rwanda.*

1. #sum(#od1(operation Turquoise) Rwanda)

**Nr 267. Topic 053:** *Vilka gener har identifierats som är upphov till eller bidrar till att orsaka sjukdomar eller utvecklingsstörningar hos människor?*

1. #sum(gener sjukdomar människor)

2. #sum(#syn(gener generna gen genen) #syn(sjukdomar sjukdomarna sjukdom sjukdomen) #syn(människor människorna människa människan))

3. #sum(#syn(gener generna gen genen arvsanlag arvsanlaget arvsanlagen arvsfaktor arvsfaktorn arvsfaktorer arvsfaktorerna DNA-segment DNA-segmentet DNA-segmenten) #syn(sjukdomar sjukdomarna sjukdom sjukdomen sjuka sjukan sjukor sjukorna lidande lidandet lidanden lidandena ont besvär besväret besvären opasslighet opassligheten åkomma åkomman åkommor åkommorna krämpa krämpan krämpor krämporna krankhet krankheten ohälsa ohälsan sjuklighet sjukligheten smitta smittan smittor smittorna infektion infektionen infektioner infektionerna inflammation inflammationen inflammationer inflammationerna katarr katarren katarrer katarrerna sot soten) #syn(människor människorna människa människan väsen väsendet väsenden väsendena varelse varelsen varelser varelserna individ individen individer individerna person personen personer personerna människoliv människolivet människoliven själ själen själar själarna man kött köttet kropp kroppen kroppar kropparna kotte kotten kottar kottarna käft))

4. Ingen sökning på snävare termer, pga. för högt antal termer (35 st snävare termer till sjukdom. Se appendix 3 för samtliga av dessa termer).
**Inga snävare till gener, människa**

5. #sum(#syn(gener generna gen genen DNA gendefekter gendefekterna gendefekt gendefekten kromosomer kromosomerna kromosom kromosomen ärftlighet ärftligheten)

#syn(djursjukdomar djursjukdomarna djursjukdom djursjukdomen epidemier epidemierna epidemi epidemin epidemiologi epidemiologin etiologi etiologin komplikationer komplikationerna komplikation komplikationen pandemier pandemierna pandemi pandemin sjukdom sjukdomen sjukdomar sjukdomarna växtsjukdomar växtsjukdomarna växtsjukdom växtsjukdomen) #syn(människor människorna människa människan))
**Inga relaterade till människa**

6. #sum(#syn(gener generna gen genen DNA gendefekter gendefekterna gendefekt gendefekten kromosomer kromosomerna kromosom kromosomen ärftlighet ärftligheten) #syn(djursjukdomar djursjukdomarna djursjukdom djursjukdomen epidemier epidemierna epidemi epidemin epidemiologi epidemiologin etiologi etiologin komplikationer komplikationerna komplikation komplikationen pandemier pandemierna pandemi pandemin sjukdom sjukdomen sjukdomar sjukdomarna växtsjukdomar växtsjukdomarna växtsjukdom växtsjukdomen) #syn(människor människorna människa människan))
**Inga bredare**

**Nr 258. Topic 039:** *Sök exempel på företag som gör eller har gjort investeringar i Östeuropa eller Ryssland efter upplösningen av det forna Sovjetunionen.*

1. #sum(företag investeringar Ryssland Östeuropa)

2. #sum(#syn(företag företaget företagen) #syn(investeringar investeringarna investering investeringen) #syn(Ryssland Rysslands) #syn(Östeuropa Östeuropas))

3. #sum(#syn(företag företaget företagen rörelse rörelsen rörelser rörelserna firma firman firmor firmorna bolag bolaget bolagen affär affären affärer affärerna affärsrörelse affärsrörelsen affärsrörelser affärsrörelserna industri industrin industrier industrierna verksamhet verksamheten verksamheter verksamheterna anläggning anläggningen anläggningar anläggningarna etablissemang etablissemanget) #syn(investeringar investeringarna investering investeringen #od1(placera kapital) #od1(placera kapitalet) #od1(satsa pengar) #od1(satsa pengarna) placera placerade placerat satsa satsade satsat #od1(plöja ner) #od1(plöjde ner)) #syn(Ryssland Rysslands) #syn(Östeuropa Östeuropas))

4. Ingen sökning på pga. för högt antal snävare termer till företag (58 st) Se appendix 3 för samtliga termer som tagits bort.

5. #sum(#syn(företag företaget företagen företagens affärsverk affärsverket affärsverken bolag bolagen bolag bolaget företagsformer företagsformerna  företagsform företagsformen företagsverksamhet företagsverksamheten företagsverksamheter företagsverksamheterna karteller kartellerna kartell kartellen) #syn(investeringar investeringarna investering investeringen #od1(direkta investeringar) #od1(direkta investeringarna) #od1(direkt investering) #od1(direkta investeringen) #od1(utländska investeringar) #od1(utländska investeringarna) #od1(utländsk investering) #od1(utländska investeringen) deinvesteringar deinvesteringarna deinvestering deinvesteringen investeringsreserveringar investeringsreserveringarna investeringsreservering investeringsreserveringen placeringar placeringarna placering placeringen placeringsverksamhet placeringsverksamheten

placeringsverksamheter placeringsverksamheterna) #syn(Ryssland Rysslands) #syn(Östeuropa Östeuropas))

6. #sum(#syn(företag företaget företagen företagens affärsverk affärsverket affärsverken bolag bolagen bolag bolaget företagsformer företagsformerna  företagsform företagsformen företagsverksamhet företagsverksamheten företagsverksamheter företagsverksamheterna karteller kartellerna kartell kartellen) #syn(investeringar investeringarna investering investeringen #od1(direkta investeringar) #od1(direkta investeringarna) #od1(direkt investering) #od1(direkta investeringen) #od1(utländska investeringar) #od1(utländska investeringarna) #od1(utländsk investering) #od1(utländska investeringen) deinvesteringar deinvesteringarna deinvestering deinvesteringen investeringsreserveringar investeringsreserveringarna investeringsreservering investeringsreserveringen placeringar placeringarna placering placeringen placeringsverksamhet placeringsverksamheten placeringsverksamheter placeringsverksamheterna) #syn(Ryssland Rysslands) #syn(Östeuropa Östeuropas))
**Inga bredare**

# Appendix 3: Borttagna termer

**Borttagna termer från nr 267, topic 053: Snävare termer till sjukdomar (35 st)**

Andningsorganens sjukdomar
Arbetsrelaterade sjukdomar
Bakteriesjukdomar
Barnsjukdomar
Blod- och lymfkärlssjukdomar
Blodsjukdomar
Cirkulationsorganens sjukdomar
Endokrina sjukdomar
Följdsjukdomar
Hjärnans sjukdomar
Hjärt- kärlsjukdomar
Hudsjukdomar
Invärtesmedicinska sjukdomar
Kvinnosjukdomar
Könssjukdomar
Matsmältningsorganens sjukdomar
Mentala störningar
Mun- och tandsjukdomar
Nervsjukdomar
Näringsrubbningar
Parasitsjukdomar
Rörelseorganens sjukdomar
Svampsjukdomar
Smittsamma sjukdomar
Strålningssjukdomar
Tropiska sjukdomar
Tumörer
Urogenitalsystemets sjukdomar
Virussjukdomar
Zoonoser
Ålderdomssjukdomar
Ämnesomsättningsrubbningar
Ärftliga sjukdomar
Ögonsjukdomar
Öronsjukdomar

**Borttagna termer från nr 258, topic 039: Snävare termer till företag (58 st)**

Affärskedjor
Andelslag
Andrahandsaffärer
Antikvitetsaffärer
Apotek
Bensinstationer
Bevakningsföretag
Bokhandlar
Bokföringsbyråer
Byggnadsfirmor
Centralaffärer
Elaffärer
Familjeföretag
Fastighetsskötselbolag
Flerbranschföretag
Flygbolag
Fotoaffärer
Fotoateljéer
Frisersalonger
Handelshus
Hotellbranschen
Industriföretag
Ingenjörsbyråer
IT-företag
Järnhandlar
Koncerner
Konsortier
Konsultbyråer
Kreditinstitut
Kunskapsföretag
Kursföretag
Lantbruksföretag
Medelstora företag
Mediebyråer
Monopol
Multinationella företag
Näringsställen
Offentliga företag
Oljebolag
Partiaffärer
Rederier
Reklambyråer
Resebyråer
Samföretag
Serviceföretag
Skrotningsanläggningar
Små och medelstora företag
Småföretag
Sociala företag
Speditionsfirmor

Storföretag
Städfirmor
Teknologiföretag
Teleoperatörer
Transportföretag
Truster
Turismföretag
Uthyrningsfirmor