

Kvalitativ analys av skriftlig text

- en utvärdering av datorprogram som hjälpmedel

Beatriz Dorriots*
Kristina Jonäll**
Inga-Lill Johansson**

* University of Göteborg
Department of Linguistics
Box 200
405 30 Göteborg, Sweden
Beatrice@ling.gu.se
Telephone: (46)-31-7734479

** Göteborg School of
Economics and Commercial Law,
Department of Business
Administration Box 610
405 30 Göteborg, Sweden
Kristina.Jonall@handels.gu.se
Fax: (46)-31-7734853
Telephone: (46)-031-7731479
Fax: (46)-31-7735877
Inga-Lill.Johansson@handels.gu.se
Telephone: (46)-031-7734331
Fax: (46)-31-7731847

Innehållsförteckning

Introduktion	3
Det mänskliga språkets natur.....	3
Att analysera korpusar med hjälp av datorprogram.....	4
Kvalitativ textanalys	4
Det oundvikliga första steget: att finna nyckelord.....	4
Manuell textanalys.....	4
Maskinell textanalys	6
Tillgängliga programvaror.....	6
Vårt urval av tillgängliga programvaror	6
Vårt urval av 14 programvaror	6
Utvärdering av datorprogram som hjälpmedel för textanalys	12
Vårt urval av tre testprogram.....	12
Användning av de tre testprogrammen för en specifik forskningsuppgift	13
Slutsatser.....	16
Referenser	
Bilagor	

Introduktion

Datorprogram har använts för kvantitativa analyser under lång tid (s.k. statistisk bearbetning av textmaterial). Fokus i det aktuella projektet har varit på kvalitativa analyser av fenomen och processer. I sådana analyser är syftet att upptäcka hur dessa fenomen och processer är beskaffade, att upptäcka deras mening. För ett sådant syfte kan datorprogram förefalla ha allvarliga begränsningar. I denna rapport utvärderas ett antal datorprogram som hjälpmedel för kvalitativa analyser. Projektet har genomförts parallellt med avhandlingsprojektet "Redogörandeaspekter på VD'n har ordet texter".

Arbetet har utförts i följande steg:

Steg 1: Identifiering av analyskategorier lämpliga vid analys av rådata som t.ex..

- a) pressmeddelanden
- b) årsredovisningar, emissionsprospekt,
- c) mötesprotokoll, transkriptioner av video/audioinspelningar av möten och intervjuer
- d) tidskriftsartiklar om företagsekonomiska forskningsresultat, företagsanalyser.

I projektet har "VD'n har ordet"-texter ur årsredovisningarna för 1999-2002 för ett svenskt börsnoterat företag använts.

Steg 2: Första test av 14stycken identifierade programvaror (se bilaga 1) med syfte att välja ut tre stycken för en djupgående test avseende tillämpningar på ovan nämnda rådata.

Steg 3: Djupgående test av de tre valda (i Steg 2) programvarorna genom applicering av de identifierade analyskategorierna (i Steg 1) på ett redan insamlat material (rådata). På detta sätt kan förekomster av lexikaliska fält kontrolleras och ett urval av data erhållas. Jämförelsen av programvarorna utvärderar följande aspekter:

- a) hanteringen av rådata, krav på organisering, kodning etc.
- b) styrningen av urvalet av data, kontroll av förekomster etc.
- c) presentation av resultat
- d) tidsåtgång vid en första test av programvaran
- e) pris

Det mänskliga språkets natur

Det mänskliga språket har två specifika kännetecken som är avgörande för i vilken utsträckning datorprogram kan användas för kvalitativ textanalys.

1) Ett väsentligt sådant drag är att språket är tvetydigt till sin natur - en utgångspunkt som är tagen för given inom lingvistik. När människor talar eller skriver är de ofta vaga och indirekta. Detta är dock inte enbart negativt, ibland är dessa drag helt funktionella i kommunikationen (se Steiner, 1998). Denna tvetydighet som karakteriserar språket får emellertid direkta konsekvenser för hur yttranden ska tolkas eftersom "the process of meaning-making depends upon what is explicit in a text but also what is implicit" (Fairclough, 2003:11). Att tolka yttranden innebär alltså att komplettera det explicita med alla s.k. 'missing links' (Brown & Yule, 1983). Här är ett datorprogram, som endast kan

arbeta med det explicita, inte till någon hjälp. Eftersom det "icke sagda" är den mest väsentliga dimensionen av ett budskap är detta en allvarlig begränsning i användbarheten hos datorprogram för analys av text. En manuell¹ innehållsinriktad textanalys krävs för att upptäcka implicita dimensioner i budskapet. Vid en sådan analys undersöker forskaren vad som finns "bakom" orden i ett segment av texten. (med hjälp av kontexten, egen tolkningsram, etablerad kunskap, etc.). Efteråt kodas segmentet på ett sätt som underlättar maskinell analys.

2) Ett annat kännetecken för mänskligt språk är att "whatever can be expressed by whatever linguistic means". Exempel på detta är s.k. 'rewording' och på vår förmåga att uttrycka nästan 'samma' budskap med olika ord genom att använda synonymer, ekvivalenta uttryck, mm. Detta kännetecken visar sig också i att samma ord kan ha olika betydelser (ibland även motsatta) beroende på sammanhanget och på de varierande extraverbala resurser människan har. Denna komplexitet i språket kan inte hanteras i ett datorprogram och är en av orsakerna att en manuell textanalys alltid måste föregå en maskinell behandling av textmaterialet.

Att analysera korpusar² med hjälp av datorprogram

En korpus är en samling verkliga yttranden. Intresset för att arbeta med korpusar har sin grund i ett av vetenskapens klassiska krav: att respektera empirin, att ha konkreta data på vilka teorierna grundas och mot vilka teorierna ska prövas. Stora korpusar eftersträvas ofta för att få en säkrare induktiv bas. Det är dock viktigt att framhålla att en stor korpus inte i sig är en garanti för att uppnå valida resultat. Ju större korpus, desto längre söknings- och bearbetningstid krävs och desto större blir behovet av att använda datorprogram som hjälpmedel i analysen.

Kvalitativ textanalys

Det oundvikliga första steget: att finna nyckelord

Eftersom datorprogram behöver bli styrda med minimala verbala enheter bör stor uppmärksamhet och möda ägnas åt uppgiften att hitta 'givande', 'täckande' nyckelord. Detta kräver viss kreativitet från forskaren initialt.

Redan 1935 talade den engelske lingvisten J.R. Firth om "focal or pivotal words" och poängterade hur viktigt det är att i en samhällsanalys att titta på "sociologically important verbs" vilkas användning skulle studeras "in actual contexts in the heart of everyday life". Konkordansanalysprogram (se nedan) är hjälpmedel som underlättar denna typ av analys. Mer avancerade program tillåter hantering av kombinationer av nyckelord.

Manuell textanalys

På grund av språkets specifika drag måste - som tidigare påpekats - en noggrann innehållsinriktad kvalitativ analys alltid inleda ett arbete med en korpus. Man kan t.ex. välja ett begränsat material som betraktas som särskilt rikt och givande i relation till ett

¹ "Manuell" är - en inte helt lyckad - benämning för analysen som forskaren själv gör utan tekniska hjälpmedel.

² En korpus utgör en samling språkligt material lagrat på dator, insamlat efter en genomtänkt design och representativ för ändamålet.

specifikt problem, det problem som senare kommer att analyseras med större material som grund och med hjälp av datorprogram och maskinell analys. Den innehållsinriktade analysen syftar till att identifiera ett antal nyckelord. Det är väsentligt att börja med en tydlig bild av syftet med analysen även om det senare kan komma att förändras. När ett preliminärt urval av nyckelord är gjort är det också intressant att titta på synonymer eller besläktade begrepp (som identifieras med hjälp av t.ex. ideologiska lexicon, thesaurus). Sådana procedurer kan avslöja eller utvidga relevanta semantiska fält³ och ge nya ledtrådar i sökandet efter lämpliga nyckelord (s.k. cohesive networks).

EXEMPEL - Demonstration av ansvar

Som sagts ovan, syftet ska vara tydligt formulerat och avgränsat så att man kan leta efter nyckelorden med viss - om än relativ - säkerhet. Låt oss ta ett exempel från en konkret forskningsuppgift så som den formuleras i avhandlingsprojektet "Redogörandeaspekter på VD'n har ordet texter". Syftet med den pilotstudie som inledningsvis genomfördes var att identifiera dimensioner viktiga för hur ledarskap och ansvar demonstreras samt hur på redogörandetryck responderas på. I pilotstudien analyserades 4 texter, "Till aktieägarna" 1999; "Till aktieägarna" 2000; "Brev från koncernchefen" 2001, and "Brev från koncernchefen" 2002, samtliga delar av Ericssons årsredovisningar för respektive år.

Efter att ha genomfört en innehållsinriktad kvalitativ analys av detta avgränsade material upptäcktes två centrala semantiska fält: "agency"- och åtagandefältet. Ansvarstagande (eller motsatsen) kan uttryckas på ett oändligt antal olika sätt, men avgörande är hur talaren/skribenten kodar in sin egen attityd (graden av säkerhet, engagemang, etc. eller tvärtom) inför det som han/hon överför (the "propositional content of the utterance", d.v.s. yttrandets sakliga innehåll). Det finns ett antal områden som lämnar avslöjande information i detta avseende:

1. Modaliteten ("the speaker's judgment of the probabilities, or the obligations, involved in what he is saying", Halliday 1994) manifesteras bl. a. via s.k. modala verb (kan/kunde, bor/borde, skulle), också via s.k. 'private verbs' (att tro, tänka, hoppas, förväntar sig, etc.) i kombination med pronomen (vi/oss, jag/mig) som anger explicit "agency".
2. Explicita performativa verb/verbala fraser (att påstå, fordra, försäkra, lova, besluta, bestämma, vidta åtgärder etc. med första person pronomina) är ofta mindre frekventa men avslöjar självklart engagemang.
3. Det är välkänt att åtagande och "explicitness" är starkt kopplade till varandra därför att det som ligger implicit i ett yttrande kan förnekas av talaren/skribenten ("Det menade jag inte"), men inte det som faktiskt sagts/skrivits. Tyvärr manifesterar människan sitt åtagande oftast på mer indirekta vägar.
4. Distansering inför ansvar kan läsas ut av sättet att inkoda "den andre/de andra". Vissa substantiv är intressanta: marknad, åtgärder, situationen, etc. Man förskjuter på det sättet det egna direkta ansvaret.
5. Passiva verbuttryck markerade med suffix -s (om detta är möjligt att söka med hjälp av ett datorprogram kan diskuteras) är talande därför att när något presenteras i passiv form osynliggörs agenten och ansvaret kringgås.

³ Ett semantiskt fält består av olika besläktade begrepp som hör till samma betydelsesfär. Ex: fältet för färgsfären, fältet för släktsfären.

6. Andra markörer för avståndstagande kan vara användningen av citationstecken, av litotes (försvagnings/förminskningsresurser: lite).

Därutöver bör naturligtvis alla nyckelord eller fasta uttryck analyseras i kontexten för att slutgiltigt avgöra om ett och samma uttryck är tecken på ansvar respektive åtaganden eller tvärtom.

Maskinell textanalys

Den manuella analysen förser forskaren med de nödvändiga ledtrådarna (nyckelord/ uttryck) för att behandla omfattande material (större korpusar). Användandet av datorprogram innebär betydande kvantitativa och tidsmässiga vinster. Som tidigare sagts utvidgas den empiriska basen tillämplig för att dra slutsatser. Följaktligen stärks resultaten förutsatt att övriga metodologiska krav respekteras.

Det kan finnas även andra vinster. "When computer methods are used to study large corpora, they may confirm what was suspected or known all along: but even such confirmation will usually provide vastly more detailed information than would otherwise be possible. [...]" (Stubbs, 1996:232). Men Stubbs påpekar också att dessa undersökningar kan generera "not only new facts, but facts of previously unsuspected kinds" (ibid.) och menar att dessa nya fakta kan leda till "far-reaching new hypothesis" (ibid.).

Korpus-lingvistiska metoder har börjat tillämpas bl a inom s.k. "culture studies", dels för att belysa hur olika kulturer skiljer sig åt i olika avseende men också för att diskutera om dessa metoder kan användas för att beskriva och jämföra kulturer (see e.g. Gerbig, 1997; Leech & Fallon, 1992). Forskarna inom andra discipliner (ekonomi, t.ex. i vårt fall) kan också använda dessa metoderna för att upptäcka oväntade aspekter inom sina respektive fält. De kan avslöja oväntade associationer/ relationer/dimensioner och erbjuda nya perspektiv. Datorprogram för textanalys/korpusanalys kan i så fall bli en inspirationskälla och inte bara ett verktyg för att fånga något förutsägbart. De kan då bli även 'more than a mere tool', men för att undvika orimliga förväntningar och välgrundad kritik ska man 1) ställa rimliga krav och 2) vidta metodiska åtgärder redan i den inledande fasen av arbetet.

Tillgängliga programvaror

Vårt urval av tillgängliga programvaror

Vi har gjort en sökning på Internet för att identifiera vilka programvaror som finns för kvalitativ analys av textmaterial. Det finns en uppsjö av olika program. Vissa kan man testa gratis under en kortare period innan man köper in dem.

Vårt urval av 14 programvaror

En tidigare rapport från Handelshögskolan i Stockholm (*Programvaror för bearbetning av kvalitativa data*, 2003) som innehåller en presentation av 21 stycken olika programvaror gav ledning i urvalet av de program som skulle ingå i första steget av vår utvärdering. Intervjuer med användare av programmen för forskningsändamål bidrog till att reducera antalet till 14 stycken. Vissa program visade sig vara överlappande och vissa mindre lämpliga för den typen av analys som vi avsåg att göra.

De flesta textanalysprogram bygger på att koda, kategorisera och “mindmappa”⁴ den text som skall analyseras. Till de främsta räknas NVivo och N6, Atlas-TI, Etnograph och Qualrus. Dataprogrammen erbjuder möjligheten att spara, organisera och kontrollera det textmaterial som skall analyseras. De underlättar också grundliga sökningar i den text som studeras. Efterhand som forskaren arbetar med texten kan materialet utökas, utvecklas och förändras allt efter forskarens egna behov. Det är lätt att växla mellan detaljer och kontext. Analysen blir också mer flexibel eftersom det är lätt att börja om på nytt om det visar sig att det spår som följdes var fel. Med ett textanalysprogram blir det också lättare att behandla materialet mer konsekvent. De tankeutvecklingar forskaren har kan följas och dessutom verifieras. Ett sådant analysätt blir mer transparent för den som står utanför att följa och kontrollera.

Det som programmen främst är till för är att organisera det material som forskaren har. De fungerar som ett hjälpmedel för att kunna se likheter och olikheter i materialet. Dessutom kan forskaren finna mönster och strukturer som annars kanske inte hade varit lika tydliga. De ger också ett stöd för kreativt tänkande. I och med att de flesta program erbjuder ett obegränsat antal koder får forskaren också oändligt många analysmöjligheter. Det är dock viktigt att vara medveten om begränsningarna i användbarheten hos datorprogrammen för analysen.

1. NVivo och 2. N6

NVivo och N6 är båda utvecklade från Nudist. NVivo är ett bra program att använda när behovet är ett analysverktyg som klarar av att hantera detaljerade analyser. Dessutom har NVivo den fördelen att det finns möjligheter till undersökning och tolkning för jämförelse av olika dokument. NVivo kan också skapa modeller och matriser för att resultatet skall bli mer överskådligt.

Om stora mängder text skall analyseras är NVivo ett av de bättre programmen på marknaden. NVivo i sig är inget analysprogram men det underlättar analysen betydligt om det används på rätt sätt. I programvaran finns olika verktyg inbyggda som underlättar för forskaren att förstå textmaterialet. Viktigt är dock att veta att det inte finns någon inbyggd funktion som gör att forskaren utan egen klassificering kan hitta minsta gemensamma nämnare i en text eller någon slags lösning till ett formaliserat problem. Programmet plockar inte fram “sanningen” för forskaren utan underlättar bearbetningen av texten.

Forskaren skapar själv de koder som skall användas i analysen och det är möjligt att använda ett flertal koder på olika nivåer. Ur dessa skapar man sedan “träd” av information som forskaren själv analyserar utifrån. Varje dokument som skall analyseras måste läsas igenom manuellt och kodas separat eftersom detta inte är något som programmet självt klarar av. Har man väl kodat filerna och texten kan man dock få ut en mängd data genom att kombinera de olika koderna med varandra och på så sätt erhålla olika dokument för varje kod eller “träd” som forskaren skapat.

N6, som liksom NVivo är en utveckling av Nudist och av samma programutvecklare, är mer lämpligt att använda på mycket stora datamängder och då man arbetar i forskningsteam med många inblandade. N6 ger inte samma möjlighet till ”trädstrukturer” som NVivo.

⁴ Ofta utgår programmen från en metod som liknar “mind-mapping” och det tankesätt som användningen av sådan teknik bygger på.

Internetadress: <http://www.qsrinternational.com>

3. ATLAS-Ti

Programmet kan användas för att göra kvalitativa analyser på stora mängder material bestående av text, grafer, bild (video) eller ljud. Programmet klarar av att göra systematiska bearbetningar av "mjuka" data där man inte kan analysera på ett meningsfullt sätt genom vanliga formella statistiska analyser. Programmet underlättar för forskaren att avslöja dolda komplexa fenomen.

ATLAS-Ti fungerar på liknande sätt som NVivo, genom att forskaren själv kodar texten, skapar noder, "träd" och "familjer". Ur dessa kan forskaren sedan få fram en slags bearbetning av allt material som kodats. Även här kan forskaren få ut en oändlig mängd av olika kombinationer mellan olika dokument och material. Detta underlättar betydligt. Det tidskrävande momentet är här, liksom i NVivo, genomläsning och kodning av materialet. Därefter går det snabbt att göra olika kombinationer av analyser inom och mellan de olika dokumenten.

Programmet är utvecklat i samarbete mellan psykologer, lingvister och blivande användare.

Internetadress: <http://www.atlasti.de>

4. Ethnograph v5.0

I Ethnograph v5.0 kan textbaserad kvalitativ data importeras direkt in i programmet. Allt som skrivits i något ordbehandlingsprogram kan användas. I programmet söker forskaren upp och noterar segment som är av värde för analysen. Dessa segment kodas på olika sätt. De olika koderna kan sedan kombineras för att bilda underlag för bearbetning. Programmet är mest lämpat att applicera vid analys av material som ligger till grund för etnografiska studier.

Internetadress: <http://www.qualisresearch.com>

5. Qualrus

Qualrus är ett program som kan användas för att göra kvalitativa databearbetningar på text-, ljud- och bildmaterial. Qualrus är utvecklat för att användas främst i forskning och undervisning. Programmet kan användas i många skiftande discipliner och är inte utvecklat för någon speciell teori eller analysmetod.

Programmet har inga direkta begränsningar avseende hur stora filer som kan bearbetas. Till de olika objekten, projekten och källorna kan användaren länka egna anteckningar eller stödord. Det finns verktyg som gör att användaren grafiskt kan göra "mind-maps" för att lättare kunna visualisera textavsnitt. Det finns också möjligheter att länka ihop olika projekt och arbeten med varandra. Programmet skapar ett slags semantiskt nätverk där de olika koderna och utvalda textdelarna visualiseras översiktligt. Det finns möjligheter att lätt navigera i detta nätverk. Ett flertal olika analysverktyg kan användas för att finna mönster i det kodade materialet. Bland dessa finns det några som är direkt analytiska och specifikt designade för att användas i kvalitativa studier.

Qualrus är ett flexibelt program som lätt går att anpassa till olika sorters forskning och olika forskares behov. Det kan användas för alla sorters kvalitativ analys.

Internetadress: <http://www.qualrus.com>

6. CDC EZ – TEXT

CDC EZ-TEXT är ett program som utvecklats för att kunna skapa och bearbeta semi-strukturerade kvalitativa databaser. Programmet är upphovsrättsligt skyddat men gratis att använda och ursprungligen utvecklat för beteendeforskning inom HIV/AIDS och sjukvårdsfältet. Detta begränsar dock inte användningen till endast dessa forskningsfält. Programmet kan användas av enskilda forskare eller i samarbete mellan forskare i större projekt.

Textfiler kan kopieras direkt in i databasen. Data och frågeformulär kan skräddarsys efter de egna behoven. Vid stora ostrukturerade etnografiska studier är programmet inte användbart. Det material som forskaren arbetar med kodas på ett liknande sätt som övriga program. De olika koderna samlas i textfiler som sedan kombineras bearbetas.

Internetadress: <http://www.cdc.gov/hiv/software/ez-text.htm>

7. C-I-SAID

C-I-SAID är ett integrerat program användbart främst för att analysera intervjuer och dialoger. Det tillåter även djupanalyser av texter och kan användas för analyser av all text från datainput, intervjuformulär till färdiga rapporter. Dokumenten som analyseras kan bestå av text, ljud eller bild. Hela textdokument kan matas in och analyseras inom loppet av några sekunder. Programmet kan användas för att koda materialet på olika sätt.

Programmet används mycket inom psykologi, sociologi, medicin och i marknadsundersökningar, men kan också användas inom andra discipliner och på bredare fält. I programmet finns ett inbyggt lexikon vilket underlättar kodningen av text. Dessutom finns ett verktyg som kan mäta ljudvolym, röstläge samt hastighet på tal för att användaren skall kunna få ut statistik på dessa parametrar ur till exempel intervjuer.

Internetadress: <http://www.code-a-text.co.uk>

8. HyperRESEARCH

Programmet hanterar olika sorters källmaterial såsom text, grafik, ljud och bild. Det finns tillgängligt för både PC och Mac.

HyperRESEARCH kan användas inom olika användningsområden där texter behöver analyseras och gås igenom. Det har mest använts inom de samhällsvetenskapliga disciplinerna. I programmet finns ett verktyg för auto-kodning i flera filer samtidigt. Forskaren får då specificera vilken sorts filer som skall kodas till vilka händelser liksom vilka ord eller uttryck programmet skall söka efter samt vilka koder det skall applicera.

Programmet är uppbyggt efter något som kallas "Point & Click" för att användaren skall kunna koncentrera sig på sin text och inte behöva lägga ner mer tid på att lära sig programmet än på att arbeta med det material som skall analyseras.

Internetadress: <http://www.researchware.com>

9. MAX qualitative data analysis

MAX qualitative data analysis är ett kraftfullt program som kan analysera textbaserad kvalitativ data. Dokument kan importeras i programmet i Rich Text Format (RTF). Storleken på dokumenten har ingen betydelse utan programmet kan hantera stora mängder av datamaterial. Text kan dras direkt från Internet Explorer och direkt in i programmet. Programmet klarar också av att importera och exportera information från program som SPSS och Excel.

Tabeller visar alla kodade segment så att de kodade variablerna och minnesanteckningarna kan visas när som helst under arbetets gång. Kodsystelet tillåter att forskaren bygger hierarkiska strukturer på upp till tio nivåer. På så sätt kan användaren med lätthet hantera flera hundra koder. Programmet erbjuder i princip två sätt att hantera och analysera data på. Dels erbjuds lexikala undersökningar då forskaren söker på ord och finner nyckelord i kontexten, dels erbjuds uppdelning av textmaterialet som sedan kan sättas ihop igen. Programmet fungerar snarlikt som NVivo och Atlas-Ti.

Internetadress: <http://www.maxqda.com>

10. Kwalitan 5

Kwalitan 5 är ett program som stöder analys av text, ljud och bildfiler. Det finns så gott som inga begränsningar för hur många dokument som kan analyseras. Dokumenten kan koda manuellt eller automatiskt med hjälp av programmet. Vid den manuella kodningen går forskaren själv igenom materialet och markerar ord och fraser samt tilldelar dessa de koder som önskas. Med hjälp av koderna beskriver forskaren vilka delar som är av vikt i materialet. I varje segment kan upp till 50 sök-/analytiker användas samtidigt. Koderna används för att specificera vad texten handlar om eller vilka tolkningar forskaren gjort. Koderna kan användas för att beskriva hela segment eller specifika delar av ett segment. Det automatiska kodsystelet som finns inbyggt i programmet underlättar delar av arbetet. Vid automatisk kodning anger forskaren en sökterm, bestämmer hur stor del av texten som omgärdar söktermen som skall kodas och sedan kodar programmet automatiskt hela dokumentet efter detta. Programmet kan lista de koder eller kodfragment som kodats på samma sätt.

En annan funktion i programmet är ordfunktionen. Denna fungerar på samma sätt som konkordansprogrammen (se nedan under Concordance).

Det finns också vissa filter i programmet som forskaren kan använda sig av för att välja ut eller ta bort vissa delar som är av speciell vikt eller som inte tillför något till forskningen.

Internetadress: <http://www.kwalitan.net>

11. AnSWR

AnSWR, *Analysis Software for Word-based Records*, är ett program som kan hantera stora komplexa kvalitativa databaser. I programmet finns funktioner som organiserar, visar, bearbetar, sammanfattar och tolkar textbaserad information.

I programmet kodar användaren texten, varvid hierarkiska kodstrukturer kan användas. Länkar kan skapas mellan olika dokument och projekt. Programmet är ett system som är designat för att hantera data efter tre principer. Det skall stödja forskningen och de behov

som finns där, det skall gå att göra analyser och organisera olika komplexa kvalitativa databaser och det skall vara objektorienterat.

Programmet är utvecklat av och fungerar på liknande sätt som CDC EZ – TEXT ovan.

Internetadress: <http://www.cdc.gov/hiv/software/answr.htm>

12. Visual text

Visual text är ett program i vilket forskaren skräddarsyr programmet efter egna behov. Det finns vissa förprogrammerade applikationer med vilka användaren startar arbetet och sedan bygger vidare på för att de skall passa in på det egna materialet och det som forskaren vill få ut.

Eftersom programmet är verktyg med vilket användaren bygger sina egna applikationer kräver det en del förkunskaper i programmering. Programmet har ett eget programmeringsspråk som bygger på C++. Har användaren dessa förkunskaper finns det stora möjligheter till flexibilitet i analyserna. Utan sådana blir programmet svårt att använda.

Internetadress: <http://www.textanalysis.com>

13. Wordstat

Wordstat är ett textbaserat verktyg som kan användas som hjälpmedel för att analysera text, allt från intervjuer till manuskript eller andra textbaserade dokument. Programmet kan även användas på viss kvantitativ data ur till exempel excel-filer eller olika databaser. Analysen bygger på att ordlistor komponeras eller importeras, vilka sedan används för att kategorisera ord och fraser på flera nivåer.

Programmet analyserar ord, fraser, kategorier eller koncept. Forskaren kan välja att manuellt koda texten med hjälp av nyckelord eller låta programmet söka ord att koda efter. Liknande ord samt ord med ändelser sorteras in i samma grupp. Programmet kan också söka efter hela fraser eller sammanhängande begrepp och sortera utifrån detta.

En nackdel är att Wordstat inte klarar större filer. Texter som innehåller specialtecken eller liknande måste i många fall anpassas för att kunna analyseras och bearbetas.

Internetadress: <http://www.simstat.com>

14. Concordance

På marknaden finns ett antal kommersiella konkordansanalysprogram som liknar varandra till sin funktion. Vid Göteborgs universitet, Humanistiska fakultet, finns ett egenutvecklat konkordansanalysprogram som använts som referenspunkt i vår utvärdering. Ett konkordansanalysprogram producerar så kallade textkonkordanser, d.v.s. förekomstställen av ett visst ord i ett textsammanhang. För att kunna göra en konkordans krävs en korpus, en textmängd som skall analyseras.

Det finns flera olika användningsområden för det man får ut av konkordansprogrammen, ett exempel är underlag för analys av ett specifikt problem, ett annat exempel är underlag för jämförelse mellan två eller flera korpus som skiljer sig åt i stil eller har upprättats vid olika tidpunkter.

Konkordansanalysprogram finns i flera utföranden. Alla ger forskaren möjlighet att göra innehållsanalyser av texter. Med programmen kan man göra ordlistor, frekvenslistor och index. Varje ord redovisas dels efter hur ofta det förekommer, dels efter i vilket sammanhang det förekommer.

Fördelen med programmet är att det blir lätt att hantera ett stort antal dokument och göra indexeringar i dessa. Det går att göra sökningar på ord i hela den samlade databasen som forskaren har byggt upp liksom att förfina sökningen på ett flertal sätt och söka på ordkombinationer.

Internetadress till det konkordansanalysprogram, Concordance, som vi testat:
<http://www.rjcw.freeseerve.co.uk/index.htm>

Utvärdering av datorprogram som hjälpmedel för textanalys

Syftet med utvärderingen var att göra en djupgående test (Steg 3) av ett mindre antal programvaror genom att applicera identifierade analyskategorier på ett redan insamlat material (rådata). Jämförelsen av programvarorna i detta steg fokuserade 1) hanteringen av rådata, (d.v.s. krav på organisering, kodning etc.), 2) styrningen av urvalet av data (kontroll av förekomster etc. samt 3) hur resultaten presenteras. Därutöver innehåller utvärderingen en jämförelse av tidsåtgången vid testen och kostnaden för anskaffningen av programmen.

Vårt urval av tre testprogram

Med hänsyn till att utvärderingen i detta steg skulle genomföras m.h.a. empiriska data som insamlats för ett specifikt forskningsändamål var det viktigt att välja de programvaror som var direkt lämpade för analys av den sortens text, i detta fall "VD'n har ordet" – texter. Denna typ av text utgör exempel på s.k. frivillig information som lämnas i företags årsredovisningar.

Urvalet av de tre programvaror som slutligen kom att utvärderas gjordes efter konsultation med forskare som har erfarenhet av motsvarande användning som vi tänkt oss. De tre programvaror som valdes ut var Concordance, NVIVO och Kwalitan 5.

Då Concordance visade sig ge snabb och överskådlig information var det önskvärt att testa ett program som eventuellt kunde kombinera konkordansprogrammets enkelhet och överskådlighet med ett textanalysprogramms djupare och flerdimensionella möjligheter. NVivo och Kwalitan 5 är sådana program. Det senare ger dessutom möjligheter till både manuell och automatisk kodning.

Utvärdering av Concordance

Concordance visade sig vara ett program som var översiktligt och lätt att lära sig använda. Filerna sparades som html vilket innebär att de går att öppna i vilken webbläsare som helst. Programmet gav snabba resultat av ordsökningar och orden kunde ses i sin kontext. Storleken på kontexten före och efter sökordet kunde väljas och ställas in i programmet. Det var dessutom möjligt att bläddra vidare om man ville se mer av kontexten kring ordet direkt på skärmen.

Utvärdering av NVivo

NVivo är ett avancerat program som kräver mycket förberedelser för inläring av programmet och för att man skall kunna göra bruk av alla möjligheter och funktioner. I demonstrationsversionen kan data inte sparas varför man hela tiden får börja om. Detta ger inte de bästa förutsättningarna när det gäller inläringen av programmets funktioner.

Det var svårt att få överblick över resultaten med de något svårtolkade dokumenten som produceras. Programmet krävde mycket arbete med kodning av textdokumenten. Forskaren måste själv manuellt koda alla textavsnitt i dokumenten. En fördel är att de olika koderna och dokumenten kan länkas till varandra om man finner samband som man vill kunna tydliggöra och kunna följa i analysen.

Resultatfilerna går endast att läsa i programmet eftersom det bygger på att specifika filer skapas som inte kan öppnas av något annat program. Det går dock att klippa in resultaten i ett annat textdokument men detta innebär ett tidsödande arbete.

Utvärdering av Kwalitan 5

Kwalitan 5 är ett program som kombinerar Concordanceprogrammets ordsökning med NVivos kodning. Introduktionstexten gjorde att man ganska snabbt kunde sätta sig in i programmet. I demonstrationsversionen var inte alla hjälpavsnitt åtkomliga, vilket var en brist. En utförlig manual saknades också när man väl kommit in i programmet och vill få ut mer av de olika funktionerna. Programmet gav överskådliga resultat och det var lätt att följa kodningen. Kodningen kan göras manuellt eller automatiskt med hjälp av programmet. Den automatiska kodningen är mycket tidsbesparande eftersom programmet självt söker upp ord eller fraser och sätter koder till dem. Vad forskaren sedan behöver göra är att bläddra igenom koderna och se vad som kommit med, samt eventuellt justera för sådant som man inte vill ha med. Nackdelen med demonstrationsversionen var också att den nederländska texten i programfunktioner, hjälpfunktioner och knappar ibland inte hade bytts ut till engelska i den engelska versionen.

Precis som i NVivo kan resultatfilerna i Kwalitan 5 endast läsas i programmet eftersom det bygger på att specifika filer skapas som inte kan öppnas av något annat program. Det går dock - i likhet med i NVivo - att klippa in resultaten i ett annat textdokument men detta innebär ett tidsödande arbete.

Användning av de tre testprogrammen för en specifik forskningsuppgift

För alla tre testprogrammen laddades en demonstrationsversion ner från Internet. Det som erhöles för dessa var lite olika. För NVivo erhöles en fullständig version men utan möjligheter att spara och för Concordanceprogrammet erhöles en fullständig version som fritt kunde testas under en begränsad tid. Kwalitan 5 har utvärderats genom en version utan fullständiga hjälpfunktioner.

Hanteringen av rådata, krav på organisering, kodning etc.

	Hantering av rådata (filsort)	Krav på organisering (textens organisering)	Kodning (förberedande bearbetning)
Concordance	Textfiler *.txt	Ingen	Ingen
NVIVO	Rtf filer *.rtf	Ingen	Forskaren måste själv koda texten. ⁵
Kwalitan 5	Rtf filer *.rtf	Ingen	För ordsökning krävs ingen förberedande bearbetning, kodningen kan göras både manuellt och maskinellt. Den maskinella kodningen innebär att grupper av ord först manuellt ges en kod som därefter används för att maskinellt identifiera samtliga förekomster av dessa grupper av ord i texten.

Styrningen av urvalet av data, kontroll av förekomster etc.

	Urvalet av data (nyckelord)	Kontroll av förekomster	Sammanhang
Concordance	Görs av forskaren. ⁶	Görs av programmet.	Storleken på kontexten bestäms av forskaren efter behov.
NVIVO	Görs av forskaren. ⁶	Görs av forskaren. ⁷	Storleken på kontexten bestäms av forskaren efter behov.
Kwalitan 5	Görs av forskaren. ⁶	Görs av programmet och forskaren.	Storleken på kontexten bestäms av forskaren efter behov.

⁵ Concordanceprogrammet skulle kunna utnyttjas för att göra det första steget i kodningen

⁶ Se avsnitten "Att finna nyckelord", samt "Maskinell textanalys".

⁷ Forskaren letar efter nyckelorden inom den mikrokontext som programmet identifierat och kan därefter manuellt summera antalet förekomster.

Presentation av resultat.

	Filformat för resultatpresentation	Organisering av resultatpresentationen (överskådlighet)	Tolkningsvärde
Concordance	.html	Mycket överskådlig.	Endimensionell ⁸
NVIVO	Filformat specifikt för NVivo.	Begränsad överskådlighet.	Flerdimensionell: möjligheter med ”träd”, skapar hierarkiska semantiska fält ⁹ .
Kwalitan 5	Filformat specifikt för Kwalitan 5.	Överskådligheten beror på vilken del av presentationen som granskas ¹⁰ .	Flerdimensionell: möjligheter med ”träd”, skapar hierarkiska semantiska fält ⁹ . Därutöver: endimensionell på samma sätt som Concordance.

Tidsåtgång¹¹

	Att bekanta sig med programmet	Att använda funktionerna
Concordance	Relativt lätt att sätta sig in i programmet. Snabbt att komma igång med (ca 4 timmar).	Lätta funktioner, snabba att använda. Texten går att bearbeta snabbt.
NVIVO	Ganska krävande förberedelser innan man behärskar programmets funktioner (minimalt 2-3 dagar).	Många funktioner, överskådliga men texten måste bearbetas länge.
Kwalitan 5	Enkel introduktionsmanual med hjälp av vilken det går snabbt att sätta sig in i programmet. (ca 4 timmar) För att få ut mesta möjliga krävs dock en del övning innan programmet behärskas (någon dag) ¹² .	Många funktioner, överskådliga, beroende på vad forskaren vill få ut krävs det olika tid för bearbetning.

⁸ Nyckelorden visas ett i taget och forskaren får själv etablera sambanden

⁹ Forskaren kan etablera interrelationer mellan olika nyckelord efter eget intresse. Dessa kan sedan länkas ihop i flera dokument och på så sätt ge flerdimensionella möjligheter till analys. Dessa dokument kan presenteras bredvid varandra, vilket underlättar analysen.



¹⁰ Resultaten presenteras under rubrikerna 1) “Words in context”(mycket överskådligt), 2) “Textfragments belonging to a code” (mycket överskådligt), 3) “Print segments” (mindre överskådligt). Exempel återfinns i Bilaga 2.

¹¹ Dessa bedömningar motsvarar en person som har datorvana.

¹² I demonstrationsversionen finns inte fullständiga hjälpverktyg tillgängliga varför alla funktioner säkert inte upptäckts eller kunnat utnyttjas. Programmet är ursprungligen på holländska och i demonstrationsversionen är vissa menyval och instruktioner på detta språk varför det kan vara lite krångligt att förstå.

Pris

	Kostnad för anskaffning	Kommentar
Concordance (England)	Pris i SEK ¹³ 1 användarlicens 522 andra användarlicensen 225 därefter 315 för varje ytterligare användarlicens	Programmet laddas ner från hemsida, licensnummer skickas med e-post. Manual kan laddas ner.
NVIVO (Australien, kontor i England)	Pris i SEK, exklusive VAT ¹⁴ 1 användarlicens 2.959 5 användarlicenser 10.407	I priset ingår CD och fullständiga manualer.
Kwalitan 5 (Nederländerna)	Pris i SEK, exklusive VAT ¹⁵ 1 användarlicens 2.837 2 användarlicenser 5.629 3-5 användarlicenser 8.647 6-10 användarlicenser 11.484 11-20 användarlicenser 14.366 21-30 användarlicenser 17.248 31-40 användarlicenser 20.131	Framgår inte hur programmet levereras.

Slutsatser

Om vi jämför de tre programmen konstaterar vi att NVIVO inte erbjuder någon tjänst som inte Kwalitan 5 gör angående hanteringen av rådata, krav på organisering, kodning etc. Fördelen med Concordance är att programmet hanterar enkla, obearbetade textfiler. Nackdelen är att analysmöjligheterna begränsas till endast ordsökning. Kwalitan 5 kräver Rtf-filer men fördelen är att man kan göra en djupare analys genom en kodning såväl manuellt som maskinellt.

Angående kontroll av förekomster görs det automatiskt av Concordance. Kwalitan 5 erbjuder möjligheten att kontrollera förekomster inom vissa "mikroinstanser" men då krävs insatser av forskaren. Det sista gäller också för NVIVO men detta program erbjuder ingen automatisk kontroll av förekomster.

När det gäller presentation av resultaten kräver både NVIVO och Kwalitan 5 specifika filformat medan Concordance presenterar resultat i html-format vilket innebär att det inte krävs någon speciell programvara för att kunna läsa resultatfilerna, det räcker med en webläsare. Överskådligheten i presentationen av resultaten är hög hos Concordance därför att komplexiteten i analysen som kan utföras är lägre. Hos NVIVO och Kwalitan 5 är överskådligheten i presentationen lägre på grund av högre komplexitet i analysen men när Kwalitan 5 används för en enklare analys är presentationen lika överskådlig som hos Concordance.

¹³ Uppgraderingar ingår i priset

¹⁴ Frakt tillkommer, priset är för användning i utbildnings och forskningssyfte, för 12 månaders fri uppgäradering tillkommer 85 USD för en licens och 380 USD för 5 licenser.

Priset för kommersiell användning är 735+140 för en licens respektive 2.580+600 för 5 licenser.

¹⁵ Priset är enligt hemsidan per oktober 2001, dock är alla andra uppgifter uppdaterade, så hur väl detta datum stämmer kan ifrågasättas. För beställning från Europeiskt land utanför Nederländerna tillkommer 5 Euro i porto.

Naturligtvis är Concordance med sin enklare typ av analys ett program som man lätt lär sig att använda. Både NVIVO och Kwalitan 5 som erbjuder en rikare typ av analys kräver mer tid men det senare är enklare att börja bekanta sig med. När det gäller reell användning är NVIVO betydligt mer komplicerat.

Sammanfattningsvis visar jämförelsen av de tre programmens funktioner att Kwalitan 5 är överlägset de två övriga. Med tanke på de funktioner som erbjuds är det naturligt att Concordance är billigast men en jämförelse av priset mellan NVIVO och Kwalitan 5 utfaller även den till Kwalitan 5's fördel. För en enklare typ av analys rekommenderas dock Concordance som förutom att det är billigt, är enkelt att hantera, inte kräver någon särskild filsort för rådata och presenterar resultaten i html-format.

Det är värt att till sist nämna att vi med ovanstående utvärdering som grund har anskaffat Kwalitan 5 och att den använda analysmodellen har applicerats på motsvarande material från två andra stora svenska företag, TeliaSonera och Skandia. Resultatet av denna analys avrapporterar i Baldvinsdottir et al (2005) "*Demonstration of what? Rhetorical approaches in chief executive's reviews in times of crisis*", skickad till European Accounting Association 28th Annual Congress", Göteborg 18-20 May, 2005.

Referenser

Brown, G. & Yule, G., (1983), *Discourse Analysis*, Cambridge, Cambridge University Press.

Fairclough, N., (2003), *Analysing discourse, Textual analysis for social research*, London: Routledge.

Firth, J. R., (1935), *The technique of semantics. Transactions of the Philological Society.*

Gerbig, A., (1997), *Lexical and grammatical variation in a corpus: a computer-assisted study of discourse o the environment*, NY: Lang.

Handelshögskolan i Stockholm, (2003), *Programvaror för bearbetning av kvalitativa data.*

Halliday, M., (1994), *An Introduction to Functional Grammar*, London: Edward Arnold.

Leech, G. & Fallon, R., (1992), Computer Corpora-what do they tell us about culture? *ICAME Journal*, 16, 29-50.

Steiner, G., (1998), *After Babel*, Oxford University Press.

Stubbs, M., (1996), *Text and Corpus Analysis: computer-assisted studies of language and culture*, Cambridge, Mass.: Blackwell.

Vårt urval av tillgängliga programvaror

1. NVivo och 2. N6 - Internetadress: <http://www.qsrinternational.com>
3. ATLAS-Ti - Internetadress: <http://www.atlasti.de>
4. Ethnograph v5.0 - Internetadress: <http://www.qualisresearch.com>
5. Qualrus - Internetadress: www.qualrus.com
6. CDC EZ – TEXT - Internetadress: <http://www.cdc.gov/hiv/software/ez-text.htm>
7. C-I-SAID - Internetadress: <http://www.code-a-text.co.uk>
8. HyperRESEARCH - Internetadress: <http://www.researchware.com>
9. MAX qualitative data analysis - Internetadress: <http://www.maxqda.com>
10. Kwalitan 5 - Internetadress: <http://www.kwalitan.net>
11. AnSWR - Internetadress: <http://www.cdc.gov/hiv/software/answr.htm>
12. Visual text - Internetadress: <http://www.textanalysis.com>
13. Wordstat - Internetadress: <http://www.simstat.com>
14. Concordance - Internetadress: <http://www.rjcw.freemove.co.uk/index.htm>