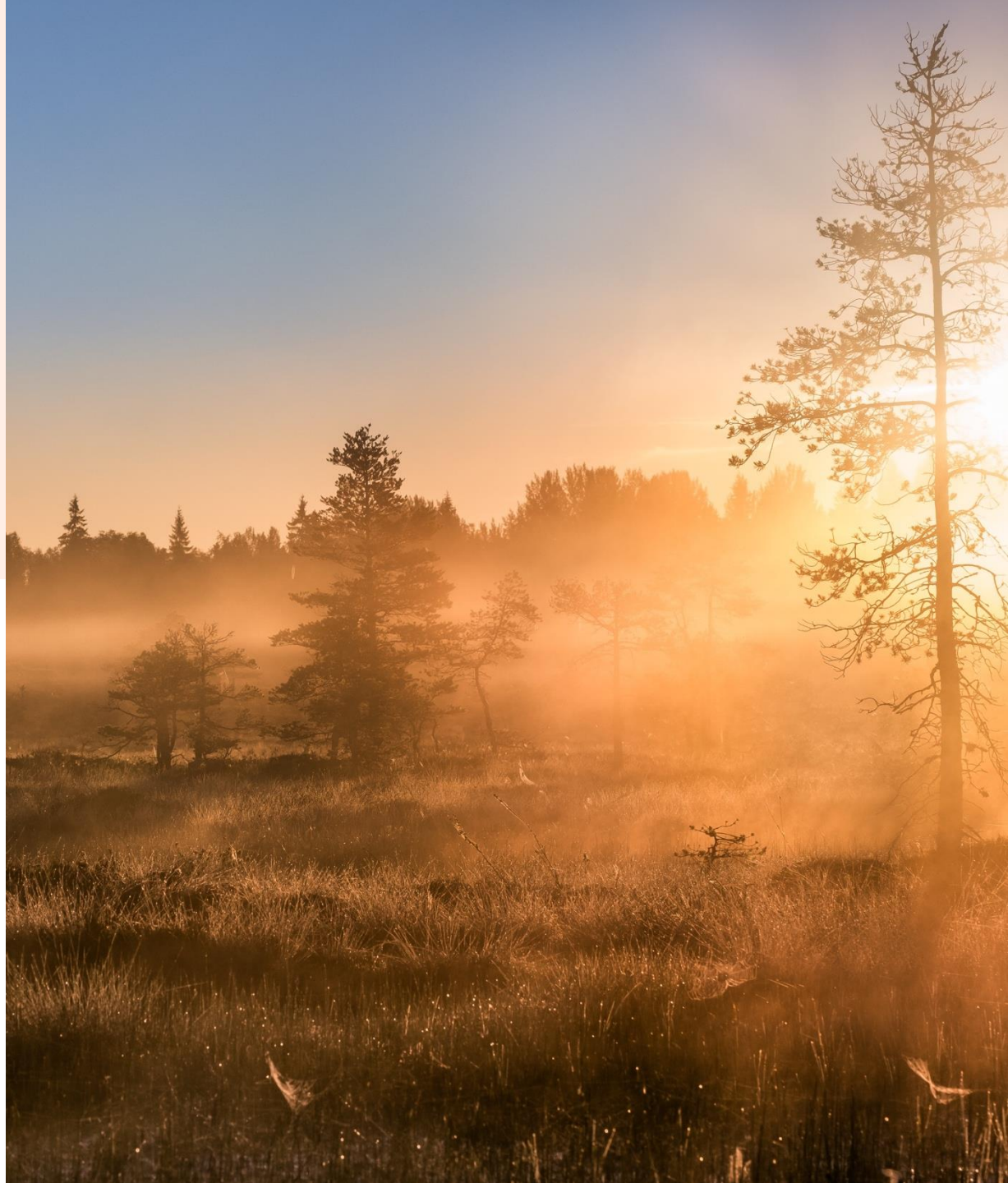
Pilottitutkimus: Avointen vastausten analysointi

Jason Theodoropoulos

24.3.2026

Agenda

1. Tausta
2. Analyyseista, metodologiaa
3. Tulokset ja yhteenveto

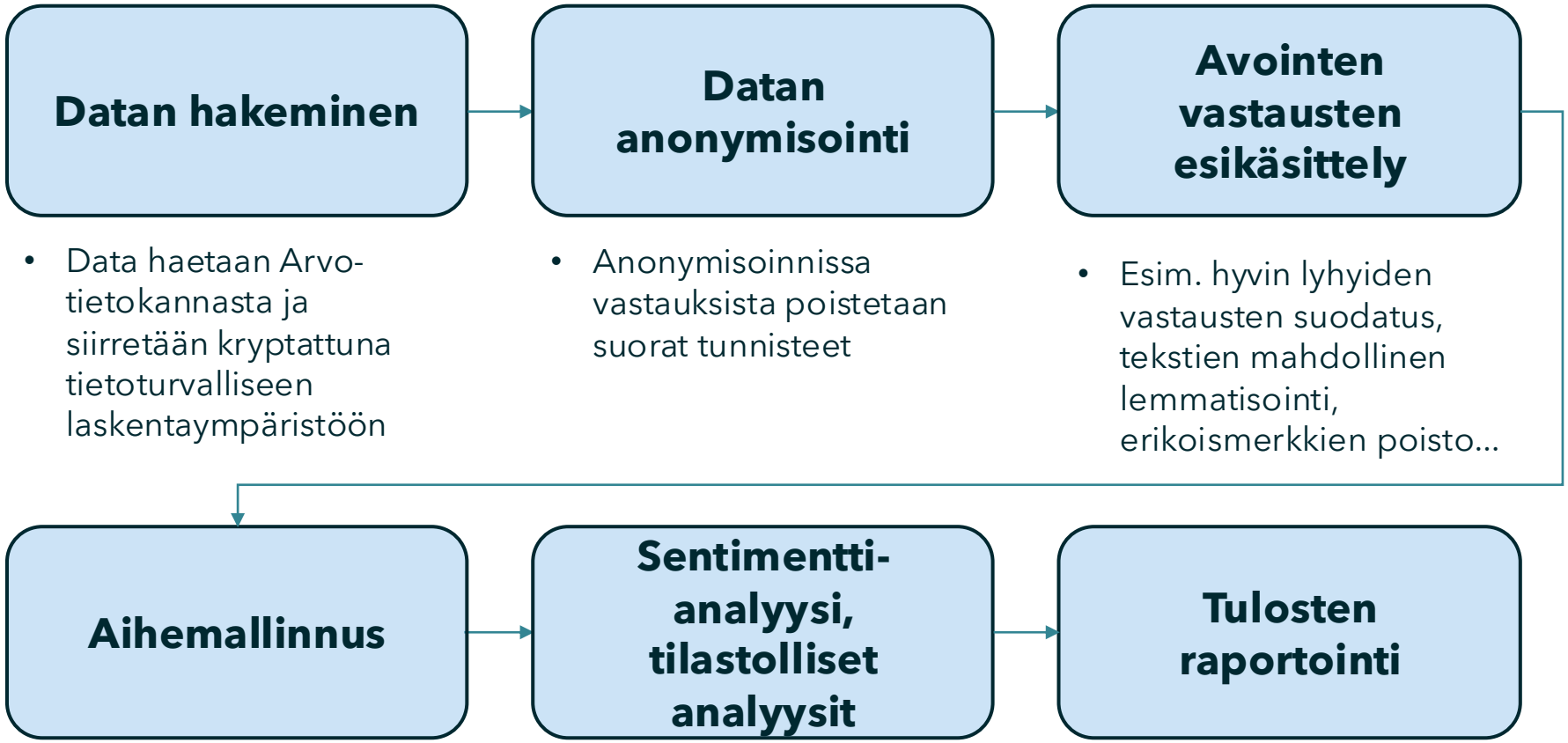


Tausta

Taustaa

- CSC:llä toteutetaan jo uraseurantakyselyiden monivalintakysymysten vastausten analysointi ja raportointi, mutta avointen vastausten analysointia ei ole vielä tehty valtakunnallisen tason aineistolla
- Pilotin tavoitteena oli tarkastella koneoppimis- ja tekoälymallien soveltuvuutta luotettavan ja toistettavan analyysin toteuttamiseen avoimen lähdekoodin ratkaisuin ilman julkipilvessä toimivia työkaluja
  - Aineistona hyödynnettiin kaikkia vuoden 2024 kyselyn vastauksia, ja työssä kehitettiin ratkaisuja, joilla analyysin voi toistaa vertailukelpoisesti myös muiden vuosien aineistoille
- Analyyseissa pyrittiin tunnistamaan vastauksissa esiintyviä aiheita (aihemallinnus/~klusterointi), ja edelleen tarkastella esimerkiksi minkä sävyistä palautetta aiheesta on annettu, ja ovatko tietyt aiheet tyypillisempiä tietyille vastaajaryhmille
  - Pilotissa tarkasteltiin myös työkalujen ja aineiston mielekkyyttä analyysiin yleisesti; onko aineistosta tunnistettavissa mielekkäitä aiheita yhteismitallisesti esimerkiksi eri korkeakoulujen ja kielten yli
- Analyysia kehitettiin yhteistyössä uraseurannan parissa työskentelevien asiantuntijoiden kanssa

Analyysin vaiheet karkeasti esitettynä



- Data haetaan Arvo-tietokannasta ja siirretään kryptattuna tietoturvalliseen laskentaympäristöön
- Anonymisoinnissa vastauksista poistetaan suorat tunnisteet
- Esim. hyvin lyhyiden vastausten suodatus, tekstien mahdollinen lemmatisointi, erikoismerkkien poisto...
- Vastausten ryhmittely yhteisten aiheiden perusteella, aiheiden tulkitseminen / luotettavuuden arviointi
- Muiden analyysien toteuttaminen, myöskin esim. ammattinimikkeiden vertailu ajanjakson yli
- Raportti tuloksista ja datan käsittelystä PDF + diamuodossa, myöskin ymmärrys ja valmius pilotin monistamiseen / laajentamiseen

Analyyseista, metodologiaa

Anonymisoinnista

Tekaistu esimerkkipalaute

Opetus loistavaa, kiitos opetuksesta etenkin Essi Esimerkkinen! XAMK oli mainio opinahjo, työllistyin pian valmistumisen jälkeen ja nykyään asun Tampereella.

FINBERT-NER →

Anonymisoitu esimerkkipalaute

Opetus loistavaa, kiitos opetuksesta etenkin **[poistettu]**! **[poistettu]** oli mainio opinahjo, työllistyin pian valmistumisen jälkeen ja nykyään asun **[poistettu]**.

- Avoimet vastaukset anonymisoitiin Kansallisarkiston ja TurkuNLP-ryhmän kehittämällä FINBERT-NER –mallilla, joka tunnistaa suomalaisia erisnimiä
- Mahdollisten epäsuorien tunnisteiden poistaminen vastauksista on kuitenkin mahdotonta, joten aineistoa käsiteltiin anonymisoinnin jälkeenkin tietoturvallisessa ympäristössä

Aihemallinnus

Tyypillisiä vastaustyyppejä

Työllistyin tutkintoa vastaavaan työhön heti...

Työskentelen yhä kouluprojektin kautta saamassani työssä...

Työllistyin opiskelujen yhteydessä...

•
•
•

→

Kuvaavimmat sanat vastauksissa

työllistyin
heti
pääsin
jälkeen
työpaikan
ennen
edennyt
valmistuttuani
työllistynyt

•
•
•

→

Aiheen tunnistaminen

”1_työllistyin_pääsin_heti_sain”
-> Nopeasti työllistyneet

- Aihemallinnuksessa koneoppimismalli ryhmittelee vastauksia samankaltaisen sanaston perusteella yhteisiin aiheisiin
- Pilotissa BERTopic-työkalua hyödynnettiin monikielisellä BERT-kielimallilla, joka tunnistaa samankaltaiset sanat myös niin suomeksi, ruotsiksi, kuin englanniksikin kirjoitetuista vastauksista
- Lopulta tyypillisimmistä vastauksista ja sanoista voidaan päätellä aiheen sisältö, ja nimetä aihe

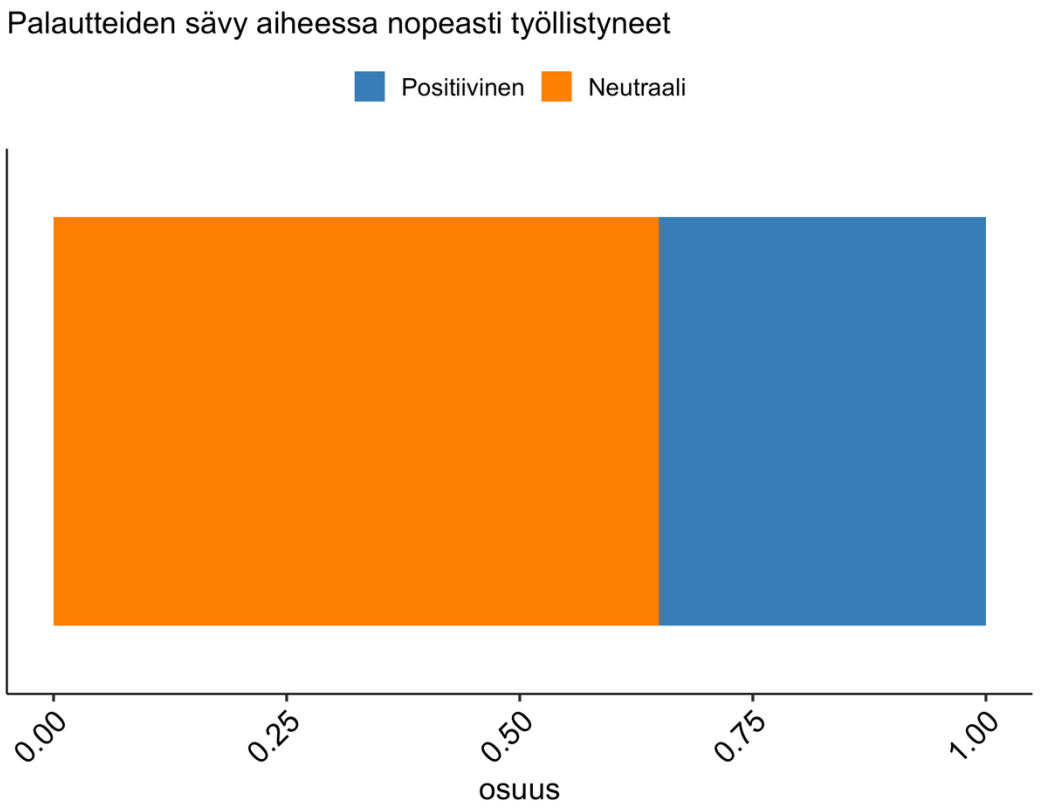
Aihemallinnus

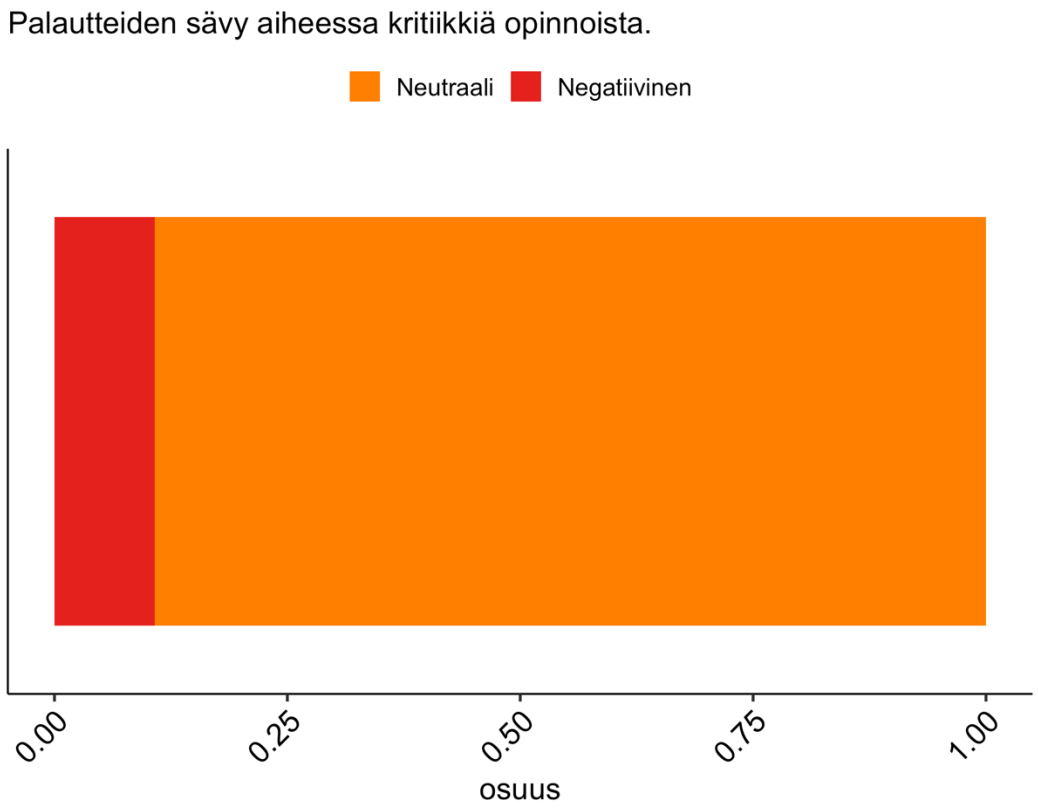
”2_ilman_tutkintoa_degree_saanut”
”3_antoi_hyvän_tutkinto_pohjan”
”10_avasi_uusiin_parani_palkkaus”

Tutkinto auttanut uralla

- Osassa aiheista toistuu samankaltaiset palautteet eri sanoituksin ja aiheita on myös yhdistelty tuloksien selkeyttämiseksi

Sentimenttianalyysista



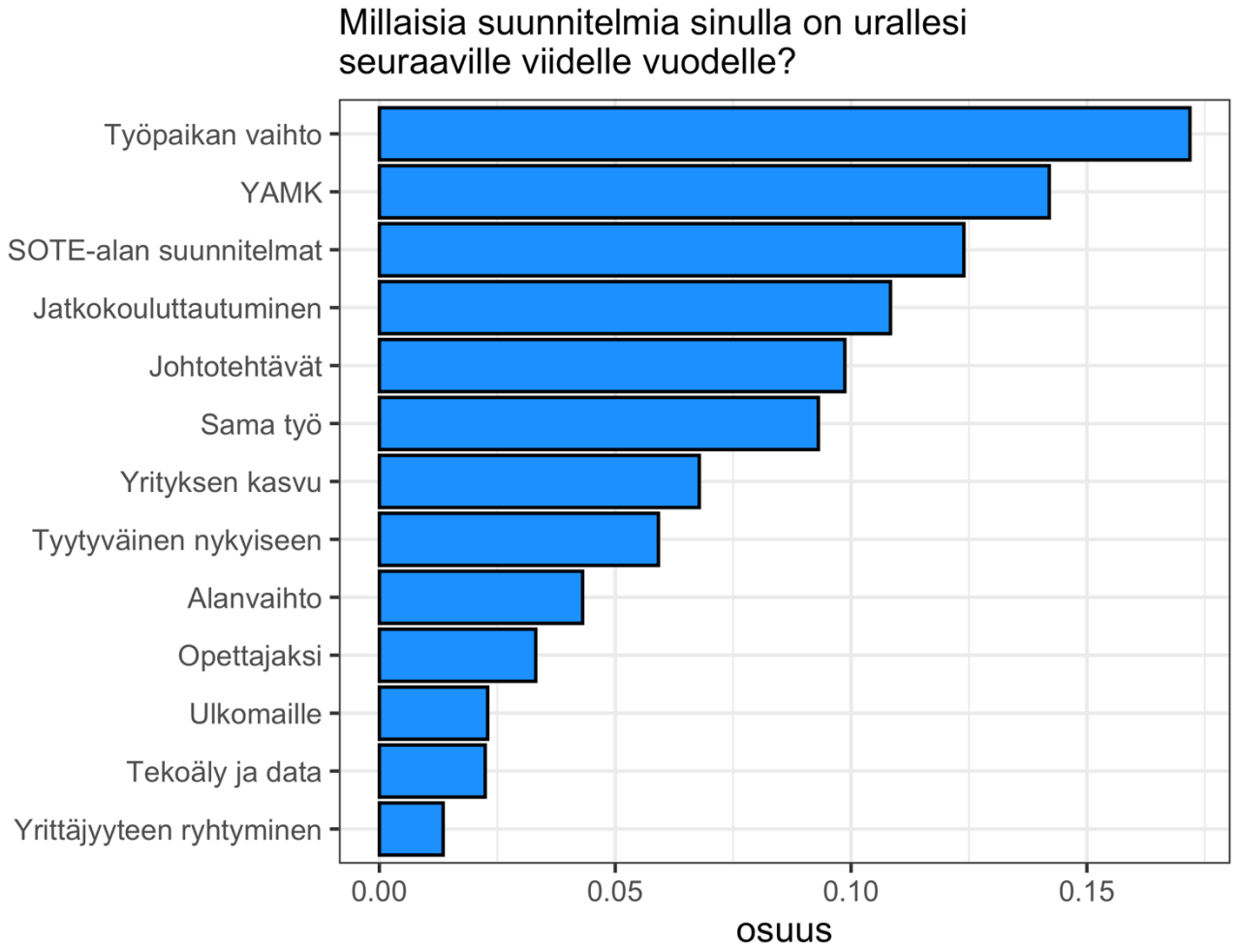


- Aiheita voidaan myös analysoida teksin sentimentin, siis sävyn, perusteella
- Esimerkiksi valituissa aiheissa palautteiden sävyn voisi arvatakin: nopeasti työllistyneet ovat antaneet suopeamman sävyistä palautetta kuin opintoja kritisoineet

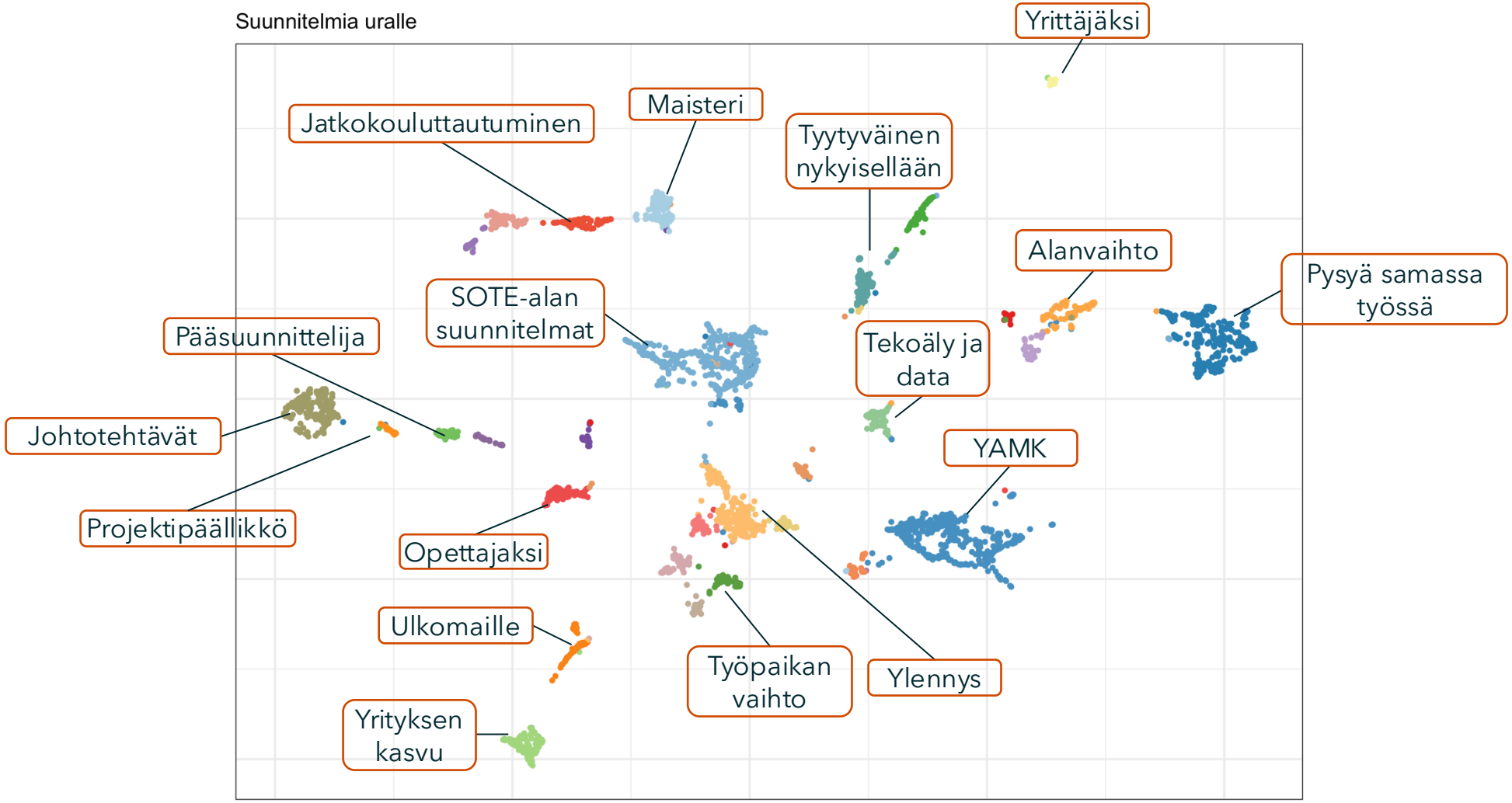
Tulokset

Aihemallinnuksen tuloksista

- Aihemallinnusta hyödynnettiin eri kysymysten vastauksiin, ja pilotissa valittiin tarkempaan tarkasteluun urasuunnitelmiin liittyvä kysymys
  - Kysymyksestä tunnistettiin selkeitä ja intuitiivisia aiheita
- Osaa aiheista pitäisi pilkkoa edelleen pienemmiksi: esim. SOTE-alan suunnitelmat sisältää varmasti monia muista aiheista, mutta yhdistävänä tekijänä mallille on ollut ammattisanasto
  - Mallia voisi kehittää jättämään ammattispesifin sanaston huomiotta, jolloin vastaukset luokiteltaisiin muihin aiheisiin

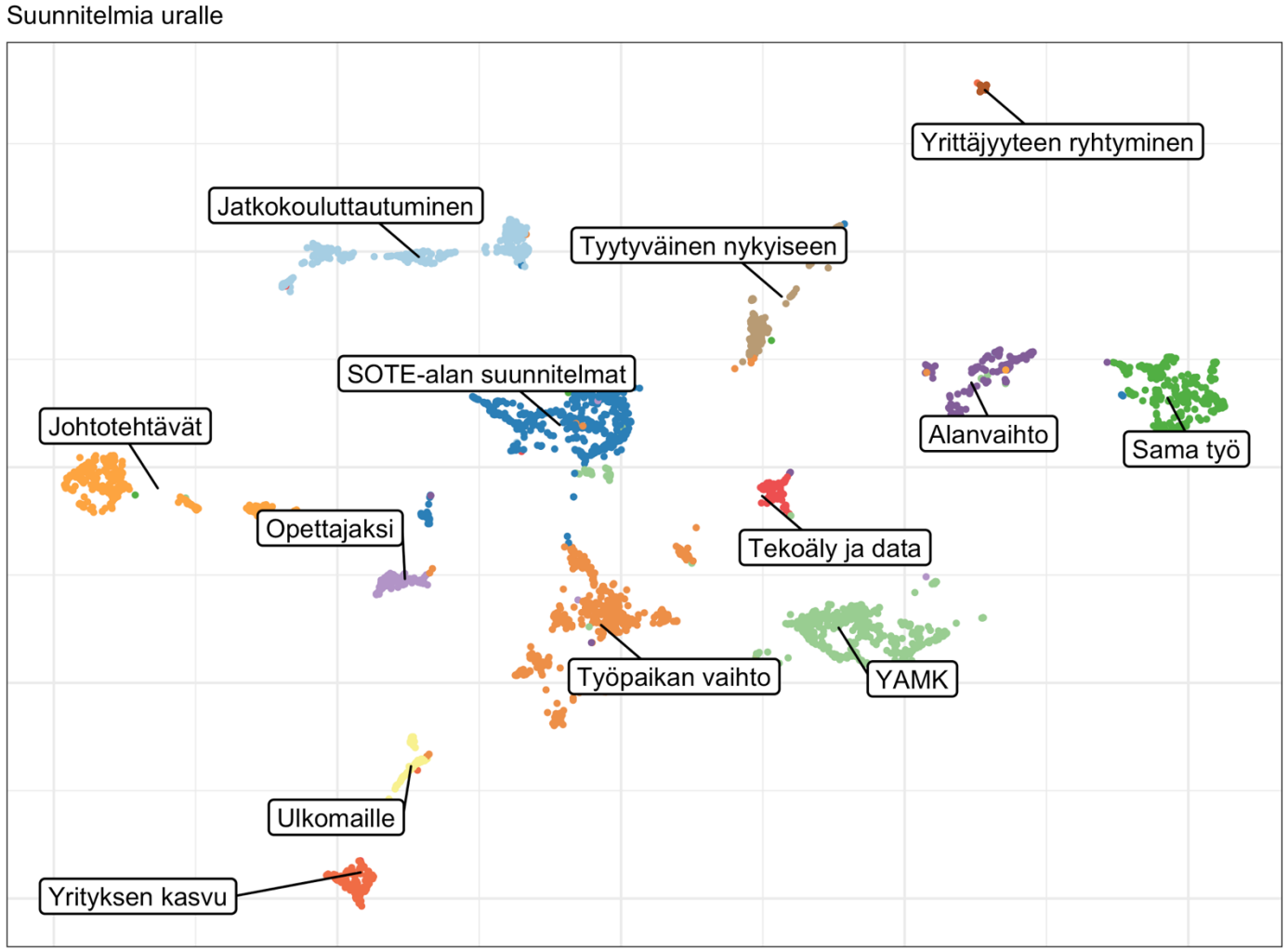


Aihemallinnuksen tuloksista



- Yllä on visualisoitu aihemallinnuksen tuloksia kysymykseen seuraavan viiden vuoden urasuunnitelmista. Kuvaajassa yksi piste vastaa yhtä vastausta. Visualisointi on hyödyllinen myös sopivien karkeampien klusterien määrän arvioimikseksi

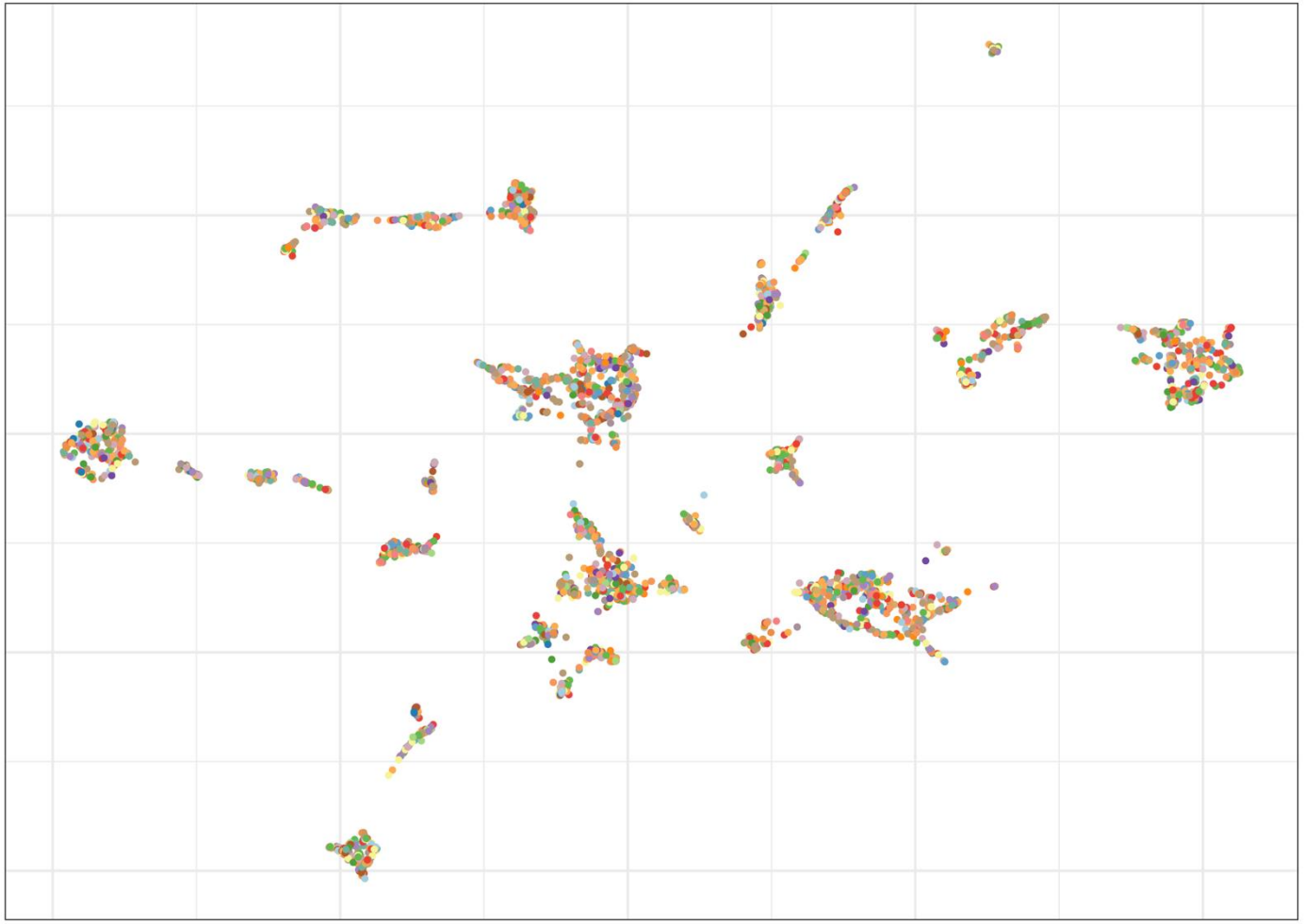
Aihemallinnuksen tuloksista



- Edellisen dian klustereista on luotu yleisempiä, vastauksia kuvaavia, klustereita, joita voidaan tarkastella tarkemmin

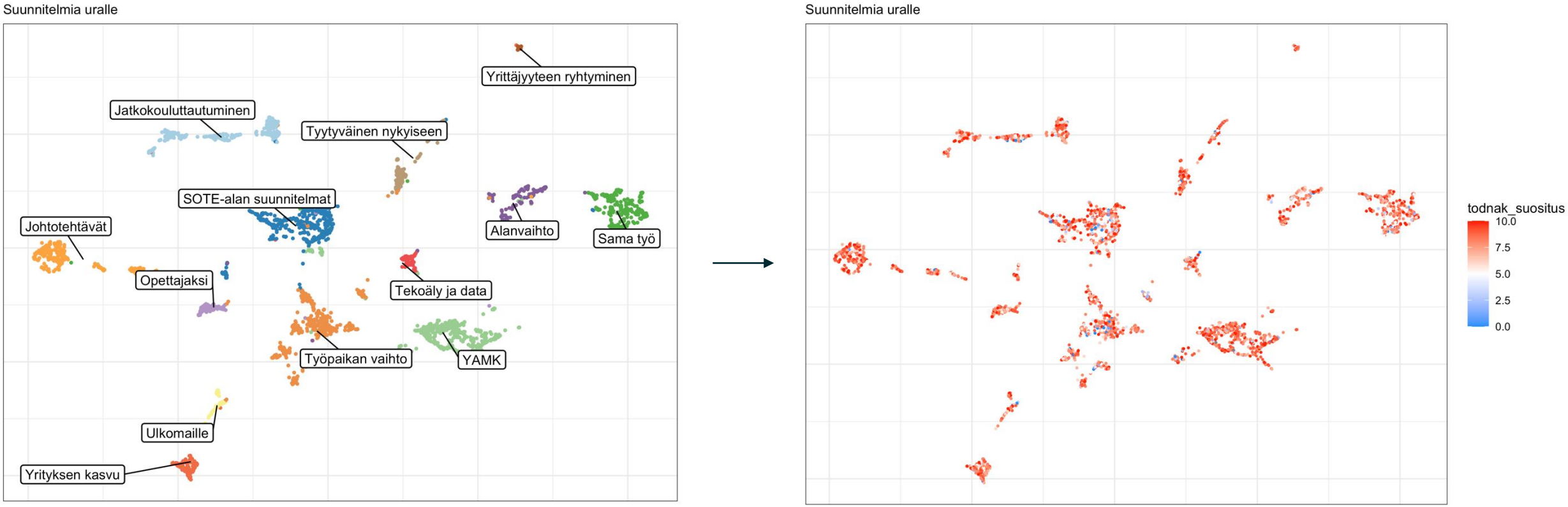
Aihemallinnuksen tuloksista

Suunnitelmia uralle, väritetty AMK:n mukaan



- Yllä olevassa kuvassa jokainen vastaus on väritetty vastaajan AMK:n mukaan, ja siitä voi nähdä aiheiden sisältävän monien korkeakoulujen vastauksia

Aihemallinnuksen tuloksista



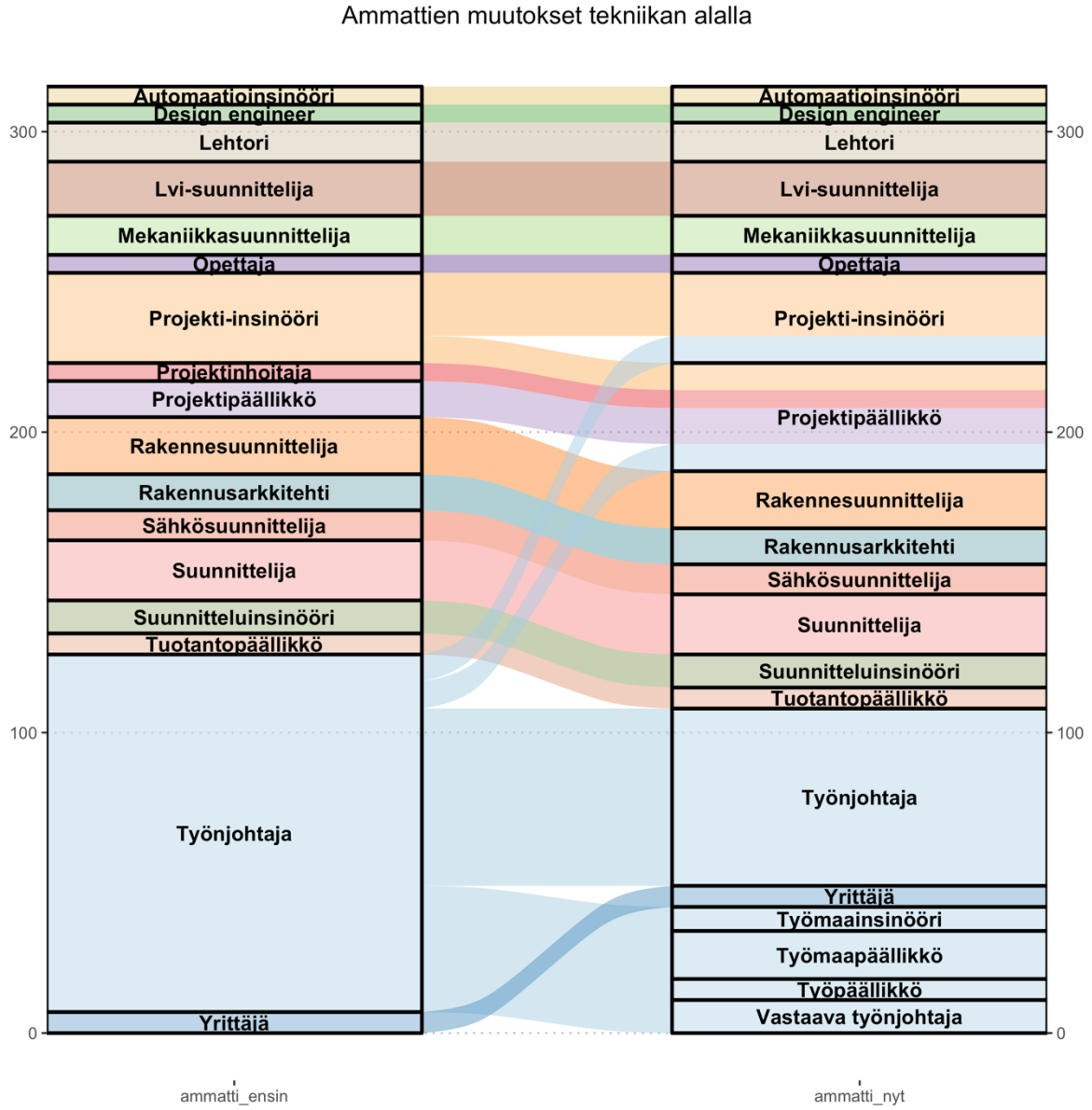
- Visualisoinnein voidaan myös havainnollistaa monivalintakysymysten vastauksia eri aiheissa
- Yllä olevassa kuvaajassa väri on kauttaaltaan pitkälti punainen, eli eri aiheiden vastaajat suosittelisivat koulutustaan muillekin

Aihemallinnuksen tuloksista

- Aiheista voidaan myös luoda tilastollisia analyyseja kuvaamaan sitä, onko koulutusalalta tullut aihetta koskevia vastauksia keskiarvoa enemmän tai vähemmän
- **Tulokset ovat vielä alustavia,** mutta analyysissa voi jo nähdä odotettuja linjoja: ICT-alalla korostuu tekoälyyn liittyvät suunnitelmat, kasvatusaloilla tähtääminen opettajaksi

Ammattinimikkeiden muutoksista

- Kysymykset ammattinimikkeiden muutoksista mahdollistavat niiden tarkastelun analyysissa
- Pilotin puitteissa ammatteja yhdisteltiin samankaltaisimpien kirjotusasujen mukaan
  - Aineisto sisältää monia harvinaisempia ammattinimikkeitä, joita voisi edelleen yhdistää samankaltaisiin ammatteihin
- Suurin osa vastaajista on ilmoittanut saman nimikkeen sekä ensimmäisessä työssä valmistumisen jälkeen että nyt
- Oikealla olevassa kuvaajassa on esitetty tekniikan alalta vastausyhdistelmät, jotka esiintyivät aineistossa vähintään viisi kertaa
  - Kuvaajasta voi huomata ammattinimikkeiden pysyneen pitkälti samoina, mutta esimerkiksi työnjohtajien edenneen erilaisiin päällikkötehtäviin



Yhteenveto

Yhteenvetoa pilotin analyyseista

- Johdonmukaisten aiheiden tunnistaminen onnistuu uraseurannan avoimista vastauksista koneoppimistyökaluin. Monikieliset mallit kykenevät yhdistämään samanlaiset kontekstit myös ruotsin- ja englanninkielisistä vastauksista
  - Suuri mahdollinen este analyysille olisi ollut esimerkiksi pidempien vastauksien vähäinen määrä, joka olisi saattanut estää sopivien aiheiden löytämisen
- Löydetyt aiheet eivät rajoitu vain yhteen ammattikorkeakouluun, vaan aineiston käsittely onnistuu valtakunnallisella tasolla
  - Mallia pitäisi jonkin verran hienosäätää purkamaan ammattispesifit aiheet
- Aihemallilla voidaan luotettavasti luokitella myös uusia vastauksia jo tunnistettuihin aiheisiin, analyysi siis mahdollistaisi vuosittaisten erojen tarkastelemisen
  - Malliin jäi vielä hiottavaa, esimerkiksi osa haastavasti luokiteltavista vastauksista suodatettiin vielä pois
- Aihemallinnuksen tuloksia voidaan myös raportoida taulukkomuotoisena, joka mahdollistaisi aiheiden ristiintaulukoinnin taustamuuttujien ja muiden vastausten kanssa
- Ammattinimikkeistä valtaosa on ajanjakson aikana pysynyt samana, mutta erojen tarkastelu aikapisteiden yli osoittautui mielekkääksi. Analyysia saisi edelleen niin hiottua, kuin kätevästi laajennettua muiden vuosien aineistoihin

Kiitos!

Jason Theodoropoulos
Datatieteilijä

etunimi.sukunimi@csc.fi

Follow us

LinkedIn

Instagram

Facebook

YouTube

csc.fi