

OLLI LEHTONEN

FT, apulaisprofessori

Itä-Suomen yliopisto, Historia- ja maantieteiden laitos

Tiedonlouhinta luo mahdollisuuksia maaseudun kehittämiseen

<https://doi.org/10.51807/maaseutututkimus.129322>

Tiedonlouhinta tarkoittaa isojen aineistojen ja niitä käsittelevien algoritmien yleistymistä tiedon tuottamisessa. Kun aineiston määrä ja koko kasvavat, eivät perinteiset analyysi- ja aineistonhallintatyökalut enää riitä aineistojen käsittelyyn eivätkä varsinkaan niiden sisällön ymmärtämiseen. Siksi isojen aineistojen analysointiin tarvitaan uusia menetelmäteknikoita, jotka jalostavat aineistoista helpommin käsiteltävissä olevaa tietoa erilaisista ilmiöistä tai niiden välisistä suhteista (Han ym. 2012).

Tässä kirjoituksessa havainnollistetaan tiedonlouhinnan mahdollisuuksia analysoimalla kuntastrategioiden sisältöä. Tavoitteena on kuntastrategioiden tekstinlouhinnalla havainnollistaa tiedonlouhinnan mahdollisuuksia – mutta samalla myös tarkastella kahdella esimerkillä sitä, miten maaseutu ilmenee kuntastrategioissa ja millaisia eroja kuntaluokkien välillä on kuntastrategioissa elinvoiman kehittämisessä. Kuntastrategian tarkoitus on ohjata kunnissa tapahtuvaa suunnittelua, päätöksentekoa ja kehittämistä, ja se on laadittava kunnissa vähintään kerran valtuuston toimikaudessa. Kuntalain (410/2015) mukaan kuntastrategiassa tulee huomioida muun muassa kunnan asukkaiden hyvinvoinnin edistäminen, palvelujen järjestäminen ja tuottaminen sekä elinympäristön ja alueen elinvoiman kehittäminen.

Kuntastrategioita on Suomessa tutkittu paljon viime vuosina (esim. Mötönen & Kettunen 2015; Kurvinen ym. 2022), mutta kuntien suuren

lukumäärän ja analysoitavan aineiston ison koon vuoksi yleensä on tarkasteltu vain osaa kuntastrategioista. Tekstianalytiikalla voidaan ilman rajoitteita analysoida suuria tekstiaineistoja, luokitella tekstidokumentteja ja tunnistaa niistä erilaisia ilmiöitä (Feinerer ym. 2008). Siten tekstinlouhinnan avulla voidaan kuntastrategioista löytää asioita ja kokonaisuuksia, jotka muuten peittyisivät suureen tekstimassaan tai eivät kävisi ilmi yksittäisten kuntastrategiadokumenttien analysoinnista.

Kuntastrategioiden tekstinlouhinta

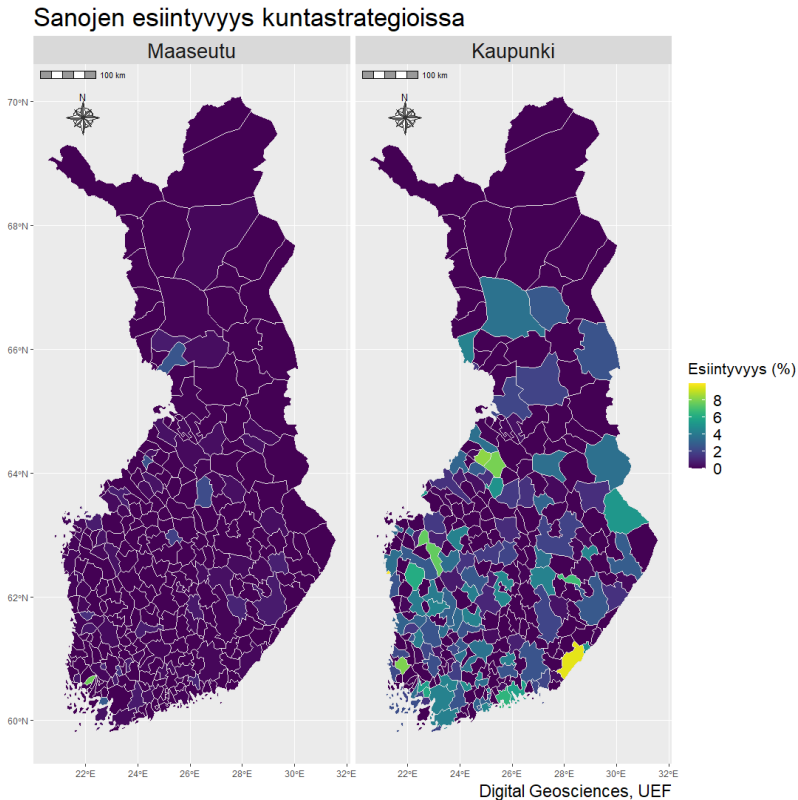
Tämän kirjoituksen esimerkissä kuntastrategioista luotiin korpus eli tekstidokumenttikokoelma, joka sisälsi 293 Manner-Suomen kuntastrategiaa. Kokoelmasta jätettiin pois Uusi-Kaarlepyy ja Ahvenanmaan kunnat, koska niiden strategiat olivat saatavilla vain ruotsinkielisinä. Keskimäärin kuntastrategioissa on 562 sanaa, mutta niiden pituus vaihtelee paljon. Pisimmässä kuntastrategiassa on 5 746 sanaa, ja lyhimmän pituus jää 38 sanaan. Laaja vaihteluväli strategioiden pituudessa paljastaa, että osa kuntastrategioista on varsin tiiviitä kuvauksia ja osa on puolestaan laajoja katsauksia. Kunnat näyttävätkin sijoittavan resursseja strategiatyöhön eri tavoin, mikä on puute kuntastrategioiden analysoinnista tehtäville tulkinnoille. Louhittavat kuntastrategiat ovat voimassa olevia, ja tyypillisin strategiakauden alkuvuosi on 2022.

Kuntastrategioiden muodostama tekstidokumenttikokoelma on lukumäärältään niin suuri, että sitä ei voida käsitellä perinteisin tutkimusmenetelmin, koska esimerkiksi aineiston lukeminen vaatisi paljon aikaa. Dokumenttikokoelman laajuuden vuoksi aineiston käsittelyssä käytetään tekstinlouhinnassa ohjelmointia, jotta isosta aineistosta saadaan irti oleellinen tieto. Tekstinlouhinnan etuja ovatkin aineiston koon rajoittamattomuus, käsittelyn nopeus ja analyysien toistettavuus (Kwartler 2017). Tekstinlouhinta suoritettiin R-tilasto-ohjelman tm-paketilla (Feinerer & Hornik 2023). Tekstidokumenttikokoelman muokkaamisen yhteydessä teksteistä poistettiin yksittäisiä kirjaimia, lyhenteitä, hukkasanoja (yleiset tai analysoinnin kannalta hyödyttömät sanat) sekä numeroita. Muokkauksen myötä erilaisten sanojen lukumäärä aineistossa väheni 10 956 sanasta 7 215 sanaan.

Kaupunki- ja maaseutu-sanojen esiintyvyys kuntastrategioissa

Ensimmäisessä esimerkissä tarkastellaan kaupunki- ja maaseutu-sanojen esiintyvyyttä kuntastrategioissa. Esiintyvyyttä kuvataan suhteellisenä muuttujana siten, että se kertoo sanojen suhteellisen osuuden kuntastrategian kaikkien sanojen kokonaismäärästä. Maaseutua kuvaaviksi sanoiksi valittiin ”maaseutu”, ”maaseudun” sekä ”maaseutukunta” ja kaupunkia kuvaaviksi ”kaupunki” ja ”kaupungin”.

Tekstinlouhinnan perusteella sana maaseutu esiintyy kuntastrategioissa selvästi harvemmin kuin sana kaupunki (kuvio 1). Kunnista peräti 216 jättää mainitsematta maaseudun kokonaan strategiassaan, kun vastavasti sanan kaupunki jättää mainitsematta vain 159 kuntaa. Keskimäärin sana maaseutu esiintyy jokaisessa kuntastrategiassa 0,8 kertaa ja vastavasti sana kaupunki 7,3 kertaa.



Kuvio 1. Sanojen maaseutu ja kaupunki esiintyvyys kuntastrategioissa. Kuvasta puuttuvat Uusi-Kaarlepyy ja Ahvenanmaan kunnat, koska näiden kuntien strategiat on julkaistu vain ruotsin kielellä.

Maaseutukäsitteen vähäisyys kuntastrategioissa ilmenee myös kuntatyyppeittäin. Kaupunkinimitystä käyttävistä kunnista 21 % ja kuntanimitystä käyttävistä kunnista 29 % mainitsee kuntastrategiassa sanan maaseutu. Vastaavat osuudet sanalle kaupunki ovat kaupungeissa 99 % ja muissa kunnissa 16 %. Paikkatietopohjaisen kuntaluokituksen kunnista (luokituksesta enemmän, Suomen ympäristökeskus 2020) maaseutu-sanana mainitsee strategioissaan 23 % harvaan asutun maaseudun kunnista, 25 % kaupunkien läheisen maaseudun kunnista ja 26 % ydinmaaseudun kunnista. Tulkintaa maaseudun vähäisestä esiintymisestä strategioissa vahvistaa se, että myöskään sanaa kylä ei yleensä mainita kuntastrategioissa, sillä sekin puuttuu 232 kuntastrategiasta. Vain ydinmaaseudun kunnissa kylä-sana esiintyy useammin kuin sana maaseutu, sillä se mainitaan 33 prosentissa ydinmaaseudun kuntien strategioita.

Havaintojen maaseutukäsitteen esiintyvyydestä voi tulkita tarkoittavan, että maaseutua ei nähdä kuntastrategioissa keskeisenä tekijänä kunnan hyvinvoinnin edistämiseksi, palvelujen tuottamisessa tai elinvoiman kehittämiseksi. Esimerkki myös havainnollistaa isojen aineistojen analyysin tuottaman lisäarvon: yksittäisten kuntastrategioiden sisältö muuttuu hyödylliseksi silloin, kun isosta tekstimassasta tunnistettuja asioita kuvataan kokonaisuutena. Havaintoa maaseudun vähäisestä esiintyvyydestä kuntastrategioissa olisi työlästä todentaa ilman tekstinlouhintaa.

Kuntaluokkien erilaiset painotukset elinvoiman kehittämisessä

Toisessa esimerkissä kuvataan tekstinlouhinnalla kuntastrategioiden sisältöä paikkatietopohjaisissa kuntaluokissa. Tarkastelussa hyödynnetään vertailusanapilveä, joka nostaa esille vain tietyille kuntaluokille tunnusomaisia sanoja (Fellows 2022). Kuntaluokittaisia eroja tulkitsemalla vertailusanapilvestä voidaan tunnistaa esimerkiksi niitä asioita, joiden varassa kuntien elinvoimaa kehitetään kuntaluokissa. Vertailusanapilvi myös varmentaa edellisen tarkastelun havaintoja. Sana maaseutu ei nouse esille yhdessäkään maaseudun kuntaluokassa, kun taas sanat kaupunki ja kaupunkistrategia ovat tyypillisiä kaupunkien kuntastrategioille (kuvio 2).

Vertailusanapilvi paljastaa, että kuntaluokissa painotetaan kuntastrategioissa eri sanoja elinvoiman kehittämisessä (kuvio 2). Tämä on luonnollista, koska eriytyvät haasteet ja mahdollisuudet elinvoiman kehittämisessä johtavat valintoihin, jotka näkyvät vertailusanapilvessä käsitteiden painotuseroina kuntaluokittain. Esimerkiksi kansainvälisyys painottuu kaupunkeihin kuten myös ympäristöasioita kuvaavat sanat hiilineutraalisuus ja monimuotoisuus. Myös koulutus ja osaaminen ovat strategioissa tunnusomaisia kaupungeille – samoin kasvu, vetovoima ja houkuttelevuus. Kaikki edellä luetellut elinvoimaan kytkeytyvät sanat ovat liitettävissä asioihin, joihin ladataan oletuksia tulevaisuuden talouskasvusta.



Kuvio 2. Vertailusanapilvi kuntastrategioista kuntaluokittain. Sanat on esimerkissä tyypistetty R-tilasto-ohjelman Snowball-paketilla ja kuva tehty Wordcloud-paketilla.

Maaseutukuntien kuntastrategioissa elinvoima rakentuu vertailusanapilvessä eri sanoin kuin kaupungeissa (kuvio 2). Harvaan asutun maaseudun kunnat painottavat strategioissaan esimerkiksi matkailua, luontoa ja luonnonvaroja, jotka viittaavat perinteisiin toimialoihin harvaan asutun maaseudun kuntien elinvoiman kehittämisessä. Ydinmaaseudun kunnille puolestaan tunnusomaista ovat etätyö, elin- ja asuinympäristöt, koulu, lähipalvelut, joustavuus ja kaavoitus, jotka kytkevät kuntien elinvoiman kehittämisen niiden houkuttelevuuden parantamiseen. Kaupunkien läheisen maaseudun kunnissa korostuvat vuorovaikutus, maaseutu, taajama, digitalisaatio, tieto ja markkinointi. Näistä käsitteistä useimmat korostavat elinvoiman kehittämisessä kuntaluokan sijaintia lähellä kaupunkeja ja kytkentää kaupunkeihin.

Kuntastrategioista voidaan tekstinlouhinnalla selvittää myös kaikille kunnille yhteisiä sanoja kuntien elinvoiman kehittämiseksi. Louhinnan perusteella tunnusomaisimpia elinvoiman kehittämiseen liittyviä sanoja olivat edellytysten luonti, elinkeinoelämä, maankäyttö ja investoinnit. Nämä sanat viittaavat kuntien yleisiin elinkeinopolitiikan osa-alueisiin.

Isojen aineistojen hyödyntämiseen keskittyvä ”teemaverkosto”?

Kaikkiaan aineiston analyysin tuloksia voi tulkita niin, että maaseutukunnissa kuntastrategiat painottavat kuntien paikallisia ominaisuuksia ja vahvuuksia, mutta maaseutua ei nähdä merkittävänä osana kuntastrategioita tai paikallisen elinvoiman kehittämistä. Merkittävä havainto kuntastrategioista on, että tulevaisuuden kasvutekijät, kuten kansainvälisyys ja useat ympäristöön liittyvät käsitteet, näyttävät olevan tunnusomaisia vain kaupungeille. Maaseutukuntien kuntastrategioissa tunnistetaan ja niihin valikoidaan paikallisia vahvuuksia, mutta strategioiden tekstinlouhinnan perusteella kytkentää megatrendeihin, kuten vihreään siirtymään, tulisi vahvistaa.

Kuntastrategioiden analyysi havainnollistaa uusien analyysimenetelmien ja aineistojen potentiaalia maaseutututkimuksessa ja -kehittämisessä, kun suuria aineistoja voidaan tarkastella kokonaisuutena. Tämä muuttaa määrällisen ja laadullisen tutkimuksen suhdetta, sillä laadullisessa tutkimuksessa käytettyjen näytteiden tai otoksien sijaan voidaan hyödyntää kokonaistutkimusta, kun aineistojen koko ei pakota rajaamaan aineistoa. Näin tiedonlouhinta luo mahdollisuuksia, joilla voidaan viestiä nykyistä kokonaisvaltaisemmin maaseudun asemasta laajoihin empiirisiin

aineistoihin tukeutuen. Isoista aineistoista jalostettu tieto monipuolistaa tilannekuvaava maaseudun olosuhteista, kehityksestä ja kehittämisessä tehdyistä valinnoista. Se myös nostaa esille keskusteluun maaseudun kuntastrategioiden mahdollisia puutteita.

Tähän asti maaseutupolitiikan verkostot ovat toimineet erilaisten ilmiöiden ja teemojen ympärillä. Isojen aineistojen ja erilaisten tiedonlouhintamenetelmien yleistyessä katsetta voitaisiin suunnata myös aineistoihin ja niiden hyödyntämisen tehostamiseen, jotta ne voidaan valjastaa palvelemaan maaseudun kehittämistä. Voisiko tässä olla tilaisuus uusien data-analytiikan suuntauksia korostavalle ”teemaverkostolle”? Koordinoitu verkosto kokoaisi yhteen eri organisaatioissa hajallaan olevaa osaamista, vahvistaisi maaseutututkimusta ja syventäisi tiedolla johtamista.

Kirjoitus on osa maa- ja metsätalousministeriön rahoittamaa tutkimushanketta *Julkisten rahavirtojen kohdentuminen maaseutu- ja kaupunkialueille*. Rahoitus on myönnetty Makeran valtakunnallisiin maaseudun tutkimus- ja kehittämishankkeisiin suunnatuista varoista.

Kirjallisuus

- Han, J., M. Kamber & J. Pei 2012. Data mining – Concepts and techniques. Morgan Kaufmann, Waltham.
- HeFeinerer, I., K. Hornik & D. Meyer 2008. Text mining infrastructure in R. *Journal of Statistical Software* 25:5, 1–54. <https://doi.org/10.18637/jss.v025.i05>
- Kurvinen, A., A. Jolkkonen, V. Lemponen & P. Kahila 2022. Pienten kuntien strategiat ja luovat ratkaisut. Kunnallisan alan kehittämistäitiön Tutkimusjulkaisu-sarja, nro 114.
- Kwartler, T. 2017. Text mining in practice with R. John Wiley & Sons.
- Möttönen, S. & P. Kettunen 2015. Kuntien strategiadokumentit: Systemiteoreettista pohdintaa suurten kaupunkien strategioista. *Kunnallistieteellinen aikakauskirja* 43:2, 126–143.

Lähteet

- Feinerer, I. & K. Hornik 2023. Package tm. Saatavissa: <https://cran.r-project.org/package=tm> [Viitattu 28.4.2023.]
- Fellows, I. 2022. Package Wordcloud. Saatavissa: <https://cran.r-project.org/package=wordcloud> [Viitattu 28.4.2023.]
- Kuntalaki 410/2015. Saatavissa: <https://www.finlex.fi/fi/laki/ajantasa/2015/20150410> [Viitattu 27.4.2023.]
- Suomen ympäristökeskus 2020. Kaupunki-maaseutuluokitus. 30.11.2020. Saatavissa: <https://www.ymparisto.fi/kaupunkimaaseutuluokitus> [Viitattu 28.4.2023.]