

Jonas Svensson

Temasektion

Mönster i rappakaljan

Ett test med datorassisterad, urvalsgenererande fjärrläsning av automatiskt transkriberade salafistiska predikningar på svenska

Nøgleord

Abstract Föreliggande artikel är resultatet av ett test som undersöker om det går att använda ett högst bristfälligt resultat av automatisk transkribering via Googles tjänst Speech-to-Text av 208 inspelade föreläsningar hämtade från den svenska salafistiska hemsidan Islam.nu för att finna återkommande mönster. Syftet är att få en översikt över innehållet i ett material som är både omfattande och i sitt ursprungsformat svåröverskådligt. Artikeln är en genomgång steg för steg av en process som kombinerar datorassisterade metoder för fjärrläsning och nätverkanalys. Experimentets resultat tyder på att datorassisterad analys av automatiskt transkriberade texter kan fylla en funktion framför allt som en del i en process att identifiera relevant material för djupstudier.

Den grupp av svenska salafistiska aktivister som har skapat och administrerar webbsidan *islam.nu* benämns ibland “Järva-salafisterna”, efter den förort till Stockholm där de främst är verksamma. Gruppen har en tydligt utåtriktad och missionerande agenda. Flera studier av svensk salafistisk aktivism på senare tid har identifierat denna grupp som inflytelserik och populär i vad som kanske kan betecknas som ett större svenskt salafilandskap. Såväl deras metoder för utåtriktad verksamhet som innehållet i deras mission har varit föremål för analys (Olsson, 2020; Ransorp et al., 2018: 139–140; Sorgenfrei, 2021a, 2021b, kommande).

Till den utåtriktade verksamheten hör produktion och distribution av ett omfattande digitalt material på svenska, i olika

Jonas Svensson is a professor in the study of religion specialized within Islamic studies at the Linnaeus University. Early in his career, he has primarily been interested in gender and Islam, modernist Islam interpretations, Islamic education in East Africa, and coultural historical perspectives on the Quran and Mohammad. Lately, he has primarily been occupied with theory and method in the study of Islam and more specifically the cognitive science of religion and its consequences for the study of Islam; in addition to this he is exploring the potentiallity of computer assisted methids in humanistic research

format och via olika plattformar. De senare innefattar ovan nämnda webbsida, en YouTube-kanal, flera Instagramkonton och en Facebooksida. Det digitala materialet består till exempel av texter (e-böcker och blogginlägg) och videoinspelningar, både av längre föreläsningar och av kortare uppmaningar riktade direkt till en publik av muslimska “bröder och systrar”. På hemsidan finns också en stor uppsättning ljudfiler med inspelade föreläsningar. Det är detta senare material som behandlas i föreliggande artikel.

Materialet – potential och problem

Vid en inventering den 15 februari 2021 fanns 1 566 föreläsningar i mp3-format tillgängliga för nedladdning på www.islam.nu. Enligt uppgift är den första publicerad den 11 april 2009 och den senaste den 1 februari 2021. Ett slumpvis urval av 1 procent av dessa (15 föreläsningar) hade en genomsnittlig längd på 54 minuter. Även om det naturligtvis ligger en stor osäkerhet i att generalisera utifrån en så liten andel,¹ pekar detta på en total mängd inspelat material på närmare 1 400 timmar. Med andra ord utgör enbart dessa föreläsningar en mycket stor mängd primärmaterial om samtida salafistisk aktivism i Sverige.

Som forskningsmaterial är ljudfilerna tacksamma. De är lättåtkomliga och kan enkelt lagras externt, vilket gör att man kan undvika ett av de stora problemen med att använda webbmaterial, nämligen dess benägenhet att plötsligt, utan förvarning, försvinna. Till yttermera visso är de “naturalistiska data” (Potter & Shaw, 2018). Materialet är, som helhet betraktat, och till skillnad från “frambringad data” som till exempel intervjuer eller enkäter, oberoende av forskarens aktivitet.

Mängden av dylikt naturalistiskt material, relevant för islamologer, har växt kraftigt som ett resultat av modern kommunikationsteknologi, och fortsätter att expandera i en rasande fart. En av de tidigaste forskarna på fältet «”Internet och islam”, Gary Bunt, publicerade 2000 boken *Virtually Islamic*. Att i dag, tjugo år senare, skriva en motsvarande bok, ämnad som en allmän översikt över vad Bunt då benämnde “Cyber Islamic Environments” är förmodligen en omöjlighet. Internet är i dag en så integrerad del av människors vardag, att företaget att skriva en akademisk översikt av typen “islam och internet” är ungefär lika hopplöst som att skriva en akademisk översikt över “islam

¹ Anledningen till detta begränsade urval var att information om ljudfilernas längd inte på ett enkelt sätt gick att inhämta automatiskt, utan i stället fick göras manuellt.

och samhället’.

Problemet som islamforskare i dag står inför, givet den stora mängden potentiellt relevant material, är brist på metoder för att orientera sig i detta material, och att göra relevanta urval. Det är detta problem som föreliggande artikel vill bidra till att lösa.

Traditionell, humanistiskt orienterad islamforskning är ofta inriktad på analys av ett mindre omfattande, och därmed mer överskådligt, material. Metoderna är detaljstudier (”närläsning”) och hermeneutiskt inkännande tolkning där historisk och kulturell kontextualisering är en central del av processen. Arbetssättet är tidskrävande och den mängd material som kan analyseras blir därmed också med nödvändighet begränsat. Metoder för att göra rimliga urval blir viktiga. I en situation där mängden potentiellt relevant material är begränsat – till exempel i de fall då endast en liten samling historiska texter finns bevarade för eftervärlden, eller då studierna inriktas mot material vars historiska eller samtida betydelse är allmänt erkänt i forskarsamhället – är urvalet inte något större problem. Men i mötet med stora mängder nytt material, som ofta är fallet i studier av nutida islam, där det saknas sekundär information som kan användas för att avgöra materialets relevans, finns hela tiden risk för godtycklighet. Mängden material gör också att även metoder för slumpvist urval är svåra att tillämpa, då ett sådant urval, för att kunna anses representativt, tenderar att bli mycket stort.

En väg ut ur dilemmat är naturligtvis att helt enkelt hävda att oavsett vad som väljs ut för närmare studier så är det valda materialet relevant, på grund av att det är resultatet av mänsklig kulturproduktion. Ett vanligt sätt att försvara ett sådant förfarande är att hävda att undersökningen utgör en ”fallstudie”. Detta är inte riktigt tillfredsställande, eftersom det inte löser problemet med varför material X är bättre lämpat som ”fall” än det lika lättillgängliga (och enligt resonemanget lika lämpliga) material Y.

Finns det då vägar ur denna problematik? Kanske. Modern dator- och kommunikationsteknologi har lett till att mängden användbart material för humanistiskt orienterad islamforskning, och dess tillgänglighet, har ökat dramatiskt. Det handlar om både historiskt material och material som återspeglar en pågående inom-muslimsk diskussion. Samtidigt har samma teknologi också frambringat verktyg för att hantera insamling, organisering, lagring och översiktlig analys av stora mängder material.

Föreliggande artikel är ett bidrag till en större, expanderande tradition av digital humaniora och dess tillämpning inom religionsvetenskap. Digital humaniora är i dag ett brett forskningsområde som innefattar till exempel humanistiska studier av digitalt arkiverat material, skapandet av sådana digitala arkiv, kritiska filosofiska och etiska analyser av digitaliseringsprocesser och, inte minst, forskning med hjälp av datorassisterade metoder för insamling, organisering och analys av data samt distribution och presentation av resultaten av sådana analyser (se t.ex. Schreibman, Siemens & Unsworth, 2016; Hansson & Svensson, 2020; Jarlbrink & Norén, 2021). Expansionen inom fältet har också nått olika religionsvetenskapliga discipliner, inklusive islamologin. Här kan till exempel nämnas förlaget De Gruyters bokserie *Introductions to Digital Humanities*, som i skrivande stund utgörs av fem volymer publicerade mellan 2019 och 2022, med titlar som *Digital Humanities and Research Methods in Religious Studies: An Introduction* (Christopher & Cantwell, 2021), *Digital Humanities and Christianity* (Hutchins & Clivaz, 2021) och *Digital Humanities and Material Religion* (Clark & Lindsay, 2022). Ännu finns ingen volym i serien som specifikt berör islam, men redan 2016 gav samma förlag ut samlingsvolymen *The Digital Humanities and Islamic & Middle East Studies* (Muhanna, 2016). I en nyligen publicerad översiktsartikel över nya tendenser inom islamologi noterar religionsvetaren Aria Nakissa digital humaniora som en sådan (Nakissa, 2021). Av hans genomgång framgår att det som hittills publicerats inom fältet huvudsakligen behandlar datorassisterad textanalys, och då primärt analyser av historiskt källmaterial, inklusive Koranen och hadithlitteraturen. Analyser av mer nutida material är framför allt fokuserat på studier av olika grupper, framför allt extremistiska sådana, propagandamaterial.

I viss mån faller även denna artikel inom kategorin datorassisterad analys av texter i digital form. I dag finns en rad etablerade metoder för att skapa överblick över stora mängder text och finna mönster och tendenser. Ibland benämns detta "fjärrläsning" (*distant reading*) (Jänicke, Franzini, Cheema & Scheuermann, 2015; Moretti, 2013; Slingerland, Nichols, Neilbo & Logan, 2017), vilket blir en kontrast till den ovan nämnda traditionella humanistiska "närläsningen". Färdig programvara eller specialkonstruerad programmeringskod – eller en kombination av de båda – kan användas till exempel för att närmare kartlägga stora textmassors karaktäristika, klassificera texter, jäm-

föra texter med varandra och skapa modeller över innehåll eller teman, på engelska benämnt “topic modelling” (se t.ex. Golub, 2020 för en översikt).² Detta görs oftast med utgångspunkt i mer eller mindre avancerade frekvensanalyser av ord och ord-sammansättningar. Dessa analyser är i allmänhet av en sådan karaktär att de hade varit om inte omöjliga, så åtminstone extremt tidskrävande att genomföra manuellt. Fjärrläsning kan ibland vara ett mål i sig, till exempel om syftet med forskningen är att studera förändringar i språkbruk över tid. I det som följer är jag dock intresserad av den roll som fjärrläsning kan ha som led i en urvalsprocess, som ett första steg i en forskning där en humanistiskt tolkande och kontextualiserande närläsning alltså är målet.

För att återgå till det material som står i fokus för denna artikel, det vill säga ljudfiler med salafistiska föreläsningar, är problemet att dessa inte är texter, åtminstone inte i en vardaglig förståelse av ordet. Det finns inte några etablerade och lättillgängliga metoder för “fjärrlyssning” av föreläsningar i ljudfilformat motsvarande metoder för fjärrläsning. Som nämnts ovan gjorde jag den 15 februari 2021 en inventering av befintliga inspelade föreläsningar på webbplatsen *islam.nu* och skapade en databas med den metadata som återfanns där: en föreläsningstitel, det datum den hade publicerats, vem talaren var, kort beskrivning av innehållet, länkar till sidan med föreläsningen och till ljudfilen samt information om i vilken kategori föreläsningen placerats. Denna metadata innehåller text, om än i begränsad omfattning, och skulle därmed kunna fungera som underlag för någon form av datorassisterad fjärrläsning. En sådan skulle emellertid knappast vara mer effektiv än en manuell genomläsning av samma material, givet det begränsade antalet föreläsningar. Jag har dock gjort en motsvarande analys av metadata för närmare 7 000 videor på svenska salafistiska YouTube-kanaler i ett annat sammanhang (Svensson, 2021), där fjärrläsning av titlar och beskrivningar användes för en tentativ gruppering av materialet. Även om vissa mönster kunde skönjas var resultatet mycket osäkert, inte minst för den begränsade informationen i metadatan. Detsamma kan sägas gälla för föreläsningarna på *islam.nu*. Tre meningars beskrivande text kan knappast täcka in innehållet i en föreläsning som varar i över en timme.

För att kunna tillämpa etablerade metoder för datorassisterad fjärrläsning på föreläsningarna måste de alltså först göras om till text. Att manuellt ombesörja detta är uteslutet. Alla som

2 Medie- och kommunikationsvetarna André Baltz och Fredrik Norén föreslår det svenska ordet “temamodellering” för “topic modelling” (Baltz & Norén 2021). För ett religionsvetenskapligt praktiskt exempel, se Linnéa Jensdotter nyligen publicerade avhandling *Religion och politik i hybrida mediemiljöer: En analys av kommentarer till nyheter om Miljöpartiet, Kristdemokraterna och Sverigedemokraterna på Facebook* (Jensdotter, 2021). För ett likaså nyligen publicerat religionsvetenskapligt exempel på fjärrläsning av stora mängder religiöst material, se Agersnap et al., 2020, där närmare 12 000 manuskript för predikningar inom Danska kyrkan mellan åren 2011 och 2016 har samlats in och bearbetats med datorassisterad textanalys.

någon gång har transkriberat till exempel intervjuer vet att det är en mycket tidskrävande process. Men finns det möjligen en annan väg?

Ett första intryck av automatisk transkribering genom Google Speech-to-Text

Automatisk transkribering av tal, baserat på tekniker för röstigenkänning, är en relativt ny teknik och ännu i sin linda. I dag finns en uppsättning verktyg och tjänster som erbjuder så kallad “tal-till-text” -teknologi (*speech-to-text*). Jag beslöt mig för att testa en sådan tjänst på föreläsningar från *islam.nu*, närmare bestämt Googles Speech-to-Text (Google, 2021).

Jag är inte först med att göra detta. Organisationen Pew Research Center publicerade 2019 resultatet av en omfattande, datorassisterad analys av innehållet i predikningar som kunde återfinnas på webbplatser tillhörande 6 431 olika kristna samfund i USA. Material inhämtades genom avancerad så kallad “webbskrapning” och utgjordes inte bara av texter, utan också av ljud- och videomaterial. Det senare automattranskriberades med företaget Amazons tjänst Amazon Transcribe (Pew Research Center, 2019a). I metodgenomgången framgår inte hur väl denna transkribering fungerade (Pew Research Center, 2019b). Mitt eget första försök med ljudfiler från *islam.nu* visade emellertid att Googles transkriberingstjänst lämnade en del i övrigt att önska.

Här är resultatet från inledningen av föreläsningen “Bulugh ul Maram – Bönens beskrivning Del 9” av predikanten Abdulwadud Frank, publicerad på *islam.nu*:s hemsida den 5 november 2018 (Islam.nu, 2018):

salam aleikum innan vi fortsätter att gå igenom vita bulldoggar en av det här Vi har kommit fram till hadith 325 där författaren kommer nämna ett antal hade gifter som handlar om Electrolux [...]

över fyra stycken här Deezer som handlar om du har eldklot och för att vi ska få en inblick i vad vi kommer att tala om så är det väldigt viktigt att förstå begreppet Du är en prutt prutt på arabiska kan ha flera innebörder det kan betyda lydnad till alla att dricka alla Du kan också

betyda att en person står i bönen.

Innebörden i ovanstående är långt ifrån glasklar. Snarare framstår texten som ren rappakalja. Här följer min manuella transkribering av samma passager:

[paus] *salam alaykum wa rahmatullahi wa barakatuhu.*
 [paus] *na'am.* Vi fortsätter att gå igenom *Kitab Bulugh al-Maram* av Hafiz ibn Hajar al-Asqalani *rahmat allahi ta'ala. Rahma wasi'a.* Vi har kommit fram till hadith 325 där författaren kommer nämna ett antal hadither som handlar om *al-qunut* [paus] *du'a al-qunut.* [...]

Här har vi då fyra stycken hadither som handlar om *du'a al-qunut*. Och för att vi ska få en inblick i vad vi kommer att tala om, så är det väldigt viktigt att förstå begreppet *du'a al-qunut*. Ordet *qunut* på arabiska kan ha flera innebörder. Det kan betyda lydnad till Allah. Att dyrka Allah. Det kan också betyda att en person [paus] står i bönen

Flera saker är värda att notera. Först och främst, när det gäller de delar som innehåller vardagliga ord på svenska är transkriptionen imponerande korrekt. Google Speech-to-Text lyckas till och med fånga upp ordet "hadith" i singular och hälsningsfrasen *salam alaykum*. Detta är särskilt imponerande då kvaliteten på inspelningen inte är den bästa.

De misstag som görs i transkriberingen är kopplade till ord, namn eller uttryck som är främmande för (eller ovanliga i) det svenska språket. Dessutom, när sådana återkommer transkriberas de på olika sätt. "Hadith" i singular transkriberas alltså korrekt, men pluralformen "hadither" transkriberas både som "hade gifter" och "här Deezer". Detsamma gäller det centrala begreppet *du'a al-qunut* som transkriberas på inte mindre än tre olika sätt: «"Electrolux", "du har eldklot" och "du är prutt".

Problemet, från en forskningshorisont, är att de feltranskriberade orden är betydelsebärande. De är så centrala för innehållet att transkriptionen är näst intill obegriplig utan dem. Den blir dessutom än mer obegriplig, och komisk, med de uttydningar som algoritmen levererar. De senare är förmodligen baserade på någon form av sannolikhetsbedömning. Hur osannolika transkriberingarna "du har eldklot", "Electrolux" och "du är prutt" än är, så är de på det stora hela förmodligen mer sannolika i en svensk kontext (i bokstavlig mening av "med-text") än *du'a al-qunut*. Det är rimligt att anta att motsvarande



Figur 1. Ordmoln över föreläsningen “Bulugh ul Maram – Börens beskrivning Del 9”.

problem var mindre i fallet med Pew Research Centers ovan nämnda insamling och analys.

Alltså, som en ersättning för manuell transkribering av just denna typ av texter, på svenska och speglade en islamisk-svensk “religiolekt” (Hary, 2011) fungerar inte Google Speech-to-Text. Det gör förmodligen inte heller någon annan motsvarande transkriberingstjänst byggd på motsvarande teknik, som till exempel IBM:s Watson, ovan nämnda Amazon Transcribe eller Microsofts Azure Speech to Text. Betyder detta att automatisk transkribering av ljudmaterial likt det som *islam.nu* publicerar är oanvändbar? Det var min första bedömning efter att ha prövat tekniken på fem föreläsningar, med motsvarande nedslående resultat. Dock gjorde jag ändå, som ett test, en enkel ordfrekvensanalys. Orden i transkriptionerna reducerades till sina grundformer (så kallad lemmatisering) med hjälp av verktyget Sparv (Borin et al., 2016), vilket finns tillgängligt i Språkbanken vid Göteborgs universitet. Vanliga ord i svenska språket (så kallade stopppord) rensades också bort. Därefter mättes de återstående ordens frekvens. Resultatet för de hundra vanligaste orden i ovan nämnda föreläsning om “Börens beskrivning” syns i figur 1. Storleken på de olika orden i ordmolnet representerar deras relativa frekvens.

Möjligen med undantag för enstaka ord som “alice” och “prutt” framstår inte de hundra mest frekventa orden, efter min fackmässiga bedömning som islamolog, som orimliga i en salafistisk föreläsning om “Börens beskrivning”. Ord som “profeten”, “bön”, “korrekt”, “hadither”, “koranen” och “ramadan” före-

faller snarare högst rimliga. Analysen av resultatet från de övriga fyra transkriptionerna gav samma intryck: vad som i en råversion framstod som rappakalja blev – efter utrensning av såväl ovanliga som mycket vanliga ord samt en ordfrekvensanalys – vad som möjligen, om än på en mycket generell nivå, kunde spegla ett faktiskt innehåll i föreläsningarna.

Detta väckte en tanke. Den automatiska transkriberingen resulterade inte i ett textmaterial som var användbart för vidare, detaljerad datorassisterad analys av innehållet i enskilda föreläsningar. Men skulle dessa texter ändå, om de jämfördes med varandra, kunna användas i ett annat syfte: för att skapa en översikt över materialet som helhet? En sådan översikt skulle sedan, i ett nästa steg, kunna användas för ett ändamålsenligt urval.

Innehållsorienterad nätverksanalys av 208 transkriberade föreläsningar

Google Speech-to-Text är en av flera resurser på plattformen Google Cloud. En registrering där gav mig 3 000 kronor i kredit. Jag beslöt mig att använda denna kredit för mitt test. Pengarna räckte till transkribering av totalt 245 föreläsningar. Transkribering av en timmes föreläsning tog ungefär tio minuter. Transkriberingarna var olika framgångsrika. Några föreläsningar returnerade inga resultat alls, eller endast ett fåtal ord. Efter att ha rensat bort alltför korta texter återstod 208 transkriptioner med en genomsnittslängd på 5 500 ord. Fördelningen runt medelvärdet var jämn (standardavvikelsen var 1 334).

Orden i texterna lemmatiserades och texterna sammanställdes i en textsamling (korpus). Därefter genomfördes en tf-idf-analys på samtliga 208.

Tf-idf (*term frequency – inverse document frequency*) är ett vanligt analysverktyg; något av en förfining av en enkel frekvensanalys av texter som ingår i större textsamlingar. Det som skiljer en tf-idf-analys från en vanlig frekvensanalys är att den relativa frekvensen av ett ord i en text (tf) viktas mot i hur många av de texter i den större textsamlingen ordet förekommer (idf). Ett ord som “hadith” kan till exempel vara mycket frekvent i en salafistisk text. En enkel frekvensanalys av ett dokument ger då för handen att detta ord är centralt. Men om man beaktar att “hadith” är ett ord som förmodligen förekommer i de flesta

salafistiska texter, säger den höga frekvensen i den undersökta texten kanske inte så mycket om just denna texts särdrag i relation till en större samling salafistiska texter. Då är förmodligen andra ord, med lägre frekvens i den undersökta texten men som helt saknas i samtliga eller de flesta andra texter av samma typ som man jämför med, mer talande. Tf-idf är alltså ett sätt att sätta en siffra på ords relativa signifikans i ett dokument givet det faktum att dokumentet ingår i en större textsamling som man jämför med (för en mer ingående men lättillgänglig beskrivning av tf-idf, se Lavin, 2019).

Syftet med denna första tf-idf-analys var dock inte – åtminstone inte i första läget – att identifiera särdrag hos enskilda transkriptioner, utan att jämföra dessa texter med varandra. Tf-idf-analysen kan nämligen utgöra en grund för en bedömning av hur lika varandra två texter är.

För varje dokument beräknades vart och ett av de däri förekommande ordens tf-idf-värde. Till detta fördes tf-idf-värdet för alla ord som inte förekom i den specifika texten, men som förekom i någon av de andra texterna (och som därför hade tf-idf-värdet 0). För varje dokument skapades genom detta en serie (eller med ett mer tekniskt språk en mångdimensionell vektor) av tf-idf-värden för alla unika ord i samtliga dokument sammantaget. När denna serie av värden för ett dokument jämförs med motsvarande serie för ett annat dokument kan man skapa ett mått på likheten mellan de två dokumenten. Det mått jag här använde benämns *kosinuslikhet* och är ett av de mer robusta likhetsmått. Processen, och matematiken bakom, är något invecklad och kan inte beskrivas i detalj här (för en förtjänstfullt begriplig genomgång, se Ladd, 2020), men slutresultatet blir ett mått mellan 0 och 1, där 1 motsvarar total likhet (med avseende på ordförekomst) och 0 total avsaknad av likhet (inte ett enda av samtliga unika ord i alla texter förekommer i båda texterna).

Alla 208 texter jämfördes med varandra med avseende på kosinuslikhet. De par som hade ett kosinuslikhetsvärde över 0,2 (ett godtyckligt val, men betingat av ett behov att inte göra grafen nedan alltför svårtydd) fördes in i en nätverksrepresentation.

Nätverksanalys är en sedan länge vanligt förekommande metod för att kartlägga sociala relationer (Scott, 1988) och har gynnats mycket av framväxten av nya digitala verktyg. Nätverksanalys är i dag också en vanlig metod vid fjärrläsning av texter i till exempel kartläggning av relationer mellan ord eller ord-



Figur 2. En första visualisering av nätverket över likhet mellan 193 föreläsningar baserat på ord-förekomst och efter tillämpning av layout-algoritmen Force Atlas.

sammansättningar. I en grundläggande introduktion till metoden skriver David Gunnarsson Lorentzen i generella ordalag att “[n]ätverksanalyser är relevanta för att studera relationer mellan saker i olika sammanhang” (Gunnarsson Lorentzen, 2021:164). “Sakerna” i sammanhanget kan alltså vara av olika slag. I föreliggande artikel är “sakerna” alltså de transkriberade föreläsningarna från *islam.nu*. I enlighet med gängse terminologi utgör dessa “noder” i nätverket. En relation mellan två noder – av Gunnarsson Lorentzen kallad “koppling”, men på engelska ofta *edge* eller “kant” – upprättas alltså om kosinuslikheten överstiger 0,2. Kopplingarnas styrka får olika vikt beroende på nivån av kosinuslikhet.

Figur 2 är en visualisering av nätverket med hjälp av den kostnadsfria mjukvaran Gephi (Gephi.org, 2008–2017). Varje punkt i grafen representerar en nod (en transkriberad text) och kopplingarna mellan noderna representerar en kosinuslikhet

som överstiger tröskelvärden 0,2. Ju tjockare linje desto större likhet. Avståndet mellan noderna och deras placering är resultatet av tillämpning av layout-algoritmen Force Atlas, en av flera möjliga i Gephi för att visuellt åskådliggöra närhet och avstånd mellan noder, baserat på både mängden och styrkan hos inbördes kopplingar. Det är alltså ett sätt att visualisera gruppering av transkriberade föreläsningar baserat på likhet. Den uppmärksamme läsaren noterar att det inte är 208 noder i nätverket. 15 noder hade inte kopplingar till några andra noder med en kosinuslikhet över tröskelvärden. Dessa är bortplockade från grafen. 193 noder återstår.

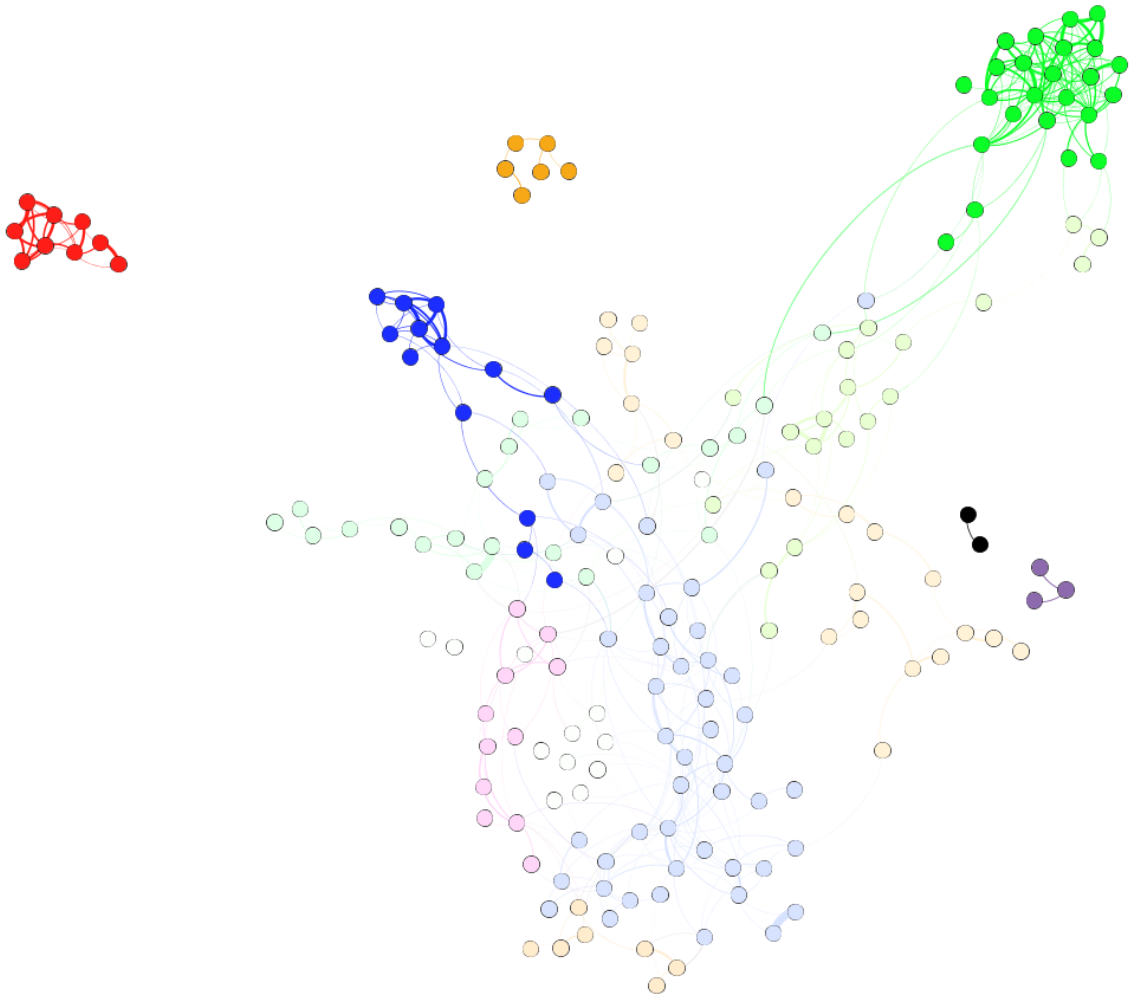
Redan här kan man möjligen skönja några mönster. Fyra grupper av transkriberade föreläsningar saknar kopplingar till det stora komplexet i mitten av visualiseringen. Den största gruppen består här av nio föreläsningar och återfinns längst upp till vänster i grafen. Längst upp till höger återfinns också en grupp som visserligen har kopplingar till den stora gruppen i centrum, men som ändå uppvisar speciellt starka kopplingar till varandra.

Dessa eventuella mönster är dock inte helt tydliga i grafen. Gephi erbjuder emellertid en algoritm för gruppering av noder i ett nätverk (nätverkets "modularitet"). Användaren anger ett värde för "upplösning" (*resolution*), och utifrån detta söker algoritmen det bästa sättet att dela upp nätverket, baserat på relationen mellan noderna. Antalet grupper som skapas är beroende av den angivna upplösningen. Figur 3 är resultatet av en tillämpning av denna funktion, med standardupplösningen 1,0. De olika färgerna representerar olika grupper.

De fem grupper som kunde noteras innan grupperingsalgoritmen tillämpats kvarstår, i grafen färgade med rött, orange, grönt, svart och lila. Men algoritmen identifierar ytterligare åtta grupper. De flesta av dessa ingår i ett tätare nätverk, inte så tydligt skilda från varandra, och har här markerats med något ljusare färger. Dock framträder en ytterligare grupp som något avskild från det tätare nätverket i centrum. Det har markerats med mörkare blått i grafen.

Åter till de transkriberade föreläsningarna

Vad är nu poängen med detta? Jo, bakom visualiseringen ligger data, närmare bestämt text från transkriberade föreläsningar.



Figur 3. En andra visualisering med tillämpning av Gephis grupperingsalgoritm (modularity).

Det är nu möjligt att med Gephis gruppering i bagaget gå tillbaka till denna data och gräva djupare med utgångspunkt i nätverksanalysen. Vad är det för ord som ger upphov till de olika grupperna och kan dessa ord säga något om vad som skiljer dessa grupper åt?

Uppdelningen i 13 grupper baseras alltså på en analys av likheter i ordförekomst mellan de 193 transkriberade föreläsningarna, vilket i sin tur vilar på en tf-idf-analys av var och en av dessa föreläsningar. Som noterats ovan används tf-idf-analys annars ofta för att identifiera *skillnader* mellan texter i en textsamling, eller mer korrekt vilka ord som kan sägas vara signifikanta för en specifik text i relation till andra texter i en textsamling. Det är denna egenskap hos tf-idf-analysen som nu kan brukas i ett andra steg av analys, för att jämföra de samlade textmassorna för transkriberade föreläsningar i de 13 grupperna.

Som noterats ovan innehåller de ursprungliga transkriberade föreläsningarna "brus" i form av feltranskriberingar. För likhetsanalysen spelade detta mindre roll. Ord som förekommer enstaka gånger i enstaka transkriptioner får inga märkbara effekter i den parvisa jämförelsen mellan texter. Nu när syftet i stället är att finna signifikanta ord för de 13 grupperna kan feltranskriberingar som "bulldogar", "Electrolux" och "prutt" störa genom att de får ett högt idf-värde, det vill säga orden förekommer i en enstaka text, men inte i de andra. Som den enklare frekvensanalysen ovan visade verkar problemet med brus försvinna om ord med låg frekvens, bland annat idiosynkratiska feltranskriberingar, rensas bort. För den vidare analysen gjordes därför ett val att rensa bort alla ord med en relativ frekvens lägre än en promille från var och en av de 193 transkriptionerna. Denna utrensning gjordes innan texterna i var och en av grupperna sammanfördes till 13 helheter.

Tabellen nedan redovisar resultatet av tf-idf-analysen av de 13 grupperna. I varje cell återfinns de tio mest signifikanta orden i varje grupp i fallande rangordning. Antalet texter i varje grupp anges inom parentes.

Vad kan man då säga om detta resultat? För det första verkar några transkriberingsfel fortfarande spöka. Orden "kyrka", "kyrk" och "gurka" i grupp 0 är misstänkt lika. Här har vi ett exempel där domänspecifik islamologisk kunskap ger tolkningsnycklar. Mot bakgrund av att gruppen som producerat och sprider föreläsningarna är salafistisk är en inte helt orimlig gissning att bakom dessa tre ord döljer sig ett arabiskt: *shirk*, eller "avgudadyrkan". Ordet "marianne" i grupp 11, där flera egennamn på profeter förekommer, kan tänkas vara en feltranskribering av "Mariam", mor till profeten 'Isa (Jesus).

Ovan noterades hur sex grupper skiljde ut sig i den grafiska representationen av nätverket. Dessa är 1 (lila), 3 (grön), 5 (blå), 7 (svart), 8 (röd) och 12 (orange). Om man ser till listorna över de tio mest signifikanta orden i var och en av dessa grupper så framträder möjligen ett mönster. I den största gruppen (3) återfinns ord som verkar ha med bön att göra. Jag tillåter mig här en gissning, också på basis av mina förkunskaper, att orden "rak", "ben" och "böna" är resultat av feltranskriberingar ("rak" kan då vara *rak'a*, eller "runda", i bönen). I den näst största gruppen (5) verkar det likaledes framträda ett tema: äktenskap och familj.

En snabb genomgång av metadatan för dessa föreläsningar

0 (11)	1 (3)	2 (19)	3 (24)	4 (20)
avsikt	ramadan	minnas	bön	ljuga
kyrka	grej	sammankomst	rak	svära
uppriktighet	aktivitet	hår	moské	förbannad
gud	saker	förlåtelse	rad	karaktär
kategori	sverige	kväll	obligatorisk	grupp
rättighet	ting	sova	natt	ärlig
förälder	månad	svår	böna	främmande
utföra	hjälpa	ali	åsikt	förbanna
kyrk	föreläsning	morgon	ben	lugn
gurka	utav	säng	mage	bevara

5 (13)	6 (13)	7 (2)	8 (9)	9 (45)
gifta	tålmod	bild	äterberätta	sanning
äktenskap	motgång	avbilda	ovanför	sändebud
fru	tålmodig	vägg	bevis	grav
förälder	öde	måla	tron	straff
familj	drabba	hund	skapelse	ali
skyldighet	sjukdom	staty	handling	uppenbarelse
fria	medicin	otur	uttalande	eld
hjälpa	hård	hus	äterberättning	skicka
baktala	nöjd	stjärna	bekräfta	mosa
försörja	hämnas	framtiden	himmel	gud

10 (21)	11 (7)	12 (6)
medelväg	ibrahim	lärd
bojkotta	marianne	regn
fattig	berättelse	uppenbarelse
kärlek	son	himmel
hjälpa	far	liknelse
skada	zacharia	islamisk
guld	ismail	överdriva
extrem	gud	makt
rikedom	föda	ljus
tyst	bokstav	likna

tyder på att metoden fungerar åtminstone i dessa två fall (grupp 3 och grupp 5).

16 av 24 föreläsningar i grupp 3 är, av titeln att döma, från samma serie, där talaren föreläser om ovan nämnda bok, *Bulugh al-maram*. Denna serie omfattar totalt 23 föreläsningar, men endast de 16 som återfinns i grupp 3 behandlar, av metadatan att döma, olika aspekter av bön. Av de resterande 8 föreläsningarna i grupp 3 har endast en ordet "bön" i titeln. Det framkommer dock av de något längre beskrivningarna av föreläsningarna att i samtliga är bön antingen huvudtema, eller ingår som ett av flera teman.

I grupp 5 utgörs 8 av 13 föreläsningar av delar i en serie, "Det lyckade äktenskapet". Dessa åtta är dock inte samtliga i materialet förekommande föreläsningar från denna serie, utan en föreläsning saknas. En närmare titt på dess beskrivning indikerar dock att temat här inte är äktenskap specifikt, utan hur man ska välkomna och namnge nyfödda barn. Av metadata för fem föreläsningar i gruppen framgår en koppling till äktenskap och relationer endast i en beskrivning. Här kan man dock i grafen notera att grupp 5 (mörkblåa punkter) är mindre samlad än grupp 3 (gröna punkter) och att flera noder befinner sig ganska långt ifrån vad som skulle kunna sägas vara gruppens centrum.

Det är möjligt att grupp 1 (lila) och grupp 7 (svart) båda representerar tematiska orienteringar, mot fastemånaden ramadan respektive mot avbildningar. När det gäller de två återstående grupperna som skiljer ut sig från den stora mängden, grupp 12 (orange) och grupp 8 (rött), är det svårare att gissa sig till en eventuell gemensam tematik utifrån signifikanta ord. Om man tar en närmare titt på metadata för föreläsningarna så får åtminstone 8 som avskild grupp en möjlig förklaring. För de 208 föreläsningar som ingick i det ursprungliga urvalet anges i 205 fall vem som talar. Den överlägset vanligast förekommande talaren är Abdulwadud Frank (188 föreläsningar). Utöver detta finns sex föreläsningar där Frank och Abdallah al-Sueidi tillsammans är talare, och för två anges Moosa Assal. Det mest intressanta för föreliggande analys är dock att de återstående nio föreläsningarna, som också är de nio föreläsningar som utgör grupp 8, har en och samma talare: Imran Sheikh. Anledningen till att dessa föreläsningar hamnar i en och samma grupp, tydligt skild från övriga föreläsningar, kan naturligtvis vara tematik. Det är dock också tänkbart att en analys baserad på ordförekomst inte bara fångar upp innehållsliga teman, utan också unika drag i enskilda talares vokabulär. Underlaget i form av transkriberade

föreläsningar och antalet talare är dock för litet och för skevt för att i detta skede tillåta några tydliga slutsatser i denna riktning.

Fortfarande är det dock svårt att utifrån tf-idf-analysen dra några slutsatser om möjliga orsaker till att grupp 12 skiljer ut sig. Av beskrivningarna i metadata framgår dock att tre av de sex föreläsningarna också ingår i en serie och att deras tema är "kunskap". De tre övriga saknar närmare beskrivning. De är alla *tafsir* av specifika kapitel eller verser i Koranen.

Utifrån tf-idf-analysen kan det framstå som om föreläsningar inom några av de andra grupperna också delar tematik, till exempel ett tema av sjukdom och medicin i grupp 6 och profethistorier i grupp 11, men dessa grupper tillhör båda den mer samlade gruppen av noder i mitten av nätverket. Skillnaderna mellan de olika grupperna är inte så tydliga.

Slutord

Vad kan man då dra för slutsatser från genomgången ovan? Först och främst vill jag betona att det handlar om ett test. Syftet med den här artikeln har inte varit att tillhandahålla en helhetsbild över föreläsningar på webbplatsen *islam.nu*. Av ekonomiska skäl har endast en mindre del av dessa föreläsningar ingått i analysen och dessa utgör heller inte ett slumpvist urval avsett för generaliseringar om helheten. Ambitionen har varit att testa en möjlig metod för "fjärrlyssning" (via fjärrläsning) av stora mängder inspelat tal i syfte att underlätta ett ändamålsenligt urval av föreläsningar för fördjupat studium. Resultatet tyder på att redan en relativt enkel analys av likheter mellan de automatiskt transkriberade föreläsningarna, med avseende på ordförekomst, kan identifiera mönster. Därmed har en sådan analys potential att fungera som vägledning i en urvalsprocess för ett omfattande material vars ursprungskaraktär annars inte på ett enkelt sätt tillåter detta.

Fokus ovan har legat på indelningen av föreläsningar i grupper och eventuella likheter inom och skillnader mellan dessa grupper. Det är dock också fullt möjligt att i en urvalsprocess i stället först identifiera enskilda föreläsningar och sedan göra ett "snöbollsurval" på basis av styrkan i kopplingar till andra föreläsningar i nätverket, för att på så sätt skapa ett rimligt underlag för en fallstudie. Det är också fullt möjligt att göra urval på basis av ett specifikt ords signifikansnivå, baserat på den

första tf-idf-analysen av samtliga 208 föreläsningar. Här kan forskaren till exempel på ett enkelt sätt identifiera de föreläsningar där ordet “äktenskap” framträder som särskilt signifikant (det vill säga har ett relativt sett högt tf-idf-värde) i jämförelse med alla föreläsningar.

Det bör noteras att den analys som gjorts av transkriptionerna är en förhållandevis okomplicerad form av fjärrläsning. Som nämnts ovan finns det långt mer avancerade metoder för datorassisterad textanalys att tillgå. Det faktum att redan de enklare former av analys som tillämpats i denna artikel verkar ge meningsfulla resultat motiverar att i framtida studier även beakta möjligheten att andra, mer avancerade metoder för fjärrläsning kan ge ännu bättre utdelning. Problemet när det gäller mer avancerade innehållsanalyser är kvaliteten på transkriptionerna, där viktiga och för en islamologisk analys betydelsebärande ord helt enkelt inte blir en del av underlaget.

Tekniker för automatisk transkribering av tal har kommit en bra bit på vägen, men det är tydligt att de i nuläget inte kan ersätta manuell transkribering, särskilt i sammanhang där det tal som ska transkriberas avviker från det dominerande språkbruk som algoritmerna tränats på. Detta gäller också den form av “islamisk svenska” som återfinns i föreläsningarna på *islam.nu*. Dessutom är automatisk transkribering i nuläget kostsam. Denna kostnad minskar måhända i framtiden när tekniken blir mer etablerad och konkurrensen mellan olika aktörer som tillhandahåller tjänster inom området ökar. För forskare är avvägningen mellan nytta och kostnad i nuläget dock rimlig att göra. I vissa fall är det kanske heller inte nödvändigt att fjärrläsa *innehållet* i material som de här studerade föreläsningarna. I de fall det förekommer tillräckligt utförliga metadata kan dessa vara ett fullgott alternativ i ett urvalsförfarande, eller kanske inte helt fullgott.

För den forskare som är intresserad av till exempel hur temat äktenskap och relationer behandlas av “Järva-salafisterna” kan naturligtvis titlarna i serien “Det lyckade äktenskapet” vara tillräckliga för att göra ett rimligt och motiverat urval. De ingår också nästan samtliga i grupp 5 ovan. Men här ingår också fem ytterligare föreläsningar. Åtminstone en av dessa kan ha viss relevans för temat, även om metadatan inte ger någon indikation i den riktningen. Föreläsningen med titeln “Riyâdh-us-Sâlihîn del 151” (Islam.nu, 2019) åtföljs av följande beskrivning: “Bismillah. Boken Riyâdh-us-Sâlihîn skriven av imam an-Nawawî

skrevs som en vägledning för muslimerna. Denna föreläsning handlar om: Vid vilka tillfällen är det tillåtet att baktala en person? Faran av namima/svaller.” En stor del av föreläsningen ägnas åt hustrumisshandel, vilket förmodligen förklarar varför denna föreläsning hamnar i grupp 5.

I beskrivningen ovan har jag medvetet fokuserat på processen som helhet, snarare än på detaljer för olika moment i den datorassisterade insamlingen, organiseringen och analysen av data. Även om det inte har framgått så har samtliga av dessa moment innefattat löpande produktion av ändamålsanpassad kod i programmeringsspråket Python för diverse operationer, inklusive kommunikation med externa tjänster som Sparv och Google Speech-to-Text. I detta har jag också nyttjat flera externa så kallade “bibliotek” med kod som andra skapat. Så har till exempel biblioteket Beautiful Soup (Richardson, 2019) använts för extrahering av data från hemsidor; Pandas (Pandas Development Team, 2020) för organisering och analys både av metadata och transkriberad text; TextBlob (Loria, 2020) och WordCloud (Mueller, 2020) för frekvensanalys och ordmolnsvisualisering; scikit-learn för tf-idf-analys (Pedregosa et al., 2011); NetworkX (NetworkX Developers, 2014–2021) för generering av nätverk och skapande av fil för import till Gephi och visualisering.

Sammantaget utgör det försök som beskrivits ovan ett exempel på vad medievetaren Nick Montfort har benämnt “explorativ programmering inom humaniora” (Montfort, 2016), en underavdelning inom det större fältet digital humaniora. Med utgångspunkt i vad som i grunden är klassiska frågor inom humaniora rörande mänsklig kulturproduktion, dess karaktär, bakomliggande drivande processer och förhållande till en större samtida och historisk kontext, används datorkraft och ändamålsenlig, ofta rudimentär, programmering “on-the-fly” för att enkelt samla in, organisera och analysera stora mängder material. Syftet kan vara att jämföra, att testa hypoteser, att finna mönster eller helt enkelt att skapa en överblick. Explorativ programmering inom humaniora är ingen ersättning för detaljerade studier av tolkning och kontextualisering, utan ett komplement och en hjälp med att hantera grundläggande problem, likt det problem som stått i fokus för denna artikel. Problemet kan sammanfattas i den återkommande jobbiga, men också i högsta grad berättigade, frågan: “Varför har du valt att närstudera X, när du lika gärna kunde ha närstuderat Y?”

Abstract

The present article is the result of a test. This test investigates if it is possible to use the highly defective results from an automatic transcription using Google's service Speech-to-Text of 208 audio-files containing religious lectures gathered from the Swedish Salafi webpage www.islam.nu, in order to detect recurring patterns. The aim is to gain an overview of the content of a material that is both large and in its original form difficult to assess. The article is a stepwise presentation of a process that combines computer assisted methods for distant reading and network analysis. The results indicate that a computer-assisted analysis of automatically transcribed texts, even when the transcription is flawed, may have a role to play especially concerning how to proceed in selecting material for more in-depth studies.

Referenser

Källor

- Islam.nu 2018. *Bulugh ul-Maram. Bönens beskrivning. Del 8*, <https://islam.nu/forelasningar/bulugh-ul-maram-bonens-beskrivning-del-8> [besökt 21-09-23]
- Islam.nu 2019. *Riyadh us-salihin. Del 151*, <https://islam.nu/forelasningar/riyadh-us-salihin-del-151/>, [besökt 21-09-23].

Anförda kodbibliotek, programvaror och on-line-resurser

- Gephi.org. 2008–2017 *Gephi* (version 0.9.2), [<https://gephi.org>]
- Google 2021. *Cloud Speech-to-Text*. [<https://cloud.google.com/speech-to-text>]
- Loria, Steven, 2020. *TextBlob* (version 0.16.0). [<https://textblob.readthedocs.io/en/dev/>]
- Mueller, Andreas 2020. *WordCloud* (version 1.8.1). [https://amueller.github.io/word_cloud/]
- NetworkX Developers 2014-2021. *NetWorkX. Network Analysis in*

- Python* [<https://networkx.org>]
- Pandas Development Team 2020 *Pandas* (version 1.2.0). [<https://pandas.pydata.org>]
- Richardson, Leonard, 2019. *Beautiful Soup 4* (version 4.8.1) [<https://pypi.org/project/beautifulsoup4/>]
- Scikit-learn developers, 2007-2020. *Scikit-learn* (version 1.1.dev0) [<https://scikit-learn.org>]

Anförd litteratur

- Agersnap, Anne., Ross Deans Kristensen-McLachlan, Kristine Helboe Johansen, K. , Uffe Schjødt, & Kristoffer Laigaard Nielbo. 2020. "Sermons as data: Introducing a corpus of 11,955 Danish sermons". *Journal of Cultural Analytics*, 1(1), 1-27, doi: 10.7910/DVN/WIQ6CB.
- Baltz, André & Fredrik Noren. 2021. "Tematisk innehållsanalys med temamodellering". I I J. Jarlbrink & F. Norén (red.) *Digitala metoder i humaniora och samhällsvetenskap*. L

- und: Studentlitteratur, s. 211–234.
- Borin, Lars, Markus Forsberg, Martin Hammarstedt, Dan Rosén, Roland Schäfer & Anne Schumacher 2016. *Sparv: Språkbanken's Corpus Annotation Pipeline Infrastructure*. Paper presenterat vid The Sixth Swedish Language Technology Conference (SLTC), Umeå universitet, Umeå, url: http://www8.cs.umu.se/~johanna/sltc2016/abstracts/SLTC_2016_paper_31.pdf.
- Clark, Emily. S., & Lindsey, Rachel. M. B. (red.). 2022. *Digital Humanities and Material Religion: An Introduction*. Berlin: De Gruyter.
- Christopher, D. Cantwell och Kristian Petersen (red.). 2021. *Digital Humanities and Research Methods in Religious Studies: An Introduction*. Berlin. De Gruyter.
- Gunnarsson Lorenzen, David. 2021. «Nätverk - analys, mått och tillämpningar». I J. Jarlbrink & F. Norén (red.) *Digitala metoder i humaniora och samhällsvetenskap*. Lund: Studentlitteratur, s. 161–186.
- Golub, Koraljka. 2020. "Automatic Identification of Topics: Applications and Challenges". I J. Hansson & J. Svensson (red.) *Doing Digital Humanities: Concepts, Approaches, Cases*. Växjö: Linnaeus University Press, s. 5–26, url: <http://lnu.diva-portal.org/smash/get/diva2:1459889/FULLTEXT01.pdf>.
- Hansson, Joachim & Jonas Svensson. 2020. *Doing Digital Humanities: Concepts, Approaches, Cases*. Växjö: Linnaeus University Press. [Tillgänglig on-line från <http://urn.kb.se/resolve?urn=nbn:se:lnu:diva-97624>].
- Hary, Benjamin. 2011. "Religiolect". I Deborah D. Moore, Anita Norich & Joshua L. Miller (red), *Critical Terms in Jewish Language Studies*. Ann Arbour: Frankel Institute for Advanced Judaic Studies, s. 43–35.
- Hutchings, Tim & Claire Clivaz (red.). 2021. *Digital Humanities and Christianity: An Introduction*. Berlin: De Gruyter, doi: 10.1515/9783110574043.
- Jarlbrink, Johan & Fredrik Norén, (red.). 2021. *Digitala metoder i humaniora och samhällsvetenskap*. Lund: Studentlitteratur.
- Jensdotter, Linnea. 2021. *Religion och politik i hybrida mediemiljöer : en analys av kommentarer till nyheter om Miljöpartiet, Kristdemokraterna och Sverigedemokraterna på Facebook*. Uppsala: Acta Universitatis Upsaliensis. [Tillgänglig online från <http://urn.kb.se/resolve?urn=nbn:se:uu:diva-435271>]
- Jänicke, Stefan, Greta Franzini, Muhammad Faisal Cheema & Gerik Scheuermann. 2015. *On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges*. Paper presenterat vid EuroVis (STARs), doi: 10.2312/eurovisstar.20151113.
- Ladd, John. R. 2020. "Understanding and Using Common Similarity Measures for Text Analysis". *Prgramming Historian* (9), doi: 10.46430/phenoo89. [Inga sidangivelser i originalet].
- Lavin, Matthew J. 2019. "Analysing Documents with TF-IDF". *Prgramming historian* (8), doi: 10.46430/phenoo82. [Inga sidangivelser i originalet].
- Muhanna, Elias (red.) 2016. *The Digital Humanities and Islamic & Middle East Studies*. Berlin: de Gruyter, doi: 10.1515/9783110376517.
- Montfort, Nick. 2016. *Exploratory Programming for the Arts and Humanities*. Cambridge, Massachusetts: MIT Press.
- Moretti, Franco. 2013. *Distant Reading*. London: Verso Books.
- Nakissa, Aria. 2021. "Cognitive and Quantitative Approaches to Islamic Studies: Integrating Psychological, Socioeconomic, and Digital-Cultural Statistics". *Religion Compass*, 15(12), 1-25, doi:10.1111/rec3.12424
- Olsson, Susanne. 2020. "Advising and Warning the People. Swedish Salafis Against Violence". I S. Stjernholm & E. Özdalga (red.), *Muslim Preaching in the Middle East and Beyond: Historical and Contemporary Case Studies*. Edinburgh: Edinburgh University Press, s. 155–172.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, et al. 2011. "Scikit-Learn: Machine Learning in Python". I *The Journal of Machine Learning Research* 12: 2825-2830, url: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>.
- Pew Research Center. 2019a. *The Digital Pulpit. A Nationwide Analysis of Online Sermons*. Washington: Pew Research Centre. [Tillgänglig online från <https://www.pewforum.org/2019/12/16/the-digital-pulpit-a-nationwide-analysis-of-online-sermons/>]
- Pew Research Center. 2019b. *The Digital Pulpit. A Nationwide Analysis of Online Sermons. Methodology*. <https://www.pewforum.org/2019/12/16/methodology-30/>. [Arkiverad på Wayback Machine 2021-12-16].
- Potter, Jonathan & Shaw, Cloe. 2018. "The Virtues of Naturalistic Data". I Uwe Flick (red.) *The Sage Handbook of Qualitative Data*

- Collection*. London: SAGE, 182-199.
- Ranstorp, Magnus, Filip Ahlin, Peder Hyllengren & Magnus Normark (2018). *Mellan salafism och salafistisk jihadism: Påverkan mot och utmaningar för det svenska samhället*. Stockholm: Försvarshögskolan. [Tillgänglig online från <https://rib.msb.se/filer/pdf/28763.pdf>].
- Schreibman, Susan, Raymond G. Siemens och John Unsworth (red). 2016. *A New Companion to Digital Humanities*. Chichester: John Wiley & Sons.
- Scott, John. (1988). "Social Network Analysis". *Sociology* 22(1), 109–127.
- Slingerland, Edward, Ryan Nichols, Kristoffer Neilbo, och Carson Logan 2017. "The Distant Reading of Religious Texts: A 'Big Data' Approach to Mind-Body Concepts in Early China". *Journal of the American Academy of Religion* 85(4), s. 985–1016, doi: 10.1093/jaarel/lfw090.
- Sorgenfrei, Simon. (2021a). "Crowdfunding Salafism. Crowdfunding as a Salafi Missionising Method". I *Religions* 12(3), 209, doi: 10.3390/rel12030209.
- Sorgenfrei, Simon (2021b). "Perhaps We See It in Negative Terms, But, Ultimately it is Positive". *Tidskrift för Islamforskning* 14(2), s 40–62, doi: 10.7146/tifo.v15i2.125959.
- Sorgenfrei, S. kommande. Branding Salafism: Salafi Missionaries as Social Media Influencers. *Method and Theory in the Study of Religions*.
- Svensson, Jonas, 2021. «Mönster i svensk YouTube-salafism. Datorassisterad metadataanalys som urvalsmetod i religionsvetenskapliga studier av videomaterial online», *Religionsvetenskapelig Tidskrift* 73, 62–85.