OSINT

AVOINTEN LÄHTEIDEN INTERNET-TIEDUSTELU

PETRI NURMI

Aaalto EE TJK 14
Aalto-yliopisto
Turvallisuusjohtamisen koulutusohjelma
Syksy 2015 - Kevät 2017

Sisällysluettelo

1. JOHDANTO

1.1 Alkusanat

Muistan hämärästi TV 1 –kanavalla vuosituhannen vaihteessa esitetyn Noutajat.net –tietokilpailun, missä kilpailukysymyksiin etsittiin vastaukset Internetistä. Kilpailija nojasi lepotuolissaan ja ratkaisi sellaisia tehtäviä kuten: ”*Mikä on presidentin kanslian puhelinnumero?*” Kilpailijat olivat voittopuolisesti nuorisoa, olihan kyseessä ennen kaikkea nuorisolle suunnattu ohjelma. Katselin kilpailua 24-vuotiaan humanististen alojen opiskelijan silmin hieman huvittuneena. Koko touhu ei vaikuttanut kovinkaan kiinnostavalta. Internet ei vaikuttanut kovin kiinnostavalta. Sen sijaan ohjelman juontaja Mari Kakko vaikutti hyvinkin kiinnostavalta.

Ohjelma oli jäänyt jollakin tavalla kuitenkin elämään muistiini, sillä se juolahti mieleeni tätä kirjoittamaan ryhtyessäni. Googlasin ohjelman ja löysin Ilta-Sanomien artikkelin, missä Mari Kakko (os. Mari Sainio) muisteli ohjelmaa. Altavistan lisäksi kilpailijat hakivat kilpailussa tietoa muun muassa Ask Jeevesistä (nykyisin Ask.com) ja Ihmemaasta (www.fi). ”Ensin piti valita hakukoneista yksi ja sitten lähteä etsimään vastausta. Siihen aikaan tietokoneet olivat hitaita ja tiimalasi saattoi vain pyöriä ruudulla menemättä mihinkään”, Sainio muistelee. ”En tajunnut ollenkaan, miten tämä internet toimii. Luulin, että se on tavallaan kuin puhelinluettelo.”[1]

Itse asiassa Sainio ei kaukanakaan verkon perusolemuksesta: sitä voi tosiaankin verrata valtavaan kirjaan, jonka kattavasta hakemistosta käy ilmi missä mikin tieto sijaitsee. Toki internet on paljon muutakin, olisi helpompi luetella mitä se ei ole. Ensinnäkään internet ei ole kenenkään omaisuutta, vaan se on pikemmin käsite kuin jotakin konkreettista. Internet ei ole myöskään polarisoitunut, vaan se on hajautettu ympäri maailmaa. Jokainen Internetin tietokone on riippumaton. Operaattorit voivat valita, mitä Internet-palveluja ne käyttävät. Internet on verkostojen verkosto tai verkostoitumisen infrastruktuuri. Se yhdistää maailmanlaajuisesti miljoonia tietokoneita yhteen, jotka muodostavat verkoston. Verkostossa mikä tahansa tietokone voi kommunikoida minkä tahansa tietokoneen kanssa, kunhan ne vain ovat molemmat yhteydessä Internetiin.

Samalla kun päivittelee Twitter-tiliä tai pelaa online-pelejä internetissä, on syytä muistuttaa itseään, että internetin rahoittajana ja kätilönä toimi alkujaan Yhdysvaltain asevoimien tutkimusorganisaatio Defence Advanced Research Projects Agency (DARPA). Internetin alkuperäisenä tarkoituksena oli toimia osana sotilaallista jatkuvuussuunnittelua ydinasehyökkäyksen sattuessa. Internet on siis Yhdysvaltojen puolustusvoimien luomus, jonka syntymisellä oli merkittävä korrelaatio geopoliittisiin voimasuhteisiin.

World Wide Web, tai yksinkertaisesti Web, ei ole suinkaan sama asia kuin internet, vaan *tapa porautua internetiin*. World Wide Web on tiedon jakamisen malli, joka on rakennettu

1 http://www.iltasanomat.fi/viihde/art-1288596309269.html

varsinaisen Internetin päälle. Neljännesvuosisata sitten englantilainen tietojenkäsittelytieteilijä Tim Berners-Lee antoi keksintönsä World Wide Webin vapaasti kaikkien käyttöön. Berners-Lee kehitti World Wide Webiä Euroopan hiukkastutkimuskeskuksessa Cernissä Sveitsissä Ensimmäinen verkkosivu julkaistiin 6. elokuuta 1991, mutta aluksi se oli vain Cernin tiedemiesten luettavissa. Sivulla kerrotaan lyhyesti mikä World Wide Web on. Elokuussa vuonna 1991 World Wide Web tehtiin kaikille avoimeksi ja sen kunniaksi vietetään nykyään internautti-päivää, eli internetin käyttäjien päivää. Sana tulee internetin ja astronautin yhdistelmästä.

Web käyttää HTTP-protokollaa, joka tosin on vain yksi Internetissä dataa välittävistä kielistä. HTTP on tuttu URL-osoitteista. Tosinaan osoitekentän URL-osoite alkaakin kirjaimilla HTTPS, mikä taas tarkoittaa kryptattua viestiä. Mikäli kuuntelisit pariisilaisen kahvilan pöydässä viereisen pöydän puhetta ranskankielen ummikkona, niin tällöin kyseessä olisi HTTPS. HTTPS on perustavaa laatua oleva tietoturvakerrostuma, jota käyttävät sellaiset verkkosivut kuten PayPal, Facebook tai Gmail. Googlen teettämän tutkimuksen mukaan verkossa on yhä valtava määrä tunnettuja ja paljon käytettäviä sivustoja, jotka eivät käytä HTTPS:aa kuten New York Times, IMDB tai Wired.[2]

Web hyödyntää meille tuttuja selaimia kuten Internet Explorer, Chrome tai Mozilla Firefox, auttaakseen meitä pääsemään käsiksi Web-asiakirjoihin, joita kutsutaan verkkosivuiksi. Web on vain yksi niistä tavoista, joilla tietoa voidaan levittää Internetissä. Internetiä, ei Webiä, käytetään myös sähköpostiliikenteessä, Usenet-uutisryhmissä, pikaviesteissä ja FTP:ssä (File Transfer Protocol). Web on siis vain osa internetiä. Nämä kaksi termiä eivät kuitenkaan ole synonyymeja toisilleen.

Internetin runko muodostuu palvelintietokoneista eli servereistä. Palvelin on teknisesti PC:tä muistuttava tietokone, jossa on kiintolevy ja käyttömuisti, minkä avulla ohjelmat suoritetaan. Palvelimet sisältävät internetissä esitetyn tiedon ja kaikki käyttäjälle tarjotut palvelut, joten palvelimen edellytyksenä on paljon suurempi muistikapasiteetti kuin tavallisella tietokoneella. Palvelinkoneet on kytketty toisiinsa nopean tietoliikenneyhteyden mahdollistavilla kaapeleilla. Palvelimet ovat usein osa jotakin lähiverkkoa, joka koostuu esim. oppilaitoksen tai yrityksen lähiverkoksi liitetyistä PC-koneista.

Kun käyttäjä muodostaa yhteyden Internetiin esim. sähköposti- tai selainohjelman avulla, ottaa ohjelma yhteyden palvelinkoneessa olevaan palvelinohjelmistoon. Palvelinohjelmisto lähettää palvelupyynnön kun haet tai avaat jonkun internet-sivun ja vastaanottaa palvelimen palauttaman tiedon sekä tulostaa sen käyttäjälle, eli näytölle avautuu haettu internet-sivu. Yhteyden muodostaminen Internetiin tarkoittaa siis yhteyden muodostamista palvelimelle, josta voidaan edelleen muodostaa yhteys toisiin palvelimiin.

2 https://www.wired.com/2016/03/https-adoption-google-report/?utm_content=buffer46336

HTML-kielellä[3]tehdyt asiakirjat eli tuntemamme internet-sivut sijaitsevat palvelimissa, jotka voivat fyysisesti sijaita missä tahansa. Sivut voivat sisältää linkkejä, jotka ohjaavat käyttäjän toisille sivustoille. Täten linkkejä seuraamalla voidaan Internetissä liikkua palvelimelta toiselle eli "surffata". Internet toimii asiakas/palvelin- periaatteella, jolla tarkoitetaan sitä, että yhteys asiakkaan ja palvelimen välille muodostetaan vain tiedonsiirron ajaksi. Tiedonsiirron loputtua yhteys katkaistaan. Jatkuvaa yhteyttä ei siis ole.

Älypuhelimien seurauksena internetistä tuli erottamaton osa arkeamme. Verkko-osoitteesta http://www.internetlivestats.com/ voi konkreettisesti todeta silmitöntä vauhtia kelaavien laskureiden kautta päivittäiset Internetin käyttäjät, blogi-päivitykset, Google-haut, Youtube-videokatselut yms. Päivittäisten Internetin käyttäjien määrä liikkuu jossakin 3,259,035,200 tietämillä, joka edustaa 40 prosenttia koko maailman väestöstä. Myös uusien verkkosivujen syntymisestä löytyy oma laskurinsa, lukumäärä lähestyy miljardia. Useat meistä ovat kantaneet kortensa kekoon rakentamalla verkkosivut vaikkapa paikalliselle avantouimarikerholle tai vuokrattavalle loma-asunnolle helppokäyttöisillä Weeblyn tai Wixin kaltaisilla suosituilla verkkosivujen rakennusohjelmilla. Interaktiivisten verkkosivujen rakentaminen oli vielä muutama vuosi sitten IT-maallikolle sula mahdottomuus, sillä se edellytti vähintäänkin edistyneitä koodaustaitoja. Nykyään rakentaminen käy käden käänteessä jopa harjaantumattomalta.

Koko internetin mikrokosmos on tänä päivänä jatkuvasti käsiemme ulottuvilla ja me osaamme koko ajan hyödyntää sitä sujuvammin. Vaikka verkon suomat mahdollisuudet tulevat meille jatkuvasti tutummaksi, monet meistä kokevat sen olevan yhä suuri arvoitus. Me tiedämme verkon sisältävän fantastisia määriä tietoa ja me tiedämme sen olevan luonteeltaan kuin ovi, jonka takaa paljastuu uusi ovi, jonka takaa paljastuu uusi ovi ja niin edelleen. Tunnettu analogia on, että Yahoon ja Googlen kaltaiset selaimet ovat kuin valtameressä raahautuvia nuottia, joihin takertuu kyllä kaikenlaista hyödyllistä, mutta internetissä vallitsee myös pinnanalaisia informaation ekosysteemejä. Syvänetin eli Deep Webin on arvioitu olevan 500 kertaa suurempi kuin perinteisillä hakukoneilla tavoitettavat 550 miljoonaa verkkosivua. Olisiko siis sittenkin järkevämpää käyttää nuotan sijaan kalastuspaikasta ja –tilanteesta riippuen sopivaa viehettä?

3 HTML (Hypertext Markup Language).´Hypertekstin merkintäkieli´ on avoimesti standartoitu merkintäkieli , jolla voidaan kuvata hyperlinkkejä sisältävää tekstiä . HTML:llä voidaan myös merkitä tekstin rakenne eli esimerkiksi, mikä osa tekstistä on otsikkoa ja mikä leipätekstiä. HTML tunnetaan erityisesti kielenä, jolla internet-sivut on kirjoitettu.

1.2 Tutkimuskysymykset

- **Tutkimuskysymys 1**: Mitä on avointen lähteiden (internet)tiedonhankinta (OSINT)?
- Alakysymys 1: Millainen on avointen lähteiden tiedonhankinnan tiedonkäsittely- ja analysointiprosessi?
- **Tutkimuskysymys 2**: Millainen informaatio-teknologinen ympäristö internet on infosec-näkökulmasta?
- Alakysymys 1: Millaisia tiedonhankintatekniikoita ja freeware-ohjelmistoja on mahdollista hyödyntää avointen lähteiden tiedonhankinnassa?

1.3 Tutkimuksen tavoitteet ja rajaukset

Avoimet lähteet ovat käsitteenä tiedonhankinnan kontekstissa hyvin laaja. Se käsittää lukuisia eri medioita ja julkaisuja. Avoimiksi lähteiksi voidaan lukea:

- Radio
- Televisio
- Uutistoimistot
- Tieteelliset julkaisut
- Ajankohtaisohjelmat
- Kuvat
- Tietokannat
- Kartat
- hallitusten julkiset tiedotteet
- Kirjastot
- Kirjallisuus
- Yritykset
- Yliopistolliset julkiset raportit
- Ns. Harmaa aineisto[4]

Tärkein rajaus koskee edellä kuvattua laaja-alaista aineistoa; kehitysprojektissa keskitytään yksinomaan tietokantoihin ja verkossa tapahtuvaan tiedonhankintaan. Tavoitteena on johdatella avointen lähteiden tiedonhankinnan menetelmiin ja tekniikoihin internetissä, kuten kehittyneisiin hakukoneisiin ja selaimiin, tehokäyttömenetelmiin, verkkosovelluksiin sekä OSINT-työkaluihin kuten Maltegoon, Focaan, Creepyyn, ExifTooliin, sosiaaliseen verkkoanalyysiin ja erilaisiin kuvahaun sovelluksiin. Monia erinomaisia työkaluja jää tyystin esittelemättä, jotka ansaitsivat myös perehtymistä kuten SearchDiggity, Reconing, TOR, i2p, Shodan, Reconing,

[4] Harmaa aineisto on erittäin merkittävä avointen lähteiden tyyppi. Kyseessä on tietolähde, jota ei ole julkaistu virallisesti, mutta jotka eivät myöskään ole salaisia.

Metagoofil, MAT ja monet muut. Ainoa puolustukseni tähän on, että käsillä oleva tutkielma on pituudeltaan rajallinen.

Kehitysprojekti sisältää myös kevyen johdannon joihinkin teknisiin aiheisiin kuten meta- ja referenssitietojen louhintaan, mutta esimerkiksi Python-ohjelmointi, jolla on mahdollista luoda kokonaan omia tiedonhakuoperaattoreita tai muokata jo olemassa olevia ohjelmointirajapintoja (API)[5], sivuutamme tyystin ja jätämme aiheen edistyneemmille verkkodekkareille.

Toinen keskeinen rajaus koskee julkisten tietolähteiden ja OSINT-työkalujen tavoitettavuutta ja maksullisuutta. Kaikkien lähteiden tulee siis olla kaikille avoimia ja ilmaisia. Tämä merkitsee useiden erittäin käyttökelpoisten, mutta maksullisten lähteiden rajaamista pois tutkielman käytettävissä olevasta lähdemateriaalista kuten

- Ajoneuvohallintakeskus, josta saa tietoja auton rekisterinumeron perusteella
- PHR, josta saa tiedot yritysten tilinpäätöksistä
- Kauppalehden ePortti, joka tarjoaa kattavasti tietoja yrityksistä ja henkilöistä
- Väestörekisterin Internet-osoitepalvelu, josta löytyy kattavammin tietoa kuin Eniron henkilöhausta

Kuten tiedetään, markkinoilla on lukuisia ällistyttävän tehokkaita verkostoanalyysi-työkaluja, joista osa on myös tiedusteluviranomaisten käytössä. Tässä tutkielmassa keskitymme kuitenkin vain joko kokonaan ilmaisiin tai sovellusten ilmaisversioihin. TOR-verkosta on saatavilla sinänsä ilmaisia, mutta selkeästi jo rikollisten tarpeisiin räätälöityjä hakkerointityökaluja; ne eivät myöskään kuulu tutkielmalle relevanttiin materiaaliin. Tämä ns. *Darknet* vaatisi laajuutensa ja kompleksisuutensa vuoksi kokonaan oman tutkimuksensa. TOR-verkossa puhutaankin OSINT:in sijaan useammin *doxingista*, vaikka ero saattaa olla joskus hiuksenhieno. Doxing-termi tulee englanninkielen sanoista ´documents´ ja ´docs´ ja tarkoittaa henkilötietojen louhimista internetistä erilaisista sähköisistä dokumenteista. Wikipedia-artikkelin mukaan doxingilla on negatiivinen sointi[6], koska sitä on käytetty esimerkiksi erilaisiin kostotoimenpiteisiin.

Internetissä tapahtuva systenmaattinen tiedonhankinta herättää myös joukon eettisiä kysymyksiä kuten: ”Saako netissä vierailla kaikilla sivuilla syyllistymättä rikokseen?” tai ”Saako ladata itselleen tekstejä, kuvia tai videoita ellei niitä jaa eteenpäin?” Nämä ovat valideja

[5] API on rajapinta, joka määrittelee ohjelmistokomponenttien vuorovaikutuksen muodot viitaten operaatioihin, syötteisiin, tulosteisiin ja tietotyyppeihin. Toisin sanoen API on määritelty ja dokumentoitu sopimus siitä, miten ohjelmistokomponentit kommunikoivat keskenään ja käskevät toisiaan.

6 https://en.wikipedia.org/wiki/Doxing

kysymyksiä. Perussääntö kumminkin on, että pelkkä katseleminen tai lukeminen ei ole laitonta. Vastuu hurjastakin aineistosta on julkaisijalle, ei katsojalla tai lukijalla. Vaikka esimerkiksi verkon vuotosivuilla on jaossa valtioiden salaisiksi julistamia asiakirjoja, ei niiden lukeminen ole kiellettyä: laitonta on kopion valmistaminen ja levittäminen. Tässä suhteessa *WikiLeaks* rikkoo yksiselitteisesti lakia whistelblower-tarkoitusperistä huolimatta.[7] Sama koskee tekijänoikeuden alaista materiaalia. Eli tällä puheella voit vapaasti lehteillä verkosta helposti saatavaa yli tuhatsivuista järkälettä *Encyclopedia of Jihadia*, jossa on sellaisia lukuja kuin ”How to Kill”, ”Explosive Devices”, ”Manufacturing Detonators” ja ”Assassination with Mines” ilman että tuntisit olosi syylliseksi.

Tässä kohden onkin hyvä tehdä rajanveto kahden kokonaan eri asian välille. *Verkkovakoiluun* kuuluu oleellisesti sitä varten suunniteltu teknologia, joka mahdollistaa puhelujen kuuntelemisen, näppäimistön käytön tallentamisen reaaliaikaisesti, paikantaa mobiilikäyttäjä ja lukea kaikki mitä kohde tekee älypuhelimella. Verkkovakoilun mahdollistavaa teknologiaa kutsutaan sanalla *hyökkäysteknologia*. OSINT taas on taiteenlaji. Sitä on kutsuttu myös eettiseksi hakkeroinniksi eli hakkeroinniksi missä ei rikota lakeja tai käytetä hyväksi psykologista tai fyysisistä manipulaatioita tiedon hankkimiseen. OSINT edellyttää suorittajaltaan hyvää verkkoympäristön tuntemusta, kattavaa keinovalikoimaa ja ongelmanratkaisutaitoa.

Toinen eettinen kysymys koskee myös kerätyn tiedon käsittelyä. Kerätessämme henkilöistä dataa, on muistettava että yksityisten henkilörekisterien luominen on laitonta. Kielto ei ole kuitenkaan täysin ongelmaton, sillä henkilötietolaki[8] on ajalta ennen Googlea ja nykyään kaikki verkkoon kirjoitettu on käytännössä henkilörekisterissä. Laki tulisi arvioida kokonaan uudelleen Big Datan ja internetin aikakautena. Helpompi on siisi lähestyä kohdetta projektiluonteisesti, tällöin laittoman henkilörekisterin luominen on vähemmän epätodennäköisempää.

Eräs henkilökohtaisista tavoitteistani kehitysprojektin liittyen on ollut selkokielinen ja helppolukuinen esitystapa, jolla pyrin välittämään keskeisen tiedon sellaisellekin lukijalle, joka ei ole perehtynyt tekniseen termistöön tai internetin infrastruktuuriin. Pyrin avaamaan maallikolle vieraampia käsitteitä sitä mukaan, kun niitä tulee eteen. Itse olen myös ICT-maallikko, joten tutkimustyö on opiskelua itsellenikin. Tunnettu sanonta IT-kirjallisuuden alalla kuuluukin, että jokainen yhtälö puolittaa kirjan myyntilukuja.

7 https://fi.wikipedia.org/wiki/WikiLeaks

8 http://www.finlex.fi/fi/laki/ajantasa/1999/19990523

2. AVOINTEN LÄHTEIDEN INTERNET-TIEDUSTELU

2.1 Mitä on OSINT?

”Informaation ei tarvitse olla salaista ollakseen arvokasta.” CIA:n verkkosivujen lause kiteyttää avointen lähteiden tiedustelun pähkinänkuoreen: OSINT sitoo avoimista lähteistä saatavan hajallaan olevan tiedon systemaattisen analysointiprosessin kautta tiiviiksi ja informatiiviseksi asiakokonaisuudeksi.[9] Internet ja etenkin sosiaalisen median kanavat ovat mullistaneet viime vuosina OSINT:n ja lisänneet analysoitavan tiedon määrää merkittävästi. Sosiaalisen median verkostoituva perusluonne on johtanut modernissa OSINT:ssä erilaisiin graafisiin esitystapoihin ja visualisointeihin sekä asioiden yhteyksien yhä parempaan tunnistamiseen.

Ajallisessa kontekstissa OSINT voidaan jäljittää II Maailmansotaan ja yhdysvaltalaiseen tiedusteluinstanssiin nimeltä *Foreign Broadcast Monitoring Service* (FBMS), joka nimensä mukaisesti seurasi ulkomaista uutisvirtaa.[10] Kuitenkin vasta 9/11 terrori-iskut ja internetin arkipäiväistyminen nostivat avointen lähteiden tiedustelumetodologian marginaalista valtavirtaan. Terrori-iskujen jälkimainingeissa perustettiin *Open Source Center* (OSC) – tiedustelukeskus, joka korvasi edeltäjänsä FBMS:n. Keskus perustettiin samoihin tiloihin edeltäjänsä kanssa Virginian Restoniin. Open Source Centerin tehtäväksi tuli kääntää, suodattaa ja analysoida uutisia ympäri maailmaa Yhdysvaltojen turvallisuusviranomaisten tarpeisiin terrorismin vastaisessa työssä. Epäilemättä avointen lähteiden tiedustelu on siirtynyt sitten keskuksen perustamisen vuosi vuodelta koskemaan ennen kaikkea verkkoympäristöä, vaikka keskus yhä mainitseekin lähteikseen radion, television, paperimedian, sekä muut perinteiset mediat.

Angloamerikkalaisessa tiedusteluterminologiassa tiedonkeräysmenetelmistä puhutaan eri –INT:teina (lyhenne sanasta ’intelligence’). Nämä –INT:it taas on jaettu perinteisesti neljään kategoriaan, jotka ovat: henkilötiedustelu, signaalitiedustelu, kuvaustiedustelu ja avointen lähteiden tiedustelu.

Henkilötiedustelussa (Human Intelligence, HUMINT) tiedon lähteenä ja kerääjänä toimii ihminen. Tieto itsessään voi olla suullinen, kirjallinen tai sähköinen. Henkilötiedustelun piiriin kuuluvat he, jotka puhuvat ohi suunsa väärälle taholle ymmärtämättään tai päätyvät asiansa osaavan tiedustelijan jututtamiksi. Kyseessä on tietysti tiedustelun tunnetuin ja kenties vaativin muoto

9 https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html

10 https://www.afio.com/publications/Schauer_Storger_Evo_of_OSINT_WINTERSPRING2013.pdf

sen inhimillisen luonteen vuoksi. Inhimillisten tietolähteiden käyttäminen on pitkä prosessi, missä ensin pyritään löytämään sopivat kohteet, kontaktoidaan ja rekrytoidaan tietolähteiksi.

Signaalitiedustelussa (Signals Intelligence, SIGINT) tietoa kerätään sähkömagneettisista signaaleista esim. radioliikenne, sähköinen viestintä ja tutkasignaalit. Signaalitiedustelu on - jos mahdollista - henkilötiedusteluakin julkisuudelta varjellumpaa toimintaa. Signaalitiedustelun rinnalle voidaan nostaa Measurement and Signatures Intelligence, MASINT). SIGINT tiedustelee signaaleja kokonaisuutena ja MASINT niiden pieniä eroja, jolloin voidaan tunnistaa laiteyksilöitä. Signaalitiedustelusta puhutaan yleensä ennen kaikkea sotilastiedustelun yhteydessä. Kun käsitellään vaikkapa terrorismin uhan torjuntaa, niin kyse on monessa tapauksessa nimenomaan signaalitiedustelusta. Esimerkiksi Ruotsissa signaalitiedustelua hoitaa FRA (Försvarets radioanstalt), jolla on käytössään supertietokone. Suomessa Tikkakoskella sijaitseva Viestikoelaitos (VKoeL) on puolustusvoimien signaalitiedustelua harjoittava organisaatio.

Kuvaustiedustelulla (Imagery Intelligence, IMINT) tarkoitetaan tiedustelua, joka tapahtuu joko ilmasta, satelliitista tai tutkalla. Kuvaustiedustelu on tiedustelulaji, jonka merkitys on viime vuosina kasvanut. Ilmakuvauksen merkitys on korostunut tekniikan tuodessa yhä parempia ja tarkempia kuvauslaitteita. Merkille pantavaa on, että sotilaskäyttöön valjastettujen satelliittien rinnalle ovat tulleet kaupalliset satelliitit, jotka tuottavat sota-alueilta kuvakavalkadia internetiin kenen tahansa katseltaviksi.

Avointen lähteiden tiedustelussa (Open Source Intelligence, OSINT) kerätään tietoa julkisista tai muutoin laillisesti saatavista lähteistä. Avoimien lähteiden tiedustelulla on eri merkityksiä riippuen tarkastelutavasta: CIA:lle sillä tarkoitetaan tietoa, joka on saatavissa ulkomaisista uutislähteistä, asianajajalle se saattaa tarkoittaa informaation keräämistä julkisista dokumenteista. Useimmille ihmisille se kuitenkin tarkoittaa tietoa, joka on haettavissa verkosta hakukoneiden avulla. Tällä tavoin muotoiltuna se ei kuullosta kovinkaan ihmeelliseltä. Sen sijaan kaikki se tieto, mitä OSINT-tekniikoilla on ammattilaisten toimesta on mahdollista saada, on ihmeellistä.

Mennäksemme vielä asteen verran syvemmälle avointen lähteiden tiedonhankinnan määritelmiin, on lainattava käsikirjaa vuonna 2001 julkaistua *Nato Open Source Intelligence Handbookia*.[11] Sen mukaan on olemassa neljä erityistä avointen lähteiden informaation ja tiedustelun alakategoriaa:

Open Source Data (OSD) on raakadataa. Se on puhutun sanan, lähetyksen, printin tai muussa muodossa olevaa ensikäden informaatiota. Se voi olla kirje, valokuva, nauhoitus tai kaupallisen satelliitin ilmakuva. Hyvä esimerkki tästä on sotareportteri, joka on tehnyt näkemästään muistiinpanoja kynällä, ottanut valokuvia ja nauhoittanut sotilaiden kanssa käytyjä keskusteluja. Ennen kuin materiaali on tavalla tai toisella käsiteltyä tai suodatettua, se on raakadataa.

11 http://www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf

Open Source Information (OSI) koostuu raakadatasta, jota on mahdollista yhdistellä ja muokata erillisen editointiprosessin kautta, johon kuuluu osana myös datan jakamisen hallinta. OSI on geneeristä informaatiota, jota saatetaan levittää hyvinkin laajalti. Sanomalehtiartikkelit, kirjat ja uutislähetykset koostuvat tällaisesta käsitellystä datasta, johon törmäämme päivittäin.

Open Source Intelligence (OSINT) on informaatiota, joka on tarkoituksella haettu, se on jäsennelty, suodatettu ja välitetty valikoidulle yleisölle. Yleensä se vastaa johonkin kysymykseen tai erityiseen tarpeeseen esim. jäljittää jokin sosiaalisen median päivitys tiettyyn IP-osoitteeseen tai maantieteelliseen sijaintiin. OSINT on siis laajaan avointen lähteiden tarjontaan kohdentuva tiedonkeruun prosessi, joka luo tiedustelutietoa.

Validated OSINT (OSINT-V) on nimensä mukaisesti varmennettua avointen lähteiden tiedustelutietoa, jonka validoinnin on suorittanut useimmiten salassa pidettäviin rekistereihin ja tietojärjestelmiin pääsyn omaava analyytikko. Määritelmä kytkeytyy siis vahvasti valtioiden turvallisuusviranomaisiin, kuten poliisiin tai puolustusvoimiin. OSINT-V voi olla myös livekuvaa jostakin tapahtumasta tai tilanteesta, jonka aitoudesta ei ole epäilystäkään. Pääsääntöisesti OSINT-V vaatii kuitenkin tiedon todenperäisyyttä varmentavaa, salassa pidettävää rekisteritietoa, joka tukee avoimista lähteistä louhittua informaatiota

Internet on luonteeltaan kuin elävä organismi: verkkosivustoja suljetaan, dataa voidaan siirtää, pilkkoa tai piilottaa. Tämän vuoksi omien löytöjen välitön dokumentointi on erityisen tärkeätä, eikä suinkaan pidä luottaa siihen että informaatio löytyy verkosta ensi viikollakin. Hakukoneiden tietokannat eivät ole koskaan ajan tasalla, vaan ne käytännössä tarjoavat aina hieman vanhentuneita hakutuloksia. Tämän takia esimerkiksi tuoreimpien uutisten hakuun hakukoneet eivät sovellu, vaan tällöin kannattaa käyttää jonkin uutispalvelun omaa sisäistä hakua – tai esimerkiksi uutisten hakemiseen erikoistuneita hakukoneita (esimerkiksi Google News).

2.2 UNDOC: Tiedustelusykli

Avoimiin lähteisiin perustuva tiedustelusykli vastaa eri vaiheiltaan perinteistä tiedustelusykliä. Keskeisin ero on tiedustelutiedon ja analyysien lähteenä käytetyn materiaalin julkaisu- ja keräystavassa, jotka perustuvat avoimiin lähteisiin perustuvassa tiedustelussa julkisiin ja vapaasti kerättäviin tietoihin. Avoimiin lähteisiin perustuva tiedustelu pyrkii tarjoamaan samanaikaisesti lopputuotteen käyttäjälle vapaan pääsyn analyysissä käytettyyn lähdemateriaaliin sekä analyytikon lähdemateriaalista tuottamaan analyyttiseen lopputuotokseen. Prosessin yhteydessä puhutaan neljästä D:stä, jotka ovat Discovery, Discrimination, Distillation ja Dissemination. Tarkoituksena on selvittää tutkimuksen kannalta keskeiset lähteet, erotella prosessin tavoitteiden kannalta keskeiset teemat ja löytää näiden teemojen kannalta keskeiset asiat sekä keskeiset toimijat.[12]

Ensimmäisenä avoimiin lähteisiin perustuvan tiedusteluprosessin vaiheena on toimeksiantajan antama tutkittavan ongelman ja tähän liittyvien tiedustelutarpeiden määrittely. Toimeksiantokuvauksen tarkoituksena on ohjata analyytikon työtä muissa prosessin vaiheissa. Koska toimeksiantokuvaus toimii keskeisenä tietotarpeen määrittelijänä ja tätä kautta analyytikon työ tä ohjaavana tekijänä, tulee toimeksiantokuvauksen määrittely tehdä yksityiskohtaisesti ja tarkasti. Analyytikon tulee tietää lopputuotteeseen liittyvät odotukset ja tietotarpeet, jotta lopputuotteen on mahdollista vastata toimeksiantajan tietotarpeeseen. Toimeksiantajan olisi suositeltavaa myös ohjata, osallistua ja antaa palautetta tiedusteluprosessista ja sen tuloksista prosessin eri vaiheissa.[13]

Toisena vaiheena avoimiin lähteisiin perustuvassa tiedusteluprosessissa on tiedustelutiedon kerääminen. Tämän vaiheen tarkoituksena on löytää keskeisiä tietotarpeeseen liittyviä lähteitä, joiden avulla voidaan luoda toimeksiantajan tietotarpeeseen vastaava lopputuote. Tiedon onnistuneen keräämisen kannalta on keskeistä, että tietotarve on ensin muunnettu tiedusteluvaatimuksiksi, jotka toimivat tietotarpeeseen vastaavana toimenpidelistana. Tätä listaa hyödynnetään tiedonkeräyksen strategian luomisessa, tavoitteisiin soveltuvien lähteiden valitsemisessa ja tiedon varsinaisessa keräämisessä.

Tiedustelutarpeet voivat olla analyytikko- tai tapahtumavetoisia tai ne voivat perustua ennalta määritettyyn aikatauluun. Siinä missä analyytikkovetoinen tietotarve perustuu asiakkaalla oleviin yksilöityihin tarpeisiin ja vaatimuksiin, tapahtumavetoinen tietotarve puolestaan syntyy yleensä jonkun tietyn keskeisen tapahtuman seurauksena. Kolmantena vaihtoehtona on ennalta määrättyyn aikatauluun perustuva tiedustelutarve, jossa voidaan seurata jonkin tietyn tapahtuman kehitystä säännöllisin väliajoin. Koska tiedon keräämiseen käytetty aika on pois

12 Open Source Intelligence Handbook 2001, 15

13 Ibid, 16 – 17

tiedon analysointiin käytettävästä ajasta, on tiedustelutiedolta ja lopputuotteelta edellytettävä tarkkuus ja yksityiskohtaisuus määriteltävä ennen tehtävän aloittamista.[14]

Tiedon prosessoinnin ja hyödyntämisen vaihe on prosessin keskeisin vaihe, jossa analyytikko käyttää arvostelukykyään ja analyysitekniikoita kerätyn tiedon merkityksen arvioimiseen. Selkeät analyysimallit eri tyyppisen ja –tasoisen tiedon analysoimisessa helpottavat toisaalta analyytikon työtä ja toisaalta tuovat esiin kerätyssä tiedossa olevia puutteita tai sen yhteyksiä muuhun tietoon. Avoimista lähteistä kerätyssä tiedossa ja sen analysoimisessa on useita ongelmia, jotka liittyvät muun muassa analyytikon ennakkoasenteisiin sekä julkaisijan ja julkaistun tiedon luotettavuuteen. Näistä syistä on keskeistä että analyytikko selvittää mahdollisimman tarkasti alkuperäisen tiedon lähteen sekä arvion tiedon luotettavuudesta kerätessään tietoa. Ilman riittävää lähteen ja kerätyn tiedon luotettavuuteen liittyvää arviota analyytikko voi käyttää ennakkoluuloista tai vääristynyttä tietoa osana lopputuotetta, joka johtaa joko osittain tai kokonaan virheelliseen lopputulokseen. Tyypillisinä tiedon arviointikokonaisuuksina käytetään tarkkuutta, luotettavuutta, ja auktoriteettia, ajankohtaisuutta, objektiivisuutta sekä relevanttiutta.

Kerätyn ja analysoidun tiedustelutiedon lopputuotteet jaetaan neljään eri ryhmään. Ensimmäisenä ryhmänä ovat raportit, jotka ovat tyypiltään yleisluotoisia tilannekuvauksia tai määritellyn tiedustelutarpeen täyttäviä kuvauksia. Raportin varsinaiselle sisällölle ei ole asetettu vaatimuksia ja se voidaan rakentaa kunkin tiedustelutarpeen mukaisesti. Raportin tulisi kuitenkin sisältää analyyttinen yhteenveto raportin lähdeaineiston tiedoista, lyhyet yhteenvedot raportin eri teemoista sekä suoria lainauksia lähteenä käytetystä materiaalista. Samoin raportista tulisi käydä ilmi sen kattama ajanjakso sekä milloin tiedustelutiedon kerääminen raporttia varten on päättynyt.[15]

Toisena keskeisenä tyyppinä on linkkitaulukko, jonka pääasiallinen tarkoitus on toimia tukiaineistona tiedustelutietoa keräävälle analyytikolle. Taulukko voidaan rakentaa tarpeiden mukaan mutta sen tulisi sisältää kuvaukset lähteen arvosta, lähteen verkko-osoite sekä lyhyt yhteenveto lähteen sisällöstä. Kolmantena tyyppinä on etäopiskeluportaali, joka sisältää organisaation toiminnan kannalta merkittäväksi luokiteltua tietoa. Portaalin tarkoituksena on toimia lähtökohtana muun tiedustelutiedon keräämiselle. Neljäntenä ja viimeisenä tyyppinä ovat joko suljetut tai avoimet keskusteluryhmät. Ryhmiin kerätään eri aloja ja alueita tuntevia asiantuntijoita, jotka osallistuvat tiedon analysoimiseen ja vastavuoroisesti saavat hyödyntää ryhmän tuottamaa materiaalia.[16]

Avoimiin lähteisiin perustuvan tiedustelun viimeisenä vaiheena on lopputuotteen jakelu. Koska lopputuotteeseen kerätty tieto perustuu avoimiin lähteisiin, sitä voidaan jaella vapaasti eri

14 Open Source Intelligence Handbook 2001, 17 – 19

15 Ibid, 29 – 30

16 Ibid, 31 – 32

tahoille lopputuotteen tilaajan harkinnan sekä tehtävänannon tarkoitusten mukaisesti. Tyypillisiä tahoja ovat yksityisen sektorin kumppaniyritykset ja eri kansalaisjärjestöt. Varsinainen jakelu voidaan suorittaa eri organisaatioille muun muassa hyödyntämällä Internetin palveluita.

2.3 Tiedon luotettavuuden arviointi

Tiedon arviointivaihe on keskeinen osa tiedusteluprosessin sykliä, sillä tehtyjen päätelmien tarkkuus riippuu olennaisesti päätelmien tukena olevan tiedon laadusta ja tarkkuudesta. Tärkeää on että tiedon laatua ja tarkkuutta arvioidaan välittömästi tietoa kerättäessä, jotta keräämisessä voidaan huomioida se minkälaisessa yhteydessä kyseinen tieto on kerätty.

Toisin kuin *NATO: Open Source Intelligence Handbook* -ohjeistuksessa, samana vuonna (2001) julkaistussa *UNODC Criminal Intelligence: Manual for Analysts*-ohjeistuksessa käydään läpi useita eri tapoja joiden avulla kerätyn tiedon luotettavuutta voidaan arvioida ja analysoida. Ohjeistuksessa käydään läpi kaksi erilaista tiedon luotettavuuden arviointimenetelmää. Läpikäydyt tiedon arviointimenetelmät ovat 4x4- ja 6x6-menetelmät, joilla arvioidaan tietolähteen luotettavuutta tähän liittyvien ominaisuuksien ja tunnuspiirteiden kautta.[17]Tämän jälkeen arvioidaan varsinaisen kerätyn tiedon luotettavuutta hyödyntämällä tiedon lähteen ja varsinaisen tiedon välistä suhdetta.

Lähteenä toimivan tahon sekä lähdeaineiston luotettavuus tulee arvioida molemmat erillisissä vaiheissa, ja arvioinnin tulee perustua ammattitaitoiseen ja perusteltuun näkemykseen eikä tiedon arvioijan henkilökohtaisten tunteiden tai mielipiteiden saa antaa vaikuttaa arvioinnin tulokseen. Arviointi tulee tehdä myös mahdollisimman lähellä alkuperäistä lähdeaineistoa.[18]

A	Lähde osoittautunut aikaisemmin luotettavaksi tai lähteen aitoudesta, eheydestä, luotettavuudesta tai pätevyydestä ei ole epäilystä
B	Lähde, jolta tieto on kerätty on osoittautunut suurimmassa osassa tapoauksista luotettavaksi.
C	Lähde, jolta tieto on kerätty, on osoittautunut suurimmassa osassa tapuksista epäluotettavaksi.
X	Lähteen luotettavuutta ei voida arvioida.

Taulukko 1: 4x4-järjestelmän lähteen luotettavuuden arviointikriteerit

17 Criminal Intelligence 2011, 25

18 Ibid, 25

Taulukko 1 on ensimmäinen neliosainen 4x4-järjestelmässä käytetty kriteeristö tietolähteen luotettavuuden arvioimiseksi. Asteikko on neliportainen, jossa A kuvaa luotettavaksi arvioitavaa tietoa ja X tietoa jonka luotettavuutta ei voida arvioida. Väliasteet B ja C kuvaavat joko sitä, että tieto on aikaisemmin osoittautunut tyypillisesti joko luotettavaksi tai epäluotettavaksi. 4x4-järjestelmässä korostetaan tietolähteestä aikaisemmin kerättyjen tietojen luotettavuutta. 1) Tiedon tarkkuudesta ei ole epäilystä. 2) Lähde tuntee tiedon, mutta tiedon välittänyt taho ei. Tieto on loogista ja yhdenmukaista muun aiheeseen liittyvän tiedon kanssa. 3) Lähde ei tunne tietoa henkilökohtaisesti, mutta muu kerätty tieto tukee kyseisen tiedon todenmukaisuutta. 4) Lähde ei tunne tietoa henkilökohtaisesti eikä tiedon oikeellisuutta pystytä varmistamaan muista lähteistä.

1	Tiedon tarkkuudesta ei ole epäilystä.
2	Lähde tuntee tiedon, mutta tiedoin välittänyt taho ei. Tieto on loogista ja yhdenmukaista muun aiheeseen liittyvän tiedon kanssa.
3	Lähde ei tunne tietoa henkilökohtaisesti, mutta muu kerätty tieto tukee kyseisen tiedon todenmukaisuutta.
4	Lähde ei tunne tietoa henkilökohtaisesti eikä tiedon oikeellisuutta pystytä varmistamaan muista lähteistä.

Taulukko 2: 4x4-järjestelmän tiedon luotettavuuden arviointikriteerit

Taulukko 2 arvio kerätyn tiedon luotettavuutta ja tarkkuutta. Vaikka kriteeristöllä arvioidaan tiedon luotettavuutta, painotetaan siinä samalla myös tietolähteen roolia tiedon välittämisessä. Asteikko on neliportainen, jossa 1 määrittää tiedon luotettavaksi ja 4 määrittää tiedon sellaiseksi ettei tietolähteen luotettavuutta voida varmistaa. UNODC:n esittämä toinen tiedon arviointijärjestelmä, 6x6-järjestelmä, jakautuu myös kahteen eri osa-alueeseen, jossa kummassakin on kuusi arviointikriteeriä.

A Täysin luotettavaa	Lähde osoittautunut täysin luotettavaksi. Ei epävarmuutta lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan.

B Yleisesti luotettavaa	Lähde osoittautunut yleisesti luotettavaksi. Hieman epävarmuutta lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan.
C Kohtalaisen luotettavaa	Lähde osoittautunut kohtalaisen luotettavaksi. Epävarmuutta lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan (kaksi tai useampi kohta).
D Usein epäluotettavaa	Lähde osoittautunut menneisyydessä usein epäluotettavaksi. Perusteellista epäluuloa lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan.
E Epäluotettavaa	Lähde osoittautunut epäluotettavaksi menneisyydessä. Varmuus lähteen aitoudessa, eheydessä, luotettavuudessa tai pätevyydessä olevista puutteista.
F	Lähteen luotettavuutta ei voida arvioida.

Taulukko 3. 6x6-järjestelmän lähteen luotettavuuden arviointikriteerit

Taulukossa 3 nähdään 6x6-järjestelmän lähteen luotettavuuden arviointiin käytettävä kriteeristö. Arviointikriteeristö noudattaa useammasta kohdasta huolimatta pääpiirteittäin 4x4-järjestelmän tietolähteen arvostelukriteerejä. Asteikko on kuusiportainen, ja siinä A-luokitellusta lähteestä kerättyä tietoa pidetään täysin luotettavana ja F-luokitellun tietolähteen luotettavuutta ei ole pystytty arvioimaan.

1 Vahvistettu todenmukaiseksi	Tieto on vahvistettu luotettavaksi itsenäisten lähteiden kautta. Tieto loogista ja yhtenäistä muun aiheesta kerätyn tiedon kanssa.
2 Todennäköisesti todenmukaista	Tieto on loogista ja yhteneväistä muun aiheeseen liittyvän tiedon kanssa, mutta sen luotettavuutta ei ole vahvistettu muista riippumattomista lähteistä.

3 Mahdollisesti todenmukaista	Tiedon todenmukaisuutta ei ole vahvistettu, mutta se on kuitenkin loogista ja ainakin osittain yhtenevää muun aiheeseen liittyvän tiedon kanssa.
4 Todenmukaisuus epävarmaa	Tieto ei ole epäloogista. Tiedon todenmukaisuutta ei ole vahvistettu eikä siihen ole uskottu tietoa vastaanotettaessa, mutta tieto voi kuitenkin pitää paikkansa.
5 Todenmukaisuus epätodennäköistä	Tieto on epäloogista tai se on ristiriidassa muun aiheeseen liittyvän vahvistetun tiedon kanssa.
6	Tiedon luotettavuutta ei voida arvioida

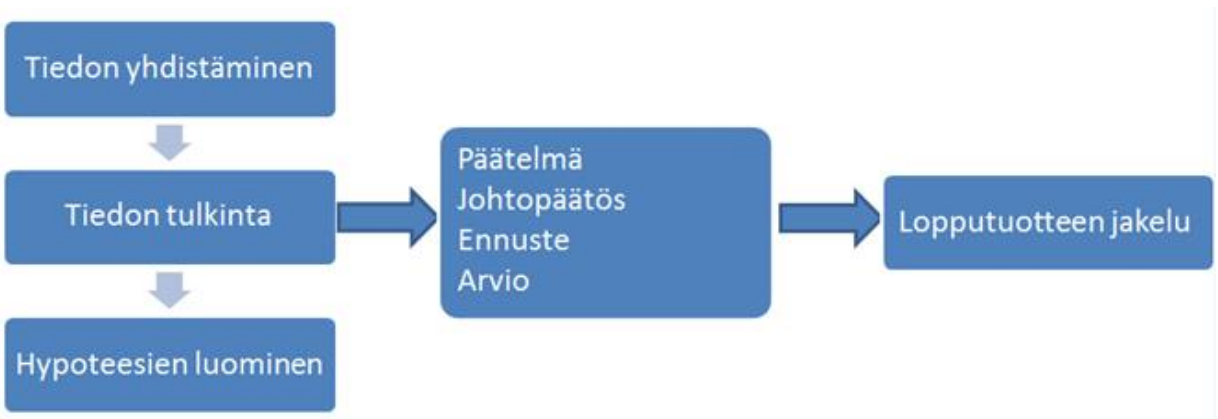
Taulukko 4: 6x6-järjestelmän tiedon luotettavuuden arviointikriteerit

Taulukossa 4 on kuvaus 6x6-järjestelmän tiedon luotettavuuden arviointikriteeristöstä. Tässä kohtaa voidaan havaita erot 4x4- ja 6x6-järjestelmien välillä. 4x4-järjestelmän tiedon luotettavuuden arviointikriteeristö korostaa 6x6-järjestelmää enemmän tietolähteen itsensä tunteman tiedon merkitystä. Kun taulukon 2 ja 4 kriteeristöjä verrataan, voidaan havaita että 6x6- järjestelmän tiedon arviointikriteeristö käsittelee tiedon luotettavuutta erillisenä asiana lähteen luotettavuudesta. 4x4-järjestelmä puolestaan korostaa tiedon luotettavuuden arviointivaiheessa tietolähteen omakohtaisesti tuntemaan tietoa, ja arvioi toisen käden tietoja kriittisemmin.

2.4 Analysointiprosessi ja dokumentointi

Tiedon analysointivaihe on lopputuotoksen onnistumisen kannalta keskeinen vaihe, sillä vasta sen avulla on mahdollista tunnistaa analysoitavassa tiedossa olevia heikkouksia sekä vahvuuksia. Tulokset ohjaavat tarvittaessa koko tiedusteluprosessin suuntaa. Prosessin ensimmäisenä vaiheena toimii kerätyn tiedon yhdisteleminen eri kokonaisuuksiin ja teemoihin sekä näiden perusteella alustavien hypoteesien rakentaminen. Koska kyseessä on syklinen prosessi, ensimmäinen vaihe toimii myös tiedon keräystoimintaa ohjaavana tekijänä, paljastaen kerätyssä tiedossa olevia puutteita ja ohjaten tiedon keräystä eteenpäin tietotarpeiden mukaisesti.[19]

Seuraavana analyysin vaiheena on kerätyn tiedon tulkitseminen. Siinä kerätylle tiedolle annetaan merkitys ja konteksti analyysin kohteessa. Tässä vaiheessa yksittäinen päivämäärä, henkilö tai tapahtuma yhdistetään osaksi isompaa analysoitavaa tapahtumaa, jonka kautta voidaan havaita tällaisen yksittäisen tiedon merkitys suuremmassa yhteydessä. Kerätyn tiedon ja tiedon merkityksen perusteella voidaan luoda hypoteeseja analysoitavasta kohteesta. Tyypillisesti hypoteesit sisältävät runsaasti spekulatiivista tietoa, joten hypoteesin luotettavuuden vahvistamiseksi näitä varten tulee kerätä sekä hypoteesia tukevaa että sen kanssa ristiriidassa olevaa aineistoa. Hypoteesia tukevan tai sen kanssa ristiriidassa olevan materiaalin perusteella voidaan erottaa toisistaan hypoteesit joita voidaan pitää todennäköisenä ja hypoteesit jotka ovat epätodennäköisiä.



Kuva 1: Analyysiprosessin vaiheet

Kuten kuvasta 1 voidaan havaita, varsinainen analyysivaihe koostuu kerätyn tiedon käsittelystä ja valmistelusta tiedon analysointia varten. Kerätyn ja käsitellyn tiedon pohjalta luodaan

19 Criminal Intelligence 2011, 29 – 30

varsinainen analyyttinen tuote, joka koostuu analyytikon käsittelemän tiedon ja hypoteesien pohjalta tekemistä päätelmistä, johtopäätöksistä, ennusteista ja arvioista.

Suoritettaessa tiedonhankintaa internetissä löydösten dokumentointi on vähintään yhtään tärkeätä kuin analysointiin tarvittavat menetelmät. Toisinaan pelkkä PrtScn-napin painaminen, URL-osoitteen tallentaminen Suosikkeihin tai leikkaustyökalun käyttäminen eivät ole riittäviä tapoja tallentaa ja arkistoida löydökset. Vai löytyykö kenties takataskustasi tapaa taltioida kokonainen verkkosivusto? Erityisesti TOR-verkossa toimiessa tulee olla tarkkaavainen sen suhteen mitä materiaalia kerää talteen, miten ja milloin. TOR-verkossa verkkosivusto saatetaan deletoida, siirtää toisaalle tai lukita yhä useampien tuunusten taakse nopeastikin. Suosittelen tallentamaan OSINT-työkalupakkiin joitakin verkkopohjaisia sivustoja, jotka auttavat taltioimaan verkossa tehtyjä löydöksiä erilaisin menetelmin. Tässä muutamia esimerkkejä:

- FreezePage (https://www.freezepage.com/)
- Screen Catpure (https://ctrlq.org/screenshots/)
- WebReaper (http://www.webreaper.net/)
- Save Web Files (https://ctrlq.org/save/)
- Printwhatyoulike (http://www.printwhatyoulike.com/)
- Youtube Grabber (http://www.videograbby.com/)
- Awesome Screenshot (https://www.awesomescreenshot.com/)
- OCR Service (https://www.onlineocr.net/

3. HAKUKONEET

3.1 Google

Yhdysvaltalainen fyysikko Larry Smarr aloitti mustien aukkojen törmäämistä koskevan tutkimuksensa 1980-luvun alussa. Näinkin monimutkaisten prosessien mallintaminen vaati kuitenkin sen aikaisilta tietokoneilta valtavan määrän laskentatehoa. Smarrista tulikin pian ensimmäisten supertietokoneiden pioneeri onnistuen lobbaamaan National Science Foundationin perustamaan ensimmäisen supertietokone-keskuksen akateemiseen ympäristöön. Hän päätyi johtamaan keskusta nimeltä National Center for Supercomputing Applications (NCSA) Illinoisin yliopistossa, minne hän perusti tietojärjestelmäkehitystiimin palvelemaan akateemisten tutkijoiden tarpeita. Tuon tiimin jäsenet Marc Andreessen ja Eric Bina suunnittelevat graafisen selaimen nimeltä Mosaic, josta myöhemmin kehittyi Netscape, ensimmäinen maailmanlaajuista tunnettuutta saavuttanut internet-selain. Tällä tavalla kosmologinen tutkimus johti kiehtovaa väylää pitkin ensimmäisten alkeellisten internet-selainten syntyyn.[20]

Kun ajattelemme hakukonetta, synonyymi sille tänä päivänä on Google, josta on muodostunut verbi *googlata*, joka tarkoittaa tiedonhakua verkosta yli hakukonerajojen. Alkuperäiseltä nimeltään *BackRub* –niminen selain näki päivänvalon Stanfordin yliopiston verkkosivuilla. Tänä päivänä kaikista maailman verkkohauista 80 % tehdään Googlella. Itse käytin pitkään tavan vuoksi Yahoo-hakukonetta, kunnes annoin periksi Googlelle, joka on monin tavoin monipuolisempi ja käyttökelpoisempi. Huomionarvoista on, että vaikka Google on ylivoimaisesti maailman käytetyin hakukone, se ei suinkaan ole kaikkialla maailmassa se ainoa ja oikea hakukone. Windows-puhelimia käyttäville tuttu on Googlen kanssa Yhdysvaltojen markkinoita hallitseva Microsoftin Bing. Mainittujen Yahoon ja Bingin lisäksi huomioimisen arvoisia hakukoneita ovat jo aiemminkin mainitut Ask ja AltaVista.

Itärajan takaa löytyy hyvä mittapuu Googlelle: Yandex hallitsee tällä hetkellä Venäjän internetiä. Sen markkinaosuus on noin 60 prosenttia venäläisistä Internetin käyttäjistä. Toiseksi käytetyimpänä hakukoneena Yandexin jälkeen tulee Googlen venäjänkielinen versio. Näiden hakukoneiden käyttäjät poikkeavat hieman toisistaan. Google on liike-elämän ihmisten ja opiskelijoiden suosiossa ja Yandexin käyttäjät puolestaan ovat "perusvenäläisiä". Tämä ero näkyy myös hakukoneissa käytettävissä hakusanoissa. Googlen avulla haetaan tietoa asioista, kun taas Yandexilla haetaan erittäin paljon kaupallisia palveluja, tietoa arkielämän askareista ja kuluttamisesta. Huomioitavaa on, että sekä Google että Yandex eivät ole yksinomaan hakukoneita. Ne ovat myös kanavia, jotka tarjoavat koko joukon palveluja, kuten kartta-, musiikki-, raha-, sähköposti- ja liikennetietopalveluja. Siksi on luontevampaa puhua Yandexista ja Googlesta sähköisenä kanavana kuin pelkästään hakukonepalveluna.

20 http://blogs.scientificamerican.com/observations/how-black-holes-led-to-the-creation-of-web-browsers/

Hakukoneet voidaan jakaa yleisiin hakukoneisiin, metahakukoneisiin ja aihehakemistoihin. Yleiset hakukoneet perustuvat tietokantoihin. Ne ovat ohjelmia, jotka etsivät verkosta jatkuvasti uusia sivuja, analysoivat ne ja liittävät ne hakemistoonsa. Tällaisia ovat esimerkiksi Google, AllTheWeb ja HotBot. Metahakukoneet lähettävät haun monelle hakukoneelle samanaikaisesti ja keräävät niiden vastauksista linkkilistan. Metahakukoneet antavat paljon tuloksia, mutta ne voivat olla epätarkkoja. Metahakukoneita ovat esimerkiksi MetaCrawler ja WebCrawler. Aihehakemistot hakevat omista tietokannoistaan ja esittävät hakutulokset eräänlaisena indeksoituna linkkikirjastona, jossa linkit on luokiteltu.

Täysin avoimeen lähdekoodiin perustuvat hakukoneet kuten Duckduckgo ovat kasvaneet viimeisen kahden vuoden aikana. Duckduckgo vahvuutena on sen käyttäjälle takaama yksityisyys, Duckduckgo ei tallenna käyttäjän aikaisempia hakuja ip-osoitteen tai laitteen tunnuksen perusteella ja optimoi niitä käyttäjälle sopivammaksi. Edellä mainitulla on myönteiset ja kielteiset puolensa: juuri Googlen hakutottumustesi pohjalta tekemä optimointi tekee siitä niin helppokäyttöisen, toisaalta Google rekisteröi hakuhistoriaa ja profiloi käyttäjää jopa kiusaannuttavassa määrin (oman hakuhistoriansa statistiikan voi tarkistaa osoitteesta: www.google.com/history). Anonymiteetin vuoksi Duckduckgo on erityisen käyttökelpoinen esimerkiksi kiellettyjen aineiden, prostituoitujen ja aseiden etsimiseen, koska käyttäjän suljettua välilehden, mikään tieto ei ole tallella.

Toinen sekä anonyymi että sensuroimaton hakukone on Gibru, joka on taitava etsimään verkosta verkkosivujen jäänteitä ja niiden heijasteita. Gibru löytää tiedostoja ja sivustoja, jotka ovat joskus poistettu, mutta jääneet välimuisteihin eräänlaiseen välitilaan. Tällä hakukoneella löytää sopivilla hakusanoilla esimerkiksi räjähteiden teko-oppaita ja taistelukaasujen teko-ohjeita, jotka on saatettu poistaa verkosta vaarallisen sisältönsä vuoksi. Tämän linkin takaa löydät kymmenen selainta, jotka eivät suurella todennäköisyydellä ole sinulle ennestään tuttuja: http://sixrevisions.com/tools/10-web-browsers-you-probably-havent-heard-of. Jokaisella selaimella on omat vahvuutensa. Esimerkiksi Flock on erinomainen selain sosiaalisen median tehokäyttäjille, jotka haluavat kirjatua some-tileilleen yhdestä ja samasta paikasta.

Useimmat ihmiset eivät varsinaisesti osaa *käyttää* Googlea. Hakusanoja osataan toki hakukenttään ja navigoida hakutulosten kautta verkkosivustoille. Useimmat käyttäjät eivät edes vaadi hakukoneelta tämän enempää. Hyvä niin. Google-hauille ja Google-hakkeroinnille on omistettu kuitenkin monta hyllyrivillistä kirjoja, joihin perehtymällä pääsee kiinni Googlen sielunelämään. Nuo kirjat vievät kuitenkin meidät liian kauaksi itse tarkoituksesta. Usein meille saattaa riittää, että hihasta löytyy muutama näppärä hakukikka. Esittelenkin seuraavaksi joitakin Google-haun perusominaisuuksia, joita soveltamalla tiedonhausta tulee välittömästi nopeampaa ja tarkemmin rajattua.

3.2 Miten hakukoneet toimivat?

Hakukoneet eivät haravoi World Wide Webiä täysin esteettä, vaan jokainen penkoo omia tietokantojaan, jotka se on kerännyt ja tallentanut välimuistiin. Joka kerta kun käyttää hakukonetta, etsii ikään kuin luolan peräseinään langennutta varjokuvaa alkuperäisestä verkkosivusta. Kun hakukoneen hakutulosten linkkejä klikkaa, pääsee käsiksi web-sivun sen hetkiseen versioon.

Hakukoneen tietokannat ovat valikoitu ja rakennettu ”hämähäkeiksi” kutsutuilla hakuroboteilla. Nämä robotit etsivät web-sivuja jo tietokannassa olemassa olevien linkkien kautta. Hämähäkit eivät ole luovia; mikäli verkkosivulle ei pääse jo tietokantaan tallennetun verkkosivun linkin kautta, hakukonerobotit eivät myöskään löydä sitä. Tällöin ainoa tapa uuden verkkosivun lisäämiseen tietokantaan on sen linkittäminen tietokannassa olevaan verkkosivuun tai että ihminen kirjaa URL-osoitteen käsin.

Kun robotit löytävät verkkosivuja, ne välittävät tiedot indeksoinnista vastaavalle sovellukselle (Googlen kohdalla indeksoijarobottia kutsutaan Googlebotiksi). Tämä ohjelma tunnistaa tekstiä, linkkejä, sekä muuta sisältöä ja tallentaa ne hakukoneen tietokantaan siten, että tietokannasta voidaan etsiä sivuja tietyillä hakusanoilla tai muilla edistyneemmillä hakumenetelmillä.

Seuraavaksi Google louhii hakua vastaavat sivut ja tulokset, jotka ovat sen mielestä kaikkein osuvimpia ja presentoi ne käyttäjälle. Osuvuus määritetään yli 200 tekijän avulla, joista yksi on sivun PageRank-arvo. PageRank on 0-10 väillä oleva numeerinen arvo, ja se muodostetaan muilta sivuilta tulevien linkkien perusteella. Yksinkertaistettuna jokainen linkki sivuston sivulle toisesta sivustosta kasvattaa sivuston PageRank-arvoa. Erityisesti hakukoneoptimointia tarjoavat yritykset ovat kiinnostuneita kasvattamaan asiakkaan PageRankia erilaisin menetelmin verkkonäkyvyyden vuoksi.

Useimmat verkkosivut on suljettu pois kokonaan monien hakukoneiden hakutuloksista jo lähtökohtaisesti. Monet internetiin kytkeytyvät tietokannat eivät ole hakukoneiden tavoitettavissa, koska hakukoneiden robotit eivät voi käyttää niitä, kuten kirjasto- ja artikkelitietokannat. Tähän materiaaliin viitataan usein sanalla "näkymätön internet" tai ”Shadow Web”, koska tähän internetin elementtiin ei pääse käsiksi hakukoneiden kautta.

Hakutulosten toimittamisen kolme tärkeintä prosessia siis ovat:

1) Lukeminen: tietääkö Google sivuston olemassaolosta? Voimmeko löytää sen?

2) Indeksointi: voiko Google indeksoida sivuston?

3) Näyttäminen: onko sivustossa hyödyllistä sisältöä, joka vastaa käyttäjän hakua?

Googlen sivulta http://www.google.com/insidesearch/howsearchworks/thestory/ on mahdollista bongata erinomainen interaktiivinen ja havainnollinen esitys siitä miten Google-haku toimii.

Mikäli haluat tietää, että mitä Google tarkkaan ottaen sinusta tietää, niin ohessa on muutama linkki, joiden takaa löytyy taltioituna digitaalista jalanjälkeäsi:

1) Hakuhistoria – Täältä löydät koko Googlen hakuhistoriasi.

   (https://myactivity.google.com/myactivity)

2) Mainokset – Täältä löydät kaikki ne kiinnostuksen kohteen, jotka Google on päätellyt hakuhistoriasi perusteella.[21]

   (https://www.google.com/settings/u/0/ads/authenticated)

3) Sijaintitiedot – Tänne on listattu kaikki ne paikat, joissa olet tavalla tai toisella hyödyntänyt Googlea (edellyttänyt käytössäsi on ollut Sijaitihistoria-lisäosa).

   (https://www.google.com/maps/timeline?pb)

4) Omat tiedot – Kopioi yhdeksi tiedostoksi kaikki tiedot, jotka Googlella on minusta.

   (https://takeout.google.com/settings/takeout)

5) Hallintapaneeli – Hallintapaneeli kertoo kaikki ne Googlen ominaisuudet, jotka sinulla on käytössä.

   (https://myaccount.google.com/dashboard?pli=1)

6) Youtuben hakuhistoria – Nimensä mukaisesti Youtuben hakuhistoriasi.

   (https://www.youtube.com/feed/history/search_history)

7) Luvat ja oikeudet – Täältä näet kootusti kaikki luvat ja oikeudet, jotka olet antanut lisäosille ja verkkosivustoille

   (https://myaccount.google.com/permissions?pli=1)

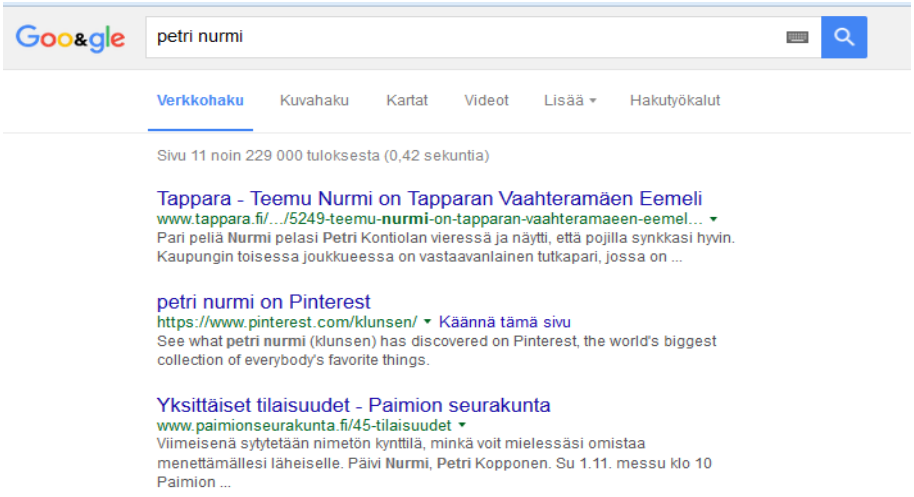
Suosittelen lämpimästi tutustumaan omaan digitaaliseen jalanjälkeen verkossa . Se edesauttaa kokonaisvaltaista ymmärrystä oman persoonallisen murupolun muodostumisesta verkkoympäristössä..

[21] Omalle kohdalleni oli kiinnostuksen kohteeksi mm. ´country-musiikki´, josta rohkenen olla eri mieltä.

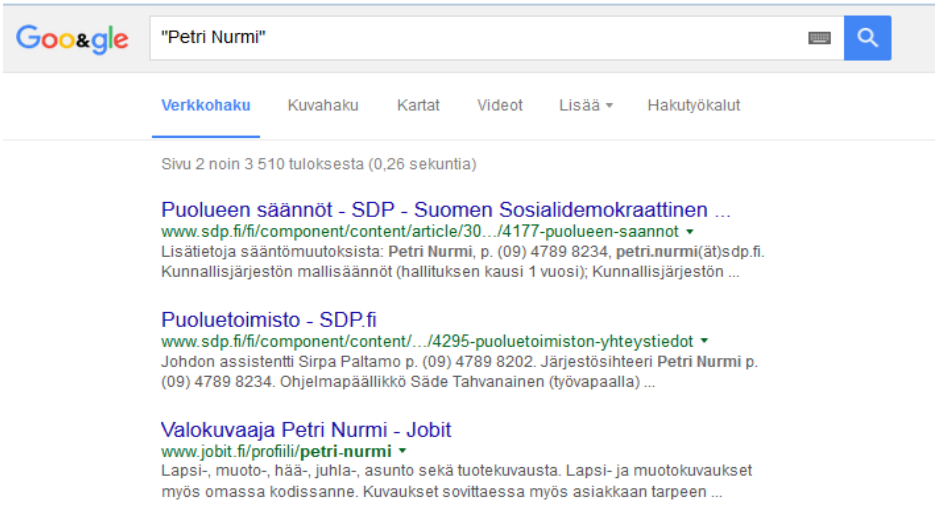
3.3 Googlen perushakutoiminnot

hakasulut

Haettavan kohteen hakasulkuihin laittaminen tekee selvän eron hakutulosten tarkkuuteen. Kirjoittamalla hakukenttään *Petri Nurmi*, saan vastaukseksi 226 000 osumaa, jotka ovat voittopuolisesti hakutavoitteeni kannalta hyödyttömiä, sillä suurimmassa osassa tuloksista sanat *Petri* ja *Nurmi* ovat erillään toisistaan. Teknisesti hakukone on toki suoriutunut tehtävästään moitteettomasti, onhan molemmat haettavat sanat mainittu hakutuloksessa, jos kohta useimmiten eri nimien yhteydessä (Kuva 1).

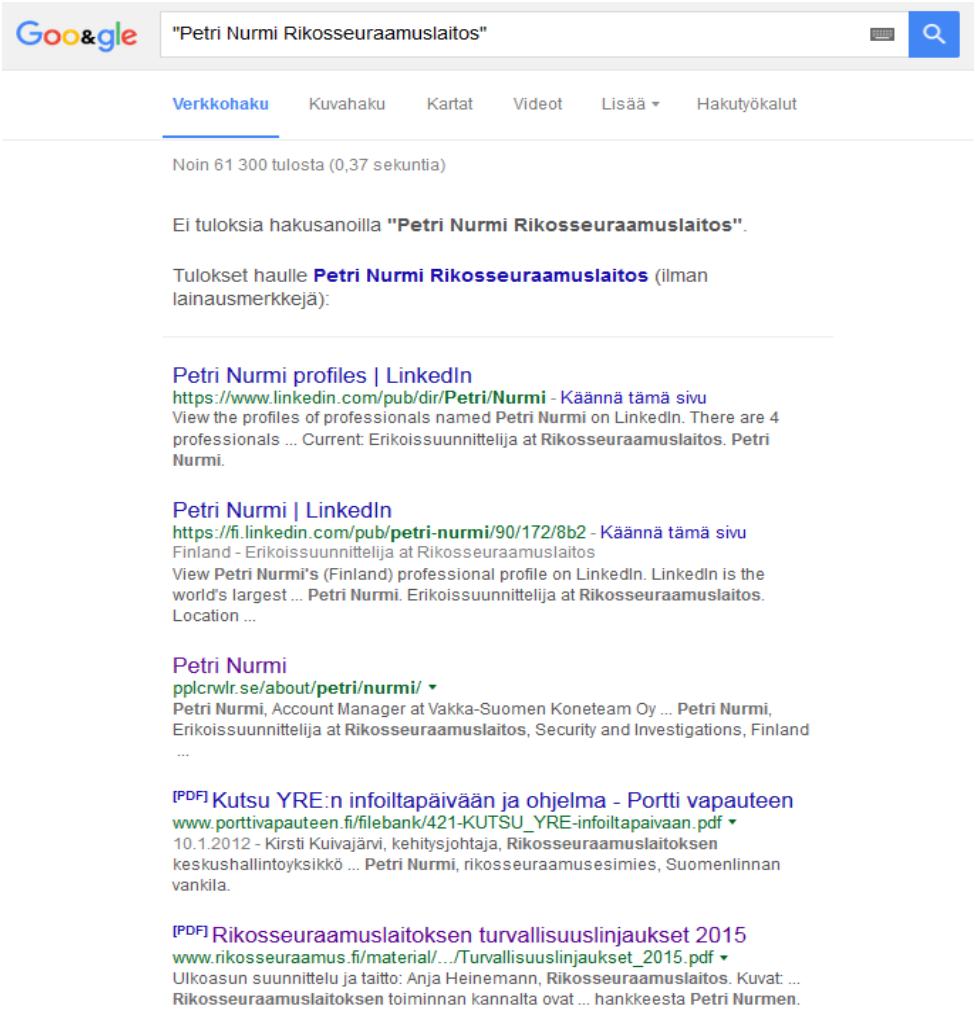


Kuva 1: Hakutulos ilman hakasulkuja



Kuva 2: Hakutulos hakasulkujen kanssa

Sen sijaan jos näppäilen nimeni ympärille hakasulut, tulos on jo kokonaan toisenlainen: hakutuloksia on tällä kertaa 3510 (Kuva 2) eli huomattavasti vähemmän kuin aiemmin. Sekä itsestäni, että nimikaimoistani löytyy nyt pienemmällä vaivalla eksaktimpaa tietoa. Tämä hakutekniikka on erityisen hyödyllinen haettaessa sähköpostiosoitteita tai käyttäjänimiä. Mikäli 3510 hakutulosta tuntuu yhä liian lukuisalta, voi lisätä nimeni perään hakuani koskevan lisätiedon: ”Rikosseuraamuslaitos”. Osumia on hakukriteerien lisääntymisen vuoksi taas selkeästi enemmän, mutta nyt löydän jo useita suoraan itseeni viittaavia hakutuloksia.



Kuva 3: Hakutulokset haulla ”Petri Nurmi Rikosseuraamuslaitos”

kysy Googlelta kysymyksiä

Aina tarvitse kirjoittaa hakukenttään ainoastaan avainsanoja, vaan haun voi muotoilla myös suoraan kysymykseksi. Mikäli haluat tietää vaikkapa kuinka pitkä välimatka on Jyväskylästä Helsinkiin, hakukenttään ei tarvitse kirjoittaa: ”välimatka, Jyväskylä, Helsinki”. Kenttään voi reilusti kirjoittaa kysymyksen, johon haluat saada vastauksen: ”Kuinka pitkä matka on Jyväskylästä Helsinkiin?”

(-), and, or

Googlen algoritmi on erittäin sopeutuvainen ja osaa hakea oikeanlaista tietoa, vaikka käyttäjä ei välttämättä olisi itsekään täysin varma siitä mitä etsii. Silloin kun käyttäjä tietää tarkalleen mitä etsii, on mahdollista syöttää hakukenttään tiettyjä perusmuuttujia. Mikäli haluat hakea tietoa Amazonista ja kirjoitat hakuknettään ”Amazon”, luultavammin saat hakutuloksena enimmäkseen markkinointitarkoituksessa tehdyn hakukoneoptimoinnin seurauksena enimmäkseen tuloksia www.amazon.com –verkkosivun sisällöstä. Jos sen sijaan kirjotat hakukenttään ”amazon and rainforest”, niin hakutulos on täsmällisempi, sillä hakutuloksissa tulee olla sisällytetynä nuo kaksi muuttujaa.

hakuoperaattorit

Hakuoperaattorit ovat hakuihin lisättäviä sanoja, joilla voidaan rajata hakutuloksia. Ohessa muutamia Googlen hakuoperaattoreita, jotka tuottavat tavoiteltuja hakutuloksia tavallista hakua tehokkaammin ja täsmällisemmin.

cache: www.kaalimato.com

Googlen välimuisti verkkosivulle www.kaalimato.com

link: www.kaalimato.com

Lista verkkosivuista, jotka on linkitetty www.kaalimato.com

related: www.kaalimato.com

Lista verkkosivuista, jotka muistuttavat verkkosivua www.kaalimato.com

info: www.kaalimato.com

Presentoi kaiken informaation, mitä Googlella on sivusta www.kaalimato.com

site: www.kaalimato.com

Lista kaikista verkkosivuista, jotka hallinoidaan sivulta www.kaalimato.com

allintitle: kaalimato

Rajaa hakutulokset niihin otsikoihin, joissa on sana ´kaalimato´

intitle: kaalimato

Rajaa hakutulokset niihin dokumentteihin, jotka sisältvät sanan ´kaalimato´

inurl: kaalimato

Rajaa hakutulokset niihin URL-osoitteisiin, joissa mainitaan sana ´kaalimato´

allinurl: kaalimato vibraattori

Rajaa hakutulokset niihin URL-osoitteisiin, joissa mainitaan molemmat hakusanat

world * kaalimato

Sanat ´kaali´ja ´mato´erottaa vain yksi sana

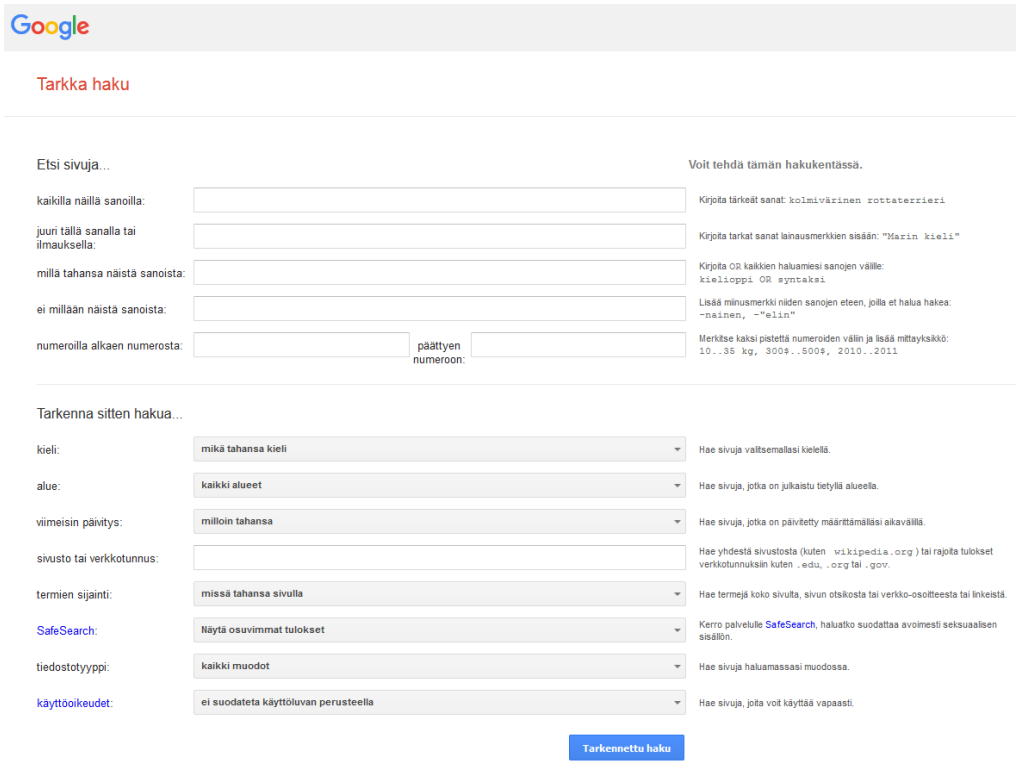
Nokia Lumia €100…300

Haku ainoastaan Nokia Lumia puhelimista, jotka asettuvat hinnaltaan välille €100-300

safesearch: kaalimato

Haku ´kaalimatoa´koskevaan materiaaliin, josta on karsittu pois aikuisviihdesivustot

Käyttäjän ei tarvitse opetella hakuoperaattoreita ulkoa, sillä voit luoda edellisen kaltaisia täsmähakuja myös Tarkennettu haku-sivulla. Operaattoreista ”site:” toiminto on erityisen käyttökelpoinen. Operaattorilla on kaksi hyödyllistä ominaisuutta: se hakee sekä tietoa ainoastaan erikseen nimetyltä verkkosivulta että kaikki hakutulokset verkkosivulta, jossa hakutermi on mainittu. Esimerkiksi mikäli haluaa hakea OSINT:iin liittyvää tietoa ainoastaan Ilta-Sanomien verkkosivulta, hakukenttään voi kirjoittaa esim. ”site:iltasanomat.fi osint”. Tarkennettu haku-näytöllä hakuoperaattorit on jäsennetty näppärästi omiin hakukenttiinsä seuraavaan tapaan (Kuva 4):



Kuva 4: Googlen tarkennettu haku

3.5 Henkilöhakukoneet

Siinä missä Google ja Bing erikoistuvat internetissä olevan sisällön etsimiseen, henkilöhakuja varten on olemassa erikoistuneita hakukoneita. Nämä hakukoneet hyödyntävät Googlea ja Bingia kokoamaan dataa ja esittämään sen helposti pureskeltavassa muodossa. Jokaisella alla listatulla hakukoneella on vahvuutensa ja heikkoutensa. Hakukoneet ovat ilmaisia käyttää, mutta usein mainostuloja pääasiallisena rahoituskanavanaan käyttäviä, joten niillä operoimista saattavat haitata erilaiset ponnahdusikkunat ja välkkyvät mainokset. Useimmat henkilöhakuihin erikoistuneet hakukoneet ovat valitettavasti suunnattu Yhdysvalloissa asuviin henkilöihin eivätkä näin ollen tuota hakutuloksia suomalaisesta, albanialaiseta tai andorralaisesta yksityishenkilöstä. Olenkin valinnut kattavasta valikoimasta viisi suomalaiseen henkilöhakuun parhaiten soveltuvaa henkilöhakukonetta.

Pipl (https://pipl.com)

Tämä sivusto väittää olevansa verkon kattavin henkilöhakuja tekevä hakukone. Hakukone on kiistatta tehokas. Hakutuloksia tulee runsaasti, joten niitä on syytä rajata vaikkapa paikkunta-parametrilla. Profiilit-osio esittää nimihaun perusteella sosiaalisen median alustoille luotuja profiileja, kuten Facebook, Twitter, MySpace, Youtube ja Meetup. Henkilöhakuja tekevistä hakukoneista Pipl on kenties se hakukone, jota on syytä pitää oletushakukoneena ihmisten etsimiseen.

Yasni (www.yasni.com)

Päällisin puolin Yasni vaikuttaisi olevan melko standardinomainen henkilöhakuja tekevä hakukone. On kuitenkin joitakin osa-alueita, joissa Yasni toimii muista henkilöhakukoneista poikkeavalla tavalla. Lähtökohtaisesti Yasni antaa kolme hakuoptiota: 1) Etsi henkilöitä, jotka tietävät 2) Tarjoan, voin, olen… 3) Mitä netti tietää… OSINT-tarkoitusperiin viimeinen hakuvaihtoehto on tarpeellisin. Hakutoiminnallisuudessa on mahdollista käyttää joko käyttäjänimeä tai oikeaa nimeä ja tulokset paljastavat erilaisia linkkejä, jotka ovat yhdistettävissä kohteeseen. Itse ainakin löysin LastFM-profillini tätä kautta, joita ei muista hakukoneista ole tullut esille.

Peek You (www.peekyou.com)

Peek You analysoi yli 60 sosiaalisen media sovellusta, uutissivustoja, kotisivuja, blogialustoja ja tunnistaa henkilön näiden takaa. PeekYou tekee tämän jäljittämällä verkossa hajallaan olevia

digitaalisia jalanjälkiä ja luo niistä kokonaisvaltaisen online-identiteettejä. Epäilemättä yksi parhaista tarkoitukseen luoduista hakukoneista.

Knowem (www.knowem.com)

Joskus ainoana johtolankana kohteesta saattaa olla vain käyttäjänimi. Käyttänimen ja henkilön yhteen linkittämisessä auttaa hakukone nimeltä Knowem. Knowem on kokonaisvaltaisin hakukone käyttänimen ja profiilin linkittämiseen. Käyttäjänimen perusteella hakukone tarkistaa käyttäjänimen saatavuuden kaikista sosiaalisen median alustoista. Mikäli käyttäjänimi ei ole käytössä jossakin some-palvelussa, tämä saattaa indikoida sitä, että kohteella saattaa olla profiili luotuna ko. sovelluksessa. Tämä tulos edesauttaa eteenpäin kohteen some-profiilin löytämisessä.

Name Chk (www.namechk.com)

Itse pidän kuitenkin enemmän seuraavasta, samaan tarkoitukseen tehdystä hakukoneesta. Se ei tee hakuja yhtä kattavasti kuin edellinen sovellus, mutta se on käyttöliittymältään selkeämpi. Name Chk –sovelluksella saatavien hakutulosten etu on, että klikkaamalla ”taken” tulosta, on mahdollista siirtyä suoraan siille verkkosivulle ja siihen profiiliin, joka käyttäjänimeä käytttää.

IntelTechniques (https://inteltechniques.com/menu.html)

Jos käytössäsi on vain puhelinnumero, niin suositeltavaa on käyttää Inteltechniquesin työkalua. Työkalun avulla on mahdollista paikallistaa sellainen verkkosivu tai some-profiili, joka kyseistä numeroa käyttää. Hakukentät on luotu lähtökohtaisesti Iso-Britannian alueella tapahtuvaa numerohakua varten, mutta mukana on myös kansainväliseen numerohakuun soveltuva hakukenttä. Itse ainakin testasin onnistuneesti puhelinnumeroani, joka on varsin tiukasti salattu. Mainittakoon myös, että IntelTechniques on sivusto, joka täytyy varsin monipuolisesti OSINT-tietdonhakua harjoittavan verkkodekkarin tietotarpeet räätälöidyillä hakuoperaattoreillaan. Sivustoon viitataankin tässä tutkielmassa useasti.

4. FACEBOOK JA TWITTER

4.1 Sosiaalinen media

Tämän tekstin kirjoittamisen aikana päättyi yksi sosiaalisen median aikakausi: Jyväskylän yliopisto lakkautti irc-palvelimensa.[22] Heinäkuussa 2016 palvelin ei enää vastannut yhteydenottopyyntöihin. Suomalaisten kehittämä irc-keskusteluverkko oli 1990-luvun sosiaalinen media. Erityisesti tunnettuutta sai nk. irc-galleria, jonka toiminta-ajatuksena oli mahdollistaa keskustelukumppanin näkeminen muutoin tekstipohjaisessa irc-keskustelussa. Jyväskylän yliopiston irc.jyu.fi-palvelin salli avoimena palvelimena yhteydet muillekin suomalaiskäyttäjille kuin Jyväskylän yliopiston opiskelijoille ja henkilökunnalle. Palvelu oli vuosien mittaan vähitellen jäänyt tarpeettomaksi some-palvelujen kehityksen myötä sekä verkossa yleisesti laajentuneen sosiaalisen median palvelutarjonnan myötä.[23]

Edellisestä hyppy vuoden 2016 kevään otsikoihin nousseeseen uutiseen, missä Facebook määrättiin lopettamaan palvelua käyttämättömien ihmisten seuraaminen verkossa. Brysseliläinen oikeusistuin velvoitti Facebookin noudattamaan määräystä belgialaisten osalta 250 000 euron sakon uhalla jokaisen päivän osalta, kun määräystä ei noudatettu.[24] Kysymys on yhdestä tietystä evästeestä eli cookiesta[25], jonka Facebook tallentaa nettikäyttäjän selaimeen heidän käydessään *Facebook.comissa* tai, kun he klikkaavat Facebookin *tykkää*-nappia muilla verkkosivuilla.

Facebook perusteli evästettä turvallisuudella ja väitti, että sillä ei voi tunnistaa ihmisiä, mutta sillä voi tunnistaa hyökkäyksiä palvelimia vastaan. Facebook näyttää tätä nykyä mainoksia niillekin käyttäjille, jotka ovat pysytelleet palvelun ulkopuolella. Samalla yhtiö on laajentanut tapaa, jolla se näyttää mainoksia Facebookin ulkopuolisilla sivuilla. Facebook-tunnuksen omistavat verkon käyttäjät voivat muokata näitä mainosasetuksia. Samalla he voivat ottaa

22 IRC:n kehitti Jarkko "Wiz" Oikarinen vuoden 1988 kesällä korvaamaan OuluBox BBS:n MUT ("MultiUser Talk") -ohjelman. IRC:n edeltäjä ja inspiraatio oli Bitnet Relay Chat -niminen ohjelma, joka mahdollisti reaaliaikaisen keskustelun kansainvälisten akateemisten verkkojen muodostamassa BITNET-verkossa. Ensimmäinen IRC-palvelin aloitti toimintansa Oulussa elokuussa 1988, osoitteenaan tolsun.oulu.fi.

23 http://www.tivi.fi/Kaikki_uutiset/kayttakaa-whatsappia-ja-facebookia-yliopisto-sammuttaa-legendaarisen-irc-palvelimensa-6555896

24 http://www.itviikko.fi/uutiset/2015/11/11/kello-tikittaa-facebookille-kohta-pitaa-maksaa--eika-ihan-vahan/201514831/7

25 Evästeet ovat dataa, jota web-palvelin tallentaa käyttäjän tietokoneelle. Selain lähettää tiedon takaisin kyseiselle palvelimelle joka pyynnön yhteydessä. Evästeiden avulla voidaan tallentaa ja välittää käyttäjän tekemiä personointivalintoja. Evästeitä on toisaalta myös pidetty riskinä käyttäjälle, mikä johtuu siitä, että niitä voidaan tietyissä tapauksissa käyttää sivustojen ulkopuolella ja ne mahdollistavat käyttäjän seurannan ulkopuolisien palvelimien kautta.

selvää, millä perusteella minkälaisia mainoksia näytetään. Navigoimalla verkko-osoitteeseen www.facebook.com/ads/preferences/edit, on nähtävissä minkä aihepiirin mainoksia sinulle näytetään ja miksi. Syitä voivat olla esimerkiksi sivusta tykkääminen, mainoksen klikkaaminen – tai Facebookin omat päätelmät aiheen kiinnostavuudesta sinulle.

Joka kerta kun vierailet verkkosivulla, sivu istuttaa evästeen tietokoneeseesi tai puhelimeesi. Näiden tiedostojen kautta on mahdollista jäljittää sinut ja digitaalisen jalanjälkesi verkossa. Lisäksi jokaisella käyttämälläsi älylaitteella on oma uniikki digitaalinen tunnisteensa, joka mahdollistaa jäljittämisesi. Yksilöivät tekniset tunnisteet kuten IP-osoite[26], MAC-osoite[27]ja puhelimesi IMEI- tai IMSI-koodi[28] mahdollistavat verkkopalveluja tarjoavien yritysten tietää tarkalleen mitkä laitteet (ja käyttäjät) käyttävät heidän palvelujaan. Mikäli et itse tiedä yhteytesi ja päätteesi tietoja (jonka näppäimiä parhaillaan painelet), voit selvittää sen klikkaamalla tätä linkkiä: http://www.computerhope.com/cgi-bin/systeminfo.cgi.

Kaikki tämä data jäljitetään, yhdistellään ja hyödynnetään verkkopalveluyritysten markkinointitarkoituksiin. *Wall Street Journalin* vuonna 2010 teettämän tutkimuksen[29] mukaan yksi nopeimmin kasvavista yritystrendeistä on internetin käyttäjien vakoileminen. Raportissa kävi muun muassa ilmi, että 50 suosituinta verkkosivustoa istuttivat käyttäjän laitteelle keskimäärin 64 evästettä. Verkkosivu, joka istutti eniten evästeitä oli yllättäen niinkin yleishyödyllinen sivusto kuin www.dictionary.com. Sivusto latasi kaikkiaan 234 evästettä käyttäjän päätteelle jokaisella vierailulla. Kaikki nämä evästeet kytketään sitten Facebookin Tykkää-painalluksiin ja twiitteihin. Hiljalleen nuo evästeet paisuvat melkoiseksi ”cookie-monsteriksi”, jotka paljastaa sinusta sellaista tietoa, jota et koskaan halunnut tulevan edes julki.

Toinen erittäin suosittu Facebookin omistama sosiaalisen median sovellus *WhatsApp* suljettiin toukokuussa 2016 brasilialaisen oikeuden päätöksellä kolmeksi vuorokaudeksi.[30] Syynä tähän oli, että WhatsApp ei suostunut luovuttamaan kryptaamaansa viestinvaihtoa huumerikostutkintaa varten. WhatsApp perusteli päätöstään sillä, että he eivät ensinnäkään ylläpidä tämänkaltaista

---

26 IP-osoite (Internet Protocol) on numerosarja, jota käytetään IP-verkkoihin kytkettyjen verkkosovittimien yksilöimiseen. Kaikki Internet-verkon tietoliikenne kulkee IP-paketteina. IP-osoitteen perusteella IP-paketti löytää perille ja vastaukset tulevat takaisin. Tavallisesti IP-osoite esitetään neljän pisteellä erotetun luvun sarjana, esimerkiksi *145.97.39.155*. IP-osoite ei yksilöi käyttäjää.

27 Mac-osoite (Media Access Control) on verkkosovittimen lähi- tai langattomassa verkossa yksilöivä osoite.. Se on useimmiten fyysisesti kirjoitettu jo tehtaalla kortille, mutta sitä voi myös vaihtaa ohjelmallisesti jälkikäteen. Osoite koostuu kuudesta kaksinumeroisesta heksadesimaalisesta luvusta, joista kolme ensimmäistä on valmistajan itselleen varaama etuliite ja kolme jälkimmäistä on juokseva sarjanumero.

28 Imei-koodi (International Mobile Equipment Identity) on viisitoista merkkiä pitkä matkapuhelimen laitetunnus, jolla voidaan tunnistaa matkapuhelin verkosta. Laitetunnuksiin siirryttiin, koska matkapuhelimia varastettiin ja niiden SIM-kortit vaihdettiin toisiin, minkä vuoksi pelkkä SIM-tunnuksella laitteen tunnistaminen verkosta ei ollut riittävä tapa vaikuttaa puhelimen käyttöominaisuuksiin. Koodin saa näkyviin näppäilemällä #06# numeronäppäimistöllä.

29 http://www.wsj.com/articles/SB10001424052748703940904575395073512989404

30 http://www.engadget.com/2016/05/02/brazilian-government-whatsapp-block-72-hours/

seurantaa ja vaikka pitäisivätkin, niin viestien kryptaus ei ole murrettavissa. WhatsApp ei myöskään halunnut vaarantaa miljardin käyttäjänsä tietoturvaa tämänkaltaisten poikkeustapausten perusteella.

Vain kymmenen vuotta Facebookin luomisesta vuonna 2004, käyttäjien määrä lisääntyi nollasta 1,3 miljardiin. Nykyinen käyttäjäjäluku lähentelee 1,6 miljardia. Joka päivä Facebookiin ladataan noin 350 miljoonaa valokuvaa ja postauksia peukutetaan kuusi miljardia kertaa. Sosiaalinen media kronikoi vuosi vuodelta tarkemmin elämäämme: meidän matkojamme, lastemme syntymiä, työpaikkojamme, uusia lemmikkejämme, avioliittojamme ja erojamme. Mark Zuckerbergiin Harvardin yliopiston opiskelijaboksissa kehittelemä palvelu on oma lukunsa tavallisten ihmisten globaalissa verkostoitumisessa. Facebookin kirjautumistunnukset toimivat federoitumisen[31] kautta useiden muiden verkkopalvelujen tunnuksina jopa siinä määrin, että niistä on tullut verkossa sähköpostiosoitteemme kaltainen passimme verkkoympäristössä.

Some-aktiivisuudella on nykypäivänä hieman negatiivinen konnotaatio ja sen kautta se on myös helppo piikittelyn kohde siitä vähemmän piittaaville tai muuten some-kriittisille. Nykypäivää kuitenkin on, että työpaikat jopa kannustavat some-aktiivisuuteen ja monelle Facebook korvaa lähestulkoon muut viestintäkanavat. On myös harhaluulo, että sosiaalinen media yksiselitteisesti kaventaisi inhimillistä kanssakäymistä, totuus on päinvastainen: erilaiset kortteliyhteisöt ja saunaseurat mainostavat kohdeyleisölle tempauksiaan, vanhat heilat löytävät uudelleen toisensa, omaa yhteisöään etsivä ihminen löytää kaltaisiaan ja mannerten väliset kaverukset ylläpitävät ystävyyttään vielä vuosienkin päästä.

Sosiaalinen media voi toimia myös keskeisenä geopoliittisten globaalien muutosten instrumentteina kuten näimme 2010 Arabikeväänä.[32] Googlen tietokoneinsinööri ja internet-aktivisti loi Facebook-sivuston, jonka tarkoituksena oli tuoda esiin Hosni Mubarakin sisäisen turvallisuuden joukkojen tekemä egyptiläisen kansalaisaktivistin raaka murha. Kahden minuutin jälkeen sivustolla oli 300 seuraajaa. Kolme kuukautta myöhemmin seuraajien määrä oli kasvanut 250 000 seuraajaan. Myös muut sosiaalisen median palvelut osallistuivat voimakkaasti Tunisian, Iranin ja Libyan kansannousuihin todistamalla, että ne voivat toimia myös positiivisena voimana.[33]

Useimmat meistä viettävät niin ison osan vapaa-ajastamme sosiaalisessa mediassa, että emme välttämättä edes halua tietää tarkalleen, että kuinka pitkään. Kun emme selaile Facebookin tai

31 Federoitumisella tarkoitetaan joukkoa verkkopalvelujen tarjoajia, jotka sopivat yhteisesti tierityisistästandardeista. Termiä voidaan käyttää kuvattaessa kahden erillisen, virallisesti erillään toimivan, ohjelmiston yhteenliittymää. Esimerkiksi Spotifyn federoituminen Facebookiin mahdollistaa palveluun kirjautumisen Facebookin kautta.

32 Arabikevät on mielenosoitusten, kansannousujen ja väkivaltaisuuksien aalto, joka alkoi Tunisian vallankumouksesta vuodenvaihteessa 2010–2011. Arabikevät oli vastareaktio paikoilleen pysähtyneille ja itsevaltaisille hallituksille. Arabikevään pohjimmaisena syynä olivat tavallisen kansan huonot elinolot, muun muassa työttömyys ja ruoan hintakriisi.

33 Goodman, 2015, s. 58

vastaavien some-palveluiden uutisvirtaa, niin luemme iltapäivälehtien uutisvirrasta julkisuuden tai muiden silmää tekevien henkilöiden some-päivityksistä. Donald Trumpin äkkiväärät Twitter-päivitykset ovat pakottaneet viimeisetkin somea vastustaneet mohikaanit perustamaan oman tilin, jotta pääsisisvät lukemaan ajantaisaiseti vapaan maailman johtajan viimeisimpiä nokkeluuksia. Mutta oletko koskaan pysähtynyt miettimään, että miksi Twitter, Google tai Facebook ei lähetä kaikesta tästä lystistä laskua? Nämä yritykset antavat meille mahdollisuuden tallentaa ja jakaa videoita ja valokuvia, perustaa sähköpostilaatikoita, seurata alati päivittyvää uutisvirtaa, mutta mitä he saavat tästä vastineeksi? Useimmat meistä arvelevat, että varmaankin yritykset saavat näiltä palveluilta mahdollisuuden mainostaa tuotteitaan oikeille kohderyhmille. Tämä on kuitenkin vain osatotuus. Vastineeksi nämä some-palvelut saavat meistä sekä kiinnostuksemme kohteista auringontarkkaa tieoa esimerkiksi päivityksiini liittämillä maantieteellisellä sijainneilla, ystävilläni, sekä nk. somekoodauksella.

Ihmisen perustarpeisiin kuuluu tarve tulla kuulluksi ja nähdyksi. Somekoodaus tukee tätä perustarvetta. Somekoodaus on itse asiassa elementti, joka auttaa merkittävästi tutkijan kohdehenkilöön kohdistamaa tiedonhankintaa. Yksi keskeisimpiä syitä sosiaalisen median voittokulkuun on se, että nykyihmisen ilman sitä viesit hukkuvat sosiaalisessa mediassa eivätkä ne eivät ankkuroidu oikeisiin teemoihin. Erityisesti Twitterissä ja Instagramissa, jossa kilvoitellaan seuraajien määrässä ja informaatiovaikuttamisessa, avainsanojen linkittäminen viesteihin on välttämätöntä. Linkittämällä avainsanan hastagilla (#) pystyy linkittämään viestin laajempaan kontekstiin.

Hashtag, suomeksi *avainsana* tai *aihetunniste*, on kokonaisuus, joka muodostetaan ristikkomerkillä ja sitä seuraavalla sanalla tai merkkijonolla. Metatietotunnisteisiin kuuluva hashtag helpottaa viestintää; sen avulla viestit ohjautuvat oikeaan paikkaan ja sen avulla voi hakea tiettyyn aiheeseen liittyviä viestejä. Hashtagin perimmäinen tarkoitus on saattaa saman aihealueen viestit yhteen, jotta tiedonhakijan olisi helpompi tarkastella ja yhdistellä tietoja. Sosiaalisen median sovelluksissa, kuten Twitterissä, käyttäjä voi muodostaa hashtagin tagaamalla viestin sisältöä kuvailevan sanan tai fraasin, esim. “Aalto-yliopiston #turvallisuusjohtaminen"

Toinen paljon somekoodauksen tapa on kohdentaa viesti @-merkillä eli jos haluan kohdentaa viestini erityisesti Jari Sarasvuolle, niin kirjoitan viestiini *@jarisarasvuo* edellyttäen tietysti, että Jari Sarasvuo käyttää Twitterissä juuri tuota käyttäjänimeä tunnuksenaan (huomioitavaa on, että Twitter ei ymmärrä suomen kielen taivutusmuotoja). Viimeksi Facebookin Messenger-sovellus otti käyttöön kyseisen tavan linkittää viestiin henkilöitä.

4.2 Facebook (www.facebook.fi)

Facebook-tiedon keräämisen esivalmistelut

Facebookin käyttäjät tapaavat suojata tietonsa muita some-alustoja paremmin. Jo oletuksena Facebookin käyttäjän tulee määritellä yksityisyysasetuksensa jo profiilia luodessaan. Monet näistä asetuksista eivät kuitenkaan edesauta mainittavasti yksityisyyttä ja jättävät valtavasti tietoa kaikkien nähtäville. Viimeaikaisten uudistusten myötä Facebookin anonymiteetti on muuttunut vaikka anonyymi profiilien selailu on kuitenkin ollut alusta alkaen yksi tärkeimmistä piirteistä. Nyt jokainen päivityksen Facebookissa julkaissut näkee, ketkä sen ovat katsoneet. Aiemminkin on ollut mahdollista lähdekoodista todeta oman profiilinsa katselijat.

Profiilin lähdekoodiin pääsee klikkaamalla profiilissasi hiiren oikeaa näppäintä ja valitsemalla valikosta *Näytä sivun lähdekoodi*. Lähdekoodi-näytöllä paina CTRL+F ja hae hakusanalla ”InitialChatFriendsList”. Nyt löydät hakusanan maalattuna lähdekoodista ja pitkän listan FB ID-numerosarjoista, jotka edustavat profiilissasi vierailleita. Kopioi jokin numerosarjoista ja syötä se Facebookin URL-osoitteeseen tähän tapaan:

www.facebook.com/1049857393

Tämän menetelmän opeteltuasi saat tietoosi jokaisen profiilissasi vierailleen FB-profiilin halutessasi. Edellä kuvattu ominaisuus edellyttää siis varotoimenpiteen tiedonhankintaa silmällä pitäen. Onkin syytä perustaa oma profiilinsa tarkoitusta varten, jolla ei ole minkäänlaista yhteyttä henkilökohtaiseen käyttäjäprofiiliisi. Apuna voi käyttää vaikkapa verkkosivua nimeltä www.fakenamegenerator.com/, joka generoi sukupuolen, kansallisuuden ja nimen perusteella hyvin yksityiskohtaiset tiedot valeprofiilille.

On olemassa kirjaimellisesti satoja tapoja rajata ja täsmentää hakuehtoja. Kärsivällisyydellä ja systemaattisuudella on tässäkin toiminnassa suuri arvonsa. Hyvin harvoin kohdehenkilö on asettanut yksityisyysrajoituksensa niin tiukoiksi, että käyttökelpoista tietoa ei ole mahdollista saada. Tällaisessa tapauksessa on syytä vain muuttaa tulokulmaa ja lähestyä kohdetta FB-kaverilistojen kautta. Aloittelijan virhe onkin, että työrukkaset tiippuvat välittömästi maahan, mikäli kohteen FB-profiili on salattu. Tämä ei ole suinkaan este, ainoastaan hidaste.

Tiedonhankinnan aloittaminen on palkitsevampaa aloittaa kohteeseen liittyvistä asioista, eikä niinkään itse kohteesta. Mikäli kuitenkin on niin onnellisessa asemassa, että käytössä on jo lähtökohtaisesti puhelinumero tai vaikkapa sähköpostisoite, niin kannattaa hyödyntää Netbootcampin työkalua www.netbootcamp.org/facebook.html FB-profiilin löytämiseen, sikäli kun sellainen on käyttäjälle luotu.

Viimeinen, mutta kenties tärkein esivalmistelun toimenpide, on selvittää kohteen FB-profiilin ID-numero. Sen selvittäminen on hyvin yksinkertaista. Verkosta löytyy muutamia sovelluksia kuten www.findfacebookid.com tai www.findmyfbid.com, jotka selvittävät ID-numeron kädenkäänteessä syöttämällä hakukenttään profiilin URL-osoitteen.

Henkilökohtaisten lisätietojen louhiminen

Yksilöivää ID-tunnistetta on mahdollista hyödyntää Facebookin Graph Search-ominaisuudessa. Ominaisuus herätti aikanaan keskustelua muun muassa siitä syystä, että ihmiset pääsivät näkemään toistensa tykkäykset ja kommentit kootusti yhdessä listassa. Tämän vuoksi Graph Searchin käyttötapaa on muutettu hankalampaan suuntaan. Perinteisellä tiedonhaullahan Facebookissa on omat heikkoutensa. Jos vaikkapa kirjoitat Facebookin hakukenttään hakemasi aiheen, henkilön tai ryhmän, niin sovellus antaa tuloksen ystäviesi sekä sosiaalisen verkostosi läheisyyden perusteella. Tämä johdattaa varmasti harhaan mikäli kohde ei ole vaikutuspiirisi lähettyvillä.

Graph Searchissa hakuja tehdään URL-kentässä. Logiikka on varsin yksinkertainen, jos kohta ei kovin käyttäjäystävällinen: kohteen Facebook-profiilin URL-osoite on seuraavanlaisessa muodossa:

www.facebook.com/käyttäjänimi

Mikäli haluamme nyt hyödyntää URL-osoitteen hakuominaisuuksia selvittämällä mistä kohde on tykännyt, missä tämä on vieraillut tai mistä sivuista on vieraillut yms., niin kirjoitamme FB ID-numeron hakukenttään ja muokkaamme URL-kenttää seuraavanlaisesti:

www.facebook.com/search/ 100013441634036/places-visited

Tämän jälkeen vain muunnellaan viimeisen osion muuttujaa tietotarpeiden mukaan esim.

stories-commented

stories-tagged

groups

relatives

places-liked

photos-by

photos-liked

pages-liked

photos-of

photos-commented

videos

videos-by

videos-of

videos-liked

apps-used

friends

events

events-joined

stories-by

Ystävien tiedot

Kuten mainittu, toisinaan kohteen oman profiilin tarkastelu ei aina tuota toivottua informaatiota. Tällöin kohteen ystäväpiirin intressien analysointi edesauttaa kokonaiskuvan hahmottamista. Helpoiten tämä tapahtuu soveltamalla taas kerran Graph Search-ominaisuutta viemällä URL-hakua askeleen pidemmälle. Edelleen käytössämme on yllämainittu lista spesifeistä some-aktiivisuutta mittaavista mittareista. Tällä kertaa URL-osoitekenttä näyttäisi seuraavanlaiselta mikäli haluamme tietoa kohteen ystävien kiinnostuksen kohteista:

www.facebook.com/search/1089577639/friends/pages-liked

Taas kerran voimme soveltaa yllä olevaa osoitekenttää muuntelemalla osoitteen viimeistä hakuparametriä edellisen kappaleen listauksen mukaisesti esim.

www.facebook.com/search/105859bce0439/friends/groups

Yhteiset hakutulokset

Mikäli taas haluamme lisätietoa kahden käyttäjäprofiilin välisestä aktiivisuudesta tai yhtymäkohdista, niin menemme taas askeleen pidemmälle Graph Searchin hyödyntämisessä. Toisinaan saattaa olla hyödyllistä liputtaa paikkoja, tapahtumia, valokuvia tai ryhmiä, joissa kaksi profiilia ovat kohdanneet. Nämä tiedot eivät välttämättä ole näkyvissä kunkin henkilökohtaisissa profiileissa. Tämä tapahtuu linkittämällä URL-osoitteeseen kaksi profiilia ja lisäämällä osoitteeseen hakuparametriksi sanan ”intersect” seuraavaan tapaan:

www.facebook.com/search/profiili1/pages-liked/profiili2/pages-liked/intersect

OSINT-tutkijat käyttääneet yllä mainittua hakutekniikkaa muun muassa siihen, että kaksi henkilöä ovat väittäneet olevansa tuntemattomia toisilleen, mutta Facebookista liputetut valokuvat ovat kertoneet toista tarinaa. Molemmat ovat olleet merkattuina samaan valokuvaan, harrasteryhmään tai tapahtumaan.

Yhteiset ystävät

Facebook esittää usein käyttäjäprofiilin ystäväverkoston, jos tämä ei ole erikseen sitä salannut. Ystäväverkosto ei kuitenkaan kerro sitä, että mitkä ystävät ovat kahden käyttäjäprofiilin yhteisiä. Tämän tiedon hankkiminen onnistuu yksinkertaisella URL-tempulla:

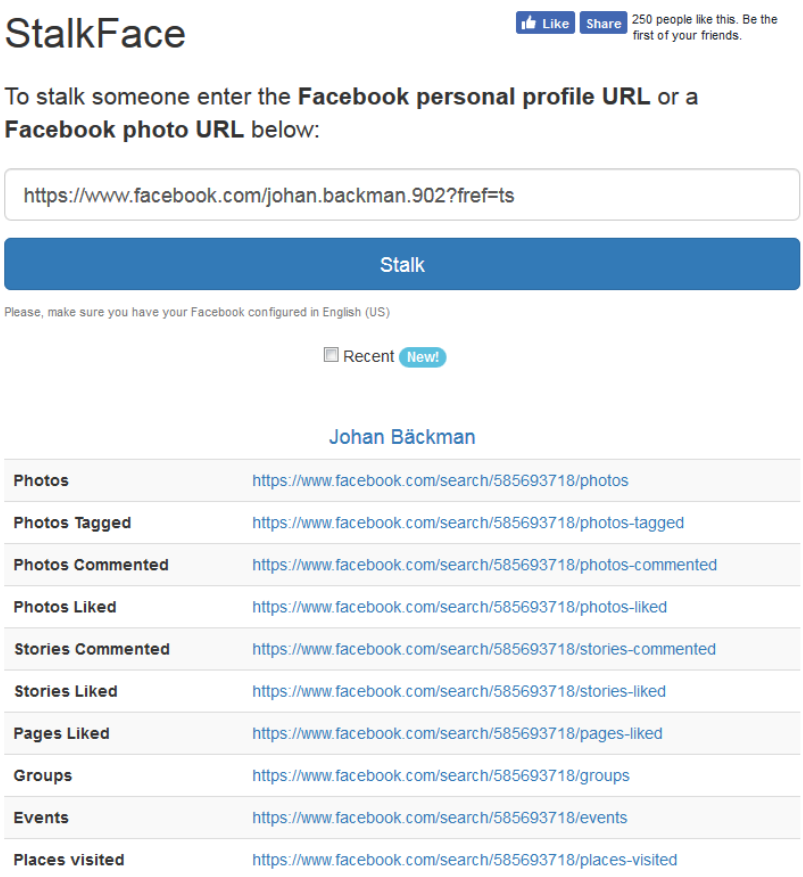
www.facebook.com/friendship/Käyttäjä1/Käyttäjä2

Tämä haku ei kuitenkaan toimi täysin toivotulla tavalla sikäli, että hakutulos esittää vain pari esimerkkiä yhteisistä kavereista, vaikka yhteisiä kavereita saattaa olla toista kymmentä. Muuta ratkaisua kattavamman hakutuloksen saamiseksi minulla ei ole kuin toistaa haku muutamaan kertaan, sillä esimerkit yhteisistä kavereista muuttuvat aina hakua toistettaessa.

Stalk Scan (http://www.stalkscan.com)

Stalk Scanin avulla voit tarkistaa kaiken, mitä Facebook sinusta julkisesti näyttää. Eli käytännössä kaiken sen tiedon minkä juuri saimme ulos käsivälitteisesti Graph Searchilla ja vähän enemmän. Stalk Scan listaa listaa samaan tapaan kuvasi, videosi ja tapahtumasi jos ne on tehty toisen käyttäjän julkiseen päivitykseen, mutta paljon vähemmällä työpanoksella. Astetta syvemmälle pääsee vieläpä tykkäys- ja kommenttitietojen kautta. Kermana kakun päälle voi vielä käydä läpi muun muassa tykätyt sivut, poliittiset puolueet ja uskonnot sekä kaverit ja kaverien kaverit.

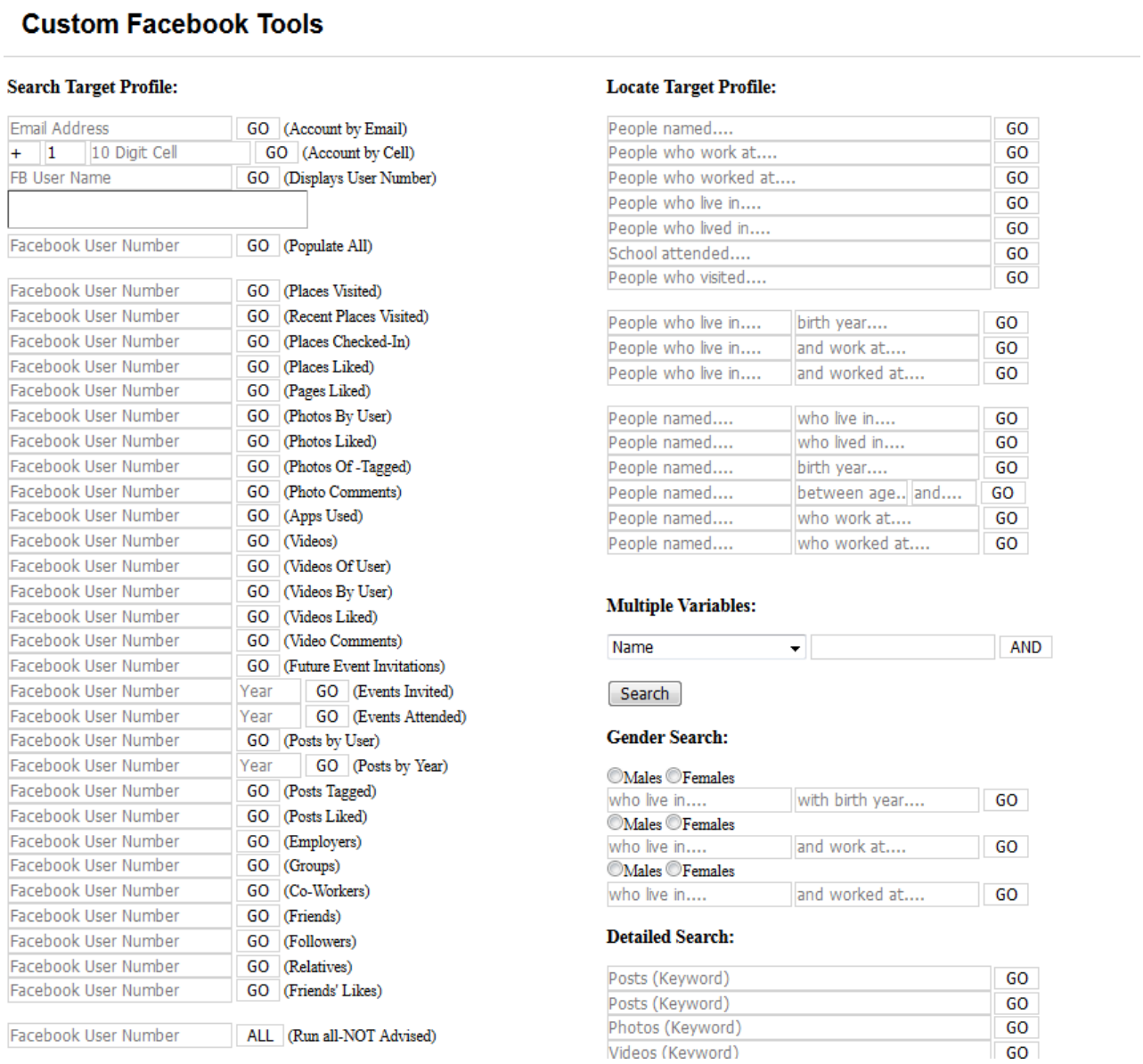
Luultavimmin Facebook haluaa ajaa alas Stalk Scan -palvelun melko nopeasti. Kuten Graph Searchiin, Stalk Scaniin voi syöttää myös kaikkien muidenkin profiilin, jolloin sovellusta voi tietty käyttää muiden Facebook-käyttäjien stalkkaamiseen. Oikeastaan samanlaisesta ohjelmasta on kyseessä myös Stalk Face (https://stalkface.com). Stalk Face on kenties hieman rajatumpi hakutuloksissaan, eikä yhtä pilkottu. Ymmärrettävsti joku voisi kuitenkin pitää Stalk Facea helpompana käyttää (Kuva 1).



Kuva 1: Stalk Face

Facebook OSINT-työkalu (https://inteltechniques.com/osint/facebook.html)

Lopuksi on esiteltävä Facebookin OSINT-työkalu, joka on pitkälti ainoa työkalu, jonka FB-tiedonhakuun tarvitsee. Mikäli on tarve vieläkin edistyneeseempään tiedonhankintaan Facebookista, niin tulemme nopeasti ohjelmointirajapinnan äärelle, joka taas on osa edistyneempien OSINT-dekkareiden keinovalikoimaa. Ohjelmointirajapinnan kanssa operoitaessa on mahdollista ohittaa verkkosivu ja komminikoida suoraan API-rajapinnan kanssa, jolla on suora linkitys servereissä säilytettävään dataan. Inteltechniquesin työkalulla on mahdollista kaikkia edellä kuvaltuja hakua yksinkertaisessa formaatissa. Työkalun vasen puoli mahdollistaa kaikki edellä mainitutu Graph Search-hakutoiminnallisuudet (Kuva 2). Oikeastaan tämä työkalu ei vaadi sen kummempia selittelyjä, koska käyttöliittymä on niin suoraviivainen ja ohjaava.



Kuva 2: Inteltechniquesin Facebook OSINT-tylökalu

4.3 Twitter (www.twitter.com)

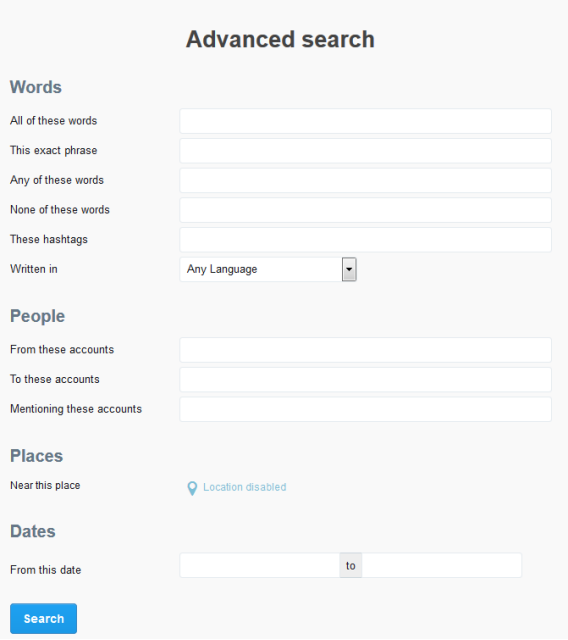
Twitter on mikroblogipalvelu, jonka päivityksen virkerajoitus on rajattu 140 merkkiin. Vuonna 2014 Twitterissä tehtiin 500 miljoonaa päivitystä eli tweettiä päivässä. Twitterin periaate on, että käyttäjä luo profiilin ja twiittaa ajatuksiaan jostakin päivän polttavasta teemasta, sijaintipaikastaan, iltasuunnitelmistaan tai jostakin muusta aiheesta, jonka kokee tärkeäksi. Käyttäjä voi seurata toisia käyttäjiä ja nähdä reaaliaikaisesti mitä seurattavat käyttäjät päivittävät palveluun. Useimmat käyttäjät hallinnoivat tiliään mobiililaitteiden välityksellä, jotka paljastavat päivityksen maantieteellisen sijainnin. Varmennetun tilin merkki Twitterissä on sininen. Twiitit, jotka ovat saaneet eniten tykkäyksiä ja jakoja, suositut puheenaiheet lähialueella tai maailmalla näkyvät eri aggregaateissa ja nousevat uutisvirtaan.

Twitter-Search (www.twitter.com/search)

Tämä on Twitterin virallinen hakukäyttöliittymä. Tulokset voivat olla ylitsevuotavan runsaita, eivätkä kovin kohdennettuja. Tämä hakuominaisuus soveltuu lähinnä ajankohtaisen uutisvirran, juorujen ja trendien löytämiseen Twitteristä.

Twitter Advanced Search (www.twitter.com/search-advanced)

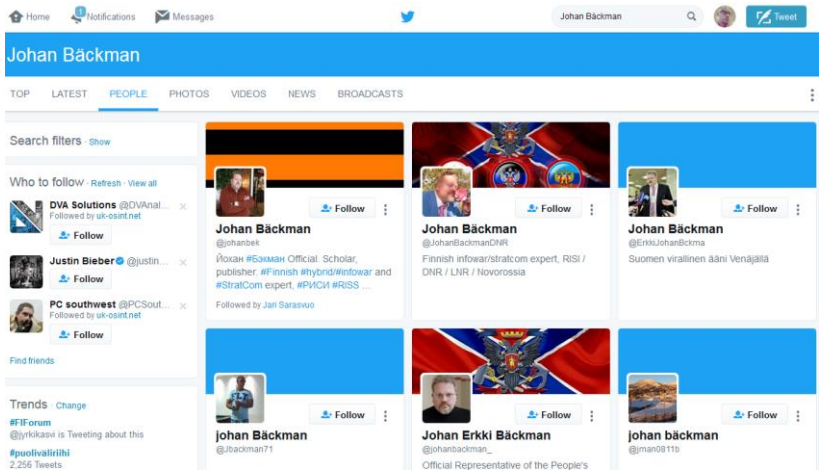
Tämä hakukone mahdollistaa tiettyjen kohdehenkilöiden, avainsanojen ja sijainnin hakemisen. Ongelmana usein, että hakukohde on rajattu koskemaan korkeintaan viimeiseen kymmeneen päivään. Henkilökohtaisten profiilien osalta twiitit ovat nähtävissä niin pitkälle kuin on halukkuutta skrollailla. Edistynyt haku on käyttökelpoinen tapa hakea dataa viime päiviltä, mutta haettavan aiheen tai teeman arkistoituneeseen dataan ei tällä hakuominaisuudella pääse käsiksi. Edistyneen haun hakupaneeli on Googlen ja Facebookin tarkennettuun hakuun verrattavissa oleva helppokäyttöinen moduuli (Kuva 3).



Kuva 3: Twitterin tarkennettu haku

Twitter Person Search (www.twitter.com/#!/who_to_follow)

Kohdehenkilön Twitter-profiilin paikantaminen saattaa olla ongelmallista. Toisin kuin Facebookissa ja Google+-palvelussa useimmat Twitterin käyttäjät eivät käytä oikeaa nimeään profiilissaan. Näin ollen on suositeltavaa käyttää Twitterin ”Who to follow” hakuelementtiä. Alla olevalla näytöllä (Kuva 4) esitellään hakutuloksia nimelle ”Johan Bäckman”. Hakutulokset esitellään visuaalisella ja informatiivisella tavalla, joten kohteen löytäminen hakutulosten joukosta on usein melko vaivatonta.



Kuva 4: Twitter Person Search

Google Cache (www.google.com)

Joskus saattaa tullaa eteen tilanne, missä kohde on twiitannut jotakin, mutta poistanut sen nopeasti tai profiili on saatettu poistaa Twitterin sääntörikkomuksen eli tiwiitin vuoksi. Tämä tieto saattaa olla vaikkapa tutkivan journalistin näkökulmasta hyvin mielenkiintoista. Poistetun tiedon palauttamiseen itsesi nähtäville on olemassa muutamia tapoja, joista esittelen nyt yhden.

Ensin on haettava Googlesta kohteen Twitter-tili. Sen sijaan että klikkaisin tuloksista Twitter-tilin linkkiä, klikkaankin pientä vihreää nuolta, joka osoittaa alaspäin. Siitä avautuu valikko, jossa on kaksi vaihtoehtoa: samankaltaisia ja välimuistissa. Vaihtoehtoista valitaan *Välimuistissa*, minkä jälkeen on mahdollista nähädä kohteen sensuroimaton versio, missä on eritelty myös poistetut twiitit.

Ongelmana tässä hakutekniikassa on, että tällä menetelmällä on mahdollista tarkastella ainoastaan hiljattain poistettuja twiittejä. Useimmiten kiinnostukseni voi liittyä kohteen toimintaan Twitterissä pidemmällä aikavälillä. Taklatakseni tämän ongelman käytän Googlen hakuoperaattoria seuraavalla tavalla:

site: twitter.com/johanbek

Hakuoperaattori ohjeistaa Googlea hakemaan tulokset ainoastaan verkkosivulta site:twitter.com/johanbek. Koska Twitter luo oman yksilöllisen sivunsa jokaiselle twiitille ja jokainen näistä sivuista alkaa osoitteella twitter.com /käyttäjänimi, niin saamme ehkä jopa satoja hakutuloksia. Hakutulokset sisältävät useita poistettuja twiittejä. Nähdäksesi poistettuja Twitter-kommentteja tai twiittejä, on taas kerran hyödynnettävä välimuistia edellä kuvatulla tavalla jokaisen yksittäisen hakutuloksen kohdalla.

Topsy (www.topsy.com)

Topsya käytetään yleisesti sosiaalisen median liikenteenseurantaan aihealueen perusteella. Toinen mielenkiintoinen käyttötarkoitus on palauttaa katseltavaksi aiemmin poistettu Twitter-profiili. Sen sijaan, että Topsyssa haettaisiin kohdetta yksinomaan Twitter-nimen perusteella, haku tehdäänkin URL-osoitteen muodossa kuten aiemmin Facebookin Graph Searchissa:

htttp://twitter/ryback22

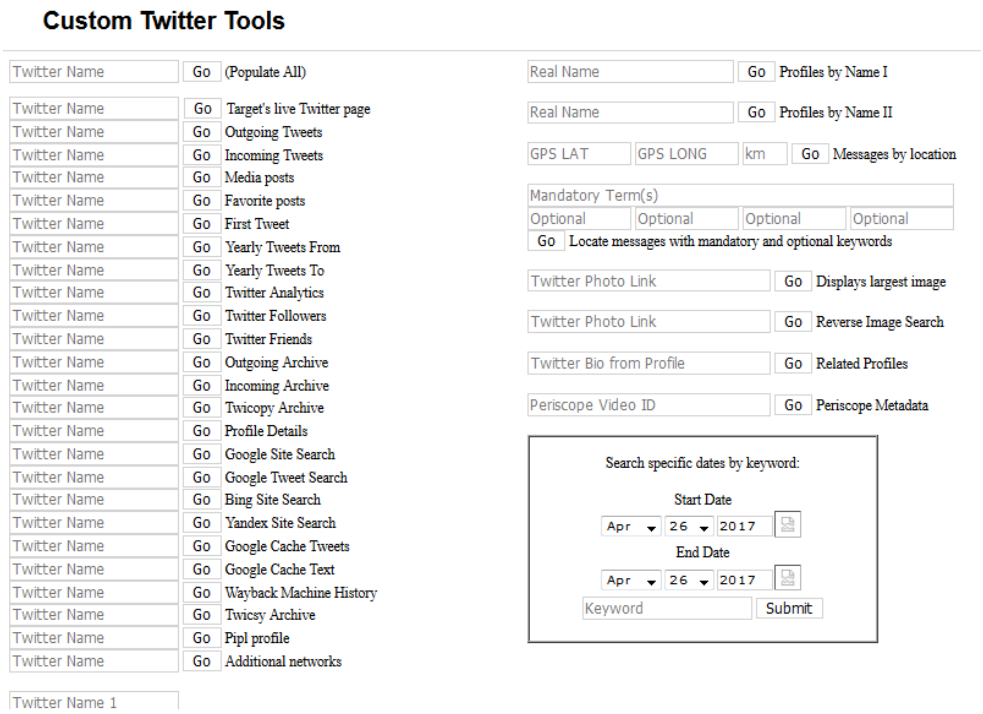
tai tarkempi muoto on itse asiassa

htttp://topsy.com/twitter/ryback22

Tällä metodilla on mahdollista saada välittömästi sellaisia hakutuloksia, jotka muutoin saattaisivat olla tavoittamattomissa. Datan analysointi on helppoa, jos tili on kokonaan poistettu tai se on ylläpitäjän toimesta pantu karenssiin.

Twitter OSINT-työkalu (https://inteltechniques.com/osint/twitter.html)

Henkilökohtaisesti huomaan käyttäväni usein edellä kuvattuja manuaalisia työkaluja, mutta joku saattaisi haluta vältellä samojen osoitteiden jatkuvaa kirjoittamista ja silloin on suositeltava jo aiemmin viitattua Inteltechniques-sivustoa ja siellä sijaitsevaan räätälöityä Twitterin hakutyökalua. Sivusto sisältää sulautetun Javascriptin[34] , joka suorittaa hakuja asetettujen hakuparametrien avulla. Tässä työkalussa kaikki on all-in-one, joten useimmat hakutarpeet Twitteristä voidaan hoitaa tällä hakupaneelilla (Kuva 5).



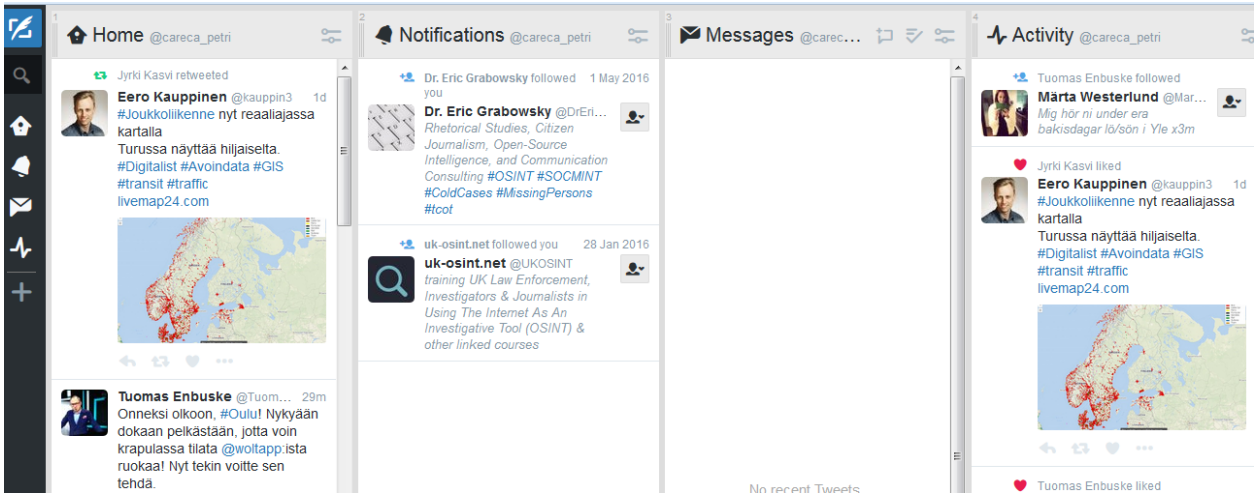
Kuva 5: Inteltechiquesin Twitter-hakupaneeli

[34] JavaScript on alun perin Netscape Communications Corporationin kehittämä pääasiassa Web ympäristössä käytettävä dynaaminen komentosarjakieli. JavaScriptin tärkein sovellus on mahdollisuus lisätä Web-sivuille dynaamista toiminnallisuutta.

Tweet Deck (www.tweetdeck.com)

TweetDeckin omistaa Twitter ja se hyödyntää Twitterin ns. “firehosea” eli “paloletkua”. *Firehousella* tarkoitetaan Twitterin valtavaa datasyöttöä, joka sisältää kaiken palvelussa postattavan datan. Monilla Twitterin palveluilla ei ole pääsyä tähän Twitterin laskimosuoneen ja hakutulokset jäävät rajallisiksi. TweetDeck edellyttää Twitter-tilille kirjautumista, mutta se ei ole sama kuin Twitter-tili. TweetDeckin etu verkkodekkarin näkökulmasta on, että sillä on mahdollista seurata hyvin systemaattisesti ja reaaliajassa sellaisia ihmisiä, teemoja tai hakuparametreja, jotka kulloinkin kiinnostavat. Twitterin komentopöytä on vieläpä helposti modifioitavissa.

TweetDeckillä on jopa mahdollista hyödyntää GeoSearchia eli toiminnallisuutta, jolla on mahdollista hakea sijaintitietoja seuraavalla tavalla: ”geocode:43.430242,-89.736459,1km”. Edellinen hakutermi esittää reaaliaikaista datasyöttöä twiiteistä erikseen määritellystä maantieteellisestä sijainnista kilometrin säteellä. Yleensä pelkkä maantieteellinen sijainti on liian ylimalkainen hakuparamteri, hyödyllisintä on lisätä jokin avainsana hakutulosten filtteröimiseen.



Kuva 6: Twitter Deck eli Twitterin komentopöytä

5. META- JA REFERENSSIDATA

5.1 Mitä meta- ja referenssidata on?

Metatiedolla tarkoitetaan tietoa kuvaavaa tietoa. Metatieto on käsite joka saa aikaan pelkoa ja inhoa, mutta myös suunnatonta innostusta ja sekavuutta. Ohjelmistokehityksessä metatiedot halutaan ottaa käyttöön ja kuvata sillä vaikka mitä, mutta useasti käy ilmi että ensimmäisenä asiana tarvitaan yhteinen ymmärrys siitä, mitä metatieto oikein tarkoittaa. Varsinainen metatiedon sisältö vaihtelee tiedostokohtaisesti, mutta yleisesti ottaen metatieto sisältää muun muassa tiedostotyypin ja tiedostonimen, tiedoston sisällön tai sen ominaisuuksien muuttamiseen liittyviä aikaleimoja sekä tiedoston käyttöoikeuksiin liittyviä määrityksiä.

Kuvatallenteessa metadataa olisivat esimerkiksi tallennusvälineen merkki, värikompositiot, sovellustiedot tms. Kuvien erilaiset tiedostomuodot sisältävät erilaista metadataa. Jotkin formaatit, kuten BMP, PPM, ja PBM sisältävät hyvin vähän muuta tietoa kuin kiinteästi kuvaan liittyviä tietoja. Sen sijaan JPEG sisältää tavallisesti hyvinkin kattavasti erilaista tietoa, kuten kameran merkki ja malli, polttoväli ja aikaleimat. PNG-tiedostot sisältävät yleensä hyvin vähän tietoa ellei kuvaa ole konvertoitu JPEG-muotoon tai muokattu Photoshopilla. Konvertoidut PNG-tiedostot voivat sen sijaan sisältää metadataa tiedostomuodossa.

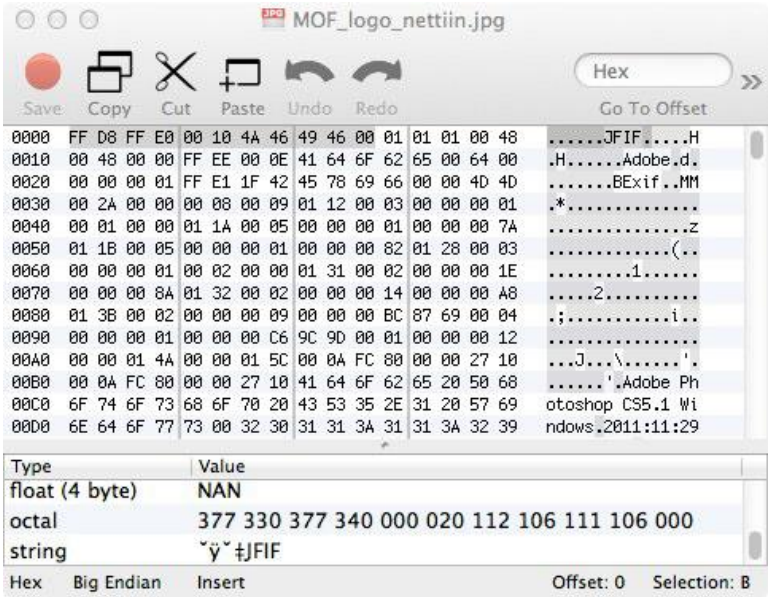
Metatieto kohdistuu aina johonkin tietovarantoon, joita ovat muun muassa dokumenttienhallintajärjestelmän dokumentit tai verkko-oppimisympäristön tietokanta. Tietovarannot sisältävät yleensä suuren määrän informaatio-objekteja ja niiden sisältämää tietoa. Jotta oikean tiedon tai informaatio-objektin löytäminen olisi mahdollista, tarvitaan metatietoa. Metatiedon avulla voidaan tehostaa objekteihin kohdistuvia hakuja, auttaa oikean tiedon löytämisessä ja ymmärtämään sekä tulkitsemaan objektin sisältö ilman itse varsinaiseen sisältöön tutustumista.

Tietokone käsittelee tietoa binääritasolla, jossa tiedon kuvaamiseen käytetään 1- sekä 0-numeroita. Kyseessä on base2-numerojärjestelmä jonka nimi tulee numerojärjestelmässä käytettävissä olevien numeroiden määrästä. Heksadesimaalinumerojärjestelmä on puolestaan base16-numerojärjestelmä. Siinä käytetään numeroita 0-9 ja kirjaimia A-F, joista desimaalilukuina A on 10 ja F on 15. Kahdella heksadesimaaliluvulla muodostetaan yksi tavu, joka koostuu kahdeksasta bitistä eli 1-tai 0-numerosta. (Carrier 2005, 17 – 21.) Heksadesimaalieditori on tietokoneohjelma joka näyttää tiedoston rakenteen sekä bitti että tavutasolla.[35]

Kuvassa 1 on näkymä ”MOF_logo_nettiin.jpg”-tiedoston sisällöstä 0xED-nimisen heksadesimaalieditorin kautta katsottuna. Vasemmassa reunassa pystysuoralla linjalla nähdään heksadesimaaleina kuvattu offset-luku, jolla määritetään sijainti tiedostossa. Ensimmäisellä

[35] Hex Editor Definition 2006

rivillä luku on ”0000”, koska ensimmäinen ”0xFF”-tavu sijaitsee offsetissä 0. Seuraavalla rivillä luku on ”0010”, koska rivin ensimmäinen ”0x00”-tavu sijaitsee offsetissä 16.



Kuva 1: JPEG-kuvatiedoston sisältöä heksadesimaalieditorin kautta katsottuna

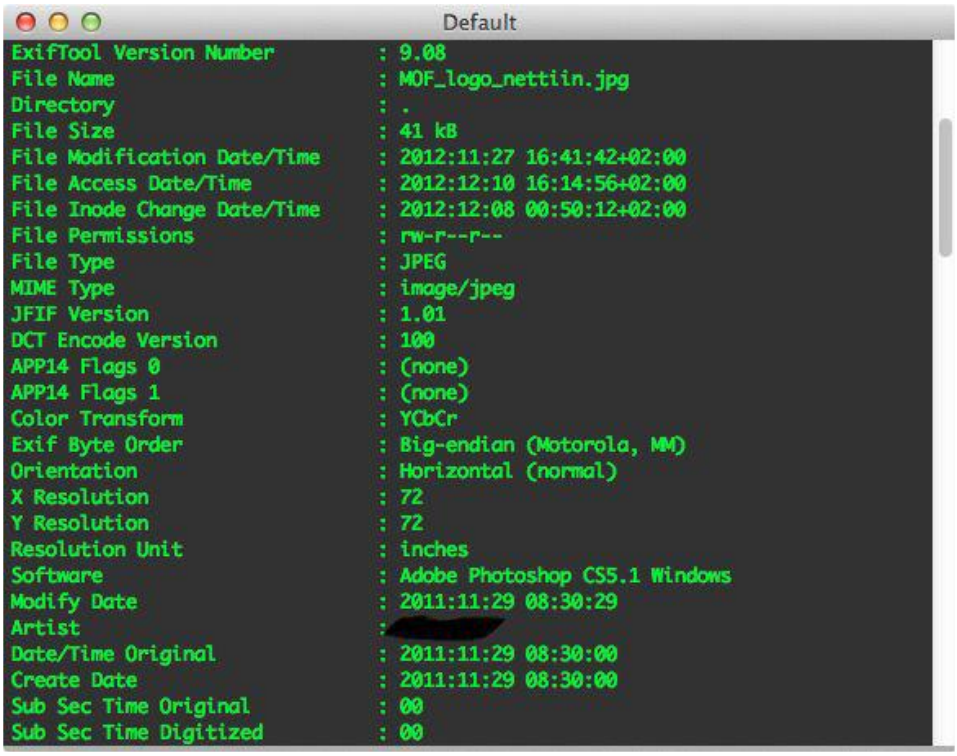
Metadatan louhimiseen on saatavilla lukuisia erilaisia ilmaisia avoimen lähdekoodin tai kaupallisia sovelluksia. Jotkut sovellukset tukevat vain yhtä tiedostotyyppiä (esimerkiksi JPEG), kun taas toiset tukevat useita tiedostomuotoja. ExifTool-niminen ohjelma (www.sno.phy.queensu.ca/~phil/exiftool/) louhii EXIF-metadataa. Lyhenne tulee sanoista *exchangeable image file format* . Toinen metadatan tyyppi on *international press telecommications council* (lyh. IPTC). EXIF-metadata liittyy tekniikkaan, ei itse kuvaan tai sen sisältöön. Lisäksi tätä metadataa ei voida jälkikäteen muokata, mutta se kuitenkin voidaan poistaa kokonaan.

IPTC-metadata on taas muokattavaa itse kuvaan liittyvää metadataa. ITPC-metadata on alun perin kehitetty uutistoimistojen käyttöön helpottamaan kuvien arkistointia ja julkaisua. Monet kuvankäsittelyohjelmat tukevat ITPC-metadataa.[36] Tiedonhankinnan kannalta EXIF on mielenkiintoisempi, koska sitä ei voi jälkikäteen editoida. Valokuviin EXIF-metadata syntyy, jos kamerassa on standardoidun metadatan luonnin mahdollistava ominaisuus otettu käyttöön.

Esimerkkinä tutustumme ”MOF_logo_nettiin.jpg”-tiedoston sisältämään metatietoon ExifTool-ohjelman avulla. Ohjelma toimii tutkimalla tiedoston rakenteisiin tallennettua metatietoa.

[36] ITPC 2010 ja Metadata working group 2015

Merkintöjen perusteella ohjelma tulostaa näytölle yksinkertaisen näkymän tiedoston sisältämästä metatiedosta (Kuva 2).



Kuva 2: Tiedoston metadataa kuvattuna ExifTool-ohjelmalla

Kuvasta 2 nähdään, että ”MOF_logo_nettiin.jpg”-tiedosto sisältää runsaasti siihen tallennettua metatietoa. Huomattavaa on muun muassa alkuperäisen tiedoston luontiajankohta sekä siihen käytetyn ohjelman nimi ja versionumero. Tiedostoon on tallennettu myös sitä muokanneen käyttäjän nimi, joka on tummennettu pois kuvasta.

5.2 Foca

FOCA tarkoittaa espanjaksi hyljettä, mutta se tarkoittaa myös espanjalaislähtöistä ohjelmistosovellusta, joka louhii dokumenteista metadataa. FOCA nimen kerrotaan monessa lähteessä olevan lyhenne sanoista *Fingerprinting Organizations with Collected Archives*. Chema Alonso, joka on yksi FOCA:n kehittäjistä, on kuitenkin sanonut, että nimi on itse asiassa vitsi: kirjainyhdistelmällä viitataan FOCA:n Francisco Oca -nimiseen koodaajaan.

FOCA:n tarkoituksena oli alun perin v. 2009 siivota erään turvallisuusyrityksen dokumenttien metadataa. Metadatalla tarkoitetaan informaatiota, joka on tallentunut dokumentin rakenteisiin: tekijä, organisaatio, luomis- ja muokkauspäivä etc. Metadataa voi vapaasti muokata. Useimmille meille on tuttua, että kadotamme jonkun dokumentin kovalevylle. Kenties tallennusvaiheessa emme ole olleet riittävän tarkkana tallennuskansiosta tai muusta syystä. Jotta löytäisimme tuon dokumentin helpommin, luotiin metadata.

Sen sijaan dokumentin referenssitietoa ei voi paljain silmin nähdä tai muokata, vaan se toimii ikään kuin teknisenä lokituksena, joka rekisteröi sellaisia asioita kuten käytetty template eli mallinne, tulostukseen käytetty printteri tai dokumenttiin liitetyt verkkolinkit. Edellisen lisäksi FOCA pystyy louhimaan dokumentista myös tietoa, joka saattaa olla kytkeytynyt siihen erehdyksessä. Esimerkiksi liittäessäsi Internetistä kopioimasi valokuvan dokumenttiisi, se tuo mukanaan oman viitetietonsa ja metadatansa.

Metadatan siivoaminen dokumenteista ei ole erityisen vakiintunutta edes julkishallinnon alalla ja sen julkistetuista dokumenteista on mahdollista saada FOCA:n kaltaisella louhintaohjelmalla paljonkin mielenkiintoista tietoa, kuten sen että kuka mitäkin dokumenttia on käsitellyt ja missä. On olemassa esimerkkejä siitä, miten dokumentin käsittelytiedoista on käynyt ilmi arkaluonteisia tietoja, kuten vuonna 2003 jolloin Tony Blair kertoi julkisesti että Irakin sotaa käsittelevää dokumenttia ei olut muokattu tai manipuloitu. Dokumentin metadata kertoi kokonaan toista tarinaa.[37]

Metadataa tutkittaessa tulee huomioida, että kaikki tiedostotyypit eivät välttämättä tue metadatan lisäämistä. Samoin tulee huomata se, että metadatan lisääminen, poistaminen tai muokkaaminen eivät vaikuta millään tavalla tiedoston tarkoitettuun toimintaan.

Mitä kaikkea sitten on mahdollista FOCA:lla louhia?

- Käyttäjät: dokumentin luojat, editoijat
- Käyttöjärjestelmät

37 http://dfir.com.br/wp-content/uploads/2014/02/blair.htm

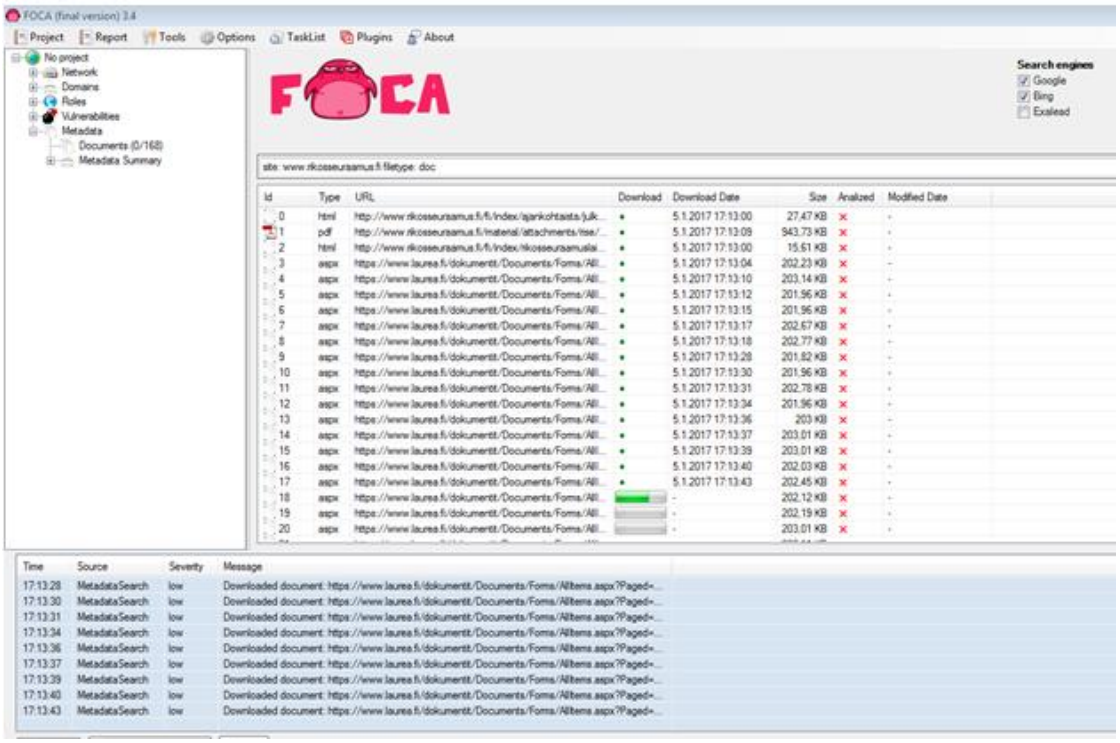
- Paikalliset- ja etäprintterit
- Dokumentin digitaalinen jalanjälki
- Jaetut tiedostot
- Sisäiset palvelimet: IP-osoite, domain-nimi etc.
- Tietotietokannat
- Laiteinformaatio: mobiililaitteet, kamerat etc.
- Henkilöinformaatio
- Käyttöhistoria
- Ohjelmistopäivitykset

Seuraavaksi teemme lyhyen testin, missä tutkimme ohjelman käytettävyyttä peilaten sitä verkkosivulle tallennettujen dokumenttien meta- ja referenssidataan.

5.3 Testi

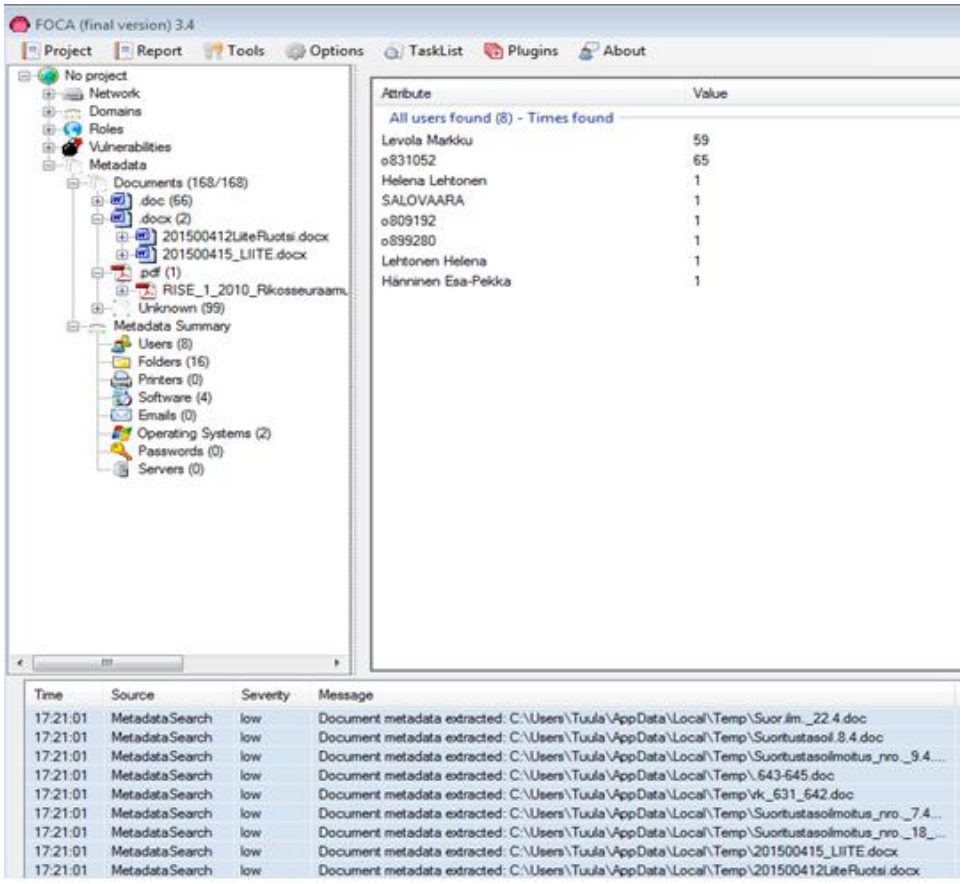
Haluan tutkia Rikosseuraamuslaitoksen verkkosivun doc. -formaatissa olevien dokumenttien metadataa.

Käynnistettyäni ohjelman, klikkaan vasemmassa sivuvalikossa Metadata-painiketta. Luon kustomoidun haun käyttämällä Googlen hakuoperaattoreita: ”site: www.rikosseuraamus.fi filetype: doc”. Painan ”Search” ja Foca hakee kaikki verkkosivustolle tallennetut dokumentit, joita on kaikkiaan 167 kappaletta. Kun olen skannannut verkkosivun lomakkeet, valitsen ”Download all”. Tähän saattaa kulua muutamia minuutteja riippuen dokumenttien lukumäärästä (kuva 3).



Kuva 3: FOCA skannaa verkkosivun lomakkeita

Voin valita listauksesta vain yhden dokumentin, jonka metadataa louhia, mutta tällä kertaa valitsen kaikki dokumentit valitsemalla *Extract all metadata*. Louhinta tuottaa kelpo tuloksen tunnistaen muun muassa kahdeksan eri käyttäjää/käyttäjätunnusta. Kansioita löytyi kaikkiaan 16 kappaletta. Verkkosivujen dokumenttien luontiin/editointiin oli käytetty neljää erilaista ohjelmaa sekä kahta käyttöjärjestelmää : Windows XP ja Windows 7 (Kuva 4).



Kuva 4: FOCALLA tehdyn verkkosivun dokumenttien louhinnan lopputulos

6. KUVAHAKU

6.1 Kuvahaku ja käänteinen kuvahaku

Jotakin OSINT:n vaikutusmahdollisuuksista kertoo sen osuus Ukrainan ja Syyrian sisällissodissa. Propagandan ja vastapropagandan ristipaineessa avoimet lähteet ovat tuottaneet realistista tilannekuvaa sekä paljastaneet propagandatarkoitukseen luotua tarkoitushakuista ”uutisointia”. Järjestö nimeltä Bellingcat[38] on vihkitytynyt tähän tarkoitukseen. Sosiaalisen median sivustoille venäläisten sotilaiden sosiaaliseen mediaan lataamat selfiet alkoivat ensi kertaa Ukrainassa raottaa sumuverhoa Venäjän läsnäolosta Krimin niemimaalla samaan aikaan kun Euroopan Unionin oma tiedusteluorganisaatio empi vahvistaa Venäjän aktiivisuutta. Euroopan Unionin sotilasesikunnan tiedustelupäällikkö Georgij Alafuzoff on myöntänyt itsekin, että Ukrainassa on eletty voimakasta informaatiosodan aikaa ja informaatiosodan sumu vaikeutti tilanteen arviointia kaikilla tasoilla.[39] Varsinaisten sotatoimien rinnalla käytävä informaatiosota Krimillä ja Donetskissa ovat osoitus siitä, että tiedusteluorganisaatiot eivät voi enää luottaa samalla tavalla perinteisiin tiedustelulähteisiin. Ei ole mitenkään suuri yllätys, että OSINT on tärkeä osa NATO:n tulevaisuuden visiota.

Visuaalisuus ja valokuvat ovat tunnetusti voimakkaita tehostekeinoja disinformaation levitykseen. Myös Ukrainassa informaatiosodassa on käytetty laajalti internetistä ladattuja valokuvia, joita on hyödynnetty joko sellaisenaan harhaan johtavassa kontekstissa taikka käsiteltyinä. Hyödyntämällä Google Earthin tai muiden kaupallisten satelliittien satelliittikuvia on mahdollista saada arvokasta tietoa kuvan autenttisuudesta suhteessa esitettyyn kontekstiin. Ne eivät ole kuitenkaan ainoita tietolähteitä, jotka auttavat selvittämään, missä kuva tai video on otettu.

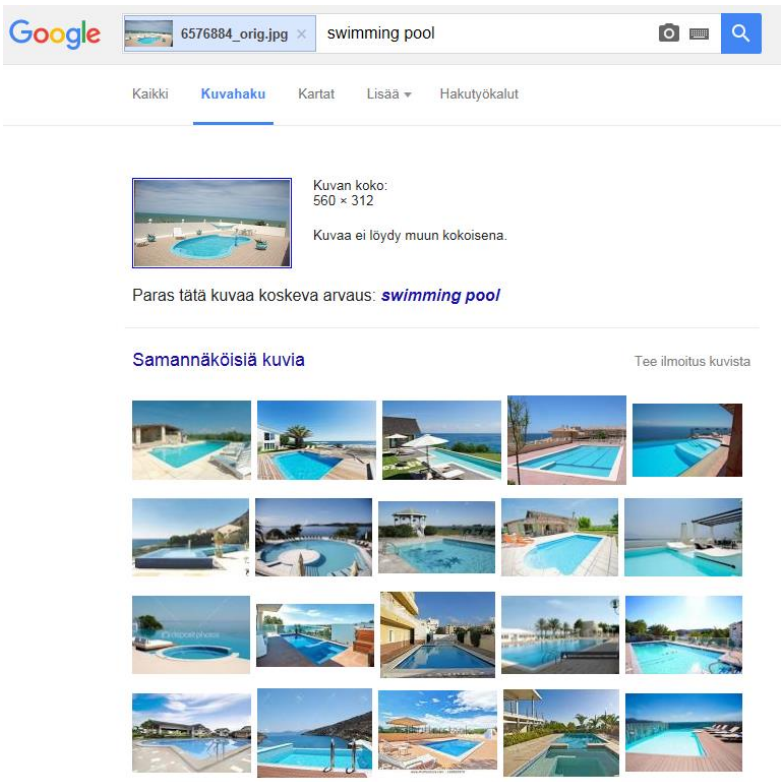
Erityisesti Venäjällä kuvamanipulaatiot tuntuvat olevan erityisen suosittuja. Verkossa on lukuisia käänteisiä CBIR-tekniikkaan (content-based image retrieval) pohjautuvia kuvahakuja tekeviä hakukoneita, joihin kuvia syöttämällä voi päästä alkuperäisen kuvan alkulähteelle. Tällaisia ovat *TinEye* tai Venäjän oman hakukoneen Yandexin *Sibir*. Muun muassa näiden hakukoneiden avulla voidaan helposti todistaa kuvan olleen irrotettu asiayhteydestään, minkä tietävää varmasti myös propagandakoneisto. Taustalla kumminkin lienee se, että vaikka kuvat myöhemmin osoitettaisiinkin päälle liimatuiksi, niin niistä saatava hyöty ylittää kuvaväärennöksen todistusarvon (Kuva 1).

38 Bellingcat on työttömän bloggarin Eliot Higginsin perustama tutkivan kansalaisjournalismin ryhmä, jonka suomalaisjäsen on Veli-Pekka Kivimäki. Järjestö tekee vapaaehtoistyötä mm. sotaprapogandan paljastamiseksi avoimia lähteitä hyväksikäyttämällä.

39 http://www.verkkomedia.org/news.asp?mode=4&id=10412

Kuva 1: Asiayhteydestään irrotettu valokuva, jota on käytetty sotapropagandan tarkoituksiin

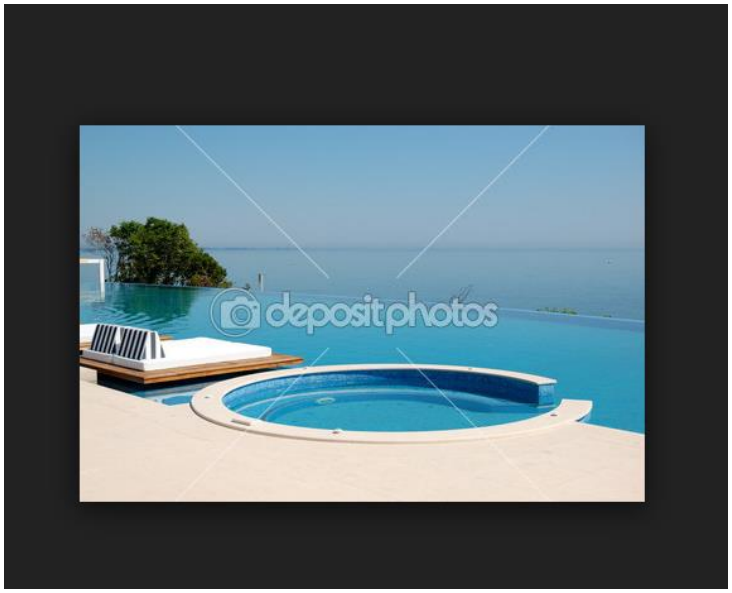
Käänteisellä kuvahaulla haetaan verkosta sivustoja, joissa haettava kuva joko esiintyy tai on sen kanssa samankaltainen. Helpoin tapa soveltaa käänteistä kuvahakua on käyttää vuonna 2011 julkaistua Googlen kuvahakua. Googlen etusivun oikeassa yläkulmassa lukee *kuvahaku*, jota klikkaamalla pääsee käänteisen kuvahaun hakukenttään. Kuvahaun voi käynnistää kahdella eri tapaa: joko kopioimalla kuvan URL-osoitteen tai lataamalla kuvan omalta tietokoneelta. Tarkkaan ottaen on vielä kolmaskin käynnistystapa, mikäli käytössäsi on Chrome tai Firefoxin versio 4 tai sitä uudempi versio. Näissä selaimissa on mahdollista myös vetää kuvia tietokoneeltasi hakukenttään. Esimerkkinä syötän kuvahaun hakukenttään tietokoneeltani kuvan kesälomakuvastani, jossa esiintyy uima-allas (Kuva 2).



Kuva 2: Hakutulos kuvan kanssa

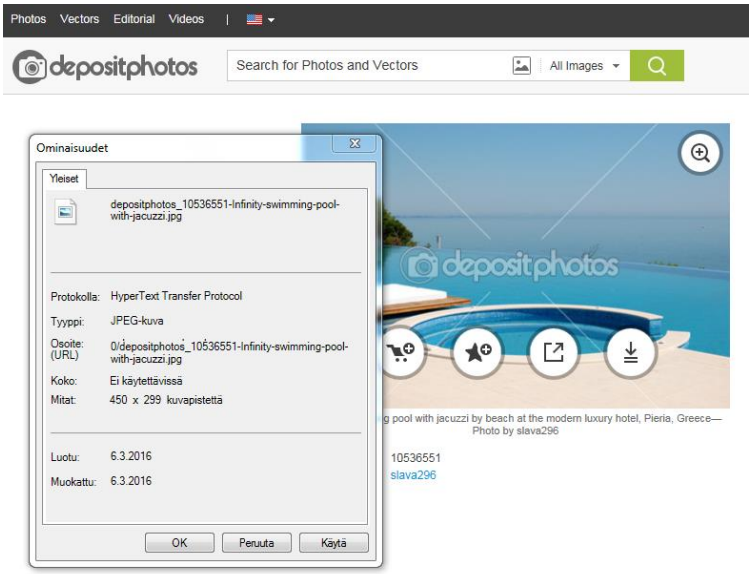
Sen lisäksi, että Google hakee tietokannoista samoja tai samankaltaisia kuvia, se pyrkii myös tunnistamaan kuvan keskeisen aiheen tai teeman. Googlen paras arvaus onkin tässä tapauksessa aivan oikein: *swimming pool*. Sivuhuomautuksena mainittakoon, että syötettyäni oman valokuvani kuvahakuun, niin Googlen arvaus kuvan aihepiiristä oli *eyewear*. Jotkut käänteisen kuvahaun hakusivustot ovat yrittäneet myös tunnistaa valokuvan henkilön sukupuolta ja ikää. Tällaiset hakusovellukset olivat joskus maksullisia, mutta nyttemmin niistä on saatavissa myös ilmaisversioita, joista yksi esimerkki löytyy alla luetelluista sivuista.

Vertailun vuoksi suoritan kuvahaun kuvan URL-osoitteen perusteella. Tällä kertaa haluan URL-osoitteen eräästä toisesta verkosta löytämästäni uima-altaan kuvasta (Kuva 3).



Kuva 3: Haettava kuva

Aivan aluksi navigoin sille verkkosivulle, jolla kyseinen valokuva on klikkaamalla kuvaa. Ei riitä, että kopioin sen työpöydälleni, vaan minun tulee löytää myös kuvan URL-osoite, joka tapahtuu siten, että klikkaan valokuvaa hiiren oikealla näppäimellä ja valitsen valikosta *Ominaisuudet*. Seuraavaksi avautuu ponnahdusikkuna, josta on kopioitavissa kuvan URL-osoite. Lopuksi syötän URL-osoitteen Googlen hakukenttään ja haen kuvaa Googlen kuvatietokannasta (Kuva 4).



Kuva 4: URL-osoite kopioitavissa Ominaisuudet-ponnahdusikkunassa.

Mihin käänteistä kuvahakua sitten voidaan sitten käyttää? Käänteinen kuvahaku helpottaa monissa arkipäiväisissä asioissa, kuten vaikkapa löydettyäsi netistä herkullisen ruokakuvan, mutta et sen reseptiä, niin voit syöttää ruokakuvan hakukenttään löytääksesi kuvan *sekä* reseptin. Samalla periaatteella voit toimia kiinnostavan vaatteen tai huonekalun osalta. Käänteisen kuvahaun avulla on mahdollista myös selvittää onko kuviasi käytetty ilman lupaa muilla sivustoilla. Monet nuoret naiset ovat todenneet sosiaalisen median profiilikuviaan käytetyn myös muissa yhteyksissä verkossa.

Open source-tiedonhankinnan näkökulmasta kaksi tärkeintä käänteisen kuvahaun tarkoitusta ovat kuvan alkuperäisen lähteen löytäminen sekä valokuvassa esiintyvän henkilön henkilöllisyyden selvittäminen. Kuvan alkuperäisen lähteen löytäminen on oivallinen verkkotyökalu esimerkiksi sotapropagandan paljastamisessa. Sotapropagandassa tai vaikkapa luonnonkatastrofeja kuvaavissa tunteellisissa uutiskuvissa voidaan kuvahuijaukset jakaa kahteen kategoriaan:

1) kuvat, jotka on oikeasti otettu toisesta paikasta ja/tai toisena ajankohtana

2) ja photoshopatut eli väärennetyt kuvat.

Jos kuva on jo viraali eli julkaistu monilla eri sivustoilla, alkuperäistä voi olla vaikea jäljittää. Tuloksia tulee liikaa, ja ensimmäinen julkaisu hukkuu niiden sekaan. Tällöin kuvahakuun voi lisätä aikarajauksen tai tiettyjä hakusanoja. Jos taas on syytä epäillä, että kuvaa on manipuloitu, niin tuolloin on suositeltavaa käyttää Fotoforensics-työkalua (www.fotoforensics.com), joka kykenee tunnistamaan kuvasta manipuloinnin jälkiä.

Käänteisessä kuvahaussa on toki myös ongelmia. Ensinnäkin vain yhtä kuvaa voi hakea kerrallaan. Esimerkiksi kaikkien blogikuvien läpikäymiseen menisi paljon aikaa. Työtä voisi helpottaa hakemalla vain suosituimpien päivitysten kuvat, koska niitä ihmiset todennäköisimmin kopioivat. Toiseksi Google tallentaa kaikki kuvat, joita palveluun on ladattu. Omien sanojensa mukaan Google käyttää kuvia tuotteidensa ja palveluidensa parantamiseen (tämä on Googlen perustelu kaikille keksimilleen kyseenalaisille Google-ominaisuuksille).

6.2 Kuvahaun verkkosovelluksia

Googlen käänteinen kuvahaku on vain yksi kyseistä palvelua tarjoavista tahoista. Venäläisellä Yandexilla, amerikkalaisella Bingillä ja kiinalaisella Baidulla on myös erittäin kehittyneet kuvahaun ominaisuudet. Seuraavaksi käyn lyhyesti läpi viisi mielestäni kiinnostavinta – ei ehkä parasta – kuvahaun hakukonetta Googlen kuvahaun lisäksi:

TinEye (www.tineye.com)

TinEye tulokset fokusoituvat kuvaduplikaatin etsimiseen, mikä tietysti näkyy myös hakutulosten määrässä. Hakutulosten määrää rajaa myös Googlea selvästi vähäisempi indeksoidun kuva-aineiston osuus, joka TinEye-sovelluksella liikkuu kahden miljardin tietämillä. TinEyellä on kuitenkin yksi etu verrattuna Googleen: sillä on API. Tämä on syy miksi yritykset kuten Ebay ja Istockphoto ovat integroineet ohjelmistonsa tähän sovellukseen.

PicTriev (http://www.pictriev.com)

Tämä sovellus suorittaa kuvahakuja ainoastaan ihmisten kasvoista, ei esineistä tai maisemasta. Verrattuna TineEyeen ja Googlen kuvahakuun, PicTriev pyrkii automaattisesti tunnistamaan valokuvassa olevan henkilön iän ja sukupuolen. PicTriev ehdottaa myös valokuvamallin nimeä, mikäli kyseessä on julkisuuden henkilö. PicTriev on toki myös käyttökelpoinen ei-julkisuudessa olevien henkilöiden osalta, mutta varsin rajatusti. OSINT:n kannalta tämä haku ei ole kovin hyödyllinen, mutta yhtä kaikki kiinnostava sillä se tuo esiin käänteisen kuvahaun eri mahdollisuuksia.

Image Manipulation (www.fotoforensiscs.com)

Usein kuvamanipulaatio on ilmeistä, eikä edellytä tarkempaa analyysiä. Tekijänä saattaa olla kuitenkin erityisen kyvykäs photoshoppaaja, jonka editoinnin tuloksia on maallikolle vaikeata huomata visuaalisesti. Kuvan lataamisen jälkeen *Fotoforensics* esittää kuvan siinä muodossa kuin se on ladattu hakukenttään. Alkuperäisen kuvan alla on tummennettu versio valokuvasta, jossa manipuloidut osuudet näkyvät korostetusti vaaleampina alueina. Huomioitava seikka on taas, että ladatut kuvat linkittyvät osaksi verkkosivun omaa kuvatietokantaa. Vaikka kuvien URL-osoite saattaisikin olla hankalasti saatavissa, se saattaa silti luoda turvallisuusriskin joillekin arkaluonteisille kuville.

Izituru (www.izituru.com)

Edellinen sovellus pystyy tunnistamaan kaikki ne kuvan kohdat, joita on saatettu manipuloida. Mikäli kuvan editointi osuus ei ole ilmeistä, kuvaa voi analysoida myös Iziturussa. Izituru kykenee tunnistamaan, mikäli kuva on esimerkiksi tallennettu sen sijaan, että se olisi kopioitu. Tämä jo pelkästään saattaa antaa aiheen kyseenalaistaa valokuvan autenttisuuden. Izituru-sovellus myös itse tuo selväsanaisesti ilmi mahdollisen epäilyn kuvan epäluotettavuudesta.

Sibir (www.yandex.com/images)

Venäjän suosituimman hakukoneen Yandexin Sibir tarjoaa hyvän vaihtoehdon Googlen kuvahaulle. Sibirin visuaalisen tunnistamisen tekniikka on vähintäänkin Googlen tasoa. Hakukoneina Google ja Yandex ovat kehittyneet rinnakkain, joten erot eivät ole suuria. Sibirin etuna muihin kuvahakuihin on sen venäläisyys. Mikäli kuvalla on jonkinlaisia liittymäkohtia itänaapuriimme, niin on ehdottomasti ajettava kuva myöskin Sirbirin käänteisen kuvahaun läpi.

7. SIJAINTITIEDON LOUHIMINEN

7.1 Mihin sijantitietoa voidaan tarvita?

Loppukeväästä 2016 moni suomalainen Windows-käyttäjä yllätettiin perinpohjaisesti: Windows 10 päivittyi koneisiin yllättäen ja äkkiä arvaamatta. Monen kohdalla päivitys ei mennyt kuin Strömsössä: sähköpostitilit katoilivat tai lakkasivat kokonaan toimimasta, jotkut koneet menivät takalukkoon. Microsoftin käyttätuki ruuhkautui hetkessä. Jälkikäteen ryhdyttäessä selvittämään sotkua, Microsoft muistutti ilmoittaneensa päivtyksestä englanninkielisessä blogissaan (!)[40] Kohun alle hautautui kulmakarvoja nostattava Windows 10:n uudistus: ohjelmisto kerää jatkossa käyttäjästä yhä enemmän dataa, lataa wifi-salasanoja ystäville ja tuttaville sekä paikantaa käyttäjän sijainnin.[41] Paikannusmekanismit voidaan toki kytkeä pois päältä, mutta kuinka moni peruskäyttäjä oikeasti ymmärtää tai osaa tehdä niin? Vaikka Microsoft varmasti muuta väittää, niin koko päivitysoperaatio antaa vaikutelman lailliselta tietoturvamurrolta, missä hyväksikäytetään tavallisen tietokoneen käyttäjän harjaantumattomuutta.

Samanlaista harjaantumattomuutta on käytetty hyväksi sekä hyvään että pahaan kuten seuraavassa kahdessa esimerkissä tuon esille. Esimerkeissä yhteistä on juuri sijaintitietojen merkitys.

Esimerkki 1

> Harvardin yliopiston opiskelijat Paul Lisker ja Michael Rose havaitsivat monien verkossa toimivien huumekauppiaiden syyllistyneen amatöörimäiseen tietoturvamokaan julkaistessaan valokuvia tuotteistaan myynti-ilmoitustensa ohessa. Opiskelijat julkaisivat aiheesta kattavan raportin, joka nosti esille mielenkiintoisia huomioita.[42] Osassa huumediilerien darknetissa olevien myynti-ilmoitusten valokuvissa oli mukana digikameran tai älypuhelimen tallentama tarkka sijaintitieto.

Suurin osa älypuhelimista tallentaa gps-paikantimensa avulla valokuviin karttakoordinaatit kuvanottopaikasta. Paikannus tapahtuu useimmiten vain muutaman metrin tarkkuudella. Enemmistö diilereistä näytti poistavan paikkatiedot kuvistaan, mutta eivät läheskään kaikki. Tutkijat kaivoivat esiin 229 valokuvaa, joiden sijainti esitetään vuorovaikutteisella kartalla. Kolme kuvaa – ilmeisesti samalta myyjältä – on otettu Helsingissä.

Kartta näytti kuvien sijainnin vain noin 1,6 kilometrin tarkkuudella. Parivaljakko vakuutti, että nämä ovat erittäin tarkkoja sijaintitietoja, joiden avulla kuvan ottopaikan voi jäljittää aina yksittäiseen taloon saakka. Suuret sosiaalisen median palvelut, kuten Facebook ja Instagram,

40 http://www.iltasanomat.fi/digitoday/art-2000001185316.html

41 http://www.zdnet.com/article/want-to-limit-windows-10-tracking-there-is-an-app-for-that/

42 https://medium.com/@roselisker/illuminating-the-dark-web-d088a9c80240#.65xn63oar

poistavat automaattisesti paikkatiedot kuvista. Mutta monet vanhemmat foorumiohjelmistot eivät yhäkään näin tee. Valokuvan metadataan automaattisesti tallentuva sijaintitieto on kuitenkin vain yksi tapa linkittää tapahtuman maantieteellisiä koordinaatteja. Yleisempää on tänä päivän kytkeä vapaa-ehtoisesti oma sijaintitietonsa twiittiin, Facebook-päivitykseen tai Periscope-videoon.

Esimerkki 2

> Vuonna 2007 neljä Yhdysvaltalaista AH-64 ”Apache” helikopteria tuhottiin Irakissa käyttäen älypuhelimen kuvien Geotag-ominaisuutta. Helikoptereissa matkustaneet sotilaat ottivat itsestään ja helikoptereista kuvia asematason ja hallien alueella. Kuvat ladattiin sosiaalisen median sovellukseen, jonka metatiedostoja tutkimalla vastustaja paikansi helikoptereiden tarkan sijainnin lentotukikohdan sisällä. Tietoa käyttäen vastustaja suoritti tuli-iskun heittimillä tuhoten helikopterit.[43]

Esimerkkitapaus paljastaa suuren haavoittuvuuden älypuhelimien kautta jaettavasta mediasta. Tätä kautta videot, kuvat ja muut jaetut tiedostot voivat sisältää paikkatietoa metatiedon sisällä. Käyttäjällä on mahdollisuus kontrolloida gps-signaalin käyttöä tiettyyn pisteeseen asti, mutta jotkut sovellukset voivat käyttää ggps-tietoa ilman käyttäjän lupaa tai tämän tiedostamatta

Google on voittanut ”mitä” kysymyksen jo aikoja sitten voimakkailla algoritmeillaan. Facebook voitti ”kuka” kilpailun: se saattaa tuntea sinut ja sosiaalisen verkostosi paremmin kuin kukaan ystävistäsi. Sen sijaan ”missä” kysymyksen osalta tilanne on edelleen avoin ja kilpailu käy kuumana nykyisten ohjemistojättiläisten sekä uusien start-up-yritysten kesken. Aiemmin teknologia oli laahannut paikannustiedon suhteen perässä, mutta Mooren lain nähtyä taas muutaman uuden kevään, paikannustiedon louhiminen on erittäin kilpailtu ala.

”Missä” voidaan selvittää muutamilla eri tekniikoilla: puhelimesi gps-antennilla, rajaamalla sijaintisi kolmen tukiaseman muodostaman kolmionmallisen alueen avulla, sekä sen perusteella mihin wifi-verkkoon kytkeytyy. Mutta kuten jo aivan alussa johdattelin, paikannustietoa lisätään koko ajan enenevässä määrin verkon tietoliikenteeseen ja datan metatietoihin. Joidenkin sovellusten osalta pyyntö sijaintitietoihisi on ymmärrettävää ja loogista kuten Google Mapsin tai gps-navigointi työkaluihin, mutta suurimmalta osin sijaintitietojesi saaminen on sovellusten kehittäjille vain yksi tapa myydä tietojasi paremmalla hinnalla. Lähes poikkeuksetta käyttäjälle ei ole mitään etua jakaa sovellukselle sijaintitietoaan (Kuva 1).

Päivityksesi Facebookissa, twiittauksesi ja hakusi Yelpissa käyttävät kaikki paikannustietojasi. Lisäksi koko ajan viriää uusia lbs-paikannukseen (location-based service) eli verkkoperustaiseen paikannukseen keskittyviä start-up-yrityksiä, jotka halkuavat sisällyttää paikkatietosi kaikkeen kiinteistömarkkkinoinnista Tinderin ja Grindrerin kaltaisiin deittipalveluihin. Mainostoimistot eivät ole kiinnostuneet yksinomaan siitä, että missä sillä hetkellä olet, vaan myös siitä missä olit

43 http://defensetech.org/2012/03/15/insurgents-used-cell-phone-geotags-to-destroy-ah-64s-in-iraq/

eilen ja missä olet huomenna. Paikannustietosi kertovat yllättävän paljon markkinointikoneistoille. Esimerkki lbs-paikannuksen hyödyntämisestä voisi olla sellainen, missä nainen menee älypuhelimineen gynekologin vastaanotolle ja kiinnostava lokaatio rekisteröityy mobiilimarkkinoinnin ekosysteemiin. Mutta kun sama nainen kolme viikkoa myöhemmin menee BR-lelukauppaan, datan syvempi merkitys mahdollisesti avautuu.

MIT:n ja Oxfordin yliopiston tutkijat ovat osoittaneet, että vain kahdeksan twiitin perusteella on mahdollista selvittää, missä kohde asuu ja käy töissä.[44] Vaikka Twitterin paikannuspalvelu on oletusarvoisesti pois päältä, monet käyttäjät aktivoivat palvelun. Tutkijat käyttivät todellisten Twitter-käyttäjien twiittejä Bostonin alueella, jotka olivat suostuneet antamaan tietonsa tutkimuskäyttöön sekä vahvistivat samalla kotinsa ja työpaikkansa osoitteet, työmatkareittinsä ja lomakohteensa, joista ne olivat twiitanneet.

Twiittien aika- ja paikkatiedot esiteltiin 45 henkilön tutkijaryhmälle, joita sitten pyydettiin päätelemään, twiittien perusteella käyttäjien kotien, työpaikkojen ja vapaa-ajankohteiden sijainnit. Tehtävä ei ollut helppo. Karttapohjaisten esitysten perusteella tutkijat tunnistivat Twitter-käyttäjien kodin sijainnin käyttäjistä noin 65 prosentin kohdalla ja työpaikan 70 prosentin kohdalla.

"Monet ihmiset ajattelevat, että vain erilaisten algoritmien ja muiden koneoppimisen tekniikoiden avulla on mahdollista löytää merkityksellisiä kaavoja sijaintitietojen perusteella, ja he tuntevat olonsa turvalliseksi, koska kaikilla ihmisillä ei ole tähän teknistä osaamista", sanoo MIT:n tutkijana toimiva Ilaria Liccardi, "Halusimme vain osoittaa, että verkkoperustaisten paikkatietojen linkittäminen sekundäärisenä informaationa twiittiin, antaa jopa vähäisellä teknisellä tieto-taidolla varustetulle ihmiselle hyvät mahdollisuudet selvittää missä käyttäjä liikkuu tai asuu.”

44 http://www.pcworld.com/article/3072372/security/got-privacy-if-you-use-twitter-or-a-smartphone-maybe-not-so-much.html

7.2 IP-ositteen geopaikannus

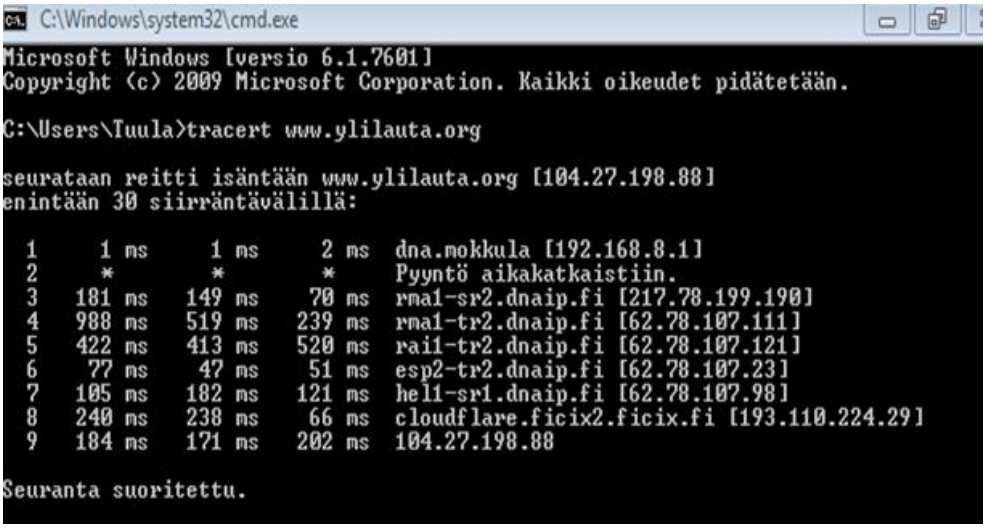
Ennen kuin suoritamme IP-osoitteen geopaikannuksen, on ensin tarpeellista esitellä cmd.exe. Cmd.exe on Windows-käyttöjärjestelmän komentotulkki, verkkoanalyytiikkaa pääsäsnöisesti analysoiva työkalu. Se on tarkoitettu sellaisten tietokoneohjelmien ajamiseen, joilla ei ole graafista käyttöliittymää tai joiden kuvaketta ei ole asennettu Käynnistä-valikkoon. Komentotulkissa voidaan suorittaa rutiininomaisia töitä, kuten varmuuskopiointeja tai skriptejä. Cmd.exe on DOSsin COMMAND.COMin seuraaja. Ennen kuin DOS integroitiin kiinteästi käyttöliittymään Windows 95:ssa, piti graafinen käyttöliittymä käynnistää komentoriviltä, mutta nykyään cmd.exe on graafisesta käyttöliittymästä ajettava apuohjelma.

Komentotulkin käyttö on vähentynyt graafisten ohjelmien yleistyttyä. Moni on kyllä nähnyt - vaikka ohimennen - cmd.exe -apuohjelmaa käytettävän, mutta harva on siihen tutustunut. 80-luvun Commodore 64 -kotietokoneiden käyttäjät muistavat hämärästi siniselle näytölle näppäiltävät komentosanat, joilla käynnistettiin kasetilla tai lerpulla olleita tietokonepelejä: LOAD*”COMMANDO”, 8, 1* tai ”RUN”. Periaate on sama.

IP-osoitteen geopaikannus on helppoa jo pelkän URL-osoitteen perusteella. Mikäli haluamme jäljittää vaikkapa kiistanalaisen keskustelusivuston Ylilaudan URL-osoitteen takana olevan IP-osoitteen, toimimme seuraavalla tavalla. Kopioimme verkkosivun URL-osoitteen, joka tässä tapauksessa on www.ylilauta.org. Tämän jälkeen painamalla Windows-kuvaketta työpöydän vasemmassa alakulmassa, avautuu hakurivi ”Hae ohjelmista ja tiedostoista”. Kirjoita riville ”cmd” minkä jälkeen klikkataan ”cmd.exe ”.

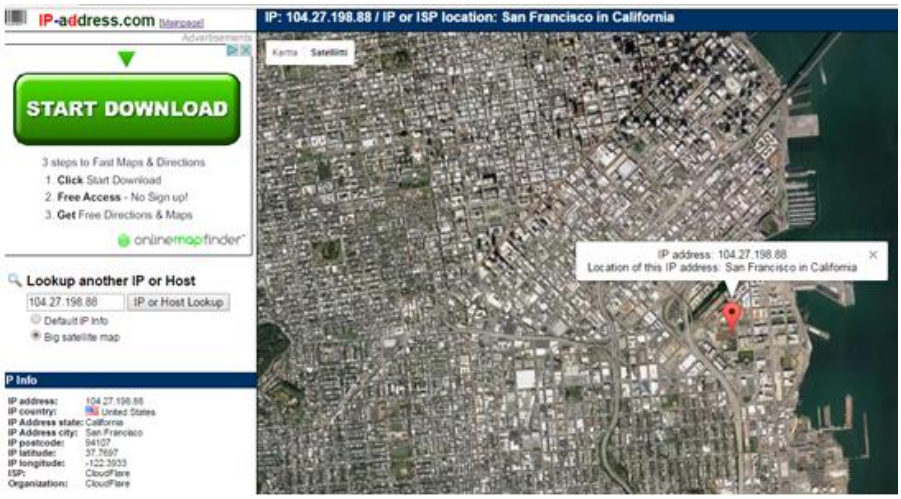
Seuraavaksi kirjoitamme komentoriville: ”tracert” ja kopioimme URL-osoitteen (tai IP-osoite) komennon perään ja painamme enteriä. Tämän ohjelma käy läpi kaikkien reitittimien[45]sijainnit, joita verkkosivu käyttää ja palauttaa niiden IP-osoitteet. Haun jälkeen cmd.exe-näytön tulisi näyttää tällaiselta (Kuva 2).

45 Reititin on tietoverkkoja yhdistävä laite. Reitittimen tehtävä on välittää tietoa tietoverkon eri osien välillä. Reitittimen siis pitää tietää, missä suhteessa eri tietoverkot ovat toisiinsa ja se osaa tehdä tietoliikenteelle reittivalinnan. Reititin on osallisena aina vähintään kahdessa verkossa.



Kuva 2: Cmd.exellä saatu IP-osoitteen reititinlistaus

Seuraavaksi vain kopioimme viimeisimmän IP-osoitteen ja siirrymme verkkosivulle www.ip-adress.com. Kopioimme IP-osoitteen hakukenttään ja hakukone kertoo IP-osoitteen maantieteellisen sijainnin. Kuten alla olevasta hakutuloksesta on todettavissa, IP-osoite paikantuu San Franciscoon (Kuva 2). Koska kyseessä on supisuomalainen sivusto, todennäköistä on että sitä ei ylläpidetä San Franciscossa, vaan verkkosivu käyttää välityspalvelinta eli proxya. Internetissä välityspalvelimet ovat tavallisia. Todennäköisesti jokainen sivu kulkee välityspalvelimen kautta ennen päätymistään selaimeen. Välityspalvelinta käytetään muun muassa IP-osoitteen eli nettiosoitteen salaamiseen jos ei haluta julkiseksi omaa nettiosoitetta, vaikkapa sähköpostissa tai nettikeskustelupalstan valvojille. Jälkimmäisestä syystä on väärinkäytöksien vuoksi usein estetty monet proxy-palvelimet joillekin nettisivustoille.



Kuva 3: IP-osoitteen jäljityksen tulos sivustolla www.ip-address.com

7.3 Creepy

Tietoturvatutkija Yiannis Kakavas on julkaisi tammikuussa 2011 Python-pohjaisen Creepy-ohjelman, jonka avulla on mahdollista tarkistaa millaisia paikkatietoja heistä on tallentunut Twitter- ja Flickr-palveluihin. Creepy kerää verkkopalveluiden tallentamien paikkatietojen ohella samoja tietoja myös käyttäjien jakamista kuvista, joihin lisätyt EXIF-metatiedot voivat paljastaa mm. kuvan ottopaikan koordinaatit.

Ohjelmalla on kaksi käyttötarkoitusta; ensisijaisesti Creepyn tarkoitus osoittaa kuinka helposti tällaisia paikkatietoja saa kerättyä, jotta ihmiset ymmärtäisivät suojata yksityisyyttään paremmin ja toissijaisesti ohjelmaa voidaan käyttää yritysten turvallisuusanalyyseihin. Vaikka Creepy on kelpo työkalu esimerkiksi penetraatiotestaukseen, niin selvää on että ohjelma soveltuu hyvin myös verkkorikollisten työkalupakkiin.

Creepyn käyttöliittymä on hyvin yksinkertainen. Twitter- tai Flickr-tunnuksen kertomalla ohjelma selvittää tarvittavat tiedot automaattisesti, minkä jälkeen ne esitellään kartalla. Aiemmin Creepyllä pystyi selvittämään myös Instagram-julkaisujen paikkatietoja, mutta Instagramin muutettua tiedonhjaussuunnitelmaa, niin Creepy *sandboxattiin*[46]. Karttapohjinaan Creepy käyttää neljää eri karttasovellusta mm. Google Mapsia. Paikat saa esiin aikajärjestyksessä, joten niiden avulla voi mahdollisesti päätellä käyttäjien päivärutiineja. Creepy on tällä hetkellä saatavilla Windows- ja Linux-käyttöjärjestelmille. GPL-lisenssin[47] myötä halukkaat voivat muokata ohjelmaa tarvittaessa esim. lisäämällä siihen tuen muillekin sosiaalisille verkkopalveluille. Creepy tarjoaa seuraavanlaisia hakuominaisuuksia

- Twiittien sijainnit
- Koordinaatit, sikäli kun twiitit on lähetetty mobiililaitteesta
- Paikkatieto ts. maantieteellinen nimi, joka on peräisin käyttäjän ip-osoitteen metadatasta. Tämä on mahdollista siinä tapauksessa, mikäli twiitti on lähetetty web-käyttöliittymästä. Paikka kääntyy koordinaateiksi verkkosivustolla: www.geonames.com

[46] Tietoturvallisuuden termeillä ”sandbox” on suojausmekanismi ohjelmien erottamiseksi. Sitä käytetään usein testaamattomien tai epäluotettujen ohjelmien tai koodien suorittamiseen, mahdollisesti epäluotetusta tai epäluotettavasta kolmannesta osapuolesta, käyttäjistä tai verkkosivustoista vaarantamatta isäntäkoneeseen tai käyttöjärjestelmään kohdistuvaa vahinkoa.

[47] GNU General Public License (GNU-hankkeen yleinen lisenssi) eli lyhennettynä GNU GPL tai pelkkä GPL on vapaiden ohjelmistojen julkaisemiseen tarkoitettu lisenssi, joka antaa kenelle tahansa oikeuden käyttää, kopioida, muuttaa ja jakaa edelleen ohjelmia ja niiden lähdekoodia.

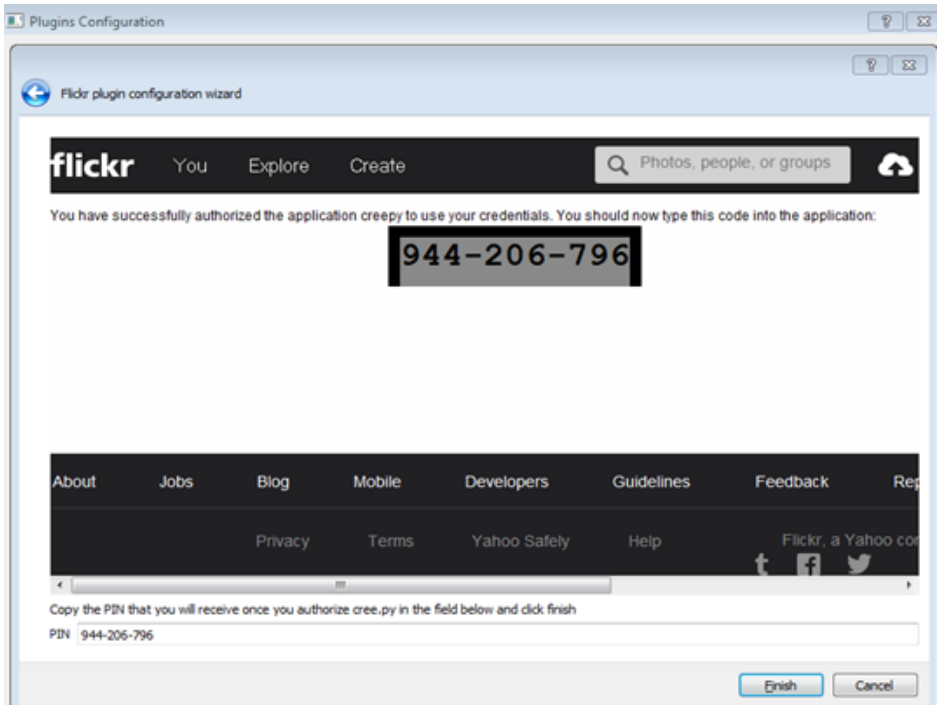
- Rajoituskenttä, joka on peräisin käyttäjän ip-osoitteesta lähetettäessä twiittejä web-käyttöliittymästä. Epätarkka lähde, koska rajoitusruudun nurkka valitaan satunnaisesti.
- Geopaikannustiedot, joihin pääsee käsiksi kuvankäsittelyohjelmien API-liittymän kautta
- EXIF-tunnisteet lähetetyistä kuvista

7.4 Testi

Luo sijaintikartta Jari Sarasvuon twiiteistä Creepya hyödyntämällä

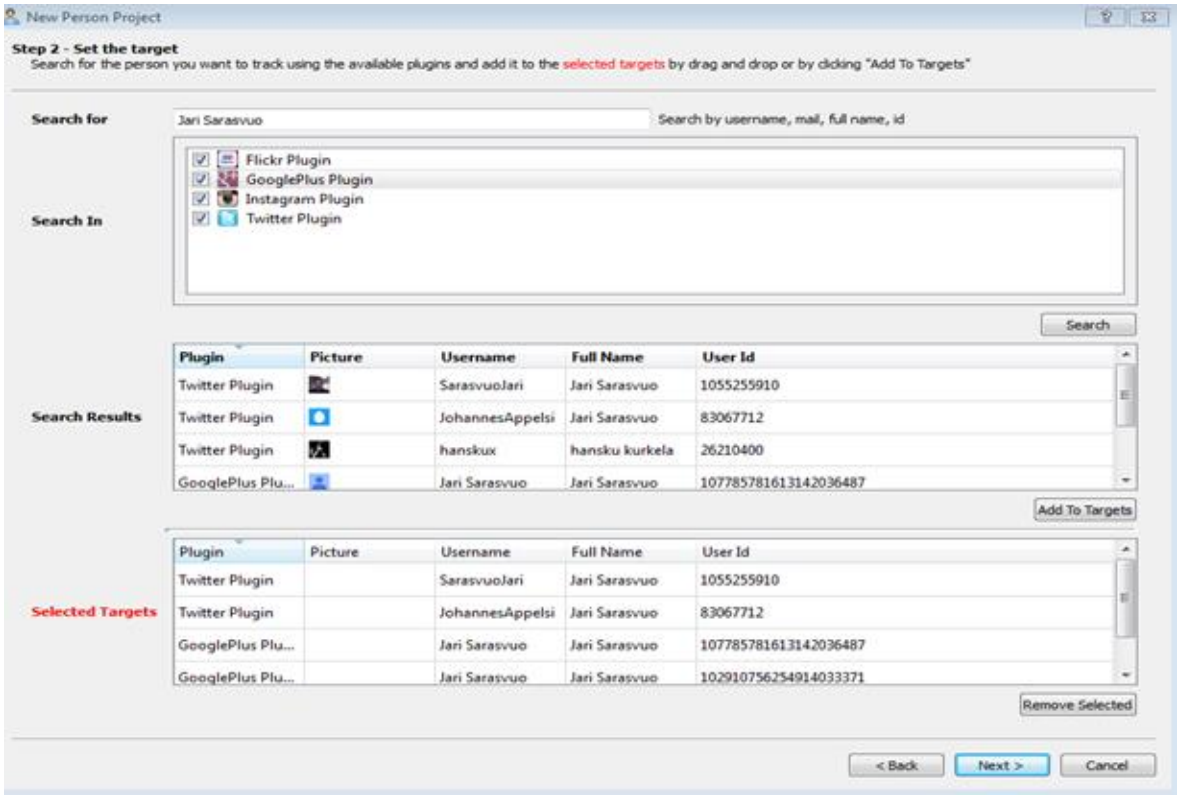
Creepyn käyttö vaatii hieman opettelua ja esivalmisteluja, joten johdattelen Creepyn käyttöön tavallista yksityiskohtaisemmin. Aivan ensimmäiseksi täytyy konfiguroida plug-in, suomeksi liitännäinen tai ”plugari” kuten ohjelmointipiireissä sitä kutsutaan. Se onietokoneohjelma, joka toimii vuorovaikutuksessa isäntäsovelluksen, kuten verkkoselaimen tai sähköpostiohjelman kanssa, tarjotakseen tietyn toiminnon tarvittaessa. Tässä tapauksessa asennamme Twitter-liitännäisen ja asentaminen käynnistetään yläpalkissa olevassa *Plugins configuration* –painikkeesta.

Seuraavaksi valitaan Twitter-liitännäinen valikosta ja painetaan *Run configuration wizard*. Tämän jälkeen käyttäjän tulee kirjautua Twitteriin ja oikeutetaan Creepy käyttämään Twitter-tiliäsi. Tämän jälkeen saat koodin, jonka kopioit sille osoitettuun kenttään Twitter-liitännäisen konfigurointikentässä. Oheisessa kuvassa on kyseessä Flickr-liitännäisen koodi, mutta periaate toimii samalla tavalla (Kuva 4).



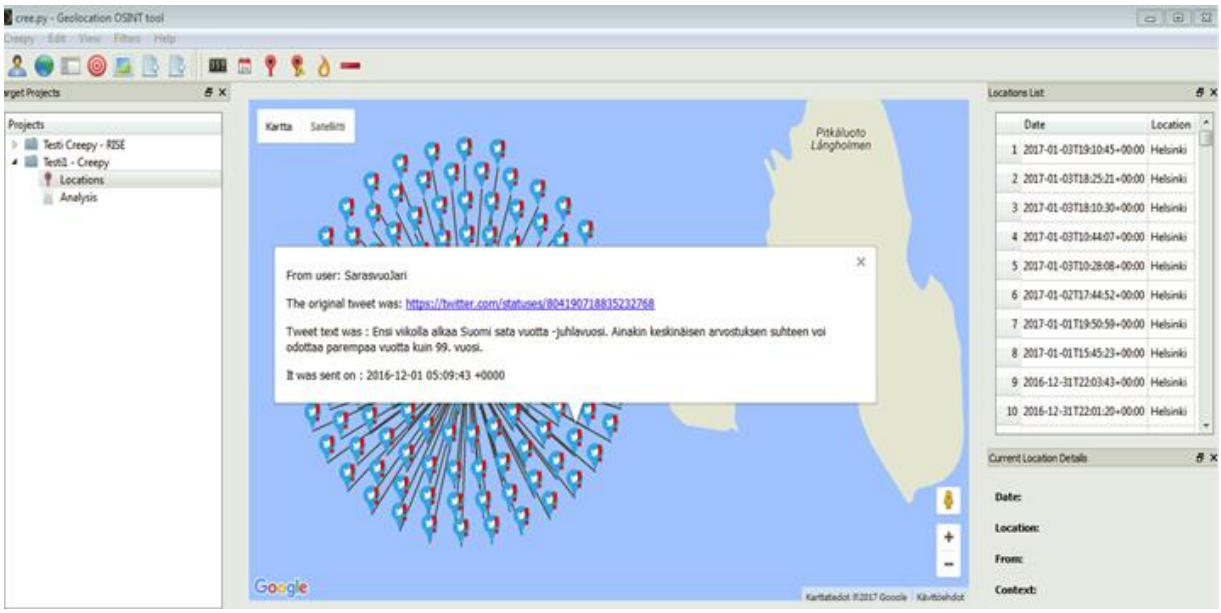
Kuva 4: Flickr-koodi liitännäisen konfiguroimista varten

Kun Creepy ilmoittaa liitännäisen olevan konfiguroitu, voit periaatteessa alkaa käyttämään Creepya. Tämä tarkoittaa sitä, että aloitat ´uuden projektin´ ja se tapahtuu painamalla vasemmasta yläkulmasta ikonia, jonka heijasteessa lukee *Person based project*. Täytettyäsi projektin perustiedot kuten nimen ja kuvauksen, voitkin jo kirjata hakukenttään kohteesi käyttäjänimen, tässä tapauksessa *Jari Sarasvuo*. Seuraavaksi vain valitsen konfiguroimalla käytössä liitännäiset, josta haluan hakuja tehtävän. Esimerkissä on käytössä Flickr, GooglePlus, ja Twitter. Instagram on myös mukana, mutta se ei jo aiemmin selitetystä syystä (Kuva 5).



Kuva 5: Creepyssa luodut hakuparametrit Jari Sarasvuo-hakusanoja varten

Creepy tarjoaa useita hakutuloksia tileistä, joista yksinkertaisesti valitaan drag-and-drop-tyyppisesti alimmaiseen kenttään kohteen tilejä vastaavat hakutulokset. Tällä tavalla rajataan epätoivotut hakutulokset pois lopullisesta projektista. Tämän jälkeen edetään eteenpäin takaisin aloitusnäyttöön, missä voimme todeta uuden projektin syntyneen vasempaan sivupalkkiin. Sitten vain painetaan *Analyze project* ja lopputulemana saadaanerilaisia visualisoituja esityksiä twiittien paikkatiedoista sekä tarkat ajankohdat, jolloin ne on tehty. Jari Sarasvuon kohdalla lähes kaikki twiitit on tehty Espoon saaristossa, missä hänellä on tunnetusti mökki, jossa Trainers´ Housen henkilökuntaa ja vieraita on lukuisia kertoja kestitetty Sarasvuon omien kertomusten mukaan (Kuva 6).



Kuva 6: Jari Sarasvuon Twiittien sijaintitiedot Creepyn mukaan

8. VERKOSTOANALYYSI

8.1 Verkostojen visualisoinnin merkitys

Avointen lähteiden tiedustelun keskeisimpiä tavoitteita on asioiden, ihmisten ja tapahtumien välisten suhteiden tunnistaminen. Ongelmana on, että tietoa on saatavilla niin valtavia määriä. Ihmisen aivoilla on suuria vaikeuksia nähdä hahmottomia yhteyksiä näennäisesti toisiinsa liittymättömän datan välillä. Helpompi on nähdä yhtäläisyyksiä kahden – graafisesti kuvatun – kappaleen kuin pelkän datan välillä. Verkostoanalyysin tarkoituksena onkin esittää analyysin kohteeseen liittyvien eri organisaatioiden ja henkilöiden välisiä suhteita ja tapahtumia visuaalisessa muodossa.

Visuaalinen suhdeverkosto paljastaa usein huomattavasti enemmän kuin tekstipohjainen analyysi. Analyysin avulla saadaan tietoa organisaatioiden ja henkilöiden välisistä suhteista sekä näiden suhteiden merkityksestä ja mahdollisesta sisällöstä. Sillä voidaan havaita myös verkostossa tapahtuvia muutoksia, jotka viestivät verkoston toiminnan, jäsenten tai tavoitteiden siirtymisestä. Verkostoanalyysi on riippuvainen vähintään yhdestä luotettavasta lähteestä. Ongelmana on usein analysoitavan verkoston rajaaminen, joka liian laveana voi johtaa analyysin kohteen paisumiseen tai liian suppeana verkostoanalyysista saatetaan karsia tahattomasti analyysin kannalta keskeisiä osia pois.[48]

Verkostoanalyysi jaetaan tyypillisesti useampaan vaiheeseen. Heuer & Pherson esittelevät kolmiportaisen jaon, jossa ensimmäisenä vaiheena esitellään verkostokaavion rakentaminen, jossa analyysin kannalta keskeiset toimijat ja paikat tunnistetaan ja näiden välille merkitään toimijoiden ja paikkojen suhteita sekä suhteiden laatua kuvaavat yhteydet. Toisena vaiheena on ensimmäisessä vaiheessa luodun kaavion analysointi, jossa toimijoiden ja paikkojen väliset suhteet lajitellaan niiden tyypin mukaan. Tämän jälkeen niistä etsitään analyysin kannalta keskeisiä säännönmukaisuuksia. Viimeisenä vaiheena on sosiaalisen verkoston analyysi, jossa toimijoiden ja paikkojen yhteyksien välimatkoja toisistaan sekä suhdetyyppejä mitataan matemaattisesti. Tarkoituksena on kerätä tarkempaa tietoa verkoston eri osapuolten välisistä suhteista sekä näihin liittyvästä osapuolten toisiinsa käyttämästä vaikutusvallasta.

UNODC:n verkostoanalyysimallissa on seitsemän vaihetta. Ensimmäinen vaihe on raakatiedon kokoaminen analyysiä varten, jossa analyytikko kerää kaiken analyysin kannalta tarpeellisen tiedon yhteen. Toisessa vaiheessa eli verkostokaavion tarkoituksen määrittelyssä analyytikko tunnistaa raakatiedossa olevia ja analyysin kannalta keskeisiä henkilöitä, paikkoja tai muita vastaavia tunnistetietoja, joihin verkostoanalyysissä keskitytään. Kolmas ja neljäs vaihe sisältävät toimijoiden suhteiden ja niiden laadun rakentamista matriisiin, jonka tietoja käytetään myöhemmin hyväksi varsinaista verkostokaaviota piirrettäessä. Viidennessä vaiheessa

48 Heuer & Pherson 2010, 68 – 69

analyytikko laskee toimijoihin liittyvien yhteyksien määrät. Kuudes ja seitsemäs vaihe käsittävät varsinaisen verkostoanalyysikaavion piirtämisen kerättyjen tietojen pohjalta.[49]

Verkostojen automaattinen louhinta tietojärjestelmistä, sosiaalisen mediasta sekä muista avoimista lähteistä on nykypäivän tiedustelun peruselementtejä: excel-listauksella ja sopivalla graafisella tietojenkäsittelyohjelmalla on mahdollista saada aikaan hätkähdyttäviä tuloksia. Verkosta on löydettävissä lukuisia *big dataa*[50] hyödyntäviä, sosiaalisia verkostoja kartoittavia ohjelmia. Jotakin tiedonhankintamenetelmän merkityksestä nykytiedustelusta kertoo se, että NSA käyttää sitä terroristiverkostojen kartoittamiseen.

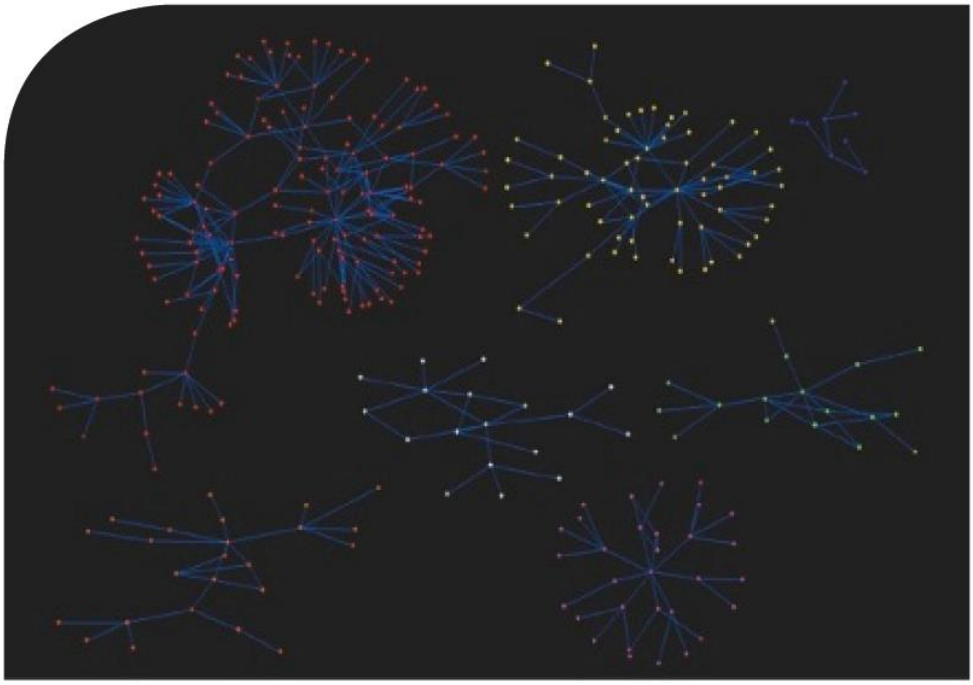
Visualisointitapoja on lukuisia ja usein ohjelmissa voi myös varioida esitystapaa tarpeiden mukaan. Toisinaan yksinkertaisin mahdollinen esitystapa saattaa olla paras. Olen halunnut esittää yhden hieman yksityiskohtaisemman esimerkin verkostoanalyysista, joka yksinkertainen mutta täydellisen toimiva, sen nimi on *Panther-rmalli*.

[49] Criminal Intelligence 2011, 35 – 40

[50] Big data on erittäin suurten, järjestelemättömien, jatkuvasti lisääntyvien tietomassojen keräämistä, säilyttämistä, jakamista, etsimistä, analysointia sekä esittämistä tilastotiedettä ja tietotekniikkaa hyödyntäen.

8.2 Panther-malli

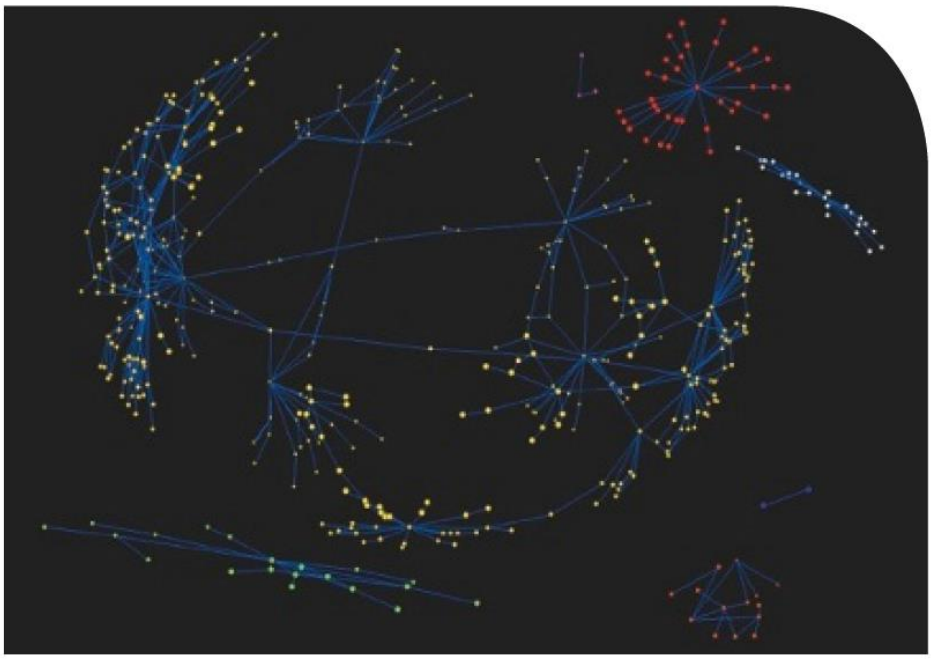
Tukholmalaisen tutkija-poliisin Amir Rostamin katujengien muodostumista ja orgaanisuutta tutkinut raportti havainnollistaa visuaalisen verkostoanalyysin luomat hyödyt. Ensimmäisessä kuvassa (Kuva 1) on kuvattu sekä jäseniä että myötävaikuttajia tuomiotietoihin perustuvassa suhdeverkostossa vuosina 2001-2003, josta myöhemmin muodostuisi merkittävä tukholmalainen katujengi.[51]



Kuva 1: Tukholmalainen katujengi alkutekijöissään.

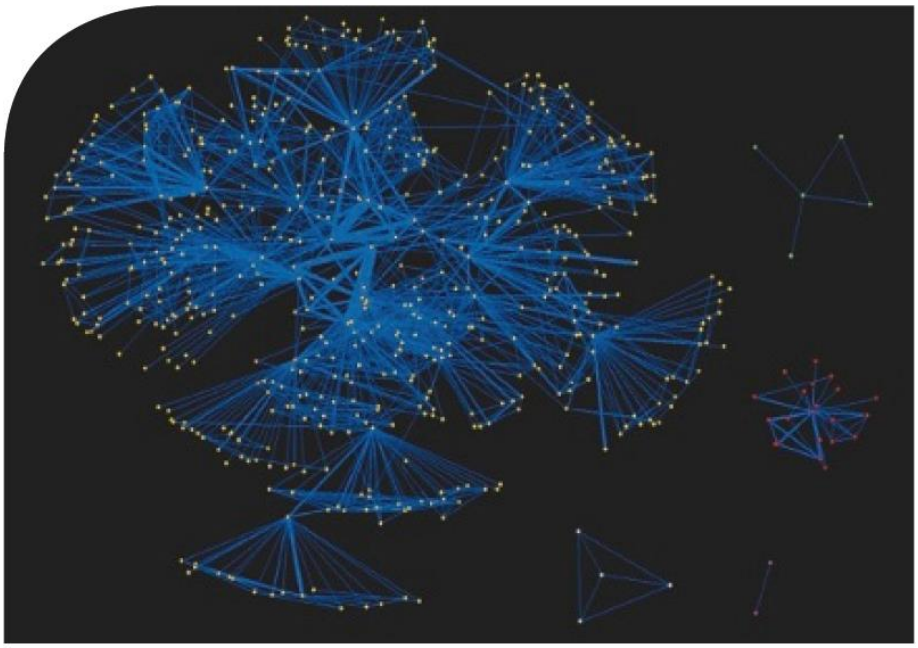
Kuten verkostosta on todettavissa, kyse on toistaiseksi itsenäisesti toimivista rikostentekomielessä aktiivisista ryhmittymistä. Jokainen kahden pisteen välinen viiva kertoo rikoskumppanuudesta. Seuraava kuva (Kuva 2) on vuodelta 2006, jolloin on jo nähtävissä ryhmittymien välistä kontaktointia. Huomionarvoista, että tietyt yksilöt muodostavat yhteyksiä ryhmittymien välille toimien viestinviejinä.

[51] Leinfelt & Rostami 2012, 118-121



Kuva 2: Eri ryhmittymät ovat alkaneet lähentyä toisiaan yhdyshenkilöiden kautta.

Edellisessä kuvassa alkaa jo selkiytyä tiettyjen yksilöiden merkitys organisaatiossa. Graafinen esitystapa edesauttaa avainhenkilöiden löytämistä jengiorganisaation muodostumisessa. Viimeinen kuva (Kuva 3) on visualisointi jengistä neljä vuotta myöhemmin vuonna 2010. Jengin muodostumisen kannalta tärkeimpien henkilöiden selvittäminen on edelleen mahdollista, mutta yhä vaikeampaa.



Kuva 3: Tukholmalainen katujengi loppuvuodesta vuonna 2010.

Edellä kuvattu graafinen esitys ilmentää niitä hyötyjä, joita graafisella verkostoanalyysillä on saavutettavissa. Se on myös erinomainen esimerkki siitä, miten visuaalinen esitystapa palvelee tiedonhankinnan tarkoitusperiä.

8.3 Maltego

Freeware-ohjelmistoksi olen valinnut *Maltegon*, joka kuuluu OSINT:ia tekevien ihmisten perustyökaluvalikoimaan. Patervan Maltego on interaktiivinen datan louhintaohjelmisto ja siitä on mahdollista ladata verkosta sekä maksullinen että ilmainen versio. Käytän esimerkissäni Maltego CE ilmaisversiota. Aivan kuten maksullinen Maltego Classic tai Maltego XL, Maltego CE:n avulla voi kartoittaa sosiaalisten verkostojen uhkakuvia sekä tunnistaa niiden haavoittuvuuksia ja poikkeamia. Maltegon painopiste on sellaisissa reaalimaailman suhteiden analysoinnissa, joista on julkisesti saatavilla tietoa internetistä.

Pääsyllä ajankohtaiseen tietoon on aina ollut suuri rooli tietojärjestelmien turvallisuuden ylläpitämisessä. Tämä tieto on avainroolissa hyökätessä tai puolustettaessa mahdollista kohdetta. Tietojen keräämisessä on ratkaisevan tärkeätä varmistua siitä, että kerää juuri oikeaa tietoa. Jotta hyökkäys olisi mahdollista, tulee tietää missä kohde sijaitsee. Suoraa hyökkäys ei aina ole paras ratkaisu, viisaampaa on hyökätä sinne, missä turvallisuus on heikoimmillaan. Löytääkseen kohteen heikon kohdan, vaatii se mahdollisimman paljon tietoa kohteesta ja kohteen verkostosta.

Maltego on avoimen lähdekoodin tiedustelu- ja analysointi-ohjelmisto. Se mahdollistaa hämmästyttävän hyvät valmiudet datan louhintaan ja keräämiseen sekä presentoida tämä helposti sulatettavassa muodossa. Maltego voi tunnistaa kohteen keskeisen suhdeverkoston sekä tunnistaa ennestään tuntemattomia suhteita. Maltego käyttää client/server-arkkitehtuuria määrittäkseen sirpaleisen tiedon välisiä yhteyksiä verkostojen ja reaalimaailman sosiaalisten yhteyksien avulla kuten:

- Ihmiset
- Sosiaaliset verkostot
- Yritykset
- Organisaatiot
- Web-sivustot
- Internetin infrastruktuuri kuten:
    - Verkkotunnukset
    - DNS-nimet
    - Blogit

- IP-osoitteet

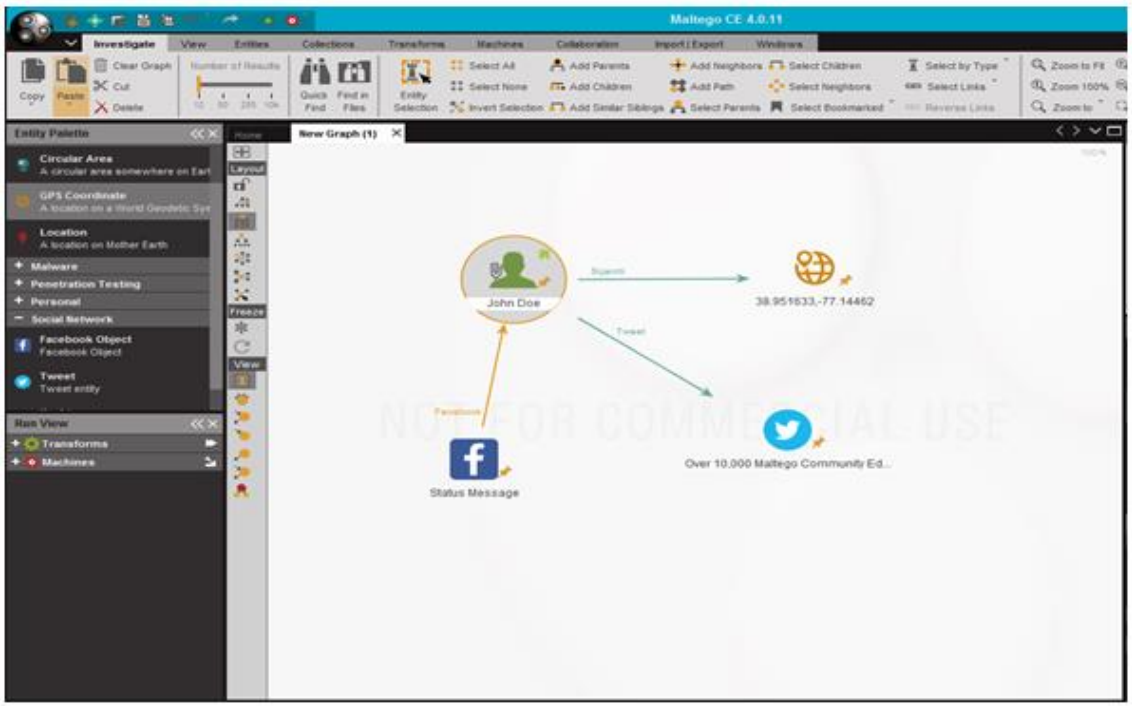
❖ Lausekkeet ja puheenaiheet

❖ Dokumentit ja tiedostot

Graafisen käyttöliittymän avulla (GUI) on mahdollista todeta suhteet verraten vaivattomasti - vaikka ne eivät olisikaan suorassa suhteessa toisiinsa. Maltegosta tekee erityisen hyödyllisen sen tehokkuus sekä joustavuus, jonka avulla tiedonhaun kustomointi on mahdollista. Maltego saattaa ensimmäisellä vilkaisulla hieman säikäyttää tottuman käyttäjän. Maltego sujuva käyttö edellyttääkin tietyn tason kokemusta ja harjaantumista, minkä jälkeen sitä on toki helppo käyttää ja ymmärtää.

Aivan ensimmäiseksi luomme uuden projektin, jonka voi tehdä vasemmasta ylänurkasta löytyvästä valikosta. Seuraavaksi avautuu kokonaan uusi näkymä, jonka sivupalkissa on seitsemän entiteettiä: *devices, infrastructure, locations, malware, penetration testing, personal, social network.* Entiteettejä klikkaamalla kunkin alta avautuu edelleen uusi valikko, joista on mahdollista valita elementtejä projektiin.

Esimerkkikuvassa projektissani on kohdehenkilön (John Doe) lisäksi kolme muuta entiteettiä: Facebook, Twitter ja Location -entiteetit. Olen hiirtä käyttämällä raahannut yhdistävän viivan kohdehenkilön ja kolmen entiteetin välille (Kuva 4).

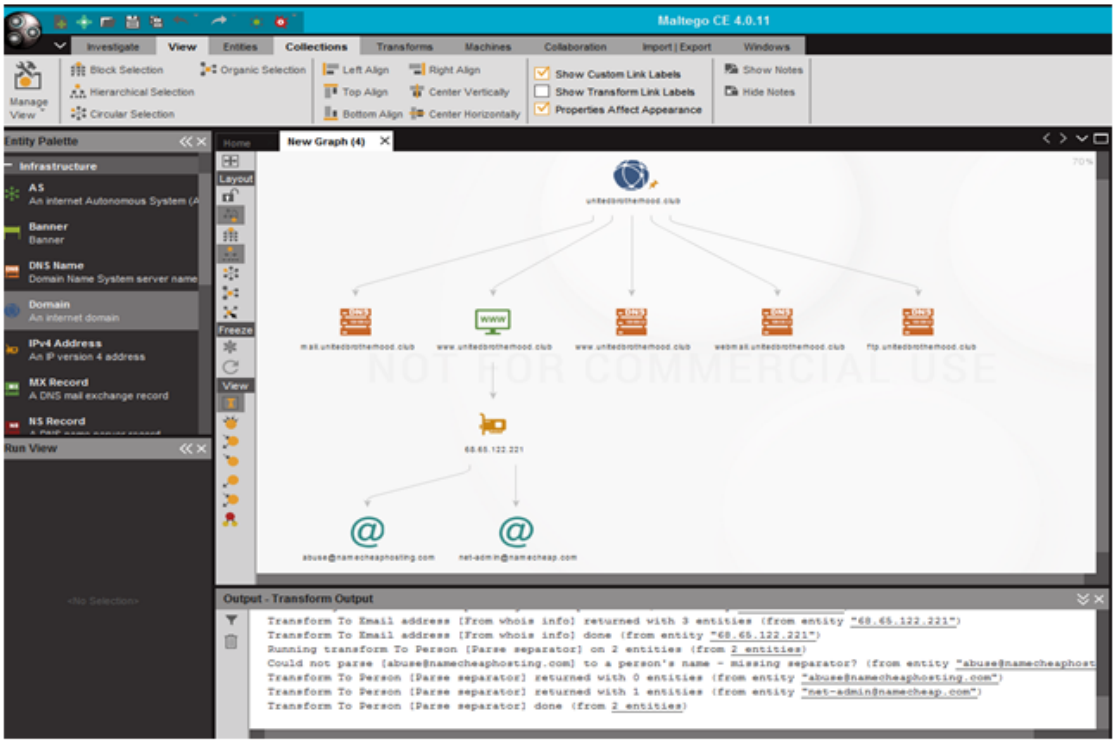


Kuva 4: John Doe johon liitetty kolme entiteettiä

8.4 Testi

Luo internet domain-osoitteen unitedbotherhood.club internet-infrastruktuuri

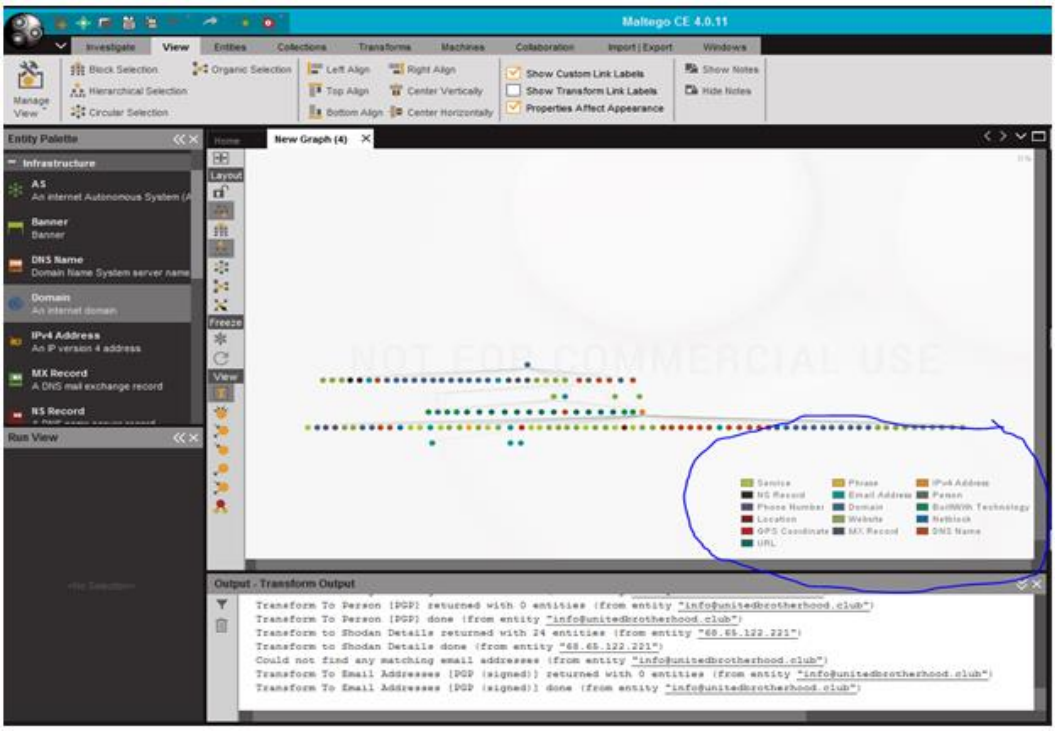
Ensimmäisenä rahaamme sivupalkista ”internet domain”-logon ja kirjoitan logoon unitedbrotherhood.club. Painamme ”Run transform(s)”, jonka alta avautuu uusi valikko. Voimme alkaa purkamaan internet verkkotunnusta pala palalta. Itse aloitan yleensä ylätason verkkotunnuksista (TLD)[52]ja lähden siitä pikku hiljaa alaspäin (Kuva 5).



Kuva 5: Ylätason linkityksiä verkko-osoitteelle www.unitedbrotherhood.club

Säästääksemme aikaa päätämme kuitenkin suorittaa haun ”Run all transforms” ja lopputuloksena saamme verraten kattavan kuvauksen verkkotunnuksen internet-infrastruktuurista ja linkityksistä, jossa on mukana valtava määrä detaljitason tietoa aina sähköposti-osoitteista, nimistä, linkitetyistä verkko-osoitteista etc (Kuva 6).

52 Ensimmäisen tason domain on juuri, joka on pelkkä piste. Juuresta seuraavan tason tunnuksia nimitetään *ylätason verkkotunnuksiksi*(*Top-Level Domain*eli *TLD*). Kaikki ylätason verkkotunnukset ovat säädeltyjä, ja lähes kaikilla ylätason verkkotunnuksilla on oma, erillinen rekisterinsä.Ylätason verkkotunnuksia on kahta päätyyppiä, maatunnuksia (*Country Coded TLD* eli *ccTLD*) ja yleisluontoisia tunnuksia (*Generic TLD* eli *gTLD*). Esimerkiksi Suomessa käytettävä maatunnus on fi.



Kuva 5: Kartoituksen kaikista linkityksistä ja verkostoista, jotka sivustoon on kiinnittynyt

Olen ympyröinyt sinisellä kaikki värikoodein järjestellyt entiteetit, jotka verkkosivuun ovat liittyneet: ip-osoitteet, sähköposti-osoitteet, domain-osoitteet, URL-osoitteet yms. Linkityksiä on siinä määrin runsaasti, että saadakseen yhdellä silmäyksellä käsityksen niistä, olen ollut pakotettu zoomaamaan ulos.

9. PÄÄTÄNTÖ

Useimmat meistä avaavat muutamassa sekunnissa yhteyden verkkoon omalla älylaitteellaan, kirjoittavat hakukoneen tietokenttään avainsanat etsimäänsä informaatioon ja samassa hetkessä heillä 64 100 osumaa lohen savustuksesta. OSINT:n kannalta oleellista on, että älylaitteen omistaja on verkossa käytännössä 24/7, mikäli laite ei ole ns. lentotilassa.[53] Tähän perusajatukseen OSINT nojaa ja hyödyntää interentin käyttäjän harhakuvaa siitä, että oma digitaalinen jalanjäli on verkossa niin hajallaan, että punaisesta langasta ei saa kukaan kiinni. Oikeastaan tämän tutkielman päätavoite on ollut osoittaa miten väärässä tässä suhteessa voikaan olla.

Avointen lähteiden rooli sotilastiedustelussa on havaittu erittäin merkittäväksi, jonka osoituksena OSINT-tiedustelua syvällisemmin käsittelevä dokumentaatio – jota tässäkin tutkielmassa on käytetty – on sotilastiedustelulle suunnattua. Suurimmaksi esteeksi julkisten lähteiden hyödyntämiseksi sotilastiedustelussa on havaittu perinteisen tiedusteluprosessin teolliselta aikakaudelta peräisin oleva toimintamalli. Tiedusteluprosessin tulisikin muuttua kohdekeskisemmäksi ja verkostomaisemmaksi pystyäkseen vastaamaan kasvaneisiin vaatimuksiin.

Avointen lähteiden tiedonhankinnan tiedustelumetodista on tunnistettavissa neljä edellytystä, jotka edesauttavat tehokasta ja systemaattista internet-tiedonhankintaa:

1. Kerää lähteitä

   Julkisia lähteitä on lukemattomia, joten on perustellumpaa lähteä liikkeellä joistakin vakiintuneista, hyväksi havaituista lähteistä ja vasta sitten alkaa kerätä case-kohtaisia lähteitä. Hyvä lähdetietopankki ja niihin liittyvät toimivat tiedonhankintatekniikat ovat kaiken a ja o.

2. Valitse alue tai aihe

   Tiedonhankinnan kohteena on yleensä hyvä olla selkeästi rajattu kohde taikka maantieteellinen alue. OSINT on ennen kaikkea tiedonhaun rajaamisen taito. Informaatiota on tarjolla yli tarpeen, merkityksellisen tiedon paikallistaminen on taitolaji. Keskity siis yhteen alueeseen tia asiaan kerrallaan.

3. Yhdistele pisteet

   Kun sinulla on käytössä riittävästi tietoa, pyri löytämään murupolkuja asioiden välillä. Ole luova. Pyri löytämään asiayhteyksiä, merkityksellisiä henkilöitä ja tapahtumia. Esim. Joskus

[53] Lentotila on asetus, jonka avulla mobiililaitteen kaikki langattomat yhteydet voi nopeasti poistaa käytöstä. Langattomiin yhteyksiin sisältyvät Wi-Fi, mobiililaajakaista, Bluetooth, GPS tai GNSS, Near Field Communication sekä kaikki muut langattomat yhteystyypit.

kohteena olevan some-profiilin ystävät ja ystävien ystävät on merkityksellisempiä tiedonhankinnan näkökulmasta kuin varsinainen kohde.

4. Seuraa uusia tuulia

OSINT-työkalut eivät ole vakiintuneita, vaan esimerkiksi Facebook, Twitter tai LinkedIn reagoivat jatkuvasti palveluihinsa kohdistuviin tiedonhankintatyökaluihin. Ne saatetaan kokea uhkana palveluja käyttävien henkilöiden yksityisyydelle. Uusia työkaluja kuitenkin ilmaantuu tilalle samaan tahtiin kuin vanhoja poistuu käytöstä. Osaamista on siis syytä alati päivittää.

Vaikka edellä onkin lueteltuna neljä askelmerkkiä hyvään avointen lähteiden tiedonhankintaan, niin totuus on että OSINT:iin ei voi soveltaa kovin standardia työprosessia tai *workflowta*. Jokaista projektia, jokaista kohdetta on lähestyttävä yksilöllisesti. Paljon merkitystä on sillä millaiset lähtötiedot prosessille on: onko käytössä nimi, domain-nimi, käyttäjänimi, sähköpostisoite, puhelinnumero, some-profiili…. Lähtiötiedot sanelevat lopulta lopullisen tiedonhankinta-prosessin kulun.

Vaikka lähtökohtaisesti OSINT liitetäänkin systemaattiseen valtiojohtoiseen sotilastiedusteluun, se voi olla arvokas resurssi myös infosec-ammattilaisille, jotka pyrkivät kartoittamaan parhaita tapoja jäsentää runsasta verkosta saatvailla olevaa dataa. Monessa tapauksessa erilaiset turva-alan toimijat, pentraatiotestaajat ja tietoturva-ammattilaiset eivät kuitenkaan kiinnitä riittävästi huomiota turvallisuusarvioinnin ensimmäiseen vaiheeseen eli tiedusteluun. Usein avoimesti esillä olevat tiedot saattavat olla jopa yhtä kriittisiä kuin luottamukselliset tiedot. Hyvänä esimerkkinä mainittakoon, että kyberhyökkäystä valmisteltaessa, avointen lähteiden tiedonhankinta on esivalmistelujen ensimmäinen vaihe. OSINT:lla pyritään luomaan kuva kohteesta, tunnistaa hyökkäyksen kohteena olevan instanssin henkilöhierarkia, avainhenkilöt, järjestelmien pääkäyttäjät, heidän lähiverkostonsa, kiinnostuksen kohteensa, heikkoutensa ja pelkonsa.

Arkisempi käyttötarkoitus OSINT-tekniikoille voisi olla huolestuneet vanhemmat, jotka voivat käyttää kuvattuja tohjelmaia ja menetelmiä etsiessään tietoa lastensa käyttäytymisestä sosiaalisessa mediassa. Oma lukunsa ovat tietysti tutkivat journalistit, joille avoimet lähteet ovat suorastaan kriittisiä tiedonlähteitä. Järjestäytyneen rikollisuuden ja terrorismin torjunnassa OSINT on nykypäivänä suorastaan korvaamaton tiedonhankintamenetelmä. OSINT:lla on saatu suuria hyötyjä rikostaustaltaan nuhteettomien taustavaikuttajien tunnistamisessa ja linkittämisessä osaksi laajempaa rikollisverkostoa.

Toivon että esitetyt esimerkit ovat sytyttäneet lukijassa kipinän aiheeseen syvällisempään perehtymiseen. Tutkielmaa tehdessäni olen tiedostanut, että tässä esitetyt metodit ja ohjelmat ja verkkosivut vanhentuvat nopeasti, joten paperimuotoinen esitystapa ei ole aheesta kovin kiitollinen, vaan toimivampi malli olisi esimerkiksi päivittyvä verkkosivu. Suosittelenkin seuraamaan ylläpitämääni Facebookin OSINT Finland –sivua: www.facebook.com/osint.finland,

jonne päivitän säännöllisesti tietoa ja uutisia uusista tuulista OSINT:in alueella sekä seuraamaan myös sellaisia sivustoja kuin:

- The OSINT Journal Review (www.osintjournal.wordpress.com)
- Bellingcat (www.bellingcat.com)
- The OSINT Journal (http://theosintjournal.blogspot.fi)
- IntelTechinques (www.inteltechniques.com)
- ONSTRAT (http://www.onstrat.com/osint)

Itse olen käytänyt säännöllisesti ja menestyksekkäästi *StumbleUpon* –nimistä ohjelmaa: http://www.stumbleupon.com , jota selailemalla hakusanalla ´Internet Tools´, olen löytänyt uusimpia verkkoon ilmestyneitä verkkotyökaluja, joita on ollut mahdollista soveltaa avointen lähteiden internet-tiedonhankintaan.

Mikäli tutkielma ei herättänyt lukijassa kipinää varsinaiseen toimintaan, vaan pikemminkin säikäyttänyt pahan päiväisesti, niin step-by-step -ohjeet virtuaaliseen verkkoitsemurhaan löytyvät tästä verkko-osoitteesta:

http://www.topinfopost.com/2013/12/12/how-to-disappear-on-the-internet-infographic

10. LÄHTEET

Kirjalliset lähteet

Criminal Intelligence: Manual for Analysts. 2011. United Nations. UNODC. Viitattu 28.7.2012 http://www.unodc.org/documents/organized-crime/Law Enforcement/Criminal_Intelligence_for_Analysts.pdf

Europol: Analyysiopas. 2001. Euroopan analyysiyksikkö. 2. versio. Suomenkielinen versio.

Goodman, Mark. 2016. Future Crimes. Inside the Digital Underground and the Batlle for Our Connected World. Anchor Books. New York.

Heuer, Richards J., Jr. & Pherson, Randolph. 2010. Structured analytic techniques for intelligence analysis. Washington: CQ Press.

Heuer, Richards J. 1999. Psychology of Intelligence Analysis. Viitattu 15.1.2017 https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/booksand monographs/psychology-of-intelligence-analysis/PsychofIntelNew.pdf

Hex Editor Definition. 2006. The Linux Information Project. Viitattu 10.12.2016 http://www.linfo.org/hex_editor.html

Leinfelt, Fredrik, Rostami, Amir (toim.). 2012: The Stockholm Gang Model: PANTHER. Stock-holm Gang Intervention & Prevention Project 2009-2012. Stockholm University.

Metadata working group 2015. Specifications. Viitattu http://www.metadataworkinggroup.org/ specs.

Open Source Intelligence Handbook. 2001. NATO. Viitattu 4.12.2016 http://www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NAT O%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf

WWW-sivustot

CIA WWW-sivut <https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html> 15.1.2016

DefenceTech WWW-sivut <http://defensetech.org/2012/03/15/insurgents-used-cell-phone-geotags-to-destroy-ah-64s-in-iraq/> 1.2.2017

Dfir.com WWW-verkkosivut <http://dfir.com.br/wp-content/uploads/2014/02/blair.htm> 10.10.2016

Engadget.com WWW-sivut <http://www.engadget.com/2016/05/02/brazilian-government-whatsapp-block-72-hours/> 10.10.2016

Finlex <http://www.finlex.fi/fi/laki/ajantasa/1999/19990523> 15.12.2016

Ilta-Sanomien WWW-sivut <http://www.iltasanomat.fi/viihde/art-1288596309269.html> 3.11.2015

Ilta-Sanomien WWW-sivut <http://www.iltasanomat.fi/digitoday/art-2000001185316.html> 1.2.2016

IT-viikko WWW-sivut <http://www.itviikko.fi/uutiset/2015/11/11/kello-tikittaa-facebookille-kohta-pitaa-maksaa--eika-ihan-vahan/201514831/7> 13.3.2017

Medium.com WWW-sivut <https://medium.com/@roselisker/illuminating-the-dark-web-d088a9c80240#.65xn63oar> 12.4.2016

PC-world WWW-sivut <http://www.pcworld.com/article/3072372/security/got-privacy-if-you-use-twitter-or-a-smartphone-maybe-not-so-much.html> 14.2.2016

Scientific America Journal WWW-sivut >http://blogs.scientificamerican.com/observations/how-black-holes-led-to-the-creation-of-web-browsers/> 12.3.2017

Tivi.fi WWW-sivut <http://www.tivi.fi/Kaikki_uutiset/kayttakaa-whatsappia-ja-facebookia-yliopisto-sammuttaa-legendaarisen-irc palvelimensa-6555896> 13.3.2017

Verkkomedian WWW-sivut <http://www.verkkomedia.org/news.asp?mode=4&id=10412> 12.12.2016

Wikipedia – Doxing <https://en.wikipedia.org/wiki/Doxing> 2.1.2016

Wikipedia – WikiLeaks <https://fi.wikipedia.org/wiki/WikiLeaks> 2.1.2016

Wired-lehden WWW-sivut <https://www.wired.com/2016/03/https-adoption-google-report/?utm_content=buffer46336> 15.4.2017

Zsdnet.com WWW-sivut <http://www.zdnet.com/article/want-to-limit-windows-10-tracking-there-is-an-app-for-that/> 2.2.2017

WSJ.com WWW-sivut <http://www.wsj.com/articles/SB10001424052748703940904575395073512989404> 30.2.2016


YOU ARE BEING
MONITORED