

HUOM! Tämä on alkuperäisen artikkelin rinnakkaistallenne. Rinnakkaistallenne saattaa erota alkuperäisestä sivutukseltaan ja painoasultaan.

Käytä viittauksessa alkuperäistä lähdettä:

Kauttonen, J., Puhakainen, E. & Vehkaperä, M. (25.05.2022). Korona muutti MaRa-alan työvoimatarpeita. eSignals PRO. <http://urn.fi/URN:NBN:fi-fe2022052538799>

PLEASE NOTE! This is an electronic self-archived version of the original article. This reprint may differ from the original in pagination and typographic detail.

Please cite the original version:

Kauttonen, J., Puhakainen, E. & Vehkaperä, M. (25.05.2022). Korona muutti MaRa-alan työvoimatarpeita. eSignals PRO. <http://urn.fi/URN:NBN:fi-fe2022052538799>



Copyright: © 2022 by the authors and Haaga-Helia University of Applied Sciences. Licensed under the terms and conditions of the Creative Commons Attribution (CC BY NC SA) license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Korona muutti MaRa-alan työvoimatarpeita

Kirjoittajat: Janne Kauttonen, Eeva Puhakainen ja Meri Vehkaperä

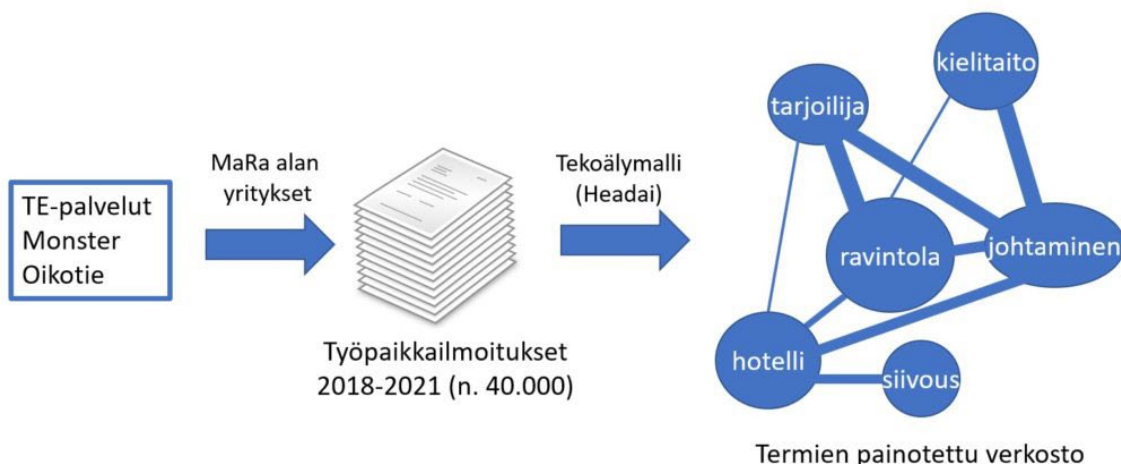
Työpaikkailmoitukset peilaavat hyvin työmarkkinoiden tarpeita. Data on täysin avointa ja sitä on paljon. Tämä oli erinomainen lähtökohta tekoälyn hyödyntämiselle MODULE-hankkeessa, kun halusimme selvittää, miten korona vaikutti MaRa-alan yritysten osaamistarpeisiin vuosina 2018–2021.

Tässä artikkelissa analysoidaan MaRa-alan työpaikkailmoituksia tekoälyn tuottaman datan avulla. Aineisto tuotettiin MODULE-tutkimushankkeen yhteydessä. Euroopan sosiaalirahaston rahoittamassa hankkeessa tunnistetaan MaRa-alan nykyiset ja tulevaisuuden osaamistarpeet sekä kehitetään joustavia modulaarisia koulutuksia vastaamaan näihin tarpeisiin. Hankkeessa luotu koulutusmalli tukee matkailusektorin työvoiman osaamisen kehittämistä sekä nuorten työllistymistä COVID-19 kriisissä, ja tarjoaa syksyllä 2022 ilmaisia koulutuksia työntekijöille.

Työpaikkailmoitusten tekstit verkostomuotoiseksi

Dataa koostimme ja laskimme yhteistyössä suomalaisen tekoäly-yrityksen, [Headain](#) kanssa. Ensin MaRa-alan yrityksistä tehtiin koko Suomen kattava lista (noin 3500 suurinta yritystä), minkä jälkeen kolmen työpaikkafoorumin (Monster, MOL ja Oikotie) materiaaleista poimittiin MaRa-alan yritysten jättämät työpaikkailmoitukset vuosilta 2018-2021. Työpaikkailmoituksia tunnistettiin seurantajaksolla yhteensä noin 40 000 kappaletta.

Työpaikkailmoitukset analysoitiin Headain tekoälymallilla, joka tuotti jokaisen kuukauden datasta niin sanotun painotetun verkoston. Verkot sisälsivät sanoja ja käsitteitä (termistö), jotka liittyvät työelämään ja osaamisiin. Kaikilla termeillä ja termien välisillä suhteilla on määritelty painotus, joka kertoo miten tärkeä tai vahva kyseinen termi tai niiden välinen kytkös on. Termejä oli yhteensä vajaat kolmetuhatta, joista merkittävä paino oli noin parilla sadalla. Prosessia on havainnollistettu kuvassa 1.



Kuva 1. MaRa-alan työpaikkailmoitusten kerääminen ja tekoälymallin tuottama verkostodata. Kuvassa näkyvä verkosto on vain havainnekuva, oikea verkosto sisälsi tuhansia termejä.

MODULE-hankkeessa on jo aiemmin sovellettu Headain tekoälymallia haastatteluaineiston sisällönanalyysissä (ks. aiempi blogiteksti). Tällä kertaa saatu data oli kuitenkin monimutkaisempaa, koska siinä oli mukana sekä aikariippuvuus että sanojen väliset riippuvuudet. Painotettua verkostoa voidaan analysoida monilla eri tavoilla, kuten visualisoimalla, ryhmittelyllä ja termien aikasarjoja tutkimalla.

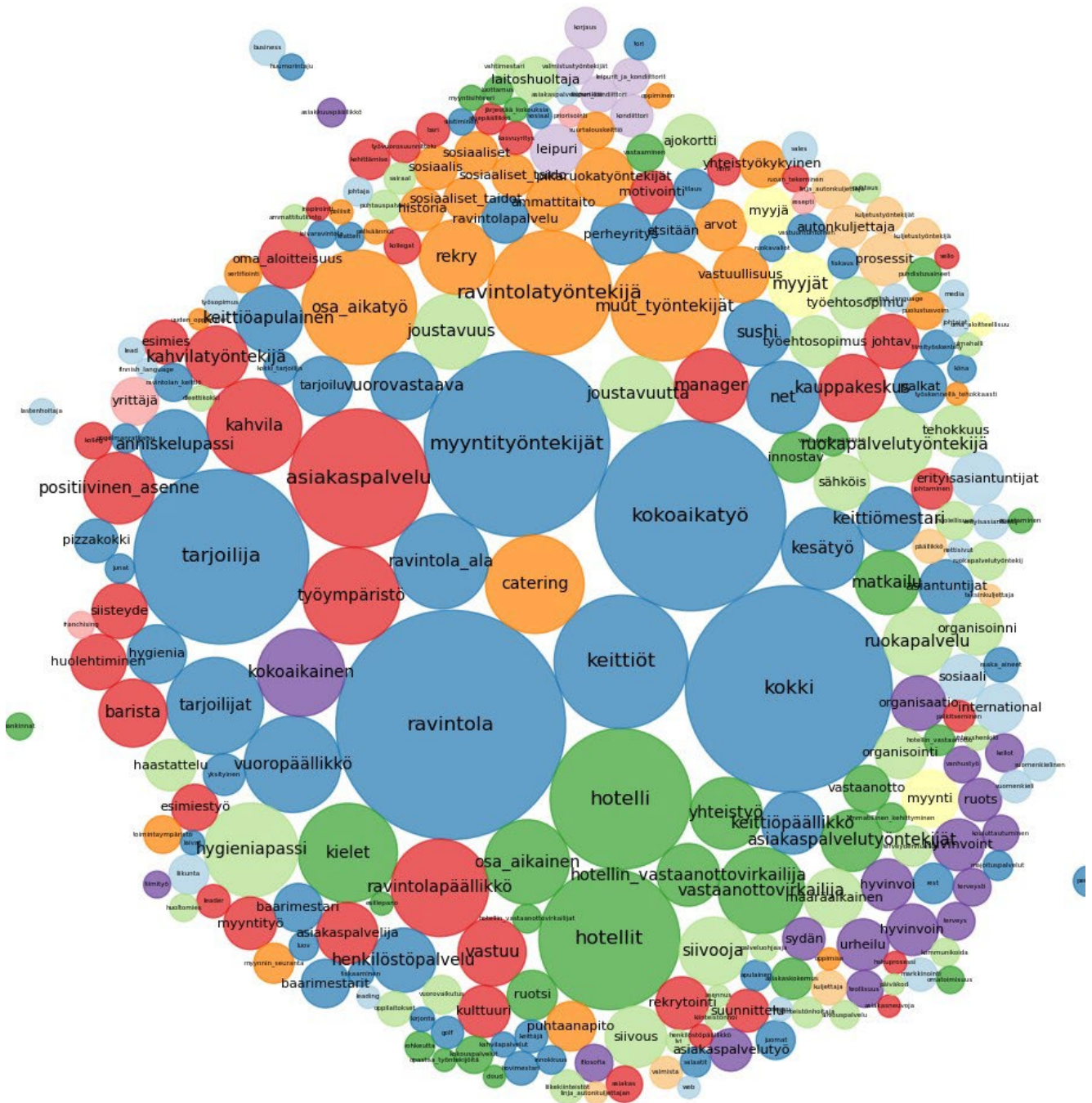
Sisällönanalyysia on perinteisesti käytetty tilanteissa, jossa laajoja dokumenttiaineistoja käsitellään määrällisen, jopa tilastollisen analyysin avulla. Usein aineistosta koodataan sisältöä koskevia luokkia, joiden välisiä suhteita sitten tarkastellaan. Tämän tyyppistä analyysia tehdään nykyään yhä useammin tietokoneavusteisesti eli erilaisten algoritmien avulla. (Kallinen & Kinnunen 2022.)

MaRa-alan yleiskuva verkostanalyysin keinoin

Lähtökohtana oli kuvata, millaisina MaRa-alan työvoimatarpeet näyttäytyivät keskimäärin vuosina 2018–2021. Syntynyt kuukausittainen keskiarvoverkko visualisoitiin käyttäen hyväksi simulaatiomallia (ks. <https://d3js.org>), jossa termit esitetään kaksikulotteisina kiekkoina. Tässä simulaatiossa kiekot ja kiekkojen väliset vetovoimat määräytyvät verkon painotuksista ja kiekot hakeutuvat niiden voimien määräämällä tietyille paikoille. Kiekot eivät saa olla toistensa päällä.

Lopputuloksena saadaan selkeä visuaalinen kuvaus, joka noudattaa hyvin datan ominaispiirteitä. Heikkoutena on menetelmän laskennallisten työläys ja se, että jokainen simulaation lopputulos on aina hieman erilainen. Tämä on seurausta siitä, että laskennallisesti yhtä hyviä loppuratkaisuja – eli kiekkojen koordinaatteja – on useita (ns. lokaalin minimin ongelma).”

Lisäksi käytimme erityistä ryhmittelyalgoritmia (<https://github.com/vtraag/leidenalg>), joka automaattisesti ryhmittelee termit siten, että vahvimmin toisiinsa kytköksissä olevat termit muodostavat oman joukkonsa, eli kiekkoilla oli myös ”väri”. Simulaation lopputuloksena tärkeimmät termit eli ilmoitusten yleisimmät sanat (kiekot) hakeutuvat aina kuvan keskelle, sekä saman ryhmän kiekkojen läheisyyteen. Vähemmän tärkeät termit jäävät reunoille. Simulaation ja ryhmittelyn lopputulos näkyy kuvassa 2.



Kuva 2. MaRa-alan ydintermit vuosien 2018–2021 työpaikkailmoituksissa

Kuvasta 2 huomataan, että keskeisimmät termit ovat olleet esimerkiksi **tarjoilija**, **ravintola**, **kokki**, **kokoaikatyö**, **asiakaspalvelu**, **myyntityöntekijät**, **hotelli ja catering**. Tulos oli odotettu ja kuvaa hyvin MaRa-alan ydintä. Monet tärkeistä termeistä ovat toisaalta generisiä ja alasta riippumatonta työpaikkailmoitusten sanastoa, kuten **kokoaikatyö**, **osa-aikatyö**, **rekry** ja **organisaatio**.

Tekoäly ei osaa erottaa, mikä osa työpaikkailmoituksen tekstistä koskee hakijalta edellytettäviä osaamisvaatimuksia ja mikä puolestaan on yrityksen esittelyä tai markkinointitekstiä, eli se ottaa mukaan kaikki ilmoituksessa olevat sanat.

MODULE-hankkeen kannalta oli erityisen kiinnostavaa tarkastella koronan vaikutusta, eli tapahtuiko vuosien 2018–2019 ja 2020–2021 välisenä aikana merkittäviä muutoksia. Analyysi toistettiin sen vuoksi jokaiselle vuodelle erikseen.

Yleisesti ottaen eri vuosien kiekkokuvat ovat keskenään samankaltaisia, eli aiemmin kuvassa 2 nähdyt tärkeät termit (isot kiekot) ovat keskeisiä. Verkoston ydin pysyi siis pääosin samanlaisena, mutta tuloksissa tunnistimme kaksi selkeää muutosta. Mukaan ovat ilmestyneet sanat **terveys**, **urheilu**, **hyvinvointi** ja **vanhustyö** omana selkeänä ryhmänään, sekä **vastuullisuus** ja **joustavuus**, jotka ovat nousseet tärkeämpään rooliin eli siirtyneet hieman keskemälle. Muutokset voivat johtua paitsi koronapandemian tuottamasta aidosta muutoksesta kysynnässä, myös siitä, että työmarkkinoille on ilmestynyt uusia rekrytoivia yrityksiä, joiden ilmoituksissa kyseiset sanat korostuvat. Siksi termeille tehtiin myös erillinen trendianalyysi, jolla voidaan vahvistaa havaittu trendimuutos myös kvantitatiivisesti.

Trendianalyysi ja korona-ajan vaikutus yksittäisiin termeihin

Saimme viitteitä siitä, että tiettyjä korona-ajan tuottamia eroja on havaittavissa. Trendianalyysia varten laskimme jokaiselle termille kuukausittaisen esiintymistiheyden vuosille 2018–2021, minkä jälkeen aikasarjalle tehtiin paikallisesti sovitettu ja painotettu regressiosovitus (LOWESS-menetelmä). Sovitus tehtiin vain sellaisille termeille, jotka esiintyivät riittävän usein aineistossa. Tällaisia ydintermejä oli noin 300 kappaletta.

Sovituksen perusteella voidaan arvioida termin esiintymistiheyskehitystä lineaarisesti eteenpäin kuusi kuukautta (kesä 2022). Lopuksi tätä ennustetta verrattiin perustasoon, joka termillä oli ennen koronaa eli vuosina 2018 ja 2019. Lopputuloksena syntyi jokaiselle termille prosentti, joka kuvaa termin trendin muutosta positiiviseen tai negatiiviseen suuntaan.

Viidenkymmenen tärkeimmän nousevan ja laskevan trendin termit on koostettu kuviin 3 ja 4. Näissä kuvissa termin sijainnilla ei ole merkitystä, mutta sen sijaan tekstin koko on suhteessa muutoksen prosenttiin (isompi viittaa suurempaan muutokseen). Kuvaan 3 on lisäksi merkitty edellä mainittuja terveyden ja vanhusten hoitoon liittyviä mielenkiintoisia termejä. Analyysin perusteella näyttää siltä, että nämä ovat tosiaan vahvassa nousussa.



Kuva 3. Top-50 nousevat termit. Näiden termien yleisyys MaRa-alan työpaikkailmoituksissa kasvoi korona-aikana verrattuna sitä edeltävään aikaan. Mitä suurempi fonttikoko, sitä suurempi kasvu. Kuvaan on ympyröity kiinnostavia termejä.

Laskevat termit liittyvät todennäköisesti suoraan korona-ajan ravintolasulkuihin ja muihin rajoituksiin (esimerkiksi etätyöt). Joukossa on kuitenkin myös yllättäviä termejä. Näyttää siltä, että korona-aikana ei kaivata **johtajia**, mutta **johtaminen** oli nousussa.

Eriyisasiiantuntijat ja **asiiantuntijat** saavat väistyä, kun alalle kaivataan **työkokemusta**, **asiakaspalvelutaitoja** ja **asennetta**.



Kuva 4. Top-50 laskevat trendit. Näiden termien yleisyys MaRa-alan työpaikkailmoituksissa laski korona-aikana verrattuna sitä edeltävään aikaan. Mitä suurempi fonttikoko, sitä suurempi lasku.

Kuten jo aiemmin mainitsimme (ks. [aiempi blogiteksti](#)), tekoälymalli on agnostinen analysoimansa tekstin kontekstin suhteen. Termeissä on siis mukana myös paljon ei-kiinnostavia ja erikoisiakin termejä, kuten termi sydän, joka nousi esiin vahvasti kasvavana trendinä kuvassa 3.

Muutosta selittää, että työpaikkailmoituksissa toistui usein esimerkiksi seuraavia fraaseja: ”Työntekijämme ovat toimintamme sydän”, ”Etsimme juuri sinua, jolle hyvän ruoan valmistaminen on sydäntä lähellä” ja ”Vastaanotto on hotellimme sydän”. Eli tässä tapauksessa termi viittaakin muutokseen työpaikkailmoitusten viestinnässä (kenties tiettyjen yritysten osalta), ei varsinaisissa osaamisissa tai taidoissa. Kuten kaikessa muussakin data-analyysissä, myös tekoälymallin tapauksessa tulokset on aina syytä käydä kriittisesti läpi ja datan erityispiirteet on tunnettava.

Monipuoliset menetelmät antavat luotettavan kuvan

Tässä kuvattujen menetelmien soveltaminen ja tulosten tulkinta vaatii asiantuntemusta, eikä kirjoituksessa ole avattu kaikkia pieniä, mutta oleellisia välivaiheita. Yleensä nämä liittyvät datan esikäsittelyyn, algoritmien ja parametrien valintaan, ja niillä voi olla merkittävä vaikutus lopputuloksiin. Valmiista tekoälyn laskemasta verkostodatasta oli vielä verrattain pitkä matka varsinaisten päätelmien ja tulkintojen tekemiseen.

Yhteenvetona voidaan todeta, että työpaikkailmoitukset ovat hyvä datalähde kvantitatiiviseen ja tekoälypohjaiseen analyysiin. Tämän aineiston pohjalta saimme mielenkiintoisen näkökulman MaRa-alaan, ja analyysissä paljastui monia kiinnostavia muutoksia ja trendejä, joita voimme huomioida koulutuksen suunnittelussa.

Toisaalta työpaikkailmoitukset tarjoavat vain epäsuoraa kuvaa työmarkkinoiden kysynnästä ja tarpeista, ja vain sillä hetkellä kuin ilmoitus on laitettu. Etenkin jos halutaan katsoa pidemmälle tulevaisuuteen, on syytä hyödyntää muitakin lähteitä, kuten alan asiantuntijoiden haastatteluja ja kyselytutkimuksia.

Lähde:

Kallinen, T. & Kinnunen, T. 2022. [Laadullinen sisällönanalyysi](#). Teoksessa Jaana Vuori (toim.) Laadullisen tutkimuksen verkkokäsikirja. Tampere: Yhteiskuntatieteellinen tietoarkisto.