

NORDICOM

Medie- och kommunikationsforskning i Norden

INFORMATION

DISTANSLÄSNING OCH NÄTVERKSANALYS

**Digitala metoder
för gamla och nya medier**

Redaktörer: Johan Jarlbrink & Ingela Wadbring

2016 nr 3 | årgång 38

Redaktörer

Johan Jarlbrink och Ingela Wadbring,
Nordicom, Göteborgs universitet, Sverige

Anvarig utgivare

Ingela Wadbring,
Nordicom, Göteborgs universitet, Sverige

Redaktionsråd

Karin Poulsen, Nordicom,
Göteborgs universitet, Sverige

Eija Poteri, Nordicom,
Tammerfors universitet, Finland

Mogens Verstergaard Kjeldsen, Nordicom,
Statsbiblioteket i Århus, Danmark

Friðrik Þór Guðmundsson, Nordicom,
Íslands universitet

Koordinering och korrektur

Karin Poulsen

Teknisk redigering och layout

Per Nilsson

Prenumerationsärenden

Anne Claesson,
anne.claesson@nordicom.gu.se

Nordicom, Göteborgs universitet

Tryck

Ale Tryckteam AB, Bohus 2016



ISSN 0349-5949

ISBN 978-91-87957-45-1

Nordicom-Information är en tidskrift på de nordiska språken, som utges både tryckt och online. Hösten 2015 bytte den redaktörer, form och innehåll.

Ambitionen med *Nordicom-Information* är att den ska kunna fungera som en mötesplats för akademien, mediebranschen och politiken inom de nordiska länderna. Det finns såväl likheter som skillnader mellan de fem nordiska länderna samt Grönland, Färöarna och Åland. På ett övergripande plan är det möjligt att tala om en stark samhörighet som bland annat manifesterar sig i det nordiska samarbetet. Skillnaderna är ofta inte större än att det finns ett intresse och en potential att lära av varandra.

Nordicom-Information utkommer två gånger per år. Innehållet är tematiskt, och skribenterna är personer verksamma inom akademien, politiken och/eller medieindustrin, beroende på ämnesområde.

Utöver de två reguljärt utkommande utgåvorna varje år är det möjligt att ge ut specialnummer av *Nordicom-Information*. Även dessa är tematiska, men helt forskningsinriktade. Avsikten är att forskare inom de nordiska länderna ska kunna publicera sig på sitt eget modersmål, med en formell granskningsprocess. Sådana specialnummer kan utkomma flera gånger per år.

Nordicom-Information publiceras som open access på <http://nordicom.gu.se/sv/publikationer/nordicom-information>.

Innehåll

- 3 Förord
- 5 *Johan Jarlbrink*
Inledning – digitala metoder
- 9 *Fredrik Norén*
Information som lösning, information som problem.
En digital läsning av tusentals statliga utredningar
- 27 *Johan Jarlbrink, Pelle Snickars & Cristian Colliander*
Maskinläsning. Om massdigitalisering, digitala metoder
och svensk dagspress
- 41 *Jonas Andersson Schwarz & Johan Hammarlund*
Kontextförlust och kontextkollaps.
Metodproblem vid innehållsanalys av sociala medier
- 57 *Pernilla Severson*
Medvetna frågor och val. Utveckling av metoder för att undersöka
användargenererat material i digitala kontexter
- 71 Humanistlaboratorier i Norden

Recension och presentation av litteratur

- 75 *Maria Eriksson:*
Richard Rogers: Digital Methods
- 79 Bøger om digitale metoder: et udvalg 2013-2016

Förord

Sommaren 2015 gick Nordicom ut med ett Call for Papers om digitala metoder för studier av digital(isera)t mediematerial till forskare i de nordiska länderna. I callet stod att läsa:

Digitalt tillgängligt mediematerial möjliggör användandet av digitala metoder för att studera innehåll, vare sig detta innehåll har digitalt ursprung eller har gjorts digitalt tillgängligt i efterhand. Under senare år har bibliotek, arkiv och museer digitaliserat stora delar av sina historiska samlingar. Samtidigt skapar mängden material på internet och i sociala medier nya förutsättningar för insamling och bearbetning. Mängden material ställer forskare inför stora utmaningar, men med det digitala formatet finns det samtidigt goda möjligheter att använda digitala redskap i analysarbetet. Såväl texter, som visuella material och nätverk låter sig bearbetas. Metodologiska perspektiv som 'topic modeling', 'distant reading' och 'cultural analytics' erbjuder alla en rad förslag till metodologiskt nytänkande. Detta specialnummer av Nordicom-Information bjuder in forskare att utforska, tillämpa, diskutera och kritisera de digitala metoder som står medieforskare till buds. Artiklarna kan kretsa kring, men är inte begränsade till, följande områden:

- Innehållsanalyser av samtida eller historiska mediematerial
- Nätverksanalyser
- Digitala metoder inom omvärldsbevakning och 'media monitoring'
- Datavisualisering
- Digitala metoders problem och möjligheter

Ett antal forskare hörsammade callet, och fyra forskningsartiklar presenteras i detta nummer av Nordicom-Information, tillsammans med presentationer av litteratur och forskningsmiljöer inom området.

De fyra artiklarna har genomgått sedvanlig granskning för forskningsartiklar, så kallad double-blind peer review. I arbetet med specialnumret har Johan Jarlbrink ansvarat för sakkunskapen, medan vi båda har samarbetat kring den praktiska delen av arbetet.

Nordicom-Information är en tidskrift på de nordiska språken, som utges både tryckt och online, av Nordicom vid Göteborgs universitet. Sedan hösten 2015 utkommer två ordinarie nummer per år som är ämnestematiska, där även andra än forskare inbjuds att delta. Dessa ordinarie nummer innehåller *inte* granskade artiklar, utan är tänkta att fungera som en mötesplats för forskning, politik och bransch och kan snarast karaktäriseras som populärvetenskapliga. Därutöver ges tematiska specialnummer ut – som detta om digitala metoder – och de innehåller i huvudsak granskade artiklar.

Göteborg i december 2016

Ingela Wadbring
Föreståndare, Nordicom

Inledning – digitala metoder

I en tid när det mesta medieinnehåll föds digitalt och när historiska samlingar samtidigt digitaliseras skapas möjligheter att utforska medier med hjälp av digitala metoder. En metodologisk uppdatering kan innebära såväl förändrade frågor och upptäckten av nya studieobjekt, som digitala redskap att använda i forskningen (Jensen 2011). Datorernas förmåga att beräkna och sortera gör det möjligt att få fram frekvenser och samförekomster i stora materialsamlingar, hur texter och medieanvändare relaterar till varandra, och hur det enskilda förhåller sig till den större massan av data. Samtidigt behöver digitala analysstrategier inte nödvändigtvis vara kvantitativa. Som Rogers (2013) betonat kan digitala metoder även innebära att följa och utforska hur exempelvis en sökmotor organiserar data och villkorar tillgången till den. Internationellt finns ett växande intresse för digitala forskningspraktiker och ambitionen med detta temanummer är att presentera aktuella metoddiskussioner inom fältet.

Helt nya är metoderna inte – särskilt inte inom medieforskningen. Redan under 1960-talet användes datorer för att hitta kvantitativa mönster i exempelvis tidningstexter (Jarlbrink 2015). Och inom datavetenskap, informatik och språkvetenskap har det under flera decennier utvecklats metoder för att analysera texter och nätverk digitalt. Humanistiskt och samhällsvetenskapligt orienterade medie- och kommunikationsforskare har goda möjligheter att bygga vidare på denna forskning och använda de verktyg som redan finns. Samtidigt finns det förutsättningar att urskilja och

ställa andra typer av frågor än dem som ofta intresserat föregångarna inom andra discipliner (Burdick m.fl. 2012). Hur kan man exempelvis förstå digitala nätverk ur ett maktperspektiv? Och hur kan man förklara diskursiva förändringar i tiotusentals texter över tid? I detta ligger också en utmaning: Med så många forskningsmöjligheter till hands är det ibland svårt att komma på vilka frågor som är angelägna att ställa.

Att anamma en digital forskningspraktik får emellertid inte innebära att forskare bortser från de problem som digitala metoder kan föra med sig. Med tillgången till och möjligheten att undersöka stora datamängder kommer ofta föreställningen om att storskaligheten representerar en högre form av vetande (Boyd & Crawford, 2012). Som påpekats av bland andra Lagoze (2014) och Gitelman & Jackson (2013) bygger även stora datamängder på urval som kan snedvrider resultaten. Data är aldrig 'rå' utan har alltid skapats och samlats med specifika syften. Tillgången både till samtida flöden i sociala medier och till digitaliserade historiska samlingar är dessutom ofta reglerad och beroende av urvalsmekanismer som enskilda forskare har svårt att kontrollera och få insyn i. Det är heller inte säkert att innehållet i datamängderna är vad det utges för att vara – automatiskt genererade samlingar innehåller ofta 'rutten' data såsom felkategoriserade texter, spam och avbrutna länkar (Boellstorff 2013). Samtidigt som digitala metoder gör stora datamängder lätt hanterbara riskerar de att urskilja mönster som inte existerar. Av forskare krävs ett kritiskt förhållningssätt både till

hur data har tillkommit och samlats in och till de digitala verktygens sätt att bearbeta dem. Här utöver bör det finnas beredskap att kombinera kvantitativa metoder med kvalitativa.

Men icke desto mindre: En medievetenskap som vill utforska medielandskapet i all dess mångfald behöver behärska en bred metodarsenal – och här blir digitala metoder en allt viktigare del. Alla metoder bär spår av de medier som de utvecklats för att studera. Att överföra analoga metoder till digitala sammanhang kan därmed innebära att forskningen förbiser det digitalt mediespecifika (Hayles 2004). Kommersiella aktörer använder redan digitala verktyg för att få syn på trender, nätverk och beteenden. Medieforskningen behöver hitta sina egna angreppssätt och verktyg, men för att en sådan utveckling ska ta fart krävs att fler forskare lär sig hantera och modifiera de digitala metoder som redan finns.

Artiklarna som samlats i detta nummer behandlar och utvecklar flera av de teman som antytts ovan. Några presenterar fallstudier där de digitala metodernas möjligheter och problem illustreras, andra diskuterar digitala metoder mer principiellt. Två av studierna beskriver hur digitala metoder för textanalys kan användas för att undersöka digitaliserat historiskt material, medan två studier berör metodfrågor i anslutning till flöden i sociala medier.

Fredrik Norén undersöker hur begreppen 'information' och 'samhällsinformation' blir en del av den svenska efterkrigstidens politiska diskussioner. Utgångspunkten är ett stort antal utredningar, utgivna i serien Statens Offentliga Utredningar (SOU, 1922-), som nyligen gjorts digitalt tillgängliga. Norén inleder sin studie med att ta reda på hur frekvenserna av ordet 'information' förändras från 1920-talet till 1990-talet, och undersöker därefter hur ordet användes och vilka politikområden det blev en del av, från 1960-tal till 1980-tal. Med hjälp av verktyget MALLET visar Norén hur samhällsproblem under den undersökta perioden börjar beskrivas som informationsproblem, samtidigt som samhällsinformationen blir viktig inom flera av statens verksamhetsområden.

Johan Jarlbrink, Pelle Snickars och Cristian Colliander beskriver hur digitala verktyg för textanalys kan användas i studier av äldre tidningspress. I artikeln tecknar de den mediehistoriska bakgrunden till bibliotekens pågående digitaliseringsprojekt och beskriver den omvandling som historiska tidningar genomgår när de digitaliseras. När det pappersbaserade kulturarvet har OCR-tolkats finns nya möjligheter att hitta mönster i de texter som tidigare har varit svåra att söka i och utforska. Digitaliseringen av äldre tidningspress är emellertid förbunden med flera problem: kvaliteten på OCR-läsningen är ofta bristfällig, samtidigt som stora delar av materialet är fortsatt otillgängligt på grund av upphovsrättsskäl. För att illustrera de möjligheter som ändå finns presenterar författarna en fallstudie om hur det nya mediet Internet presenterades för 1990-talets tidningsläsare.

Jonas Andersson Schwarz och Johan Hammarlund behandlar i sin artikel de mediala och sociala kontexter som digitala analyser av twitterdata både kan framhäva och riskerar att förvränga och osynliggöra. Utifrån ett antal fallstudier av twitterdebatter diskuterar de hur digitala metoder kan användas för att analysera användare och nätverken de bildar, både inom forskningen och inom omvärldsbevakningsbranschen. Kvalitativa analyser riskerar här ofta att framhäva enskilda exempel som inte nödvändigtvis är representativa för den större datamängden. Kvantitativa analyser riskerar å andra sidan att föra samman olika användare på ett sätt som skapar sammanhang som egentligen inte finns. För att hantera kontexter på ett rimligt sätt krävs att forskare kombinerar kvantitativa metoder med en kvalitativ lyhördhet för materialets egenart.

Pernilla Seversons artikel behandlar hur användargenererat material på internet kan förstås och undersökas utifrån tre forskningsperspektiv och referenspunkter: framhävandet av publikens njutning inom kulturstudieforskningen, designforskningens persona-begrepp, samt medieinnehållet betraktat som data, så som det framträder i ett datavetenskapligt perspektiv. Med publikens njutning i åtanke framstår det som intressant att studera medie-

användarnas egna drivkrafter, samtidigt som det med designforskningens betoning av medieplattformarnas gränssnitt och struktur är viktigt att undersöka hur användargenererat material villkoras av de tekniker som samtidigt utgör dess förutsättning. Med medieinnehåll i digitalt format blir det möjligt att samla in och analysera det med hjälp av digitala verktyg, men Severson betonar samtidigt att detta kan

föra med sig såväl etiska som analytiska problem. Kvantitativa metoder bör därför kombineras med kvalitativa.

Härutöver innehåller temanumret presentationer av digitala forskningsmiljöer, en recension av Richard Rogers bok *Digital Methods*, samt en översikt över nyutkommen litteratur inom området.

Referenser

- Boellstorff, Tom (2013). Making Big Data, in Theory. *First Monday*, 18(10): 1-17.
- boyd, danah & Crawford, Kate (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5): 662-679.
- Burdick, Anne, Drucker, Johanna, Lunenfeld, Peter, Presner, Todd & Schnapp, Jeffrey (2012). *Digital Humanities*. Cambridge & London: MIT-Press.
- Gitelman, Lisa & Jackson, Virginia (2013). Introduction, pp. 1-14 in Gitelman, Lisa (ed.) *"Raw Data" Is an Oxymoron*. Cambridge & London: MIT-Press.
- Hayles, Katherine N. (2004). Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis. *Poetics Today*, 25(1): 67-90.
- Jarlbrink, Johan (2015). Historievetenskapens mediehantering, pp. 225-247 in Hyvönen, Mats; Snickars, Pelle & Vesterlund, Per (eds.) *Massmedieproblem: Mediestudiets formering*. Lund: Mediehistoriskt arkiv.
- Jensen, Klaus Bruhn (2011). New Media, Old Methods: Internet Methodologies and the Online/Offline Divide, pp. 43-58 in Consalvo, Mia & Ess, Charles (eds.) *The Handbook of Internet Studies Edited*. Chichester: Blackwell Publishing.
- Lagoze, Carl (2014). Big Data, Data Integrity, and the Fracturing of the Control Zone. *Big Data & Society* 1(2): 1-11.
- Rogers, Richard (2013). *Digital Methods*. Cambridge & London, MIT-Press.

Johan Jarlbrink är docent i mediehistoria och lektor i medie- och kommunikationsvetenskap vid Umeå universitet. Han forskar om 1800-talets tidningspress i alla dess skepnader: tidningen som informationsbärare i arkiv och klippböcker, som avfall och omslagspapper, som forskningsdata i digitala databaser.

Fredrik Norén

Information som lösning, information som problem

En digital läsning av tusentals statliga utredningar

Abstract

This article investigates how the Swedish information politics emerged in the twentieth century by analysing thousands of Swedish Governmental Official Reports (8000 reports since 1922). By using methods such as topic modelling with LDA/MALLET, the result shows that a governmental information discourse arose in the 1960s, infiltrating a growing number of political interests, and that the content of the discourse has changed over time.

One important conclusion highlights digital text methods as a more inductive way of doing content analysis than is commonly practised in media and communication studies. A machine can find patterns in texts that no single person could read in a lifetime. Furthermore, the quantitative approach can generate results that problematise previous – qualitative – research. Digital and quantitative analysis of a huge corpora does not neglect close reading; on the contrary this study shows the necessity of bridging both methods in order to better understand the results.

Keywords: information politics, topic modeling, digital methods, Swedish governmental official reports

[...] man kan märka en viss informations-
trötthet hos många människor vad gäller
vissa former av information.

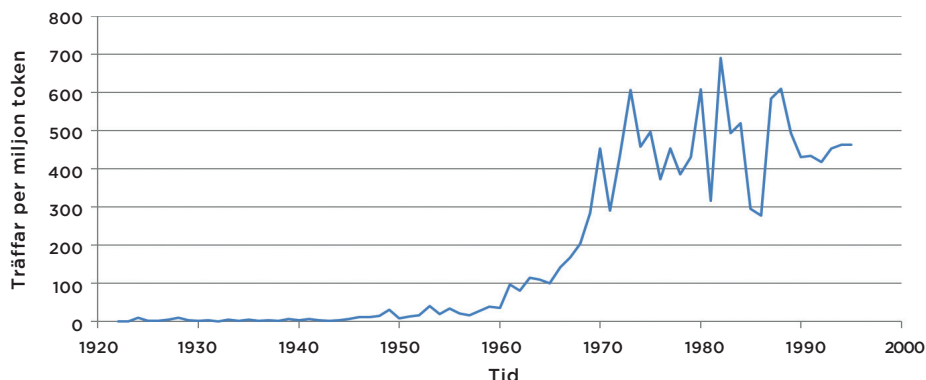
(SOU 1972: 52 *Rapport angående
kommunal information m.m.*)

En enkel linje kan fånga en hel samhällsförändring. I en *Science*-artikel från 2011 analyseras språkliga förändringar i miljontals digitaliserade böcker, vid tidpunkten fyra procent av alla böcker som någonsin tryckts. Studien visar till exempel hur frekvenserna av orden 'men' och 'women' förändras över tid: först två linjer med ett avgrundsdykt avstånd mellan varandra i början av 1800-talet, för att i slutet av 1900-talet närma sig varandra tills 'women' slutligen går om 'men' i frekvens (Michel m fl 2011). Utgångspunkten för min artikel är linjen i Figur 1, och hur statens in-

formationspolitik, språkligt, växte fram under framförallt 1960-talet och 1970-talet.¹ Vad var det egentligen som hände vid kurvans dramatiska ökning?

Forskare och utredningar har beskrivit 1960-talet och 1970-talet som perioden då vår tids informationspolitik växte fram. Men hur denna utveckling har diskuterats och framförallt mätts har präglats av tydliga metodologiska begränsningar. Perioden lyfts ofta fram som samhällsinformationens tid: myndigheter skulle informera medborgarna om sina verksamheter samtidigt som en ny informationsteknologi och samtida kommunikationskanaler bar på nygamla löften – och hot – om demokratins framtid och om ökat informationsutbyte mellan medborgare och myndighet (se t ex Kjellgren 2002; Hyvönen, Snickars & Vesterlund 2015; SOU 1969: 48; SOU 1984: 68). Men hur

Norén, Fredrik (2016). Information som lösning, information som problem. En digital läsning av tusentals statliga utredningar. *Nordicom-Information* (38)3: 9-26.



Figur 1. Frekvensförändring av ordet 'information' (lemgram) i Statens offentliga utredningar

kan en sådan bred och spretig utveckling studeras med ett samlat grepp?

Forskare har hittills hänvisat till vaga eller problematiska källor när de diskuterat intensifieringen av samhällsinformationen under 1960-talet och 1970-talet, om det har refererats överhuvudtaget. Det finns några förklaringar till varför det förhåller sig på det viset. Ibland har utvecklingen tagits för given eftersom alla vet att det var så. I andra fall har den betraktats som ett osynligt politiskt fält och inte ansetts nödvändig att problematiseras. En tredje förklaring är att förändringen av statens informationspolitik och informationssamhällets framväxt, helt enkelt har varit *svår att mäta* (Webster 2002: kap 2). Vad gäller den sista förklaringen refereras det på sin höjd till enskilda och 'talande' exempel, ökade medlemstal i PR-föreningar eller till äldre, självuppskattande enkätundersökningar, där myndigheter själva har fått ange hur stor andel av sin budget som har lagts på extern informationsverksamhet (se t ex Kjellgren 2002; Larsson 2005; *Nordicom-Nytt* 1984: 2-3). Även om dessa mått kan vara en indikation på förändring ger de inte sällan läsaren fler frågor än de bär på svar. Och ju mer tid som förflyter från händelsernas centrum, desto fler och större blir eftervärldens frågetecken.

Den här artikeln visar hur digitala metoder och stora samlingar text tillsammans kan

språkliggöra politiska strömningar i historien på andra sätt än vad som tidigare har varit möjligt, generellt såväl som partikulärt. I detta specifika fall används tusentals betänkanden i serien Statens offentliga utredningar (SOU) för att synliggöra den statliga informationspolitikens förändring, från 1960-talet till 1980-talet. Den informationspolitik som här åsyftas är de statliga myndigheternas informationsaktiviteter, i vid bemärkelse.

Utifrån dessa förutsättningar ställs följande frågor: vilken information var det som diskuterades och hur såg informationsdiskursen ut? Vilka politiska verksamhetsområden växte informationsdiskursen in i och hur förändrades den över tid? Studiens explorativa anslag bygger på övertygelsen att digitaliseringen av stora textsamlingar har skapat behov av andra tillvägagångssätt för att studera samhällsförändringar. Det öppnar också upp för möjligheter som ännu inte har en tillräcklig metodologisk standardisering, inte minst inom medie- och kommunikationsforskningen, och därför står också den metodologiska processen i studiens centrum.

En diskussion om material och definitionen av statens röst

Innan regeringen ska utforma en proposition anses det ibland vara nödvändigt att ordentligt utreda lagförslagets olika alternativ. Reger-

ingen, eller ett enskilt departement, tillsätter då en kommitté som ges i uppdrag att utreda frågan vars resultat utgörs av ett så kallat slutbetänkande, en statlig offentlig utredning. Betänkandet skickas därefter ut på remiss innan regeringen slutligen tar ställning till utredningsförslagen och remissvaren. Ibland följer propositionen kommitténs förslag till punkt och pricka, i andra fall bara delvis med hänsyn till remissvaren och i vissa fall väljer regeringen att bordlägga frågan på obestämd tid. Efter klassisk svensk modell består kommittémedlemmarna i regel av en grupp personer som representerar olika samhällsintressen i den aktuella utredningsfrågan. Medlemmarna är under utredningstiden anställda av staten och kommittén räknas som en egen myndighet, men utan ekonomisk frihet (Bäck & Larsson 2006: 213-217). Den kontinuerliga, gedigna och ämnesbreddade SOU-produktionen utgör som helhet en unik källa för att studera de generella trenderna i den svenska statens natur, intressen och attityder över tid. Det är därför möjligt att använda det samlade materialet som en representation av 'statens röst'.

2015 avslutade Kungliga biblioteket (KB) digitaliseringen av SOU-materialet, från 1922 till och med 1999 (regina.kb.se/sou/; resterande betänkanden finns digitaliserade på data.riksdagen.se/). 1900-talets SOU:er utgör en korpus på över 20 miljoner meningar fördelade på 5600 utredningar. Antalet betänkanden har ökat över tid, inledningsvis omkring 50 per år för att sedan successivt bli fler och fler, och i dag går årsproduktionen aldrig under 100. En utredning kan vara allt från ett tiotal sidor till uppåt 1000-sidiga dokument. Materialet, i sin digitala form, är tämligen utforskat med undantag för några explorativa exempel (se t ex Kullenberg 2015). Många frågeställningar återstår med andra ord att bli besvarade.

Kvaliteten på det digitaliserade materialet, det vill säga hur väl det överensstämmer med källmaterialet (de tryckta SOU:erna) är relativt god, men knappast perfekt och olika problem har uppstått när det scannats och genomgått Optical Character Recognition (OCR). Typiska fel är till exempel att ma-

skininläsningen inte registrerat mellanslag och därför gett korrupta ord som till exempel 'tillkommakonstitutionsutskott'. I andra fall misstas en glyf (teckenbild) för en annan, 'förhållande' blir 'ferhållande', och ibland uppstår mer svårtolkade fel som 'tjenestemandsvædtaegter'. Ett annat problem som uppstått under digitaliseringen är hanteringen av text som återfinns i tabeller, listor eller diagram. Då har ordens ordning ofta kastats om eller bilder översatts till oläsbara ordföljder. Men överlag har stickproven av materialet visat på en kvalitet som är tillräckligt god för att utföra digitala och kvantitativa textanalyser, där det enskilda ordet blir underställd den totala textmängden (jmf Jarlbrink, Snickars & Colliander 2016).

Språkbanken, den språkteknologiska forskningsenheten vid Göteborgs universitet, har tillgängliggjort KB:s digitaliserade SOU-samling via sitt konkordansverktyg, Korp. Den öppna molntjänsten har olika sökalternativ som artikelns första analysdel grundar sig på (Borin, Forsberg & Roxendal 2012). Min undersökning använder sig även av Språkbankens xml-transkriberingar av SOU-materialet, vilket utgör en grund för att maskinellt kunna urskilja olika ordklasser och hantera ords ändelser (spraakbanken.gu.se/swe/resurs/sou). De Python-script som används i analysen är framtagna i dialog med Roger Mähler på Humlab vid Umeå universitet.

SOU-materialet har tidigare använts som en viktig källa för forskning om politisk-byråkratiska attityder till olika frågor. Oftast har det handlat om kvalitativa läsningar av enskilda SOU:er. Ett sådant exempel är statsvetaren Hanna Kjellgrens avhandling om den statliga informationspolitikens utveckling i Sverige utifrån från nerslag i några av 'de mest centrala faserna i dess utveckling', framförallt 1940-talet och 1970-talet (Kjellgren 2002: 43f). Både tidsglappen och de på förhand valda tidsperioderna, med tillhörande politiska dokument, skapar en något fragmentarisk helhet, vilket blir en naturlig konsekvens av närläsningssmetoden. Olika metoder hjälper oss att se olika saker och med dagens digitala teknik hade Kjellgrens undersökning kunnat

göras annorlunda, vilket analysen längre fram kommer att visa.

Avståndsläsning

Att studera hur språkliga förändringar över tid relaterar till samhälleliga skiftningar har sedan länge praktiserats genom kvalitativa metoder. Ryan Heuser och Le-Khac Long lyfter fram Raymond Williams *Culture and Society* (1958) som ett sådant exempel där förändringar i samhället studerades utifrån skönlitterära närläsningar. Men eftersom Williams saknade både digitala metoder, och storskaliga digitala korpusar, försvårades möjligheten att avgöra kvalitativa trendförändringars kvantitativa signifikans (Heuser & Long 2012: 2: 19). Litteraturvetaren Franco Moretti har gjort liknande studier utifrån ett kvantitativt perspektiv och ett byråkratiskt textmaterial. Ett exempel är hans samarbete med sociologen Dominique Pestre där de analyserade Världsbankens årsrapporter från 1946 till 2012. Deras undersökning visar hur bankens språk över tid har blivit allt mer tjockt och abstrakt, som ett sätt att både dölja och förena motstridiga värden (Moretti & Pestre 2015).² Världsbankstudien heter talande nog *Bankspeak*, och anspelar på dokumenten som bankens röst. På samma sätt används här Statens offentliga utredningar som en representation för den svenska statsapparaten.

Avståndsläsning eller *distant reading* som myntades av Moretti, anspelar på motsatsen till *close reading*. Genom att ta ett kliv tillbaka från själva texten kan forskaren upptäcka mönster som annars hade varit omöjliga att studera. Morettis idéer utvecklades främst som en motreaktion till världslitteraturens kanon och som ett sätt att i stället närma sig dess fält, inte utifrån dess separata delar (enskilda romaner som anses mer betydelsefulla än andra) utan utifrån dess helhet (alla romaner) (Moretti 2000; Algee-Hewitt m fl 2016; Jockers 2013). Varje metodval innebär att något annat trängs bort, vilket också utgör en av Morettis polemiska huvudpånger:

If we want to understand the system in its entirety, we must accept losing some-

thing. We always pay a price for theoretical knowledge: reality is infinitely rich; concepts are abstract, are poor. But it's precisely this 'poverty' that makes it possible to handle them, and therefore to know. (Moretti 2000: 57f).

Men utifrån ett metodologiskt perspektiv är storskaligheten bara ena sidan av det Janus-ansikte som det digitala skiftet står för. Det digitala möjliggör en flexibel växelverkan mellan å ena sidan det makroskopiska och å andra sidan det mikroskopiska. Katharine Hayles menar därför att denna maskinella läsning inte ska ses som en negligering av – eller hot mot – studier av det partikulära. Snarare utgör avståndet och närheten en syntetisk interaktion, som i bästa fall kan skapa ett ömsesidigt beroende som stärker båda perspektiv, vilket även har poängteras av Morettis egna kritiker (Hayles 2012: 80f; Ross 2014: paragraf 6).

Maskinella kvantitativa textanalyser har sedan länge praktiserats inom språkvetenskapen och utgjorde den första inriktningen inom fältet som senare kom att kallas humanities computing, och som nu bytt namn till digital humanities. Startskottet brukar dateras till prästen Roberto Busas samarbete med IBM under 1940-talet för att maskinellt indexera alla ord som skrivits av St. Thomas Aquinas, något som 1974 resulterade i ett hålkorts-konkordansprogram. Denna typ av korpuslingvistik har sedan 1960-talet fått ett allt större genomslag och fokuserar traditionellt på frekvensfigurer (hur enskilda ord förändras över tid) samt konkordanser (hur ords sammanhang förändras över tid) (Hockey 2004; Lindquist 2009: kap 1; Biber, Conrad & Reppe 2006: kap 1)

Även inom medie- och kommunikationsforskningen finns en lång tradition av kvantitativa textstudier. Men det digitala har ändrat förutsättningarna för dessa metoder. I ämnets läroböcker får vi till exempel lära oss att före en genomgång av empirin måste det väljas teori och därefter variabler som ska studeras (t ex Berger 2014: kap 11). Digitala metoder vänder upp och ner på detta och erbjuder ett mer induktivt tillvägagångssätt. Maskinläs-

ning behöver nämligen minimal input för att på egen hand identifiera olika diskurser i en större textmassa. Forskaren kan således i efterhand identifiera vad som anses intressant och välja teoretiskt ramverk. En algoritm kan med andra ord ge andra resultat än den mänskliga förhandskodningen.

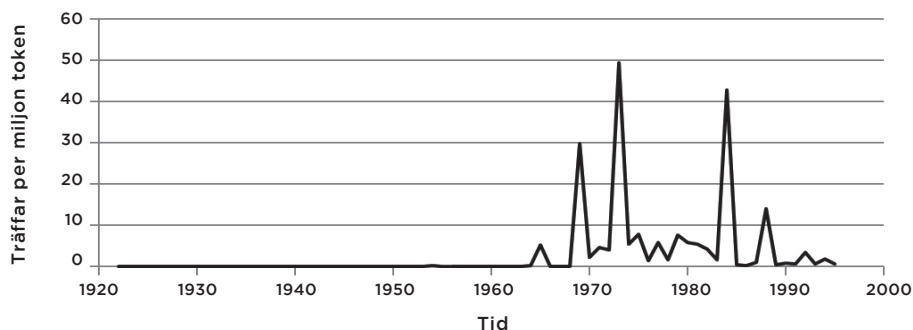
Begreppen *signal* och *koncept* kan användas som ett sätt att beskriva rörelsen från siffror till mening. En signal utgör i detta fall något som kan identifieras, spåras och analyseras maskinellt. Koncept är å andra sidan det som vi bedömer att en signal kan stå för. För forskaren är det koncepten som verkligen betyder något, och i slutändan handlar det om att försöka bygga en bro från signal till koncept (Heuser & Long 2012: 46-48). I mitt fall utgör signalerna ord, både individuella och i kluster, medan informationspolitikens förändring utgör konceptet.

De senaste decenniernas digitaliseringsprojekt har gjort oss än mer medvetna om det banala faktum att det helt enkelt finns för mycket att läsa (Nicholson 2012: 238). Matthew L. Jockers menar att flera länders digitaliseringsproduktion har börjat göra det möjligt och fruktbart att närma sig stora textsamlingar utifrån digitala metoder och kvantitativa perspektiv (2013: 4). Denna tipping point gäller inte bara skönlitteraturen, som Jockers och Moretti främst studerar, utan även den offentliga byråkratin. Särskilt i länder som Sverige, med en lång tradition av offentlighetsprincip i kombination med en relativt långt kommen digitaliseringsutveckling, finns en potential för omfattande digitala studier av den offentliga textproduktionen (se texdata.riksdagen.se). Detta öppnar upp för andra perspektiv och resultat, både för forskare som tidigare studerat materialet kvalitativt och de som har ett kvantitativt fokus. Medan litteraturvetenskapen står inför problemet att det finns för många böcker att läsa i förhållande till vad som skulle vara önskvärt att läsa under en livstid, står andra fält inte inför samma problematik. Till skillnad från 1800-talets samlade romankorpus är till exempel 8 000 statliga utredningar något som ingen ensam person kan läsa eller, troligen, vill läsa.

Ett ord tar över samtalsämnet

I Sverige, liksom i andra västerländska länder, växte ett ökat politiskt intresse för *information* fram decennierna efter andra världskriget. Den samhällsvetenskapliga forskningen brukar benämna utvecklingen som framväxten av vårt moderna informationssamhälle (ekonomiskt, teknologiskt, socialt, rumsligt så väl som kulturellt), med start under 1960-talet (Webster 2002: kap 2). Men det handlade också om en begreppslig förändring av ordet *information*, från att vara något som skapade mening åt en saks essens till att, med Shannons kommunikationsmodell från 1948, bli en sak i sig; något som var kvantitativt mätbart och möjligt att finna i alla möjliga sammanhang (Dahling 1962; Peters 1988; Gleick 2011: 10-12, 270f). Idén om att *information* var något som överfördes från sändare till en mottagare spreds snabbt till samhällsvetenskaperna och kom i sin tur att påverka hur staten bedrev informationspolitik. *Information* blev tidens melodi, både som mål och medel, lösning och problem för efterkrigstidens olika samhällsutmaningar. Således blev *information*, och *information* som styrmedel (genom överföring av *information*), en politisk fråga som hamnade allt högre upp på den politiska dagordningen under 1960-talet.

Vid en genomgång av SOU-titlarna från perioden märks det tydligt hur *information* och medier blir allt mer närvarande inom utredningsväsendet. Under 1960-talet sker också ett kommunikativt kraftprov för den svenska statsapparaten, som påverkade informationspolitikens praktiska riktning: koordineringen av högertrafikomläggningen. Den föregicks av en lång och omfattande, statligt organiserad, informationskampanj som blev normbildande för hur både privat och offentlig sektor skulle komma att bedriva kampanjer. Samma år som omläggningen ägde rum, 1967, tillsatte regeringen en informationsutredning för att undersöka hur den statliga informationen på bästa sätt skulle samordnas. När kommittén lade fram sitt betänkande var det begreppet *samhällsinformation* som blev centrum för diskussionen (SOU 1969: 48). Figur 2 visar frekvensen för 'samhällsinformation', framväx-



Figur 2. Frekvensförändring av ordet 'samhällsinformation' (lemgram) i Statens offentliga utredningar

ten av ett nyord. Kurvans toppar följer beskedligt de specifika utredningar som behandlade just samhällsinformationen.

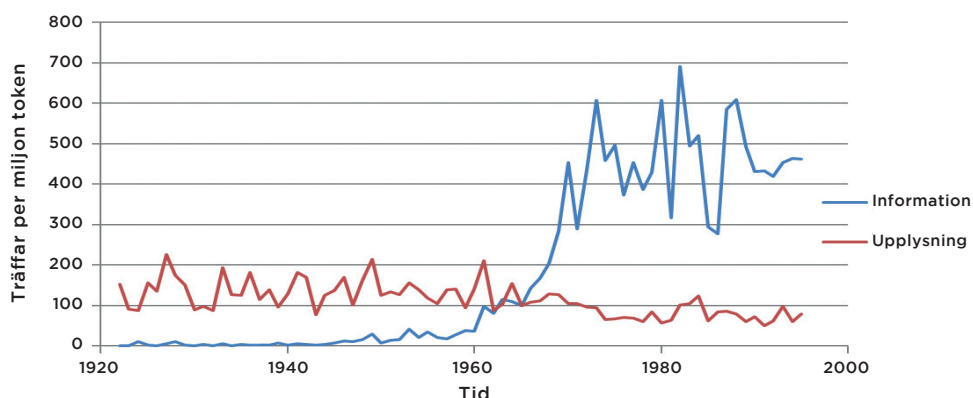
Det mindre specifika ordet 'information' sätter ett mer precist finger på framväxten av en samhällsutveckling. 'Information' är i princip helt frånvarande i kanslisvenskan före 1960-talet men användandet exploderar därefter. Från att ha varit ett ganska perifert substantiv på 1950-talet (plats #1378) stiger det kraftigt under 1960-talet (#306) och 1970-talet (#68) för att sedan stabilisera sig på 1980-talet (62). På 1970-talet förekommer 'information' oftare än substantiv som till exempel 'kvinna' (#72), 'samhälle' (#83) och 'elev' (#87).

'Upplysning' användes tidigare som synonym till 'information' och skiftet syns i SOU-

materialets titlar: SOU 1947: 76 *Kommunal upplysning* och SOU 1972: 52 *Rapport angående kommunal information* och motsvarigheten till 1960-talets informationsbetänkande hette på 1940-talet 1949:31 *Social upplysning*.

Men en jämförelse av 'upplysning' och 'information' visar att det inte bara handlade om ett ord som ersatte ett annat, utan snarare ett nytt politisk-byråkratiskt samtal som växer fram (Figur 3).

Kjell Nowak och Karl-Erik Wärneryd satte 1969 ord på tidsandan: "Ordet informera och kommunicera har blivit alltmer använda i vardagsspråket. En person 'informerar' i stället för 'berättar' eller 'talar om', han 'kommunicerar' i stället för att 'meddelar sig med.'" och lite längre fram tillägger de: "Ordet information används här i ungefär samma betydelse



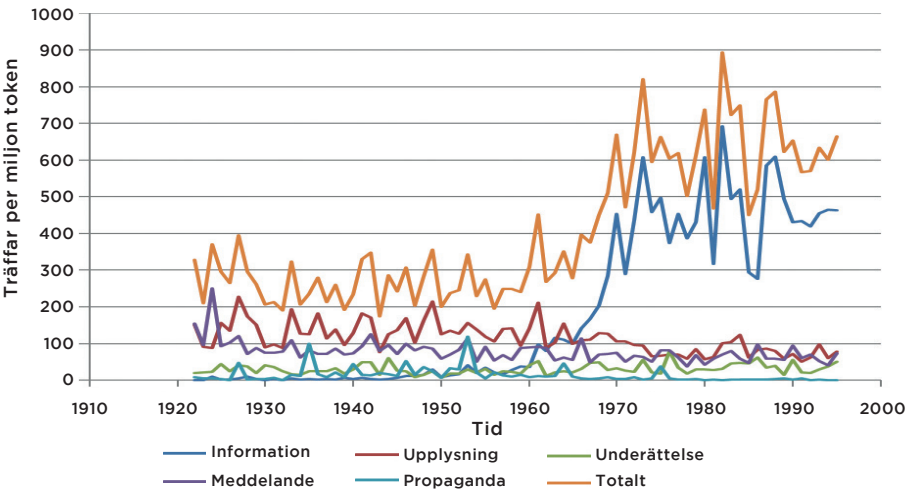
Figur 3. Frekvensförändring av orden 'information' och 'upplysning' (lemgram) i Statens offentliga utredningar

som innehåll i meddelanden.” (24). Men även om vi lägger till andra synonymer som ’underrättelse’, ’meddelande’ och ’propaganda’ kvarstår den radikala ökningen.³ Kurvan problematiserar Hanna Kjellgrens metod att lyfta fram fördefinierade ’guldåldrar’, som till exempel 1940-talet med Statens informationsstyrelse (1940-1945). Utifrån den induktiva grafen i Figur 4 är det snarare utvecklingen på 1960-talet och 1970-talet som utgör den verkliga förändringen i informationspolitiken under 1900-talet.

KWIC och den normativa informationen

I Språkbankens Korp finns funktionen ’Ordbild’ för att studera enskilda substantivs närliggande preposition, attribut och verb i en samlad korpus (i detta fall alla betänkanden mellan 1922-1995), vilket ger en inblick i idévärlden för ordet ’information’.

I Figur 5 ser vi att verben, både med ’information’ som subjekt och objekt, kretsar kring framförallt tre aspekter: *insamla* information, *sprida* information och *motta* information. Av



Figur 4. Frekvensförändring av orden ’information’, ’upplysning’, ’underrättelse’, ’meddelande’ och ’propaganda’ samt sammanslagen frekvens (lemgram) i Statens offentliga utredningar

Preposition		Attribut		information		Efterställt Attribut				information		verb				Verb		information	
1. av	6438	1. relevant	204	1. i prognosfrågor	104	1. lämna	515	1. ge	2229	1. för	2769	3. bra	375	3. om verksamhet	140	3. sprida	159	3. lämna	1087
2. om	1877	2. god	582	2. om arbetsmiljörisker	92	2. ge	637	2. få	1874	4. med	1558	4. fortlopande	200	4. till allmänhet	111	4. vara	1681	4. sprida	448
3. för	2769	3. bra	375	3. om verksamhet	140	5. om förhållande	150	5. nå	107	5. genom	541	5. värdefull	199	5. om förhållande	150	6. kunna	547	5. skaffa	253
4. med	1558	4. fortlopande	200	4. till allmänhet	111	6. om risk	89	6. erhålla	448	6. på	1395	6. statistisk	257	6. om risk	89	7. förmedla	61	6. förmedla	195
5. genom	541	5. värdefull	199	5. om förhållande	150	7. om kemiskaprodukter	52	7. förmedla	195	7. till	1573	7. skriftlig	222	7. om kemiskaprodukter	52	8. behöva	139	8. samla ²	173
6. på	1395	6. statistisk	257	6. om risk	89	8. lämna	129	8. behöva	139	8. _för	20	8. muntlig	180	8. lämna	129	9. inhämta	76	9. samla	173
7. till	1573	7. skriftlig	222	7. om kemiskaprodukter	52	9. finnas	153	9. inhämta	76	9. utifrån	34	9. saklig	158	9. finnas	153	10. hämta	56	10. skaffa sig	145
8. _för	20	8. muntlig	180	8. lämna	129	10. från tillsynsverksamhet	42	10. hämta	56	10. via	37	10. tillgänglig	178	10. från tillsynsverksamhet	42	11. behövas	71	11. inhämta	160
9. utifrån	34	9. saklig	158	9. finnas	153	11. om utveckling	77	11. behövas	71	11. utöver	44	11. aktuell	213	11. om utveckling	77	12. förbättra	64	12. innehålla	203
10. via	37	10. tillgänglig	178	10. från tillsynsverksamhet	42	12. finna	153	12. förbättra	64	12. förutom	35	12. ömsesidig	110	12. finna	153	13. samla	52	13. tillhandahålla	120
11. utöver	44	11. aktuell	213	11. om utveckling	77	13. om bl	81	13. samla	52	13. i form	31	13. tillräcklig	181	13. om bl	81	14. samla ²	52	14. hämta	96
12. förutom	35	12. ömsesidig	110	12. finna	153	14. om läromedel	39	14. samla ²	52	14. kring	26	14. intern	132	14. om läromedel	39	15. samlas	44	15. föra ut	62
13. i form	31	13. tillräcklig	181	13. om bl	81	15. kunna	294	15. samlas	44	15. emot	18	15. allsidig	91	15. kunna	294				

Figur 5. Ordbild av ’information’ (lemmat) i Statens offentliga utredningar

attributen går det föga oväntat att dra slutsatsen att den information som staten lyfte fram var den *goda* informationen, som gärna skulle vara relevant, saklig och cirkulera fortlöpande.⁴

Även om tabellen i viss mån är talande för hur staten diskuterade informationsfrågan är ordbildsverktyget för trubbigt för att ge en konkret uppfattning om hur detta tog sig uttryck i själva texterna. Men Språkbankens Korp har även funktionalitet för att praktisera pendelrörelsen mellan avstånd och närhet till texten och vi kan identifiera utredningar där ordet 'information' i större utsträckning användes. Nedan återfinns ett citat från SOU 1970:41 *Företag och samhälle*, en utredning som inte primärt avhandlade informationsfrågan och just därför utgör ett extra talande exempel för Figur 5.

Mot bakgrund av direktiven och vad som anförts i det föregående har det varit en naturlig uppgift för utredningen att överväga former för **förbättrad** information och vidgat samarbete i planeringsfrågor mellan olika samhällsorgan och de enskilda företagen. Utredningen presenterar i det följande ett förslag till informationssystem som är avsett att **ge snabb och smidig** information å ena sidan om företagens lägesbedömningar och framtidsplaner i första hand till kommunala och till offentliga regionala organ och å andra sidan om sådana samhällsorgans verksamhet och planer till företagen. Det föreslagna systemet syftar till att **ge** både samhällsorgan och företag **bättre** underlag för planering och beslut i näringspolitiska frågor. Därvid har stor vikt fästs vid önskemålet om **ökad** information om företagens planering. (s. 61, min fetmarkering)

Att en önskan om att informationen skulle öka var dock en sanning med modifikation. Redan på 1970-talet levde diskussionen om informationsöverflödet och svårigheten att nå ut i informationsbruset. Önskemålet om mer information handlade så klart om att den goda informationen skulle öka. Attributen i Figur 5 och citatet ovan ger en bild av information som en *deliberativ kraft*. Detta bekräftas av de

två informationsutredningarna från 1969 respektive 1984. Båda behandlade frågan om hur medborgare, framförallt de som fanns på den fattiga sidan av 'informationsklyftan', skulle nås av rätt information för att kunna förhålla sig till rättigheter och skyldigheter samt ges möjlighet att fatta självständiga beslut. Behovet av att informera om nya reformer för att de skulle ge önskad effekt förstärktes med välfärdsstatens expansion (Kjellgren 2002: 136-139). Kravet på att nå ut intensifierades av upplevelsen att samhället alltmer präglades av ett ökat antal informationskanaler och en hårdare konkurrens om medborgarnas uppmärksamhet (SOU 1969:48; SOU 1984:68). På andra sidan myntet fanns myndigheternas oro för medborgarnas oförmåga att skaffa rätt information, vilket blir tydlig i SOU 1974:29 *Att utvärdera arbetsmarknadspolitik*. Citatet nedan ger en såväl kvävande som komisk inblick i hur måna myndigheterna var om medborgarnas informationsstrategier:

Den arbetssökande kan ägna tid och andra resurser åt att skaffa mer allmän arbetsmarknads**information** (inklusive **information** om **informationskällor**), åt att skaffa extensiv **information** och åt att skaffa intensiv **information**. I varje tidsperiod måste den arbetssökande göra en bedömning av vilken eller vilka typer av **information** som kan beräknas ge högst avkastning och av vilken **informationskälla** eller kombination av **informationskällor** som i relation till kostnaderna kan förväntas bäst generera denna **information**. Troligen finns det också en optimal placering av aktiviteterna över tiden. T. ex. kan en arbetssökande genom att i början av sökandet ägna resurser åt att skaffa allmän **information** om arbetsmarknad och om **informationskällor** i nästa tidsperiod öka produktiviteten i insamlandet av extensiv **information**, vilken i sin tur utgör en förutsättning för insamlandet av intensiv **information**. (s. 121, min fetmarkering)

Citatet är även ett exempel på hur information både utgjorde medlet och målet för att lösa

samhällets och medborgarnas problem. För läsaren blir det också tydligt att samtidigt som information ansågs vara en lösning skapade information också problem (till exempel förvirring). Härutöver märks en tendens att reducera andra problem till informationsproblem.

Topic modeling och informationsdiskursens klubbighet

Frekvensstudier av enskilda ord kan vara en signal för förändringar i ett samhälle, men metoden har en tydlig begränsning. David Mimno sammanfattar styrkan med att i stället analysera kluster av samförekommande ord snarare än enskilda ord: "Individual words are informative, but are often not as meaningful as small groups of related words. In different contexts, the same word type could have very different connotations." (Mimno 2012: 6; Quinn m fl 2010: 218). Även konkordansstudier, som ordbilden i Figur 5, fångar bara upp en strikt avgränsad kontext. Metoden är än mer begränsande om man vill analysera vilka politiska områden som informationsdiskursen växte in i under andra halvan av 1900-talet. En signal är inte ett koncept, enskilda ord är inte diskurser; de har olika ontologier. Ett allsidigt ord som 'information' kan utgöra en signal för flera olika politiska diskurser eller i värsta fall ointressant *brus* (Block 2006). Men som kluster tillsammans med andra ord skapas olika informationsdiskurser, till exempel ger orden 'information', 'säkerhet' och 'övervakning' tillsammans en annan konnotation än orden 'information', 'konsument' och 'produkt'. En övergripande metod för att analysera diskurser i stora korpussamlingar är *topic modeling*. Namnet syftar på det språkliga mönster, i form av underliggande teman (topics), som metoden modellerar och som sammantaget ger en representation av den totala textmassan. Till skillnad från att leta efter specifika nyckelord söker vi med topic modeling övergripande teman, likt de i ordföljdexemplen ovan (Blei & Lafferty 2006: 1; Blei, Carin & Dunson 2012: 77).

En modell som ger meningsfulla och intuitiva topics är Latent Dirichlet Allocation (LDA), en generativ sannolikhetsmodell där

stora textsamlingar antas bestå av ett antal topics. Det LDA gör är att dels klustra ord till topics och dels klustra dokument till blandningar av topics. Formeln bygger på en bayesiansk slutledningsmodell som associerar varje dokument med en sannolikhetsfördelning av topics, där topics är sannolikhetsfördelning av ord. Modellen ger ett resultat där varje dokument representeras av en slumpmässig blandning av latent topics, där varje topic kännetecknas av en fördelning av ord, genererade enligt en sannolikhet för respektive topics. Detta innebär att ett ord kan ingå i flera topics med olika sannolikhetsgrad (Blei & Jordan 2003: 996; Griffiths och Steyvers 2004; Underwood 2012). (Till skillnad från KWIC tar LDA hänsyn till större textfiler, inte bara meningar där ett visst nyckelord förekommer.) Ett open source-baserat och användarvänligt LDA-verktyg är MACHINE LEARNING FOR LANGUAGE TOOLKIT (MALLET), som styrs genom *bash command line arguments*. Verktöget ligger till grund för den följande analysen.

För att avgöra om det fanns en informationsdiskurs med en lika dramatisk kurva som den i Figur 1 gjordes en MALLET-genomgång av samtliga SOU:er från 1960 till och med 1989 (2298 betänkanden). Jag prövade materialet på 100, 500 och 1000 topics. Med 100 topics framstod representativiteten av topics för låg när jag jämförde med 500 topics. 1000 topics hade varit ett gångbart alternativ, jag fick till exempel ut ett snarlikt informations-topic som i Figur 6 (15/19 toppord var samma), men den stora diversifieringen av topics bedömdes inte vara nödvändig för min studie. LDA-körningen programmerades därför till 500 topics och begränsades till att endast omfatta substantiven i materialet, reducerade till sin grundform. Begränsningen till substantiv förfinar topics diskursiva karaktär, då ordklassen är effektiv för att leta efter tematik (Jockers 2013: kap 8). Men det går utmärkt att inkludera samtliga ordklasser och samtidigt få tillbaka meningsskapande topics. Urval av ordklass och reduktion till grundform (stemming) utgick från versioner av materialet (i form av XML-filer) där orden annoterats (märkts upp) med, bland annat, ordklasser och grundformer, och

som fritt finns att ladda ned från Språkbankens hemsida. Detta gör att till exempel ett sammansatt ord som 'informationsverksamhetens' får en stemming för prefixet (information) och en för suffixet (verksamhet) och får därför den naturligt förväntade grundformen 'informationverksamhet' (se topic 285 i figur 6). Ett script utvecklades (i Python) som extraherar alla ord i dess grundform för valda ordklasser (i detta fall endast substantiv) från materialet och skriver orden i textfiler om max 1000 substantiv per fil (omkring 57 000 filer). Anledning till att betänkningarna delas upp i mindre textdelar är att LDA-modellen antar att en text, till exempel en SOU, är uppbyggd av olika topics. Resultatet blir därför mer förfinat om samförekomstanalysen baseras på mindre textblock. Valet av just 1 000-ordsfiler bygger bl.a. på Matthew Jockers LDA-studier (Jockers 2013: kap 8; Weingart 2011; Blades 2012). Stoppord rensas bort av MALLET genom en stoppordslista som kan designas efter eget behov.

En del av resultatet visas i Figur 6, ett utdrag från MALLET-filen 'tutorial_keys.txt' som listar samtliga 500 topics. Den blåmarkerade texten utgör ett relativt starkt topic, av mig identifierat som en allmän informationsdiskurs. Rådatat ska läsas på följande sätt. Det första numret (285) identifierar topic-numret, det andra värdet (0,04028) utgör dess sannolika vikt i den totala textsamlingen. Sist följer, i storleksordning, de 19 mest sannolikt frekventa orden som ingår i topic 285 (antalet bygger på en förinställning i MALLET).

De 500 'påsar' med ord skapar hos uttolkaren intuitiva associationer till olika dis-

kurser (i detta fall både politiska sakfrågor och utredningsgenrens språktrender). Eller som Scott Weingart har uttryckt det: "[...] generating what really *feels* like topics as we know them [...]" (2011). Men inte alltid, i enstaka fall dyker så kallade korrupta topics upp som utgörs av mer eller mindre obegripliga ordföljder. Till exempel topic 375:

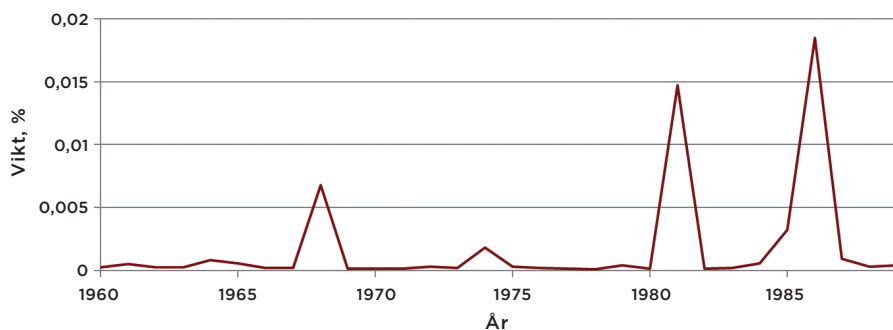
375 0,01669 nr juni meddelande ang stad
kvarn anm medd namn not laxfiske ålfiske
län bil beskaffenhet avd tidskrift serie bruk

Metoden utgör ett induktivt sätt att bedriva kvantitativ forskning, som inte utgår från att forskaren behöver förbestämma hur kodningen ska gå till eller ens inledningsvis ha bestämt frågeställning och hypotes. Men för att forskaren ska kunna tolka resultatet från LDA-beräkningen krävs däremot både kunskap och tillgång till de enskilda texterna (Newman & Block 2006: 758; Weingart 2011). Ta till exempel topic 286, som vid första ögonkast framstår som ett korrupt topic. Men med hjälp av ett Python-script kan vi med filen 'tutorial_composition.txt' från MALLET-körningen skapa tidsdiagram, vilket ger följande resultat i Figur 7. (Det ska tilläggas att LDA-algoritmen inte är det starkaste verktyget för att studera topics över tid. David M. Blei har till exempel även utvecklat modellen Dynamic Topic Models som bättre fångar en sådan förändring (Blei & Lafferty 2006; Blei, Carin & Dunson 2012). Men i denna artikel väljer jag att konsekvent använda LDA/MALLET.)

När vi spårar grafens toppar tillbaka till SOU-materialet återfinns vid de olika tidpunkterna betänkningarna i form av psalmboken, tillägg till *Gamla testamentet* med mera.

```
284 0,00681 försäkring sjukförsäkring arbetsgivaravgift sjukförsäkring inkomst
förmån kassa arbetsgivare yrkesskadaförsäkring folkpensionering tilläggs-pensionering avgift
sjukpenning utgift riksförsäkringsverket försäkringskassa sjukdom tilläggs-pension
285 0,04028 information allmänhet kontakt fråga kunskap material organisation form upplysning
informationverksamhet syfte behov broschyr massmedium intresse innehåll myndighet uppgift typ
286 0,00651 not ord folk son människa bror kung namn tid far herre jude stad ande kor mack
vishet liv profet
287 0,01378 förbud överträdelse dispens strid ingripande slag reglering sanktion risk verkan
bud åläggande efterlevnad medgivande förbudsregel straff gräns vite väg
288 0,06494 riksdag år fråga förslag utskott motion proposition departementschef utredning
betänkande maj prop anledning skrivelse ändring utlåtande nr åtgärd motionär
289 0,02043 fastighet ägare fastighetsägare byggnad jordbrukfastighet taxeringsvärde fall
egendom delägare försäljning villa ändamål mark kostnad äganderätt omfattning bostad innehav
upplåtelse
290 0,00241 parkering plats bilplats parkeringsplats trafikföreskrift parkeringanläggning
uppställning bil parkeringutrymme trafik utrymme trafiknämnd fordon skyldighet gata tomtmark behov
byggnadsnämnd gatamark
```

Figur 6. MALLET-filen tutorial_keys.txt

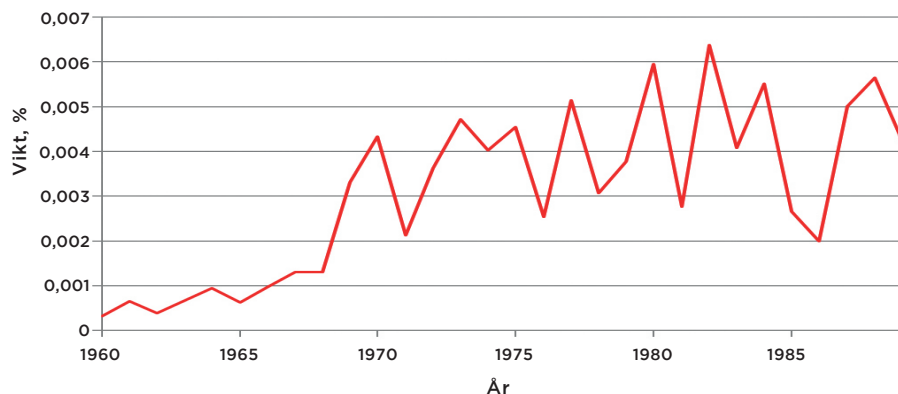


Figur 7. Frekvensförändring av topic 286 (baserad på vikt)

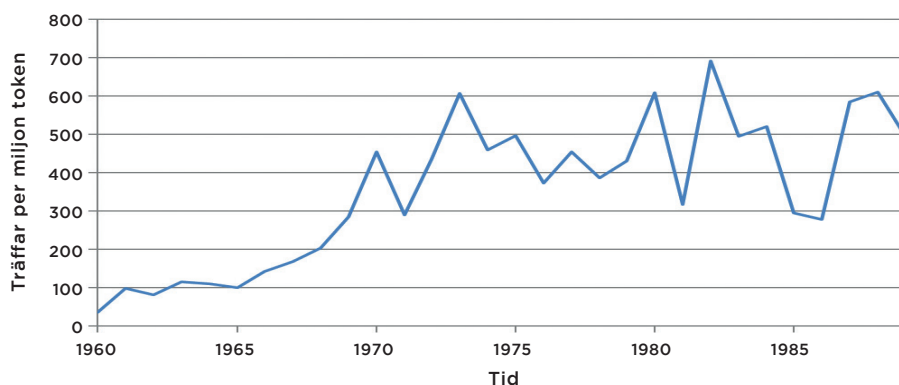
När Svenska kyrkan fortfarande var en statlig myndighet, utreddes de religiösa skrifterna och ceremonierna inom samma ramar som när skattesystemet skulle ses över. Liksom ordet 'samhällsinformation' i Figur 2 följer signalen för topic 286 beskedligt de år då svenska kyrkans verksamhet utreddes, ett specifikt utred-

ningsobjekt, och grafen lämnar inte mycket utrymme åt fantasin.

Tidsdiagrammet över topic 285 visar något helt annat: en generell informationsdiskurs, en trend som är ökänslig för enskilda utredningar (Figur 8). En jämförelse med Figur 9 visar snarare att det är ordet 'information' som är



Figur 8. Frekvensförändring av topic 285 med trendlinje (baserad på vikt)



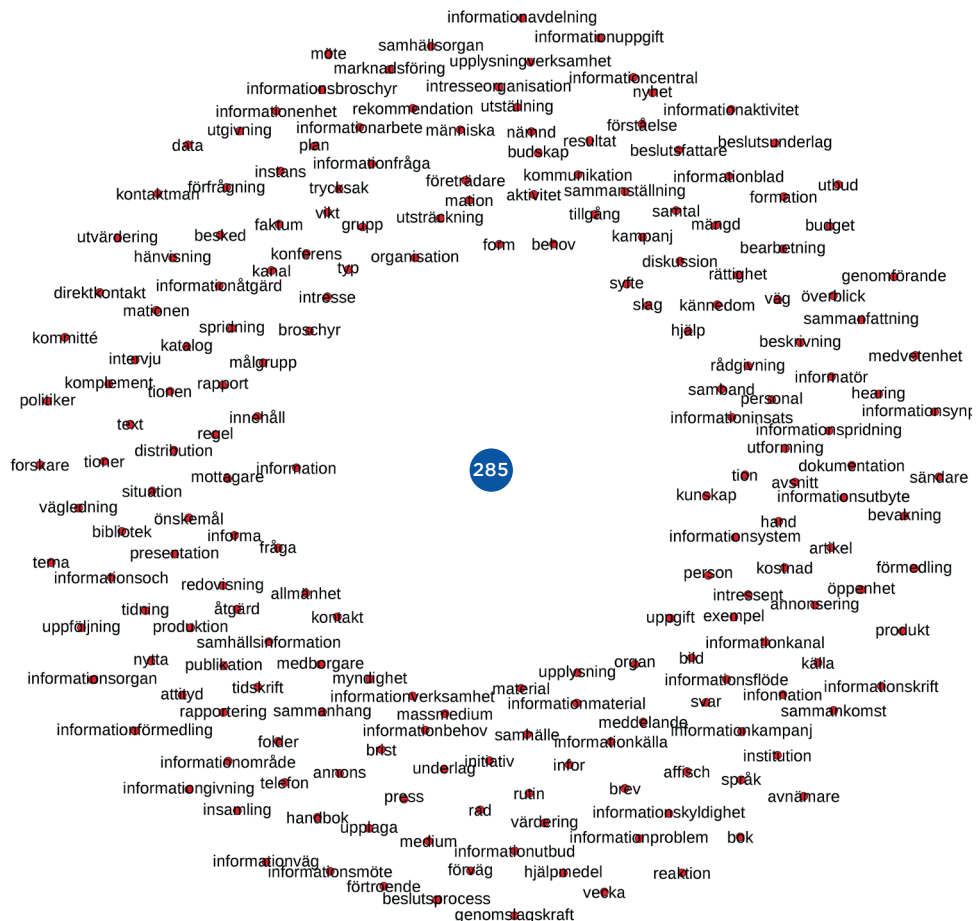
Figur 9. Frekvensförändring av ordet 'information' med trendlinje (lemgram)

styrande för diskursens utveckling. Vi kan dra den slutsatsen genom att studera frekvensen av andra substantiv i topic 285. De sannolika frekvenserna ges av en resultatfil från MALLET som sammanställts via ett script i Python. Ordet 'information' är 11 gånger mer sannolikt förekommande (48 871 träffar) än det näst mest förekommande substantivet, 'allmänhet' (4 303), och 13 gånger mer förekommande än tredjeplatsen 'kontakt' (3 667).

Genom att expandera innehållet i ett topic kan forskaren definiera dess centrum och periferi (Jockers 2013: kap 8; Jarlbrink, Snickars & Colliander 2016: 38). Utav sannolikhetsfrekvenserna som ges av en resultatfil från LDA-genomgången av SOU:erna från 1960-1989, sammanställt via samma Python-script, generades de 200 mest sannolikt förekommande substantiven i topic 285. Resultatet visualiserar-

des med hjälp av open source-verktyget Gephi i Figur 10, där orden närmast mitten utgör diskursens kärna, de mest sannolika substantiven att förekomma i samband med topic 285.⁵ Även om resultatet hade framgått av en enkel tabell ger Figur 10 en bättre pedagogisk känsla för ett topics centrum och periferi.

Av figur 10 kan vi avläsa att topic 285 utgör en generell informationsdiskurs även i expanderad form, inifrån och ut: det handlade om att sprida information genom en uppsättning kanaler från och till olika sändare och mottagare. Diskursen rör sig från den väldigt generella kärnan (#1 'information', #2 'allmänhet', #3 'kontakt') för att i ytterkanterna bli något mer specifik (#171 'informationsmöte', #175 'kontaktman', #194 'politiker'). I topic 285 återfinns även en uppsättning ord som utgjorde informationspolitikens mediekkanaler



Figur 10. Topic 285 expanderat och visualiserat i Gephi

(#13 'broschyr', #57 'tidskrift', #97 'annons', #104 'brev').

Däremot återfinns endast ett av de efterställda attributen från Figur 5: 'allmänhet' (#2). Efter en liknande topic modeling-genomgång av SOU-materialet 1960–1989 som även inkluderade verb och adjektiv, fångade motsvarigheten till topic 285 in flertalet av verben och adjektivattributen från ordbilden. Jämförelsen visar på olikheterna mellan KWIC och LDA (topic 285). Till skillnad från KWIC-metoden kan den senare ringa in en större informationstematik om sådant som rör mediekanalet och avsändare/mottagare. Språkbankens KWIC-verktyg (givet sin hårt avgränsade kontext) visar i stället förmågan att direkt peka ut vissa politiska frågor som var kopplade till nyckelordet 'information'. Vi kan därför dra slutsatsen att den generella karaktären på topic 285 helt enkelt är för stark för att släppa in influenser från andra topics än det som strikt handlade om information och medier. Men, utifrån LDA-modellens principer, hindrar det inte att olika topics kan samexistera sida vid sida, vilket vi kan utnyttja för att analysera hur en diskurs kan infiltrera en annan.

Efter att ha slagit fast att topic 285 utgör en allmän informationsdiskurs, vars signal tydligt följer frekvensförekomsten av ordet 'information', är det ett rimligt antagande att diskursen klibbade sig fast vid andra topics där substantivet är frekvent förekommande. För att spåra en sådan framväxt delades SOU-materialet upp per decennium (1950-talet, 1960-talet, 1970-talet och 1980-talet). Därefter utfördes tre topic modeling per SOU-decennium för att avgöra hur och till vilken grad som topics förändrades vid varje genomgång. Det visade sig att skiftningar förekom i resultaten, vilket är ett naturligt inslag i LDA-modellen, men variationen bedömdes (av mig) inte heller vara för omfattande. För varje resultat (tre per årtionde) valdes de topics ut som innehöll ordet 'information' bland de 19 mest förekommande substantiven, antingen som enskilt ord, prefix eller suffix. Sedan utfördes ett manuellt tolkningsarbete där varje topic tilldelades en tematisk rubrik. Sist gjordes

en gallring baserad på följande kriterier: enbart topics innehållande svenska ord behölls, uppenbara dubletter sorterades bort (den dublett med störst vikt behölls), korrupta topics ströks liksom de med för generellt eller ottydligt innehåll för att kunna sorteras in i ett politiskt område (dessa topics avslöjar dock hur informationsdiskursen växer sig in i utredningssvenskans resonerande prosa). I den slutliga uppställningen, som framgår av Tabell 1, framträder en liknande framväxt som i Figur 1, men med skillnaden att antalet topics som innehåller ordet 'information' fortsätter att öka under 1980-talet. (Topic-rubriken 'information/allmänt' bedömdes vara ett motsvarande topic 285.)

Det exakta antalet topics skiljde sig något åt vid varje MALLT-genomgång, en naturlig konsekvens av LDA-algoritmens sannolikhetsprincip, men trenden var den samma; nämligen ökningen av antalet olika politiska diskurser som inkluderade ordet 'information'. Hanna Kjellgren menar att den offentliga informationen i hög grad blir ett eget problemområde under 1960-talet med informationsutredningen och Nämnden för samhällsinformation som resultat (Kjellgren 2002: 289). Men Tabell 1 visar också att den offentliga informationen allt mer blev en integrerad del i olika befintliga och nya politiska verksamhetsområden. Liksom ordet 'terrorism' spred sig in i allt fler politiska verksamheter i USA efter elfte september, ser vi en liknande diversifiering av topics som informationsdiskursen klistrade sig fast vid (Quinn m fl 2010: 218).

Det som växer fram är inte nödvändigtvis nya politiska områden, men flera kännetecknas av kontinuitet som över tid blir allt mer infiltrerade av informationsdiskursen. Ett exempel är den statliga statistiken. Det existerade ett starkt statistik-topic även på 1950-talet, men då hamnade 'information' på omkring plats #40,⁶ för att på 1960-talet kvala in på topp 20 och på 1970-talet och 1980-talet lägga sig på omkring topp fem bland de mest sannolikt frekventa substantiven. Vi kan även illustrera förändringen genom att jämföra två någorlunda likvärdiga statistikbetänkanden från olika tidsperioder. I SOU

Tabell 1. Manuellt kodade topic-rubriker där 'information' ingår bland de 19 mest sannolikt förekommande substantiven

1950-talet	1960-talet	1970-talet	1980-talet
Information/allmänt Högertrafik Krig	Arbetsförmedling Bibliotek Databehandling Fastighetsregister Försvar/säkerhet Information/allmänt Information/myndighet Konsumentupplysning Läkemedel Målstyrning Statistik Transport Utland/bistånd Yrkesvägledning	Arbete/demokrati Arkiv/medier Bibliotek Folkrörelser/utbildning Forskning Information/allmänt Konsumentupplysning Miljövård Partistöd Personregister Radio/TV Reklam Snatteri Statistik Utland/kultur	Aktiehandel Arbete/inflytande Bibliotek Datasystem Forskning/etik Information/allmänt Kemikalier Konsumentupplysning Kultur/musik Lantmäteri/karta Läkemedel Livsmedelsforskning Miljö/datasystem Nya medier Personregister Produktion/tillväxt Radio/TV Sekretess/offentlighetsprincip Statistik Utland/bistånd Utland/kultur

1959: 33 *Organisatoriska riktlinjer för svensk statistik* förekommer 'information' (40 träffar med lemmgram, prefix och suffix inräknade) och 'upplysning' (14) sammanlagt 54 gånger, eller 0,0008 procent fördelat på antalet tokens i utredningen. I SOU 1983: 74 *Framtida statlig statistik* utgör siffran för 'information' (321) och 'upplysning' (16) sammanlagt 337 träffar och i relativa tal 0,0035 procent. En tydlig intensifiering av informationstemat 24 år senare.

En annan metod för att studera diskursiva förändringar och kontinuiteter över tid är att jämföra likvärdiga topics över olika tidsperioder. Det finns ett tydligt 'information/allmänt'-topic från 1960-talet, 1970-talet respektive 1980-talet (även om de skiljer sig lite åt), som utgör motsvarande topic 285 för hela perioden. Medan orden i topic 285 formar en statisk informationsdiskurs över 30 år (även om

dess styrka förändras över tid) kan vi med en periodisk jämförelse se hur diskursens innehåll förändrades. I Figur 11 visualiseras hur de 200 mest sannolikt frekventa substantiven i respektive decenniums 'information/allmänt'-topic förhåller sig till varandra.⁷

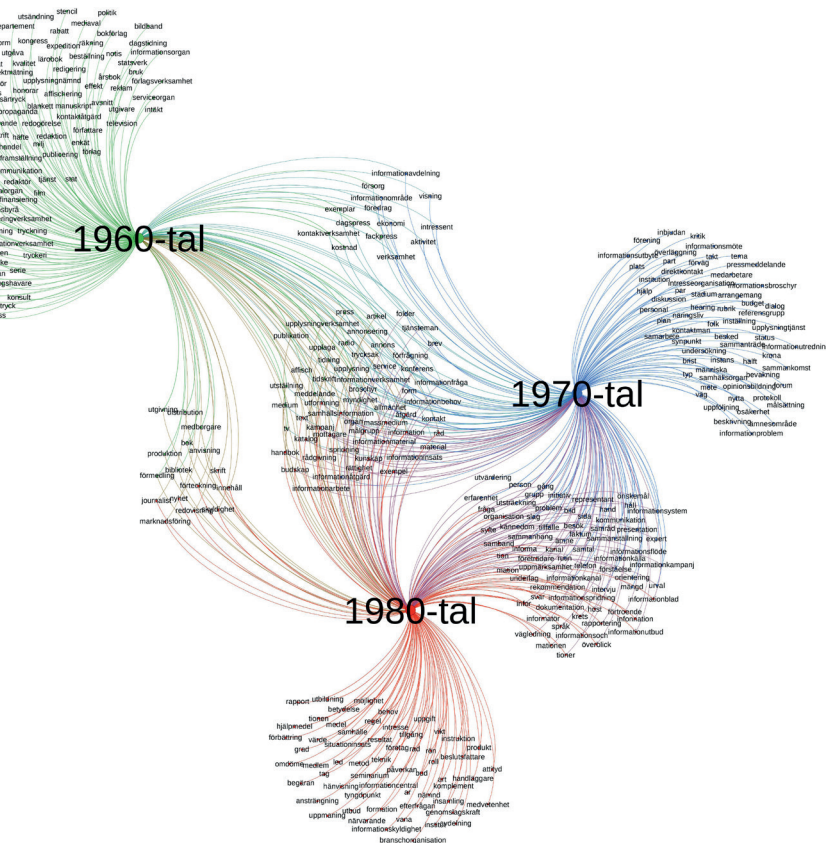
Återigen är det trenderna som först slår betraktaren. Varje topic karaktäriseras av en uppsättning unika ord för respektive årtionde (klustren som sträcker sig utåt från vardera topic), en annan samling ord förenar samtliga tre tidsperioder (klustret i mitten) samt tre ordmoln som förenar två topics. Av de 200 mest sannolikt frekventa substantiven från topic 285 är det endast 25 som inte finns representerade i något decenniums 'information/allmänt'-topic. Substantiven i centrum formar en avskalad version av topic 285, en diskursiv kontinuitet över tid, där också en

betoning på textbaserade medier synliggörs ('broschyr', 'tidskrift', 'katalog', 'brev', 'affisch'). Denna vikt stämmer överens med det fokus som kännetecknade de båda informationsutredningarnas betänkanden från 1960-talet respektive 1980-talet, framförallt den senare (SOU 1969: 48; SOU 1984: 68).

Utifrån Figur 11 synliggörs en diskursiv gemenskap mellan 1970-talet och 1980-talet, vilket skulle kunna kopplas till den politik som SOU 1969: 48 ledde fram till; och när informationspolitiken stramades åt gjorde språket detsamma. Inom denna tidsperiod återfinns bland annat modeord som 'kommunikation' och värdeladdade substantiv som 'problem', vilket genast skapar uppslag för intressanta studier och nya frågor. Ordet 'problem' har till exempel en mer specifik form i 1970-talets 'information/allmänt'-topic, nämligen 'informationsproblem'. Nämnden för samhällsinfor-

mation (NSI) gav till exempel ut diskussions-skriften med det talande namnet *Samhälls-informationens principiella problem* (1977). Kanske kan detta ses som en signal för en del av det 'massmedieproblem' som karaktäriserade mediastudiets formering under 1970-talet (Hyvönen, Snickars & Vesterlund 2015). Kontinuiteten över decennierna kan även i detta fall spåras till informationsbetänkandet från 1969 (med efterföljande proposition) som lade grunden till statens informationspolitik under 1970-talet och 1980-talet (Kjellgren 2002: 332–334).

Övergången från 1960-talet till 1970-talet skedde med en något större diskursiv brytning. Även om Figur 11 främst tecknar en generell trend går det också att använda resultatet som en utgångspunkt för att fördjupa sig i specifika frågeställningar. 1960-talets informationsdiskurs hade uppenbarligen ett an-



Figur 11. Informationspolitikens diskursiva förändring mellan 1960-talet, 1970-talet och 1980-talet

nat fokus än under 1970-talet och 1980-talet, och i den tidigare perioden kan vi identifiera tidsspecifika ord som sedan helt försvinner (och som inte ens fick plats i topic 285). Varför utgör till exempel 'konsult' ett av de 100 mest sannolikt frekventa substantiven under 1960-talets information/allmänt-topic (mellan #37 och #72 i de tre olika MALLET-körningarna), för att sedan helt försvinna under de följande två decennierna? Det motsatta hade varit mindre överraskande, framförallt då 1980-talet associeras med tiden då den offentliga styrningen börjar influeras av näringslivet och ordet 'konsult' bär på en marknadsorienterad konnotation. För att söka svar på en sådan fråga krävs att vi återigen vänder tillbaka till de enskilda informationsutredningarna. Av den senare framgår mycket riktigt att konsultfrågan har en betydligt mindre närvaro i den senare utredningen. I SOU 1969: 48 *Vidgad samhällsinformation* finns en uttalad vilja att föra den statliga informationspolitiken närmare det privata genom att föreslå en statlig reklambyrå som centralorgan för samhällsinformationen (79-87). I SOU 1984: 69 *Samordnad samhällsinformation* föreslogs i stället ett råd i form av en stiftelse (87-107). Även om dessa förslag aldrig resulterade i konkret politik väcks nya frågor: är denna tendens applicerbar i andra politiska områden? Finns det anledning att vidare problematisera våra historiska föreställningar om 1960-talet respektive 1980-talet?

Resonemanget ovan implicerar, återigen, att den här typen av analys kräver att forskaren både har tillgång till – och kunskap om det ursprungliga empirimaterialet. Det är också ett exempel på hur fruktbart och friktionsfritt det kan vara att röra sig mellan det kvantitativa och det kvalitativa. Konsult-exemplet illustrerar hur avståndsläsning kan användas som en utgångspunkt för att, i dialog med de individuella texterna, problematisera vedertagna föreställningar om historien.

Avslutande diskussion

Avståndsläsning innebär inte bara en metodologisk verktygslåda, det utgör även ett teo-

retiskt perspektiv, 'a condition of knowledge' som Franco Moretti uttryckte det (2000: 57). Våra frågor om världen behöver inte nödvändigtvis förändras men avståndsläsning kan möjliggöra *andra* perspektiv och resultat. Således har undersökningen kunnat visa hur en informationsdiskurs expanderar mellan 1960-talet och 1980-talet, medan 1940-talet, som annars också brukar lyftas fram som en expansion för den statliga informationspolitiken, inte får någon (språklig) uppmärksamhet här. Både Språkbankens KWIC-verktyg och LDA topic modeling-verktyget MALLET kunde på olika sätt lyfta fram vad den statliga informationsdiskussionen handlade om under ett långt tidsspänn. MALLET visade sig också vara ett användbart verktyg för att spåra hur informationsdiskursen över tid växte *in* i andra politiska diskurser och hur dessa politiska områden ökade från 1960-talet och in på 1980-talet. Och med hjälp av verktyget Gephi har diskursiva förändringar och kontinuiteter visualiserats som ett alternativt sätt att avståndsläsa en oöverblickbar textsamling.

Vissa av resultaten bekräftar den tidigare kvalitativt baserade forskningen, andra nyanserar och väcker nya forskningsfrågor. Det är just förmågan att kunna säga något kvalitativt annorlunda om ett visst studieobjekt som främst motiverar valet av kvantitativ eller kvalitativ metod. Poängen med artikeln har därmed inte varit att avfärda andra metoder, snarare att peka på nya metodologiska möjligheter, som ibland kan komplettera och ibland visa sig vara bättre lämpade än andra.

Med avståndsläsning är det trenderna och det som avviker från dessa som är det intressanta. Men med hjälp av avståndsläsning kan vi upptäcka saker som annars hade gått forskaren förbi. Konsult-exemplet i slutet av analysen visar hur vi genom att ta ett par kliv bort från de enskilda texterna också kan problematisera föreställningar om historien, för att sedan åter vända tillbaka och söka svar i det partikulära. Faktum är att undersökningen har visat att det inte bara går utan även krävs att forskaren också dyker ner i det specifika för att klä det generella med mening och substans, i en ständig rotation av kikartuben.

Att arbeta med digitala metoder och digitalt material innebär också risker som måste hanteras under processens gång. Ett varnande finger är till exempel på sin plats för risken att både som forskare och läsare tolka kvantitativa grafer och tendenser för hårt och att inte frestas att säga mer än vad resultatet och förutsättningarna tillåter. En utmaning under arbetets gång har varit att hela tiden ställa sig kritisk till det som mäts för att inte riskera att kvantifiera det uppenbara. I slutändan innebär det att forskaren ofta behöver lyfta på flera stegar innan något av verkligt intresse upptäcks, och det färdiga resultatet ser alltid enklare ut än processen bakom. En annan aspekt är att skärskåda de digitala verktygens egna premis-

ser och karakteristika (att, så att säga, öppna den 'svarta låda' som sådana verktyg i många avseenden utgör), och att kvalitetsgranska de digitaliserade dokument och resultat som programmeringsscript och mjukvara spottar ut. Dessa avslutande varningsord syftar inte till att skrämja bort forskare från digitala verktyg och material, tvärtom. I en tid som allt mer präglas av det digitala, i stort som i smått, finns det ett ökat behov av fler hum/samforskare med kompetens att studera vad den digitala övergången innebär för ett samhälle och inte minst hur vi kan använda oss av digitala metoder i vår forskning (t ex Fölster & Felländer: 2016; Snickars 2015).

Noter

1. Figur 1-4 bygger på resultat från Språkbankens konkordansverktyg Korp, <https://spraakbanken.gu.se/korp> 2016-05-27. Lemgram är ett ords eller ett flerordsuttrycks samtliga böjningsformer och token är sammansatta tecken separerade av mellanslag. Normaliseringen är gjord på antalet träffar (x) per miljon token för varje specifik tidsenhets totala mängd token (y), dvs $(x/y) \times 1\,000\,000$. Vid tidpunkten för analysen hade Språkbanken endast xml-transkriberat alla utredningar fram till och med 1995, vilket utgör grunden för hela analysen.
2. Två liknande projekt som pågår är en språklig studie av den amerikanska högsta domstolens miljömål sedan 1970-talet och förändringen av begreppet 'security' i tidskriften *Foreign Affairs* 1922-1939 (<https://litlab.stanford.edu/current-projects/>).
3. Ordet 'kommunikation' valdes bort då det ofta syftar på fysisk infrastruktur.
4. När analysen gjordes var Språkbankens xml-kodningen av SOU-materialet under omarbetning vilket gjorde att ordföljden i flera SOU:er kastades om. Resultatet i Figur 5 får trots allt bedömas som trovärdigt. Övriga analyser i artikeln påverkas inte av detta tillkortakommande.
5. Figuren bygger på följande inställningar i Gephi: 1) Layout: Force Atlas, Repulsion strength 100 000, Adjust by Size aktiverad, 2) nodernas storlek är inte proportionella.
6. I en MALLET-genomgång fanns inte ens 'information' med bland de 200 mest sannolikt frekventa substantiven.
7. Visualiseringen är skapad i Gephi, med hjälp av ett python-script och MALLET-filen 'word_topic_counts.txt'. Figuren bygger på följande inställningar i Gephi: 1) Forced Atlas med Repulsion strength 200 000, Gravity 1 samt aktiverad Adjust by sizes, 2) noderna är rankade efter Closness-centrality.

Referenser

- Algee-hewitt, Mark; Allison, Sarah; Heuser, Ryan; Gemma, Marissa; Moretti, Franco & Walser, Hannah (2016). Canon/Archive. Large-Scale Dynamics in the Literary Field Mark, in *Stanford Literary Lab Pamphlets*, online. Hämtad den 28 juni 2016. <https://litlab.stanford.edu/pamphlets/>
- Berger, Arthur Asa (2014). *Media and Communication Research Methods. An Introduction to Qualitative and Quantitative Approaches*. Los Angeles: SAGE Publications.
- Biber, Douglas, Conrad, Susan & Reppe, Randi (2006). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge Univ. Press.
- Blades, Robert (2012). Mining the Open Web with 'Looted Heritage', online. Hämtad den 28 juni 2016. <https://electricarchaeology.ca/2012/06/08/mining-the-open-web-withlooted-heritage-draft/>
- Blei, David M., Carin, Lawrence & Dunson, David (2012). Probabilistic Topic Models, in *IEEE Signal Processing Magazine* 27(6).
- Blei, David M. & Lafferty, John D. (2006). Dynamic Topic Models, in *Proceedings of the 23rd International Conference on Machine Learning (ICML 2006)*.
- Blei, David M., Ng, Andrew Y. & Jordan, Michael I. (2003). Latent Dirichlet Allocation, in *The Journal of Machine Learning Research*, 3(1).
- Block, Sharon (2006). Doing More with Digitization. An Introduction to Topic Modeling of Early American Sources, online. Hämtad den 28 juni 2016. <http://www.common-lacearchives.org/vol-06/no-02/tales/>
- Borin, Lars, Forsberg, Markus & Roxendal, Johan (2012). Korp. The Corpus Infrastructure of Språkbanken, in *Proceedings of LREC 2012. Istanbul: ELRA*.
- Bäck, Henry & Larsson, Torbjörn (2006). *Den svenska politiken. Struktur, processer och resultat*. Malmö: Liber.
- Dahling, Randall L. (1962). Shannon's Information Theory. The Spread of an Idea, in *Studies of Innova-*

- tion and of Communication to the Public. Stanford: Institute for Communication Research.
- Fölster, Stefan & Felländer, Anna (2016). Det digitala paradigmskiftet kräver ny ekonomisk politik, i *Dagens Nyheter* den 8 maj 2016.
- Gleick, James (2011). *The Information. A History, a Theory, a Flood*. London: Fourtrth Estate.
- Griffiths, Thomas L. & Steyvers, Mark (2004). Finding Scientific Topics, in *Proceedings of the National Academy of Sciences of the United States of America* 101 Suppl. 5228–35.
- Hayles, N. Katherine (2012). *How We Think. Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.
- Heuser, Ryan & Long Le-Khac (2012). A Quantitative Literary History of 2,958 Nineteenth- Century British Novels. The Semantic Cohort Method, in *Stanford Literary Lab Pamphlets*, online. Hämtad den 28 juni 2016. <https://litlab.stanford.edu/pamphlets/>
- Hockey, Susan (2004). The History of Humanities Computing, in Susan Schreibman, Ray Siemens & John Unsworth (ed.) *A Companion to Digital Humanities*. Malden, MA: Blackwell Pub.
- Hyvönen, Mats; Snickars, Pelle & Vesterlund, Per (ed.). (2015). *Massmedieproblem. Om mediestudiets formering*. Stockholm: Mediehistoriskt arkiv.
- Jarlbrink, Johan; Snickars, Pelle & Colliander, Christian (2016). Maskinläsning. Om massdigitalisering, digitala metoder och svensk dagspress. *Nordicom Information* (38)3.
- Jockers, Matthew L. (2013). *Macroanalysis. Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Kjellgren, Hanna (2002). *Staten som informatör eller propagandist? Om statssyners betydelse i svensk informationspolitik*. Göteborg: Komapniet Aidla Trading AB.
- Kullenberg, Christopher (2015). Sorlet från Statens offentliga utredningar, i *Intensifier. Pluralism = Monism*, online. Hämtad den 28 juni 2016. <http://www.christopherkullenberg.se/> (May 3, 2016).
- Larsson, Larsåke (2005). *Uppllysning och propaganda. Utvecklingen av svensk PR och information*. Lund: Studentlitteratur.
- Lindquist, Hans (2009). *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.
- Michel, Jean-Baptiste; Shen, Yuan Kui; Aiden, Aviva Presser, et al. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books, in *Science* 331(6014).
- Mimno, David (2012). Computational Historiography. Data Mining a Century of Classics Journals, in *Journal on Computing and Cultural Heritage* 5(1).
- Moretti, Franco (2000). Conjectures on World Literature, in *New Left Review* (1).
- Moretti, Franco & Pestre, Dominique (2015). Bankspeak. The Language of World Bank Reports, 1946–2012, in *Stanford Literary Lab Pamphlets*, online. Hämtad den 28 juni 2016. <https://litlab.stanford.edu/pamphlets/>
- Newman, David J. & Block, Sharon (2006). Probabilistic Topic Decomposition of an Eighteenth-Century American Newspaper, in *Journal of the American Society for Information Science and Technology*, 57(6).
- Nicholson, Bob (2012). Counting Culture; Or, How to Read Victorian Newspapers from a Distance, in *Journal of Victorian Culture*, 17(2).
- NIS (1977). *Samhällsinformationens principiella problem. En diskussionspromemoria*. Stockholm: Nämnden för samhällsinformation (NSI).
- Nowak, Kjell & Wärneryd, Karl-Erik (1969). *Kommunikation och påverkan*. Stockholm: Prisma.
- Peters, John Durham (1988). Information. Towards a Critical History, in *Journal of Communication Inquiry*, 12(2).
- Quinn, Kevin Mc; Monroe, Burt L.; Colaresi, Michael, et al. (2010). How to Analyze Political Attention with Minimal Assumptions and Costs, in *American Journal of Political Science* 54(1).
- Ross, Shawna (2013). 'In Praise of Overstating the Case. A Review of Franco Moretti, Distant Reading', in *Digital Humanities Quarterly*, 008(1).
- Samhällsinformation. Ett temanummer från Nordicom-Sverige 2-3(1984)* Göteborg: Nordicom-Sverige.
- Snickars, Pelle (2015). Så trångsynt att man blir mörkrädd, i *Mittmedia* den 11 maj 2015.
- SOU 1969:48 *Vidgad samhällsinformation*.
- SOU 1970:41 *Företag och samhälle. Betänkande. D. 2, hearings med företrädare för samhällsorgan, företag, löntagarorganisationer, politiska partier m.fl.*
- SOU 1974:29 *Att utvärdera arbetsmarknadspolitik*.
- SOU 1984:68 *Samordnad Samhällsinformation. Stat – Landsting – Kommun*.
- Underwood, Ted (2012). Topic Modeling made just Simple enough, in *The Stone and the Shell*, online. Hämtad den 28 juni 2016. <http://tedunderwood.com/2012/04/07/topicmodeling-made-just-simple-enough/>
- Webster, Frank (2002). *Theories of the Information Society* Second Edition. New York: Routledge.
- Weingart, Scott (2011). Topic Modeling and Network Analysis, online. Hämtad den 28 juni 2016. <http://www.scottbot.net/HIAL/?p=221> (May 3, 2016).

Fredrik Norén är doktorand i medie- och kommunikationsvetenskap, med inriktning mot digital humaniora, vid Umeå universitet och affilierad till Humlab. Han forskar om samhällsinformationen på 1960-talet och 1970-talet och använder olika metoder för att studera vilka idéer och förutsättningar som formade varför och hur staten kommunicerade med medborgarna.

Johan Jarlbrink, Pelle Snickars & Cristian Colliander

Maskinläsning

Om massdigitalisering, digitala metoder och svensk dagspress

Abstract

This article highlights the media historical possibilities to analyse linguistic patterns in massive amounts of texts using digital methods. Our starting point is the fact that The National Library of Sweden has made over 12 million newspaper pages available in digital format. An important question is how to research them. The article presents a media history of the Swedish newspaper digitisation, as well as new ways of conducting historical newspaper research using digital methods. A case study is presented where the conceptualisation of a new media technology (the internet) in newspapers from the 1990s is tracked with a digital tool searching for word co-occurrences. The possibilities of digital methods are often incredible, but we should not underestimate the problematic aspects of using digital tools to explore digitised newspapers. The poor quality of the OCR (Optical Character Recognition) is described as one of the major challenges facing historical newspaper research in a digital environment.

Keywords: media history, digitized newspapers, OCR, digital humanities, text analysis

Hösten 2014 samlades ett sjuttioal bibliotekari-
er, forskare och representanter för tidnings-
branschen på British Library i London för en
konferens om digitalisering av äldre dagspress
inom ramen för kulturarvsportalen Europeana.
'Europeana Newspapers' har ambitionen
att digitalisera 18 miljoner tidningssidor, va-
rav drygt hälften ska vara sökbara i fulltext
genom så kallad 'optisk teckenigenkänning'.
Just OCR, 'Optical Character Recognition', var
ett av konferensen mest omdiskuterade teman.
Det är inte förvånande eftersom OCR på flera
sätt är den teknik som gör digitaliserad dags-
press till en helt ny typ av forskningsmaterial.
OCR är en process i flera steg som konverterar
dagspressen som medium – från tidning till
maskintext. Det sker genom en närmast cirku-
lär textuell migreringsprocess där tidningssida
blir inskannad bild, vilken därefter omkodas

till maskinläsbar text. Ur den senare kan dels
skapas en sorts inherent data, extrapolerad ur
det inskannade materialet – vilket är en av di-
gitaliseringens stora fördelar – dels möjliggör
den maskinkodade texten storskaliga kvanti-
tativa och jämförande analyser av exempelvis
ords samförekomster.

Samtidigt är OCR-tekniken fortsatt behäf-
tad med problem, vilket London-konferen-
sen också resonerade kring. Har vi nått 'peak
OCR' var det flera deltagare som undrade. Sär-
skilt tröstlöst tycktes läget vara i länder där
frakturstilen dominerat. 1930 publicerades
exempelvis fortfarande omkring 60 procent
av alla trycksaker i Tyskland i fraktur, och
frakturstilen var även vanlig i Sverige fram
till 1800-talets slut. Men även den latinska
tryckstilen, antikva, orsakar igenkännings-
problem menade andra. Det marknadsledande

företaget ProQuest, och deras inskanning av 472 brittiska tidskrifter mellan 1681 och sent 1930-tal, ger exempelvis fantastisk access till ett omfattande källmaterial. Samtidigt har forskare påpekat att till och med ProQuest sofistikerade OCR-algoritmer fortsatt inte riktigt går att lita på. I en genomgång av viktoriaiska texter om Kuba i det inskannade tidskriftsmaterialet, har Albert D. Pionke bland annat visat på betydande svårigheter för ProQuests algoritmer att urskilja ordet 'Cuba' från ord som: 'cash', 'Cheap', 'Colours', 'cube' eller 'cure' (Pionke 2014: 391). Det säger sig självt att sådana 'bad matches' kan leda till felaktiga slutsatser och forskningsresultat. OCR-kodad text är med andra ord synnerligen användbar – men långt ifrån felfri.

Den här artikeln innehåller dels en mediehistoria kring den svenska dagspressens digitala förhistoria (med fokus på mikrofilm), dels resonerar den kring de möjligheter som nya digitala metoder ger den mediehistoriska forskningen. Utifrån en mediehistorisk infallsvinkel uppmärksammar texten den digitala teknikens möjligheter att på ett generellt plan analysera stora textmängder på jakt efter lingvistiska mönster, bland annat genom så kallad distansläsning, en sorts data-metodologisk läsning av stora textmassor som populariserats av bland andra litteraturvetaren Franco Moretti med fokus på 1800-talets skönlitteratur (Moretti 2013). För att illustrera metodernas möjligheter och problem görs en teststudie på ett lättillgängligt material från 1990-talet.

Dagspressen remedierad – från mikrofilm till OCR

I bibliotekskretsar, kunde man läsa i 1949 års direktiv rörande biblioteks- och arkivfilmning, uppmärksammas numera den så kallade 'mikrofilmningen' vilken innebär "att man i stark förminskning fotograferar den handling, som man önskar kopiera. ... Vad biblioteken beträffar är i detta sammanhang spørsmålet om bevarandet av de dagliga tidningarna särskilt betydelsefullt." Möjligheten att kopiera dagstidningar var kring 1950 svaret på två brännande frågeställningar när det

gällde informationshantering: dels vållade det tråhaltiga pappret i dagstidningar att de (i regel) hade mycket kort livslängd, dels utgjorde tillväxten av antalet tidningar ett skrymmande utrymmesproblem (SOU 1951: 36: 7).

Mikrofilmning av svenska dagstidningar startade i privat regi redan i slutet av 1940-talet genom företaget Rekolid/Cefab (Centrala filmarkivet AB). Verksamheten var delvis ett sätt att svara upp mot tidningsbranschens behov, men även Kungliga biblioteket började snart att beställa filmade tidningskopior. Rekolid/Cefab kom under 1950-talet att samarbeta med både KB och Lunds universitetsbibliotek. Sammantaget mikrofilmade Rekolid/Cefab cirka 360 tidningstitlar fram till 1970-talet (Widholm 2011). I en rapport från Statskontoret 1969 påtalades att mellan 1850 till 1950 hade cirka 30 procent av alla tidningssidor mikrofilmats, och för perioden därefter uppmot 45 procent (Statskontoret 1969). När en ny utredning gjordes vid mitten av 1970-talet bedömdes att ungefär hälften av alla dittills utkomna tidningar hade filmats. En del av de befintliga mikrofilmerna var dock av så dålig kvalitet att omfilmning krävdes. Eftersom tidningarna var inbundna var det ibland svårt att få hela sidan läsbar på bild. Ofta var det innerspalten som var oläslig, och där placerades bland annat ledarartiklar och presskommentarer, av tradition centrala texter i presshistorisk forskning. Överföring från ett analogt medium till ett annat innebär i princip alltid förlust av data, även om överföringen i övrigt förlöpt som det är tänkt. Vad gäller mikrofilmningen av dagspress anfördes också att den premierade texter på bildernas bekostnad; egenskaper som storlek och papperskvalitet gick naturligtvis också förlorade (Tidningsfilmningskommittén 1976).

Såväl mediehistoriskt som informationshanteringsmässigt kan man notera att om mikrofilmtekniken redan kring 1950 framför allt knöts till tidningsmediet – både som ett lovande bevarandeinstrument och som ett sätt att tillgängliggöra presshistorien – så förknippades automatisk databehandling (ADB) och tidiga digitaliseringsinsatser inom bibliotekssektorn under 1970-talet snarare till

bokmediet. Anledningen har delvis att göra med bristfällig OCR-teknik där pressläggen innehöll alltför mycket brus, för att tala med en informationsteoretisk vokabulär. Boksidor var grafiskt renare, och därför enklare att digitalisera (åtminstone framstod de så i teorin). Mikrofilm och digitalisering är med andra ord relaterade till varandra på flera sätt, och under lång tid existerade de följaktligen som parallella informationshanteringstekniker. Inte förrän i den så kallade KB-utredningen 2003 påtalades att KB borde ges ”uppdraget att kartlägga fördelar och nackdelar med att ersätta mikrofilmningen med digital leverans och lagring” (SOU 2003: 129: 142).

KB avslutade sin reguljära mikrofilmning av svensk dagspress så sent som 2013. Det var en verksamhet som i egen regi hade påbörjats 1979 (genom att delvis införliva Rekolid/Cefab), detta parallellt med att Arkivet för ljud och bild påbörjade sin verksamhet – baserad på en utvidgad pliktexemplarslag för audio-visuella medier – där framför allt etermedier kopierades på video- och magnetband. *Medietransfer* var här den gemensamma nämnaren. Mediernas materialitet nedprioriterades, samtidigt som innehållet uppgraderades till att bli vad som primärt värdesattes för forskning och som framtida kulturarv. Det är en underskattad förändring i mentalitetsarkivariskt hänseende – innehåll ersatte form. Omställningen i den tankemässiga och praktiska hanteringen av medialt kulturarv som verksamheten på ALB introducerade förklarar också varför den institutionen (vilken sedermera bytte namn till Statens ljud- och bildarkiv) var långt mer framgångsrik (än KB) på att ställa om sin verksamhet i digital riktning. Om innehåll kopierades till magnet- och videoband eller till digitala bärare var strängt taget av mindre betydelse – innehållet var allt, formatet intet.

Att tidningar och böcker skulle digitaliseras och bibliotekens kortkataloger göras tillgängliga över nätverk kan från dagens horisont förefalla vara en självklarhet. Men teknikomställningen inom bibliotekssektorn har varit långsam. Under 1960-talet var det snarare genom datoriseringen av bank- och försäkringsbranschen som digitaliseringens

möjligheter gradvis uppenbarades. Just maskinläsning – som enligt Computer Swedens datalogiska språkwebb bör vara den svenska översättningen av OCR – har en i sammanhanget intressant medie-ekonomisk historia. Under 1960-talet började nämligen checkar och inbetalningskort att OCR-behandlas. För det krävdes att speciella typsnitt användes för att tekniken skulle fungera, först i mitten på 1970-talet utvecklade datavisionären Ray Kurzweil en mjukvara som kunde OCR-läsa i princip alla typsnitt. Det var en produkt som kopieringsföretaget Xerox köpte 1978 för att underlätta överföring av elektronisk information till papper, och därefter har tekniken kommersialiserats på bred front (Schantz 1982).

Att just Xerox övertog patent på tekniken är inte konstigt eftersom all OCR-behandling innebär ett slags omkopiering, där algoritmen i ett första steg omformar färg eller gråskala i originaldokumentet till bi-tonala bilder i svart-vitt – och därefter läser om den analoga texten till binär data. Det är den processen som ofta orsakar problem då alla tecken inte uppfattas på exakt rätt sätt av programvaran. Beträffande dagspress är det i regel så att ju äldre tidningsmaterialet är desto sämre fungerar OCR-tolkningen. En äldre tidningssida kan exempelvis ha blivit nött av användning eller missfärgad, vilket inte sällan resulterar i en ojämn mörkerskala på satsytan. Den senare är svår att exakt OCR-läsa, även om programvaror som exempelvis Tesseract använder sig av algoritmisk efterbearbetning för skannade ark. ‘Optical Character Recognition’ handlar som termen antyder om *igenkännbarhet* på teckenbildsnivå, vad som på svenska kallas ‘glyf’. När teckenbilderna är grafiskt svårläsliga – som i äldre presslägg med frakturstil – blir de också svåra att maskinläsa. Det gäller naturligtvis också om tidningens satsyta hamnat snett i skannern; OCR-algoritmen är helt beroende av raka, linjära rader. OCR-algorithmens svårigheter gäller därför dels *urskiljandet* av specifika glyfer, dels *överensstämmelser* mellan olika teckenbilder, som exempelvis *a* respektive *ä* – vilka är två glyfer som representerar samma tecken. Att grafisk igenkänning kan

vara knepig är egentligen uppenbart för var och en; skillnaden mellan glyfen 'I' och 'l', 'I' eller '↑' är trots allt inte så stor – speciellt inte för en dator. Under det senaste två decennierna har både kommersiella företag liksom olika nationalbibliotek därför strävat efter att optimera OCR-kodningen, inte minst genom införandet av dynamiska begreppsordböcker (tesaurus) som OCR-tolkningen matchas mot. Generellt brukar det för äldre dagspress hävdas att om mer än tio procent av innehållet är felkodat så är OCR-tolkningen av dålig kvalitet. Över 90 procent OCR-överensstämmelse är däremot för de flesta aktörer ett acceptabelt resultat (Holley 2009).

KB digitaliserar

Genom en tilltagande datorisering av biblioteks- och arkivväsendet under 1970- och 1980-talet uppenbarades digitaliseringens potential för alla inom ABM-sektorn. I Sverige tog den digitala omställningen dock lång tid – inte förrän 1997 introducerade KB exempelvis genom Libris sitt så kallade webbsök. Ett år senare påbörjades projektet, 'Tiden – Digitalisering av historiska tidningar i Norden', ett skandinaviskt samarbetsprojekt för att utveckla metoder för digitalisering av tidningar. I projektet prövades olika OCR-läsningstekniker av både *antikva* och *fraktur* för att göra tidningarna sökbara. Men verksamheten syftade främst till att höja kompetensnivån kring tidningsdigitalisering. Intern och tekniskt orienterad metodutveckling har på många sätt varit ledstjärna för i princip samtliga av KBs digitaliseringsinsatser – snarare än resultat i form av digitalt tillgängligt material för forskare och andra intresserade. Efter projektet 'Tiden' följde under 00-talet ytterligare två försök att få fart på tidningsdigitaliseringen: 'Digitalisering av svensk dagspress', samt ett EU-finansierat projekt 'TELplus'. Resultaten var ånyo magra; inalles skannades bara omkring 200 000 sidor tidningstext från 24 svenska dagstidningar i ett försök att visa en helhetsbild av svensk dagspress. KB tog under denna period föga intryck av Googles omtalade bokskanningsprojekt.

Snarare präglades även de två nya digitaliseringsprojekten av ett internt sökande efter lika optimala som utsiktslösa metoder och tekniker, där KB bland annat provade att skanna dagspress från mikrofilm. Eftersom pressmaterialet till största delen bestod av äldre lägg från 1700- och tidigt 1800-tal (då frakturstilen dominerade) blev OCR-läsningen därefter. Stickprov ger vid handen att knappt en enda mening tolkades rätt.

På detta minst sagt valhanta sätt fortsatte KB att hantera frågan om digitalisering av dagspress, ett av nationalbibliotekets allra mest använda forskningsmaterial. Så sent som i årsredovisningen för år 2008 kunde riksbibliotekarie Gunnar Sahlin påpeka att "dags- och veckotidningar behöver digitaliseras i stor skala. KB har inlett ett arbete men ytterligare resurser krävs" (KB årsredovisning 2008: 5). Vid denna tidpunkt hade de flesta nationalbibliotek i Europa redan digitaliserat stora delar av sin dagspress. I digitaliseringshänseende var Sveriges nationalbiblioteket kring 2010 ett av Europas minst utvecklade. Genom en beviljad ansökan om medel ur EU:s strukturfonder kopplat till Riksarkivets mediekonverteringscentral (MKC) i Västernorrland lyckade KB emellertid få till stånd ett pilotprojekt – återigen dock med syfte att "utveckla kompetens, arbetsrutiner och teknik för digitalisering av dagstidningslägg." Projektet, påtalades det i årsredovisningen 2009, "lägger grunden för framtida digitalisering av svenska dagstidningar i stor skala" (KB årsredovisning 2009: 11). Genom att även andra finansiärer gick med, bland annat Schibstedtkoncernen, lades grunden för digitaliseringsprojektet Digidaily.

Digidaily har under de senaste åren (tillsammans med Kungliga bibliotekets migrering av audiovisuella medier) varit nationalbibliotekets digitala flaggskepp. Efter snart två decennier av idel tveksamheter har Digidaily faktiskt inte enbart resulterat i tekniska metoder och kravspecifikationer, utan också i konkreta resultat i form av massdigitaliserade tidningslägg. I skrivande stund är mängden sökbara tidningssidor strax över 12 miljoner – men majoriteten av dem kommer från nutida tidningar, från 2012 och framåt. Sök-

tjänsten på tidningar.kb.se, 'Sök bland svenska dagstidningar', innehåller för närvarande hela *Svenska Dagbladet* och *Expressen* och nästan hela *Dagens Nyheter* och *Aftonbladet*. Därtill finns *Dagens Industri* från 1983 till idag samt *Göteborgs Aftonblad* mellan 1888 och 1918. KB tolkar emellertid upphovsrätten strängt, och material publicerat senare efter 1900 är upphovsrättsskyddat och kan endast läsas lokalt i Stockholm.

Digidaily är inte minst ett lyckat projekt eftersom kostnaden per digitaliserad tidningssida ligger långt under den internationella prisbilden. 2010 redovisade Collection Trust till EU-kommissionen en kostnadsrapport (Poole 2010) av vilken det framgår att kostnaderna per digitaliserad tidningssida i projekt som bedrivs vid British Library ligger på cirka £1. En genomsnittssida i Digidaily produceras däremot för sju svenska kronor. Resultaten från Digidaily-projektet visar dock att priset är starkt beroende av tidningssidans beskaffenhet; ju sämre kondition en sida befinner sig i desto högre kostnad och desto sämre kvalitet på OCR-resultatet.

Genom Digidaily är de fysiska förutsättningarna för digitaliseringsprocessen färdiggetablerade när det gäller historiska dagstidningar. KB:s tidningsarkiv skiljer sig från många andra nationalbiblioteks eftersom man i början av 2000-talet beslöt att ta hand om Uppsala universitetsbiblioteks tidningssamling. Det betyder att KB idag förfogar över en nästan komplett dubbeltsamling, vilken med fördel kan användas för digitalisering genom att läggen på intern jargong 'slaktas' – bara så kan digitaliseringspriser pressas och en kritisk massa av skannad dagspress uppnås. Noterbart är att KB vid flera tillfällen äskat extra medel från regeringen (bland annat till 2012 års forskningsproposition) för att försätta att digitalisera merparten av den svenska dagspressen inom ramen för Digidaily. Den totala dagstidningssamlingen på KB omfattar 122 miljoner sidor, och nationalbiblioteket har kalkylerat att det reguljära anslaget skulle behöva ökas med omkring tio miljoner per år (i tio år) för att digitalisera det för forskningen mest värdefulla tidningsmaterialet.

Digitala metoder – från närläsning till distansläsning

Vad kan ni göra med en miljon inskannade böcker? Eller en miljon tidningssidor? Sådana frågor har sedan 2009 ställts av ett antal anglosaxiska forskningsstiftelser i utlysningen om forskningsmedel inom den så kallade 'Digging into Data Challenge'. En rad utlysningar har gjorts, och den övergripande frågan har gällt hur ny tillgång till enorma mängder digitaliserat (kulturarvs)material påverkar den humanistiska forskningen – praktiskt såväl som teoretiskt. Vilka nya frågor (och svar) kan genereras av datorkraft applicerad på ett slags kvantitativ empirisk skalbarhet? Termen 'data mining' anförs ofta i sammanhanget, men metodfrågor har också varit vanliga. Givetvis är det utmärkt att digitalisering förenklar tillgängligheten till forskningsmaterial; för mer human-tekniskt orienterade forskare ställer digitaliseringsprocessen dock framför allt frågor kring nya digitala metoder liksom en uppdaterad digital forskningspraktik (Rogers 2013). Från ett forskningsperspektiv ligger digitaliseringens stora betydelse i att vetenskapligt kvantitativa forskningsperspektiv på allvar kan tas i bruk inom humaniora, och i så måtto både utveckla och komplettera de traditionellt kvalitativa hermeneutiska läsningarna utifrån förmodade nyckeltexter.

Att utföra sökningar av stora datamängder var länge förbehållet naturvetare och samtidsorienterade samhällsvetare, men med historisk dagspress i massdigitaliserad form kan även humanister (tillsammans med kulturarvsinstitutioner) utföra storskalig analys av enorma mängder data; det gäller både 'hårda' mätbara uppgifter kring diverse ekonomiska förhållandet såväl som 'mjuka' kring begreppsanvändning och diskursiva frågor. Inom det växande forskningsfältet kring digital humaniora kan man just skönja hur modern datateknik och växande digitala samlingar av historiskt material håller på att förändra (delar av) historievetenskapen i en storskalig och kvantitativ riktning (Svensson 2011, Svensson 2012). I korthet, handlar det om att bedriva historisk forskning som datorbaserad vetenskap, snarare än som mer eller

mindre godtycklig hermeneutisk praktik. Vad som förr räknades som heroiska insatser med år av mikrofilmsrullande kan nu göras av en dator på nolltid – åtminstone i teorin. Frågeställningar har givetvis skiftat, men en betydande fördel med digital humaniora-inspirerad metod och teknik är att nya storskaliga frågor kan ställas och besvaras.

Den presshistoriska forskningen i Sverige och internationellt har framförallt dominerats av studier som ägnat sig åt enskilda genrer, skribenter eller texter knutna till specifika händelser. Om större grepp har tagits har det ofta varit tidningsföretagen snarare än innehållet som stått i fokus. Undantagen utgörs framförallt av ett antal kvantitativa studier från 1970-talet. Symptomatiskt är att de använde sin tids datormöjligheter för att söka och kvantifiera. Eftersom OCR-tekniken inte fått något bredare genomslag kvävdes emellertid betydande mänskliga resurser för att ex-cerpera och koda (Jarlbrink 2015). I övrigt har de kvalitativa studierna dominerat. En viktig anledning till detta är att presshistorisk forskning ofta haft rötterna i litteraturvetenskapliga traditioner. Att närgranska kanoniska texter såsom följetonger, ledare och kulturartiklar, skrivna av tongivande skribenter, ligger i linje med traditionella litteraturvetenskapliga angreppssätt. Ofta har forskningen kretsat kring 'bra' artiklar, där det funnits ett behov av att visa att även den flyktiga dagspressen innehållit texter av litteraturhistoriskt värde. En annan förklaring till detta något snäva perspektiv är att det har varit svårt att orientera sig i materialet. Antalet tidningar och texter har varit överväldigande. Ingen har kunnat läsa allt, och den som sökt efter något specifikt har sällan haft några register att gå efter. Vad som har funnits är bibliografier för några få titlar och år, men de har haft flera begränsningar. Bibliografiernas upplägg har ofta verkat styrande för vad som varit möjligt att hitta och beforska. *Svenskt pressregister*, som en gång initierades på Litteraturvetenskapliga institutionen i Lund, har exempelvis förtecknat poesi och prosa, kulturartiklar och signerade texter, det vill säga sådant som intresserar den litteraturhistoriskt orienterade pressforskaren.

Osignerade texter i andra genrer – som alltid dominerat i spalterna – har valts bort och osynliggjorts. I det läget har många forskare valt att koncentrera sig på några få kanoniska tidningstitlar, på enskilda genrer eller författarskap, eller på texter som varit möjliga att hitta eftersom de publicerats i samband med någon specifik händelse.

Mot bakgrund av ovanstående kritik av traditionerna inom den presshistoriska forskningen ska det samtidigt medges att den forskning som försökt analysera presshistoriskt material med hjälp av digitala metoder än så länge inte alltid har varit mycket bättre. I flera studier har forskare experimenterat med metoder, men resultaten har ofta varit triviala. En forskargrupp har exempelvis använt 'topic modeling', en metod för att analysera ords samförekomster, för att bearbeta 232 000 sidor ur en rad tidningar från Texas mellan 1865 till 1930. Ett resultat är att det skrivs mycket om bomull i dessa tidningar – vilket de flesta presshistoriker hade kunnat räkna ut utan avancerad programvara (Yang, Torget & Michalcea 2011).

En anledning till att den här typen av forskningsresultat än så länge ofta framstår som närmast banala ur ett presshistoriskt perspektiv, är att forskningsprojekten i regel drivs av informations- och datavetare. De är mer intresserade av att testa och utveckla programvara än att åstadkomma mediehistoriskt intressanta analyser (t.ex. Hendrickx et al. 2013; Kushkuley 2012). Men givetvis existerar det undantag. Som en mer presshistoriskt driven pionjär framstår exempelvis Sharon Block, som tillsammans med datavetaren David Newman använt så kallad 'pLSA' (Probabilistic Latent Semantic Analysis) för att kartlägga samförekomster av ord och de ordkluster – 'topics' – som samförekomster av högfrekventa ord bildar. Materialet bestod av 80 000 artiklar och annonser i *Pennsylvania Gazette*, från 1728 till 1800. Block och Newman kunde visa hur enskilda ordkluster (exempelvis rörande statstyre, religion och brottslighet) ökade eller minskade över tid, hur enskilda ordkluster relaterade till varandra, och vilka kluster ett enskilt ord med störst sannolikhet

ingick i. Exempelvis visade det sig att ämnet 'statsstyre' blev mer och mer marginellt från 1730-tal till 1750-tal, för att därefter öka kraftigt igen under 1700-talets andra hälft. Ämnet 'textilier' gjorde en motsatt resa. Analyserna av vilka ordkluster ett enskilt ord med störst sannolikhet ingick i, kunde dessutom synliggöra hur marginaliserade vissa grupper av människor var. Block och Newman kunde till exempel visa att ordet 'Negro' nästan enbart förekom inom klustret tjänare/slav – och även ordet 'woman' återfanns oftast inom samma kluster (Newman & Block 2006).

Block och Newman har visat att det är fullt möjligt att använda digitala metoder för att synliggöra mönster som hade varit mycket svåra att upptäcka med traditionella metoder. Men de har också påpekat att metoderna mycket väl kan kombineras, att den digitala analysen kan visa på samband som kan undersökas vidare med andra metoder (Newman & Block 2006, Block 2006). Få senare studier har visat de digitala metodernas fördelar lika framgångsrikt som Block och Newman, även om en del forskare försökt. Bob Nicholson (2012) har med enkla sökningar i British Librarys digitala tidningsarkiv visat hur samförekomsterna av orden 'konkurrens' och 'Amerika', respektive 'Frankrike' och 'Tyskland', varierar under 1800-talets andra halva – och att artiklarna om utländsk konkurrens tycks samvariera med spannmålsskördarna. Colella (2013) och tidigare anförda Pionke (2014) har använt något mer avancerade metoder för att undersöka så kallade 'nyckelord i kontext' ('Key Word in Context'). I Collellas fall handlar det om hur affärslivet beskrivs i brittisk 1800-talspress. I Pionkes fall om handel och slavarbete i artiklar om Kuba. Huijnen et. al (2014) och Eijnatten et. al. (2014) använder en liknande metod för att kartlägga diverse ämnen i framförallt holländsk press. Verktøget Texcavator erbjuder här användaren att ange söktermer, precis som en vanlig online-databas, men presenterar sökresultatet i ett ordmoln som visar vilka frekvent använda ord som förekommer tillsammans med sökordet. Detta möjliggör komparationer mellan olika tidsperioder, tidningar och länder. En del av

deras forskning relaterar till medicinhistoriska frågor, och här har Huijnen et. al. (2014) bland annat visat hur eugenik går från att tillhöra en medicinsk diskurs under 1800-talets mitt, till att bli en del av en rasdiskurs i början av 1900-talet. Detta förutsätter dock att forskarna får direkt tillgång till tidningsmaterialet utan att behöva gå via biblioteksportalernas gränssnitt. Förmodligen är det en viktig anledning till att antalet presshistoriska studier som använt digitala metoder än så länge är få till antalet.

Inom delar av det litteraturvetenskapliga forskningsfältet har man hunnit betydligt längre. Orsaken är att de skönlitterära textmängderna är mer begränsade, att upphovsrättsproblematiken är enklare att hantera, samt att skönlitterära texter (vilket tidigare påtalats) som publicerats i bokform är renare, enklare att digitalisera och analysera jämfört med tidningslägg. Matthew Jockers, en av de tongivande forskarna på fältet, presenterar i boken *Macroanalysis* (2013) en rad analyser av engelskspråkiga 1800-talsromaner. Traditionellt har den engelskspråkiga litteraturvetenskapen ägnat sig åt en kanon som bara utgör en mindre del av allt som publicerats. Jockers tar ett större grepp och behandlar sammanlagt 3 346 romaner. Genom att använda verktyg för topic modeling visar han hur olika ämnen ökar och minskar i popularitet, hur ämnesvalen skiljer sig mellan olika länder, och mellan kvinnliga och manliga författare. Genom att mäta och jämföra ord- och ämnesval har Jockers även studerat avståndet mellan olika romaner och författare. Här visar det sig exempelvis att kvinnliga och manliga författare beskriver helt olika världar, och använder olika språk (Jockers 2013, kap. 8-9). De digitala metoder som används gör det möjligt att fånga mycket övergripande mönster, men också att visa likheter och skillnader på detaljnivå. Mänskliga läsare missar förmodligen den relativt höga frekvensen av ordet 'like' i bildningsromaner. När sådana mönster på mikronivå synliggjorts med digitala redskap kan en mänsklig läsare ta vid och försöka förklara vilken betydelse ordvalet har i texten (Jockers 2013, 89ff). Det är såtillvida i samspelet mellan människa och

maskin som den verkliga kunskapsutvecklingen sker.

Digitala lägg

I projektet, 'Digitala lägg. Om pressens gränssnitt 1800' (finansierat av Torsten Söderbergs stiftelse) avser vi att testa de metoder och forskningsperspektiv som beskrivits ovan. Ambitionen är att både utveckla metoderna och att generera mediehistoriskt intressanta forskningsresultat. Som projekttiteln anger är det i första hand 1800-talspressen som kommer att stå i centrum. På grund av strikt tolkad upphovsrättslagstiftning är det tidningsmaterial från 1800-talet som det är lättast att få tillgång till. Detta äldre material är emellertid svårast att arbeta med rent tekniskt. De flesta metoder för topic modeling och kartläggning av samförekomster kräver att varje text är tydligt avgränsad från de omgivande texterna på tidningssidan. När tidningssidor digitaliseras används en programvara som genomför en automatisk segmentering. Eftersom denna i flera fall är bristfällig krävs en manuell rättning av felsegmenterade texter. Digitaliseringen som KB stått för har emellertid hoppat över detta steg, vilket innebär att så gott som varje sida innehåller fel: halva följetongen har klumpats ihop med annonsen för galoscher, ledartikeln har styckat upp i två separata texter, etcetera.

Materialet kan ändå användas, men kräver speciallösningar. Istället för att utgå ifrån en befintlig – segmenterad – artikelenhet (som alltså ofta är felaktig) kan man exempelvis avgränsa analysen till textsjok eller 'fönster' som innehåller specifika nyckelord. För att prova möjliga tillvägagångssätt har vi genomfört en mindre förstudie på ett mer lättarbetat tidningsmaterial. Vår tanke har varit att denna studie ska kunna tjäna som metodologisk modell för senare undersökningar, samtidigt som resultatet också ska kunna ingå i komparativa analyser. Den övergripande frågeställningen som väglett arbetet har varit hur ett nytt medium språkliggörs när det introduceras och diskuteras offentligt inför en publik av tidningsläsare. Mer konkret har vi velat

undersöka vilka ord som används frekvent i tidningstexter om ett nytt medium, samt hur dessa ord relaterar till varandra. Frågeställningen undersöks med hjälp av ett program som kartlägger ords samförekomster.

Tidigare forskning om nya medier i historien framhäver just introduktionsfasen som särskilt intressant. I ett sådant initialt skede är det inte givet hur mediet ska användas, vilken betydelse det ska få, vem som ska (tillåtas) använda det, hur det ska regleras (Thorburn & Jenkins 2003). När mediet är nytt pågår en sorts *förhandling* om dess betydelse och användning (Marvin 1990). Sådant som senare uppfattas som självklart måste förklaras, den 'svarta lådan' har ännu inte stängts igen (Latour 1987). Tekniker och experter får berätta om mediets möjligheter (Gitelman 2006), samtidigt som kritiker varnar för dess negativa konsekvenser (Gustafsson & Arnberg 2013).

För att undersöka kontinuitet och förändring i hur nya medier beskrivs och diskuteras är vår ambition att göra ett antal nedslag i tidningsmaterial som kan belysa introduktionen av en rad nya medietekniker: dagerrotyp i kring 1840, den elektriska telegrafan under 1850-talet, telefonen kring 1880, kinematografi under 1890-tal. Ett första metodförsök gäller emellertid ett betydligt senare medium: internet under 1990-talet. Metoden som används inbegriper ett antal steg, från insamling och urval till bearbetning och analys. För att åskådliggöra processen kommer den att presenteras kortfattat. Syftet är inte att leverera en slutgiltig analys av insamlade data utan mer att åskådliggöra metodens huvuddrag.

Samtliga artiklar ur *Expressen* och *Göteborgs-Posten* från år 1994 och 1997, tillgängliga i databasen Retriver, samlades in och indexerades med programvaran Lucene (McCandless et al. 2010). Programmet erbjuder avancerade sökmöjligheter såsom booleska sökningar, trunkering och så kallad fuzzy search. De två tidningarna utgör givetvis bara en liten del av det totala tidningsbeståndet, men representerar ändå olika tidningstyper med olika utgivningsort. Det första året har valts för att det representerar ett relativt tidigt skede i nätets (och webbens) publika historia. Mediet

var 1994 inte helt nytt – men ändå en nyhet för de flesta läsare. Det senare året, 1997, har internet och webben fått långt fler användare; då var nätet inte längre en teknik för några få entusiaster.

I de insamlade texterna – 102 616 stycken från 1994 och 131 880 från 1997 – sökte vi efter texter innehållande nyckelordet 'internet*'. Dessa texter annoterades för att möjliggöra filtrering med avseende på ordklasser. För att kunna slå ihop olika böjningar av ett och samma ord till ett enda ord användes den fritt tillgängliga lingvistikapplikationen Stagger (Östling 2013). Stagger nyttjades i huvudsak till ordklasstagning, och detta material låg sedan till grund för sökningar av ords samförekomst. Länken mellan två ord i graferna (figur 1 och 2) utgörs av det totala antalet samförekomster mellan orden inom ett givet fönster i texterna. I detta fall har fönstret utgjort tre ord: ett ord samförekommer med ett annat om det förekommer tre ord före eller efter ett annat. Styrkan i samförekomsten definieras av en normalisering baserad på frekvensen av två ords samförekomst i relation till den totala förekomsten av dessa ord i textmaterialet, enligt det så kallade Jaccard indexet (Salton & McGill 1982). Om två ord alltid förekommer tillsammans är sambandet alltså starkt, om de två orden förekommer ofta, men bara samförekommer några enstaka gånger är sambandet svagt. För att undvika brus har småord sorterats bort. Listan över stoppord innehåller sextiofem småord i stil med: 'i', 'att', 'och', 'det'. Beroende på materialet och forskningsintresset kan ord läggas till eller tas bort från stoppordlistan. Våra inledande sökningar i tidningsmaterialet visade att paratextuella ord som bildtext, teckning, Expressen och Göteborgs-Posten också behövde läggas till på listan över stoppord.

Antalet texter som innehöll nyckelordet 'internet*' var 101 från 1994 och 2107 från 1997. Samförekomster har sökts bland ord inom ordklasserna substantiv, egennamn och utländska ord. För att inte ytterligare reducera det mindre materialet från 1994, och göra det större materialet från 1997 svårbehandlat, har en minimigräns satts vid tre förekomster 1994

och sju förekomster 1997. Detta innebär att ordpar som endast samförekommer enstaka gånger (färre än tre respektive sju gånger) har sällats bort. Med dessa förutsättningar genererade sökningen 420 ordpar i materialet från 1994, och 2 662 par 1997. Samförekomstanalysen resulterar i data vilken kan representeras som ett nätverk – som i sin tur kan importeras i program för statistisk analys och grafisk visualisering. Här har vi använt visualiseringsverktyget Gephi. Visualiserat som nätverk har våra sökningar i materialet från 1994 genererat 243 noder och 410 länkar – respektive 1165 noder och 2602 länkar för år 1997.

Även med stoppordlistor, ett begränsat fönster och en spärr för lågfrekventa ordpar innehåller nätverken en hel del brus. Ofta utgörs detta brus av fristående nodkonstellationer som består av namn – typ: Rolling och Stones, Kurt och Cobain – som inte länkar till några andra noder. Dessa är det ofta önskvärt att filtrera bort. Samtidigt kan sådana fristående subnätverk vara intressanta i sig: här finns aktörer och företeelser som är en del av materialet men som samtidigt har svag koppling till nätverkets övriga delar. Att noderna 'Ines-Uusman-kommunikationsminister' till exempel helt saknar koppling till det övriga nätverket 1997 tyder förmodligen på att hon hade en ytterst begränsad roll i de texter som skrevs om internet detta år, ett resultat som inte är oviktigt i sammanhanget. Ett annat betydande exempel från 1994 är det fristående subnätverket 'William-Gibson-roman-Neuromanser-begrepp-cyberspace'. Beroende på fönstrets storlek fångar sökningen in också det som är mer eller mindre perifert. Och att såväl kommunikationsministern som den framburande cyberpunkförfattaren endast förekommer i marginalen när det skrivs om internet kan som sagt vara värdefulla resultat i sig.

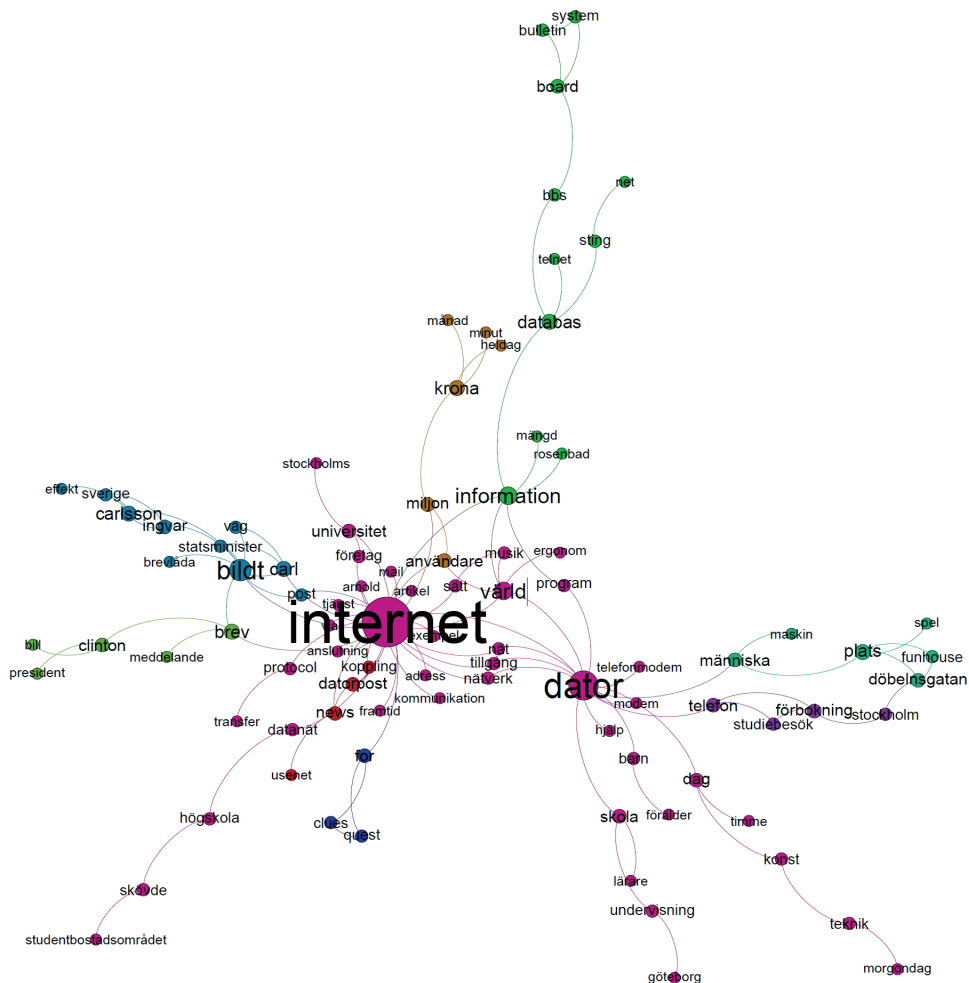
Centrala noder – i den meningen att de knyter an till flera andra noder som tillsammans bildar kluster – är det i regel klokt att framhäva. Noder med många länkar kan göras grafiskt större och olika kluster kan färgläggas för att tydliggöra vilka ord som ofta förekommer tillsammans. Klustren kan påminna om de 'topics' som skapas med andra digitala ana-

lysverktyg. De kan också ses som olika typer av gestaltningar (frames) av internetfenomenet – eller som internets diskursiva kontexter (DiMaggio et al. 2013).

De noder som dominerar är i vissa fall samma för båda år: ord som information, dag, dator, miljon, värld, et cetera. Men det finns också mycket som skiljer de båda åren åt. År 1994 har en enskild aktör en central plats: Carl Bildt. I februari detta år skickade han ett e-postmeddelande till president Bill Clinton vilket fick mycket uppmärksamhet. När de borgerliga partierna förlorade valet i september och Bildt efterträddes av Ingvar Carlsson uppstod frågan om den senare skulle få en ny 'elektronisk brevlåda' eller behålla

den som Bildt hade haft. Några få enskilda personer sätter såtillvida ett tydligt avtryck i internetrapporteringen, men de tongivande politikerna är ändå avskilda från andra delar av nätverket.

Noterbart är att majoriteten av de analyserade texterna 1994 behandlar praktiska och tekniska aspekter av internetanvändningen: Vad det är, hur man gör, vad som finns, vad det kallas, vad det kostar. Texterna lanserar och förklarar nya begrepp och företeelser som 'bbs', 'databas' eller 'modem'. Här finns också spår av diskussioner om relationen människa-maskin och datorer i skolundervisningen. År 1994 sprider sig internetanvändningen över världen – men internet är samtidigt en



Figur 1. Internet i *Expressen* och *Göteborgs-Posten* 1994

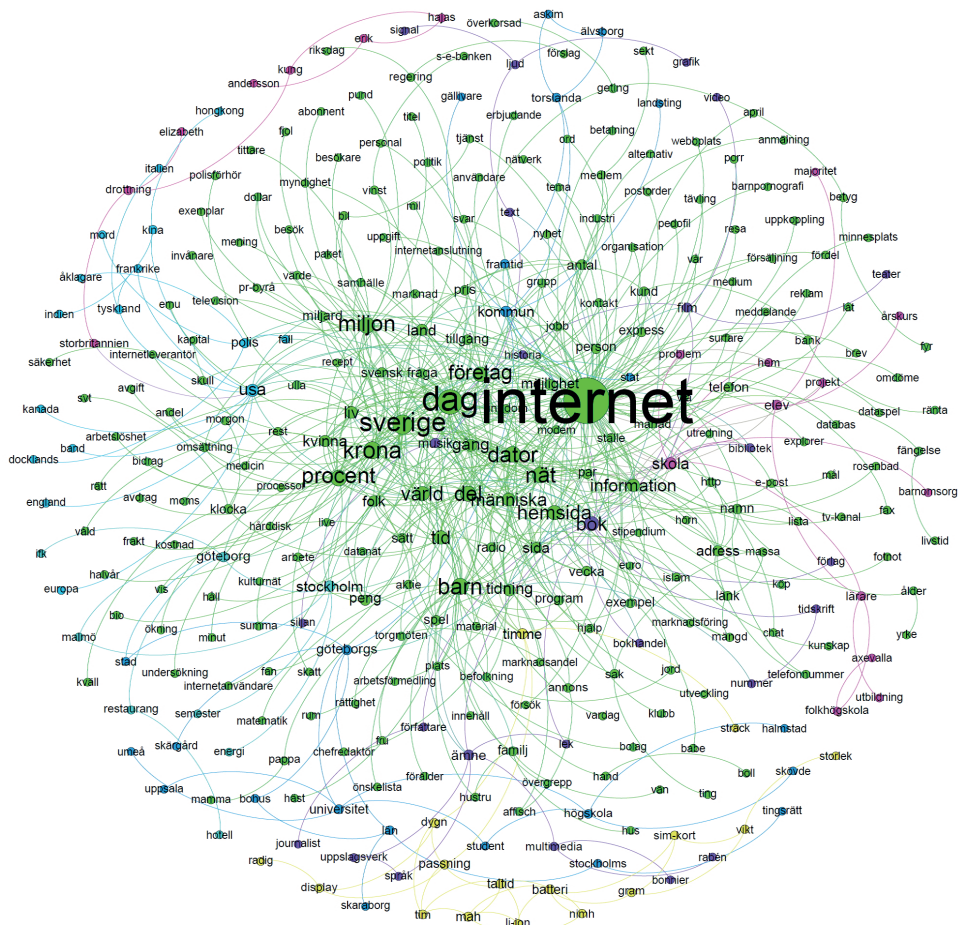
(språklig) värld i sig. Denna digitala värld är emellertid ännu ganska glesbefolkad. Enskilda exempel och verksamheter får illustrera möjligheterna: 'universitet', 'studentbostadsområdet', 'futurekids', spelhallen 'Funhouse'.

I nätverket förekommer flera ord som är synonyma: 'mail', 'post', 'brev', 'datorpost', 'brevlåda'. Om syftet är att studera kommunikationsformen snarare än ordvalen kan dessa enkelt slås ihop till en enda nod. I en analys av hur nya medier språkliggörs är det emellertid intressant att notera begreppsmångfalden. År 1994 har de journalister som skriver inte bestämt sig för en enskild benämning utan rör sig med flera olika termer.

I materialet från 1997 är Sverige en viktig del av rapporteringen. Barn har också en

central plats, liksom begrepp som 'hemsida', 'USA', 'tid' och 'bok'. Inga enskilda aktörer dominerar på samma sätt som Carl Bildt 1994. Istället förekommer en lång rad personer och institutioner. Här finns kultur och förlag, kvinnor och arbete, framtid och möjligheter, men också barnpornografi och brottslighet. Internets värld är större och språkligt mer mångfasetterad 1997.

Denna mångfald kan visserligen förblinda, men det finns flera sätt att bryta ner nätverket i mindre delar, exempelvis genom att studera ett enskilt ord och dess närmast omgivande nätverk. Barn omges exempelvis av 'föräldrar', 'spel', 'möjlighet', 'önskelista', 'tid', 'liv' och 'rättighet' – men också av 'övergrepp' och 'matematik'. Människa länkar på motsvarande sätt



Figur 2. Internet i *Expressen* och *Göteborgs-Posten* 1997

till 'värld', 'möjlighet', 'utveckling', 'kontakt'. Ord som har många länkar är mer centrala än ord med få länkar. 'Företag' väger med detta sätt att räkna 23 gånger mer än 'politik', 'kommun' tio gånger mer än 'riksdag'. Bland statsråden är internet något som berör "kommunikationsminister", "kulturminister", "skatteminister" och "stadsminister", men dessa aktörer länkar endast till några enstaka noder. USA:s inflytande är desto mer påtagligt, med länkar till 17 andra noder. Landets centrala ställning framgår också av att "Europa" inte har någon direkt länk till 'internet' i nätverkets centrum – nätverksanslutningen går via USA.

Avslutning

Med utgångspunkt i de preliminära resultat som ovanstående förstudie presenterar är det viktigt att inte övertolka de samförekomster som identifierats. Frekvenser och länkar är som påtalats resultat av sökfönstrets storlek. Med en annan optik kan bilden förändras – precis som vid all annan vetenskaplig kartläggning. Fördelen gentemot manuell kvantitativ kodning är att den digitala sökningen enkelt kan göras om på nytt om ingångsvärdena visar sig vara problematiska. Själva ordklassningen och samförekomstsökningen går i regel mycket snabbt och tar ofta mindre än en minut. Att sökverktygen är enkla att använda innebär att det finns goda möjligheter att pröva sig fram. Det är inte minst experimenterandet som gör att forskaren lär känna texterna och börjar upptäcka mönstren.

Ett alternativ till den enklare metod som valts för detta test är 'topic modeling'. Ett topic kan påminna om det begrepps nätverk som omger ett enskilt ord i nätverksanalysen ovan. Fördelen med topic modeling är att programvaran kan placera ett och samma ord i flera olika topics. Den enklare samförekomstsökningen förmår ju faktiskt inte skilja mellan, säg, schlagerartisten Nick Borgens låt 'World Wide Web' och storföretaget Atlas Copcos användning av samma webb. Självklart finns det en poäng med att visa att den norske musikern och den globala industrikoncernen samlas kring samma medieteknik. Men eftersom

sammanhanget skiljer sig åt finns också en risk med att klumpa ihop de båda. En fördel med den enklare kartläggningen av samförekomster kan emellertid vara metodens transparens. Inga komplicerade sannolikhetskalkyler ligger bakom resultaten, metoden är enkel att förstå, förklara och utvärdera, och den gör det lätt att växla mellan distansläsning och närläsning. Risken med mer förfinade verktyg är att de blir metodologiska svarta lådor vars utdata är svårutvärderade.

En styrka med digitala metoder är att de på ett enkelt sätt visar vad som är centrum och vad som är periferi. Eftersom kvantifieringen inte kräver några fördefinierade kategorier är sökningarna mer förutsättningslösa jämfört med traditionella kvantitativa metoder. Detta gör att de digitala redskapen lämpar sig väl när man vill skaffa sig översikt – samt inte minst att orientera sig i stora material. När man väl identifierat de stora dragen kan man sedan kombinera distansläsningen med närläsning av traditionell sort. Belackare av digitala humaniora brukar ofta anklaga de forskare som är verksamma i detta fält för att vara antihermeneutiska nypositivisterna med en övertro på 'Big Data'. Vad de ofta glömmat är att storskalig analys av kulturella dataset som exempelvis äldre pressmaterial ofta just handlar om att kombinera kvantitativ och kvalitativ analys. Härvidlag finns stora möjligheter att ställa helt nya frågor, att arbeta explorativt och laborativt, samt att undersöka och tänka med 'det digitala'. Som den här artikeln visat är inte minst själva metodutvecklingen (kring de forskningsfrågor som riggas) synnerligen produktiv i sig själv (ibland nästintill oberoende av de resultat som åstadkoms). Det senare gör forskningsfältet kring digital humaniora till ett av de mer progressiva inom humanvetenskaperna för närvarande.

Här finns därför, menar vi avslutningsvis, betydande metodologiska impulser som den mer traditionella medieforskningen kunde hämta näring ur. För digitala metoder behöver inte med nödvändighet appliceras på pressmaterial från 1800- eller 1900-talet. Exempelvis finns ett underskott av digitala metoder inom medievetenskapen för mätning av samtida

mediebeteenden. Ett metodologiskt problem för den medie- och kommunikationsvetenskapliga forskningen – liksom i ökande grad för de forskare som ägnar sig åt mediehistoriska frågeställningar med utgångspunkt i äldre digitaliserat mediematerial – är emellertid svårigheten att få tag på relevant data. Här handlar det inte med nödvändighet om att programmera fram nya sätt att förstå eller hantera medier som kod (även om det underlättar). Som forskningspraktik innebär digitala metoder framför allt att tänka med och förstå den mjukvara, applikationer och protokoll som idag hanterar nya (och äldre)

medier, exempelvis på exakt vilka sätt som analoga medier digitaliserats och vilka forskningsmöjligheter som då uppstår när det gäller hanteringen av filer och metadata som digitaliseringsprocessen genererar. Alla medier är sedan länge digitala. I takt med att digitaliseringen av kulturarvet fortsätter gäller det även äldre medieformer som dagspress. Att lära sig hur olika slags digitala metoder kan användas för att analysera *medier som data* för alternativa förklaringsmodeller och (ibland) ökad förståelse, framstår därför som en angelägen uppgift – både för den historiska och mer samtidsorienterade medieforskningen.

Referenser

- Arkiv- och biblioteksfilmning (1951). SOU 1951: 36. Stockholm
- Block, Sharon (2006). Doing More with Digitization: An Introduction to Topic Modeling of Early American Sources. *Common-Place*, 6(2). (www.common-place.org/vol-06/no-02/tales)
- Colella, Silvana (2013). 'That Inscrutable Something': Business in the Periodical Press. *Victorian Periodicals Review*, 46(3): 317-342.
- DiMaggio, Paul, Nag, Manish & Blei, David (2013) Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding. *Poetics*, 41(6): 570-606.
- van Eijnatten, Joris; Pieters, Toine & Verheul, Jaap (2014). TS Tools: Using Texcavator to Map Public Discourse. *Tijdschrift voor Tijdschriftstudies*, 35: 59-65.
- Gitelman, Lisa (2006). *Always Already New: Media, History and the Data of Culture*. Cambridge, Mass: MIT Press.
- Gustafsson, Tommy & Arnberg, Klara (2013). *Moralpanik och lågkultur: Genus- och mediehistoriska analyser*. Stockholm: Atlas.
- Hendrickx Iris; Düring, Marten; Zervanou, Kalliopi & van den Bosch, Antal (2013), Searching and Finding Strikes in the New York Times, pp. 25-36 in Mambrini, Francesco; Passarotti, Marco & Spörleder, Caroline (eds.), *Proceedings of The Third Workshop on Annotation of Corpora for Research in the Humanities (ACRH-3)*, Sofia: Bulgarian Academy of Sciences.
- Holley, Rose (2009). How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs. *D-Lib Magazine* 15(3/4): (www.dlib.org/dlib/march09/holley/03holley.html)
- Huijnen, Pim; Laan, Fons; de Rijke, Maarten & Pieters, Toine (2014) A Digital Humanities Approach to the History of Science. *Social Informatics: Lecture Notes in Computer Science*, 8359: 71-85.
- Jarlbrink, Johan (2015). Historievetenskapens mediehantering. Hyvönen, Mats; Vesterlund, Per & Snickars, Pelle (eds.) *Massmedieproblem*. Lund: Mediehistoriskt arkiv.
- Jockers, Matthew (2013). *Macroanalysis: Digital Methods & Literary History*. Chicago: University of Illinois Press.
- KB – ett nav i kunskapssamhället (2003). SOU 2003: 129. Stockholm,
- KB årsredovisning (2008). Stockholm,
- KB årsredovisning (2009). Stockholm,
- Kushkuley, Sophie (2012). Trend Analysis in Harper's Bazaar, pp. 84-87 in *Workshop on Computational Linguistics for Literature, Montreal, Canada, June 8, 2012*. Stroudsburg, Pa.: Association for Computational Linguistics.
- Latour, Bruno (1987). *Science in Action: How to follow Scientists and Engineers through Society*. Cambridge, Mass: Harvard University Press.
- Marvin, Carolyn (1988). *When Old Technologies Were New: Thinking about Electric Communication in the Late Nineteenth Century*. New York: Oxford University Press.
- McCandless, Michael; Hatcher, Erik & Gospodnetić, Otis. (2010). *Lucene in Action*. Greenwich: Manning.
- Moretti, Franco (2013). *Distant Reading*. London: Verso.
- Newman, David J. & Sharon Block (2006). Probabilistic Topic Decomposition of an Eighteenth-Century American Newspaper, *Journal of the American Society for Information Science and Technology*, 57(6): 753-767.
- Nicholson, Bob (2012). Counting Culture; or, How to Read Victorian Newspapers from a Distance. *Journal of Victorian Culture*, 17(2): 238-246.
- Pionke, Albert D. (2014). Excavating Victorian Cuba in the British Periodicals Database. *Victorian Periodicals Review*, 47(3): 369-397.
- Poole, Nick (2010). *The Cost of Digitizing Europe's Cul-*

- tural Heritage. A Report for the Comité des Sages of the European Commission.* (http://nickpoole.org.uk/wp-content/uploads/2011/12/digiti_report.pdf)
- Rogers, Richard (2013). *Digital Methods*. Cambridge, Mass.; MIT Press.
- Salton, Gerard, & McGill, Michael J. (1982). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Schantz, Herbert F. (1982). *The History of OCR, Optical Character Recognition*. Manchester: Recognition Technologies Users Association.
- Svensson, Patrik (2011). From Optical Fiber to Conceptual Cyberinfrastructure. *Digital Humanities Quarterly*, 5(1).
- Svensson, Patrik (2012) Envisioning the Digital Humanities. *Digital Humanities Quarterly*, 6(1).
- Statskontoret (1969). Mikrofilmning av svensk dagspress. Förstudierapport. Riksarkivet – Dataarkiveringskommittén volym E1:3.
- Thorburn, David & Jenkins, Henry (2003). Introduction: Toward an Aesthetics of Transition, pp. 1-16 in Thorburn, David & Jenkins, Henry (eds.), *Rethinking Media Change: The Aesthetics of Transition*. Cambridge, Mass: MIT Press.
- Tidningsfilmningskommittén (1976: 14). *Dagspress på mikrofilm: Filmning – arkivering – läsning*. Stockholm.
- Widholm, Christian (2011). Från massmedium till mikrosurrogat Några reflektioner kring den mikrofilmade dagstidningen som källa. *Scandia*, 77(2): 155-170.
- Yang, Tze-I; Andrew J. Torget & Rada Mihalcea (2011). Topic Modeling of Historical Newspapers, *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 96-104.
- Östling, Robert (2013). Stagger: An open-source part of speech tagger for Swedish. *Northern European Journal of Language Technology*, 3: 1-18.

Johan Jarlbrink är docent i mediehistoria och lektor i medie- och kommunikationsvetenskap vid Umeå universitet. Han forskar om 1800-talets tidningspress i alla dess skepnader: tidningen som informationsbärare i arkiv och klippböcker, som avfall och omslagspapper, som forskningsdata i digitala databaser.

Pelle Snickars är professor i medie- och kommunikationsvetenskap med inriktning mot digitala humaniora vid Umeå universitet. Han har bedrivit forskning om såväl äldre som nyare medier. För närvarande leder han två forskningsprojekt: om Spotify och strömmande medier, samt om digitala gestaltningar av samlingar på Tekniska museet.

Cristian Colliander är doktor i biblioteks- och informationsvetenskap och arbetar vid Umeå universitet som forskare på sociologiska institutionen samt som bibliometriker på universitetsbibliotek.

Jonas Andersson Schwarz & Johan Hammarlund

Kontextförlust och kontextkollaps

Metodproblem vid innehållsanalys av sociala medier

Abstract

The article proposes that social media platforms enable large volumes of user-driven circulation of media content, and argues for a combination of qualitative and quantitative considerations when analysing data from such platforms. Issues of context are vital; context must be understood both qualitatively (cultural setting) and quantitatively (statistical reference points for comparison). The authors emphasise that the possibilities of 'big data' should not tilt analyses so that sensitivities to subtler meanings are lost. By examining a recent research project of our own, examples are given of how topological network analysis can be successfully combined with close readings of strategically selected parts of the data and how, by doing so, context shifts can be identified that increase the reliability of the analysis. Consequently, it is recommended that mere number crunching is not enough, and that questions of 'how,' 'why,' and 'whether' are required in order to understand the phenomena in their societal settings.

Keywords: Twitter, data analysis, epistemology, knowledge context, competitive intelligence

Baserat på erfarenheter från ett samarbete mellan medieforskaren Jonas Andersson Schwarz och omvärldsbevaknings- och medieanalysföretagen M-Brain och Retriever vill vi här lyfta fram ett antal metodfrågor som vi ser som centrala inom databaserade metoder, både vad gäller strikt vetenskapliga tillämpningar och branschpraxis. Vi tar vår utgångspunkt i egen tillämpad forskning där vi kombinerat analys av relationer och flöden i sociala nätverk med analys av ostrukturerad text i ett projekt finansierat av Internetfonden år 2014-2015: "Vad gör en politisk utsaga framgångsrik: den användardrivna kommunikationens villkor" (Andersson Schwarz m.fl. 2015). Projektet undersökte olika aspekter av hur politiskt kodad kommunikation cirkulerar på svenskspråkiga Twitter samt hur press och sociala medier samspelar. Vi vill därmed bidra både till en forskningsmässig metodmedvet-

enhet och till att belysa omvärldsanalyssektorns mer praktiska tillvägagångssätt.

En tidig insikt under arbetet var nödvändigheten av kontext – och att kontext, för bästa resultat, bör förstås både kvalitativt och kvantitativt. Kvalitativt kan kontext förstås som *kulturellt sammanhang* men kontext måste också förstås som *referenspunkter för kvantitativa jämförelser*. Denna senare aspekt kan ibland glömmas bort, dels i renodlat kvalitativa studier där frågor om representativitet helt väljs bort, dels när forskare inom mer kvantitativt orienterad forskning upprättar mätvärden som i själva verket är mycket svåra att överhuvudtaget dra slutsatser ifrån, då dessa värden inte är självförklarande utan måste sättas i relation till andra värden. Hur mycket är 'mycket'?

Framför allt ser vi en risk att analysen går miste om både kvalitativ och kvantitativ

kontextualisering när tiden är knapp och resurserna begränsade. Ett relaterat problem är att analysförsök ofta görs trots att datan egentligen är allt för knapphändig avseende kontext, eller rentav inte tillräckligt vederhäftig till att börja med: Dagens kommersiella internetbaserade plattformar för socialt utbyte och mediekonsumtion (vi åsyftar primärt de Alphabet- och Facebookägda plattformarna) brister i transparens beträffande datans skick och mätapparatusens tillförlitlighet. Forskare tillåts inte ta reda på huruvida resultaten är präglade av olika former av dold bias; detta är ett akut kunskapsproblem och berör därmed även regleringen av dessa plattformstjänster (Andersson Schwarz 2016).

Vi diskuterar följaktligen kontext i både kvalitativ och kvantitativ bemärkelse, och vi kommer att koppla diskussionen till två radikalt skilda förhållningssätt som vi har observerat, beträffande empirisk datainsamling och -analys: Vi kallar den första för *närläsning* – en anekdotisk, abduktiv *ad hoc*-analys där analytikern ofta befinner sig 'inuti' skraddarsydda realtidsflöden, precis som vardagliga medieanvändare gör. Här är en kritisk faktor den algoritmiska filtrering som riskerar att uppstå, då varje användare är kringskuren av de sätt på vilka 'frontend'-gränssnitten är utformade, ett fenomen som populärt kallats för 'filterbubblor' (Pariser 2011). Dagens kraftfulla specialistsökverktyg gör det visserligen möjligt att spräcka dessa bubblor men de kan också i värsta fall göra det allt för enkelt att hitta utsagor som stöder ens hypotes eller föreställningar. Följaktligen krävs en hög grad av metodologisk medvetenhet och självreflektion för att man inte själv ska skapa en 'manuell' motsvarighet till de algoritmiska filterbubblorna. Det andra förhållningssättet kallar vi *fjärrläsning* – en sammanfattande, induktiv *post hoc*-analys, där analytikerns uppgift blir att uppskatta datamängder genom att retroaktivt söka eller på annat sätt erhålla urval ur infrastrukturens 'backend', det vill säga de underliggande databaserna, och där data-mängderna är för stora för manuell innehållsanalys så att analytikern således måste förlita sig på datadrivna, automatiserade bryt-

ningar av materialet (exempelvis algoritmisk sortering, regressionsanalys, 'deep learning'). Fjärrläsning ska inte ses som synonymt till det inom digital humaniora så vanliga begreppet distansläsning ('distant reading'), intimt förknippat med Franco Moretti (2013), då vi här även inkluderar mindre ambitiösa kvantitativa ansatser. Vi menar att styrkan i dessa nya metoder, i en akademisk kontext, framförallt uppstår när de kombineras med andra, mer konventionella metoder då triangulering kan generera både kontext och representativitet.

Precis som Lewis m.fl. (2013) och Gaffney & Puschmann (2014) förordar vi en pragmatisk ansats, där datalogiska och manuellt tolkande metoder kombineras; en 'kvali-kvantitativ' ansats (Latour m.fl. 2012). Genom att omsorgsfullt kombinera tekniker går det att bibehålla fördelarna med konventionell innehållsanalys (systematisk stringens och kontextuell känslighet) samtidigt som storskalig kapacitet kan erhållas genom stora datavolymer och beräkningsmetoder baserade på algoritmer (vilket kan synliggöra relationer och samvariationer, samt påvisa olika aggregationers relativa storlekar).

Vi stöder oss på ett tentativt schema över omvärldsanalysens förändrade historiska karaktär, föreslagit av Boullier (2015), samt forskningsartiklar inom fältet kritiska sociala mediastudier. Vårt teoretiska ramverk är socialkonstruktivistiskt men med stor tonvikt även på den ickemänskliga agensen; det vill säga, en cybernetiskt influerad konstruktivism där begrepp som infrastruktur och system är lika viktiga för att förstå begreppet kontext som sociokulturella kategorier. Vi har inte utrymme att ingående diskutera forskningsetik och personlig integritet men det bör framgå att en kontextkänslig analysmodell per definition aktualiserar sådana aspekter.

Varför Twitter?

Den enda av de stora, populära plattformstjänsterna som ger någorlunda goda möjligheter till översikt är Twitter, då den absoluta majoriteten av dess data är (åtminstone teoretiskt sett) offentlig, samt att den begränsade läng-

den på varje statusuppdatering (140 tecken) borgar för effektiv informationsinhämtning och jämförande analys (Tao m.fl. 2014).

Twitter är en arena där opinionsbildare, journalister och medborgare möts och interagerar med varandra, liksom även institutionella användare (organisationer, partier, företag). Det rör sig både om offentliga samtal och diskussioner inom små grupper. Som kommunikationsplattform förefaller Twitter därför ha ett påfallande samhällsinflytande; plattformen fungerar på många sätt i symbios med journalistiska massmedier (Broersma & Graham 2012, Lasorsa m.fl. 2012, Paulussen & Harder 2014). I en stor studie av användarbeteenden observerades att Twitter är mer att betrakta som en medieplattform än ett socialt nätverk; 85 procent av inläggen var nyhetsrelaterade (Kwak m.fl. 2010). Samtidigt är Twitter helt annorlunda än traditionella medier eftersom användarna agerar 'sociala sensorer' (Sakaki m.fl. 2010) som kan ge förstahandsinformation om olika aspekter av nyhetshändelser. Mycket av den kommunikation som sker på Twitter och som etiketteras som politisk kommer dock från aktörer som redan är involverade professionellt i politiken (Larsson 2014, 2015, 2016).

Twitters roll som agendasettande arena, i kombination med dess öppenhet – och den relativt lättillgängliga sociala mediedata detta medför – gör att plattformen är en tämligen välanvänd datakälla inte bara bland journalister, men även i den internationella medieforskningen (Lewis m.fl. 2013, Weller m.fl. 2014), trots att det är långt ifrån ett populärt medium rent demografiskt. Uppskattningsvis sex procent av de svenska internetanvändarna använder tjänsten dagligen (Findahl & Davidsson 2015) – samtidigt visar vår forskning att Twitter tveklöst är den sociala medieplattform som det hänvisas oftast till i dagstidningarna (Andersson Schwarz m.fl. 2015: 20).

Twittercensus, en årlig kartläggning av det svenskspråkiga Twitter (Brynnolf 2015), visualiserar dess olika nätverk i form av följarelationer. Dessa nätverk är samtidigt överlappande med mer eller mindre permanenta intressenätverk som formas kring olika fe-

nomen och händelser. De nätverk som lyftes fram i vår studie var objekt som i viss mening skapades ur dedikerade sökningar på vissa termer eller hashtaggar. Skärningar som görs av 'nätverk inom nätverken' av det här slaget går att använda som populationer i sin egen rätt, vilka de med tillgång till data och verktyg kan hänvisa till, rentav använda retoriskt som led i argumentation ('frågan X debatteras just nu flitigt på Twitter'). Huruvida de observerade deltagarna själva anser sig ingå i nätverket är ofta oklart, men en högst angelägen fråga. Marwick & Boyd (2011) dokumenterar olika avsändarstrategier som användarna har angående detta. Samtidigt bör också läsaren ha en strategi för att hantera att den information denne möter i sitt flöde (eller *dashboard*, se nedan) kan vara skapat utifrån en helt annan kontext än den användaren själv verkar inom och därför kan misstolkas.

Relevans för praktisk omvärldsanalys och teoretisk forskning

Nätverksplattformar såsom Twitter och Facebook gör olika former av närläsning och fjärrläsning av stora mängder information förrädiskt enkla. Enkla, eftersom potentiellt stora mängder information kan hämtas in med ett fåtal knapptryckningar. Förrädiskt, eftersom processerna och algoritmerna som styr vilken information som ges tillgång till är svårgenomträngliga; de saknar vad Diakopoulos (2014) kallar 'algorithmic accountability'. Medel i form av datatillgång, digitala verktyg såsom dashboards (kontrollpaneler för sökningar i datafloden), samt kompetensen att använda dem möjliggör strukturerade sökningar som genererar allt ifrån enstaka citat till stora mängder data om vissa gruppers beteenden och kommunikation. Artefakter av detta slag kan sedan omvandlas till retoriskt övertygande visualiseringar, ofta utan tillräcklig käll- och metodredovisning.

Två godtyckliga element spelar in när sociala grupperingar av det här slaget observeras: För det första är grupperingar på nätet i regel 'självrekryterade' (självalda gemenskaper) eller resultat av att aktörer med specifika intres-

sen spontant går samman. Därmed är de sällan *a priori* representativa för större populationer. För det andra innebär själva akten att välja ut en gruppering att analytikern gör en avgränsning och därmed 'skär ut' en subgrupp ur en större population. Därtill kommer frågan om huruvida datan är vederhäftig, något som blir svårt att veta på grund av avsaknaden av transparens från plattformsföretagen. Många har sagt vittnat om en så kallad 'filterbubbla', att osynliga algoritmer styr viken information vi exponeras för (Pariser 2011). De dashboards som många omvärldsbevakningsföretag erbjuder kan verka som motmedel till denna. Men utan en hög självmedvetenhet om ens egna föreställningar riskerar analytikern att själv upprätta s.k. *confirmation bias*, i förlängningen en mental bubbla inom vilken de egna förutfattade teorierna bekräftas. Skruvar du tillräckligt länge på inställningarna i verktygen för sök, analys och visualisering så ökar chansen att du till slut får 'rätt' svar: 'hårda data' som stödjer din ståndpunkt.

Både närläsning och fjärrläsning av de här slagen möjliggör på många sätt 'nya' typer av information om mänskligt beteende – på ett enkelt och ofta kostnadseffektivt sätt. De utgör därför på många sätt en sorts tredje generationens kvantitativ metod inom samhällsforskningen (Boullier 2015, se nedan). En risk är dock att den kontext i vilken information skapats går förlorad. En sådan form av förlust är kontextkollaps (Marwick & boyd 2011), vilket åsyftar de mekanismer då olika användares föreställningsvärldar och tolkningar krockar, eftersom information idag kan röra sig mellan nätverk och grupper i svårförutsedda mönster.

Kontext blir i vår diskussion i hög utsträckning synonymt med nätverk – och *nätverksanalys* blir följaktligen ett sätt att (re)kontextualisera information som ryckts loss ur sitt ursprungliga sammanhang. Det är värt att notera att inte bara individer agerar i nätverken. Mängder av institutionella aktörer är också aktiva, automatiserade konton (s.k. *bots*) därtill, och många individer tycks växla mellan att agera åt/som sig själva och åt/som institutioner. Diskurs och (digital) praktik inom de medierade sociala nätverken samverkar

vidare med de fysiska sociala nätverken i en ömsesidig växelverkan. Till exempel kunde vi, i en studie genomförd i samarbete med Institutet för mediestudier (Andersson Schwarz & Hammarlund 2016), se hur den svenska diskursen kring migration i såväl sociala som redaktionella medier (dessa är centrala aktörer även i de sociala nätverken) påverkades när relationerna i nätverken förändrades, exempelvis när drunknande flyktingar uppmärksammades i massmedia. Nätverk är, i denna bemärkelse, samtidigt både medierade och köttsliga, precis som vår tillvaro idag. De är överlappande, ständigt föränderliga och samverkar med andra nätverk i en stor svåröverskådlig väv. För att parafrasera Herakleitos så är det ej möjligt att två gånger nedstiga i samma nätverk – följaktligen kommer vår kunskap alltid att vara bristfällig och i någon mån föråldrad. De digitala nätverken lämnar visserligen hela tiden spår (mer eller mindre öppna) och dessa spår kan analyseras. Betyder det att vi genom att studera en aspekt av den nätverkande människan, exempelvis spåren av medierad kommunikation på en viss arena, kan förstå helheten? Nej, men kanske att vi har fått några nya pusselbitar.

Datatillgång

Det finns dilemman beträffande datatillgång och kontroll som möter alla som på något vis interagerar med digitala medieplattformar (Puschmann & Burgess 2014). Dessa frågor är själva grundbulten för alla andra resonemang (empiriska, epistemologiska, regleringstekniska) vi kan föra om dessa plattformar. Som boyd & Crawford (2012) och även Zelenkauskaitis & Bucy (2016) har konstaterat är det hela ytterst en fråga om makt.

Sociala medieplattformar skiljer sig åt i graden av dataåtkomst de är villiga att ge, och det är en myt att forskare fritt kan samla användargenererad data från dessa plattformar (Gaffney & Puschmann 2014, Tao m.fl. 2014, Vis 2013). Vad gäller Facebook kan majoriteten av användarkonton uppskattas vara slutna/privata i olika utsträckning (exakt hur många konton och vilken grad av öppenhet behand-

lar företaget Facebook dock som en affärs-hemlighet). Twitter, däremot, kännetecknas av en strukturell öppenhet som härrör ur en tydlig norm bland användarna att låta sina konton vara offentliga. Även Twitter har dock gradvis begränsat den fullskaliga tillgången till dess databibliotek (s.k. 'firehose access') till dataleverantören Gnip som köptes av Twitter år 2014. Även Facebook har exklusivt knutit sin kommersiella datatillgång till det London-baserade företaget Datasift. Genom s.k. applikationsprogrammeringsgränssnitt (API:er) kan begränsad tillgång till Twitters data visserligen erhållas av kunniga användare, men full tillgång till företagets data sker endast på kommersiell basis, och då genom Gnips *de facto* monopol.

Den typ av slutenhet som detta leder till drabbar framförallt tredjepartsaktörer som vill få en överblick i den sociala medieekologin: omvärldsbevakningsföretag samt akademiska forskare. För att få mer heltäckande tillgång tvingas dessa tredjepartsaktörer således att tillgå ett grossistled av dataleverantörer som för närvarande domineras av plattformsföretagen själva, en trend som kan komma att ytterligare fördjupas.

Vår undersökning

Den studie som utgör basen för vårt empiriska material (Andersson Schwarz m.fl. 2015) inleddes 2014 med huvudsyftet att få en aktuell lägesbild av vad som cirkulerar inom Twittersfären som sådan, och mellan sociala och redaktionella medier. I projektet har vi huvudsakligen använt olika kvantitativa metoder (induktiv nätverksanalys samt konventionell innehållsanalys) för att hitta mönster och bättre kunna kategorisera det som återcirkuleras inom politisk debatt och opinionsbildning, samt illustrera några viktiga aspekter av samspelet mellan Twitter och redaktionella medier.

Projektet bestod av tre separata kvantitativa studier samt den fallstudie av nyheten om Pippi Långstrump som beskrivs nedan. Projektets egenart är att det kombinerar en induktiv nätverksanalys (d.v.s. en graf som

visar det topologiska mönster som användares interaktioner genererar) med konventionell innehållsanalys (d.v.s. manuell kodning av ett stickprov). Den nödvändiga datatillgången möjliggjordes genom samarbete med omvärldsbevakningsföretagen Retriever och M-Brain. Genom detta samarbete kunde vi samla in ett råmaterial av 109 000 Twitter-inlägg (inklusive retweets) och ett tusental tidningsartiklar, vilket i analys- och visualiseringsarbetet reducerades till knappt 60 000 relevanta tweets och retweets samt drygt 400 tidningsartiklar. Studien går att upprepa, med förbehållet att vi avsiktligt studerade Twitter under 'supervalåret' 2014, en politiskt turbulent period då det politiska landskapet i Sverige polariserades (vederlagt i bland annat Bergström m.fl. 2015). Detta tycks ha cementserats även på det politiska Twitter, där beteendet att retweeta (återcirkulera) utsagor följer uppenbart grupp-dynamiska mönster.

En del av vår studie som illustrerar många av de fenomen vi vill lyfta fram var fallstudien där vi analyserade reaktionerna på nyheten att tv-serien *Här kommer Pippi Långstrump* skulle moderniseras (Andersson Schwarz 2015). Kortfattat konstaterade vi att olika kluster (användare av hashtaggen #svpol respektive övriga användare på svenskspråkiga Twitter) reagerade olika på nyheten, samt att deras kommunikativa beteenden i klustret/nätverket också skiljde sig åt. En högljudd minoritet motsatte sig redigeringen och en liten grupp inom denna minoritet var väldigt aktiva på sociala medier med delning av inlägg. Vi analyserade de 50 mest spridda inlägg som nämnde Pippi den dagen och var avsiktligt markerade med #svpol. Av dessa uttryckte två tredjedelar sentiment; samtliga inlägg som gjorde detta uttryckte motstånd mot moderniseringen. Omvänt fann vi att bland de 50 mest spridda inläggen i den mycket större samlingen inlägg ej markerade med #svpol var sentimentet överväldigande positivt till moderniseringen. Dock var användarna själva sannolikt i väldigt varierande grad av uppfattningen att de över huvud taget tillhörde ett särskilt nätverk eller kluster (utöver Twitter i sig). #svpol är visserligen en etablerad hashtag (som har en

påvisad överrepresentation av högerextrema/främlingsfientliga, se Larsson 2014) men många av de övriga som skrev till sina följare/vänner om Pippidebatten skulle sannolikt inte anse att de utgjorde en part av någon tänkt gruppering av 'motståndare' till moderniseringen av Pippi i den öppna debatten utan den typen av kategorisering skapas endast *a posteriori* – i vår analys, baserad på deras kommunikation på Twitter i just det observerade ögonblicket. Hade vi enbart valt att använda hela Twitter som population så hade vi här missat en central slutsats, splittringen mellan de olika kluster/gemenskaper som förekommer inom populationen. Omvänt hade vi med en ofullständig analys av #svpol-användarna kunnat konstatera att denna grupp motsatte sig redigeringen med en relativt hög volym av inlägg i förhållande till antalet individer/konton. Värdet i analysen kommer först när vi särskiljer dessa nätverk samt tittar på deras karakteristika – mer exakt, hur #svpol avviker. Då synliggörs plötsligt ett nätverk inom nätverket med en hög grad av anonyma konton, en markant avvikande åsikt i sakfrågan (enligt de utsagor som görs), samt ett kommunikativt beteende där en liten grupp är väldigt aktiva med att retweeta inlägg och därigenom blåser upp gruppens delningssiffror.

Likaledes, i vår större rapport, parallellt med Pippistudien, använde vi, för att göra urvalet som låg till grund för analysen, dels sökordsbaserade funktioner i insamlingsverktyget (n=109 000), dels klustringsfunktioner i visualiseringsverktyget som visar hur retweetandet hänger samman (här blev drygt en tredjedel av urvalet relevant, n=37 959). I analys av det här slaget skapas alltså kategorier som syftar till att situera utsagor som annars hade förblivit lösryckta. Att därefter bädda in funna excerpter i en ny miljö, vår egen forskningsrapport, innebär i sin tur en kontextväxling som de observerade deltagarna svårligen kan ha förväntat sig.

En liknande typ av kontextväxling illustreras också av den kedja i vilken informationen om redigeringen av Pippi rörde sig (utförligare beskrivet i Andersson Schwarz 2015). Den ursprungliga pressreleasen från SVT fångades

upp av Aftonbladet som skrev en nyhet om detta. Denna nyhet fångades i sin tur upp av en användare på Facebook (som opponerade sig mot redigeringen), som i sin tur hängdes ut av en användare på Twitter (genom att denne ironiserade över en skärmbild av Facebookinlägget). När denna tweet sedan retweetades av en av landets mest inflytelserika opinionsbildare (som själv är kopplad till public service) fullbordades en referenskedja av olika budskap, där räckvidden eskalerades närmast exponentiellt. Kedjor av det här slaget överbryggar plattformar, kontexter, normer och beteenden, samt även lagrum.

Sådana kedjor där information hoppar mellan kontexter och nätverk är illustrativa för dagens komplexa kommunikationslandskap där individers 'privata' beteenden i vissa lägen allt mer liknar organisationers beteenden. Individuella användare på Twitter och andra plattformar kan idag använda en rad olika tekniker och redskap för att strukturerat samla in information ämnen och personer eller organisationer de finner intressanta. Genom att använda Twitters egna 'backend'-funktion *Twitter Analytics* kan användare analysera sin egna kommunikativa effektivitet. När tekniker av det här slaget är allmänt tillgängliga har vi en situation där individer med tekniska medel automatiskt och genom sökfunktionalitet kan bevaka sina intresseområden, där information lätt kan flyttas från kontext till kontext, och där relativt nya hybridtillämpningar som 'citizen journalism' och 'DIY-intelligence' (och i viss mån datajournalistik) möjliggörs. Den spontana individuella omvärldsbevakningspraktik som Hamrefors (1999) beskrivit i termer av 'spontaneous environmental scanning' har idag alltså kompletterats med en mer strukturerad och delvis automatiserad praktik. Teknikutvecklingen gör denna praktik tillgänglig för en stor majoritet av internetanvändarna, även om andelen som *de facto* använder sig av dessa metoder troligen är relativt liten. En liknande, men omvänd, uppluckring ses i den företagsekonomiska kunskapsbildningen om organisering i värdenätverk (Christensen 1997) som beskriver hur företag får tillgång till kunskap via relationsbyggande i organisa-

tionsöverskridande nätverk. Organisationer kan här ses växla mellan olika roller i samspel med andra organisationer (Hamrefors 2014), och det vore värdefullt att vidare utforska hur dessa tendenser förhåller sig till varandra.

Diskussion

Sociala medier som empiriska objekt

Kvantitativ innehållsanalys kan definieras som en forskningsmetod som syftar till 'objektiva, systematiska och kvantitativa beskrivningar av manifest [direkt observerbart] kommunikativt innehåll' (Berelson 1952: 18, vår översättning). Med *manifest* menas olika texters 'uppenbara' innebörd, medan antydda eller underförstådda innebörder hänförs till den *latenta* nivån. Datorassisterade metoder kan med fördel tillämpas på manifest innehåll, då kodningen där ofta handlar om att tillämpa uppenbara logiska instruktioner (det vill säga, algoritmer), medan mänsklig arbetskraft alltså är oundgänglig när det gäller latenta dimensioner, då detta handlar om associativ tolkning (Sjøvaag & Stavelin 2012: 219).

I vårt projekt var t.ex. bevekelsegrunderna för twittrarna i fråga ej manifest. Vi kunde tydligt se omfattningen av den spridning som hade skett, men vi kunde inte med säkerhet veta varför upphovspersonerna bakom inläggen hade valt att skriva just dessa inlägg, och vi kunde inte heller med säkerhet veta varför de som hade retweetat dessa inlägg gjorde det.

Trots tjuvningen med datorbaserade beräkningsmetoder är mänsklig tolkning alltså fortfarande oundgänglig för innehållsanalys; både när det kommer till förförståelse och förmågan att inse sammanhang ('contextual sensitivity'; se Lewis m.fl. 2013: 48) och deras placering i den större, omgivande medieekologin. Medvetenheten och känslan för kontexten blir extra viktig i studier av *big data*, då de hisnande datavolymer riskerar att sporra forskare till att leta efter mönster utan att se de subtila dimensioner som finns inbäddade i medietexternas tillblivelse och cirkulation, samt att lockelsen ofta är stor att få de olika brytningar som görs av datan att passa färdiga

teoretiska modeller (boyd & Crawford 2012: 671).

Tydligt är att nätverkssamhället inte ger upphov till en jämn, ensartad population utan att mängder av olika, delvis överlappande klusterformationer och grupperingar uppstår, ofta med radikalt ojämlika styrkeförhållanden vad gäller enskilda deltagares inflytande. Huruvida de sociala gemenskaper som möjliggörs på nätet bidrar till starkare skillnader mellan grupper är fortfarande en i hög grad obesvarad fråga (Wadbring 2016). Forskning indikerar att olika åsiktskluster uppstår (Garimella m.fl. 2015). Man kan prata om fragmentisering (Bergström m.fl. 2015, Strömbäck 2015) eller rentav enklavisering (Sunstein 2007), men mycket beror på vad man lägger i begreppen. Hur stor är överlappningen mellan olika gruppformer? I vilken grad förekommer gruppöverskridande åsiktsutbyten? Även här blir kontexten avgörande för bilden av den större samhällsordning i vilken handlandet äger rum.

Det är en kvalitativ ny situation att självvalda åsikts- och intressegemenskaper kompletteras av algoritmisk filtrering där flöden moduleras (Andersson Schwarz 2015) genom infrastrukturella popularitets- och personaliseringfilter som varken är transparenta eller självvalda (Pariser 2011, Webster 2014, Zuiderveen Borgesius m.fl. 2016). I en pionjärstudie av personaliseringen av nyhetskonsumtionen fann Flaxman m.fl. (2016) motsägelserfulla resultat. Samtidigt som polariserande effekter konstaterades tycktes sociala medier bidra till gruppöverskridande exponering. Nischat medieinnehåll konsumeras normalt i tillägg till en mainstreamdiet (Bimber & Davis 2003, Trilling & Schoenbach, 2015). Kruket är att politik helt kan väljas bort (Flaxman m.fl. 2016, Zuiderveen Borgesius m.fl. 2016) med följd att den mer uttalat partiska mediedebatt som kvarstår kännetecknas av ömsesidigt repellerande grupper i högre grad än vad som annars vore fallet. De slutsatser som går att dra om huruvida medielandskapet verkligen förstärker fragmentisering är dock beroende av metod, val av fokus och val av kriterier/gränsvärden (Webster & Ksiazek 2012).

Trots heterogena populationer och svårjämförda grupper finns ett sätt att skapa åtminstone teoretiska jämförelsepunkter: att följa en och samma parameter konsekvent över tid. Oavsett mätvärdets eventuella övriga brister uppstår en jämförande tidsserie för mätvärdet i fråga. En annan jämförande dimension som ofta uppstår är möjligheten att jämföra likvärdiga artefakter. Inom kvantitativ medieanalys kan det till exempel handla om graden av förekomster av ett och samma sökord i olika miljöer. Därutöver, genom att göra kompletterande urval, till exempel av helt andra men likväl jämförbara sökord, går det att sätta förekomsterna i ett kvantitativt sammanhang där de uppmätta mängder man valt att fokusera på jämförs med andra uppmätta mängder under samma tidsperiod.

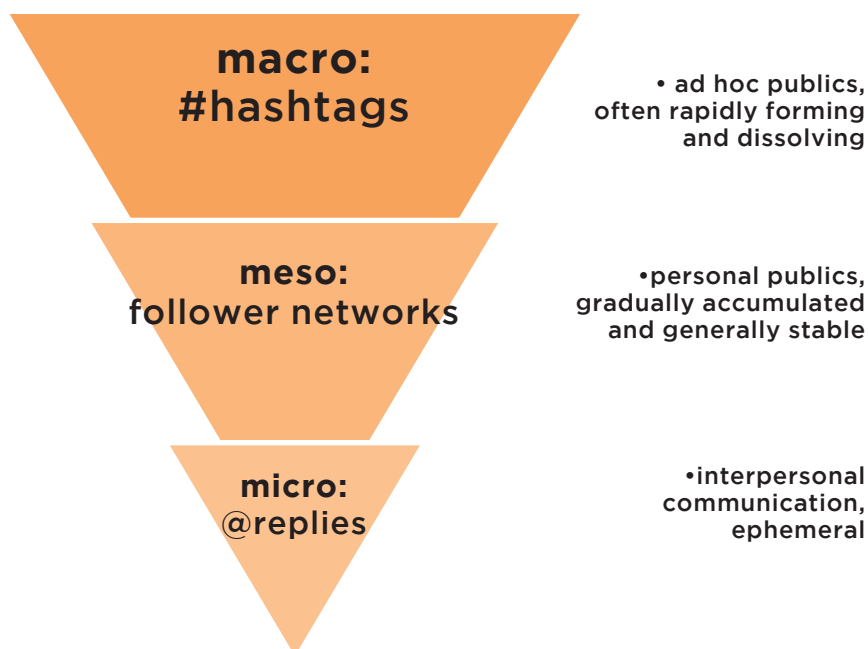
Detta förklarar våra egna metodval, då syftet var att upprätta en form av triangulering för att kunna belysa samma fenomen ur olika synvinklar, något vi generellt vill propagera för. Ju fler jämförande dimensioner som upprättas, desto rikare kontext och bättre validitet.

I studiet av Twitter kan tonvikt läggas på olika nivåer av närhet, där skilda sociala prak-

tiker hamnar i fokus beroende på vilken analysnivå som väljs. Helt olika optik erfordras; i vår studie opererade den automatiserade nätverksklusteranalysen på makronivån medan vår manuella innehållsanalys av enskilda tweets opererade på mikronivån. Våra kluster skapades dock utifrån retweetande, vilket Bruns och Moe (figur 1) beskriver som mer flyktig kommunikation på mikronivå (möjlig mesonivå). Genom att samla in stora volymer över tid skapar vi dock en aggregerad makrostruktur där vi kan se stabila mönster vad gäller såväl beteende som relationer.

Twitter och kontextkollaps

Kontextkollaps åsyftar olika konfliktartade skeenden beträffande de sociala normer som det sociala handlandet förväntas bygga på (Wesch 2009), inte minst beträffande människors förståelse av sig själva i relation till vad Mead (1934) en gång kallade 'den generaliserade andre' (förmågan att föreställa sig sitt beteende sett utifrån). När mänsklig samvaro determineras av internetplattformar som Facebook tenderar en rad sociala sammanhang, vilka tidigare varit rumsligt åtskilda, att kol-



Figur 1. Olika nivåer av kommunikativa utrymmen på Twitter

Källa: Bruns & Moe 2014: 20 (Faksimil)

lapsas till ett enda heterogent utrymme (Hogan 2012).

I onlinemiljöer försvåras möjligheterna att skilja på de sammanhang i vilka handlandet sker; digitala delpersonligheter tillåts smälta samman och/eller överlappa på andra sätt än offline. Aktörer i semi-offentliga sociala nätverk tvingas därför hantera möjligheten att deras publik rymmer ohanterbart många olika 'generaliserade andra'. Agerande som en själv anser självklart kanske ej är det i vissa andras ögon. Twitter tillåter dock en friare presentation av identitet än Facebook och Google; exempelvis kan användare ha multipla konton och agera i enlighet med de många olika delpersonligheter som alla människor har. Twitterflöden är i hög grad intressebaserade snarare än identitetsbaserade (Marwick & boyd 2011); många användare tycks ha en funktionell inställning där Twitter används för omvärldsbevakning snarare än expressivt som en skildring av ens egna vardag, även om det senare också förekommer på Twitter.

Kontext kan förstås spatialt; en plats som avgränsas i tid och rum med igenkännbara roller och sociala normer (Hogan 2012: 300). Miljöpsykologer talar om beteenderegioner (Barker 1968), medan Goffman (1959) använder scenografiska metaforer ('frontstage', 'backstage') och Feld (1981) etablerar begreppet 'fokus' inom social nätverksanalys; sociala, psykologiska eller fysiska företeelser kring vilka nätverksaktörer kan organisera gemensamma aktiviteter (ibid.: 1025). Sådana fokus begränsar i regel handlandet, men ju fler aktörer som delar samma fokus desto friare är det normativa utrymmet (Melbeck 1996: 261). Att kontexter delvis överlappar har länge varit ett sociologiskt faktum (se exv. Simmel 1922). Den rent fysiska platsen för cirkulation och reproduktion av mediediskurser påverkar förväntningarna på genre, modalitet och sanningsanspråk (Fairclough 2012). Materiellt villkorade förväntningar är förstås också sociokulturellt villkorade; de kollektivt delade koderna för läsningen och förståelsen av medietexter är starkt normativa och ideologiskt präglade (Hall 1980). Notera att detta även gäller analytikern, inte minst beträffande

de rådande normer om information som varuform (Bolin & Andersson Schwarz 2015) som påverkar läsningen av social mediedata, dashboards, infographics och visualiseringar.

Bounegru (2009) menar att digitala sociala nätverkstjänster som Twitter tycks ge upphov till en form av 'sekundär oralitet' (Ong 1982) typisk för sammansmältningen av skriven, muntlig och elektronisk kultur, där deltagarna ständigt påminns om att de verkar i ett i någon mening offentligt sammanhang, oerhört mycket större än det som är aktuellt för primär oral samvaro (ibid.: 136). Här premieras deltagande, främjande av en gemensam känsla, ett fokus på nuet, samt även förhållandevis tillrättalagda, förutsägbara, rentav stereotypa uttrycksformer. Men i och med att mediets diskurser ständigt lagras i beständiga medier – lämnar spår – blir onlinekulturen jämförelsevis mer försiktig och självmedveten i sitt uttryck än det talade ordet. Kommunikation som lämnar spår av det här slaget har därtill egenskapen att den synliggör tydliga demarkationer av gruppstillhörighet. Jämfört med muntlig kommunikation kan skriven kommunikation därigenom framstå som entydig och kategorisk, inte minst i moral- och politikfrågor.

Länge byggde Twitter på en relativ konsensus beträffande de gemensamt delade koderna och normerna för deltagandet, menar Meyer (2015, som i sin tur hänvisar till Stewart 2014). Denna gemensamma praxis byggde i sin tur på att både läsandet och skrivandet (konsumtionen och produktionen) var inflikade i varandra, något som bidrog till att stärka den ömsesidiga förståelsen av kontexterna för skrivandet. Twitter kan skämtsamt ses som en gigantisk självrekryterad webbpanel med endast en fråga, vars tonalitet låter oss observera tre generationer av Twitter (figur 2).

Utöver de grundläggande incitamenten för att över huvud taget skriva något på Twitter – vilket determineras både av form/design och av normer i den specifika gemenskap inom vilken användaren verkar – kan vi rada upp en mängd förutsättningar och försöka kategorisera dem, även dessa med förbehåll för att de diskursiva villkoren är tätt sammanlänkade

Twitter I (2006-2009)	kännetecknas av frågan "What are you doing?" vilket appellerar till urban livsstil, och vardagliga statusuppdateringar.
Twitter II (2009-2012)	kännetecknas av frågan "What's happening?" och förskjuter tonvikten till världshändelser som val, katastrofer och revolutioner. Här blev mätverktyget "trending topics" av mycket stort intresse.
Twitter III (2012-)	kännetecknas av uppmaningen "Compose new tweet" vilket avspeglar hur plattformen blivit ett generiskt verktyg för omvärldsbevakning (valrörelser, börskurser, celebriteter). Talande för denna utveckling är funktionen Twitter Analytics – ett dashboard integrerat direkt med Twitter som användarna själva kan tillgå.

Figur 2. Tre generationer av Twitter

Källa: Rogers 2013 (vår översättning och sammanfattning)

- **mediets form**
 - begränsningen i längd (140 tecken)
 - hur kedjor av tweets fungerar (det är ej alltid tekniskt uppenbart att ett inlägg utgör svar på ett annat inlägg eller är del av en längre "tråd" av inlägg)
- **villkoren för skrivandet som handling**
 - tidsliga villkor (dels att formulera sig kort, dels att skriva sitt inlägg medan debatten är aktuell)
 - skribentens upplevda behov av att vara slagfärdig för att bli retweetad och nå ut
- **kollektiva normer i gemenskapen av användare**
 - förväntningen att Twitter är ett jämförelsevis "smart" medium, för insatta användare (både tekniskt kunniga och intellektuellt kapabla att förstå ironier och andemeningar), något som under Twitters historia faktiskt rent konkret har uppmuntrats av plattformens diskursiva utformning:

En rutinmässig uppmaning till användarna vilken säger något om den förväntade interaktionen; en bokstavlig förväntning om "cleverness" hos användarna (skärmdump från 2012).

Figur 3. En typologi över Twitterspecifika kontextuella element

Not: Författarnas egna sammanställningar.

med varandra både formmässigt (det vill säga, tekniskt och språkligt) och normmässigt (socialt inom gruppen och beträffande bredare samhälleliga konventioner) (figur 3).

Om en användare har en förförståelse beträffande dessa villkor för skrivandet så påverkas rimligen dennes läsart av detta. Meyers och Stewarts resonemang är att denna delade praxis på senare år blivit allt mer ansträngd, främst på grund av Twitters expansion i antal användare. Denna expansion har inte bara gjort att enstaka kändisar får en närmast massmedial närvaro på Twitter, med miljontals följare, utan innebär också att Twitter är fragmentiserat (vilket Brynolf 2015 visar empiriskt) samt att det kommit att ske ett idogt hänvisande till Twitter inom massmedia, vilket vi visar empiriskt (Anders-

son Schwarz m.fl. 2015). Citat lyfts ofta ur sin sociala kontext på Twitter och sätts i en massmedial kontext. Enskilda uttalanden som egentligen är del av större konversationer tas ur sitt sammanhang och behandlas därefter av människor utanför samtalet som vore de entydiga politiska uttalanden. Det är inte bara twittrarens egna ideologiska motståndare som kan plocka billiga poänger på att ta slarvigt prat ur sitt sammanhang; även dennes egna supporters kan ta enskilda uttalanden ur sin kontext, till exempel för att visa på förmodad oärlighet eller illojalitet.

Utöver krockarna mellan konversionsbaserad förväntan och textbaserad tolkning sker alltså även olika former av kontextkollaps beträffande de avsedda mottagargrupperna för olika budskap, och de avsedda graderna av

formalitet som uttalandet är tänkt att läsas genom.

Slutsatser

Nya digitala verktyg kan låta analytikern dyka ner i dataflödet och med hjälp av närläsning hitta anekdoter som egentligen inte är det minsta representativa för den större popu-

lationen. Likafullt kan analytikern, efter att ha hämtat ut stora mängder ’backend’-data genom olika typer av digitala gränssnitt, göra fjärrläsningar som kan visa på samband och mönster som kan tyckas mycket egenartade i jämförelse med konventionell deskriptiv statistik.¹ Det blir uppenbart att läsningar i det Boullier (2015) identifierat som tredje generationens kvantitativa förklaringsmodeller

	Första generationen	Andra generationen	Tredje generationen
<i>Föreställning om det sociala</i>	Samhälle(n)	Opinion(er)	Vibration(er)
<i>Insamlingsapparat</i>	Folkräkning (census)	Enkäter (survey)	Spår (big data)
<i>Valideringsprincip</i>	Uttömlighet	Representativitet	Spårbarhet
<i>Kunskapsgenererande verktyg</i>	Register, förteckningar, bokföring	Publikmätningar, opinionsundersökningar	Sökmotorer, dashboards
<i>Social aktivitet i fokus för mätningen</i>	Förekomst	Uttryck	Engagemang
<i>Intressenter / finansörer</i>	Nationsstater	Massmedia	Varumärken
<i>Operativa aktörer</i>	Institut	Opinionsundersökningsföretag	Plattformsföretag
<i>Centrala intellektuella gestalter</i>	Émile Durkheim	Paul Lazarsfeld, George Gallup	Gabriel Tarde, Bruno Latour, Lev Manovich, Richard Rogers
<i>Huvudproblem som analyseras</i>	Arbetsfördelning och välfärdsstat (demografi)	Massmedialt inflytande och propaganda (publikmätning)	Spridning av kunskap (scientometri), samband av preferenser (sociometri, psykometri)
<i>Tekniska villkor</i>	Hållkortsmaskin / tabulator (Hollerith)	Informationsteknik (Turing, Shannon, von Neumann)	Internet, webb, mobila sensorer
<i>Semiotiska format</i>	Korstabeller, topografiska kartor	Normalkurvor, histogram, stapeldiagram, cirkeldiagram	Nätverksgrafer, infographics, tidslinjer
<i>Mätmetod</i>	Register över hela populationer (deskriptiv statistik)	Representativa stickprov/urval (induktiv statistik)	Topologi, frekvens, paretofördelade ej normalfördelade fenomen (Bayesiansk statistik)
<i>Tekniska kriterier för datakvalitet</i>	Relevans, exakthet, aktualitet, tillgänglighet, jämförbarhet, sammanhang	Konfidensintervall, probabilitet	Volym, variation, hastighet (big data)

Figur 4. Tre generationer av kvantitativa förklaringsmodeller

Källa: Boullier 2016: 13 (vår översättning och bearbetning)

inom samhällsvetenskapen (figur 4) ständigt behöver kompletteras med insikter genererade ur första och andra generationens.

Denna tänkta generationsväxling beträffande statistiska metoder och världsbilder återfinns även i omvärldsbevakningsbranschens diskurser: Demonstrationer av *dashboards* och *big data* åtföljs av lockande appeller till affärsdata vars tillblivelse är 'cool, easy, fast, fun', och där en sorts postmodern vändning tycks ha inträffat där närhet föredras framför distans, subjektivitet framför objektivitet och relevans framför representativitet (figur 5).

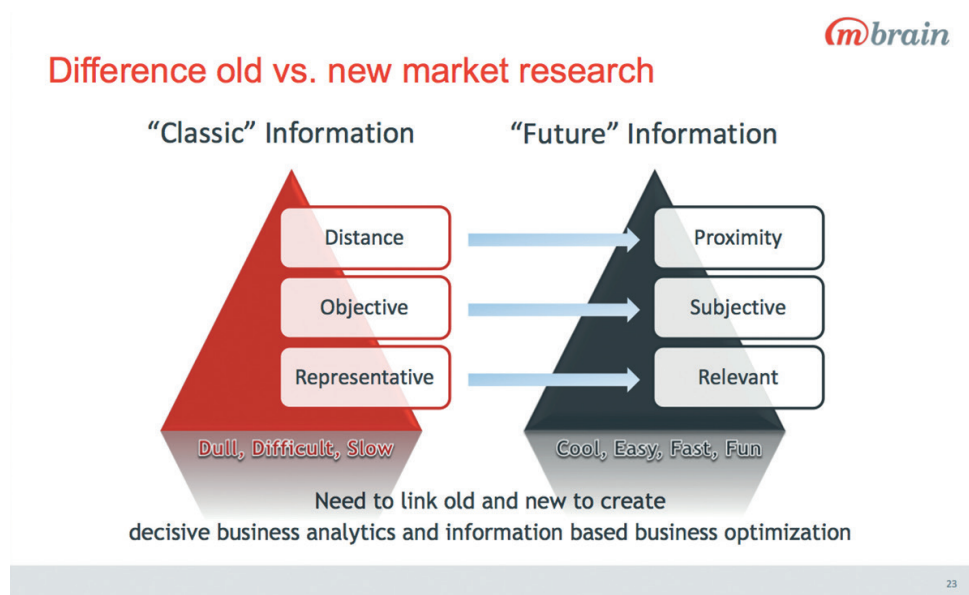
Även inom finansmarknader och politiska opinionsmätningar har sentimentanalys – m.a.o. realtidsindikationer av stämningar och känslolägen snarare än registrering av faktum eller riskmodellering baserad på statistiska normalfunktioner – kommit att dominera, något som yttrade sig i de felaktiga prediktionerna av 2016 års Brexit-omröstning i Storbritannien (Davies 2016).

Vi menar att det inte är så enkelt. Ordentlig kontextualisering innebär tyvärr arbetsmoment som faktiskt är 'dull, difficult, slow' och i regel kräver insiktsfulla analytiker. Ur

branschhänseende rör det sig snarare om förädling snarare än automatiserad vidarebefordran, en förflyttning uppåt i värdekedjan snarare än sidledes eller nedåt.

Analytikern bör ge akt på datans tillblivelse (de specifika villkor för användningen som gör de spår som lämnas särpräglade), dess tillgänglighet (huruvida man arbetar med kommersiellt införskaffad data och vad denna har för tänkbara begränsningar), samt huruvida ens slutsatser dras ur det manifest synbara eller ur latent (ej uppenbara) dimensioner i datan och vilka statistiska artefakter som skapas genom analytikerns egna avgränsningar. Den förförståelse som möjliggör tolkningar av latent innebär ingår i datans kulturella kontext – som är specifik för varje social medieteknologi.

Ser vi till den distinkta sociala topologi och kontextkollaps vi observerade på Twitter, är frekvens i många avseenden minst lika viktigt som räckvidd i sociala mediemiljöer, då hög frekvens kan ge intrycket att frågor är allstädes närvarande, och att enskilda frågor blir 'akuta' när 'alla' i ens eget flöde samtidigt lyfter dem. Därtill kan räckvidd vara olika bred. En aktör som når hundra olika användare inom olika grupper kan ha större influens än en



Figur 5. Förmodade skillnader mellan ‘gammal’ och ‘ny’ omvärldsbevakning

Källa: Utdrag ur presentationen ‘Social Media for Competitive and Market Intelligence’, Susanna Tirkkonen, Stockholm 18 november 2015 (Faksimil)

aktör som har tusen följare inom en väldigt snäv grupp. Nätverk med hög inbördes länkning är således mindre inflytelserika än vad de manifesta räckviddssiffrorna kan ge sken av. Därför rekommenderar vi att analytiker

inte ska stirra sig blinda på siffror för 'antal följare' utan snarare bör ge akt på grundfrågor såsom *huruvida*, *hur*, och *hur ofta* ett budskap cirkulerar i olika individers flöden och vilka som därmed nås.

Not

1. Läsningar av det här slaget kan exempelvis göras med hjälp av statistik- och visualiseringsverktyg såsom Python, R, Elasticsearch, NodeXL & Gephi.

Referenser

- Andersson Schwarz, Jonas (2015). Hur flöden module-ras: Vad som hamnar på agendan i sociala medier, i Carlsson, Eric m.fl. (red.) *Digital politik. Sociala medier, deltagande och engagemang*. Göteborg: Daidalos.
- Andersson Schwarz, Jonas (2016). Delningslogik och plattformisering, i Westlund, Oscar (red.) *Människorna, medierna & marknaden: Medieutredningens forskningsantologi om en demokrati i förändring*. SOU 2016:30. 133–164.
- Andersson Schwarz, Jonas; Hammarlund, Johan; di Grado, Stefan; Kjellberg, Magnus (2015). *Opinioner och offentligheter online: Slutrapport för forskningsprojektet 'Vad gör en politisk utsaga framgångsrik? Den användardrivna kommunikationens villkor'*. Stockholm: IIS.
- Andersson Schwarz, Jonas & Hammarlund, Johan (2016). Kända men inte erkända: Skildringen av Avpixlat och andra invandringsfientliga nätmedier i fem svenska dagstidningar 2013–2015, s. 172–188 i Truedson, Lars (red.) *Migrationen i medierna – men det får en väl inte prata om?* Stockholm: Institutet för mediastudier.
- Barker, Roger G. (1968). *Ecological Psychology: Concepts and Methods for Studying the Environment of Human Behavior*. Stanford, CA: Stanford University Press.
- Berelson, Bernard (1952). *Content analysis in communication research*. New York, NY: Free Press.
- Bergström, Annika; Johansson, Bengt; Oscarsson, Henrik & Oskarson, Maria (2015). Fragment, s. 11–36 i Annika Bergström m.fl. (red.) *Fragment*. Göteborgs universitet: SOM-institutet.
- Bimber, Bruce & Davis, Richard (2003). *Campaigning online: The Internet in U.S. Elections*. New York: Oxford University Press.
- Bolin, Göran & Andersson Schwarz, Jonas (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society* 2(2). doi:10.1177/2053951715608406
- Boullier, Dominique (2016). *Big data challenges for the social sciences: from society and opinion to replications*. Social Media Lab, EPFL Lausanne, maj. arXiv:1607.05034
- Bounegru, Liliana (2009). Secondary Orality in Microblogging: Orality versus literacy in the history of human consciousness. Blogginlägg, 20 november 2009. <http://lilianabounegru.org/2009/11/20/secondary-orality-in-microblogging/>
- boyd, danah & Crawford, Kate (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5): 662–679. doi:10.1080/1369118X.2012.678878
- Broersma, Marcel, & Graham, Todd (2012). Social media as beat: Tweets as a news source during the 2010 British and Dutch elections. *Journalism Practice* 6(3): 403–419.
- Bruns, Axel & Moe, Hallvard (2014). Structural Layers of Communication on Twitter, s. 15–28 i Weller, Katrin m.fl. (red.) *Twitter and Society*. New York: Peter Lang.
- Brynnolf, Hampus (2015). *Twittercensus*. <http://twittercensus.se/>
- Christensen, Clayton (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, MA: Harvard Business School Press.
- Davies, Will (2016). Thoughts on the sociology of Brexit. 24 juni. Political Economy Research Centre (PERC) blog, Goldsmiths, University of London. http://www.perc.org.uk/project_posts/thoughts-on-the-sociology-of-brexit/
- Diakopoulos, Nicholas (2014). Algorithmic accountability. Journalistic investigation of computational power structures. *Digital Journalism* 3(3): 398–415. doi:10.1080/21670811.2014.976411
- Fairclough, Norman (2012). Critical discourse analysis. Working paper. https://www.academia.edu/3791325/Critical_discourse_analysis_2012_
- Feld, Scott L. (1981). The Focused Organization of Social Ties. *American Journal of Sociology*, 86(5): 1015–1035.
- Findahl, Olle & Davidsson, Pamela (2015). *Svenskarna och internet 2015*. Stockholm: IIS.
- Flaxman, Seth; Goel, Sharad; Rao, Justin M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80: 298–320.
- Gaffney, Devin & Puschmann, Cornelius (2014) Data Collection on Twitter, s. 55–67 i Weller Katrin m.fl. (red.) *Twitter and Society*. New York: Peter Lang.
- Garimella, Kiran; De Francisci Morales, Gianmarco; Gionis, Aristides; Mathioudakis, Michael (2015). *Quantifying Controversy in Social Media*. arXiv:1507.05224v2 [cs.SI]

- Goffman, Erving (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday.
- Hall, Stuart (1980) Encoding/Decoding, s. 128-138 i Centre for Contemporary Cultural Studies (red.) *Culture, Media, Language: Working Papers in Cultural Studies, 1972-79*. London: Hutchinson.
- Hamrefors, Sven (1999). *Spontaneous Environmental Scanning. Putting 'putting into perspective' into perspective*. Doktorsavhandling. Stockholm: Ekonomiska forskningsinstitutet vid Handelshögskolan.
- Hamrefors, Sven (2014). *Kommunikativt ledarskap: Den nya tidens ledarskap i värdeskapande nätverk*. Lund: Studentlitteratur.
- Hogan, Bernie (2012). Pseudonyms and the Rise of the Real-Name Web, s. 290-308 i Hartley, John m.fl. (red.) *A Companion to New Media Dynamics*. Chichester: Blackwell Publishing.
- Kwak, Haewoon; Lee, Changhyun; Park, Hosung; Moon, Sue (2010). What is Twitter, a social network or a news media?, s. 591-600 i Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC.
- Larsson, Anders O. (2014). Everyday elites, citizens or extremists? Assessing the use and users of nonelection political hashtags. *MedieKultur*, 30(56): 61-78.
- Larsson, Anders O. (2015). "Ju mer vi är tillsammans" – partiledare, samhälleliga eliter och medborgare på Twitter under riksdagsvalet 2014, s. 91-112 i Carlsson, Eric m.fl. (red.) *Digital politik. Sociala medier, deltagande och engagemang*. Göteborg: Daidalos.
- Larsson, Anders O. (2016). Online, all the time? A quantitative assessment of the permanent campaign on Facebook. *New Media & Society*, 18(2): 274-292.
- Lasorsa, Dominic; Lewis, Seth; Holton, Avery (2012). Normalizing Twitter: Journalism practice in an emerging communication space. *Journalism studies*, 13(1): 19-36.
- Latour, Bruno; Jensen, Pablo; Venturini, Tommaso; Grauwlin, Sébastien; Boullier, Dominique (2012). 'The whole is always smaller than its parts' – A digital test of Gabriel Tarde's monads. *British Journal of Sociology*, 63(4): 590-615.
- Lewis, Seth; Zamith, Rodrigo; Hermida, Alfred (2013). Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods. *Journal of Broadcasting & Electronic Media*, 57:1: 34-52, doi:10.1080/08838151.2012.761702
- Marwick, Alice & boyd, danah (2011). I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society*, 13(1): 114-133. doi:10.1177/1461444810365313
- Mead, George H. (1934). *Mind, self and society from the standpoint of a social behaviorist*. Chicago: University of Chicago Press.
- Melbeck, Christian (1996). Relationships between neighbors and between acquaintances, s. 248-268 i Auhagen, Ann E. & von Salisch, Maria (red.) *The Diversity of Human Relationships*. Cambridge: Cambridge University Press.
- Meyer, Robinson (2015). The Decay of Twitter. The Atlantic, 2 november. <http://www.theatlantic.com/technology/archive/2015/11/conversation-smoosh-twitter-decay/412867/>
- Moretti, Franco (2013). *Distant Reading*. London: Verso.
- Ong, Walter (1982). *Orality and Literacy: The Technologizing of the World*. London & New York: Methuen.
- Pariser, Eli (2011). *The Filter Bubble: What the internet is hiding from you*. New York: Penguin Books.
- Paulussen, Steve & Harder, Raymond (2014). Social media references in newspapers: Facebook, Twitter and YouTube as sources in newspaper journalism. *Journalism Practice*, 8(5): 542-551.
- Puschmann, Cornelius & Burgess, Jean (2014). The Politics of Twitter Data, s. 43-54 i Weller, Katrin m.fl. (red.) *Twitter and Society*. New York: Peter Lang.
- Rogers, Richard. (2013). Debanalizing Twitter: The Transformation of an Object of Study. WebSci '13, Proceedings of the 5th Annual ACM Web Science Conference, 356-365. doi:10.1145/2464464.2464511
- Sakaki, Takeshi; Okazaki, Makoto; Matsuo, Yutaka (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors, s. 851-860 i Proceedings of the 19th International Conference on World Wide Web. Raleigh, NC.
- Simmel, Georg (1922). The Web of Group Affiliations, s. 125-195 i *Conflict and the Web of Group Affiliations*. New York: Free Press.
- Sjøvaag, Helle & Stavelin, Eirik (2012). Web media and the quantitative content analysis: Methodological challenges in measuring online news content. *Convergence: The International Journal of Research into New Media Technologies*, 18(2): 215-229. doi:10.1177/135485 6511429641
- Stewart, Bonnie (2014). Something is rotten in the state of... Twitter. Blogginlägg, 2 september. <http://theory.cribchronicles.com/2014/09/02/something-is-rotten-in-the-state-of-twitter/>
- Strömbäck, Jesper (2015). Social sammanhållning och medieanvändning, s. 63-80 i Bergström, Annika m.fl. (red.) *Fragment*. Göteborgs universitet: SOM-institutet.
- Sunstein, Cass (2007). *Republic.com 2.0*. Princeton: Princeton University Press.
- Tao, Ke; Hauff, Claudia; Abel, Fabian; Houben, Geert-Jan (2014). Information Retrieval for Twitter Data, s. 195-206 i Weller, Katrin m.fl. (red.) *Twitter and Society*. New York: Peter Lang.
- Trilling, Damian & Schoenbach, Klaus (2015). Investigating people's news diets: How online news users use offline news. *Communications: The European Journal of Communication Research*, 40(1): 67-91.
- Vis, Farida (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, 18(10). doi: 10.5210/fm.v18i10.4878

- Wadbring, Ingela (2016) Om dem som tar del av nyheter i lägre utsträckning än andra. I: Oscar Westlund (red.) *Människorna, medierna & marknaden: Medieutredningens forskningsantologi om en demokrati i förändring*, SOU 2016: 30, s. 463-486.
- Warner, Benjamin (2011). *Fragmentation and Social Polarization: The Character of Political Attitude Extremism in America*. Doktorsavhandling, University of Kansas.
- Webster, James (2014). *The Marketplace of Attention: How Audiences Take Shape in a Digital Age*. Cambridge, MA: MIT Press.
- Webster, James & Ksiazek, Thomas (2012). The Dynamics of Audience Fragmentation: Public Attention in an Age of Digital Media. *Journal of Communication*, 62(1): 39-56.
- Weller, Katrin; Bruns, Axel; Burgess, Jean; Mahrt, Merja; Puschmann, Cornelius (red.) (2014). *Twitter and Society*. New York: Peter Lang.
- Wesch, Michael (2009). YouTube and You: Experiences of Self-awareness in the Context Collapse of the Recording Webcam. *Explorations in Media Ecology*, 8(2): 19-34.
- Zelenkauskaitė, Asta & Bucy, Erik (2016). A scholarly divide: Social media, Big Data, and unattainable scholarship. *First Monday* 21(5). doi:10.5210/fm.v21i5.6358
- Zuiderveen Borgesius, Frederik; Trilling, Damian; Möller, Judith; Bodó, Balázs; de Vreese, Claes; Natali Helberger, Natali (2016). Should we worry about filter bubbles? *Internet Policy Review* 5(1). doi:10.14763/2016.1.401

Jonas Andersson Schwarz är lektor i medie- och kommunikationsvetenskap (Södertörns högskola) och intresserar sig för hur vardagsliv och samhällsstrukturer påverkas av den tilltagande digitaliseringen. Efter flera års forskning om delning och delningssamfund har han på senare tid skiftat fokus mot digitala plattformar som strukturellt fenomen, inte minst den datadrivna medieekologi som dessa möjliggör.

Johan Hammarlund är fristående medieforskare och gästlärare i medie- och kommunikationsvetenskap på Södertörns högskola. Han är även rådgivare inom omvärldsbevakning och medieanalys. I fokus för båda dessa verksamheter är det framväxande digitala medielandskapet där sociala medier och redaktionella medier samspelar på nya sätt.

Pernilla Severson

Medvetna frågor och val

Utveckling av metoder för att undersöka användargenererat material i digitala kontexter

Abstract

This is an article concerning the development of methods for studying user-generated content (UGC) in digital contexts. Method development can advantageously mean applying analytical reference points: a form of perspective making and navigation with a common understanding of where different disciplines can contribute. Cultural studies on 'pleasure' raises the understanding connected to how UGC creates meaning for the people who use and create the material, for the content which has different meanings and for the view that politics is broader than news. Design research on 'personas' recalls how images of users are built into the development of services as well as in research. Informatics (computer science) 'data mining' deepens understanding of the importance of the digital as research material (data) and the skills that are required to analyse the data. These three analytical points of reference work in interaction with discussions related to tools about precisely how a study may be designed.

Keywords: user generated content, method development, pleasure, personas, data mining

Den här artikeln inbjuder till att diskutera och förstå hur digitaliserat mediematerial i form av studiet av användargenererat material väcker resonemang om och möjliga perspektivbyte för metod (i bred betydelse, se nedan). Det finns olika syn på värdet av såväl metod som av olika metoder och huruvida användning av teorier spelar roll för både studiedesign och slutsatser som kan dras. Samtidigt innebär ökad medvetenhet om och diskussion kring just de här olika synsätten och tillämpningarna dels att olika studiers styrkor och svagheter artikuleras, dels möjliggör utveckling åt ett önskat håll.

Användargenererat material (user generated content) är olika typer av innehåll som användare frivilligt skapar och bidrar med i olika sammanhang. Begreppet används för att skilja mellan professionellt producerat mate-

rial och amatörproducerat material i digitala kontexter. Historiskt finns det många medieformer där amatörer på olika sätt skapat och spridit material, allt från brev och böcker till radio och tv (Lobato m.fl. 2011). Det som är speciellt i digitala kontexter är att gränserna mellan professionellt och användargenererat material blir mindre tydliga (se t.ex. Pavlik & McIntosh 2014). Ett tydligt exempel är hur tidningsläsarbrev till en redaktör eller dylikt alltid funnits, och i olika grad tagit plats i tidningen, men där webbtidningar nu innebär andra former och högre grad av interaktion och spridning.

Framväxten av sociala nätverkssajter innebär också att det användargenererade materialet inte nödvändigtvis behöver vara relaterat till professionella sammanhang. Sociala nätverkssajter (social network sites) är webbase-

rade tjänster som tillåter individer att skapa en publik eller semi-publik profil inom ett avgränsat system, där kontakter i olika former beroende på sajt kan ske med andra användare som man delar en koppling till (Boyd & Ellison 2007). Det handlar om former av sociala medier (YouTube, Facebook, Twitter etcetera), som ger upphov till begrepp som 'prosumer' och 'produsage' där användargenererat material kan förstås utifrån anspråk på större medverkan och makt (van Dijk 2009).

Det är i det här dubbla sammanhanget som den här artikeln placeras i med grundsynen att användargenererat material har potential att säga något viktigt om våra liv och vårt samhälle. Ett exempel som tydliggör det är den årliga rapporten från World Internet Institute som visar att det finns en utbredd och varierad både konsumtion och produktion av olika former av användargenererat material (Dunahee & Lebo 2015).

Syftet med artikeln är att bidra till metodutveckling för att undersöka användargenererat material i digitala kontexter. Analys och slutsatser kring användargenererat material har från mediastudiers håll ofta undersökts som en del av publikens agerande och som mindre värt än de professionella producenternas. Med den referenspunkten utmanas och utarmas professionell produktion av användargenererat material (se avsnitt "Exempel på svagheter i mediastudier av användargenererat material" nedan). Det innebär att flera betydelselager av användargenererat material utelämnas.

Det krävs ytterligare förståelse för och perspektiv på användargenererat material. I denna artikel föreslås därför lösningar som har med metodutveckling att göra. Det handlar till stor del om att det enskilda mediet med sina specifika drag får ännu större betydelse för de metoder som kan tänkas användas i en vetenskaplig studie (se särskilt Durham Peters 2015). Att överföra medie-bundna metoder från ett medium till ett annat är problematiskt, exempelvis att använda text-bundna metoder rakt av för audiella eller audiovisuella medier (se t.ex. Hayles 2004). Det är problematiskt inte minst för att metoder ofta åtföljs av en

viss förståelse som pekar mot vissa slutsatser. De metoder som används bär alltså ofta med sig en form av medieförståelse (för likartade argument om medier och materialitet, se mediet teori av särskilt Meyrowitz 1994).

Samtidigt som medie- och kommunikationsvetenskapen har medier som sitt studieobjekt har disciplinen visat begränsat intresse för att utveckla metoder som beaktar mediespecifika (se t.ex. Nyre 2010). Digitala medier har samtidigt manat fram ett intresse för metodutveckling från många discipliner och fält. Det gäller särskilt den forskning som görs om digitala metoder (Arthur & Bod 2015; Roberts, Snee & Hine 2015; Rogers 2013). Inom fältet pekas på värdet av att arbeta mer med metodutveckling för att ytterligare förstå och fördjupa kunskap om digitala medier. Kritiken som framförs från digitala metod-perspektiv är att många mediastudier inte använder de digitala medierna som metod i sig och att viktiga dimensioner och aspekter därmed förbises. Det handlar inte minst om att försöka utnyttja de metodologiska möjligheter som mediet själv tillhandahåller, att exempelvis studera hur en sökmotor ordnar information genom att använda sökningar som metod.

Metodutveckling kan se olika ut beroende på perspektiv, inriktning och grundsyn på vad metoder är. Metoder kan vara resultat av att göra (vanor i praxis) och att reflektera (medvetet utformade riktlinjer) (Goldkuhl 1993). Metod avser ett föreskrivet tillvägagångssätt som vägleder kunskap och är därmed en social företeelse som kan förstås på olika nivåer: De kan vara subjektiva i form av kunnande, intersubjektiva genom flera personers likartade kunnande, språkliga i uttryckta metodbeskrivningar och även vara de handlingar och effekter som metoder föreskriver och leder till. Metoder kan dessutom implementeras i artefakter, även i digitala artefakter, med mer eller mindre uttryckta metodregler. Metoder bygger på mer eller mindre explicita synsätt som ska komma till uttryck vid användning av metoden, vilket därmed är en form av styrd uppmärksamhet. Metoder är alltså kunskap på en generell nivå att användas i specifika situationer (ibid.).

Den här artikeln handlar om behovet av metodutveckling för användargenererat material i digitala kontexter och föreslår att det på ett värdefullt och relevant sätt kan göras med analytiska referenspunkter. Att vara analytisk innebär att förstå perspektiv, samband, orsaker och konsekvenser, ofta genom jämförelser, och leder många gånger till att lösningar på problem kan identifieras. En referenspunkt är något som används för att dels navigera, dels för att skapa och nå en gemensam uppfattning. Med analytiska referenspunkter betonas att metod inte bara handlar om verktygen för insamling och analys.

Istället för att föreslå exakta metodupplägg kommer jag här att rekommendera analytiska referenspunkter för metodutveckling för att undersöka användargenererat material i digitala kontexter. En forskare som står i begrepp att göra studier av användargenererat material kan göra en mer robust genom att föreställa sig sin studie utifrån de tre analytiska referenspunkterna som förordas i artikeln.

Jag inleder med att diskutera de svagheter som kan identifieras i tidigare mediestudier, för att därefter diskutera hur metodutveckling kan ske.

Exempel på svagheter i mediestudier av användargenererat material

Behovet av metodutveckling för att studera användargenererat material grundas i en vetenskaplig kunskapslucka. En kunskapslucka betyder att det finns något osäkert och otillräckligt i existerande vetenskapliga studier. I det här avsnittet illustreras behovet av utvalda studier av användargenererat material från traditionella medie- och kommunikationsvetenskapliga perspektiv. Dessa studier har en huvudsaklig tyngdpunkt på hur massmedieproduktion förändras av att publiken får bidra med innehåll. Dessa studier ur ett produktionsperspektiv handlar ofta om vad dagspress gör på internet och de förändringar som sker genom närvaro av användargenererat material (Wahl-Jorgensen 2015).

I studier av dagspress på internet värderas samspelet mellan journalistik och användargenererat material ofta utifrån hur det förändrar och riskerar att urholka journalistiska ideal. Ett typexempel är att undersöka hur och om det är viktigt att användargenererat material används inom nyhetsjournalistik och huruvida det lever upp till de krav som finns på nyhetsjournalistik.

Ett tidigt och illustrerande exempel är Thurmans (2008) intervjuer med redaktörer och redaktionschefer om de journalistiska debatter som förs om gräsrots- och medborgarjournalistik. Studien visar förändringar som äger rum i journalistikens förhållande till sina konsumenter. I en annan studie använder Hermida och Thurman (2008) en enkät för att undersöka användningen av användargenererat material av traditionella nyhetsorganisationer, som visar stora möjligheter för bidrag från läsare. Djupintervjuer med ledande nyhetschefer (ibid) pekar på tveksamheter om det redaktionella och kommersiella värdet av material från allmänheten. Slutsatsen som dras är att tidningar börjat med en traditionell gatekeeping i hanteringen av användargenererat material.

Andra studier av användargenererat material har gjorts genom innehållsanalyser, ibland kombinerat med enkäter. Innehållsanalyser med produktionsinriktning kan vara studier av nyhetstidningarna på internet i olika länder (Jönsson & Örnebring 2011), de största nyhetstidningarna på internet i ett land (Karls-son m.fl. 2015), eller studier av innehåll från aktörer som säger sig bedriva medborgarjournalistik (Holt & Karlsson 2015). I dessa studier jämförs användargenererat material med traditionell nyhetsjournalistik och den slutsats som dras är att användargenererat material inte lever upp till ett sådant journalistiskt ideal. Genom att betona vem som har makten över ett medieinnehåll och hur användare deltar i en etablerad medieproduktion, blir uppdelningen producent och konsument central. Där har då producenten mer makt och användargenererat material ska mer förstås som en form av aktiv konsumtion med risk för kommodifiering av användare.

Uppdelningen av innehåll i nyheter/information respektive populärkulturellt innehåll kan relateras till uppdelningen mellan det offentliga och det privata. Den politiska offentliga sfären representeras ofta av nyheter/information och adresserar publiken som medborgare medan den kulturella offentliga sfären istället tilltalar publiken som kultur- och underhållningskonsumenter (Jönsson & Örnebring 2011). Exempel på studier är där medborgarjournalisternas sätt att berätta nyheter studerats (Holt & Karlsson 2015), samt där det ställts frågor om människors benägenhet att kommentera nyhetsartiklar, vilket setts som ett tecken på publikens deltagande i offentligheten som medborgare (Karlsson m.fl. 2015).

I den här typen av studier reproduceras således föreställningen om att en viss typ av medieproduktion är värdefull för den politiska offentliga sfären, medan annan inte är det. Det blir en mer eller mindre given utgångspunkt. Ett illustrerande värderings-resonemang finns i Östmans (2012) enkätstudie med svenska ungdomar 13-17 år. I analysen av resultaten uppvärderas produktion och delning av videos, musik, film och att skriva personliga bloggar. Sociala medier-aktivitet (inklusive att göra inlägg) däremot kvalificerar sig inte som en form av medieproduktion. Demokratiskt engagemang översätts också till att betyda politisk kunskap (om det partipolitiska systemet om styre och lag och rätt) och politiskt deltagande (mer eller mindre kopplat till politiska budskap snarare än samhällsfrågor som respondenterna själva tar upp).

I de ovan nämnda studierna visas självmedvetenhet om otillräcklighet i studiernas utformning, samt den osäkerhet som därmed finns vad gäller vilka slutsatser som faktiskt kan dras. Jönsson och Örnebring (2011) argumenterar för att användargenererat material är bredare än journalistik och att demokratisk potential inte bara behöver ligga i politiskt inflytande utan även i kulturell jämlikhet. Även Karlsson m.fl. (2015) pekar på värdet av att vidga vad som undersöks, exempelvis deltagande i arenor utanför traditionella medier. Holt och Karlsson (2015) tar upp att det finns

mer spridd medborgarjournalistik att undersöka. Östman (2012) lyfter speciellt fram att framtida studier bör utforska medierande och modererande faktorer som kan påverka etablerade förhållanden mellan hur involverad användare är i att producera användargenererat material och att ha och utöva ett demokratiskt engagemang.

Svagheten är således inte att det saknas en medvetenhet om otillräcklighet i studierna. Problemet är snarare att det ändå dras långt gångna slutsatser om vilken typ av användargenererat material som är av mest och minst politiskt värde och att teknikens formande kraft i många fall inte uppmärksammas. Det finns exempelvis studier som visar att det behövs mer kunskap om hur användargenererat material kan och bör förstås som tecken på koppling till politiska aspekter som brukar värderas i relation till ett medborgerligt engagemang och ett demokratiskt samhälle. Leung (2009) undersöker vilken tillfredsställelse som finns kopplat till olika typer av innehåll och hur 'offline' och 'online' hänger ihop. Resultatet visar att psykologisk egenmakt (empowerment) kan förstärkas både av ens bidrag till online-material och av användarnas attityd till och beteende vad gäller medborgarengagemang offline. Blotta närvaron av vilket användargenererat material som helst kan därmed vara tecken på egenmakt. Det kan då vara kopplat till det som kan förstås som politiska aspekter som brukar värderas i relation till ett medborgerligt engagemang och ett demokratiskt samhälle.

Andra former av studier av användargenererat material med innehållsperspektiv utgår mer från tekniska förutsättningar och egenskaper, vilket kan göras med metoden "mix-of-attributes" (Dylko 2014). Metoden bygger på antagandet att ett medium är en kombination av olika teknologiska egenskaper, som interaktivitet och modalitet (ibid.). Klargörande av vad politiskt användargenererat material är och vilka implikationer det kan ha för politiskt deltagande blir då vad Dylko och McCluskey (2012) identifierar som fem tekniska attribut för politiskt användargenererat material: sökeffektivitet, anpassningsbarhet, innehållsma-

nipulering, kostnadsreduktion och gemenskapsorientering. Genomgående för alla de fem tekniska attributen är att de dels kopplas till betydelse och värde för tillgång till och anpassning av politisk information. Exempelvis innebär sökeffektivitet då att hjälpa användare att spara resurser kognitivt, tidsmässigt och energimässigt, vilket kan stödja dem att "more deeply process the retrieved political information, which increases political knowledge, which then increases political participation" (Dylko 2014: 504). Dels kopplas attributen också till hur de stimulerar utbyte och deltagande, där kausalitet i flera led även här finns med. Exempelvis innebär innehållsmanipulering då att redigera en wiki-sida kan stärka uppfattade politiska egenmakt, vilket kan leda till ökat politiskt deltagande.

Den här artikeln vill bidra till en ökad förståelse för det användargenererade materialet och pekar framförallt på behovet av metodutveckling.

Metodutveckling för att studera användargenererat material i digitala kontexter

Nya digitala metoder kan bidra till ökad förståelse för hur det går att undersöka användargenererat material i digitala kontexter. Rogers (2013) menar att digitala metoder inte bara handlar om internetforskning och att studera onlinekultur, inte heller är de bara ett verktyg eller handlar om hur en mjukvara fungerar. Digitala metoder är istället ett sociomateriellt sätt att närma sig digitala medier: att tänka med apparater och objekt för att forska om kulturell förändring och samhällsliga tillstånd.

Digitala metoder kan vara onlineenkäter, men metoderna kan också vara digitalt inbyggt, såsom rekommendations- och folksonomisystem. Digitala metoder handlar då om att samla in och analysera hyperlänkar, taggar, sökmotorresultat, arkiverade hemsidor och andra digitala objekt. Det innebär att lära från sökmotorer och rekommendationssystem om kulturell förändring och samhällsliga förhållanden vad gäller onlinedynamik. Samhällsvetenskapliga forskare som studerar internet skulle därför

kunna använda kunskaper från områden som människa-dator-interaktion, socialpsykologi och cyberkulturella studier (ibid.).

Att förstå användargenererat material innebär att med digitala metoder uppmärksamma att det är en form av användardata i egenskap av preferenser, sökhistorik, köphistorik och plats som samlas in och analyseras. Det går då att förstå användargenererat material som det som mjukvaran registrerar som användarnas vardagsanvändning. Webbläsarhistorik, rekommendationssystem och algoritmer framtonar då som särskilt intressant att studera. Delning av länkar kan ses som ett tecken på effekt och anseende. Rogers (2013) bygger det här argumentet utifrån att google som sökmotor behandlar hyperlänkar som tecken på att någon/något får status i betydelsen 'ryktesmarkör' och 'relevans-indikator'. Det blir då intressant att integrera en förståelse för teknisk specificitet och materialitet med deras diskurser och representationer (ibid.). Rogers (ibid.) exemplifierar den här integrerade förståelsen med att länkar kan ses som en "politics of association", en förbindelse som har en politisk betydelse. Det skulle kunna innebära att studera hur en aktör, exempelvis Greenpeace, länkar till andra webbsidor från sin egen webbsida och förstå de här länkarna som ett politiskt uttryck. Det kan också innebära att en sökmotor kan användas för att studera social förändring, så som Google Flu Trends (ibid.).

Att förstå digital metodutveckling kan också vara så som Snee (2015) betonar, att digitala metoder har potential men att de riskerar att betona det mätbara. Det digitala blir inbäddat, inte bara i våra vardagsliv utan också i metoder. Det gör att forskare behöver vara mer agila och koppla ihop flera metoder för att bygga mångfasetterade perspektiv på multimodala fenomen. Det handlar om att jämföra data erhållna på olika sätt och att kombinera dem för att undersöka komplexa och hybrida sociala situationer. Det inbegriper att studera och förstå det multimodala, i betydelsen text som innehåller flera olika kommunikationsformer som skrift, bild, rörlig bild, tal, musik. I mångt och mycket innebär det också en triangulering, att kombinera och att jämföra.

Ett annat perspektiv på att förstå digital metodutveckling kopplat till användargenererat material är det Gubrium och Harper (2013) beskriver om hur visuella och digitala metodologier kan bidra till en participatoriskt "public-engaged ethnography". Fokus är då på att ändra det traditionella förhållandet mellan akademiska forskare och samhället. Det skulle innebära att studier av användargenererat material inkluderar en uppmuntran till användare att reflektera och engagera sig i det som är viktiga frågor för dem och deras sammanhang.

Konkret kan impulserna från digitala metoder-fältet innebära att förstå användargenererat material i digitala kontexter som en typ av utveckling av digitala metoder. Ett exempel på det är hur Herring (2004) gör en historisk beskrivning av innehållsanalysens ursprung, tillämpning och utveckling och argumenterar för att nya typer av medieinnehåll borde frammana utveckling av de definitioner, de urvalsresonemang och de forskningsetiska övervägande som görs. Herring (ibid.) betonar att metodutmaningar bäst möts genom att låna från andra discipliner och anpassa till den egna. Det kan exempelvis handla om att koppla samman innehållsanalys och social nätverksanalys. I studier av internetinnehåll, som konversationer och hyperlänkar, är det värdefullt att använda flera metoder och se till att anpassa metoderna till datormedierade miljöer. Social nätverksanalys är då bäst lämpad för att studera konversationer och länkmönster som kan besvara frågor om community-formering, auktoritet och påverkan (i betydelsen att centralitet anses betyda denna prestige, dvs. om en aktör refereras till av många andra webbsidor) samt webbsajttrovärdighet och webbgenrestruktur. Etablerade forskningsmetoder utmanas även av att det kan vara svårare att göra slumpmässiga urval på webben och att vedertagna statistiska modeller därmed framstår som problematiska. Användningen av i förväg fastslagna kodschemata försvåras dessutom av att innehåll i nya medier ofta har karaktären av framväxande fenomen som uppträder och förändras snabbt.

I utvecklingen av metoder för att studera material i digitala kontexter är en diskussion

om digitala verktyg viktig – men diskussionen bör inte begränsas till det. De i artikeln nedanstående föreslagna analytiska referenspunkterna fångar upp grundläggande ståndpunkter inom digitala metoder-fältet och ger dem ett sammanhang där de kan förstås i relation till varandra.

"Pleasure" som analytisk referenspunkt

Kulturstudiernas betoning av "pleasure", njutning, används här som en grund för ambitionen att vidga betydelsen och värdet av det politiska. Det har mindre att göra med användardata och mer med att bygga mångfasetterade perspektiv på ett användargenererat materials värde för politik och demokrati. Njutning har varit en del av mediepublikforskning sedan 1980-talet (O'Connor & Klaus 2000). Kulturstudier analyserade och begripliggjorde publiken som de såg som aktiva och skapande. Njutning var för kulturstudierna ett sätt att hantera och uttrycka värdet i populärkulturen. Tidigare forskning hade handlat mer om högkultur än lågkultur.

O'Connor och Klaus (2000) visar på hur två distinkta spår finns inom receptionsanalysen, och argumenterar för att de bör överbryggas både på teoretisk och empirisk nivå. De nämner dels det allmänna kunskapsprojektet (public knowledge) med fokus på ideologi, offentliga sfären, medborgarskap och mediers informativa roll med nyheter och samhällsprogram som direkt kopplade till det Corner (1991) formulerar som "the politics of information and the viewer as citizen" (Corner 1991: 268). Den andra distinkta vägen inom receptionsanalysen urskiljer O'Connor och Klaus (2000) vara populärkultur-projektet (popular culture), med fokus att utforska njutning av fiktionsgenrer. Även här exemplifierar de med en beskrivning från Corner (1991) som menar att populärkultur handlar om hur medier som underhållningskälla – en industribaserad populärkultur – har implikationer för social medvetenhet, vilket därmed är kopplat till smak och njutning som i sin tur är klass- och könsbundet.

O'Connor och Klaus (2000) menar att det förra spåret försummar frågor om njutning och det senare frågor om ideologi. De fortsätter resonemanget med att lyfta att det allmänna kunskapsprojektet också har en inbyggt maskulin bias, med referens till Hermes (1997). Uppdelning i offentliga sfärer som associeras med män, produktion och rationalitet och i privata sfärer som i sin tur associeras med kvinnor, konsumtion och känslor, är både begränsande och sexistisk (Fraser 1994; Benhabib 1992).

Problemet är att studier av användargenererat material i digitala kontexter fortfarande verkar dras med den här uppdelningen att fiktionsgenrer enbart är underhållning och skvaller och bara hör till den privata sfären. O'Connor och Klaus (2000) visar att det är en grundläggande felaktighet i att skilja mellan verklighet och fantasi, mellan kognitiv rationalitet och emotionell sensualitet i processen att tillägna sig och förstå kunskap. Det finns länkar mellan njutning och kunskap. Populärkulturella genrer har för publiken realistiska moment och kognitiva dimensioner. Njutning präglar publikens användning även av aktualitets-, informations- och allmänt kunskapsinnehåll. Det visar sig också att lägre värderade nyhetsartiklar som historier om katastrofer, brott eller mänskliga tragedier erbjuder publiken faktisk njutning. Och i användandet av koder som människor är välbekanta med är det mer sannolikt att de förstår och finner upplevelsen angenäm:

This goes beyond the dutiful citizen position, and actually conflicts with it, since the discourse of the dutiful citizen gains its legitimacy precisely in the idea of social obligation rather than pleasure (Dahlgren 1988: 296-297)

Det finns därmed en koppling mellan emotion och kognition och det är i sin tur betydelsefullt för hur vi kan förstå nyheter och dess del i, och koppling till, den politiska dimensionen (ibid.). Journalistik är heller inte enbart en rationell domän och inte helt skild från underhållning (ibid.). En sådan insikt innebär att fokus förflyttas från ideologi till

betydelse- och meningsskapande (O'Connor & Klaus 2000). Här kan berättelseanalysen vara en brygga som förbinder studier av såväl fakta som fiktion. Dessutom är det ett sätt att förstå varför och hur det användargenererade materialet ofta efterliknar det professionellt producerade. Berättarstrategierna framstår helt enkelt som universella (ibid.).

Njutning är den känslomässiga, sensuella och föreställda känsla som gör att publikerna aktivt vänder sig till och bearbetar ett givet innehåll. Njutning är länkat till den offentliga sfären genom kulturellt medborgarskap (cultural citizenship), som är Hermes (1997) idé lånat från Allor och Gagnon (1994) (O'Connor & Klaus 2000). Kulturellt medborgarskap är i det sammanhanget länkat både till producenter och konsumenter. Det gör det möjligt att dela på praktiker och identiteter i betydelsen att blanda frågor om njutning med frågor om politik. Det stödjer också omdefinierande av gränser av det offentliga och det privata som något närvarande i vardagen och ömsesidigt konstituerande (Hermes 1997).

Vad det innebär att studera användargenererat material som uttryck för en form av kulturellt medborgarskap är inte klart och färdigdefinierat. Men det skulle kunna tas på allvar som en del av den offentliga sfären, vilket då inkluderar subjektivitet och självkänedom där även hopp, fantasier och utopier äger giltighet (ibid.). Användargenererat material bär i det perspektivet på estetiska och emotionella aspekter som adderas till, och inte är åtskilt från rationell och moralisk argumentation. Det är ett resultat av vardagliga aktiviteter och kommunikativa händelser genom vilka människor bekräftar sin gemensamma kultur, rekonstruerar sin sociala identitet och arbetar om normer och värderingar som reglerar beteende.

Genom denna bekräftelse ställs andra viktiga frågor: Hur stimulerar genrekonventioner i användargenererat material innehållsspecifikt njutning och kommunikativ involvering? Hur är användargenererat material del av interaktion, skvaller och ömsesidig uppmuntran? Hur är användargenererat material en del av användarnas motivation att både producera

användargenererat material och ta del av andras? Hur kan användargenererat material vara en plats för motstånd mot exempelvis patriarkala krav? Vad är det för känslor som användargenererat material skapar hos publik och professionella? Är villkoren desamma för användargenererat material som för såpoppor, att njutning beror på trovärdigheten i karaktärerna, deras känslor och deras interaktion och på sannolikheten att de händer i verkliga livet – och – därmed att det är många samhällsfrågor och politiska problem som tas upp?

Kulturstudiernas utgångspunkter och sätt att undersöka njutning på, visar vilket tolkningsarbete som krävs för att på ett relevant sätt meningsgöra vad det är vi faktiskt vill undersöka i användargenererat material. Det visar på värdet att verkligen förstå användargenererat material, hur det skapar mening för människor som använder det och vilket innehåll som har vilken funktion. Förståelsen av politik bör innefatta mer än nyheter och det är en analytisk referenspunkt att ha i åtanke när användargenererat material studeras.

“Personas” som analytisk referenspunkt

En annan fruktbar analytisk referenspunkt för den digitala metodutvecklingen utgör designforskningens personabegrepp. En persona kan definieras som en modell av en användare, skapad främst utifrån kvalitativa studier, vilken används i en utvecklingsprocess för att generera bättre produkter (Cooper m.fl. 2007, Long 2009).

Med personas som analytisk referenspunkt går det att undersöka och förstå hur användargenererat material präglas av tjänster och plattformproducenter som bär på föreställningar av vilka användarna är. Personas kan också synliggöra forskares inbyggda och föreställda bild av användare. För att utmana detta handlar det om att ställa frågor av typen: Hur grundas egentligen antaganden om användargenererat material och de användare som skapar materialet?

Pruitt och Adlin (2006) visar att föreställningar om personas spänner över allt från

ytliga antagande om personas till datadrivna personas (grundade i tidigare forskning om användare med dess data från såväl kvantitativa som kvalitativa studier) och att göra nya studier av användare. Turner och Turner (2011) lyfter särskilt föreställningar om personas som sätt att undersöka och förstå hur användarmodeller riskerar att reproducera stereotyper. I metodutveckling kan deras resonemang om föreställningar och reproduktion av stereotyper användas som grund för såväl forskare som tjänsteproducenter. Resonemanget fångar upp det som Gulliksen och Göransson (2003) beskriver, att personas kan stödja kommunikation inom tekniska utvecklingsteam samt fungera som en slags dokumentation av användarnas situation. Används personas på det viset av tjänsteproducenterna? Och hur kan forskares diskussion om personas, som föreställda användare, stödja studiet av användargenererat material?

En fortsatt nyansering och fördjupning av personas kan ske utifrån Massanari (2010), som argumenterar för att användningen av personas motiveras lika mycket av politiska verkligheter inom organisationer för nya medier, som av en önskan att möta användarbehov. Genom att peka på organisationers politiska verkligheter vill Massanari peka på såväl de strukturella villkoren, exempelvis ekonomiska och juridiska, som de ideologiska för att betona att personas är ett politiskt verktyg. För metodutveckling innebär det möjligheten att skapa en känslighet inför hur användare konstrueras och att användarbehov kan vara en konsekvens av politiska verkligheter.

Hur personas används som en analytisk referenspunkt varierar mellan olika studier. I en studie där medieforskning möter såväl informatik som stadsutveckling använder Foth m.fl. (2011) personas i utveckling av lägenhetsboende, där de boende formar olika personas med olika behov. Det är alltså en aktionsforskningsinriktning där själva personametoden används för utveckling av något, vilket ligger väl i linje med betydelsen av att använda personas i designprocessen. Används persona som inspiration skulle det kunna leda till utveckling av kulturellt medborgarskap

i utforskandet och byggandet av tjänster för användargenererat material, som främjar just detta medborgarskap.

För studier av användargenererat material på dagstidningars nyhetssajter innebär persona som analytisk referenspunkt för metodutveckling en insikt om att det användargenererade materialet aldrig enbart representerar användares motiv, önskemål och reaktioner. Materialet bör snarare förstås och undersökas som en del av olika producenters politiska verkligheter. Då kan slutsatsen att användare inte är intresserade av att kommentera nyheter på nyhetssajter inte stå som enbart en konsekvens av användares motiv, önskemål och reaktioner. Slutsatsen borde inkludera en förståelse för strukturella villkor för kommenterandet, samt hur användarkommentarer även innebär en spegling av etablerade normer och värderingar och i vissa fall snarare mer tjänar som ett politiskt verktyg.

Mer konkret kan personas användas för att förstå hur materialet är samproducerat utifrån de förutsättningar som den tekniska tjänsten ger och att dessa förutsättningar är skapade av någon eller något. Vilka är de här producenterna av kommentarsfält på nyhetssajterna? Och är tjänsten en produkt av journalisters önskemål? Det här kan förstås och kallas för teknikens affordanser, som innebär att få syn på tjänstens egenskaper och förutsättningar utan att landa i teknologisk determinism (Juris 2012). Teknikens affordanser uttrycker tillsammans med personas specifika föreställningar om användare. En studie av användargenererat material på exempelvis YouTube bör då försöka förstå tjänsten och dess inbyggda syn på användare. Här kan det vara värdefullt att studera de interaktionsdesigners och programmerare som faktiskt konstruerat kommentarsfunktioner och tjänster som YouTube. Nyhetssajterna kan därmed också undersökas genom att intervjua dem som skapat användartjänsterna, vad det är för användarbehov och önskemål som byggs in i tjänsterna och vad som väljs bort.

Designforskningens ingångar till och användning av personas, är en metod utvecklad inom användbarhetsforskning och som visar på hur bilder av användare byggs in i såväl

tjänster som i studier. Perspektivet understryker att producenter av användargenererat material även är själva tekniktjänsterna och deras skapare – och deras användarmodeller.

“Data mining” som analytisk referenspunkt

För informatiken och datavetenskapen framträder det användargenererade materialet som data. Denna analytiska referenspunkt pekar tydligast på det som utgör grunden för användningen av digitala metoder: att utgå från det digitalas förutsättningar. Det innebär att data innehåller och präglas av egenskaper som inte enbart kan undersökas och förstås som direkta utsagor från publik, som rent innehåll, eller som aspirerande producentmaterial.

Utgångspunkten för resonemanget är att användargenererat material i digitala kontexter kan förstås som data. Rogers (2013) pekar på värdet av att just uppmärksamma hur webben genererar data. Durham Peters (2015) argumenterar för att digitalt innehåll snarare bör förstås som integrerade dataflöden, där det finns organisationer, positioner och metadata. Användargenererat material blir då intressant att förstå i sig själv och kanske som personifierad data eller som en särskild form av digital mediekonsumtion.

Data i det här sammanhanget innebär information som finns i en form som kan bearbetas med automatiska eller manuella metoder. Internationella standardiseringsorganisationen ISO definierar data som ”tolkningsbar representation av information på ett formaliserat sätt som lämpar sig för kommunikation, tolkning eller bearbetning (av människor eller med automatiska metoder)” (ISO 2016).

Användargenererat material som data innebär i praktiken många olika typer av data. Det innebär att ”rådata” i digitala kontexter är ett urval av datatyper som inte vanligtvis görs kända och arbetas med i humanistiska och samhällsvetenskapliga studier. Användargenererat material som data leder också till forskningsetiska överväganden, där olika datatyper som är offentliggjorda inte nödvändigtvis behöver innebära att användare ser

på sitt genererade material som ett offentliggörande där informerat samtycke för att ingå i en forskningsstudie automatiskt ges.

Jensen (2012, 2014) skiljer mellan data som hittad (automatiskt genererad) eller skapad (en meningsfullgjord kontextualisering) för att fördjupa metodkunskap och metodfärdigheter för digitala medier. Åtskillnaden ger en god grund för dataförståelse och kan vidareutvecklas. Purdam och Elliott (2015) betonar att det är viktigt att förstå data i relation till användarens godkännande. Den enda formen där användaren verkligen har uttryckt sitt godkännande är i *ortodox intentionell data* dit traditionella enkätsvar räknas in. För användargenererat material skulle det innebära att forskare använder material som användare tydligt har godkänt för användning och att materialet skapats inom ramen för en specifik undersökning där i praktiken ”fråga-svar” är grunden. Användningen av övriga former av data innebär viktiga forskningsetiska överväganden som bör ta hänsyn till hur informerat samtycke kan och bör ske. Användares syn på privatliv och integritet stämmer inte nödvändigtvis med synen på att offentligt är lika med informerat samtycke (Marwick & Boyd 2014). Det gäller även där användare själva bidrar med data, och indirekt genererar data som är sekundär till vad som görs i interaktionen – så kallade dataspar (elektroniska spar i form av datatrafik).

Att studera användargenererat material innebär möjligheten att komma åt stora datamängder (big data) vilket understryker behovet av att tillämpa och reflektera kring tekniker för datautvinning. Målet är att extrahera och analysera datamängder som är relevanta i relation till den uppgift som ska lösas. Grundläggande här är att det å ena sidan går att lägga samman en mängd data, eller låta en delmängd av all data representera något. Å andra sidan går det också att hitta kopplingar mellan data, och här finns risken att man hittar mönster som egentligen inte existerar (Dunham 2003). Klassiska statistiska metoder kan användas på små datamängder. Men vid stora datakvantiteter uppstår nya problem. Datautvinning utvecklades ur intresset för att

utveckla metoder som kan hantera den typen av data (Bertino 2001).

Datautvinning handlar ofta om att hämta information från databaser (Fayyad m.fl. 1996). Delprocesserna består av att: 1) önskad mängd av grunddata väljs ut, 2) utvald data bearbetas för att ta bort/rätta till felaktig data eller förutsäga data om den saknas, 3) att bearbetad data omvandlas för att lättare kunna användas, 4) själva datautvinningsprocessen, där algoritmer används för att generera information ur data, samt 5) visualisering/presentation (Dunham 2003). Det innebär att studier av stora mängder användargenererat material kan stärkas genom resonemang och synliggjorda val vad gäller alla dessa delar. För studier av mindre datamängder av användargenererat material krävs fortfarande en insikt om vilken grunddata som används och vilka som då väljs bort.

För att göra användargenererat material användbart för datautvinning krävs ofta att innehållet märks upp med beskrivningar, ofta baserat på maskininlärningstekniker. Ett typiskt sätt för automatisk märkning av användargenererat material är ”sentiment analysis”, en analys där känslouttryck och åsikter spåras (Moens, Li & Chua 2014). Den automatiska märkningen har begränsningar i de färdiga verktygen som finns tillgängligt för olika dataanalyser. Moens, Li och Chua (ibid.) menar att den stora utmaningen ligger i att förstå hur det går att skapa mening i innehållet så att det kan indexeras och beskrivas för ytterligare användning. Det här handlar om neurolingvistisk programmering, bildigenkänning och videoigenkänning som grund för bra datautvinningsalgoritmer. Ett problem här är att de ofta är utvecklade för ideala omständigheter och inte fungerar lika bra för användargenererat material, med exempelvis stavfel och bilder med dålig upplösning.

En annan problematik kan gälla sökorden. Det går inte att utgå från att ett begrepp som används för analys i kodschemata är samma som det sökord som används för att undersöka användargenererat material. Cambria m.fl. (2013) visar att studier av känslouttryck och ”opinion mining”, åsiktsutvinning, behöver ut-

veckla en gemensam kunskapsbas för begrepp som kan användas för att avkoda känslor och hur de styrs. Det här är ett karakteristiskt exempel på nödvändig metodutveckling där teori kan vidareutveckla datautvinning.

Visualisering är en annan viktig del att ta vidare som referenspunkt från datautvinning. Wheeldon och Ahlberg (2012) diskuterar datavisualisering inom samhällsvetenskap och visar på hur det kan ta plats i analysprocesser. Här finns kritiska frågor att vidareutveckla. Hur görs datavisualiseringar och hur sprids och används de, och vad får det för implikationer i samhället? Vilka problem och möjligheter uppkommer? Hur kan vi förstå datavisualiseringar som ett uttryck för maktintresse och maktmotstånd? Precis som i all annan forskning är det inte minst viktigt att det finns en transparens som gör det möjligt att förstå hur visualiseringar är gjorda och hur de relaterar till underliggande data.

Med datautvinning som analytisk referenspunkt betonas vikten av att förstå hur användargenererat material är olika former av användardata. Det handlar om ett särskilt (maskinunderstött) arbetssätt att ”märka” data och att söka efter denna ”märkta” data. Generellt innebär det en insikt om att användardata kräver särskilda analysprocesser som ibland går bortom statistiska mätmetoder och som inkluderar visualiseringar (både i analys och i presentation av analys). Datautvinning artikulerar slutligen vikten av ett kritiskt tänkande och den nödvändiga kunskapen av att förstå hur algoritmer fungerar.

Avslutning

Det finns metodmässiga utmaningar i att undersöka användargenererat material: Vad är det för typ av material, hur kan det förstås på ett relevant och meningsfullt sätt och inte minst, hur kan det då undersökas? Vid sidan om att undersöka användargenererat material som ett innehåll vilket som helst, kan analyserna byggas ut och stärkas genom att forskare även inkluderar den digitala kontextens betydelse.

För att hantera utmaningarna har jag i artikeln resonerat kring olika fälts bidrag i ter-

mer av analytiska referenspunkter. Genom att bredda förståelsen av det som undersöks och grunda det i såväl teoretiska som empiriska resonemang, skapas valida studier. Huvudfrågan rör vad det politiska utgörs av i användargenererat material, var det politiska finns i användargenererat material och hur det politiska uttrycks, struktureras och stimuleras i användargenererat material. I artikeln betonas överbryggandet av uppdelningen mellan det allmänna kunskapsprojektet och det populärkulturella projektet.

För en ökad förståelse av användargenererat material krävs också att uppmärksamhet riktas mot hur användarmodeller byggs och bäddas in i teknik av olika producenter i olika tjänster. Användargenererat material styrs på olika sätt av den tekniska tjänsten och i alla led, allt från skapandet av tjänsten till hur delande av material fungerar. I artikeln diskuteras detta utifrån designforskningens persona-metod.

I metodutveckling gäller det också att förstå och använda olika typer av användargenererat material även som data. Den digitala kontexten ger möjligheter till undersökningsmaterial som antingen väljs eller väljs bort. Det finns en mängd ny data som kan studeras. Här krävs en metoddiskussion om styrkan och svagheten i olika typer av forskningsmaterial och hur det väljs ut. De forskningsfält som arbetat med teknisk data i olika former kan fungera som inspiration och stöd för sådana diskussioner. I den här artikeln har frågan främst diskuterats utifrån digitala metoder som exempelvis datautvinning.

Med hjälp av de tre analytiska referenspunkterna skapas en beredskap för att närma sig digitala tjänster för insamling och analys av användargenererat material. Att studera användargenererat material i nyhetsjournalistiskt sammanhang kan innebära att undersöka exempelvis samspel mellan journalister och andra användare på Twitter. Det kan stå för en särskild ”genre” av användargenererat material som kanske rymmer förståelse för hur det offentliga samtalet sker på en arena som Twitter. Då görs först en analys av medielogiken som präglar Twitter och hur inbyggda användar-

personas strukturerar såväl journalister som andra Twitter-användare. Om syftet är att undersöka samspelet finns det möjlighet att göra en så kallad social nätverksanalys av det användargenererade materialet (se exempelvis Domínguez & Hollstein 2014). Här finns flera insamlingsverktyg att använda, baserade på öppen källkod, exempelvis Twitter Capture and Analysis. Verktyg(et) som samlar in data via Twitters API gör det möjligt att ladda ner och analysera stora tweet-samlingar. Verktyget kan också göra enkla mätningar av inlägg, svar, hashtags, länkar refererande, retweets med mera. Det är också kompatibelt med och kan användas med ett annat verktyg för social nätverksanalys, Gephi, som är än mer avancerat. Om det efterhand blir tydligt att studien kommer att hantera stora datamängder, och att visualisering av data anses vara en god hjälp för såväl analys som presentation, byts till verktyget Tableau. I en andra studie kanske intresset är det platsspecifika, vilket då innebär val av andra verktyg. Hela tiden är det dock viktigt att göra en kritisk granskning av verktygens inbyggda automatik, samt visa känslighet inför och förståelse för såväl urval som forsknings-etiska aspekter i det valda materialet.

Är det istället en studie av YouTube, kan forskningsfrågan exempelvis handla om visuellt användargenererat material och kontroverser kring det. Först krävs en teoretiskt grundad förståelse för vad kontroverser är och dess betydelse för människor och vårt samhälle (exempelvis Latour 1987 & 2005). Förståelsen ramas in av både de politiska och de populärkulturella kontroverser som kan finnas i det valda användargenererade materialet.

Betraktas det användargenererade materialet utifrån idén om persona blir de idealtyper av användare som YouTube byggt in i tjänsten intressanta.

Det användargenererade materialet kan även processas som data, exempelvis med syftet att kartlägga hur olika användargrupper agerar. Ett verktyg som Gephi kan användas för att visualisera de mönster som framträder. Eller så utforskas kombinationer av metoder där både olika former av data ingår, i en så kallad mixad metodansats som blir allt vanligare (se t ex Domínguez & Hollstein 2014).

Det jag argumenterar för i den här artikeln är värdet av att förstå metodutveckling genom analytiska referenspunkter. Metodutveckling för användargenererat material kräver diskussioner kring vad det är som efterfrågas i studierna och vad det är som analyseras, lika mycket som om vilka verktyg som kan användas.

Som all annan metodtillämpning krävs ett arbete för att situera, anpassa och rikta metoderna till det specifika sammanhanget. Den grundläggande kunskapsluckan som artikeln belyser utgörs av att medievetenskapliga metoder inte har situerats för användargenererat material i digitala kontexter. Vad de tre analytiska referenspunkterna gör, är att de fungerar som stöd åt det här arbetet. De bidrar med avgörande kunskap från relevanta fält för att begripliggöra användargenererat material i digitala kontexter. Det minskar risken att metodutveckling i digitala kontexter reduceras till ett val av verktyg. Det skapar grunden för ett öppet vetenskapligt prövande samt resonemang om metodval och dess konsekvenser.

Referenser

- Benhabib, Seyla (1992). *Situating the Self: Gender, Community and Postmodernism*. London: Routledge.
- boyd, danah & Ellison, Nancy (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1).
- Cambria, Erik, Schuller, Björn, Xia, Yunqing & Hvasi, Catherine (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems* 28(2).
- Corner, John (1991). Meaning, Genre and Context: The Problematics of 'Public knowledge' in the new audience studies, in Curran, James & Gurevitch, Michael (eds.) *Mass media and society*. London: Edward Arnold.
- Dahlgren, Peter (1988). What's the Meaning of This? Viewer's Plural Sense-Making of TV News. *Media, Culture and Society*, 10(3): 285-301.
- Domínguez, Silvia & Hollstein, Bettina (2014). *Mixed Methods Social Networks Research: Design and Applications*. Cambridge: Cambridge UP.
- Dunahee, Monica & Lebo, Harlan (2015). *The World Internet Project International Report 6th Edition*. World Internet Project.

- Dunham, Margaret (2003). *Data mining introductory and advanced topics*. Upper Saddle. River, N.J.: Prentice Hall/Pearson Education.
- Durham Peters, John (2015). *The marvelous clouds*. Chicago: University of Chicago Press.
- Dylko, Ivan (2014). Using Technological Attributes to Study Online Media: The Case of User-Generated Content. *Journal of Broadcasting & Electronic Media*, 58(4).
- Dylko, Ivan & McCluskey, Michael (2012). Media effects in an era of rapid technological transformation: A case of user-generated content and political participation. *Communication Theory*, 22: 250-278.
- Fayyad, Usama, Piatetsky-Shapiro, Gregory, & Smyth, Padhraic (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3).
- Foth, Marcus, Forlano Laura, Satchell, Christine & Gibbs, Martin (2011). *From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement*. Cambridge, MA: MIT Press.
- Goldkuhl, Göran (1993). Välgrundad metodutveckling. VITS Höstseminarium, 28-29/9. Tillgänglig på <http://www.vits.org/publikationer/dokument/479.pdf>. Hämtad den 12 oktober 2016.
- Gubrium, Aline & Harper, Krista (2013). *Participatory Visual and Digital Methods*. Walnut Creek: Left Coast Press.
- Hayles, Katherine (2004). Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis. *Poetics Today*, 25(1): 67-90.
- Hermes, Joke (1997). Gender and Media Studies: No Woman, No Cry, in Corner, John; Schlesinger, Philip & Silverstone, Roger (eds.) *International Handbook of Media Research*. London: Routledge.
- Hermida, Alfred & Thurman, Neil (2008). A clash of cultures: The integration of user-generated content within professional journalistic frameworks at British newspaper websites. *Journalism Practice*, 2(3).
- Holt, Kristoffer & Karlsson, Mikael (2015). "Random acts of journalism?": How citizen journalists tell the news in Sweden. *New Media & Society*, 17(11).
- ISO (2016), <http://it-ord.idg.se/ord/data/>, hämtad den 12 oktober 2016.
- Jensen, Klaus Bruhn (2012). Lost, Found, and Made: Qualitative Data in the Study of Three-Step Flows of Communication, pp. 435-450 in Ingrid Volkmer (ed.), *The Handbook of Global Media Research*. Hoboken, New Jersey: Wiley-Blackwell.
- Jensen, Klaus Bruhn (2014). Audiences, audiences, everywhere – Measured, interpreted and imagined, pp. 227-240 in Patriarche, Geoffroy; Bilandzic, Helena; Linaa, Jakob Jensen & Jurisic, Jelena (eds.) *Audience research methodologies: Between innovation and Consolidation*. London: Routledge.
- Juris, Jeffrey (2012). Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation. *American Ethnologist*, 39: 259-279.
- Jönsson, Anna-Maria & Örnebring, Henrik (2011). User-generated content and the news: Empowerment of citizens or interactive illusion?. *Journalism Practice*, 5(2): 127-144.
- Karlsson, Michael; Bergström, Annika; Clerwall, Christer & Fast, Karin (2015). Participatory journalism – the (r)evolution that wasn't. Content and user behavior in Sweden 2007-2013. *Journal of Computer-Mediated Communication*, 20: 295-311.
- Latour, Bruno (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, Mass.: MIT Press.
- Latour, Bruno (2005). *Reassembling the Social – An Introduction to Actor-Network-Theory*. New York: Oxford UP.
- Leung, Louis (2009). User-generated content on the internet: an examination of gratifications, civic engagement and psychological empowerment. *New Media & Society*, 11(8): 1327-1347.
- Lobato, Ramon; Thomas, Julian & Hunter, Dan (2011). Histories of User-Generated Content: Between Formal and Informal Media Economies. *International Journal of Communication*, 5: 899-914. <http://ijoc.org>.
- Long, Frank (2009). Real or Imaginary. The effectiveness of using personas in product design. *Irish Ergonomics Review*, Proceedings of the IES Conference 2009, Dublin.
- Marwick, Alice & boyd, danah (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, 16(7).
- Massanari, Adrienne (2010). Designing for imaginary friends: information architecture, personas and the politics of user-centered design. *New Media & Society*, 12(3).
- Meyrowitz, Joshua (1994). Medium Theory, p. 50-77 in Crowley, David J. & Mitchell, David (eds.) *Communication Theory Today*. Cambridge: Cambridge UP.
- Moens, Marie-Francine, Li, Juanzi & Chua, Tat-Seng (2014). *Mining User Generated Content*. Boca Raton, FL: Chapman and Hall/CRC.
- Nyre, Lars (2010). Experimenting with New Media for Journalism, in nord09media. Body Soul Society, *Nordicom Review*, (31)Special Issue.
- O'Connor, Barbara & Klaus, Elisabeth (2000). Pleasure and meaningful discourse: an overview of research issues, *International Journal of Cultural Studies*, 3(3).
- Pavlik, John & McIntosh, Scott (2014). *Converging Media: A New Introduction to Mass Communication* (4th edition). Oxford: Oxford UP.
- Pruitt, John & Adlin, Tamara (2006). *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. San Francisco, CA: Morgan Kaufman/Elsevier.
- Purdam, Kingsley & Elliot, Mark (2015). The changing social science data landscape, in Halfpenny, Peter & Procter, Rob (eds.) *Innovations in Digital Social Research Methods*. London: Sage.

- Rogers, Richard (2013). *Digital Methods*. Cambridge, MA: MIT Press.
- Snee, Helen (2015). Introduction to Part II Combining and comparing methods, in Roberts, Steven; Snee, Helene; Hine, Christine; Morey, Yvette; Watson, Hayley (eds.) *Digital Methods for Social Science: An Interdisciplinary Guide to Research*. Hampshire: Palgrave-Macmillan.
- Thurman, Neil (2008). Forums for citizen journalists? Adoption of user generated content initiatives by online news media. *New Media & Society*, 10(1).
- Turner, Phil & Turner, Susan (2011). Is stereotyping inevitable when designing with personas?. *Design Studies*, 32(1).
- Wahl-Jorgensen, Karin (2015). Cooptation, segregation and the boundaries of journalism, in Carlson, Matt & Lewis, Seth C. (eds.) *Boundaries of Journalism: Professionalism, Practices and Participation*. London: Routledge.
- van Dijk, José (2009). Users like you? Theorizing agency in user-generated content. *Media, Culture & Society*, 31(1).
- Wheeldon, Johannes & Ahlberg, Mauri (2012). *Visualizing Social Science Research: Maps, Methods and Meaning*. Thousand Oaks, CA: Sage.
- Östman, Johan (2012). Information, expression, participation: how involvement in user-generated content relates to democratic engagement among young people. *New Media and Society*, 14(6).

Pernilla Severson är lektor i journalistik och medieproduktion vid Linnéuniversitetet. Hon forskar om medieutveckling och deltagarorienterade processer och har studerat: televisionens digitalisering, lokala innovationspraktiker i mediebranschen och digitala medieprofessioners identitetsskapande.

Humanistlaboratorier i Norden

Humlab vid Umeå universitet

Stefan Gelfgren, Föreståndare, Humlab, Umeå universitet

Humlab är en fristående arbetsenhet vid Humanistiska fakulteten, Umeå universitet. Enheten har funnits sedan 2001, och får väl därmed anses vara en av Sveriges tidigaste miljöer för fältet som numera brukar benämnas digitala humaniora (ett begrepp som tål att diskuteras då det snarare handlar om ena humaniora i en digital tid, än ett sub-område inom humaniora). I uppdraget från fakulteten nämns att Humlabs syfte är ”att initiera, inspirera och utveckla relationen mellan humaniora, kultur och informationsteknik inom forskning, forskarutbildning och undervisning” – och det är inom dessa områden fokus ligger. Det är viktigt att Humlab kan vara en mötesplats mellan olika discipliner och kategorier, och flervetenskaplighet och öppenhet är en ambition som kännetecknar verksamheten, samt det internationella inslaget. Samtidigt är (och har) samverkan med närings-, kultur- och föreningsliv varit ett viktigt spår inom verksamheten genom åren. Verksamheten finansieras dels av anslag, dels av externa projektbaserade medel.

Humanistiska fakulteten har de senaste åren gjort en satsning mot digitala humaniora där doktorander med digital profil anställts, Humanisthuset har byggts om och utformats för undervisning med en högre grad av teknisk slagkraft, och inslaget av nätkurser har ökat – och i den satsningen har Humlab med sin kompetens utgjort en viktig resurs.

Forskningsmässigt finns det bland de egna forskarna å ena sidan ett tydligt humanistiskt, kritiskt och kvalitativt anslag med forskning på digitala miljöer och praktiker, och därtill med digitala metoder och verktyg. Det handlar bland annat om att studera språkrevitalisering,

’patient-empowerment’ inom hälsovården, kyrka och religiös auktoritet, representation av antiken i spel och film, och e-avfallshantering – för att nämna några områden. Å andra sidan är Humlab genom sin kompetens inom teknisk/digital utveckling inblandad som part i ett flertal forskningsprojekt där man tillsammans inom projekten utvecklar nya metoder och verktyg. I de fallen har det exempelvis handlat om att bygga och driva databaser i olika former – bland annat arkeologiska, språkliga, och i anslutning till arkivsektorn, ofta också med en GIS-/kartkomponent. Man har också utvecklat verktyg för digitalisering, textbearbetning och textanalys för stora textflöden (text kan här också omfatta till exempel bild och ljud) i externa projekt.

Humlabs pedagogiska verksamhet riktar sig till lärare och studenter, genom kursplanering, undervisning, handledning och seminarier. Humlab har exempelvis ett antal undervisningsåtaganden, en del tillsammans med program och kurser vid fakulteten, inom vilka Humlab bidrar med digital kompetens i relation till kursplaner, pedagogik i allmänhet, examinationer och mer utåtriktade aktiviteter. Viktigt i det sammanhanget är att Humlab erbjuder pedagogiskt tekniskt stöd, med bland annat tillämpning av verktyg och metoder utifrån ämnesspecifika teoretiska och metodologiska perspektiv. I Humlabs regi ges också varje år ett så kallat kortkursprogram (kurser om 3 timmar styck som ger inblick i specifika ämnen eller verktyg). Humlab har också gett, och ger, doktorandkurser inom digitala humaniora som är både av teoretisk och praktisk karaktär – och stödjer på så sätt doktoranderna i att utveckla metoder, verktyg och analys.

HELDIG Helsinki Centre for Digital Humanities

Eero Hyvönen, Director, Heldig, Helsinki Centre for Digital Humanities

Computational methods provide ever more new research opportunities in the humanities and social sciences, and digitalisation is rapidly changing the ways in which research is done. Digitalisation and its applications, e.g. the web with its data, services and applications, are closely related to the challenges of globalisation and are changing society in fundamental ways. These changes create a growing need for research and education in multidisciplinary digital humanities.

To address these challenges, a new centre Heldig for Digital Humanities has been established in Finland at the University of Helsinki (UH). Heldig is a research network and infrastructure for solving research problems in the humanities and social sciences with novel computational methods, and for studying digitalisation as a phenomenon. Heldig also supports education in digital humanities and application development. The goal is to form an ecosystem that fosters collaboration

between academic research groups, memory organisations, companies, and the general public.

The challenges of digitalisation are tackled by refocusing, joining, and coordinating the intellectual and other resources in the humanities, social sciences, and computer science across faculties, universities, and other stakeholders in the Helsinki capital area. Heldig is a major Finnish strategic effort for 2016–2020 in order to take this step towards the digital future. As part of the activities, eight new tenure-track professor positions in DH-related fields will be established at UH in seven different faculties during 2017–2020. In addition, Heldig is advancing several current strengths and initiatives at the University of Helsinki and beyond in the areas of natural language processing, computer-aided linguistics, research on the social and psychological aspects of digitalisation, smart cities, quantitative social sciences, and digitised cultural heritage.

Humanistlaboratoriet vid Lunds universitet

Marianne Gullberg, Föreståndare, Humanistlaboratoriet vid Lunds universitet

Humanistlaboratoriet vid Lunds universitet är en forskningsinfrastruktur som erbjuder forskare i humaniora och teologi fri tillgång till avancerade mätinstrument (t.ex. ögonrörelsemätning, EEG, motion capture, o.s.v.) och tillhörande metodologisk kompetens, forskningsförberedande utbildning i dessa tekniker och metoder, samt expertis kring datahantering. Labbets uppdrag är att stärka och diversifiera humanistisk forskning genom att erbjuda möjligheter till kvantitativ och experimentell forskning som ett komplement till mera traditionella kvalitativa metoder. Forskningen som bedrivs i labbet rör främst mänsklig kognition, kommunikation och kultur, men många projekt är tvärvetenskapliga och bedrivs tillsammans med samhällsvetenskap, medicin, naturvetenskap, teknik och e-

vetenskap. Labbet har därför användare från många fakulteter vid Lunds universitet.

Organisatoriskt är Humanistlaboratoriet placerat direkt under de Humanistiska och teologiska fakulteterna (HT). Humanistlaboratoriets leds av en föreståndare (med delegation från dekanus) som också är ordförande i styrgruppen. Verksamhet och tillgänglighet regleras av riktlinjer och användaröverenskommelser som specificerar ansvar och plikter. Labbets verksamhet rapporteras i årsrapporter riktade till det nationella och internationella vetenskapssamhället.

Humanistlaboratoriet är bas för världsledande forskning. Labbet står värd för flera högt kvalificerade projekt och grupperingar, bl.a. en Linnémiljö, delar av ett strategiskt forskningsområde (e-vetenskap/ eSENCE),

ett ERC starting grant, en Wallenberg Scholar, en Wallenberg Academy Fellow, ett Jubileumsanslag från Riksbankens Jubileumsfond samt ett par andra KAW/MAW-projekt. Humlabbet har spelat en avgörande roll vid tilldelningen och placeringen av dessa stora prestigefyllda anslag som annars sällan går till forskare på HT-områdena. Dessa projekt rör t.ex. förhållandet mellan språk, tanke och de landskap vi bebor (ERC *Language, Cognition, Landscape*) där utrotningshotade språk också dokumenteras för eftervärlden. Andra projekt studerar hur studenter bearbetar bild och textmaterial genom att mäta deras ögonrörelser (MAW *EyeLearn*), eller mäter hur hjärnan hos språkinlärare bearbetar nya betydelser och hur minnet då påverkas (KAW *Culture,*

brain, learning). I det Svenska Pompejiprojektet har Humanistlaboratoriet scannat ruiner och byggnader, och byggt 3D-modeller av dem som sedan kan levandegöras i virtuell verklighet, och labbet har också stått värd för ett projekt om relationen mellan kördirigenter och körsångare (*Avaco*).

Humanistlaboratoriet ingår även i nationella och internationella nätverk som t.ex. *Swe-Clarín*, ett svenskt konsortium för språkteknologi (som i sin tur är en del av det europeiska *Clarín*). Härigenom stöder labbet projekt genom att bl.a. utveckla specialiserade digitala lexikon, verktyg för automatisk textanalys samt system för audiovisuell analys, t.ex. för att länka huvudrörelser till framhävda ord i talet.

HUMLab på Det Humanistiske Fakultetsbibliotek i København

Lars Kjær, Informationsspecialist, HUMLab,
Det Humanistiske Fakultetsbibliotek i København

HUMLab er placeret midt på campus på Københavns Universitet Amager og blev åbnet i foråret 2016 for at bidrage til at udvikle et fagligt miljø omkring digital humaniora (DH) og for at styrke bibliotekets position som center for viden og information på campus. Labbet er med sine cirka 50 m² relativt lille, men vi forventer at yde et stort bidrag ved at samle folk fra små spredte miljøer til fælles aktiviteter.

Af hardware har vi stærke computere af typen HP Envy Phoenix 860-002NO, en Next Engine 3D scanner, to flatbedscannere fra Epson, et webcam, en mikrofon og fire hovedtelefoner. Af software har vi Agisoft PhotoScan, Blender og Meshlab til 3D scanning, Abbyy FineReader Pro 12 til OCR scanning, Adobe programmer til foto og videoeditering, Audacity og Camtasia til pod- og videocasts samt en række programmer til digital tekstanalyse og visualisering, f.eks. Gephi, NVivo, Python, OpenRefine og Qgis.

Vi planlægger aktiviteter i samarbejde med afdelinger inden for Det Kongelige Biblioteks Nationalbibliotek, Københavns Universitet og IT-Universitet. Vi har to former for aktivi-

teter: events og studiekredse. Events er f.eks. datasprints med udgangspunkt i omfattende digitale samlinger af kort og fotografier fra de Dansk-Vestindiske øer; en konference om Spatial Humanities og Qgis samt workshops om Netarkivet. Studiekredsene handler f.eks. om 3D modellering, WordPress, digital tekstanalyse med Python og OpenRefine, remixing med Photoshop og georeferering med Qgis.

Udviklingen inden for DH foregår forskellig steder på campus, og vi forsøger at samle entusiastiske ansatte og studerende, der har en interesse i at indgå i et fællesskab med det formål at realisere egne projekter i samarbejde og på lige fod med andre. Vi udvikler ikke egne projekter, men bidrager til projekter, som ansatte og studerende selv har en interesse i at udvikle. På denne måde mener vi, at vi bedst bidrager til at udvikle et miljø og sprede interessen for DH med de midler, vi har til rådighed.

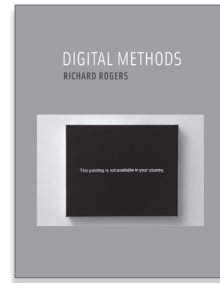
Vi tror på, at vi på kan samle et miljø, styrke bibliotekets position som center for viden og information på campus og samtidig bidrage til at understrege bibliotekets rolle på fremtidens universitet.

Recension

Richard Rogers

Digital Methods

Cambridge & London, MIT Press, 2013, 280 s.



Den allra första meningen i Richard Rogers bok *Digital Methods* introducerar läsaren för ett motsägelsefullt tillkännagivande: "This is not a methods book". Snarare, menar han, presenterar den en "methodological outlook for research on the web." Rogers kontrasterar lättheten i sina 'metodologiska utblickar' med sinnebilden av blytung klassiska metodpaket. En otymplig verktygslåda får assistera som en metafor för traditionella forskningspraktiker, och istället erbjuder Rogers ett smörgåsbord av påhittiga (och presumtivt enkla) sätt att studera det digitala.

Om Rogers aviga förhållningssätt till begreppet 'metod' handlar om blygsamhet eller begreppslig olust förblir osagt, men nog handlar hans bok – som till stor del är en sammanfattning av hans egen forskning sedan drygt 15 år tillbaka – om digitala metoder. I boken får vi exempelvis lära oss mer om Rogers experiment med att skapa dokumentära filmer av Googles hemsideshistoriska evolution mellan åren 1998-2007; hans djupdykningar i Internetarkiv för att utforska den tidiga Holländska blogosfärens utbredning; och hans strategier för att genomföra vad som liknas vid hälsokontroller av nationella Internet(s), som den Iranska webben (ett projekt som till stor del handlat om att studera förekomsten av Internetcensur, och i vilken utsträckning hemsidor fungerar och underhålls).

Boken redogör också för hur kunskapsproduktion på Wikipedia kan beforskas genom att spåra hur dess artiklar växer fram, redigeras och skiljer sig åt mellan olika nationella kontexter. Mer specifikt har Rogers jämfört hur skildringar av massakern i Srebrenica år 1995 har skrivits fram på det Holländska, Serbiska och Bosniska Wikipedia, och därmed kastat ljus över historieskrivningars politiska dimensioner i en digital kontext. Boken, som är indelad i nio kapitel, beskriver även hur Rogers har kartlagt 'avstånd' mellan olika hemsidor genom att spåra förekomsten av hyperlänkar på olika sajter och var de leder. Därigenom har bland annat företags och statliga organisationers associationsförmågor och sociala nätverk studerats, vilket inte minst berör frågor kring inkluderingens och exkluderingens politik på Internet.

Rogers, som tidigare har gjort sig känd för att bygga upp ett av Europas ledande centrum för digital metodutveckling, Digital Methods Initiative vid Amsterdams Universitet, ägnar dock inte hela boken åt att sammanfatta tidigare forskning. Han tar också ett bredare grepp om teoretiska och metodologiska frågor. Till exempel resonerar Rogers kring vikten av samtida studier som 'följer mediet' och återanvänder dess metoder. Snarare än att uppfinna nya forskningsverktyg, menar Rogers att det finns anledning att testa och utforska redan existerande digitala redskap, plattformar och funktioner, och att tänka jämte dem när nya sätt att bedriva Internetforskning utvecklas. De funktioner som finns på nätet ses därmed

Recensent: **Maria Eriksson**, Doktorand, Medie- och kommunikationsvetenskap / HUMlab, Umeå Universitet.

inte bara som studieobjekt, utan också som forskningsverktyg som kan manipuleras för att skapa ny kunskap.

Rogers har också varit delaktig i flera forskningsprojekt som mynnat ut i nya verktyg för digitala studier, såsom Issue Tracker (ett program som spårar icke-statliga organisationers kampanjbeteenden på nätet), och Election Issue Tracker (en mjukvara som spårar medieaktörers inverkan på spridningen av populism). Dessa verktyg följer delvis en sociologisk och journalistiskt orienterad medie-forskningstradition och är – tillsammans med en rad andra nedladdningsbara program – fritt tillgängliga på nätet.¹

I boken förklarar Rogers även att han ägnar sig åt 'webbepistemologiska' frågeställningar, vilket utgår ifrån en tanke om att Internet huserar en kunskapskultur som skiljer sig från andra mediers kulturella logiker. Här blir det för Rogers därför viktigt att specifikt studera hur digitala objekt (exempelvis länkar, taggar, bokmärken och IP-adresser) hanteras på nätet. Rogers väljer emellertid inte bara att fokusera på digitala föremål – hans forskning grundas också i en större ambition att bedriva "Internetforskning som inte bara handlar om Internet", som Rogers beskriver det själv. Storskaliga samhällsfrågor hamnar därför också ofta i centrum för hans projekt, som beskrivs som ett steg bort från studier av så kallade 'onlinekulturer' (ett ämne som mycket av den tidiga Internetforskningen fördjupat sig i – ofta i mer eller mindre exotifierande ordalag). I *Digital Methods* lyfter Rogers istället fram möjligheten att "diagnostisera kulturell förändring och samhällsliga förhållanden med hjälp av Internets medel" (s. 21).

Här blir det också tydligt var Rogers riktar sin politiska udd. Hos Rogers är det inte enbart mjukvara och teknik som granskas ur en kritisk vinkel – fokus riktas lika mycket utåt, mot politiska tendenser bland Internetanvändare och samhället i stort. På så sätt skiljer sig också hans forskning från andra framväxande fält som mediearkeologi och mjukvarustudier – discipliner där tekniken själv ofta hamnar helt i fokus. Rogers tar delvis också avstånd från litteraturvetenskapligt orienterade metodologiska

grenar som 'distansläsning' genom att främst förespråka studier av material som är 'fött digitalt', snarare än material som har digitaliserats (vilket traditionella 'distansläsare' ofta intresserat sig för). Där andra väljer att zooma in på texter, plattformar eller hemsidefunktioner, intresserar sig således Rogers lika mycket för att spåra och mäta handlingar hos de människor som befolkar Internet i sin helhet.

En sådan utgångspunkt innebär givetvis att avkall görs på somligt, och ofta är det djupgående filosofiska och vetenskapsteoretiska problematiseringar som går förlorade i Rogers resonemang. Ur ett metodologiskt perspektiv saknas exempelvis kritiska diskussioner kring vad forskningsmässiga approprieringar av kommersiellt genererad mjukvara och Internetdata egentligen får för politiska och ideologiska konsekvenser. Draget till sin spets, så innebär Rogers argument om att använda 'mediets metoder' att forskare ofrånkomligen hamnar i en beroendeställning gentemot kommersiella aktörers onlineverktyg. Från ett kritiskt perspektiv har också forskare som Jim Thatcher (2014) framhållit att studier som godtar, använder och utgår ifrån app- och plattformsgenererad data implicit (och oundvikligen) accepterar en kommodifiering och kvantifiering av subjekt, men om problematiken kring detta säger Rogers mycket lite. Kanske är det dock digitalt orienterade forskjutningar av hur vi mäter, spårar och förhåller oss till mänskligt handlande som dagens forskning i huvudsak borde närma sig kritiskt, snarare än att reproducera?

Digital Methods lämnar också en av den digitalt orienterade forskningens mest ödesmättade frågor därhän – frågan kring tillgång till Internets innehåll. Att mycket av Rogers egen forskning rör sig i en legal gråzon problematiseras exempelvis endast mycket kortfattat, trots att en av Rogers mest frekvent använda metoder – storskaliga skrapningar av webbsidor – ofta bryter mot Internetaktörers användaravtal. Vid ett av de (unika) tillfällen där detta diskuteras i boken, förklarar Rogers hur han och hans kollegor valt att fortsätta hämta hem information om Googles sökesultat även om företaget strypt det API som

under en period gav dem legitim tillgång till data. Enligt Rogers åtföljdes detta av en postum notifiering som förklarade att forskarna "uppskattar Googles överseende" vad gäller deras datainsamling (s. 31) – ett metodologiskt tillvägagångssätt som ger tvivelaktiga svar på hur tillgång till kommersiellt rotad data (d.v.s. en klar majoritet av allt material på nätet) egentligen bör hanteras.

Här hade det varit önskvärt att en av Europas främste forskare på det digitala området tagit tillfället i akt att på allvar lyfta frågor kring forskningsmässigt tillträde till Internet-innehåll – en fråga som är fullständigt central för alla som studerar kommersiella aktörers aktiviteter på Internet. I min egen forskning, som bland annat utforskar de algoritmiska system som styr digitala plattformar för musik, breder exempelvis gråzonerna för vilken data som egentligen är tillåten att hämta hem snabbt ut sig. Och vad får det för konsekvenser när allt mer av våra liv förflyttas till det digitala, samtidigt som den metodmässigt digitalt orienterade – och laglydige – forskarens händer är baktvunna av begränsande använ-

daravtal? Är lösningen att, som Rogers, i efterhand ödmjukt be om ursäkt för sin insamling av data och hoppas på att Google och andra nätjättar förbarmar sig över forskare? Eller är det dags att det förs en bredare debatt kring metodologiska förutsättningar för kritisk humanistisk och samhällsvetenskaplig forskning i digitala kontexter?

Trots ovan nämnda luckor är dock *Digital Methods* en bok som genom detaljerade casestudier ger många goda exempel på hur digitala verktyg kan öppna upp för forskning som metodmässigt närmar sig de skolor, hastigheter och beteendemönster som i övrigt präglar Internet. Även om det är svårt att närma sig Rogers drygt tre år gamla metodbeskrivningar som regelrätta recept på hur digitala forskningsexperiment kan kokas ihop (webbens föränderlighet är, som alltid, ett gissel för Internetforskare), så håller boken som en rik inspirationskälla för den som är intresserad av digital metodutveckling. Och kanske är det just Rogers kreativa och experimentella metodologiska ingångar som är bokens allra största styrka.

Not

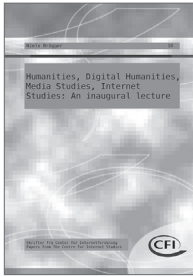
1. Se <https://wiki.digitalmethods.net/Dmi/ToolDatabase> (besökt 2016-06-16)

Referens

- Thatcher, Jim (2014). Digital Footprints, Data Fumes and the Limitations of Spatial Big Data. *International Journal of Communication*, 8: 1765-1783.

Bøger om digitale metoder

Et udvalg 2013-2016

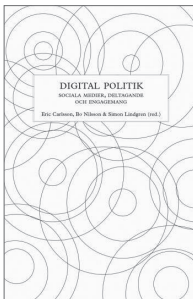


Humanities, Digital Humanities, Media Studies, Internet Studies

An Inaugural Lecture

Niels Brügger (2015). Aarhus Universitet, Center for Internetforskning. (Skrifter fra Center for Internetforskning, vol. 16), 16 s.

Today's expanding digital landscape constitutes an important research object as well as the research environment for the Humanities at the beginning of the 21st century. Taking this state of affairs as a starting point this inaugural lecture presents a vision for how the digital affects the interplay between four areas which until now to a certain extent have been separated: Traditional Humanities, Digital Humanities, Media studies, and Internet studies. The vision is followed by an outline of how it can be unfolded in concrete activities, in the form of research projects, research infrastructure, and research organisation.



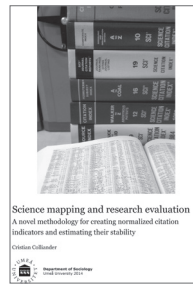
Digital politik **Sociala medier, deltagande och engagemang**

Eric Carlsson, Bo Nilsson & Simon Lindgren (red.) (2015). Göteborg: Daidalos, 213 s.

Digital politik handlar om hur sociala medier kan fungera politiskt i människors vardag. Digitaliseringen av samhället har

starkt aktualiserat frågan om förhållandet mellan medier och deltagande. I Sverige har exempelvis de olika politiska partierna lagt mycket resurser på synlighet, inte minst via sociala medier som Facebook, Twitter och Instagram. Sociala nätverksmedier medför även möjligheter till en mer vardaglig form av engagemang och deltagande i en mängd olika politiska och sociala frågor. Boken innehåller bidrag av Jonas Andersson

Schwarz, Linda Berg, Maria Carbin, Eric Carlsson, Stefan Gelfgren, Jenny Gunnarsson Payne, Anders Johansson, Anna Johansson, Anders Olof Larsson, Simon Lindgren och Bo Nilsson.



Science Mapping and Research Evaluation

A Novel Methodology for Creating Normalized Citation Indicators and Estimating their Stability

Cristian Colliander (2014). Umeå University, Department of Sociology, (Akademiska avhandlingar vid Sociologiska institutionen; 76) Dr.diss, 37 p.

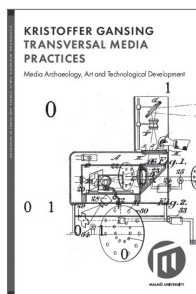
The purpose of this thesis is to contribute to the methodology at the intersection of relational and evaluative bibliometrics. Experimental investigations are presented that address the question of how we can most successfully produce estimates of the subject similarity between documents. The results from these investigations are then explored in the context of citation-based research evaluations in an effort to enhance existing citation normalization methods that are used to enable comparisons of subject-disparate documents with respect to their relative impact or perceived utility.



Digitaalinen humanismi ja historiatieteet **[Digital humanities and science of history]**

Kimmo Elo (ed.) (2016). Turku: Turun historiallinen yhdistys (Historia mirabilis; 12), 256 p.

Computer mediated research methods have become a part of history science, as well. The authors of the book writes about text mining, network analyses and increased reality. Concrete case studies expose how digital data and digital methods have been used.

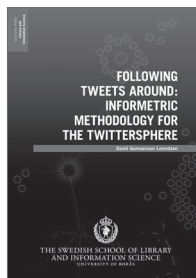


Transversal Media Practices

Media Archaeology, Art and Technological Development

Kristoffer Gansing (2013). Malmö University, School of Arts and Communication, K3 (New Media, Public Spheres, and Forms of Expression), Dr.diss., 350 p.

Transversal Media Practices work across specific situations of technological development, critically examining and redefining the terms of production in different media by bringing heterogeneous histories, institutions, actors and materialities into play with one another. This dissertation is all about trying out and refining the methodologies of such transversal media practices, in the end outlining a conceptual set of tools for further development. Following the technological hype of the "digital revolution" of the mid-1990s, the field of new media studies gained popularity over a ten year period. This dissertation takes its cue from a historical turn in new media theory, and argues that it is time leave behind strict polarisations between old and new as well as analogue and digital.



Following Tweets Around

Informetric Methodology for the Twittersphere

David Gunnarsson Lorentzen (2016). University of Borås, The Swedish School of Library and Information Science (Skrifter från Valfrid; 61), Dr.diss., 151 p.

The purpose of this thesis is to critically discuss methods to collect and analyse data related to the interaction and content on the social platform Twitter. The thesis contains examples of how networked communication can be studied on Twitter, based on the affordances of the platform considering interaction with interfaces and other users. The thesis contains examples of how networked communication can be studied on Twitter, based on the affordances of the platform considering interaction with interfaces and other users. The thesis consists of a summary essay and five articles.



Otteita verkosta

Verkon ja sosiaalisen median tutkimusmenetelmät [Grip on the Internet: research methods for studying net and social media.]

Salla-Maaria Laaksonen, Janne Matikainen & Minttu Tikka (eds.) (2013). Tampere: Vastapaino, 376 p.

Data drawn from Internet has caused questions and wonderment for researchers used to traditional surveys or printed materials. At the same time, Internet can be a data source, a place to do research, and an object for research. The book is written to answer for some of the methodological and other questions, such as differences between different net questionnaires, ethical challenges in social media research, or how to do network analyses on the net. Authors of the chapters are the following: Salla-Maaria Laaksonen, Janne Matikainen, Minttu Tikka, Susanna Paasonen, Riikka Turtiainen, Sari Östman, Pekka Räsänen, Outi Sarpila, Jani Miettinen, Kimmo Vehkalahti, Arto Selkälä, Jaakko Suominen, Hannakaisa Iso-mäki, Tiina-Riitta Lappi, Johanna Silvennoinen, Johanna Sumiala, Salli Hakala, Juho Vesa, Jukka Huhtamäki, Olli Parviainen, Antti Ukkonen, Frans Mäyrä, Vilma Lehtinen, Airi Lampinen, and Emmi Suhonen.



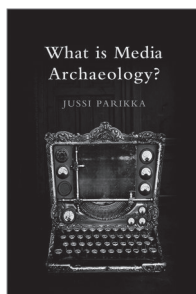
Forskerbrug af webarkiver

En kort indføring

Janne Nielsen (2015). Aarhus: NetLab, 50 s.

Denne bog er udformet i forbindelse med udviklingen af NetLabs forskerworkshops om webarkivering. I disse workshops introduceres deltagerne til at arbejde

med arkiveret web i forskning, herunder hvad webarkivering er, udfordringerne ved arkiveret web som forskningsobjekt, viden om eksisterende webarkiver og om værktøjer til mikroarkivering, så forskeren selv kan arkivere web. Formålet med denne bog er i forlængelse heraf at samle al relevant viden om forskerbrug af webarkiver, som er relevant i en dansk kontekst og tilpasset dansk forskning. Bogen vil løbende blive opdateret med relevant ny forskning inden for webarkivering.



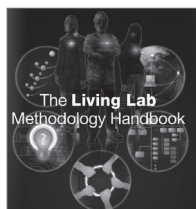
What is Media Archaeology?

Jussi Parikka (2012). Chichester: Wiley-Blackwell, 200 p.

The book offers an introduction to the emerging field of media archaeology and analyses the innovative theoretical and artistic methodology used to excavate

current media through its past.

What is Media Archaeology? examines the theoretical challenges of studying digital culture and memory and opens up the sedimented layers of contemporary media culture. The author contextualizes media archaeology in relation to other key media studies debates including software studies, German media theory, imaginary media research, new materialism and digital humanities.



The Living Lab Methodology Handbook

Anna Ståhlbröst & Marita Holst (2013). Stockholm: Vinnova, 68 p.

Most innovations come from gaps between an existing product and

customers' expectations. The technological factor is only one element of the innovation. Other elements can be better working conditions or methods of service delivery that may, or may not, have a technological component

This book is based on results from the collaboration within the project SmartIES and the process of using and evaluating the FormIT methodology in a Nordic cross-border pilot.

The goal has been to make the Living Lab Key Principles and the application of them more visible and easy to use.

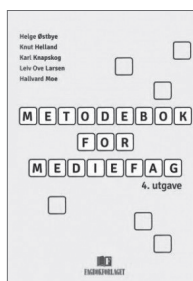


Between Humanities and The Digital

Patrik Svensson & David Theo Goldberg (eds.) (2015). Cambridge: MIT Press, 574 s.

Between Humanities and the Digital offers an expansive vision of how the humanities engage with

digital and information technology, providing a range of perspectives on a quickly evolving, contested, and exciting field. It documents the multiplicity of ways that humanities scholars have turned increasingly to digital and information technology as both a scholarly tool and a cultural object in need of analysis. The contributors explore the state of the art in digital humanities from varied disciplinary perspectives, offer a sample of digitally inflected work that ranges from an analysis of computational literature to the collaborative development of a 'Global Middle Ages' humanities platform, and examine new models for knowledge production and infrastructure.



Metodebok for mediefag

Helge Østbye, Knut Helland, Karl Knapskog & Leif Ove Larsen (2013). Bergen: Fagbokforlaget, Utg. 4., 319 s.

Boken gir en innføring i sentrale forskningsmetoder i mediefaget og viser bredden i metodiske

tilnærminger: Den presenterer både metoder som benyttes innenfor andre humanistiske og samfunnsvitenskapelige fag, og metoder som er spesielle for mediefag. Forfatterne tar opp temaer som strekker seg fra vitenskapsteoretiske grunnlagsproblemer til konkrete metoder som kan brukes for å analysere ulike medier. Den nye utgaven er gjennomgående oppdatert og med behandling av de nyeste medietekstene.

NORDICOM

Nordiskt Informationscenter för Medie- och Kommunikationsforskning



norden

NORDICOM är en institution inom
Nordiska Ministerrådet

Nordicom är ett nordiskt kunskapscenter för medie- och kommunikationsområdet. Med utgångspunkt i forskningen insamlar, bearbetar och förmedlar Nordicom kunskap till olika brukargrupper i Norden, Europa och övriga världen. Verksamheten bygger på kontakter med en rad intressenter: forskare, medieföretag, politiska beslutsfattare, myndigheter, organisationer, lärare, bibliotek med flera, inte bara i de nordiska länderna utan över hela världen.

Nordicoms arbete syftar till att utveckla kunskapen om medier i samhället och bidra till att forskningens resultat synliggörs i behandlingen av mediefrågor på olika nivåer i både offentlig och privat verksamhet. Verksamheten utmärks av förmedling inom tre områden:

Medieforskningen och dess resultat i de nordiska länderna

Nordicom utger tidskrifter och böcker, främst antologier, på nordiska språk och engelska. På så sätt erbjuder Nordicom de nordiska medie- och kommunikationsforskarna kanaler både i Norden och globalt. En engelskspråkig referee-granskad tidskrift, *Nordicom Review*, utges med stor spridning. Tidskriften *Nordicom-Information* publiceras på nordiska språk. En forskningsdatabas för forskare, litteratur och pågående forskning uppdateras löpande. Verksamheten är uppbyggd kring dokumentationscentraler i de nordiska länderna.

Medieutvecklingen och medietrender i de nordiska länderna

Nordicom utarbetar nordisk mediestatistik och gör kvalificerade analyser, bland annat i skriftserien *Nordic Media Trends*. Genom deltagande i flera europeiska och internationella nätverk/organisationer och institutioner som arbetar med medie- och kulturpolitiska frågor, ger Nordicom de nordiska länderna en samlad röst. Samtidigt samlar Nordicom in relevant omvärldskunskap från regionala och internationella arenor för vidareförmedling till nordiska brukare. Arbetet är uppbyggt kring ett nordiskt nätverk. Därtill dokumenterar den svenska delen av Nordicom på uppdrag av Kulturdepartementet den svenska medieutvecklingen, som bland annat innefattar studier av den svenska mediemarknaden och den årliga räckviddsundersökningen *Mediebarometern*.

Forskning om barn, unga och medier i världen

Nordicom startade 1997 på uppdrag av UNESCO The International Clearinghouse on Children, Youth and Media. Arbetet syftar till att öka kunskapen om barn, ungdomar och medier och därmed ge underlag för beslut, bidra till en konstruktiv samhällsdebatt samt främja barns och ungdomars mediekunnighet (media and information literacy) samt att stimulera till vidare forskning på området. Verksamheten är uppbyggd kring ett globalt nätverk med 1000-talet deltagare som representerar inte bara forskarsamhället utan även till exempel mediebranschen, politiken och frivilliga organisationer. En årsbok, rapporter och antologier utges.

www.nordicom.gu.se

NORDICOM

Nordiskt Informationscenter för Medie- och Kommunikationsforskning

Göteborgs universitet
Box 713, SE 405 30 Göteborg
Telefon +46 31 786 00 00 • Fax + 46 31 786 46 55
e-post info@nordicom.gu.se
www.nordicom.gu.se

ISBN 978-91-87957-45-1



9 789187 957451



Norden



GÖTEBORGS UNIVERSITET