

*This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.*

**Please cite the original version:** Lahtinen, H. (2019) Theseuksen louhintaa Varis.ai –tekoälyllä. Kreodi 5.

URL: <https://www.kreodi.fi/en/34/Artikkelit/598/Theseuksen-louhintaa-Varisai--teko%C3%A4lyll%C3%A4.htm>

# Theseuksen louhintaa Varis.ai -tekoälyllä

Hanna Lahtinen

**Theseus on oivallinen kohde tekoälylle. Se sisältää riittävästi tekstimassaa ja sisältöä, jota voidaan louhia monenkin kiintoisan kysymyksen selvittämiseksi. Artikkelissa kuvataan kokemuksia pilotista, jossa pyrittiin hakemaan ymmärrystä tekoälyn soveltamisesta Theseuksen tekstilouhinnassa.**

Korkeakoulut ovat saaneet viime vuosina monia yhteydenottoja tekoälysovelluksia tarjoavilta yrityksiltä. Ammattikorkeakouluissakin on pohdittu digitalisaation seuraavaa vaihetta, jossa tekoälyä koulutetaan toimimaan ihmisen kumppanina informaatiomassan jalostajana ja rikastajana. Tekoälyllä haetaan kohteita, joiden tarkkaa lopputulosta ei tiedetä tai tarkkoja vaiheita ei osata määritellä ja analysoitavana materiaalina käytetään omista tietokannoista löytyviä tekstejä ja dataa tai sosiaalisesta mediasta ja internetistä löytyvää dataa, kirjoituksia jne.

Tekoälyn roolina on käsitellä suuria määriä informaatiota, joka voi olla eri paikoissa ja eri muodoissa. Ihmisen roolina on määritellä informaation tiivistämisen tavoitteet, suunnitella tekoälyn kouluttaminen toimimaan yhä tarkemmin tavoitteen saavuttamiseksi tietyssä ympäristössä sekä tehdä jäsentyneen informaation pohjalta päätökset. Viime aikoina päätöksentekoa tai sen valmistelua on siirretty myös tekoälyn tehtäväksi, esimerkiksi rikosten ennakkoinnissa tai tehtäessä niin sanottuja automatisoituja päätöksiä esimerkiksi rikustuomioissa tai verotuksessa.

## Laurean tekoälykokeilu

Tekoälyn soveltamismahdollisuuksia pohdittiin Laureassa alkuvuonna 2018, jolloin haluttiin käytännön kokemuksia siitä, millä tavalla ihmisen ja tekoälyn yhteistyö toimii ja minkälaisia hyödyntämispotentiaalia tekoälyllä on tulevaisuudessa. Keskusteluissa oli esillä useita tekoälyn hyödyntämismahdollisuuksien kokeiluja, mm. oppimisanalytiikka eli miten voidaan ennakoida opintojen sujumista ja minkälaisia yksilöllistä tukea opiskelija tarvitsee.

Toisena aiheena oli onnistuneiden hankehakemusten tuottaminen, esimerkiksi hankesopimusten tekstien analysointi, tiivistäminen ja hakemusten osittain automaattinen kirjoittaminen. Keskustelimme miten tekoälyn avulla olisi mahdollista tukea hakemusten tekoa tunnistamalla yhteneväisiä teemoja hakemuksista ja käytössä olevista aineistoista kuten tutkielmista, kirjoista, artikkeleista ja sosiaalisesta mediasta. Pohdimme, miten kirjoitettavaa hakemusta voitaisiin verrata rahoitusta saaneisiin hakemuksiin ja löytää trendejä, mitkä seikat ovat johtaneet onnistumisiin. Pyrkimyksenä oli tukea ja helpottaa raskaiden hankehakemusten kirjoittamista. Haasteeksi osoittautui riittävän tekstimassan saaminen helposti käyttöön eri järjestelmistä lyhyen pilotin aikana.

Alustavista epämääräisistä ajatuksista ja tarpeista muotoiltiin vähitellen sopiva sisältö pilottiin, jolla saisimme käsitystä tekoälyn sovellettavuudesta myöhempiin tarpeisiimme. Halusimme selvittää tekoälyn avulla toimialoihin liittyvää vahvuuttamme, osaamistamme ja

kokeneisuuttamme hankehakemusten tueksi. Keskusteluihin osallistui Laurean TKI:n, IT:n ja kirjaston henkilöstöä sekä tekoälysovellusta tarjoava yritys [Valuemotive](#).

## Theseuksen opinnäytetyöt Laurean pilotin aineistona

Valitsimme pilotin aineistoksi opinnäytetyöt, joiden arvelimme kuvastavan ammattikorkeakoulullemme keskeistä osaamista ja vahvuutta. Opinnäytetöissä on riittävästi informaatiomassaa, joka oli helposti saatavilla ilman tunkeutumista GDPR-tietoa sisältäviin käyttäjätunnuksellisiin järjestelmiin. Informaatio oli myös yhtenäisessä muodossa ja yhdessä paikassa, mikä nopeutti pilotin tulosten saamisessa käyttöömmme. Theseuksen tekstimassa soveltui hyvin myös kumppaniksi valikoituneelle yritykselle, joka oli kehittänyt tekoälyohjelmansa [Varis.ai](#):n luonnollisen kielen tekstimassojen analysoimiseen.



Varis.ai on kokoelma tekstianalytiikan työkaluja ja teknologioita.

Pilotin toteuttamiseen osallistui Laureasta muutama kirjaston työntekijä, jotka raportoivat kokemuksista TKI- ja IT-johdolle. Pilotti aloitettiin tammikuun lopulla 2018 pohtimalla kysymyksiä, joihin halusimme vastauksia. Tämä osoittautui haasteelliseksi, koska meillä ei ollut kovin hyvin selvillä, millä tavalla operationalistaisimme pilotin sadaksemme selville ammattikorkeakoulumme vahvuudet ja osaamiset. Meillä ei ollut myöskään riittävää käsitystä tekoälystä, esimerkiksi minkälaiset kysymykset ovat tyyppisiä tai ylipäätään mahdollisia selvittää, ja erityisesti lyhyen pilotin aikana.

Ymmärsimme, että tekoälyä ei voi ostaa kaupasta ja napsauttaa päälle kuten tietokoneen ohjelmaa, vaan tekoälyn hyödyntäminen edellyttää tavoitteiden pohdintaa ja pitkäjänteistä työskentelyä ja tekoälyn kouluttamista tehtävänsä. Tässä mielessä pilotti oli jo onnistunut, ja olimme saaneet ensimmäiset kokemukset työskentelystä tekoälyn kanssa.

## Aineiston rajaus

Yhteistyössä yrityksen kanssa ja pitkän pohdinnan jälkeen päädyimme analysoimaan viiden korkeakoulun sosiaalialan ja liiketalouden opinnäytetyöt vuosilta 2016–2017. Korkeakoulut valikoituivat kumppanuusperiaatteen mukaisesti, koska meitä kiinnosti, millä tavalla naapurikorkeakoulut eroavat meistä. Valitsimme aineistoksi liiketalouden koulutusalan 2016–2017 opinnäytetyöt Laureassa, Haaga-Heliassa, Hamkissa ja Metropoliasa sekä sosiaalialan 2016–2017 opinnäytetyöt Laureassa, Diakissa, Hamkissa ja Metropoliasa.

Loppuvaiheessa pilottia teimme lisäksi vertailun Laurean uudempien 2016–2017 ja vanhempien 2010–2011 sosiaalialan opinnäytetöiden välillä. Rajaamalla aineistoa pyrimme saamaan pilotissa esiin olennaisimman tekoälyn hyödynnettävyyttä ajatelleen, emmekä niinkään mahdollisimman tarkkaa ja todenmukaista vastausta kysymyksiimme. Pitäessämme aineiston riittävän pienenä ja hallittavana välttyimme laajemman aineiston mukanaan tuomilta satunnaisilta poikkeamilta, jotka olisivat teettäneet lisäpohdintoja ja -toimintoja sekä vaatineet enemmän työstämisaikaa. Alla olevassa taulukossa Varis.ai:n analysoimat opinnäytetyömäärät.

Aineisto	Opinnäytetöiden määrä
Laurea_LIKO 2016-2017	618
Laurea_LIKO_old 2010-2011	523
Laurea_SOS 2016-2017	274
Laurea_SOS_old 2010-2011	190
Diakonia_SOS 2016-2017	434
Haaga_LIKO 2016-2017	489
HAMK_LIKO 2016-2017	125
HAMK_TER 2016-2017	189
Metropolia_LIKO 2016-2017	273
Metropolia_SOS 2016-2017	95

Taulukko 1. Opinnäytetyömäärät korkeakouluittain ja koulutusaloittain. Taulukon lähteenä on käytetty Valuemotiven raporttia (2018).

Yritys työsti ohjelman, joka latsi opinnäytetyöt ja niiden kuvailutiedot Theseuksesta. Opinnäytetyöt on tallennettu PDF-muodossa, joten niiden analysointia varten tekstisisältö irrotettiin erilliselle ohjelmalle. Tämän jälkeen data oli siinä muodossa, että erilaisten jaotteluiden käytännön toimivuutta voitiin testata.

Valuemotive teki analyysin suorittamalla vertailuryhmien välisiä tarkasteluja ja poimien kunkin ammattikorkeakoulun opinnäytetöissä tunnusomaisia eli poikkeuksellisen runsaasti esiintyviä käsitteitä. Kaikki vertailut tehtiin vertaamalla kunkin yksittäisen ryhmän opinnäytteitä vuorollaan muiden tutkimuksessa mukana olevien ryhmien yhdessä muodostamaa tausta-aineistoa vasten. Tulokset koottiin Excel-tilukoihin erillisille välilehdille.

Testien jälkeen yritys totesi, että tavoitettamme vastaavat tarkemmat tulokset saadaan, kun hakua ei kohdisteta kokotekstiin, vaan otsikkoon, avainsanoihin ja tiivistelmään. Näihin on yleensä tiivistynyt keskeisimmät opinnäytetyön aihetta kuvaavat sanat, jolloin tuloksen tarkkuus kasvaa.

Laaja tekstimassa tuo sattumanvaraista hajontaa niin paljon, että siitä on haastavampaa löytää keskeiset sanat lyhyessä pilotissa. Jos olisimme jatkaneet pilotista varsinaiseen projektiin, niin olisimme voineet edetä ja kouluttaa tekoälyä tunnistamaan yhä paremmin laajoistakin tekstimassoista olennaisia asioita. Ymmärsimme myös, että pilotissa yritys poisti meille näytetyissä tuloksissa manuaalisesti yleisimmät, ns. tyhjt sanat (ja, olla jne), koska tekoälyä ei ollut vielä koulutettu ymmärtämään näiden merkityksettömyyttä tavoitteidemme kannalta.

Helmikuun alkupuolella saimme katsottavaksemme ensimmäiset tulokset Excel-tiedostoina, jotka kävimme yrityksen kanssa läpi. Meille oli tärkeää ymmärtää, missä muodossa ja minkälaisia tuloksia pilotista oli odotettavissa. Alla olevassa taulukossa on esimerkki sosiaalialan opinnäytetöiden analysista. Avoinna on Laurean välilehti.

label	feature	p_value	ratio	yes_case	yes_control	words
1	viitekehys	0,000	2,761	137	123	viitekehyyseen, Viitekehyyksen, Viitekehyyksenä, Viitekehys, viitekehyyksen, viitekehyyksestä, viitekehyyks
2	päiväkoti	0,000	1,794	335	463	päiväkodin, koululuokkien, Ohjaamon, Ohjaamo, päiväkodeista, päiväkotit, Kioskin, päiväkodeissa, huor
3	tutkielma	0,000	6,197	45	18	Tutkielmassani, tutkielmalle, Tutkielman, tutkielman, Tutkielmastani, Tutkielmamme, tutkielmassa, Tut
4	huoltaja	0,000	5,061	49	24	Huoltajat, Huoltajilla, Huoltajien, huoltajia, huoltajalleen, huoltajat, huoltajien
5	loma	0,000	54,535	22	1	loma, lomaa, lomassa
6	teoreettinen	0,000	2,068	151	181	teoreettisena, Teoreettinen, käsitteellisiä, teoreettisen, teoreettinen, teoreettisessa, teoreettista, teo
7	iva	0,000	18,178	22	3	iva
8	kuvataide	0,000	14,254	23	4	kuvataiteen, kuvataiteeseen, kuvataidetta, Kuvataiteeseen, Mosaikki, maalaus, ojenukset, kuvataite
9	palvelu	0,000	1,772	193	270	toimivuus, palveluita, käyttötarkoituks, käyttötarkoitusta, palveluko, palveluille, palveluitaan, toimivuu
10	kehittäminen	0,000	1,615	258	396	Uudistusten, kuntoutukseen, parannuksia, kehittämiseksi, kuntoutuksessa, kuntoutusta, kuntoutuksen,
11	kasvattaja	0,000	2,309	95	102	kehittäjänä, kasvattajat, kasvattajan, suunnittelija, tuottajan, kuvittajien, kasvattajana, kasvattajilla, ka
12	luova	0,000	5,911	31	13	luova, luovien, Luovien, luovia, luovissa, luovaa, luovien, omaperäisiä, luovana, luovassa, luovan
13	kuva	0,000	2,300	90	97	valokuvien, kuvat, kuvaukset, kuvauksia, Kirjeiden, Kartta, kuvauksen, kuva, kuvaan, kartan, valokuvat,
14	työkalupakki	0,000	15,700	19	3	työkalupakkia, työkalupakin, laatikon, laetiko, laetikoon, Työkalupakki, Työkalupakin, työkalupakki, t
15	digitaalinen	0,000	4,958	32	16	digitaalisista, digitaalista, digitaalisesta, digitaalisia, digitaalinen, digitaaliseen, Digitaalinen, Digitaaliste
16	ohjekirja	0,000	8,145	23	7	ohjekirjaa, opaskirja, ohjekirja, Ohjekirja, ohjekirjasta, Ohjekirjan, Ohjekirjasta, Opaskirja, opaskirjasta,
17	ihmeellinen	0,000	8,145	23	7	vaikuttavat, Ihmeelliset, erinomaisena, erinomaisin, valtava, vaikuttavina, vaikuttavaa, erinomaisia
18	hanke	0,000	1,536	254	410	projektiin, Projektimme, tehtäviin, Hanke, Hankkeessa, uhka, hankkeelle, Tehtävänä, hankkeella, toime
19	kaupunki	0,000	1,688	173	254	kenttään, kunnissa, kunnassa, Kunnat, Kaupunki, kaupunkien, kaupungilla, kaupungille, kunnista, Kunta,
20	kirja	0,000	2,347	71	75	käsikirjaa, opaskirja, ohjekirja, Esite, Esitteessä, työkirjan, esitteistä, työkirja, kirjalta, Kirjasta, ohjekirja
21	uudistus	0,000	6,817	22	8	Uudistusten, uudistuksesta, uudistuksella, uudistus, uudistuksessa, uudistusta, uudistuksen, uudistukse
22	hakuteos	0,000	4,627	28	15	käsikirjaa, ohjekirjaa, opaskirja, ohjekirja, Ohjekirja, ohjekirjasta, Ohjekirjan, Ohjekirjasta, Käsikirjasta, C
23	työväline	0,000	2,066	80	96	työvälineiksi, Työvälineiden, Työvälineitä, työvälineiden, työväline, työvälineen, työvälineeksi, työkalu
24	oppiminen	0,000	1,951	85	108	Oppimisen, oppimiselle, opiskelu, opiskelu, oppimiseen, oppimisessa, opiskelujen, oppimisen, Opiskelu
25	draama	0,000	5,206	21	10	draama, draamaa, draaman, draamaan, Drama, Dramasta
26	sovellus	0,000	5,206	21	10	sovellus, sovellusta, sovelluksista, sovelluksen, sovelluksia, sovelluksina, sovelluksilla, sovelluksien, sovi
27	pelaaminen	0,000	3,653	28	19	Pelaamisessa, pelaamista, Pelaamisesta, pelaamisveen, pelaamisesta, pelaaminen, pelaamisen
28	leikkii	0,000	1,968	77	97	leikkivät, leikit, leikki, leikittiin, Leikin, käyttäytyvää, Leikki, leikkimällä, leikkii, leikin
29	avata	0,000	3,010	34	28	avaamme, avataan, avasi, avata, avattiin, Avaamme, avattu, avannut, Avaan, avaa

Kuva 1. Varis.ai:n analysoimat sosiaalialan (2016–2017) opinnäytetyöt. Kuva Valuemotiven tuottamista Excel-taulukoista.

Saimme taulukon lisäksi lukuohjeet, mitä eri sarakkeiden sisältö tarkoitti. Label tarkoittaa käsitteen nimeä tai yleisintä sanaa, jolla käsite esiintyy aineistossa. Feature tarkoittaa käsitteen WordNet-tunnistetta. P-value tarkoittaa todennäköisyyttä saada vähintään havaittu määrä esiintymiä otannassa esiintyvyyden tausta-aineiston perusteella olettaen, kun on suhteellisen esiintyvyyden määrä suhteessa tausta-aineistoon. Ratio tarkoittaa havaintomäärää kyseisessä aineistossa. Yes\_case tarkoittaa havaintomäärää tausta-aineistossa. Yes\_control tarkoittaa havaintomäärää tausta-aineistossa. Words kuvailee esimerkkejä sanoista, joilla käsitteestä on aineistossa kirjoitettu.

Kiinnostavaa on esimerkiksi havaita, minkälaisia erilaisia käsitteen ilmenemismuotoja Varis.ai on tulkinut tarkoittavan päiväkotia, kaupunkia tai hanketta, jotka taulukossa esiintyvät ensimmäisessä ja viimeisessä sarakkeessa. Kiinnostavaa on myös korkeakoulun aineiston (yes\_case) suhde (ratio) tausta-aineistoon (yes\_control) eli muiden korkeakoulujen opinnäytetöihin. Esimerkiksi Diakin aineistossa ratio-luku oli suuri labeleilla kuten seurakunta, kirkko, hengellinen, uskonnollinen rituaali, seremonia, jumalanpalvelus, rukous ja diakoniatyöntekijä.

Muista ammattikorkeakouluista ei niin selkeästi erottunut tietyn tyyppisiä sanoja. Metropolissa suurin ratio oli monimuotoinen-labelilla ja Laureassa loma-labelilla. Tässä vaiheessa huomasimme, että Hamkin opinnäytetyöt olivatkin sairaanhoitajan ja terveydenhoitajan koulutusohjelmista, kun muilla oli sosiaalialan koulutusohjelmista, mikä tietysti vaikutti koko analyysia vinouttavasti. Totesimme kuitenkin, ettei analyysin korjaaminen muuttaisi sitä perusasiaa, ettei tuloksena juurikaan ollut sellaisia labeleita kuin olimme toivoneet.

## Analyysi ja hyödyntäminen

Vaikka on kiinnostavaa tietää, miten oman ammattikorkeakoulun opinnäytetöiden aiheet poikkeavat vastaavista muiden ammattikorkeakoulujen, niin emme oikein tienneet, miten todennettua tietoa voitaisiin hyödyntää. Labelit tuntuivat olevan edelleen kovin yleisiä sanoja, kuten loma. Oli vaikeaa mieltää, miten se olisi oman organisaation vahvuus ja miten sitä voitaisiin hyödyntää hankehakemuksissa. Pohdimme kyllä, että tekoälyä kouluttamalla voisimme jatkossa puristaa mielekkäämpiä labeleita tekstimassasta.

Haasteena oli päättää, miten etenemme tekoälyn kanssa. Meille ei ollut muodostunut vielä riittävää käsitystä, soveltuuko tekoäly tarpeisiimme ja aiommeko jatkaa pilotista projektiin vai tyydymmekö vain esimaistiaisiin. Varsinaiset hyödyt tekoälyn kanssa toimiessa saavutetaan vasta pitkän ajan kuluessa. Oli haastavaa sitoutua varsinaiseen projektiin tietämättä, oimmeko edes oikeilla jäljillä. Tässä vaiheessa myös rahoitus alkoi vaikuttaa asiaan. On helppo sitoutua kevyeen pilottiin, joka on suhteellisen pieni kustannus sekä rahallisesti että henkilöresursseiltaan mutta pitkäkestoisempaan projektiin sitoutuminen vaikutti jo vaativammalta ponnistelulta.

Käydessämme läpi alustavia tuloksia ymmärryksemme lisääntyi tekoälyn sovellettavuudesta. Tämän jälkeen keskustelimme ja määrittelimme uudelleen tarkemmin, mitä tavoittelemme pilotista ja miten parhaiten saavuttaisimme tavoitteemme.

Yhteenvetona saimme analyysit (Valuemotive 2018):

- Laurea keskittyy sosiaalialan opinnäytetöissä lapsiin, nuoriin ja oppimiseen (päiväkotia, huoltaja, kasvattaja, oppiminen, pelaaminen, leikkiä, vantti!, ohjaaja). Hyvin vahvoina teemoina näyttäytyvät myös luovuuteen ja kokemiseen liittyvät käsitteet kuten loma, kuvataide, luova, kuva, kirja, draama, liikunta, kirjallinen teos, matka, kulttuuri, havainnointi, näkyä, kuvallinen esitys, tarina, ja kuvallinen.
- Diakonia-ammattikorkeakoulun opinnäytteissä korostuvat voimakkaasti uskonnolliset ja hengelliset teemat: seurakunta, kirkko, diakoni, hengellisyys, diakoniatyöntekijä, hengellinen, seremonia, uskonnollinen rituaali, jumalanpalvelus, rukous, ja uskonto. Myös sosiaaliseen asemaan liittyvät teemat ovat heillä vahvasti esillä, esimerkiksi diakoni, sosiaaliryhmä, suomi, sosiaalityö, apu, vanki, vankila, vertainen, syrjintä, jäsen, hankinta, lahja, vammaisuus, omaisuus, sukupolvi, maahanmuuttaja, verkosto ja työnantaja.
- Metropolian alle sata lopputyötä käsittävä aineisto oli tutkimuksen pienin. Kantavina teemoina otoksessa näyttäytyivät kuntouttavaan ja ennaltaehkäisevään toimintaan sekä asenteisiin viittaavat käsitteet, mm. asukas, yksinäisyys, sosiokulttuurinen, osallisuus, päihde, innostaminen, myönteinen, luonteenpiirre, hyvinvointi, ehkäistä,

osallistua, positiivinen, asuinpaikka, osallistuminen, vahvuus, ohjelma, kyky ja vapaaehtoinen.

Liiketalouden osalta saimme yhteenvedoksi (Valuemotive 2018):

- Laureassa korostuvat selvästi lapsiin ja nuoriin (lapsi, teini), sekä perheeseen (avioliitto, vanhempi) liittyvät käsitteet. Toisena merkittävänä ryhmänä nousivat lakeihin, niiden noudattamiseen liittyvät teemat (oikeus, oikeusturva, tuomioistuin, rangaistus, laki, oikeudenkäynti, kriminalisointi, rikosentekijä, rikoslaki, rikkomus, lainsäädäntö).
- Haaga-Helian lopputöitä luonnehtivat kansainväliseen toimintaan liittyvät maita, vientiä ja matkailua kuvaavat aiheet. Luonnollisena rinnakkaisteemana näkyivät monet logistiikan osa-alueet. Haaga-Helian lopputöissä viitattiin korostuneen runsaasti erilaisiin ajankohtiin, mikä kertonee jotakin tavasta, jolla aiheita lähestytään.
- Hamkin aineistosta esiin nousivat henkilöstön palkat, luontoisedut ja perehdyttäminen. Taloudellista tilaa ja kannattavuutta kuvaavat aiheet ovat nekin vahvasti esillä, esimerkiksi konkurssi, kirjanpito, talous, kulut, taloudellinen, lopetus, avu, kasvu, hyöty, tehokas
- Metropolia erottui muista tarkastelluista koulutusohjelmista vahvoilla sijoitus- ja johtamispainotuksillaan

Päättelimme, että tulokset heijastelivat osittain korkeakoulujen sisällöllisiä painopisteitä, mutta niissä näkyivät myös suuntautumiset esimerkiksi Laureassa rikosseuraamukseen ja Haaga-Heliassa logistiikkaan. Tulokset eivät tuoneet esille mitään yllättäviä uusia näkökulmia toimintaamme.

## **Vertailu vanhempaan aineistoon: Trendejä esille?**

Kokeilimme aineistolla vielä, voiko oman organisaation tuoreiden (2016–2017) opinnäytteiden vertaaminen varhaisempiin (2010–2011) tuoda esiin kehityskaarta, jota voisimme jatkossakin seurata ja hyödyntää. Ajattelimme löytävämme trendejä tai ajankohtaisia kiinnostuksen kohteita, jotka voisivat kertoa heikkoja signaaleja, mitä yhteiskunnassa on meneillään ja koetaan kiinnostavaksi ja mihin ammattikorkeakoulu on mahdollisesti reagoanut.

Yhteenvetona tähän analyysiin saimme (Valuemotive 2018):

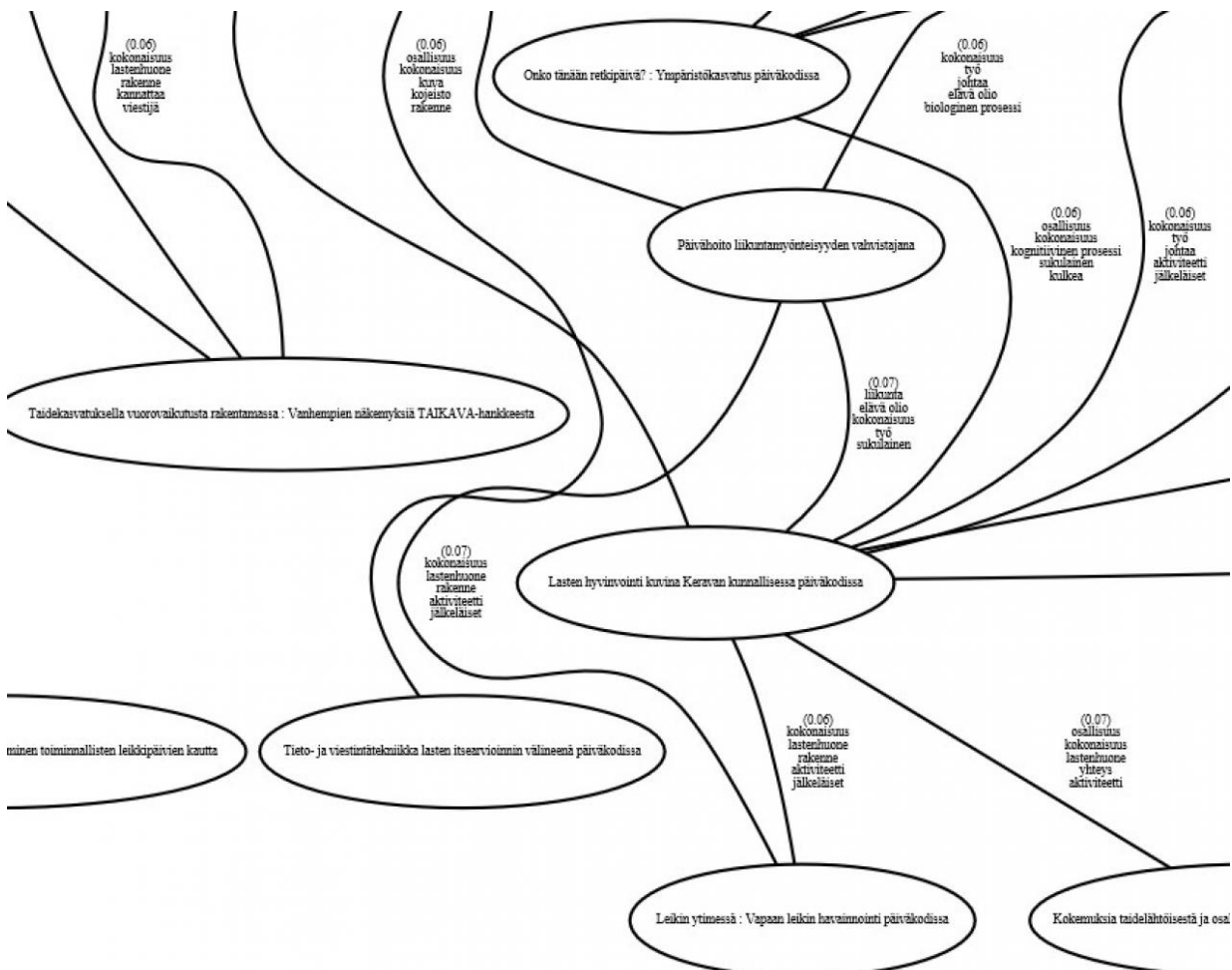
- Verrattaessa uusimpia lopputöitä Laureassa vuosina 2010–2011 julkaistuihin teksteihin havaittiin, että sekä perheeseen että lakeihin ja niiden noudattamiseen liittyvät teemat ovat syntyneet viimeisten vuosien aikana ja nousevat esiin vertailussa. Samassa vertailussa havaittiin digitaalisen sisällön entistä vahvempi mukaan tulo käsiteltyihin aiheisiin (digitaalinen, kanava, sosiaalinen, verkkosivu, blogi, video), vaikka Internet-alusta puhututti vielä enemmän varhaisemmassa aineistossa.

Mielestämme tämäkään tulos ei vielä vastannut sitä, että olisimme vakuuttuneet tekoälyn hyödyllisyydestä ja sen tuomasta lisäarvosta. Kokonaisten koulutusohjelmien laajuiset vertailut eivät ehkä yksin anna riittävän tarkkaa kuvaa toiminnan sisällöstä. Osaamisvahvuuksien löytämiseksi pohdimme yksittäisten opinnäytetöiden vertaamista toisiinsa.

Kolmantena analyysinä Valuemotive selvitti sosiaalialan oppinäytetöiden sisällöllisiä samankaltaisuuksia hyödyntäen Varis.ai:n kykyä määrittää aineistojen semanttisia etäisyyksiä. Analyysia tehtiin ensin tiukemmalla etäisyysrajalla (0.07/1) ja sitten vähän löyhemmällä (0.08/1). Tulokset esitettiin kaavioina, jossa tiukemmalla rajalla tuotettuun ja siksi yksinkertaisempaan kuvaan merkittiin näkyviin töiden väliset samankaltaisuusmitat ja esimerkkejä niitä yhdistävistä teemoista. Esitettyjen teemojen valinta ei perustu niiden erityiseen relevanssiin vaan algoritmin suoritusjärjestykseen.

Tämän analyysin tuloksena Varis.ai tunnisti *Lasten hyvinvointi kuvina Keravan kunnallisessa päiväkodissa* edustavaksi esimerkiksi varhaiskasvatuksen oppinäytetöistä, koska se liittyy läheisesti useisiin muihin töihin. Tuloksena oli siis organisaatiota kuvaava tyyppiesimerkki. Varis.ai tunnisti viisi tällaista tyyppiesimerkkiä sosiaalialan aineistosta. Emme havainneet löydöksistä olevan hyötyä hankehakemusten kirjoittamisessa, mutta sen sijaan pohdimme niiden hyödynnettävyyttä esimerkiksi oppinäytetöiden ohjauksen ja aiheidenvalinnan kehittämisessä.

Analyysin tuotoksina saimme kolme erilaista kaaviokuvaa. Alla esimerkki semanttisesta kaaviosta, jossa tekoäly on tunnistanut semanttisten etäisyyksien avulla keskeisiä oppinäytetöitä.



Kuva 2. Semanttisen analyysin perusteella tuotettu kaavio, jossa yksittäiset oppinäytetyöt on tunnistettu algoritmin avulla läheisiksi. Osakuva Valuemotiven tuottamasta kaaviosta.



## **Jatkokehitysideoita**

Pilotti ei tuonut esille kovin yllätyksellisiä asioita, joita emme olisi muutenkin jo tunnistanet omassa toiminnassamme. Lyhytaikaisessa pilotissa ei ollut mahdollista toteuttaa tekoälyn mahdollisuuksia täysimittaisesti. Tekoälyn kanssa jatkaminen olisi varmasti tuonut tarkkuutta tuloksiin ja olisimme voineet muokata tavoitteitamme ja kyselyjämme lukemattomilla eri tavoilla, ja kouluttaa Varis.ai:sta vähitellen hyvän ja tarpeellisen kumppanin itsellemme. Lisäksi käyttöliittymiä ja tulosten esittämistä olisi kehitetty helppokäyttöisemmiksi ja visuaalisemmiksi pidempiaikaisessa projektissa tai varsinaisessa toiminnassa. Pilotin toteuttamiseen oli kiinnostavaa osallistua ja se lisäsi ymmärrystä tekoälyn hyödyntämisen mahdollisuuksista ja rajoitteista.

Kokemuksista oli hyötyä, kun seuraavana vuonna osallistuimme Pepper-robotin kouluttamiseen kirjaston asiakaspalveluun. Pilotissa hankkimamme kokemuksen perusteella osasimme suhtautua toteutukseen melko realistisesti. Kokemusten myötä on myös hävennyt pelko siitä, että tekoäly tekee meidät tarpeettomiksi. Sen sijaan on oleellista pohtia, minkälaisiin tehtäviin tekoäly soveltuu ja miten omia työtehtäviämme on hyvä muuttaa.

## **Lähteet:**

Valuemotive (2018). Katsaus Laurean opinnäytteisiin. Pilotin yhteenveto.  
Valuemotive (2018). Analyysin tulokset. Excelit ja kuvat.

## **Tietoa kirjoittajasta:**

Hanna Lahtinen, johtaja tieto- ja julkaisupalvelut, Laurea-ammattikorkeakoulu