

Keskustelun tunneallot – Suomi24-hanke

■ Krista Lagus, Mika Pantzar ja Minna Ruckenstein

Helsingin yliopiston politiikan ja talouden tutkimuksen laitoksella toimiva Kuluttajatutkimuskeskus on avannut yhteistyössä Allerin, FIN-CLARINin ja CSC:n kanssa Suomi24-aineiston tutkimuskäyttöön. Kyseessä on suomalaisittain ainutlaatuinen avoimen datan hanke. Hankkeeseen on jo kytkeytynyt kymmeniä tutkijoita ja yhteistyötahoja, joten aineistoanalyysi ja niistä käyty keskustelu tulee tuottamaan tutkimusta ja uudenlaisia sosiaalisen median käyttöjä. Työn alla on esimerkiksi tutkimusaloitteita ennakoivista algoritmeista ja uudisanoista.

Kuluttajatutkimuskeskuksen lähtökohtana on, että Suomi24-aineiston avaaminen tukee yhteiskuntatieteellisen tutkimuksen uudistumista sekä uudenlaisen osaamisen ja yhteistyömuotojen rakentumista. Suomi24 on jaettu varanto, jonka ympärille tutkijat voivat kerääntyä. Keskusteluaineisto on ajallisesti pitkäkestoinen ja kansainvälisestikin poikkeuksellisen laaja sosiaalisen median aineisto. Viestejä on yli 70 miljoonaa ja niihin on tallentunut 15 vuoden ajalta suomalaista keskustelua.¹

Suomi24-aineistoa luonnehtii asiakeskeisyys. Se perustuu vähemmän oman identiteetin rakentamiseen ja kaveriverkostoihin kuin esimerkiksi Facebook. Keskustelua vievät eteen-

päin jaetut kiinnostuksenkohteet, esimerkiksi harrastukset, sukupuolinen suuntautuneisuus, perheen perustaminen tai terveyteen liittyvät ongelmat. Parhaillaan käynnissä oleva työ edistää valtavan kokaisen aineiston käytettävyyttä rakentamalla aineistonhaun käyttöliittymää. Tavoitteemme on, että aineistoa voisivat käyttää myös tutkijat, joilla ei ole data-analyysin osaamista. He tarvitsevat yksinkertaisia työkaluja aineiston analyysiin ja visualisointiin, jotta menetelmälliset edistysaskeleet olisivat mahdollisia. Data-analyysin on tuotettava tarpeeksi luotettavia tuloksia, jotta niille voisi perustaa tutkimusta. Toisaalta ymmärrämme, ettei sosiaalisen median analyysi ole luonteeltaan objektiivista tai aukotonta. Se avaa näkymiä keskustelutodellisuuteen, jotka on kontekstualisoitava joko toisten aineistojen tai laajemman tutkimuksellisen näkökulman avulla. Tämänkaltaiselle tutkimustyölle on ominaista, että löydökset nousevat aineistosta lukuisten iteraatioiden ja monimuotoisen osaamisen integraatioiden kautta.

Käyttöliittymää kehittää yhteistyössä tutkimusryhmämme kanssa Pasi Karhu (Studio Mind), joka on muun muassa tutkinut ja visualisoinut aiemmin BBC:n keskusteluaineistoista kansalaisten reaktioita maahanmuuttokysymyksiin. Sosiaalisen median keskustelut ovat aina osin riippuvaisia käytettävän palvelun informaatioarkkitehtuurista. Käyttöliittymän avulla toivomme voivamme tunnistaa ajallisesti aiheenmukaista ja yhteiskunnallista keskusteludynamiikkaa erilaisissa sosiaalisissa medioissa. Tällaista tutkimusta on jo tehty digitaalisen kulttuurin alueella, mutta kehittämämme käyttöliittymän avulla keskusteluja ohjaavan tunnesanaston ja keskustelujen intensiteetin tunnistaminen helpottuu. Tutkimushypoteesimme on, että kes-

1 Suurin osa sosiaalisen median käyttäjistä tulee Suomi24:n sivuille Googlen hakujen kautta. Keskustelijoiden taustamuuttujat mukailivat suomalaista demografiaa. Suomi24:ssä on lähes 2 900 erilaista keskusteluaihealuetta ja päivittäin Suomi24:n keskustelupalstoille kirjoitetaan yli 20 000 uutta viestiä eli vuodessa yli seitsemän miljoonaa. Joka vuosi lähes miljoona viestiä myös poistetaan, koska ne rikkovat palvelun sääntöjä (mm. kunnianloukkaukset, muuten asiattomat tai laittomat viestit). (Lähteet: TNS Metrix 2014, Google Analytics 2014.)

kusteluaineiston avulla löydämme voimia, joiden ohjaamina suomalaista mielenmaisemaa synnytetään ja tiivistetään.

Tunneaaltoihin keskittyminen sisältää ajatuksen, jonka mukaan keskustelun sisällä voi tunnistaa sitä ohjaavia aaltoja. Esimerkiksi kriitillä, raivolla tai myötätunnolla on aina alku, keskikohta ja loppu. Tekstinlouhinnan ja algoritmien avulla aineistosta voidaan tunnistaa tunnealloille tyypillisiä kestoja, kiihdytyksiä, jarruja tai rytmejä. Voimme tarkastella, minkälaiset kommentit vahvistavat tai vaimentavat Suomi24:n keskusteluissa kuljetettua raivoa tai myötätuntoa? Onko löydettävissä keskusteluillemme ominaisia ”takaisinkytkentöjä”, keskustelun tihentymistä, viestien pituuksia tai sanastoa, jotka pitävät keskusteluja yllä, hajaannuttavat tai lopettavat ne? Minkälaisia rooleja ja työnjakoa sosiaalisen median tunteiden pelissä on tunnistettavissa? Voimmeko tunnistaa suurista keskustelumassoista työnjakoa, jotka aktiivisesti järjestävät keskusteluvaruutta?

Ajatuksena on siis oppia kuvaamaan kansalaiskeskustelun tunteiden ja teemojen elinkaaria kehitettävän käyttöliittymän avulla. Jatkossa pohdimme, kuinka kehittää kielitieteilijöiden tunneluokituksesta ”tunnetilatunnistin”, joka osaa identifioida tunteisiin liittyviä ilmaisuja ja kuinka tutkimusyhteistyö voisi ylläpitää sitä tavalla, joka kunnioittaa kielen rikkautta ja kielellisen vuorovaikutuksen alati muuntuvia muotoja.

Sosiaalisen median aineistoja syntyy reaaliaikaisesti. Siksi niitä on hyvä tuntee, arvioida ja jäsentää osana tutkimusta, uuden oppimista, poliittista vaikuttamista ja osallistamista. Kun aineistosta löydetään esimerkiksi uudisanoja, voimme nähdä mihin suuntaan jonkun aiheen käsittely on menossa. Uusien jännitteiden tunnistaminen on arvokasta, jotta niitä voidaan käsitellä ennen kuin niistä tulee avoimia konflikteja. Olemme pitäneet tärkeänä sitä, että hankkeessamme on kansalaislähtöinen ote. Pyrimme edistämään data-analyysin keinoin eri tahojen välistä dialogia ja uuden oppimista, emme kansalaisten seurantaa, leimaamista ja ylhäältä ohjailua. Rakennamme hankkeen aikana aktiivisesti tarkkailusta poispyrkivää kult-

tuuria luomalla yhteistyökumppaneiden kanssa avoimuudelle ja yhteistyölle perustuvia käytäntöjä ja toimintatapoja. Olemme kutsuneet vuoropuheluun tahoja, jotka voisivat mahdollisesti hyödyntää tutkimuksemme tuloksia, ja toisaalta pyrimme edistämään laajempaa keskustelua siitä, minkälaiset tulokset ovat yhteiskunnan ja sen kansalaisten näkökulmasta hyödyllisiä.

Suomalaisen osaamisen tukena

Hankkeellamme on vahva yhteys uudenlaisen data-analyysiosaamisen synnyttämiseen. Tutkimukseen kohdistuu yhä enemmän painetta tuottaa markkinoitavia tuotteita ja taloudellisesti hyödynnettävää osaamista. Meidän näkemyksemme mukaan uusia tuotteita ja taitoja syntyy tutkimusyhteistyön sivutuotteena, jos hankkeessa osataan taitavasti lomittaa ja yhdistellä jo olemassa olevaa osaamista. Emme usko yliopistokulttuuriin, jossa tutkijoita kannustetaan vastaamaan yritysten muotoilemiin tutkimuskysymyksiin. Tutkijoiden ja yritysten yhteistyö voi kuitenkin parhaimmillaan hyödyttää molempia. Siksi olemme koonneet Suomi24-hankkeeseen eri tieteenalojen osaajia, mutta myös kaupallisia toimijoita. Pyrimme näin luomaan edellytyksiä tekstinlouhinnalle ja sosiaalisen median kehitystyölle kiinnittämällä erityistä huomiota innovatiivisen kehitystyön reunaehtoihin.

Innovaatioista puhutaan nykyään kevytmielisesti, ymmärtämättä niiden vaativan vuosien tai jopa vuosikymmenien kypsytelyä. Suuri osa kehitysideoista on esitelty aiemmin, mutta ne eivät ole saaneet aikalaisyhteisöltään tukea. Markkinoiden ja politiikan tunneliikkeiden analyysin tutkimushistorialliset juuret löytyvät 1900-luvun alusta, kun ranskalainen sosiologi Gabriel Tarde hahmotteli ”imitaatiolakeja”. Tarden mukaan keskustelu on uudistuvaa yhteiskuntaa ja taloutta ylläpitävä voima. Englannissa ja Yhdysvalloissa taas käynnistyi 1930-luvulla *Mass observation* -tutkimus kansakunnan mielenliikkeiden ja arkisten huolien ymmärtämiseksi. Nämä tutkimukselliset aloitteet ovat aktivoituneet digitaalisten työkalujen kehittymisen myötä. Tarden kirjoituksia luetaan nyt täysin uudesta näkökulmasta.

Sosiaalisen median diagnostiikka kehittyi maailmalla nopeasti. Hallitsevat yritykset, kuten Google ja Facebook, tukevat omien aineistojensa tutkimuskäyttöä. Tutkijoilla ei ole pääsyä yritysten keräämiin aineistomassoihin, joten he ovat altavastaaajia sosiaalisen median diagnostiikassa. Tekstinlouhintaa käytetään jo varsin arkisesti akateemisen yhteisön ulkopuolella, mutta näihin käyttäisiin on vaikea päästä käsiksi. Esimerkiksi Federal Reserve ja Nasdaq tutkivat ja ennakoivat talouden suhdanteita digitaalisilla menetelmillä.

Tutkimushankkeemme näkökulmasta kiinnostava tutkijavetoinen hanke on esimerkiksi Amsterdamin yliopiston Digital Methods Initiative. Lisäksi seuraamme kriittistä datatutkimusta, joka seuraa datavarantojen omistajuuden kasautumista ja sitä minkälaisia päämääriä ja tavoitteita edistetään nyt kehitteillä olevilla teknologioilla. Sosiaalisen median aineistojen analyysin eettisiä periaatteita pohditaan hankkeemme aikana useista eri näkökulmista. Yritämme täsmentää minkälaisia käytäntöjä, prosesseja, koulutusta ja laadunvarmistusta tarvitaan, jotta sosiaalisen median analyysimenetelmien ja teknologioiden kehitys tukee yhteiskunnan suotuisaa uudistumista, eikä johda esimerkiksi entistä vahvempaan valvonta- ja kontrolliyhteiskuntaan, jossa valta kasaantuu yhä harvemmalle.

Suomalaisittain on myös huomionarvoista, että aineistoanalyysi on keskittynyt lähinnä englanninkielisiin teksteihin. Suomi24-aineiston avaaminen pyrkii korjaamaan tätä vinoumaa.

Yksi hankkeemme keskeisiä kysymyksiä on, miten suomen kielen tuottama sosiaalisen median tunnepuhe ja vuorovaikutus eroaa vaikkapa englanninkielisestä. Tämän takia on tärkeää, että Suomi24 aineisto houkuttaa yhteiskuntatieteilijöiden rinnalle vuorovaikutuksen tutkijoita, lingvistejä ja kieliteknologeja. Vaikka aineistomme on suomenkielinen, kehitettävä työskentelytapa ja tilasto-matemaattiset menetelmät luovat valmiuksia kansainväliseen yhteistyöhön. Esimerkiksi IBM, joka on johtava suurten tekstimassojen algoritmien kehittäjä, on kiinnostunut pienen kielialueen vaikeahkon kielen mallintamisesta.

Uudenlaisia työtapoja

Näin iso hanke vaatii akateemiselta tutkimukselta uudenlaisia työnjaollisia malleja.² Pyrimme luomaan yhteistyömalleja tutkijoiden yhteistyöhön, mutta myös päättäjien ja kansalaisten osallistamiseen. Suomi24-aineiston yksi keskeinen arvo liittyy siihen, että se voidaan tarvittaessa yhdistää muihin aineistoihin. Sitä voi peilata esimerkiksi työttömyystilastoihin, vaalikäyttämiseen tai muihin sosiaalisen median aineistoihin. Kiinnostavan vertailukohdan aineistolle tuovat myös esimerkiksi terveystilastot. Suomi24-aineistolla voi luoda sisältöä siihen, miltä suomalaisten yleisimmät terveysongelmat, esimerkiksi masennus tai päihdeongelmat, näyttävät Suomi24:n keskustelujen tai keskustelussa käytettävien tunnesanojen kautta.

Työ etenee nyt pienimuotoisten pilottien avulla, joiden avulla tutustumme aineistoon ja esittelemme alustavia tuloksia tutkijoille ja muille yhteistyötahoille. Toivomme pilottien konkretisoivan aineistoanalyysiä, tutkimuksellisia edistysaskelia ja erilaisia sovellusalueita. Jatkossa tavoitteena on kehittää aineistosta helposti käytettäviä johdannaisaineistoja, kuten erilaisia sanastoja ja tiivistyksiä. Data-analyysitutkijoita toivomme mukaan pohtimaan, millaisilla signaalinkäsittelyn menetelmillä vuorovaikutusketjuja voitaisiin parhaiten analysoida, ryhmitellä ja visualisoida. Avoimen datan hankkeen hengessä kutsumme yhteistyöstä kiinnostuneita ottamaan yhteyttä Krista Lagukseen (krista.lagus@helsinki.fi) ja seuraamaan keskusteluja *Citizen Mindscales* Facebook-ryhmässä.

Kirjoittajat työskentelevät tutkijoina Helsingin yliopiston Kuluttajatutkimuskeskuksessa.

2 Suomen Akatemia valitsi "Kansakunnan mielenliikkeet"-hankkeen yhdeksi Digitaalisten ihmistieteiden ohjelmassa rahoitettavaksi tutkimushankkeeksi. Konsortiolle on myönnetty 1,1 miljoonan rahoitus. *Tieteessä tapahtuu* -lehden toimitus onnittelee!