Tekniker för analys av data från webben

ULRIK FRANKE, FREDRIK JOHANSSON, MAGNUS JÄNDEL,
LISA KAATI, MARIANELA GARCIA LOZANO

FOI är en huvudsakligen uppdragsfinansierad myndighet under Försvarsdepartementet. Kärnverksamheten är forskning, metod- och teknikutveckling till nytta för försvar och säkerhet. Organisationen har cirka 1000 anställda varav ungefär 800 är forskare. Detta gör organisationen till Sveriges största forskningsinstitut. FOI ger kunderna tillgång till ledande expertis inom ett stort antal tillämpningsområden såsom säkerhetspolitiska studier och analyser inom försvar och säkerhet, bedömning av olika typer av hot, system för ledning och hantering av kriser, skydd mot och hantering av farliga ämnen, IT-säkerhet och nya sensorers möjligheter.

FOI
Totalförsvarets forskningsinstitut
164 90 Stockholm

Tel: 08-55 50 30 00
Fax: 08-55 50 31 00

www.foi.se

FOI-R--3532--SE
ISSN 1650-1942

December 2012

Ulrik Franke, Fredrik Johansson, Magnus Jändel,
Lisa Kaati, Marianela Garcia Lozano

Tekniker för analys av data från webben

Titel	Tekniker för analys av data från webben
Title	Techniques for analysis of web data
Rapportnr/Report no	FOI-R--3532--SE
Månad/Month	December
Utgivningsår/Year	2012
Antal sidor/Pages	57 p
ISSN	1650-1942
Kund/Customer	FM
FoT område	Ledning och MSI
Projektnr/Project no	E36701
Godkänd av/Approved by	Lars Höstbeck
Ansvarig avdelning	Informations- och aerosystem

Bild/Cover: enciktep / Shutterstock

Sammanfattning

I den här rapporten beskrivs några av de tekniker som kan användas för analys av data från webben. Rapporten är inte tänkt att ge en uttömmande beskrivning av ämnet utan visar bara på en del av de tekniker som finns tillgängliga och kan vara användbara när man analyserar data som härstammar från webben. De tekniker som presenteras är olika former av datoriserad textanalys, analys av sociala nätverk, tekniker för att avgöra om en användare har flera alias, hur digitala spår kan användas i jakten på ensamagerande terrorister, integritetsbevarande tekniker för analys och slutligen hur man kan bedöma trovärdigheten hos data från webben.

Nyckelord: webbspaning, webbdata, textanalys, sociala nätverk, informationstrovärdighet, integritetsskyddande tekniker, informationsfusion

Summary

This report describes some of the techniques that can be used for analyzing data from the Internet. The report is not an exhaustive description of the subject; it just points to some of the techniques that are available and useful when analyzing data derived from the web. The techniques that are described are different forms of computerized text analysis, social network analysis, techniques for detecting if a user has multiple aliases, how digital traces can be used in the hunt for lone wolf terrorists, integrity preserving data mining and finally how information credibility can be judged.

Keywords: web mining, web data, text analysis, social networks, integrity preserving analysis, alias matching, information fusion

Innehållsförteckning

1 Bakgrund

Denna rapport riktar sig främst till personer inom svenska försvarsmyndigheter med intresse för inhämtning och analys av information från öppna källor på webben, men bör också kunna vara av intresse för motsvarande befattningsinnehavare på civila myndigheter. Grundläggande datavetenskapliga förkunskaper om bland annat sociala medier och internets uppbyggnad i stort förutsätts, men ambitionen är att rapporten ska vara av intresse för läsare med vitt skilda bakgrunder och förkunskaper. Rapporten ger en översikt av några av de tekniker som kan användas för att analysera webbdata och är en leverans inom FoT-projektet Verktyg för informationshantering och analys (VIA). Det bör noteras att rapporten inte ger någon fullständig överblick över alla de tekniker som kan vara av intresse för webbanalys utan fokuserar på de tekniker som studerats och forskats på inom projektets ramar.

Bland de andra avgränsningar som gjorts så har vi valt att fokusera på enbart textuell data, även om det finns enorma mängder andra typer av data av stort potentiellt intresse vid webbanalys i form av bland annat bild- och videodata[1]. Vidare så har de flesta icke-tekniska aspekter utelämnats från rapporten, trots att det exempelvis finns många intressanta och viktiga etiska och legala frågeställningar kopplade till webbanalys.

[1] För mer information om analys av bild- och videodata se: Ahlberg, J., Johansson, F, Johansson, R., Jändel, M., Linderhed, A., Svenson, P., Tolt, G., *Content-based image retrieval – An introduction to literature and applications.* FOI Rapport, FOI-R--3395--SE, 2011.

2 Inledning

Det är ingen överdrift att säga att informations- och kommunikationsteknik har förändrat världen. Många av oss arbetar, roar oss, umgås med varandra, söker information, kommunicerar med myndigheter och köper varor på helt andra sätt än vad vi gjorde för tio – för att inte tala om tjugo – år sedan. Internet har tagit plats i våra liv och våra liv tar i större och större utsträckning plats på internet.

Mänsklighetens kollektiva spår på internet är häpnadsväckande. Engelska Wikipedia har över 4 miljoner artiklar[2] och den svenska versionen har nästan en halv miljon,[3] LinkedIn har över 150 miljoner nätverkande användare,[4] på mikrobloggen Twitter skrivs det 340 miljoner ”tweets” varje dag,[5] det laddas upp 60 timmar video på YouTube varje minut[6] och Facebook hade 955 miljoner aktiva användare (på månadsbasis) i juni 2012.[7]

(Om du bara tänker minnas en sak i den här rapporten – se till att det inte är siffrorna ovan. De är nämligen redan gamla.)

Det som brukar kallas för webb 1.0 bestod framförallt av innehåll som skapades av enskilda individer, organisationer eller företag. De laddade upp sitt innehåll på hemsidor som andra sedan besökte. Besökarna tog del av innehållet på ungefär samma sätt som en läsare läser en bok eller en TV-tittare tittar på TV. Kommunikationen gick i huvudsak i en riktning: från producent till konsument.

Webb 1.0 sänkte trösklarna för att bli producent och internet växte till ett svårnavigerat nätverk där det snart växte fram sökmotorer som försökte göra reda i kaoset och peka ut intressant innehåll. Men den stora explosionen dröjde till det som kallas webb 2.0. Revolutionen brukar kallas *användargenererat innehåll* och innebär att alla användare kontinuerligt bidrar till att utveckla och förfina det som presenteras[8]. De fyra miljonerna artiklar på Wikipedia är inte resultatet av någon enskild utgivare som har satt sig ner och producerat innehåll för att sedan envägskommunicera ut det till den encyklopediskt intresserade allmänheten. Wikipedia är ett ständigt pågående samarbetsprojekt, en oupphörlig

[2] http://en.wikipedia.org/wiki/Main_Page, läst 2012-08-08.

[3] http://sv.wikipedia.org/wiki/Portal:Huvudsida, läst 2012-08-09.

[4] http://www.linkedin.com/, läst 2012-08-08.

[5] http://blog.twitter.com/2012/03/twitter-turns-six.html, läst 2012-08-08.

[6] http://www.youtube.com/t/press_statistics/, läst 2012-08-08.

[7] http://newsroom.fb.com/content/default.aspx?NewsAreaId=22, läst 2012-08-08.

[8] Kaplan, A M, Haenlein. M, *Users of the world, unite! The challenges and opportunities of Social Media*, Business Horizons 53, 59—68, 2010.

tvåvägskommunikation som suddar ut skillnaden mellan producent och konsument.

Den här sortens tvåvägskommunikation är grunden för det som kallas sociala medier. Några av de tekniska verktygen som möjliggör webb 2.0 är bloggar, RSS-flöden[9], Wikis, mashups[10] och taggar.[11] Internet är större än någonsin, och rymmer massor av information som kan användas på nya och oväntade sätt.

Den som gör en Google-sökning är i allmänhet mest intresserad av sökresultatet och tänker inte så noga på att även själva sökningen kan vara intressant. Men faktum är att sökningar på webben kan användas för att förutsäga andra fenomen – inte bara på internet, utan också i den verkliga världen. En studie visar att data om antalet sökningar på webben kan användas för att förutsäga hur mycket filmer och TV-spel säljer när de släpps och hur nya låtar kommer att rankas på topplistor.[12] I Sverige använder Smittskyddsinstitutet anonymiserad data om sökningar på vårdguiden.se för att uppskatta andelen patienter med influensaliknande sjukdom.[13] Naturligtvis använder även sökmotorerna själva sin information: om de flesta användare klickar på träff nummer två snarare än nummer ett så kommer Google att byta plats på dem efter ett tag.

Inte bara sökmönster utan även Twitter har visat sig kunna användas för att förutsäga framtida händelser. En annan forskargrupp lyckades förutsäga försäljningssiffror vid biopremiärer genom att mäta hur mycket det twittrades om filmerna i förväg.[14] Förutsägelserna var till och med bättre än den information som gick att köpa på Hollywood Stock Exchange, den informationsmarknad som branschen brukar använda sig av. Andra forskare har sedan byggt liknande Twitter-baserade modeller för att försöka förutsäga börskurser[15] med mera.

Att folk twittrar om vilken musik de lyssnar på eller söker efter information om filmer som snart har premiär är kanske inte så överraskande. Men (den upplevda)

---

[9] Really Simple Syndication, ett sätt att sprida information om nyheter på hemsidor och bloggar till ett stort antal prenumeranter som därmed inte själva behöver gå in och se om det har hänt något sedan sist

[10] Tekniker för att blanda innehåll från olika källor på en och samma plats

[11] Murugesan, S, *Understanding Web 2.0*, IT Professional, IEEE Computer Society, July/August 2007.

[12] Goel, S, Hofman, J M, Lahaie, S, Pennock, D M, Watts, D J, *Predicting consumer behavior with Web search*, PNAS, vol. 107 no. 41 17486-17490, October 12, 2010.

[13] http://www.smittskyddsinstitutet.se/publikationer/veckorapporter/webbsok/sasongen-20122013/, läst 2012-11-05.

[14] Asur, S, Huberman, B A, , *Predicting the Future with Social Media*, arXiv:1003.5699v1 [cs.CY]

[15] Bollen, J., Mao, H., Zeng, X., *Twitter mood predicts the stock market*, Journal of Computational Science, Volume 2, Issue 1, Pages 1-8, March 2011.

anonymiteten på internet gör att användare lägger ut betydligt mer personlig information än sin musiksmak. Medicinska studier har visat att det nu kan vara lättare för forskare att få detaljerad information om sjukdomsprocesser, medicinering och känsliga ämnen som sexuell hälsa genom diskussioner på internetforum än genom intervjuer med patienter.[16] Forskarna som gjorde undersökningen konstaterar att medan det förr var lättare att göra kvalitativa intervjuer än att göra observationsstudier är situationen nu den omvända: det är lättare att via internet samla in patientobservationer i forskningssyfte. Sådana exempel har skapat en debatt i vetenskaper som till exempel sociologi, där vissa forskare menar att man släpar efter och inte använder internet till dess fulla potential.[17]

Att använda information på smarta sätt är dock inte bara en fråga för forskare. I företagsvärlden lönar det sig att aktivt använda stora datamängder som beslutsunderlag. En studie av 179 stora börsnoterade företag visade att de företag som i stor utsträckning använde sig av ”datadrivet beslutsfattande” i genomsnitt har 5-6% högre produktivitet än de som inte gör det.[18]

De här exemplen väcker naturligtvis frågan om i vilken mån man kan använda information från webben för att identifiera kriminell verksamhet, spionage, terrorism, krigsplaner eller andra hot mot människors liv och egendom. Runt om i världen finns det gott om exempel på hur polisen på olika sätt använder sociala medier för att lösa – eller i alla fall försöka lösa – brott. Vanligast är att polisen lägger ut efterlysningar på Facebook. På samma sätt som man förr använde tidningar och affischer för att få allmänhetens hjälp kan man nu samla tips via sociala nätverk. I Morris, Illinois, rapporterade polisen i juli 2012 att man arresterat åtta personer med hjälp av Facebook-tips de senaste månaderna.[19] Liknande historier finns från hela världen och även den svenska polisen länkar nu sina efterlysningar via Facebook.

I ett uppmärksammat svenskt fall förlorade en person sin iPhone i Karlskrona. Men efter ett tag började bilder tagna med telefonen automatiskt att dyka upp på ägarens Flickr-konto via en koppling som fotografen sannolikt var lyckligt

[16] Seale, C, Charteris-Black, J, MacFarlane, A, McPherson, A, *Interviews and Internet Forums: A Comparison of Two Sources of Qualitative Data*, Qual Health Res vol. 20 no. 5 595-606, 2010.

[17] Farrell, D, Petersen, J C, *The Growth of Internet Research Methods and the Reluctant Sociologist*, Sociological Inquiry, Volume 80, Issue 1, pages 114–125, February 2010.

[18] Brynjolfsson, E, Hitt, L M. and Kim, H H , *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?* (April 22, 2011). Available at SSRN: http://ssrn.com/abstract=1819486 or http://dx.doi.org/10.2139/ssrn.1819486.

[19] http://www.morrisdailyherald.com/2012/07/17/morris-police-tracking-down-fugitives-via-facebook/aftd88j/, läst 2012-08-09.

omedveten om. Stöldgodset övervakade sig självt. Telefonen lämnades sedermera in till polisen av en man som hävdade att han köpt den i god tro[20].

Men det kan också gå snett. Polisen i staden Oakland i Kalifornien publicerar numera tid och plats för gripande av misstänkta. Ett privat initiativ, Oakland Crimespotting,[21] visualiserar det hela snyggt med en karta och en tidslinje på en hemsida. Men vid åtminstone ett tillfälle ledde det också till att alla enkelt kunde se att polisen patrullerade en viss gata på jakt efter olaga prostitution alla kvällar utom på onsdagar. Den taktiken ville man nog egentligen hemlighålla[22].

”Detektiven allmänheten” har dock inte obegränsat med tid och lust att hjälpa till, ens om man ber via Facebook. Att själv leta efter mönster i data från internet är en mer användbar metod för att bekämpa brott. I ett fall där en man i södra Sverige sexuellt trakasserade ett stort antal barn per telefon analyserade man mottagarnas geografiska positioner. Det visade sig att förövarens hem faktiskt låg i närheten av den region som hade pekats ut som mest trolig av den algoritm som användes[23].

Automatisk dataanalys kan också användas för att systematiskt förhindra brott. Facebook använder sådan teknik för att övervaka postningar och chattmeddelanden. Baserat på forskning om hur normalt chattbeteende ser ut (till exempel mönster i ålder, geografisk hemvist och intressen) [24] kan systemet varna för misstänkta beteenden som sedan kan granskas av en människa och anmälas till polisen. I juli 2012 ledde en sådan varning till gripandet av en man i södra Florida som chattat om sex med en trettonåring och planerat att träffa henne efter skolan.[25]

Det är inte bara ofrivillig information från sociala medier som är intressant för polisen att hålla reda på. Ibland lämnar brottslingar frivilligt information i form av hot. I spåren av Batman-skjutningen i Aurora i juli 2012 hotade en twittrare i början av augusti med att ställa till ett liknande blodbad på en Broadway-föreställning med Mike Tyson. När New York-polisen bad Twitter om twittrarens identitet fick de dock nej. Istället fick polisen skaffa fram ett

---

[20] http://www.svt.se/2.22620/1.2709566/polis_loser_brott_med_sociala_medier, läst 2012-08-08.

[21] http://oakland.crimespotting.org/, läst 2012-08-09.

[22] *Data, data everywhere*, The Economist, 2010-02-25, http://www.economist.com/node/15557443, läst 2012-08-09.

[23] Ebberline, J, *Geographical profiling obscene phone calls—a case study*, Journal of Investigative Psychology and Offender Profiling, Volume 5, Issue 1-2, pages 93–105, 2008.

[24] Singla, P, Richardson, M, *Yes, There is a Correlation - From Social Networks to Personal Behavior on the Web*, WWW 2008, April 21–25, Beijing, China, 2008.

[25] http://www.reuters.com/article/2012/07/12/us-usa-internet-predators-idUSBRE86B05G20120712, läst 2012-08-09.

domstolsföreläggande. Om Twitter (eller något annat företag) lämnar ut personuppgifter bara utifrån en förfrågan från polisen riskerar de att bli stämda.[26]

Analys av data på internet har tydlig potential att bidra till att skydda människors liv och egendom. Samtidigt är det svårt att veta vad som är mogna och användbara tekniker och vad som snarare är prototyper, koncept på idéstadiet eller ren marknadsföring. Den här rapporten syftar till att på ett överskådligt och lättfattligt sätt redogöra för teknikperspektivet på några av de mest lovande metoderna.

En svårighet är dock att framgångsrik analys av webbdata inte är en renodlat teknisk fråga. Tekniken interagerar med människor och då påverkas vårt beteende. Om tillräckligt många människor börjar googla efter filmer som de inte tänker se så går det inte längre att förutsäga biljettförsäljningen, trots att systemet fungerade bra nyss. Vad gäller just elektronisk övervakning finns det gott om exempel på hur system för övervakning bara delvis har nått upp till de höga förväntningar som man ursprungligen haft på tekniken. Detta har visats bland annat i studier om övervakning av skolbarns internet-surfande,[27] telemarketing-personals arbete[28] eller sjuksköterskors vård av patienter[29]. Dessutom uppvisar spanings- och analysinsatser typiskt det som ekonomer kallar avtagande marginalnytta – om man investerar tio miljoner i spaning och fångar tio brottslingar så fick man nog de lättaste, och för nästa tio spaningsmiljoner fångar man kanske bara två.[30]

Det är också lätt att som bedömare av potentialen med spaningstekniker luras av kognitiv bias och blanda ihop vad som är *möjligt* med vad som är *troligt.*[31] Det gäller såväl utopister som menar att internetspaning kan möjliggöra radikalt

---

[26] ABC News, http://abcnews.go.com/US/nypd-subpoena-twitter-identity-user-promising-violence-aurora/story?id=16947483#.UCIhYqNVJI3, läst 2012-08-09.

[27] Hope, A, *Panopticism, play and the resistance of surveillance: case studies of the observation of student Internet use in UK schools*, British Journal of Sociology of Education, 26:3, 359-373, 2005.

[28] Bain, P, Taylor, P, *Entrapped by the 'electronic panopticon'? Worker resistance in the call centre*, New Technology, Work and Employment, Volume 15, Issue 1, pages 2–18, March, 2000.

[29] Timmons, S, *A failed panopticon: surveillance of nursing practice via new technology*, New Technology, Work and Employment, Volume 18, Issue 2, pages 143–153, July, 2003.

[30] Danezis, G, Wittneben, B, *The Economics of Mass Surveillance*, Fifth Workshop on the Economics of Information Security, 2006.

[31] Dupont, B, Hacking the panopticon: *Distributed online surveillance and resistance*, in Mathieu Deflem, Jeffrey T. Ulmer (ed.*)* Surveillance and Governance: Crime Control and Beyond (Sociology of Crime Law and Deviance, Volume 10), Emerald Group Publishing Limited, pp.257-278, 2008.

bättre brottsbekämpning som dystopister som menar att brottslingarna bara krypterar allt ändå.

Som framgått av resonemanget ovan innebär analys av data från internet många möjligheter, men också stora utmaningar. Den här rapporten gör inte anspråk på att ge en heltäckande bild av webbanalys utan belyser ett antal specifika områden. De utvalda områdena speglar frågeställningar som FOI bedriver forskning kring, främst på uppdrag av Försvarsmakten men även av andra uppdragsgivare.

För att sätta bidragen i den här rapporten i ett sammanhang kan man göra en grov skiss av området webbanalys. Webben som sådan kan sägas karaktäriseras av följande påståenden:[32]

- Webben består av data, information och tjänster sammankopplade via hyperlänkar.
- Mängden data/information på webben är oerhört stor och växer snabbt.
- Webbdata är heterogen, den är representerad på en stor mängd olika format, exempelvis i form av filmer, bilder, texter och tabeller.
- Kvaliteten på information på webben varierar mycket och innehåller dessutom stora mängder brus i form av exempelvis reklam. Det är lätt för var och en att publicera information på webben och trovärdigheten varierar därefter.
- Webben är dynamisk, informationen på den förändras ständigt.
- Webben utgör ett virtuellt samhälle där människor, organisationer och automatiserade system interagerar med global och näst intill omedelbar effekt.

Manuell analys av webben är en förhållandevis rättfram uppgift, men med en naturlig begränsning i form av människans kognitiva kapacitet. Att analysera webben automatiskt innebär helt andra utmaningar. Merparten av den information som finns på webben är skapad i syfte att konsumeras av människor och den heterogenitet som data och information uppvisar när det gäller format och kvalitet kräver en omfattande förbehandling innan en analys kan påbörjas. Att automatiskt analysera stora datamängder i syfte att upptäcka mönster och ny kunskap kallas data mining. Applicerat på webbdata kallas det web mining. Inom web mining använder man sig av tre centrala kategorier*: hyperlänkstruktur, webbinnehåll och interaktionsdata.* Beroende på vilken av de ovanstående kategorierna som är dominerande kan man dela upp web mining processen i tre

[32] Inspirerat av Liu, B. *Web data mining Exploring Hyperlinks, Contents and Usage Data.* Springer 2006.

delar: web structure mining (analys och kartläggning av webbsidors länkstruktur), web content mining (extraktion och analys av innehåll) och web usage mining (analys av information om hur webben används)[33].

Rapporten är strukturerad på följande sätt. I Kapitel 3 behandlas olika aspekter av textanalys, som är en underkategori till web content mining där man fokuserar på fritext. Kapitel 4 till 6 använder sig av webbdata från all de tre centrala kategorierna, men med olika syften. Kapitel 4 beskriver hur webbdata kan användas för att skapa sociala nätverk av individer och organisationer. På grund av webbens inneboende kvalitetsvariationer blir dessa nätverk ofta osäkra vilket ställer speciella krav på analysmetoderna. Kapitel 5 tar upp användningen av alias och anonymitet på webben, medan Kapitel 6 beskriver hur man kan använda sig av olika tekniker för att försöka att identifiera potentiella ensamagerande terrorister. I Kapitel 7 behandlas integritetsskyddande tekniker som kan användas vid analys av data. Kapitel 8 behandlar trovärdighets aspekter av webbdata och avslutningsvis sammanfattas resultaten i Kapitel 9.

[33] Liu, B. *Web data mining Exploring Hyperlinks, Contents and Usage Data.* Springer 2006

3 Textanalys

Mänskliga analytiker är mycket väl lämpade att läsa, tolka, analysera och dra slutsatser utifrån dokument innehållande ostrukturerad text[34]. Det finns dock ett problem, nämligen att mängden data eller dokument av potentiellt intresse tenderar att ständigt öka, inte minst beroende på den explosionsartade ökningen av olika typer av internetresurser och sociala medier. Detta gör att fler och fler analytiker skulle krävas för att bearbeta all tillgänglig data. Den begränsade tillgången på välutbildade analytiker och kostnadsaspekter gör det dock omöjligt att manuellt behandla all data. En naturlig begränsning blir därför att bara ta de mest relevanta dokumenten i beaktande, men att på förhand veta vad som är relevant eller inte är allt annat än trivialt.

Människan är i dagsläget överlägsen datorer vad gäller att förstå och dra slutsatser utifrån ostrukturerad text och kommer med största sannolikhet att fortsätta vara det under en överskådlig framtid. Detta hindrar dock inte att olika typer av tekniska lösningar kan tillämpas för att underlätta analytikernas arbete och att bygga system där människan och datorn tillsammans mer effektivt kan bearbeta och analysera stora mängder ostrukturerad text. I detta kapitel kommer vi att titta närmare på några olika typer av tillämpningar där textanalys kan komma till nytta för underrättelseanalytiker.

3.1 Maskinöversättning

Ett område inom textanalys som redan idag används i en rad kommersiella tillämpningar såväl som vardagstillämpningar är maskinöversättning. Genom populära tjänster som Google Translate och Yahoo! Babel Fish har maskinöversättning fått en enorm genomslagskraft. De flesta som har använt denna typ av tjänster vet att resultaten långtifrån alltid blir perfekta, men det hindrar inte att de trots allt har en stor mängd användningsområden. Översättningarna duger inte till att ta ett litterärt verk skrivet på kinesiska och vänta sig att en automatöversatt version kommer att bli en storsäljare på något annat språk, däremot duger de vanligtvis till att få en någorlunda bra förståelse för vad en text handlar om i stora drag. Detsamma gäller även för de flesta andra typer av problem som textanalys kan tillämpas på. Generellt sett står sig resultaten inte gentemot mänskliga specialisters, men förmågan att på kort tid bearbeta stora mängder data överstiger vida specialistens förmåga.

[34] Med ostrukturerad text avses här text som inte har en klar struktur och därför inte är välanpassad för automatiserad bearbetning i en dator, d.v.s. består av fritext snarare än strukturerad data (där ett typiskt exempel på strukturerad data är innehållet i en relationsdatabas). Notera att även ostrukturerad text ofta innehåller någon form av struktur, så som rubriker, brödtext, etc.

Historiskt sett har maskinöversättning ofta grundats på lingvistiska metoder, där komplexa språkliga regler och ordlistor appliceras efter att texten som ska översättas analyserats med avseende på tempus, numerus, genus, etc. På senare år har dock statistiska metoder blivit alltmer populära, inte minst tack vare Google Translate som i mångt och mycket bygger på denna typ av teknik. Istället för att lingvistiska experter manuellt skapar komplexa regler så bygger statistisk maskinöversättning på att hitta statistiska mönster och samband i befintliga översättningar mellan språken man vill översätta från och till, såväl som stora mängder texter på de olika språken (för vilka direktöversättningar mellan språken inte behöver finnas). Ett exempel på en väldigt användbar resurs som kan ligga till grund för nya statistiskbaserade översättningar är de manuella översättningar som idag görs mellan en rad olika officiella europeiska språk av olika EU-dokument. Till detta tillkommer de enorma mängder text på olika språk som utgör mycket av innehållet på internet. Det är därför ingen tillfällighet att företag som tillhandahåller sökmotorer ligger i framkant vad gäller maskinöversättning.

Framöver kan vi förvänta oss att få se nya hybrider mellan lingvistisk och statistisk maskinöversättning som tar automatiserad översättning till nya nivåer långt bortom vad som är möjligt idag.

3.2 Entitetsextraktion

Entitetsextraktion, som i forskningslitteratur ofta benämns NER (Named Entity Recognition) kan användas för att i ostrukturerad text identifiera olika typer av entiteter, såsom personer, organisationer, adresser och telefonnummer. Detta kan exempelvis vara användbart för att i stora mängder av ostrukturerad information hitta information som är relevant att föra in i strukturerade databaser. Entitetsextraktion är ett relativt moget forskningsproblem där dagens bästa tekniker ofta resulterar i relativt hög precision. Resultaten varierar dock för olika typer av domäner, varför helautomatiska lösningar för entitetsextraktion inte alltid är rätt väg att gå. Ett annat alternativ är att använda sig av semiautomatiska lösningar. I de fallen kan systemet föreslå vad i texten som ska klassificeras som till exempel en person och användaren kan sedan kan välja att komplettera dessa (alternativt ta bort sådant som blivit felklassificerat som personer). Semiautomatiska lösningar kan vara en god idé som ger bra resultat samtidigt som det kraftigt reducerar den manuella insatsen.

Enkla entitetsextraktionsalgoritmer bygger ofta på fördefinierade listor med exempelvis personnamn, organisationsnamn och geografiska platser. Enkelheten hos denna typ av metoder är tilltalande men ofta inte tillräcklig. Anledningar till detta är att det inte fungerar för ovanliga namn eller platser som ännu inte finns tillagda i några listor eller som stavats fel. Detta kan delvis lösas genom att lägga till poster i listan efter hand, men även om detta görs kvarstår problemet med felstavningar och att något som i ett sammanhang kan vara ett platsnamn inte ska

tolkas som en plats i ett annat sammanhang. Som ett exempel på detta ska Washington i meningen ”Washington är USA:s huvudstad” tolkas som en plats, medan det i meningen ”George Washington har varit USA:s president” ska tolkas som en person. För de flesta människor är denna typ av så kallade *polysemi* där ett ord kan ha dubbla betydelser inget större problem eftersom semantisk information kan användas för att avgöra om det rör sig om en person eller inte, men att förstå språkets semantik är bortom dagens befintliga algoritmer. Istället är många av dagens bäst fungerande lösningar för entitetsextraktion ofta en kombination av listor, fördefinierade regler (såsom att ord som följer direkt efter titlar så som Mr., Dr., eller överste och som inleds med stor bokstav typiskt är ett personnamn) samt olika typer av mer komplexa samband som lärts in automatiskt via maskininlärning. Exempel på det senare kan vara när träningsdata som annoterats manuellt med olika typer av entiteter används för att träna klassificerare som lär sig använda syntaktiska särdrag (såsom ordklasser och meningsbyggnad) för att sedan använda inlärda mönster för att klassificera nya texter.

Entitetsextraktionsalgoritmer är idag tillräckligt bra för att börja dyka upp i en rad olika kommersiella programvaror. Exempel på produkter inom underrättelseanalysdomänen med stöd för entitetsextraktion är Palantir och Analyst’s Notebook. Dessa har även visst stöd för att utvinna relationer mellan entiteter i text. Detta fungerar dock inte lika bra som entitetsextraktion, eftersom relationsextraktion (i det generella fallet) är betydligt svårare än entitetsextraktion. I takt med att forskningen inom detta område utvecklas kan vi dock förvänta oss att se allt mer välfungerande lösningar i kommersiella såväl som icke-kommersiella applikationer.

3.3 Sentimentanalys

Den stora användningen av Twitter och andra typer av sociala medier under senare år har lett till att företag fått ett stigande intresse för att hålla koll på vad som sägs om det egna företaget, om konkurrenter samt om relaterade produkter i dessa sammanhang. Kritiska recensioner kan snabbt spridas vidare till en stor mängd personer inom den relevanta målgruppen och få stor negativ publicitet om de inte fångas upp och tas om hand i tid. Så länge företaget är litet kan det räcka med att göra en enkel nyckelordsbaserad sökning efter de inlägg som handlar om det egna företaget eller produkten och sedan manuellt läsa inläggen, men i och med att inläggen blir allt fler för varje dag så krävs det alltmer tid och resurser för att följa allt som skrivs. Detta gör att automatiserade lösningar för monitorering och analys av sociala medier blivit alltmer populärt. Som en komponent i sådana lösningar finns ofta algoritmer för sentimentanalys, vilka vanligtvis försöker klassificera ett inlägg som positivt, neutralt eller negativt.

De enklaste algoritmer som finns för sentimentanalys bygger vanligtvis på en lista med ord med tillhörande positiv eller negativ styrka. Som ett exempel kan ordet ”jättebra” i de flesta sammanhang vara starkt förknippat med ett positivt omdöme, medan ord som ”katastrof” kan antas ha en mer negativ koppling. Ord som inte återfinns i listan antas vara neutrala eller inte ha någon inverkan alls på klassificeringen, och genom att kombinera de resultat som fås för enskilda ord kan klassificeringar av meningar, stycken eller hela inlägg göras.

Det finns dock rejäla begränsningar med denna typ av förenklad lösning. En sådan begränsning uppstår vid användandet av negationer. ”Den här produkten är verkligen *inte* lättanvänd” skulle med en naiv algoritm klassas som ett positivt omdöme (baserat på ordet lättanvänd), medan de flesta människor skulle göra bedömningen att det är ett negativt omdöme. Aningen mer avancerade algoritmer kan på olika sätt hantera användningen av negationer, men ett större problem är att ett ord eller en mening som i ett sammanhang har en positiv innebörd kan ha en omvänd innebörd i ett annat sammanhang. Exempelvis så ska uttalandet ”Slutsatsen kan bara bli en, läs boken!” troligtvis tolkas som positivt om det rör sig om en bokrecension, men om det istället rör sig om en filmrecension så rör det sig mer sannolikt om en negativ recension av filmen. Detta visar på problemen med att skapa en stor generell lista med positiva och negativa ord som kan användas för alla möjliga typer av domäner. I praktiken visar det sig istället ge bättre resultat om man skräddarsyr listan med ord för den aktuella domänen. Att låta domänexperter identifiera ord som kan indikera positiva eller negativa omdömen kan dock ta mycket tid och resurser i anspråk. Därför används istället ofta olika typer av maskininlärning för att automatiskt lära in vilka ord som ska finnas med i listan och vilken tillhörande positiv eller negativ styrka de ska ha. Denna typ av lösningar visar sig i allmänhet ge bättre resultat än de som fås vid användning av listor med ord som domänexperter konstruerat. Ytterligare en fördel som uppnås är att man vid maskininlärning inte längre behöver förlita sig enbart på enstaka ord, utan kan ta andra typer av särdrag i beaktande. Exempelvis kan man använda sig av särdrag så som ordklasser (om ”springa” är ett verb ska det tolkas på ett sätt, om det är ett substantiv ska det tolkas på ett annat), användning av emoticons/smileys, eller bigram (par) eller trigram av ord snarare än bara enskilda ord (unigram). Som ett exempel på det senare så har trigrammet ”verkligen inte bra” en helt annan tolkning än de individuella orden för sig.

För att kunna använda sig av maskininlärning vid sentimentanalys krävs dock tillgång till stora mängder data, där en del av inläggen klassats som positiva, andra som negativa, och ytterligare andra som neutrala. Sådana finns ofta i överflöd för exempelvis hotell-, film- och bokrecensioner, varför denna typ av datamängder är populära bland forskare inom området. För helt andra typer av domäner kan det dock krävas att människor först märker ett antal tusen inlägg manuellt, så att algoritmerna sedan kan lära sig att klassificera utifrån dessa exempel.

De metoder som beskrivits ovan kan även användas för att klassificera texter eller inlägg utifrån andra klasser än bara positivt och negativt. I pågående forskning försöker FOI tillämpa samma typ av algoritmer för att upptäcka inlägg som har ett våldsbejakande innehåll eller där algoritmerna ska kunna klassificera känslor snarare än bara positivt och negativt innehåll. Relaterad forskning har även gjorts av forskare från Univeristy of Arizonas Artificial Intelligence Laboratory, där man försökt mäta hur mycket vålds-, ilska-, hat- och rasistrelaterat innehåll som uttrycks på olika webbforum genom att klassificera textinnehållet med hjälp av maskininlärning.

3.4 Buzz monitoring

Ord som ”buzz monitoring” och ”trendanalys” blir allt mer vanligt förekommande, men vad innebär det egentligen och på vilket sätt kan de vara relevanta för analytiker? Enkelt uttryckt skulle man kunna säga att det rör sig om olika typer av insamling och analys av de stora mängder data (”big data”) som produceras dagligen vid sökningar i sökmotorer, användargenererat innehåll från sociala medier, nyhetsflöden, etc. Ett vanligt förekommande exempel på ”buzz monitoring” är användning av Google Trends. Där kan man till exempel se vilka som är de mest använda söktermerna just nu, uppdelat på olika geografiska områden. När detta skrevs var sökningar efter ”twin towers” och ”9/11” populära söktermer i USA, då ytterligare en årsdag efter terroristattackerna mot World Trade Center och Pentagon passerat. Genom användning av Google Trends har man exempelvis kunnat hitta söktermer som är bra indikatorer på influensaaktivitet, och på så sätt gjort det möjligt att snabbare upptäcka influensaepidemier än vad som är möjligt genom att endast förlita sig på försäljningsstatistik över vilken typ av läkemedel som säljs på apoteken. Smittskyddsinstitutet använder sig av sökningar på vårdguiden.se för att analysera och kartlägga sjukdomsförekomster inom Sverige.[35]

Ett annat exempel på ”buzz monitoring”, som också berörts ovan, är när företag använder produkter som söker i sociala medier efter termer som relaterar till deras produkter och använder sig av sentimentanalys för att försöka avgöra huruvida det egna varumärket får fler eller färre positiva omdömen än tidigare.

[35] Hult, A., *Smittor lämnar spår på nätet* http://www.slf.se/SYLF/Moderna-lakare/Artiklar/Nummer-2-2012/Smittor-lamnar-spar-pa-natet/ , läst 2012-11-04.

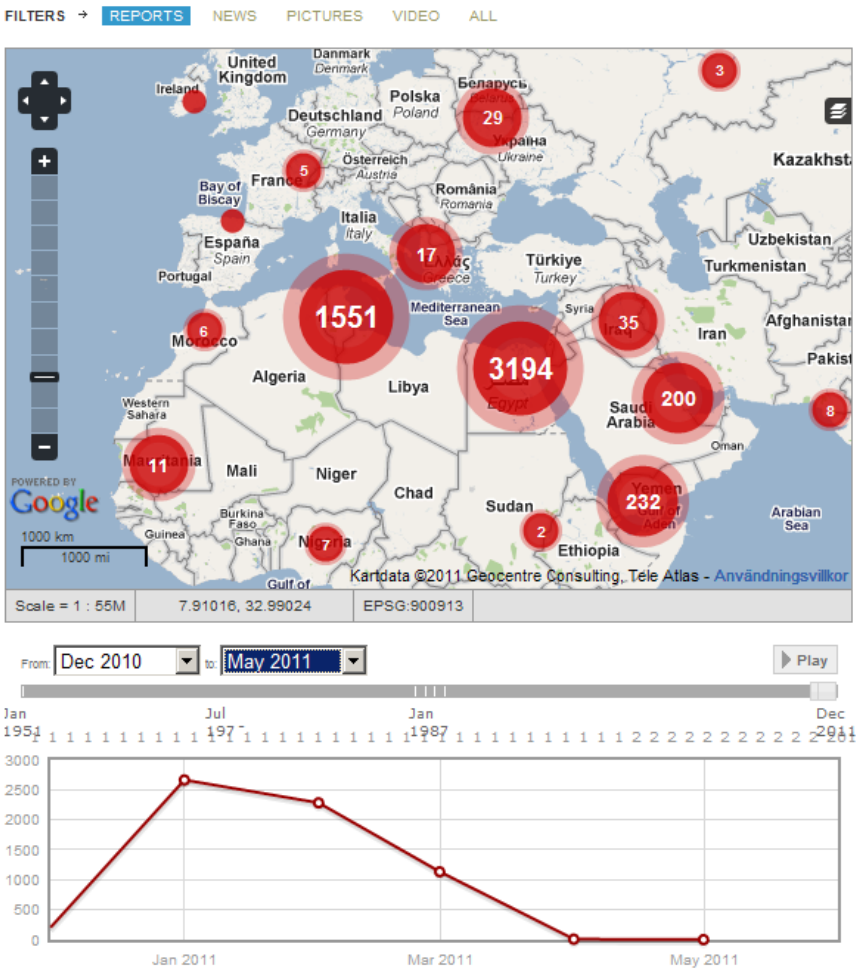
3.4.1 Trendanalys med Recorded Future och Ushahidi

Nya konflikter eller krissituationer uppstår ofta runt om i världen och alla möjligheter till tidigare förvarning skulle vara av stort värde, exempelvis i syfte att möjliggöra bättre planering och anpassad träning innan nya humanitära eller militära insatser görs eller för att skapa bättre underlag för säkerhetspolitiska analyser. Med detta som utgångspunkt har FOI tillsammans med företaget Recorded Future tagit fram ett prototypsystem för trendanalys, i vilket vi integrerar Recorded Futures produkt för insamling och analys av data från diverse olika öppna datakällor med den öppna mjukvaran Ushahidi[36]. Ushahidi används ofta i krissituationer för geospatial visualisering av insamlade ögonvittnesskildringar i syfte att ge en bättre lägesförståelse. Genom att integrera dessa system blir det möjligt att extrahera händelser som en analytiker är intresserad av (exempelvis händelser som har att göra med terrorism, hungersnöd eller demonstrationer) och visualisera dem i syfte att se uppkommande trender.[37] Systemet har använts för att samla in och analysera RSS-flöden från människorättsorganisationen Human Rights Watch och diverse nyhetsbyråer. Insamlingen gjordes huvudsakligen under första kvartalet 2011 och de rapporter som klassificerats som att handla om protesthändelser filtrerades ut (baserat på dess lingvistiska innehåll). Dessa händelser visualiserades med hjälp av Ushahidi, vilket gav användaren en möjlighet att se hur protesthändelserna spridde sig under början på den Arabiska våren (se Figur 1).

Ett problem med trendanalys av detta slag är att nyhetsmedia snabbt byter fokus, så att det som ena dagen får stor uppmärksamhet i nyhetsflödena nästa dag inte uppmärksammas alls. Ett sätt att komma tillrätta med detta är att inte ta med rapporterna från nyhetsmedier i inhämtningen utan bara fokusera på rapporter från mer nischade hjälp- och människorättsorganisationer. Ett alternativ kan också vara att i områden där användningen av sociala medier är utbredd även ta in den typen av rapporter eller användargenererat innehåll.

[36] http://www.ushahidi.com/

[37] En utförligare beskrivning finns i: Johansson, F., Brynielsson, J, Hörling, P., Malm, M., Mårtenson, C., Truvé, S., Rosell, M., *Detecting Emergent Conflicts through Web Mining and Visualization*. In Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2011.



Figur 1: Överst: Karta med händelsetypen "protest", där händelser aggregerats baserat på geografisk närhet. Underst: Histogram över antal inträffade protesthändelser i Egypten under första kvartalet 2011.

3.4.2 Analys av sociala medier för förbättrad kriskommunikation

Sociala medier har använts för kommunikation och informationsinhämtning vid en rad olika kriser, däribland Fukushima-krisen och terrorattentaten i Norge. Baserat på detta faktum pågår forskning kring hur man kan hämta in data från sociala medier som relaterar till pågående kriser och automatiskt analysera innehållet i inhämtad data för att kunna få en överblick över hur människor

reagerar på krisen och på utkommunicerade krismeddelanden. Genom användning av algoritmer för sentiment- och affektanalys kan man bilda sig en uppfattning om människors känslostämningar, vilket kan användas som beslutsunderlag när beslut ska tas om huruvida ny krisinformation ska skickas ut till befolkningen och hur den i så fall ska utformas. I en nyligen utgiven artikel[38] beskrivs ett system för analys av sociala medier med avseende på hur människor reagerar på utkommunicerade krismeddelanden.

3.5 Författarigenkänning

Författarigenkänning handlar om att avgöra vilken författare som har skrivit ett stycke text genom att studera vissa särdrag eller karaktäristiska drag i texten. Sådana karaktäristiska drag kan exempelvis vara ordval, språk, innehåll, syntaktisk stil eller kombinationer av dessa. Karaktäristiska drag för texter brukar delas upp i två kategorier: lexikala särdrag och syntaktiska särdrag. Lexikala särdrag baseras på frekvens av olika typer av ord och ordstammar medan syntaktiska särdrag baseras på olika satstyper eller andra typer av meningsuppbyggnad.

En vanlig förutsättning för att kunna göra författarigenkänning är att man utgår ifrån ett begränsat antal författare och att man för alla dessa har tillgång till texter som skrivits av författarna av intresse. Givet en ny text försöker man sedan identifiera vilken av de möjliga författarna som har skrivit texten. För att avgöra om två stycken texter har liknade karaktäristiska drag finns det ett flertal olika metoder som kan användas. Dessa metoder bygger vanligtvis på att man jämför hur ofta dessa karaktäristiska drag förekommer i texten. Om man har stora mängder data är dessa jämförelser svåra att göra beräkningsmässigt och de flesta experiment som presenteras i litteraturen har gjorts på maximalt ett hundratal författare.

Användningen av internet och texter som skrivs på internet ställer andra krav på författarigenkänning än de traditionella metoder som finns. Eftersom internet tillhandahåller möjligheten att fritt kunna publicera texter och annan typ av information under olika alias[39] så är det ofta intressant att undersöka likheter i skrivstil hos olika alias för att på så sätt kunna identifiera huruvida en författare använder sig av flera olika alias. För att hitta likheter i texter kan man använda sig av något som ibland kallas för skrivavtryck[40]. Tanken med skrivavtryck är att

---

[38] Johansson, F., Brynielsson, J., Narganes Quijano, N., *Estimating Citizen Alertness in Crises Using Social Media Monitoring and Analysis.* In Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2012

[39] Ett alias är samma sak som ett användarnamn eller en pseudonym

[40] Från engelskans writeprint

alla människor har ett mer eller mindre specifikt sätt att skriva på, vilket kan användas för att identifiera individen med, ungefär som vid användning av fingeravtryck.

Skrivavtrycket kan baseras på olika särdrag som kan extraheras från en eller flera texter. Skrivavtryck kan sedan användas för att jämföra och identifiera likheter hos författare. De särdrag som man kan använda sig av när man skapar ett skrivavtryck är exempelvis:

- Ordval
- Språk
- Syntaktiska mönster
- Återkommande felstavningar
- Användning av punkter och kommatecken
- Funktionsord (ord som används frekvent och som inte är domänspecifika)
- Uttryckssymboler (som exempelvis smileys som beskriver olika känslor)

Ny forskning visar att man genom att använda sig av skrivavtryck med relativt god säkerhet kan identifiera upp till 100 000 unika författare, [41] vilket tyder på att dessa tekniker är skalbara (d.v.s. klarar av att hantera en stor mängd författare) och därför kan vara användbara även i webbsammanhang. Experiment har gjorts på stora mängder data från bloggar där man kunnat koppla några få anonyma blogginlägg till någon av de 100 000 författare (bloggare) som fanns med i datamängden med 20% säkerhet. I ungefär 35% av fallen fanns den korrekta författaren med bland de 20 första ”gissningarna”. Detta kanske inte låter så imponerande, men kan jämföras med att sannolikheten för att slumpmässigt välja rätt författare är en på 100 000. Genom att avstå från klassificering i den hälft av datamängden där den använda algoritmen var mest osäker uppnåddes en precision på hela 80%. Viktigt att notera är också att den använda metoden bygger på ett skrivavtryck som är oberoende av domän. Om man skulle titta på andra faktorer som är signifikanta för internetanvändare (se Kapitel 4) skulle sannolikheten för att hitta rätt författare till en text troligtvis öka än mer.

[41] Narayanan, A., Paskov, H., Gong, N.Z., Bethencourt, J., Stefanoc, S., Shin, E.C.R., Song, D., On the Feasibility of Internet-Scale Author Identification. In Proceedings of the 2012 IEEE Symposium on Security and Privacy

3.6 Diskussion

Det finns en mängd utmaningar som uppkommer vid textanalys av just webbdata. En första utmaning är kvaliteten på texten som ska analyseras. Långa och välskrivna dokument är ofta att föredra vid användandet av textanalystekniker så som entitetsextraktion, maskinöversättning och sentimentanalys. Tweets och andra typer av användargenererat innehåll i sociala medier utgörs dock snarare av motsatsen: korta texter som inte sällan utmärks av en stor mängd felstavningar, förkortningar, specialuttryck och ironi. Detta gör ofta utmaningen betydligt större än för mer traditionella typer av textdokument.

En andra utmaning är den stora mängden data. Även om ett inlägg eller tweet i sig är kort, så finns det ofta en ofantlig mängd av dem. Detta ställer stora krav på de algoritmer som används i form av effektivitet och beräkningskomplexitet. Eftersom det är så stora mängder data som ska hanteras måste processandet av enskilda inlägg gå mycket snabbt.

Slutligen finns det också en utmaning i att hämta hem all data som ska analyseras. Om ambitionen exempelvis är att analysera Twitter-strömmar i realtid så kan detta ställa stora krav på både den hårdvara och mjukvara som används. För andra typer av användargenererat innehåll kan det krävas mycket arbete med specialskriven mjukvara för att automatiskt logga in på ett forum, identifiera vad på sidan som är data av intresse att ladda ned, och att slutligen ladda ned dessa data utan att lämna några spår eller störa övrig aktivitet på webbservern (exempelvis genom att för stora datamängder laddas ned på kort tid och därmed ger återverkningar på övriga anrop till webbservern).

4 Kartläggning och analys av nätverk

I detta kapitel redogörs för hur textanalysmetoder som entitets- och relationsextraktion kan kombineras med så kallad ”social network analysis” (SNA) för att analysera och kartlägga nätverk baserat på textuell information inhämtad från webben. Då denna typ av information vanligtvis är behäftad med osäkerhet presenteras först en metod för att hantera analys av osäkra sociala nätverk (vilket vi refererar till som osäker SNA).

Utöver användningen av osäker SNA för att kartlägga nätverk baserat på textuell information så är kunskap om sociala nätverk även användbart för andra ändamål kopplade till webbanalys. Till exempel finns det många sociala medier som bygger på olika former av nätverksbildningar (Facebook och LinkedIn för att nämna några). Ett annat tillfälle då SNA kan vara användbart är vid analys av olika former av individer och grupperingar på diskussionsforum eller bloggar. Som ett konkret exempel på detta kan SNA exempelvis användas för att hitta så kallade ”key influencers”, d.v.s. personer som i hög grad påverkar andra. Nedan följer först en beskrivning av ”traditionell” SNA, där vi översiktligt presenterar teorin bakom SNA och några av de centralitetsmått som ofta används. Därefter beskrivs några av de problem som gör att traditionell SNA är svår att tillämpa för webb- och underrättelseanalys. Baserat på dessa observationer föreslås osäker SNA som en lämpligare kandidat att använda inom webb- och underrättelseanalys. Vi beskriver den föreslagna metoden, samt ger exempel på hur osäker SNA kan användas på nätverk som har extraherats med hjälp av textanalys.

4.1 Analys av sociala nätverk

Sociala nätverk är en teoretisk modell som traditionellt använts framför allt inom sociologin för att studera relationer mellan individer, grupper och organisationer. Ett socialt nätverk består av en mängd aktörer och relationer mellan dessa aktörer. Aktörerna representeras vanligtvis som noder i ett nätverk, medan relationer mellan aktörerna beskrivs med hjälp av länkar (även kallat kanter) mellan noderna. Dessa nätverk kan vara komplexa, dels eftersom de kan bli väldigt stora, dels beroende på att det kan finnas flera olika typer av både aktörer och relationer.

Förutom att använda sig av sociala nätverk för att visualisera aktörer och relationer (som exempelvis vilka som är vänner eller skriver i samma diskussionsforum) kan det även vara intressant att analysera nätverken i sig, och det är detta som åstadkoms med hjälp av SNA.

Det finns framför allt två olika kategorier av analyser som kan göras av sociala nätverk. Den första kategorin är algoritmer som med hjälp av olika typer av mått

beräknar numeriska värden för varje nod i nätverket (d.v.s. för varje aktör i nätverket). Denna typ av mått kan exempelvis användas för att plocka fram de aktörer som är mest centrala i nätverket. Den andra kategorin av analyser är *klustring* av noder. I detta fall beräknar man inte ett mått för varje enskild aktör utan målet är att gruppera aktörer utefter en bestämd definition, så som att maximera antal kopplingar inom ett kluster samtidigt som man minimerar antalet kopplingar mellan olika kluster. Detta kan användas för att hitta delgrupperingar i det större nätverket.

Traditionell SNA kan användas för en mängd olika saker. Man kan till exempel använda det för att kartlägga ett företags informella och formella kontaktnät för att kunna skapa fördelar i samarbeten med samarbetspartners eller för att sätta ihop effektiva team inom organisationer. Kopplat till underrättelsedomänen används det istället ofta för att hitta centrala aktörer och delgrupperingar inom exempelvis kriminella organisationer eller terroristnätverk. Denna typ av tillämpningar kommer att beskrivas mer i detalj längre fram, men för att kunna göra detta måste vi först beskriva några av de centralitetsmått som ofta används inom SNA och på vilket sätt klustringsalgoritmer kan användas.

4.1.1 Centralitet

Centralitet är ett av de mest använda begreppen inom analys av sociala nätverk. Centralitet är ett mått på hur central en person är i ett nätverk, men det finns flera olika definitioner för vad detta egentligen innebär. Några av de mest vanligt förekommande varianterna av centralitet är degree centrality, closeness centrality, betweenness centrality och eigenvector centrality. I den här rapporten har vi valt att använda de engelska namnen på de olika centralitetsmåtten eftersom det inte finns standardiserade svenska namn. Nedan följer en kortfattad beskrivning över vad de olika centralitetsmåtten betyder och hur de kan användas.

Degree centrality är ett mått på det antal direkta länkar till andra aktörer i nätverket som en aktör har. De aktörer som har många direkta länkar till andra kan antas vara viktiga för nätverket eftersom de inte är beroende av andra aktörer och har möjlighet att nå ut till stora delar av nätverket.

Closeness centrality kan beskrivas som ett mått på hur nära de övriga aktörerna i nätverket är till den aktuella noden i fråga.[42] En aktör kan anses vara viktig om den är relativt nära resten av aktörerna i nätverket och closeness centrality är det mått som beskriver denna egenskap.

[42] För att beräkna closeness centrality beräknas först summan av alla de kortaste vägarna till alla aktörer. Inversen av summan är aktörens closeness centrality.

Betweenness centrality beskriver hur viktig en person är som länk mellan olika nätverk. En persons vänkrets på Facebook kanske inte är så stor men det utesluter inte att det är genom den personen som information passerar mellan stora nätverk som i övrigt har få eller inga förbindelser.

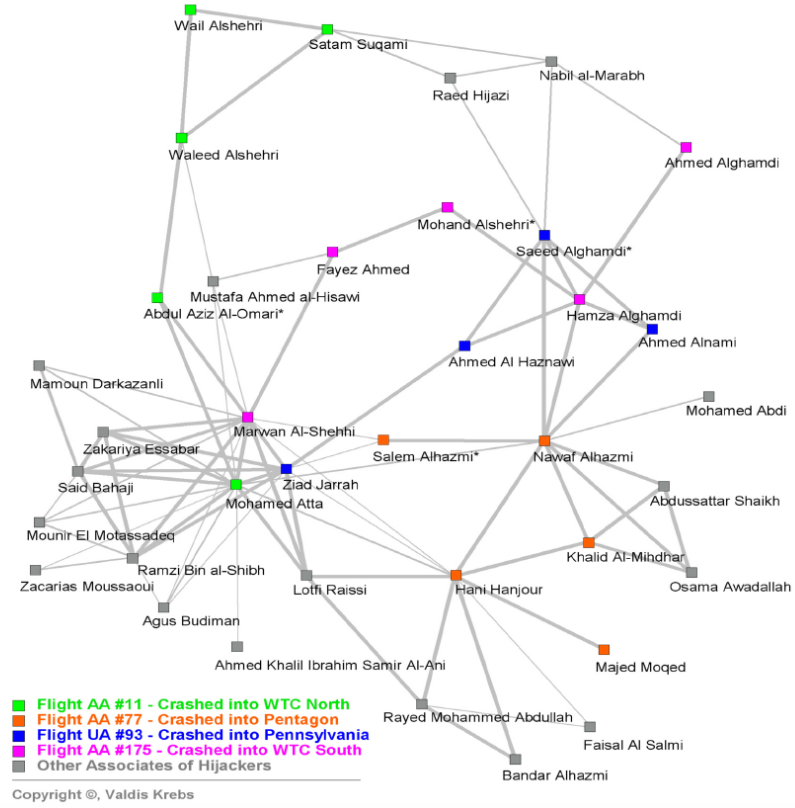
Eigenvector centrality är ett mått på hur central en aktör i ett nätverk är, med hänsyn till den globala strukturen på nätverket. Varje aktör i nätverket får en ”poäng” som beror på principen att länkar till aktörer med höga poäng ger en bättre utdelning än länkar till noder med låga poäng. Rent beräkningsmässigt åstadkoms detta genom att initialt sätta denna poäng till 1 för alla noder i nätverket. Därefter uppdateras poängen genom att varje nods poäng beräknas om som en viktad summa av poängen för nodens grannar, varpå poängen normaliseras. Detta återupprepas tills de beräknade poängen inte längre ändras, vilket innebär att de beräknade poängerna slutligen kan användas som ett mått på aktörernas eigenvector centrality.

Google använder sig av en variant av eigenvector centrality i sin PageRank-algoritm. PageRank är den algoritm som avgör hur relevanta webbsidor är för sökfrågan och hur resultaten av sökningen ska presenteras.

De olika centralitetsmåtten används vanligtvis för att identifiera olika typer av nyckelaktörer i sociala nätverk. Man kan till exempel använda sig av en kombination av olika centralitetsmått för att identifiera nyckelpersoner i ett terroristnätverk[43]. Som ett exempel på detta kan Krebs[44] välkända analys av kaparna som genomförde 9/11-attacken nämnas, vilket återfinns i Figur 2. I detta exempel kan man se hur Mohammed Atta identifierats som en av nyckelpersonerna i nätverket genom hans höga degree centrality.

[43] Berzinji, A., Kaati, L., Rezine A, *Detecting Key Players in Terrorist Networks* In Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2012.

[44] Krebs, V. E., *Mapping Networks of Terrorist Cells*, Connections 24(3): 43-52.



Figur 2: Socialt nätverk för 9/11-kaparna (från Krebs, 2002)

4.1.2 Kluster och grupperingar

Ofta räcker inte centralitetsmått till för att analysera sociala nätverk. Ett exempel är när man är intresserad av vilka personer i ett socialt nätverk som är de mest inflytelserika. Den typen av information är viktig inom marknadsföring när ett företag lanserar nya produkter eftersom ett sätt att marknadsföra nya produkter på är att ge inflytelserika personer gratisprodukter för att dessa sedan ska rekommendera produkten till sina vänner eller skriva om den i sociala medier. Forskning har visat att det inte bara är antalet kopplingar till andra aktörer i nätverket (degree centrality) som påverkar hur inflytelserik en person är.[45]

[45] Z. Katona, P. P. Zubcsek, and M. Sarvary. *Network Effects and Personal Influences: The Diffusion of an Online Social Network*. Journal of Marketing Research (JMR) 2011.

Istället verkar det spela en stor roll hur vänners vänner är kopplade till varandra, d.v.s. hur mycket någon påverkar dig beror på hur väl de känner dina vänner. Det är till exempel större sannolikhet att man påverkas av en nära vän (som känner flera av dina vänner) än en avlägsen vän som har många vänner när det gäller råd inför byte av bil eller bostad. För att kunna identifiera inflytelserika personer måste man således använda sig av klustringsalgoritmer för att först skapa grupper (kluster) i nätverket där varje grupp består av de aktörer som känner varandra. Ett exempel som är mer relaterat till underrättelsedomänen är att identifiera personer som exempelvis uppmanar till terroristattacker eller att hitta viktiga personer att påverka vid psyops-operationer.

Det finns en mängd olika klustringsalgoritmer som på olika sätt försöker dela upp ett nätverk i grupper genom att maximera antalet kanter inom varje kluster. En klustringsalgoritm som delar upp ett nätverk i mindre grupper (så kallade communities) börjar vanligtvis med att dela upp nätverket i mindre och mindre grupper tills nätverket är uppdelat i ett på förhand önskat antal grupper.[46]

4.1.3 Begränsningar med traditionell SNA

Potentialen med att använda SNA för att hitta centrala aktörer och delgrupperingar inom såväl diskussionsforum och terrornätverk torde vara uppenbar och det finns följaktligen många exempel på där SNA använts för denna typ av tillämpningar i vetenskapliga arbeten. Det finns dock en rad begränsningar och svårigheter som gör det svårt att applicera denna typ av tekniker rakt av för både webbdata i allmänhet och underrättelseanalys i synnerhet. Dessa begränsningar belyses sällan tillräckligt.

En första stark begränsning är framtagandet av själva det sociala nätverket. Detta tas ofta för givet, men i praktiken är det sällan man har perfekt kunskap om det underliggande nätverket som ska modelleras. Inom traditionell analys av nätverk inom sociologin samlas data om aktörer och relationer vanligtvis in genom utförliga studier, enkäter, observationer, etc. Inom underrättelsedomänen är det dock sällan uppenbart vilka aktörer som ska tas med i det sociala nätverket, och framförallt, vilka relationer som existerar mellan aktörer. Även sociala nätverk kopplade till sociala medier kan ofta vara svåra att kartlägga fullt ut. Om man som analytiker till exempel vill kartlägga någon gruppering på nätet är det sällan så att all information om de ingående individernas relationer finns direkt tillgänglig. Istället kan man ofta i bästa fall hoppas på att nysta upp delar av nätverket genom att vissa av de ingående individerna har sitt närmaste vän-nätverk publikt åtkomligt (i Facebook-fallet) eller genom att via en delmängd av alla tweets kunna bygga upp ett icke-perfekt nätverk av vilka personer som re-

[46] En detaljerad översikt av olika algoritmer som kan användas för att identifiera delgrupperingar i nätverk återfinns i: Dahlin, J., *Community Detection in Imperfect Networks*, 2011, FOI-S--3710--SE.

tweetar varandra (i Twitter-fallet). Dessa är bara exempel, men visar på att en analytikers modell av ett socialt nätverk ofta bara är en approximation av det riktiga underliggande nätverket. Det är viktigt att påpeka att det i vissa fall finns välstrukturerade databaser som innehåller den information som krävs för att skapa nätverket, men detta är långt ifrån alltid fallet.

En annan begränsning är att man inom traditionell SNA betraktar relationer som att de antingen finns där eller också inte, d.v.s. man tar sällan hänsyn till att relationer kan vara olika starka eller att det kan råda osäkerhet kring huruvida det faktiskt finns en relation mellan två personer. Den typen av osäkerhet är snarare regel än undantag inom underrättelsedomänen, men trots detta så negligeras denna aspekt ofta vid användandet av SNA. I vissa fall modellerar man styrkan på en relation genom att använda vikter på kanterna, men detta är långtifrån vanligt förekommande och det råder ingen enighet om hur centralitetsmått och klustringsalgoritmer ska anpassas för att hantera sådana viktade nätverk.

4.2 Osäker SNA

För att komma till rätta med problemen och begränsningarna med att tillämpa analys av nätverk för webbdata i allmänhet och underrättelsedomänen i synnerhet så har FOI utvecklat en metod för att hantera osäkerhet i nätverket. Denna osäkerhet kan gälla olika fenomen som huruvida två aliasnamn relaterar till en och samma individ (se även Kapitel 5), huruvida en aktör existerar eller ej, eller huruvida det ska finnas en relation mellan två eller flera aktörer. I detta delkapitel ges en översiktlig bild av metoden för att hantera osäkerheter vid analys av sociala nätverk.[47,48,49]

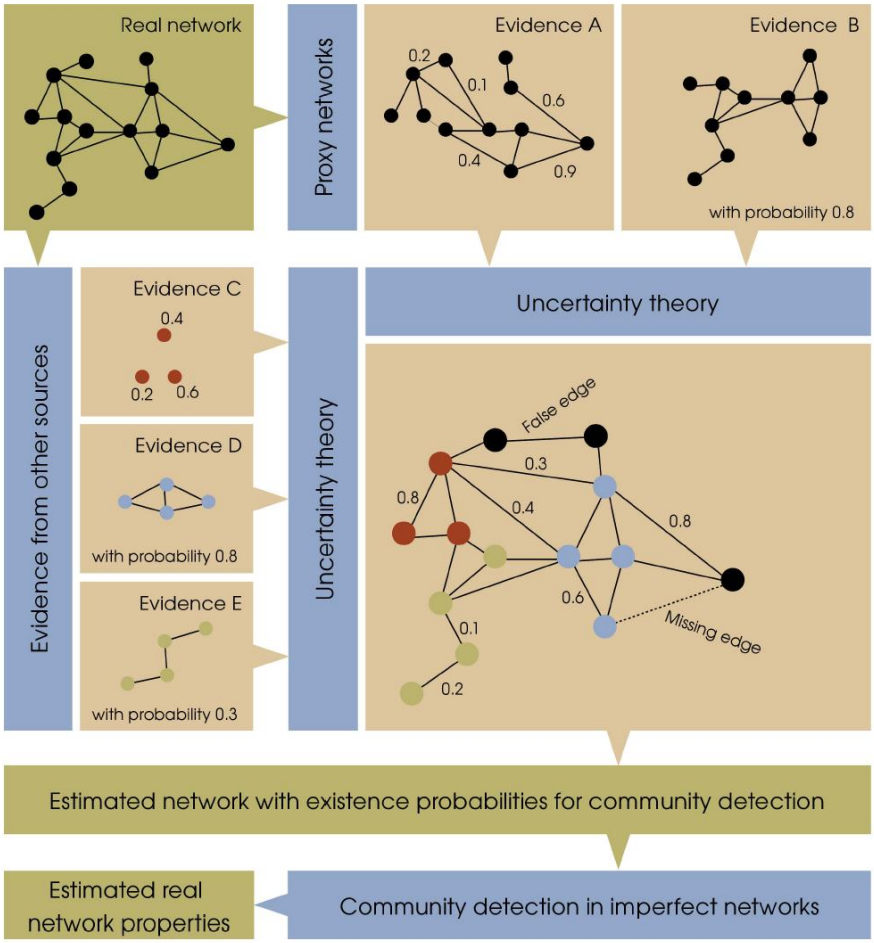
Utgångspunkten i metoden är att det verkliga (underliggande) sociala nätverket inte är direkt observerbart utan att olika så kallade proxy-nätverk istället är det som observeras och får användas som approximation av det verkliga nätverket. Ett exempel på ett sådant proxy-nätverk är när kommunikation mellan aktörer (där sådan kommunikation exempelvis kan identifieras via loggning av mailkonversationer) används som proxy för olika typer av mer abstrakta relationer av intresse. Utöver denna typ av proxy-nätverk går det även att få in annan typ av evidens från olika källor, så som att en källa säger sig ha sett två

[47] För mer information om hur osäkerheter kan hanteras vid analys av sociala nätverk, se: Dahlin, J., *Community Detection in Imperfect Networks*, 2011, FOI-S--3710--SE.

[48] Dahlin, J., Svenson, P., *A Method for Community Detection in Uncertain Networks*, In Proccedings of the European Intelligence and Security Informatics Conference (EISIC) 2011

[49] Johansson, F., Svenson, P. *Constructing and Analyzing Uncertain Social Networks from Unstructured Textual Data*. Accepterad för publicering i Studies in Mining Social Networks and Security Informatics.

aktörer kommunicera med varandra (där utsagan från källan alltså gäller små delar av nätverket snarare än hela nätverket, vilket är fallet för proxy-nätverk). Skillnaden mellan evidens i form av proxy-nätverk och andra typer av evidens illustreras i Figur 3.



Figur 3: Illustration av den allmänna metoden för att analysera nätverk med osäker data.

Figur 3 illustrerar också hur osäkerhetsteori används för att kombinera olika evidens (exempelvis är evidens C i figuren ett exempel på osäkerhet rörande noders existens och evidens D ett exempel på osäkerhet rörande en större struktur) och låta detta resultera i ett osäkert nätverk. Själva fusionen ligger

utanför denna rapports ramar,[50] men olika typer av osäkerhetsteori kan användas för detta ändamål, så som Bayesiansk sannolikhetsteori eller Dempster-Shaferteori. Den resulterande osäkerheten för relationerna kan exempelvis modelleras med hjälp av sannolikheter mellan 0 och 1 som associeras med kanterna i grafen.

När ett osäkert nätverk skapats på detta sätt är nästa steg att applicera centralitetsberäkningar eller klustringsalgoritmer på det osäkra nätverket. Det finns exempel i forskningslitteraturen på hur centralitetsmått och klustringsalgoritmer kan anpassas för viktade nätverk, men dessa ger inte ”korrekta” resultat för osäkra nätverk i de fall där osäkerheten gäller relationer mellan klickar[51] (exempelvis trianglar[52]) i nätverket, snarare än endast par. För att tydliggöra detta skulle uttalandet ”med 80 % sannolikhet är Lisa, Anders och Nils vänner” vara ett exempel på en triangel (klick av storlek 3) med tillhörande osäkerhet. Vidare så är inte dessa metoder tillämpbara i situationer där osäkerheten gäller nodernas existens snarare än kanternas existens. På grund av detta har en speciell metod för klustring i osäkra nätverk utarbetats, vilken även är tillämpbar vid beräkningar av centralitetsmått. I korthet går metoden ut på att sampla ett stort antal (säkra) nätverk från det osäkra nätverket (där dessa samples är konsistenta med informationen i det osäkra nätverket), göra en traditionell nätverksanalys av de samplade nätverken och sedan slå samman resultaten. Denna metod har testats på så väl verkliga och syntetiska nätverk med genererade osäkerheter och brister, där resultaten indikerar att metoden klarar av att återidentifiera delgrupper i nätverk där hälften eller mer av relationerna i nätverket känns till med säkerhet. Möjligheten att på detta sätt hantera osäkerhet vid analys av sociala nätverk är något som inte finns i kommersiella programvaror utan bara i forskningsprototyper.

4.3 Extraktion av entiteter och relationer från ostrukturerad text

Det finns, som nämnts ovan, tillfällen då man snabbt vill kunna kartlägga ett nätverk eller en organisation utan att ha tillgång till den välstrukturerade data som vanligtvis krävs för att tillämpa SNA. Hur ska till exempel en analytiker snabbt kunna göra en översiktlig kartläggning av en organisation som tagit på sig

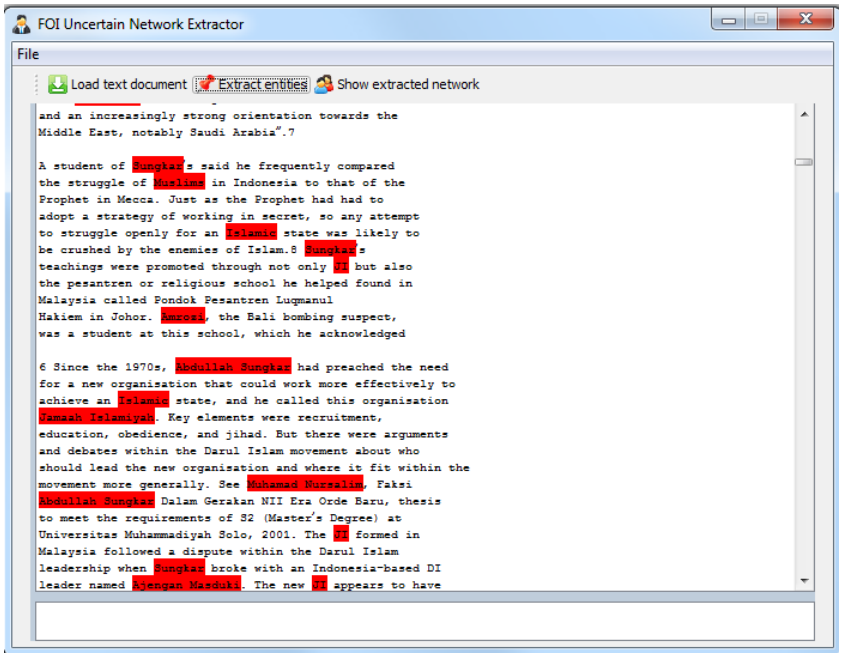
[50] Detaljer återfinns istället i Dahlin, J., *Community Detection in Imperfect Networks*, 2011, FOI-S--3710--SE.

[51] En klick är en mängd noder som bildar en subgraf i en graf, sådan att delgrafen är komplett, d.v.s. att det i delgrafen finns en kant mellan varje par av noder.

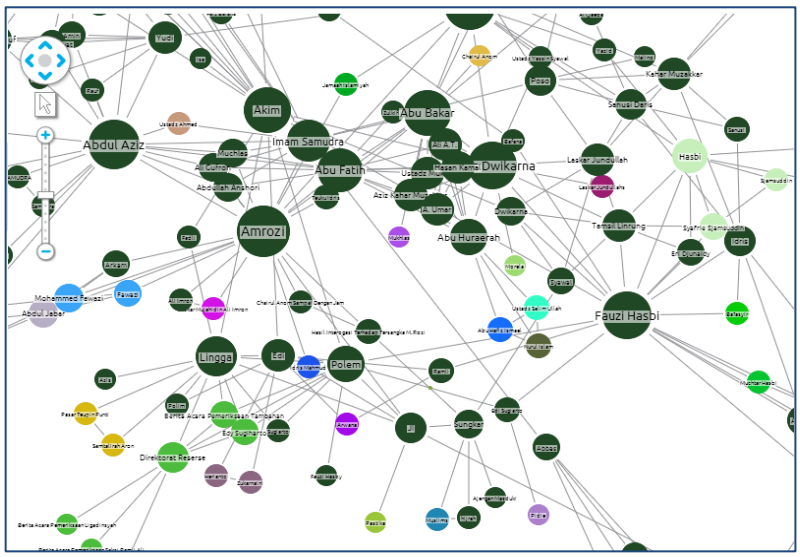
[52] En triangel är i detta sammanhang en klick av storleken 3, d.v.s. en subgraf där alla tre noder är kopplade till varandra.

ett terroristattentat och som man sedan tidigare bara har en väldigt grundläggande kunskap om?

Ett sätt att lösa den här typen av problem är att utifrån ostrukturerade textdokument automatiskt extrahera entiteter och relationer för att på så sätt generera en icke-perfekt modell (ett proxy-nätverk) av ett socialt nätverk som texten implicit beskriver. För att undvika att få med ord som blivit felklassificerade som personer (det vill säga ett ord som den automatiska extraheringen tror är en person men som i själva verket är något helt annat) går det att sätta en undre gräns för hur många gånger i textmassan som en ”person” måste nämnas för att komma med i det extraherade nätverket. Relationsextraktionen går att göra mer eller mindre komplex, men en enkel och rättfram modell är att anta att personer eller andra typer av entiteter som förekommer tätt intill varandra i meningar har starkare relationer än personer som förekommer långt isär. På samma sätt går det att väga in antalet gånger som två (eller fler) personer förekommer tillsammans när bedömningen görs av styrkan på en relation mellan dessa personer. Vid förekomst av två eller fler textdokument som beskriver samma typ av nätverk kan dessa modelleras som enskilda proxy-nätverk och sedan fusioneras samman på det sätt som beskrivits ovan. På så sätt leder användandet av entitetsextraktion och relationsextraktion fram till ett osäkert socialt nätverk. Centralitetsberäkningar och klustringsalgoritmer kan sedan appliceras på detta nätverk enligt vad som beskrivits ovan. Vid denna typ av tillämpning av osäker SNA kan det vara av avgörande betydelse att känna till sakriktigheten hos det som står beskrivet i textdokumenten samt att känna till källornas tillförlitlighet. Genom att känna till detta går det att vikta de olika textdokumenten på olika sätt utifrån deras innehåll. Hur man kan bedöma informationskvalitet och trovärdighet berörs vidare i Kapitel 8. En skärmdump av det verktyg som tagits fram baserat på resonemanget i detta stycke kan ses i Figur 4. För att exemplifiera hur verktyget kan användas har ett antal dokument som beskriver terroristorganisationen Jeemah Islamiyah laddats ned från nätet (www.crisisgroup.com) och lästs in, varpå ett resulterande osäkert nätverk har extraherats. Delar av det framtagna nätverket visas i Figur 5, där nodernas storlek beror på dess degree centrality, det vill säga antalet direkta kanter (viktat med styrkan på kanterna). Färgen på noderna illustrerar vilket kluster de tillhör (det vill säga varje färg utgör ett separat kluster), som ett resultat av den osäkra klustringsalgoritm som applicerats.



Figur 4: Skärmdump från verktyg för att extrahera osäkra sociala nätverk från ostrukturerad text.



Figur 5: Visualisering av delar av det resulterande osäkra nätverket som extraherats från textdokument som handlar om Jeemah Islamiyah.

4.4 Diskussion

Det finns många aspekter av SNA som inte fått utrymme i detta kapitel. Utöver att använda algoritmer för att beräkna centralitet och identifiera kluster finns det även många andra typer av egenskaper hos nätverket som kan beräknas, däribland densiteten och diametern hos nätverket. Även sådana egenskaper kan vara mycket användbara vid analys av sociala nätverk på webben. En annan mycket intressant aspekt är att undersöka hur nätverk utvecklas över tid. Detta har traditionellt inte getts så mycket uppmärksamhet inom SNA-forskningen eftersom det tidigare krävt att tidsödande datainsamling gjorts vid flera olika tillfällen, men för information om sociala nätverk som kan extraheras på ett mer automatiserat sätt (vilket är typiskt för webbdata) kan temporal SNA komma att få ett stort genomslag.

En annan tillämpning där man kan använda sig av SNA är när man försöker upptäcka om en användare har flera alias, vilket beskrivs i nästa kapitel.

5 Anonymitet och alias

Internet är en utmärkt plattform för att kommunicera åsikter och idéer. Internet tillåter användare att vara anonyma när de publicerar information vilket gör det svårt att få en överblick över vem som egentligen står bakom en åsikt eller ett uttalande. En ytterligare försvårande faktor är att en användare kan välja att ha ett flertal olika alias. Det finns olika anledningar till att en användare väljer att använda sig av flera alias. En orsak kan vara att en användare använder sig av ett eller flera extra alias för att stödja sina egna åsikter (så kallade alter ego-alias). Andra orsaker kan vara att användaren glömt bort sitt lösenord, att det ”favoritalias” som användaren brukar använda redan är upptaget eller att det vanliga aliaset helt enkelt inte uppfyller de formatkrav som kan ställas på ett alias på en del diskussionsforum eller andra typer av sociala medier. Ett exempel på någon som använde sig av flera alias när han kommunicerade på internet är Anders Behring Breivik[53]. I hans fall skulle det kunna ha varit intressant att koppla ihop olika typer av uttalanden som gjordes av hans olika alias. För att kunna identifiera en användare som har flera alias kan man använda sig av en teknik som kallas för aliasmatchning. Aliasmatchning bygger på att man tittar på flera olika egenskaper hos ett alias och sedan försöker avgöra om dessa tillhör en och samma användare.

De faktorer som man kan titta på för att identifiera och matcha multipla alias är framför allt det skrivavtryck (se Kapitel 3) som kan utvinnas ur de texter eller inlägg som varje alias har producerat. Förutom skrivavtrycket finns det även andra faktorer som kan användas. Några av dessa är:

1. Vilka tider på dygnet som ett alias publicerar sina inlägg och mängden inlägg som publicerats under längre tidsperioder (som exempelvis månader och år).
2. Likheter i aliasnamn
3. Huruvida meddelanden skrivs i samma diskussionstrådar
4. Likheter i nätverk av vänner för olika alias

Tid

Även om internet tillåter kommunikation dygnet runt är det mer troligt att en användare skriver sina inlägg under ungefär samma tidsperiod på grund av faktorer som regelbundenheter i arbets- och sovvanor. Genom att dela upp dygnet i ett antal olika delar kan man använda information om när på dygnet alias är aktiva för att koppla dessa till en användare. Information om när på

[53] Några av Breiviks identifierade alias är: Anders B. Breivik, Andersnordic, Andrew Berwick och Conservatism.

dygnet ett alias är aktivt kan användas både för att identifiera alter ego-alias och alias som används på olika delar av internet.

Man kan även titta på ett alias aktivitet under längre tidsperioder och använda den informationen för att jämföra olika alias. I det fallet är det framför allt anomalier som brist på aktivitet under vissa perioder som är av intresse (man kan till exempel tänka sig att sjukdom eller resor kan vara en orsak till att inget av en användares alias är aktivt under perioder). Att två alias har skrivit inlägg under ungefär samma tidsperioder säger givetvis inte att de behöver vara en och samma individ (det troligare scenariot är snarare att de bara bor i samma tidszon och har liknande dygnsrytm), men kombinerat med andra typer av information kan tidsinformation vara användbart vid aliasmatchning. Även frekvensen med vilken ett alias författar inlägg kan användas för att karaktärisera det.

Likheter i aliasnamn

Om en användares favoritalias är upptaget är det vanligt att man väljer ett liknande alias. Ofta ger den tjänst där man registrerar sitt alias förslag på likande alias om det önskade aliaset är upptaget. Ett vanligt sätt är att lägga till siffror, tecken eller några bokstäver till det önskade aliaset. För individer som inte aktivt försöker dölja att två alias tillhör en och samma användare kan likheter i aliasnamn därför vara en mycket användbar indikator vid aliasmatchning.

Likheter i trådar

Om en användare skapat ett så kallat alter ego-alias i syfte att stödja sina åsikter är det troligt att dessa alias skriver i samma diskussionstrådar eller på andra sätt är aktiva på samma ställen. Information om vart aliasen är aktiva kan i dessa fall användas för att identifiera en användare. På samma sätt som tid är trådlikhet i sig inte tillräckligt för att matcha samman alias, men kan vara användbart som komplement till andra indikatorer.

Sociala nätverk

En användare som skapat ett nytt alias av den anledningen att denne glömt sitt gamla lösenord eller att det gamla aliaset blivit bannlyst från exempelvis ett diskussionsforum vill sannolikt gärna ha kvar alla eller delar av sitt ursprungliga nätverk av vänner. För att identifiera en användare som har olika alias av en sådan anledning kan man använda sig av analys av sociala nätverk. I de fallen kan man titta på likheter i de sociala nätverken för olika alias, en teknik som kallas för strukturell ekvivalens. Genom att kombinera några eller alla av de ovanstående faktorerna kan man få en bättre bild över hur troligt det är att flera olika alias används av en och samma användare.[54] Beroende på anledningen till att en användare har flera alias väljer man ut de faktorer som är relevanta för det

[54] Mer information om aliasmatchning finns i: Dahlin, J., Johansson, F., Kaati, L., Mårtenson, C, Svenson, P., *Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Boards.* International Symposium on Foundation of Open Source Intelligence and Security Informatics 2012.

fall man analyserar. Om användaren har skapat ett alter ego-alias som används för att underbygga sina åsikter i ett diskussionsforum är det till exempel inte troligt att användaren väljer ett liknande aliasnamn men det är däremot troligt att de olika aliasen skriver i samma diskussionstrådar. På samma sätt är det mer troligt att en användare väljer ett liknande aliasnamn om användaren är aktiv på ett flertal forum eller om användaren har tappat sitt lösenord och måste skapa sig ett nytt alias.

5.1 Individen bakom ett alias

Man kan mycket sällan identifiera en unik fysisk individ enbart baserat på att ett eller flera alias av intresse har identifierats. Det finns vanligtvis goda möjligheter att skapa alias utan någon spårbar anknytning till den egna individen, och även i de fall där det krävs att användare verifierar sin identitet med en giltig e-postadress så finns det inget som hindrar att en separat e-postadress skapas enbart för detta. Även om IP-adressen som används vid skapandet av e-postkontot eller inlägget sparas och lämnas ut till polisen så finns det heller inget som hindrar att individen använder sig av någon typ av anonymitetstjänst eller ett internetcafé för att förhindra att bli spårad. Själva identifikationen av vilken individ som ligger bakom ett alias är dock utanför denna rapports ramar.

5.2 Diskussion

Att identifiera användare med flera alias är ett svårt problem, inte minst beroende på att det kan finnas många olika anledningar till att flera olika alias används. Hur pass bra resultat man kan förvänta sig av den här typen av tekniker är svårt att säga något om eftersom det i dagsläget inte har gjorts tillräckligt mycket experiment. Genom att kombinera tekniker för författarigenkänning med andra faktorer kan man i alla fall tänka sig att sannolikheten för att kunna identifiera att flera olika alias tillhör en och samma användare ökar.

En sak som man däremot kan säga någonting om att är möjligheten är att kunna identifiera att en användare använder sig av flera olika alias genom aliasmatchning är av stor vikt när man letar efter digitala spår som kan tyda på att en person planerar att begå ett terrorbrott, vilket beskrivs i följande kapitel.

6 Jakten på ensamagerande terrorister på internet

Som tidigare nämnts tar våra liv i större och större utsträckning plats på internet och detta gäller även mer ljusskygga delar av samhället. Därför är internet en viktig plats för att leta efter exempelvis kriminell verksamhet och andra typer av hot mot samhällets säkerhet. Ett påtagligt och aktuellt hot är ensamagerande terrorister vilket uppmärksammades på ett tragiskt sätt den 22 juli 2011 i Norge då Anders Behring Breivik utförde två terrorattentat som resulterade i 77 döda. Eftersom ensamagerande terrorister arbetar på egen hand kan man inte samla information om dem genom traditionellt polisarbete som exempelvis infiltrering och avlyssning av kända extremistgrupper. Ett sätt att försöka upptäcka dem i tid är att analysera data från internet i jakt på digitala spår. Det har visat sig att ett flertal terrorister har använt sig av internet för att uttrycka radikala åsikter och i vissa fall även berättat om sina planer innan sina attacker. Om dessa digitala spår skulle upptäckas och tas på allvar i tid finns alltså i varje fall en teoretisk möjlighet att förhindra terrordåd. Dessvärre är det inte lätt att upptäcka och analysera dessa digitala spår eftersom det ofta handlar om uttalanden som i sig inte är speciellt anmärkningsvärda men i kombination med annan typ av information kan ge en indikation på att de kan vara värda att analysera vidare. Digitala spår i form av uttryck för extrema åsikter eller planer på framtida terrorbrott kan liknas vid svaga signaler. Med en svag signal menar vi något som i sig själv inte betyder någonting men som genom kombination med andra signaler tillsammans kan tyda på en utveckling av något slag. Att hitta relevanta svaga signaler (som är värda att analysera vidare) är en av de mest utmanande uppgifter man kan tänka sig, eftersom de oftast kan tolkas på flera sätt och är tvetydiga.

Problemet med att spana efter potentiella ensamagerande terrorister är att det på många sätt är som att leta efter en nål i en höstack. Internet rymmer en överväldigande mängd information. Att manuellt surfa igenom alla nätforum som skulle kunna vara platser som exempelvis inbjuder till ”självradikalisering” är en omöjlig uppgift. En intressant fråga är därför om det är möjligt att med hjälp av automatiserade tekniker hitta nätforum där extrema åsikter utbyts, och om man genom att analysera en persons inlägg kan misstänka att denne planerar att begå våldshandlingar.

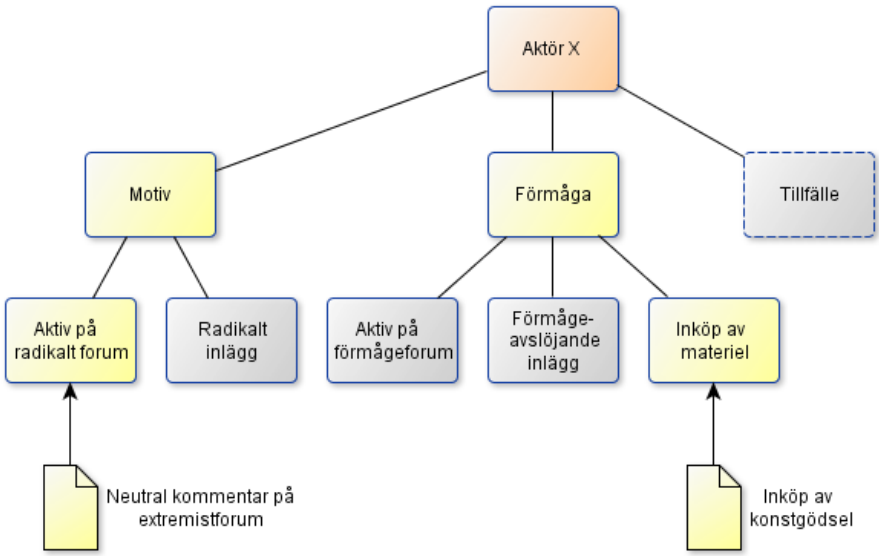
Som vi har sett i tidigare kapitel finns det en mängd olika tekniker som kan användas för att automatiskt analysera digitala spår på internet. Det är dock inte trivialt att använda dessa för att upptäcka potentiella hot. Förutom att teknikerna inte är helt tillförlitliga är också mängden tillgänglig information för stor för att det ska gå att analysera allt. Istället måste man välja ut delar av internet som ska analyseras, genom att till exempel begränsa inhämtningen till inlägg från vissa

utvalda forum som klassats som intressanta. Även efter en sådan begränsning kommer informationsmängden att analysera att vara överväldigande och en analytiker kommer att behöva en strukturerad metod och datorstöd för att klara uppgiften.

I det här kapitlet beskriver vi hur man kan använda sig av olika tekniker för att identifiera alias som kan vara av intresse för vidare undersökning och analys.

6.1 Att lösa ett komplext problem

För att kunna analysera svaga signaler behövs ett flertal komponenter. Först av allt behöver man en lämplig analysmodell. En klassisk metod för att lösa komplexa och svåra problem är att bryta ner problemet i mindre beståndsdelar, där varje del i sig är lättare att förstå och hitta en lösning till. Den här typen av nedbrytning lämpar sig väl för att analysera svaga signaler. Ett exempel på hur man kan använda sig av nedbrytning finns i Figur 6. I figuren är det grundläggande problemet huruvida aktör X planerar att begå ett terrorbrott. Problemet bryts ner i beståndsdelarna motiv (intention), förmåga och tillfälle. Alla dessa beståndsdelar måste vara uppfyllda för att det ska anses föreligga risk för att aktör X planerar ett terrorbrott. Nästa nivå i nedbrytningen exemplifierar ett antal indikatorer. En indikator är en så kallad observerbar faktor, det vill säga en händelse, företeelse, beteende eller dylikt, och kan jämföras med en svag signal. I exemplet i Figur 6 finns det en uppsättning indikatorer som kan upptäckas genom spaning på internet, samt en indikator ”Inköp av materiel” som även skulle kunna upptäckas via andra informationskanaler. I exemplet i figuren har två var för sig harmlösa signaler upptäckts, men som vid sammanvägning kan ge anledning till fördjupad undersökning.



Figur 6: Exempel på nedbrytning av ett komplext problem.

Den här typen av modeller ger ett bra stöd för att hantera komplexa problem samt för att strukturera och värdera svaga signaler. Om modellen hanteras i ett datorbaserat verktyg där information automatiskt kan sorteras och kopplas till olika individer kan man öka antalet potentiella misstänkta som kan hållas under uppsikt. Ett annat skäl till att använda ett datorbaserat stöd är att man i det specifika fallet med en aktör som planerar ett terrorbrott behöver analysera och hämta in information under långa tidsperioder, vilket ställer krav på att snabbt kunna uppdatera en tidigare analys när ny information inkommer.

6.2 Konsten att leta på rätt ställe

Den snabba utvecklingen av Internet och sociala medier har bidragit till att mängden med tillgänglig öppen information är stor. För att ha möjlighet att analysera data från öppna källor krävs både tekniker och metoder som är anpassade för att hitta rätt saker i ett hav av tillgänglig information. Ett sätt är att använda sig av en sökmotor. Sökmotorer används för att upptäcka, hämta in och indexera webbsidor. När man indexerar webbsidor skapar man ett omfattande index över alla ord som upptäckts på de sidor som inhämtats. En sökning utförs sedan i detta index eftersom det skulle ta för lång tid att söka i realtid.

Stora delar av webben är inte indexerade av sökmotorer. Detta beror delvis på att det finns webbsidor som inte går att hitta (eftersom det inte finns några länkar till dem) men också på innehåll som inte går att indexera, till exempel information

från databaser eller webbsidor med dynamiskt innehåll. Som webbplatsägare kan man förhindra sökmotorer och webbspindlar från att indexera webbsidor genom att till exempel använda sig av lösenord på webbsidan eller att använda sig av filen robots.txt (som finns på varje webbplats) för att förbjuda webbspindlar att komma till sidan. Viktigt att notera är dock att inte alla webbspindlar följer de instruktioner som finns i filen.

Det finns en hel del metoder som kan användas i syfte att bli mer osynlig på webben. De flesta metoderna bygger på att man använder sig av en så kallad proxy-server för att göra det svårare att binda en IP-adress till specifik kommunikation. TOR (The Onion Router) är en anonymitetstjänst som används i hela världen. TOR fungerar genom att frivilliga användare sätter upp servrar som fungerar som noder i nätverket. Noderna vidarebefordrar all trafik på ett slumpmässigt sätt samtidigt som den krypteras vid varje nod i nätverket.

Toppdomänen .onion är reserverad för TOR-nätverket och genom att använda sig av TOR kan man få access till en mängd olika webbplatser inom .onion-domänen. Eftersom syftet med TOR är att skapa anonymitet för webbplatsägare (och användare) finns det en mängd olika sorters webbplatser med varierande innehåll, ett exempel är Silk Road som uppmärksammats på grund av det liberala innehållet.[55] Hidden wiki är en annan webbplats under .onion-domänen. Hidden wiki innehåller länkar till andra webbplatser inom domänen. Eftersom .onion-domänen inte är sökbar för sökmotorer och adresserna inte är kopplade till en DNS server[56] är Hidden wiki och sidor på ”öppna” nätet som innehåller adresser till sidor i .onion domänen[57] nödvändiga för att man ska kunna hitta det man söker efter. En adress till en .onion-sida är en sträng bestående av 16 tecken (siffror mellan 2 och 7 samt bokstäver) som genereras genom en kryptografisk metod som kallas för hashning. Ett exempel på hur en adress kan se ut i .onion-domänen är: **ofrmtr2fphxkqgz3.onion.**

Som kuriosa kan nämnas att man för att kunna köpa saker i .onion-nätverket måste använda sig av bitcoins som är en virtuell valuta. Bitcoins kan man exempelvis köpa på MT.GOX[58].

Istället för att bara använda sig av en sökmotor (och således missa de delar av webben som inte är indexerade) för att leta efter information kan man göra en kartläggning av de delar av webben som anses vara intressanta och använda sig

---

[55] Erfarenheter kring hur det går till att köpa droger från Silk Road finns beskrivet i: Chen, A., *The underground Website Where You Can Buy Any Drug Imaginable* http://gawker.com/5805928/the-underground-website-where-you-can-buy-any-drug-imaginable, läst 2012-09-17.

[56] En datormaskin som översätter domännamn till IP-adresser och vice versa

[57] Exempelvis http://onion.is-found.org

[58] https://mtgox.com/

av en riktad inhämtning. Detta ökar chansen att man letar på rätt ställe. Tanken är då att man utgår från en mängd identifierade webbsidor och sedan använder sig av en webbspindel för att undersöka vilka sidor som är länkade från dessa sidor. Genom att använda sig av textanalys och identifierade nyckelord (det vill säga ord som man anser att sidor med potentiellt intressant innehåll ska innehålla) kan man automatiskt skapa ett nätverk av potentiellt intressanta webbplatser. Inhämtningen av information kan sedan riktas till dessa webbplatser.

6.3 Analys av inhämtad data

När man har hämtat in den information som man tror kan vara av intresse är det dags att börja analysarbetet. I det här fallet plockas alla alias ut från den inhämtade datamängden. Varje alias kopplas till en modell liknande den i Figur 6. Målet med analysen är att skapa en lista av de alias som förekommer i den inhämtade datamängden. Aliasen i listan rangordnas inbördes efter hur stor risk det är att det ska begå terrorbrott enligt modellen. Man gör också aliasmatchning för att identifiera olika alias som används av en och samma person.

För att kunna ranka aliasen analyseras texterna de har skrivit med hjälp av olika former av textanalys. De indikatorer som man kan leta efter på internet är framför allt indikatorer som tyder på att ett alias har utryckt motiv för att begå ett terrorbrott.

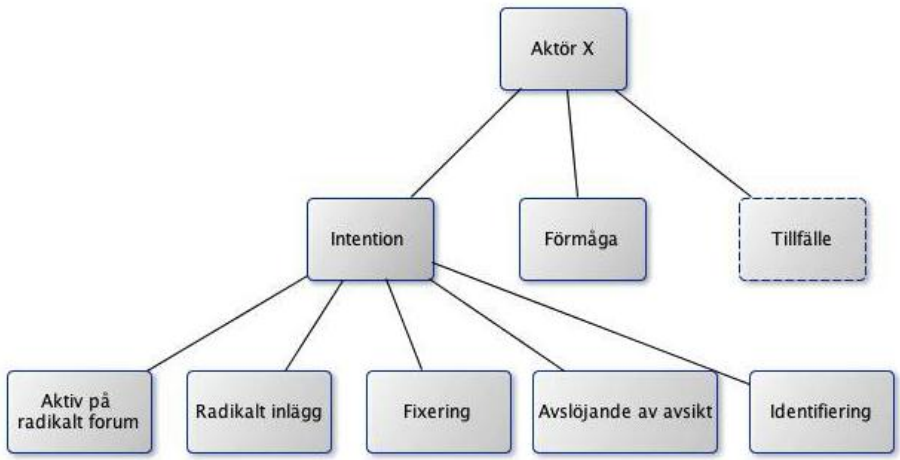
Om man fokuserar på beståndsdelen motiv i analysmodellen så kan de indikatorer som tyder på att någon har för avsikt att begå ett terrorbrott vara att personen är aktiv på en mängd extremistiska diskussionsforum. Man kan också tänka sig att en indikator för intention kan vara att personen använder sig av våldbejakande innehåll i sina inlägg. Förutom aktivitet och våldsbejakande innehåll i texter finns det vissa varningsbeteenden som man kan se ofta föregår ett terrordåd. De varningsbeteenden som är mest troliga att kunna upptäckas genom kommunikation på internet är:[59]

1. *Avslöjande av avsikt*, då gärningsmannen i förväg, mer eller mindre specifikt och mer eller mindre avsiktligt, informerar en tredje part om sin förestående handling.
2. *Fixering*, som uttrycker en extrem upptagenhet med en person eller en sakfråga.

[59] Mer information om varningsbeteenden finns i: Cohen, K., Johansson, F.,Kaati, L., Mork, J. C., *Detecting Linguistic Markers for Radical Violence in Social Media*. Accepted for publication in Terrorism and Political Violence 2013.

3. *Identifiering*, som innefattar en självbild präglad av krigar- och hjältefantasier, och/eller ett starkt intresse för vapen eller militära medel och strategier, liksom identifikation med andra radikala tänkare.

Figur 7 visar ett exempel på hur en modell med indikatorer för beståndsdelen intention skulle kunna se ut. Varje identifierat alias har en sådan modell och genom att använda sig av den modellen kan man leta efter evidens för de indikatorer som identifierats. När man har hittat evidens för någon av indikatorerna noteras detta för uppnå spårbarhet.



Figur 7 Exempel på en modell med indikatorer för intention

6.4 Tekniker för att hitta digitala spår

För att hitta avgöra om en aktör har intention kan man använda sig av olika former av textanalys. Evidens för att ett alias är aktivt på ett radikalt diskussionsforum är att personen har gjort inlägg på radikala forum (oavsett innehållet i inläggen). Ett radikalt inlägg kan identifieras genom att använda sentimentanalys och klassificera radikalitet hos text. Viktigt att notera är att den typen av klassning är domänberoende (det vill säga klassningen ser med stor sannolikhet olika ut beroende på vilka miljöer man är intresserad av) och att det är relativt obeprövat hur väl den fungerar i praktiken.

De olika varningsbeteenden som man har kunnat se ofta föregår ett terrordåd kan också identifieras genom olika former av textanalys.

Evidens för att någon gör ett så kallat avslöjande av avsikt kan upptäckas genom att dessa uttalanden vanligtvis består av ord som signalerar en avsikt (”ska”, ”vill”, etc.) i kombination med ord som beskriver en våldshandling av något slag. Eftersom det finns ett stort antal synonymer för de ord som beskriver

våldshandlingar kan man använda sig av ontologier som beskriver relationer mellan synonymer.

Evidens för indikatorn fixering kan observeras genom att en person, grupp eller ett syfte nämns i en hög utsträckning relativt andra texter som skrivits i sammanhanget. Man kan också tänka sig att vissa nyckelord som förekommer med hög frekvens kan tyda på fixering vid en idé. Fixering kan alltså upptäckas genom att man tittar på den relativa frekvensen av förekomster av ord som beskriver en person, en grupp, ett syfte eller en idé.

Evidens för indikatorn identifiering kan upptäckas genom att en viss terminologi används. Om identifieringen är med en krigare är det vanligt att en militärterminologi används, medan om identifieringen är med en radikal tänkare kan en terminologi som liknar tänkarens (ofta i kombination med direkta citat) användas.

6.5 Sammanvägning av indikatorer

När evidens inhämtas görs en sammanvägning av de indikatorer som har upptäckts. Sammanvägningen kan göras med ett flertal olika beräkningar som exempelvis Dempster-Shafer, Bayesianska metoder eller olika typer av medelvärdesberäkningar. Ett exempel på hur en sammanvägning skulle kunna se ut är att bestämma att motivnoden i modellen i Figur 7 är uppfylld om evidens för åtminstone tre av de fem indikatorerna har hittats.

6.6 Multipla alias

Varje alias har en rankning som beskriver hur intressant det är att göra ytterligare analyser av aliaset. Men som tidigare nämnts är det vanligt att en användare har ett flertal olika alias och därför är det viktigt att göra en bedömning om det finns risk för att samma användare har flera olika alias. Detta görs genom att använda sig av de tekniker för aliasmatchning som finns beskrivna i Kapitel 5. Om man misstänker att samma användare har två eller fler alias fusioneras deras analysmodeller och en ny sammanvägning utförs.

6.7 Diskussion

Att upptäcka en potentiell ensamagerande terrorist är som tidigare nämnts lika svårt som att leta efter en nål i en höstack. Men eftersom internet är en plats där våldsbejakande extremister kan mötas, inspireras och uttrycka sina åsikter, är möjligheterna att man kan hitta digitala spår som kan tyda på att någon planerar en terrorattack något som bör undersöka vidare. Datoriserade metoder kan hjälpa analytiker i deras arbete genom att ranka alias och på så sätt ge analytikerna

vägledning i vad som ska prioriteras och undersökas vidare. I det här kapitlet har vi resonerat kring olika tekniker och metoder som skulle kunna användas i jakten på ensamagerande terrorister.

Innan man kan uttala sig om hur mycket den här typen av tekniker underlättar en analytikers arbete och hur väl det fungerar i praktiken återstår en hel del arbete. Även om de olika teknikerna i sig fungerar för andra domäner behövs det mer forskning för att de ska kunna anpassas till just den här domänen.[60]

Det finns en mängd integritetsaspekter som aktualiseras när man spanar efter potentiella ensamagerande terrorister eftersom spaningen inte kan ske på grundval av misstanke om brott. I nästa kapitel beskrivs integritetskyddande tekniker som kan användas för att skydda den personliga integriteten när man analyserar data från exempelvis internet.

[60] Utförligare beskrivning över tekniker som kan användas i jakten på ensamagerande terrorister finns i: Brynielsson, J., Horndahl, A., Johansson, F., Kaati, L., Mårtenson,C., Svenson, P., *Analysis of weak signals for detecting lone wolf terrorists*. In Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2012

7 Integritetsskyddande teknik för effektivare spaning

Resultatet av ett analysuppdrag hänger ibland på att få tillgång till konfidentiella data som finns hos andra myndigheter, organisationer eller företag. Webbspaning kan peka mot en misstänkt men för att få avgörande bevis krävs ibland att integritetskänsliga register öppnas för analytikern. Etik, lagar och affärsintressen kan då lägga hinder. Detta kapitel beskriver ny teknik som kan göra det möjligt att i denna situation får fram de kritiska uppgifter som analysen hänger på och samtidigt respektera dataägarens behov av sekretess. Sökning i konfidentiella register kan med rätt teknik bli en integrerad del av webbspaning. Här är vårt fokus på de nya tekniska möjligheterna men det är viktigt att framhålla att användningen av dessa verktyg i varje enskilt fall måste styras av de lagar, policys och etiska aspekter som gäller de inblandade länderna, organisationerna och kulturerna.

I detta kapitel ska vi använda två fiktiva exempel för att på ett enkelt sätt förklara tekniken. Scenen är polisarbete men exemplen gör inga anspråk på att spegla hur polisutredningar bedrivs i verkligheten.

Rånaren: En äldre kvinna blir rånad i en gångtunnel nära Rotebro station. Förbipasserande observerar en ung man i röd munkjacka av ett känt märke som springer från platsen. På ett webbforum hittar polisen bilder från en gårdsfest som pågick i närheten av brottsplatsen samma kväll. Tre av bilderna som är tagna sju minuter efter rånet visar en förbipasserande yngling i röd munkjacka. Ynglingen identifieras med hjälp av bilderna och grips. Märket på jackan är samma som vittnena observerade. Den misstänkte nekar och vittnena och offret kan inte identifiera den gripna som gärningsmannen. Analytikern behöver nu värdera styrkan av det enda indicium som finns: munkjackan. Håller det i domstol? Hur många unga män i Rotebro gick klädda i denna typ av munkjacka under rånkvällen? Om det lokala modet föreskriver just den typen av jacka så är sammanträffandet inte så mycket värt. Men det kan också vara så att nästan ingen i Rotebro klär sig på samma sätt som rånaren. Försäljningsstatistiken hos det stora internationella postorderföretag som är de enda som säljer klädmärket skulle kunna svara på den frågan.

Mordbrännaren: En serie mordbränder har drabbat vårdcentraler och sjukhus i västra Sverige. Genom webbspaning har polisen hittat en pseudonym ”Gollum” som skulle kunna vara mordbrännaren. I ett forum på den mörka delen av internet avslöjar Gollum att sjukvården har smittat honom med HIV men inte vill ge medicin trots att han har sökt för det flera gånger. Han berättar även om att han varit tvungen att söka vård för en brännskada trots risken för avsiktlig felbehandling. Analytikern skulle nu vilja söka i sjukvårdens register efter en

person som bor i Västsverige, har sökt för HIV utan att smittan har bekräftats och dessutom har vårdats för en brännskada de senaste tre månaderna.

I båda dessa exempel är det viktigt för utredningen att komplettera webbspaning med sökning i konfidentiella databaser. Och i båda fallen är det inte självklart att utredaren får tillgång till dessa.

Det finns tre uppenbara lösningar på problemet att få tillgång till konfidentiella data.

- Dataägaren kan vara villig att öppna sin databas för analytikern direkt eller under förutsättning att det skrivs avtal som reglerar hur data får användas.
- Analytikern skulle kunna beskriva vilken sökning som ska göras och be dataägaren att utföra sökningen och lämna ut resultatet. Detta är ibland också en bra lösning men det förutsätter att analytikern kan lita på att dataägaren genomför sökningen på ett kompetent sätt och ibland också att lagar och regler tillåter detta arbetssätt. Det kan vara problematiskt att analytikern måste tala om för dataägaren vad man letar efter. I det andra exemplet ovan kan till exempel dataägaren misstänka att personen som man spanar på är den ökända mordbrännaren.
- Dataägaren kan vara villig att öppna sin databas för en mellanhand som analytikern också har förtroende för. Det kan ofta vara svårt att hitta en sådan mellanhand men ibland kan detta vara en bra lösning.

Om ingen av det tre uppenbara lösningarna är genomförbar finns det nya tekniska metoder som kan föra utredningen framåt. Metoderna brukar med ett samlingsnamn kallas *integritetsbevarande informationsutvinning* (på engelska *Privacy-Preserving Data Mining*). Ofta används den engelska förkortningen *PPDM*. Det finns två huvudtyper av metoder 1) *modifierande anonymisering* och 2) *krypterande anonymisering.* Vi ska i tur och ordning diskutera båda.

7.1 Modifierande anonymisering

Om dataägaren inte kan öppna sin databas för analytikern så kanske det går att anonymisera uppgifterna och låta analytikern jobba med det modifierade materialet. Anonymisering innebär att databasen ändras så att det inte längre går att identifiera enskilda personer. En enkel men ofta otillräcklig metod för anonymisering är avidentifiering där explicita identifierare som personnummer och personnamn tas bort. Problemet med detta är att det ofta ändå går att identifiera personer med hjälp av sökningar i bakgrundsdata. Om personnummer tas bort från vårdjournalen men det finns uppgifter om att patienten är ”man, 57 år, född i Ölme, gift med tre barn och nu boende i Bollstanäs” så kan en sökning på internet med stor säkerhet identifiera patienten. En samling av attribut som på

detta sätt med hjälp av offentliga data kan användas för att identifiera en person kallas för kvasi-identifierare. I många sammanhang finns det oftast kvasi-identifierare kvar efter avidentifiering.

Modifierande anonymisering syftar till att ta bort både explicita identifierare och kvasi-identifierare så att personlig integritet respekteras samtidigt som de modifierade data blir så användbara som möjligt för utredningen. De två huvudmetoderna för detta kallas för *k-anonymisering* respektive *slumpredigering* (*k-anonymisation* och *randomization* på engelska).

K-anonymisering går ut på att redigera databasen så att samma kvasi-identifierare delas av minst *k* personer där *k* är ett heltal. Vi kan till exempel säga att minst sex personer måste dela på samma kvasi-identifierare (*k*=6). I exemplet med patientjournalen finner en vårdmyndighet att bara en person är ”man, 57 år, född i Ölme, gift med tre barn och nu boende i Bollstanäs”. Om man tar bort uppgiften om födelseorten så blir det dock 11 personer som är ”man, 57 år, gift med tre barn och boende i Bollstanäs”. Genom att på detta sätt systematiskt redigera databasen så att varje kombination av egenskaper delas av minst sex personer får man fram en modifierad databas som är tillräckligt integritetsbevarande för att öppnas för analytikern. Det är inte självklart att den modifierade databasen är användbar för analytikern men tillgång till k-anonymiserade data kan ibland vara bättre än ingen tillgång alls. Vid k-anonymisering är det viktigt att bara ta bort eller generalisera uppgifter så att den publicerade informationen inte är falsk. Generalisering innebär att detaljerade attribut ersätts med mer generella men fortfarande sanna attribut. Födelseorten i exemplet kan till exempel generaliseras till ”Värmland”. Att manuellt redigera en stor databas för att uppnå k-anonymisering är extremt arbetskrävande men processen går att automatisera.

Slumpredigering innebär att data ändras slumpmässigt så att det inte längre går att lita på kvasi-identifierare men så att databasen fortfarande är användbar för statistiska undersökningar. Postorderföretaget i exemplet med rånaren i munkjacka har en försäljningsdatabas som vi kan tänka på som en jättelik tabell med en rad per kund. Kolumnerna i tabellen talar om hur många exemplar kunden har köpt av varje vara som lagerförs. Med slumpredigering adderas nu ett slumptal till varje fält i tabellen. Slumptalet dras ur mängden {-3, -2, -1, 0, 1, 2, 3} med samma chans för varje tal att bli draget. Nu går det inte längre att se vem som har köpt vad. Det går dock fortfarande att beräkna summor och medelvärden med gott resultat eftersom slumptalen i genomsnitt tar ut varandra. En analytiker kan med hjälp av den slumpredigerade databasen räkna ut ungefär hur många i Rotebro som har köpt en röd munkjacka av samma typ som rånaren bar.

Slumpredigering är användbar när det är summor, medelvärden och liknande statistiska mått som är intressanta för utredningen. De statistiska måtten blir mindre exakta om data är slumpredigerade jämfört med om originaldata används. Storleken av denna försämring går dock att beräkna. En nackdel med

slumpredigering är att enskilda attribut inte längre är sanna. Om den som använder databasen inte förstår detta kan slumpmodifieringen skada den personliga integriteten.

7.2 Krypterande anonymisering

Om det är viktigt för undersökningen att inga detaljer i data ändras kan krypterande anonymisering användas. Dessa metoder bygger på att dataägare krypterar data innan dessa lämnas ut, att undersökningen utförs på krypterade data och leder till ett krypterat resultat. Endast resultatet avkrypteras. Det enda som någon part får veta är just det resultat som man kommit överens om att avslöja.

För att förklara hur en undersökning med krypterande anonymisering kan gå till återvänder vi till exemplet med mordbrännaren. Landstingets centrala HIV-mottagning kan göra ett register över personnummer på alla som flera gånger har sökt för HIV men har blivit friskförklarade. Ur vårdcentralernas databaser går det att utvinna en lista över personnummer på alla som har sökt för en brännskada under den aktuella perioden. Polisen vill veta vilka personnummer som finns på båda listorna och landstinget är villigt att lämna ut just dessa få personnummer på villkor att inget annat avslöjas. HIV-mottagningens lista får inte lämnas ut till någon och detsamma gäller vårdcentralernas lista. Inte ens inom landstingets organisation får listorna lämnas ut för samkörning.

Med krypterande anonymisering kan problemet lösas. HIV-mottagningen krypterar varje personnummer med en speciell krypteringsmetod och en krypteringsnyckel som de själva väljer. Listan skickas till vårdcentralernas organisation som krypterar varje krypterat personnummer en gång till med samma speciella krypteringsmetod och sitt eget val av krypteringsnyckel. Listan på dubbelkrypterade personnummer för HIV-sökanden skickas till polisens utredare.

Vårdcentralerna gör nu om processen med sin egen lista på personnummer för brännskadepatienter. Listan krypteras med den egna kryptonyckeln och skickas till HIV-mottagningen. Där krypteras listan en gång till med HIV-mottagningens kryptonyckel och de dubbelkrypterade personnumren för brännskadepatienter skickas till polisen.

Den krypteringsmetod som användes kallas *kommutativ kryptering* och har den speciella egenskapen att det inte spelar någon roll i vilken ordning successiva krypteringsoperationer utförs. Om ett personnummer först krypteras med nyckel A och resultatet krypteras en gång till med nyckel B blir det dubbelkrypterade slutresultatet likadant som om krypteringsoperationerna hade utförts i omvänd ordning.

Polisens utredare jämför de båda listorna på dubbelkrypterade personnummer och hittar de poster som är identiska i båda listorna. I detta fall hade man tur och hittade precis en post som var gemensam för de båda dubbelkrypterade listorna. Eftersom både HIV-mottagningen och vårdcentralerna använde kommutativ kryptering motsvarar denna poster precis det enda personnummer som förekommer i båda listorna. Utredaren skickar nu det dubbelkrypterade personnumret för avkryptering först vid HIV-mottagningen och sedan hos vårdcentralernas organisation. När det avkrypterade resultatet kommer tillbaka till polisen har utredaren gärningsmannens personnummer i klartext. Ingen av deltagarna i processen har tagit del av någon integritetskränkande information utom just det avsedda slutresultatet. Med något mer komplicerade operationer går det också att helt och hållet maskera vilken information som polisen letar efter.

Det finns en mängd metoder för krypterande anonymisering som hanterar de flesta vanliga informationsutvinningsproblem. Data kan vara fördelade på olika sätt bland många olika deltagare och den information som ska utvinnas kan vara av olika typ. Gemensamt för alla dessa metoder är att man kommer överens om vilken information som ska avslöjas och att de kryptografiska metoderna ser till att inget annat än det avsedda slutresultatet avslöjas för någon av deltagarna. En nackdel med krypterande anonymisering är att kryptering kräver mycket stor datorkraft. I vissa speciella fall finns det risk att information läcker om en eller flera deltagare aktivt bryter mot överenskomna processer. Om samma deltagare gång på gång använder krypterande anonymisering går det också att utvinna känslig information genom en försåtligt arrangerad serie av frågor. Trots dessa fallgropar kan dock krypterande anonymisering vara mycket användbart, speciellt vid samarbete mellan myndigheter som litar på varandras processer men som av juridiska skäl inte får lämna ut data.

7.3 Diskussion

Integritetsbevarande informationsutvinning ger nya möjligheter att komplettera webbspaning med sökning i konfidentiella register. Detta är speciellt användbart när regelverk eller etik hindrar en direkt sökning i data eller då analytikern inte vill berätta vad man letar efter för dataägaren.

Om detaljerna i enskilda data inte är viktiga kan metoder som bygger på modifierande anonymisering användas. K-anonymisering och slumpredigering är de viktigaste metoderna för modifierande anonymisering. K-anonymisering kan dölja och generalisera enskilda data men ger en sann och så exakt bild av individuella attribut som integritetsskyddet medger. Statistiska medelvärden förvrängs dock av k-anonymisering. Slumpredigering är bra för att beräkna statistiska medelvärden med bevarad integritet men förvränger och förfalskar enskilda attribut. Om det finns mycket bakgrundsinformation i öppna källor kan

det vara svårt att tillämpa modifierande anonymisering utan att dölja och förvanska data så mycket att det inte går att utvinna någon intressant information. Orsaken till detta problem är att mycket öppen bakgrundsinformation om individerna gör att många olika kombinationer av fält i databasen bildar kvasi-identifierare som måste raderas med anonymisering.

Om analytikern söker svar på ett fåtal frågor där svaren beror på detaljerna i sekretessbelagda data kan krypterande anonymisering vara en bra lösning. Metoden kräver kompetenta utövare och stora beräkningsresurser.

Ett exempel på möjlig användning inom FM är automatisering av behandling av frågor från befattningshavare med lägre behörighet till befattningshavare med högre behörighet. Det kan exempelvis krävas hög behörighet för tillgång till ett register över misstänkta terrorister. Analytiker med lägre behörighet får då fråga en person med högre behörighet om specifika misstänkta personer finns i registret. Att sköta detta manuellt kan vara ett arbetskrävande rutinarbete speciellt om det ställs många frågor men antalet träffar är få. Integritetsbevarande informationsutvinning kan på ett säkert sätt automatisera processen så att befattningshavaren med högre behörighet bara behöver bevaka loggarna för systemet men inte själva svara på alla frågor. Det datorsystem som används för att fråga har bara kontakt med en krypterad version av registret. Frågan krypteras också vilket gör att namn på misstänkta hanteras säkert. På samma sätt kan frågor mellan organisationer, såväl inom som utom landet, som samarbetar men inte vill öppna sina register hanteras. Det är värt att notera att det inte bara är personlig integritet som kan skyddas på detta sätt, utan även andra känsliga uppgifter som inte rör personer kan skyddas enligt samma principer.

Integritetsbevarande informationsutvinning är ett ungt forskningsfält. Metoderna har begränsningar och en del säkerhetsluckor. Det saknas även lättanvända verktyg från pålitliga leverantörer. För praktisk användning av metoderna krävs därför just nu stöd från metodexperter.

8 Trovärdighet och information från webben

Den stora mängden tillgänglig information på webben gör att det finns ett ökat behov att automatisera trovärdighetsbedömningar av tillgänglig information. De mått som finns för att mäta trovärdighet på information kan grovt sett delas in i tre grupper[61]: innehållsbaserade, kontextbaserade och värderingsbaserade mått. Innehållsbaserade mått använder sig av själva informationen för att avgöra trovärdighet. Kontextbaserade mått använder sig av kontextrelaterad information för att avgöra (framför allt en källas) trovärdighet. Värderingsbaserade mått använder sig av andras bedömningar och värderingar av informationen för att avgöra trovärdigheten.

8.1 Algoritmer

Det finns flera funktioner och algoritmer som kan användas (och används!) för att bedöma trovärdighet hos information.[62,63,64] Nedan följer en kortfattad beskrivning av några av de tekniker som kan användas stöd vid trovärdighetsbedömningar.

Poängräkningsfunktioner

Enkla poängräkningsfunktioner kan man stöta på i flera sammanhang där ett enkelt betyg eftersöks. En del fungerar så att de är summan av de positiva och negativa betygen eller helt enkelt medelvärdet av givna betyg. Ett problem med den här typen av betyg är att de inte tar hänsyn till orättvisa eller subjektiva omdömen.

Kollaborativ och anseendebaserad filtrering

I både kollaborativ och anseendebaserad filtrering accepteras endast betyg från medlemmarna i en grupp. I kollaborativ filtrering anser man att personer som har gett liknande betyg antagligen har ungefär samma smak. Denna information används sedan i så kallade rekommendationssystem där saker som en person

[61] Anpassat från diskussion om kvalitetsmätning av webb baserat material i Bizer, C., *Quality-Driven Information Filtering in the Context of Web-Based Information Systems*. PhD thesis, Freie Universität Berlin, March 2007.

[62] Mer information om detta finns bland annat i: Bizer, C., Quality-Driven Information Filtering in the Context of Web-Based Information Systems. PhD thesis, Freie Universität Berlin, March 2007.

[63] Jøsang, A., Ismail, R., and Boyd, C., A survey of trust and reputation systems for online service provision. Decis. Support Syst., 43(2):618–644, March 2007.

[64] Ziegler, C-N. and Lausen, G., Propagation models for trust and distrust in social networks. Information Systems Frontiers, 7(4-5):337–358, December 2005.

gillade sedan rekommenderas till de med liknande smak, till exempel "De som gillade den här [filmen, boken, saken] gillade även de här".

I anseendebaserad filtrering accepterar man att personer i gruppen har olika smak, men försöker använda de betyg som inte anses bero på smak. Exempelvis kanske personer kommer att betygsätta filmer olika beroende på sin smak, men kvalitén på filmen (till exempel dåligt ljud eller bild) borde ha som effekt att alla (majoriteten) i gruppen reagerar på och ger ett enhetligt betyg.

Förtroende- och inflytandemodeller

Ett sätt att hantera orättvisa och subjektiva bedömningar är att man endast låter de personer man i någon mån litar på att ge omdömen. Det finns flera algoritmer och system som fokuserar på förtroende och inflytande.

TidalTrust[65] är en algoritm som syftar till att undersöka hur mycket man kan lita på någon som man själv inte känner i ett socialt nätverk. De noder/personer man är kopplad till räknas som ens grannar. För de grannar man litar på anger man ett värde på hur mycket man litar på dem. Om man själv inte har angett ett förtroendevärde för en nod räknar algoritmen ut ett värde. Detta förtroendevärde baseras på vilka förtroendevärden den nodens grannar har angett, på dessa grannars förtroendevärden, samt på hur många steg man själv befinner sig från noden i fråga. Algoritmen skulle kunna utvidgas med stöd för osäkra nätverk enligt vad som tidigare beskrivits.

SUNNY[66] är en vidareutveckling av TidalTrust i två avseenden. För det första används ett mer avgränsat förtroendenätverk där ett urval av grannarnas förtroendevärden görs. För det andra kompletteras förtroendevärdet med ett konfidensintervall som anger hur mycket man bör lita på förtroendevärdeberäkningen.

WIQA[67] är en förkortning av "Web Information Quality Assessment Framework" och är ett projekt som försöker hjälpa en användare att filtrera information baserat på förtroende-policyer, det vill säga användaren anger till exempel att endast information godkänd av någon myndighet skall presenteras.

Klout[68] är ett företag som försöker mäta folks inflytande i sociala nätverk. De räknar ut inflytande genom olika mätdata från Twitter, Facebook och Google+

---

[65] Golbeck, J., *Computing and applying trust in web-based social networks*. PhD thesis, College Park, MD, USA, 2005. AAI3178583.

[66] Kuter, U. and Goldbeck, J., *Sunny: A new algorithm for trust inference in social networks using probabilistic confidence models*. In Proceedings of the National Conference on Artificial Intelligence (AAAI,) 2007.

[67] http://www4.wiwiss.fu-berlin.de/bizer/wiqa/, läst 2012-10-14

[68] http://klout.com, läst 2012-10-20

såsom antal anhängare, ”retweets”, hur många ”döda” konton som är följare, hur inflytelserika de människor som retweetar är, unika omnämnanden, kommentarer, “likes”, och antal vänner i ens nätverk. Allt detta viktas sedan ihop i en så kallad "Klout Score". Mycket kritik har dock riktats mot dem sedan inflytelserika personer (som exempelvis presidenter) har fått lägre poäng än bloggare.

Ett annat sätt att hantera subjektivitet och orättvisa bedömningar är att använda sig av indirekta förtroendemodeller. Det mest välkända exemplet på detta är Googles PageRank[69]-algoritm (se även Kapitel 4.1). I PageRank mäts inflytande rekursivt i form av poäng baserade på antalet inkommande länkar från andra webbsidor. Inkommande länkar räknas som positiva poäng som blir större ju fler poäng de sidor som länkar själva har.

Sökmotoroptimering

Då det ofta är så att användare endast förlitar sig på de tio främsta träffarna vid en sökning finns det många företag som fokuserar på att optimera sin information så att de ska hamna högt upp i sökmotorernas listor. Vid informationssökning bör man vara uppmärksam på detta, då värdefull information kan försvinna i mängden. Det finns ett flertal mer eller mindre avancerade sätt som kan används vid sökmotoroptimering. Dessa metoder brukar benämnas med “white hat” eller “black hat” beroende på om de är accepterade/rekommenderade metoder för att sökmotoroptimera eller inte. Vissa av metoderna bygger på att utnyttja att det är webbspindlar som automatiskt söker i genom webbsidornas innehåll och inte en människa, som i det flesta fall är svårare att lura. Några av metoderna är:

- nyckelordsanvändning – man lägger till alla möjliga nyckelord som kan ha någon relevans för ämnet på webbsidan. Detta kan göras båda på själva sidan och i metataggarna.
- korslänkning – man kan dels låta webbsidor på den egna domänen länka till huvudsidan och andra undersidor för att verka mer populär, dels skapa särskilda domäner och webbsidor (så kallade *linking farms*) endast i syfte att länka vidare till de sidor man vill exponera.
- man kan använda dold eller gömd text på en sida för att öka sidans “relevans”.
- “article spinning”[70] – är en teknik för att skapa mer uppmärksamhet kring ett ämne och lura folk att tro att något diskuteras mer än vad det

[69] Page, L., Brin, S., Motwani, R. and Winograd, T., *The PageRank citation ranking: bringing order to the web*. Technical Report, Stanford University1998.

[70] http://en.wikipedia.org/wiki/Article_spinning, läst 2012-10-24

görs. Man producerar med små variationer och omskrivningar kopior på artiklar, bloggar och foruminlägg.

8.2 Diskussion

Att automatiskt bedöma trovärdighet av information i öppna källor är inte trivialt. Utvecklingen av web 2.0 med en ökad mängd användargenererad information accentuerar detta ytterligare. Här finns stor potential för att hitta användbar underrättelseinformation, men risken för felaktiga slutsatser och vilseledning ökar.

Dessa svårigheter kan angripas på flera sätt. Utöver utbildning för att ytterligare medvetandegöra psykologiska fördomar som påverkar oss vid värdering av information,[71] finns möjligheten att utveckla tekniskt stöd som kan bistå analytiker vid trovärdighetsbedömningar. Ett sådant stöd bör på grund av problemets natur endast ge vägledning; det kommer fortfarande att vara upp till analytikern att fatta beslut om den slutgiltiga bedömningen.

För underrättelseanalys kan man tänka sig flera spår för tekniskt stöd vid trovärdighetsbedömningar. Ett sätt är att följa WIQA:s ansats och filtrera ut auktoriserad information och markera denna på något särskilt sätt för analytikern. Detta fungerar dock enbart för ett fåtal källor som utsätts för extern bedömning. Nästa steg kan vara att anamma en kollaborativ eller anseendebaserad filtreringsmetod, där information eller källor kan rekommenderas av kollegor. En förfining blir sen att använda förtroendemodeller i stil med SUNNY, där probabilistiska metoder används för att få en mer exakt uppskattning av källans trovärdighet.

En mer utmanande ansats för automatiskt stöd är att ge sig i kast med de innehållsbaserade måtten, i syfte att ge analytikern indikationer på informationens sakriktighet. För detta krävs att man automatiskt kan extrahera fakta från texter (se Kapitel 3) och sedan jämföra med fakta som man tror sig veta är sann.

[71] I denna rapport har inte de psykologiska aspekterna som styr oss vid informationsutvärdering och analys nämnts. En bra referens för information om detta är Heuer, R. J., *Psychology of intelligence analysis*. US Government Printing Office. 1999

9 Slutord

Att internet har tagit en allt större plats i våra liv och att våra liv i allt större utsträckning tar plats på internet är ett faktum. Detta gör att internet i många sammanhang blir en allt viktigare källa för information. Men att analysera data från internet är komplicerat. Ett av problemen är att hitta i den stora mängden av tillgänglig information. En anledning som gör det svårt att hitta det som kan vara intressant är att stora delar av internet inte är indexerat och således inte kan genomsökas av en sökmotor. Det är också svårt att hantera så stora mängder information som internet tillhandahåller vilket ställer stora krav på de algoritmer som används för analys av data. Ett annat problem är att den information som finns tillgänglig ofta består av korta texter som är skrivna på ett speciellt sätt med förkortningar, slang, ironi, felstavningar och länkar vilket försvårar analysen.

Vi har i denna rapport presenterat ett antal olika typer av tekniker för att analysera webbdata. Översikten som gjorts är långt ifrån fullständig utan är snarare tänkt att tjäna som en introduktion till några av de problemområden där FOI bedrivit forskning inom ramen för FoT-projektet Verktyg för informationshantering och analys (VIA). För merparten av det som presenterats i rapporten finns möjlighet att göra mer tekniska djupdykningar genom att följa de referenser som ges i texten.