

Indikaattoreista ilmiöihin - hyvinvointi- ja turvallisuustiedolla johtaminen Lapissa

Raportti työpaketti 2.1 – Demon digitaaliset vaihtoehdot

Helameri Eeva
Kanto Teija
Keränen Pia
Lindholm Tuomo

Lapin ammattikorkeakoulu
2024

1. Sisällysluettelo

1. Sisällysluettelo	2
2. Taustaa	3
3. Olemassa oleviin tietojärjestelmiin ja sovelluksiin tutustuminen	4
4. Ennakointialustat	5
5. Havaintoja nykyisistä tietojärjestelmistä ja sovelluksista	8
6. Demo laadullisen tiedon hyödyntämisestä	9
7. Demon komponentti 1 – Mobiilisovellus- laadullisen ilmiöpohjaisen tiedon kerääminen	10
8. Demon komponentti 2 - Tekoälyratkaisu - laadullisen ilmiöpohjaisen tiedon analysointi	11
8.1 Ketterä kokeilu tekoälyrajapinnan käytöstä dokumenttien käsittelyssä	11
8.2 Kokeilun tuloksia	12
8.2.1 Dokumenttilähteiden kerääminen aineistoksi	13
8.2.2 Dokumenttisisältöjen normalisointi	13
8.2.3 Dokumenttisisältöjen paloittelu tekoälylle lähetettävään muotoon	14
8.2.4 Dokumenttipalasan lähettäminen tekoälyn käsiteltäväksi	14
8.2.5 Tekoälyn muodostaman analyysin vastaanottaminen ja tallentaminen	16
8.2.6 Dokumentti palasten analyysien yhdistäminen	16
8.2.7 Tiedon vieminen havaintodatan yhteyteen	16
8.3 Kokeilun tekniset tiedot	17
9. Jatkosuunnitelma	17
10. Liitteet	18
11. Lähteet	18

2. Taustaa

Lapin Arjen turvallisuuden tiekartta –hankkeessa vuonna 2021 selvitettiin mitä toimenpiteitä ja toimenpidekokonaisuuksia tulisi Lapissa toteuttaa, jotta asukkaiden arjen turvallisuus parantuisi. Hankkeen toimenpiteissä Lapin eri toimijat toivat esiin näkemyksiään kehittämistarpeista. Tunnistettiin kymmenen toimenpidekokonaisuutta, joihin tulisi jatkossa suunnata kehittämistoimia. Näistä kokonaisuuksista yksi oli tiedolla johtaminen. Tunnistettiin, että tiedolla johtamisen prosesseihin ja myös tiedon saatavuuteen tulee panostaa. Tunnistettiin myös, että pelkkä tilasto- ja indikaattoritieto ei riitä siihen, että lappilaiset hyvinvointi- ja turvallisuustoimijat osaavat tehdä ajantasaisia päätöksiä siitä, mihin resurssinsa ja yhteistyönsä suuntaavat. Tilastotieto on luonteensa vuoksi aina taaksepäin katsovaa, ja toisaalta monilla toimijoilla on paljon ns. hiljaista tietoa tai kokemustietoa, jota hyödynnetään kovin vähän eikä kovin systemaattisesti. Tarvitaan myös laadullista eli kuvailevaa, ilmiöpohjaista ja ennakoivaa tietoa, jota voidaan yhdessä eri toimijoiden kesken hyödyntää.

Tähän tarpeeseen haettiin kaksivuotista Indikaattoreista ilmiöihin - hyvinvointi- ja turvallisuustiedolla johtaminen Lapissa (INDI) –hanketta. Hankkeessa edistetään monitoimijaista, verkostomaista yhteistyötä kehittämällä laadullisen, ilmiöpohjaisen ja ennakoivan hyvinvointi- ja turvallisuustiedon keräämistä, jakamista ja analysoimista. Keinoksi valikoitui yhteiskehittämällä demon rakentaminen tiedon keräämiseen, jakamiseen ja analysoimiseen.

Verkostomaisessa työskentelyssä tiedon jakamisen ja yhteiskäytön haasteet ovat moninaiset ja voivat vaikuttaa merkittävästi toiminnan ja päätöksenteon tehokkuuteen. Tiedonkulun esteet voivat johtua muun muassa työolosuhteista, viestintäteknologian rajoituksista sekä tiedon havaitsemisen, tulkinnan ja arvioinnin vaikeuksista. Erityisesti verkostomaisissa organisaatioissa, joissa työntekijät edustavat eri taustaorganisaatioiden näkemyksiä, tiedon oikea-aikainen ja oikeellinen välittyminen on kriittistä. Tämä edellyttää selkeitä kommunikaatioprosesseja ja tehokkaita digitaalisia välineitä, jotka tukevat yhteistyötä ja tiedon jakamista.

INDI-hankkeessa näitä haasteita ratkaistaan mahdollistamalla laadullisen tiedon sujuvaa liikkumista, ja siten tuetaan päätöksentekoa verkostoissa. INDI-hanke rakentuu kahdesta työpaketista. 1. työpaketti keskittyy tiedolla johtamisen nykytilan ja tarpeiden selvittämiseen, kun taas työpaketti 2 rakentaa taustaa itse demolle, sekä toteuttaa sen. Työpaketissa 1 toteutettiin kysely Lapin hyvinvointi- ja turvallisuustoimijoille, haastateltiin 17 näistä toimijoista, sekä toteutettiin kuusi Indikaattoreista ilmiöihin -työpajaa keväällä 2024. Taustakartoituksesta hankkeen osatoteuttaja, Lapin yliopisto, toimittaa raportin, joka löytyy myös hankkeen verkkosivuilta (www.indikaattoreistailmiöihin.fi). Samalla työpaketissa 2 selvitettiin demon rakentamisen taustoja: kartoitettiin mahdolliset digitaaliset vaihtoehdot, olemassa olevat mahdolliset digitaaliset ratkaisut/alustat ja niiden rajapinnat sekä tekoälyn käytön mahdollisuuksia demon kehittämisessä sekä ennakoitiedon hyödyntämistä lappilaisessa hyvinvointi- ja turvallisuustiedolla johtamisessa. Toteutettiin selvitys, jossa selvitettiin jo olemassa olevia digitaalisia palveluita aiheen tiimoilta, kartoitettiin mikä tekoäly voisi soveltua työpaketissa 1 kartoitettuihin tarpeisiin, ja miten tekoälyä voisi hyödyntää hankkeessa toteutettavan digitaalisen työkalun demon yhteydessä. Selvitetiin myös globaalin ennakoitilustan käyttöä, millaista tietoa alustalta saa hyvinvointi- ja turvallisuus (myöhemmin hytu) johtamisen näkökulmasta ja miten sitä voitaisiin käyttää Lapissa.

3. Olemassa oleviin tietojärjestelmiin ja sovelluksiin tutustuminen

Perehdyttäessä olemassa oleviin tietojärjestelmiin ja sovelluksiin, todettiin pitkälle se, ettei hankkeen tuotoksen kaltaista sovellusta tai applikaatiota ole käytössä. Tietojärjestelmien vertailussa on otettava huomioon niiden integroitavuus olemassa oleviin prosesseihin ja järjestelmiin sekä käyttäjäystävällisyys, jotta ne palvelevat eri käyttäjäryhmiä tehokkaasti. Tietosuojalainsäädännön noudattaminen on myös kriittinen tekijä, erityisesti kun käsitellään henkilökohtaisia tietoja tekoälypohjaisissa ennustemalleissa. Tämänkaltaisen vertailun tuloksena voidaan saavuttaa parempi ymmärrys siitä, miten erilaiset tietojärjestelmät voivat tukea ennakointia ja päätöksentekoa.

Suomen Itsenäisyyden juhluvuoden lastensäätiö (Itla) on alueellisessa oppimisverkostossaan kehittänyt kumppaneiden kanssa tekoälyä hyödyntävän Näkymä -työkalun lapsiperhepalveluiden tiedolla johtamisen tueksi. Itlan edustajan, erityisasiantuntija Mari Hirvosen kanssa järjestettiin palaveri tammikuun lopulla 2024, jotta saataisiin lisätietoa sovelluksesta. Näkymäksi nimetty työkalu auttaa ennakoimaan korjaavien palveluiden kysyntää ja tunnistamaan ehkäisevän työn mahdollisuuksia. Tekoälyä hyödyntäen havaintotietoon yhdistetään mahdollisuuksien mukaan Sotkanet- ja kouluterveyskyselyn tietoja sekä paikallisuutisotsikoita. Kyseiseen Näkymä-sovellukseen alueen kenttätoimijat tuovat havaintotiedon/hiljaiset signaalit lasten, nuorten ja perheiden arjesta osaksi tiedolla johtamisen prosessia. Tieto on täysin anonymia ja se tulee hyväksyä ennen havainnon kirjaamista. Tietosuoja-asioita oli mietitty tarkkaan, pyydetty lausunnot kunnan ja hyvinvointialueen tietosuojavaltuutetuilta ja lakimiehiltä. Sovelluksessa vain taustajärjestelmän pääkäyttäjät voivat tarkastella ja käsitellä havaintoja. Työkalua ei ole suunnattu asukkaille, mutta tämä voisi olla mahdollista. Työkalu oli käytössä tammikuussa Pirkanmaan hyvinvointialueen Perhekeskuksissa ja todennäköisimmin laajenemassa Oulun seudulle syksyllä 2024. Näkymä -sovelluksen toimintaperiaate käy ilmi alla olevasta Itlan verkkosivustolta olevasta kuvasta.

Näin Näkymä toimii



Kuva 1. Näkymä-sovelluksen toimintaperiaate (Itla)

VASA 2 -hankkeessa pilotoitava Tarina-auttamisen verkkopalvelu (Poske, 2023) tarjoaa psykososiaalista tukea lapsille, nuorille ja perheiden vanhemmille. Tarina-auttamisen videot sekä chat -palvelu ovat matalan kynnyksen palveluita, joiden tarkoitus on tukea arjessa jaksamista ja parantaa varhaista pääsyä palveluiden piiriin. Tilannehuone -sovellus on taas prototyypin selvitys (Pyykkönen, 2021), jossa tarkoituksena oli seurata lasten ja nuorten hyvinvointia alueellisesti. Johdolle suunnattu palvelu oli tehty konkreettisten aloitteiden tekemiseen ja alueen työntekijöiden havainnoista työkalusta kootaan merkittävistä havainnoista kokonaisuuksia, jossa ilmiöitä ja niitä rikastetaan tilastotiedoilla ja lisähavainnoilla. Tilannehuoneen prototyyppiä ei enää ole saatavilla keväällä 2024.

Sitra esittelee vuosittain Suomen oivaltavimmat dataa hyödyntäviä ratkaisuja. Lista sisältää oivaltavia ratkaisuja eri toimialoilta, kuten esimerkiksi kuluttajapalveluista ja infrastruktuuriratkaisuista. Listalle on valittu 24 erilaista datan hyödyntämisen ratkaisua 80 ilmoittautuneiden joukosta (Sitra, 2023).

Näiden lisäksi tutustuttiin muihin, myös ulkomaisiin sovelluksiin, kuten esimerkiksi mHealth App: Bright Path, Cyberbullying Vaccine, 117 Chat ja Loneliness Advice Chatbot. Tavoitteena oli nähdä, kuinka näillä tietojärjestelmillä voidaan kerätä ja jakaa laadullisen ilmiöohjaista tietoa sekä oppia miten tietoa analysoidaan ja visualisoidaan, jotta sitä voitaisiin hyödyntää paremmin päätöksen teon tukena. Listaus perehdytyistä digitaalisista sovelluksista, on liitteenä 1.

4. Ennakointialustat

Hankkeessa perehdyttiin erilaisiin ennakointialustoihin, jotta voitaisiin ymmärtää niiden toimintaperiaatteita ja selvittää voidaanko niitä hyödyntää tulevaisuuden ilmiöiden ennustamisessa hankkeen tavoitteiden mukaisesti hyvinvoinnin ja turvallisuuden teemoissa. Erilaisten ennakointiin soveltuvien tietojärjestelmien vertailu on tärkeää, sillä se auttaa hanketoimijoita ymmärtämään kunkin järjestelmän vahvuuksia ja heikkouksia suhteessa tulevaisuuden tarpeisiin toiminta-alueellamme. Tietojärjestelmien kyky käsitellä suuria datamääriä ja tuottaa tarkkoja ennusteita on avainasemassa, kun pyritään ennakoimaan monimutkaisia hyvinvointiin ja turvallisuuteen liittyviä ilmiöitä, kuten arjen turvallisuuden kokemuksia tai syrjäytymistä.

Ennakointiin liittyen selvitettiin erilaisia mahdollisuuksia. Turun yliopiston Tulevaisuuden tutkimuskeskuksen Hanna Lakkala kertoi, että laadullisen tiedon kerääminen ei useinkaan ole nopea ja siihen käytetään lukuisia menetelmiä, mm. laadulliset raportit, sivustot, keskustelut, tapahtumat, tutkimukset jne. (Lakkala, 2024). Eri näkökulmista ennakointia tehdään mm. Sitrassa, ministeriöissä, esimerkiksi Tulevaisuusvaliokunta ja Turun yliopiston tulevaisuuden tutkimuskeskuksessa. Lakkala myös totesi, että suurin merkitys on ennakointiprosessilla, joten Lapin demon kannalta tärkeintä on sisällyttää koko tiedolla johtamisen prosessiin ja demoon ennakointitiedon ja –materiaalin hyödyntämistä niistä lähteistä, joista sitä on saatavilla.

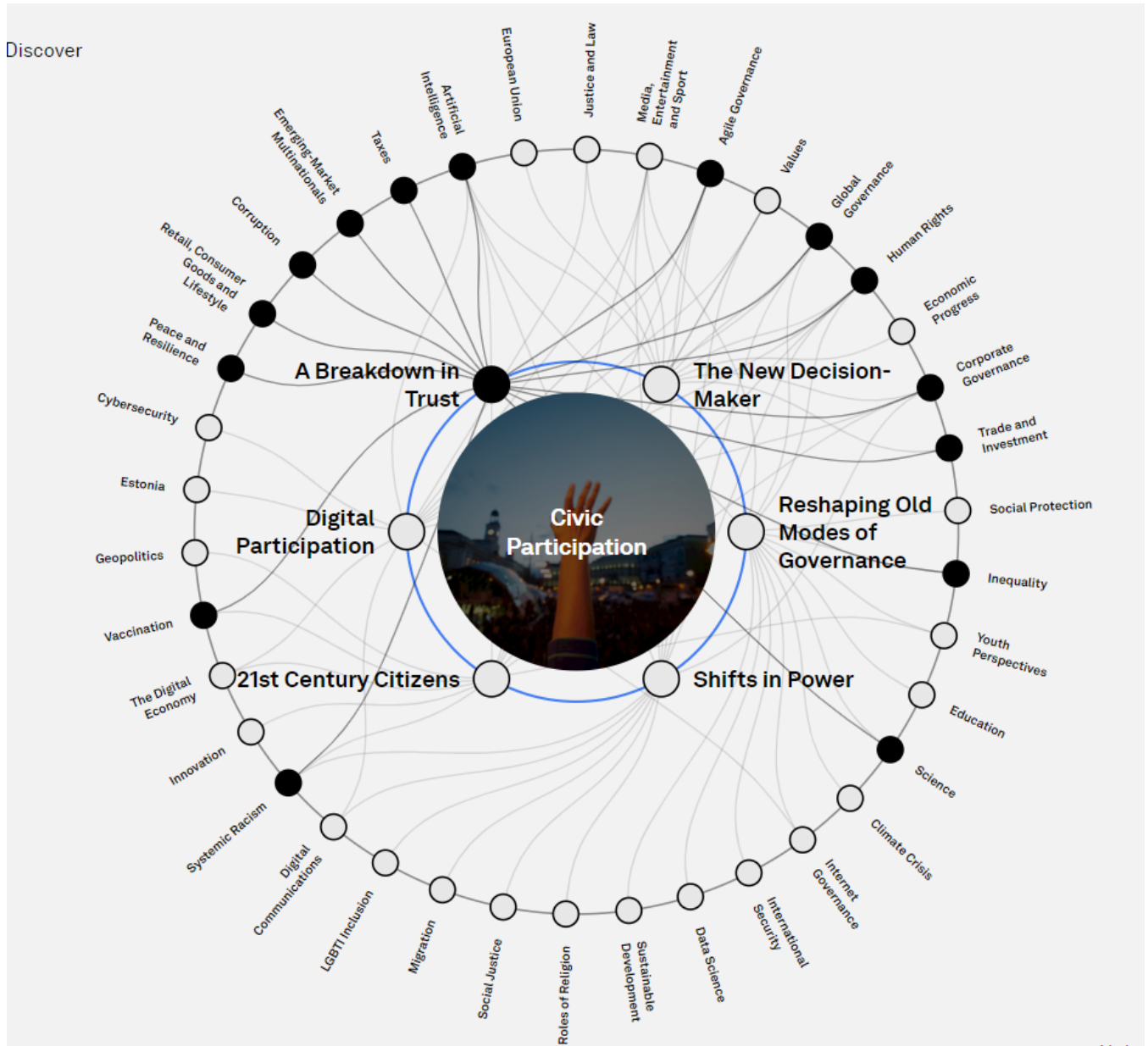
Selvityksen aikana Futures Platform erottui joukosta sen soveltuvuuden, käytettävyyden ja monipuolisten ominaisuuksien ansiosta. Kuitenkin järjestelmän kustannusten merkittävä nousu tutustumisvaiheen aikana johti siihen, että hankkeen resurssit eivät mahdollistaneet sen hankintaa. Tämä kokemus korosti ennakointijärjestelmien markkinoiden muuttuvaa luonnetta ja palveluntarjoajien kokeiluja löytää ratkaisuilensa sopivia hinnoittelumalleja.

Ennakointialustojen analyysi avasi, kuinka järjestelmät keräävät ja esittävät ilmiökohtaista tietoa, mikä on arvokasta erityisesti, kun pyritään ymmärtämään miten monimutkaisia globaaleja trendejä ja ilmiöitä koostetaan yhdeksi kokonaisuudeksi. Selvityksen perusteella järjestelmien toiminnan ytimessä on asiantuntijoiden keräämä laadullinen informaatio, eli toisinsanoin lehtiartikkelit ja muut julkaisut, joita liitetään metatietoja lisäämällä osaksi ilmiökohtaista tietomassaa. Nämä siis toimivat periaatteessa kuten muutkin alustatalouden mukaiset

tietojärjestelmät, eli pyrkivät yhdistämään tarjontaa (ilmiötieto) ja kysyntää (ennakointi ja päätöksenteko) mahdollisimman vähällä kitkalla. Informaation keräämisessä pystytään myös hyödyntämään erilaisia verkossa toimivia robotteja, joiden keräämät artikkelit ja luokittelut voidaan tarkistaa ihmisten tai tekoälysovellusten avulla. Jatkossa robottien ja tekoälysovellusten keräämä tieto auttaa meitä ymmärtämään ilmiöitä paremmin, koska ne pystyvät käymään läpi suurempia tekstimassoja kuin ihminen ja luokittelemaan niitä nopeammin. Kuten muuallakin, ihmisten rooli tulee olemaan tässä prosessissa enemmänkin tekoälyn päätöksiä vahvistava ja korjaava, kuin artikkeleita hakeva ja analysoiva.

Selvitystyön aikana opittiin myös ymmärtämään, kuinka jotkin tietojärjestelmät mahdollistavat ilmiöiden lokalisoinnin, eli millaisten toimenpiteiden avulla voidaan sovittaa globaaleja tietoja paikallisiin konteksteihin. Tämä on erityisen hyödyllistä, kun halutaan hahmottaa ilmiöiden paikallisia vaikutuksia hankkeen kohdealueen kaltaisessa ympäristössä ja seurata, kuinka tietty ilmiö etenee ja miten se muuttuu ylittäessään esimerkiksi alueellisia ja ikäryhmien välisiä rajoja. Lokalisointimenetelmät nojasivat joukkoistamisen menetelmiin. Esimerkiksi Näkymä sovelluksen lokalisointi prosessissa järjestelmän käyttäjiä pyydettiin havainnoimaan tiettyä ilmiötä tai siihen liittyviä asioita omassa ympäristössään ja kirjaamaan havainnot ylös. Tämän jälkeen käyttäjien havainnot käytiin läpi keskitetysti ja liitettiin osaksi ilmiön tietomassaa. Tällainen toimintatapa ja mahdollisuus hyödyntää laajaa asiantuntijaverkostoa ja sen moninaisia näkökulmia, rikastuttaa ennakkoinnissa käytettävissä olevaa tietoa ja se myös parantaa järjestelmän käyttäjien mahdollisuuksia tunnistaa ilmiöitä ja ymmärtää niiden vaikutuksia tarkemmin.

Selvityksen aikana kävi myös ilmi se, kuinka keskeinen osa visualisoinnilla on ennakointijärjestelmien tarjoamassa hyödyssä. Visualisointi auttoi ymmärtämään ilmiön koostumusta ja sen eri osiin liittyviä tietoja, kuten on tehty esimerkiksi World Economic Forumin Strategic Intelligence, kuva alla, ja Futures Platformin tapauksissa. Käyttäjä pystyy selaamaan ilmiöitä ja huomaamaan millaisia riippuvuussuhteita ilmiön sisällä ja ilmiöiden välillä on olemassa. Sama tieto tai havainto voi siis olla osa useampaa ilmiötä ja yhteen ilmiöön voi liittyä useita erilaisia havainnoitavia kohteita ja tietoja. Visualisoinnilla voidaan myös auttaa käyttäjää hahmottamaan vilkaisemalla esimerkiksi ilmiön vaikuttavuutta, laajuutta, muutosta ajassa ja alueella sekä mahdollista reagointitarvetta. Myös mikrotason visualisointia voidaan tehdä esim. merkitsemällä hyvin pienelle kartalle ilmiöön liittyviä tietoja. Visualisointi ei ainoastaan selkeytä ilmiöiden rakenteen hahmottamista, vaan myös edistää niiden ymmärrystä laajemman yleisön keskuudessa.



Kuva 2. Ruutukaappaus World Economic Forumin Strategic Intelligence työkalun visualisoinnista. (weforum.org 2024)

Lisäksi tunnistettiin, että monet ennakointialustat on suunnattu erityisesti tietyille kohderyhmille. Markkinointi ja tuotesuunnittelu ovat esimerkkejä aloista, joille räätälöidyt ennakointityökalut voivat tarjota merkittävää lisäarvoa. Tämä kohdennus mahdollistaa alustojen erikoistumisen tiettyjen toimialojen tarpeisiin, mikä puolestaan voi johtaa tarkempiin ja relevantimpiin ennusteisiin. Kohteenamme oli perehtyä hyvinvoinnin ja turvallisuuden ilmiöitä ennakoiviin alustoihin, näiden saatavuus on lähes olematonta. Tilastoihin perustuvia kansallisia järjestelmiä tunnistettiin ja kokeiltiin, mutta niiden hyödyntäminen hankkeen tavoitteiden mukaisesti on käytännössä mahdotonta. Oppia toki voi ottaa niiden tavasta esittää ilmiöitä, niiden koostumuksia, keskinäisiä suhteita ja vaikutuksia toisiinsa ja ympäristöön.

Kokonaisuudessaan ennakointijärjestelmien tutkimus on ollut valaiseva kokemus. Opittiin niiden toimintalogiikasta, sekä ihmisten roolista järjestelmien hyödyn tuottamisessa. Nyt pystymme paremmin

arvioimaan niiden potentiaalia tulevaisuuden ilmiöiden ymmärtämisessä ja ennustamisessa. Vaikka emme pystyneet hankkimaan Futures Platformia, saamamme opit eri ennakointijärjestelmistä ohjaavat digitaalisen ratkaisun kehittämistä ja ennakointitiedon hyödyntämistä.

5. Havaintoja nykyisistä tietojärjestelmistä ja sovelluksista tekoälypohjaisiin ratkaisuihin

Hankkeessa koottiin tietoja nykyisistä käytössä olevista työkaluista, joita voi hyödyntää tekoälypohjaisissa ratkaisuissa. Listaus työkaluista koottiin Excel-tiedostoon (liite 2) ja työkalut luokiteltiin niiden käyttötarkoituksen mukaan. Työkalulistaukset ovat omina taulukoina Excel-tiedostossa, luokittelu on seuraava:

TOOLS 1: Ohjelmistokehityksen työkaluja tekoälypohjaisiin ratkaisuihin,

TOOLS 2: Muita ohjelmistokehityksen työkaluja,

TOOLS 3: Työkaluja datan analysointiin,

TOOLS 4: Testbed-tarjontaa ja datan jakamisen alustoja.

Kaksi ensimmäistä taulukkoa ovat ohjelmistokehittäjien käyttämiä työkaluja. Tekoälypohjaisten ratkaisujen kehittäminen vaatii tekoälymallin käyttämisen datan hallinnoimisen, mallin kehittämisen, käyttöönoton ja ylläpitämisen työkaluja sekä yleisiä ohjelmistokehityksen hallinnoinnin työkaluja. Taulukossa TOOLS 1 on nimetty nämä eri työkalut seuraavasti ohjelmiston käyttötavan mukaan:

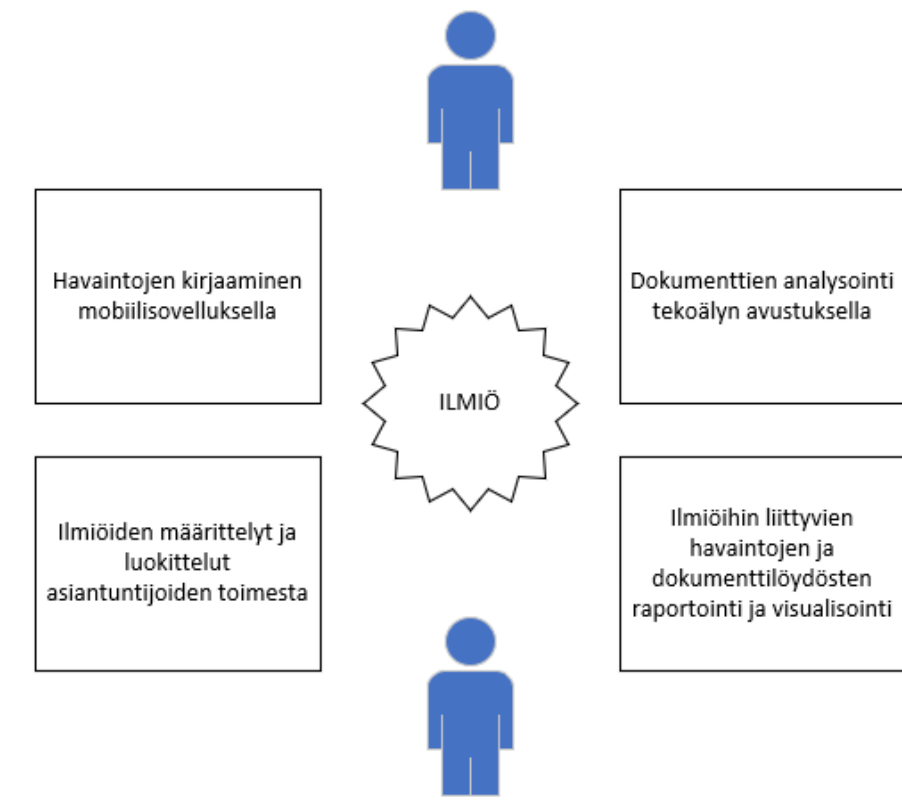
1. data management eli datan / (tiedostojen) hallintatyökalu
2. data refinery and cleansing eli datan käsittelyn automatisointi, jotta sitä voidaan hyödyntää eri järjestelmien välillä
3. data analysis eli data-analyysin työkalu
4. data integration and transformation
5. model building eli tekoälymallin kouluttamisen työkalu
6. model deployment eli tekoälymallin käyttöönoton työkalu
7. model monitoring and assessment eli tekoälymallin toiminnan valvonta ja arviointi työkalu
8. code asset management eli tietokoneen ohjelmiin liittyvien tiedostojen hallintatyökalu
9. development environment eli ohjelmistokoodin kehitysympäristö
10. execution environment eli ohjelmiston tuotanto-/käyttöympäristö

Excel-taulukossa TOOLS 3 on työkaluja data-analyysiin sekä tutkimusartikkelien hakuun ja analysointiin. Ohjelmistot on suunnattu tutkijoille. Ohjelmistoista osa on kaupallisia palveluja, jotka tarjoavat eri vaihtoehtoja lisensseille. Osa tarjoaa rajatun ilmaisen vaihtoehdon (freemium) ja osa on täysin ilmaisia palveluja.

Yhteiskehitystyöpajoissa käsitellään sovelluskehityksen tarpeita ja kokeilukohteita. Kehitettävän kohteen valinnalla voi olla useita kriteerejä kuten nopea toiminnan parantaminen, organisaation omien työkalujen hyödyntäminen, innovatiivisuus.

6. Demo laadullisen tiedon hyödyntämisestä

Hankkeessa ollaan kehittämässä eri toimijoiden ja/tai verkostojen käyttöön suunnattua demoa, joka osoittaa mobiilisovelluksen avulla tapahtuvan havaintojen kirjaamisen ja tekoälypohjaisen raporttien analysoinnin potentiaalin tiedolla johtamisen tukena. Alla kuvatun demon on tarkoitus tarjota alueellemme uudenlainen näkökulman siihen, miten laadullista dataa voidaan tuottaa, kerätä ja tuoda sitten hyödynnettäväksi päätöksenteossa eri toimijatasoilla verkoston laajuudessa.



Kuva 3 Demon toiminnallinen kuvaus.

Mobiilisovelluksen avulla käyttäjät voivat kirjata havaintojaan suoraan kentältä. Kirjaamisessa on huomioitu erilaisten ilmiöiden ja havaintojen tarpeet esimerkiksi siten, että havainnon tueksi voi halutessaan liittää tekstiä, kuvaa, ääntä ja videota.

Tekoälyn avulla pystytään analysoimaan erilaisia julkisesti saatavilla olevia raportteja, muistioita ja suunnitelmia. Näistä voidaan muodostaa haluttuun ilmiöön kytkettyjä koosteita, jotka auttavat johtajia ymmärtämään paremmin toimintaympäristöään ja reagoimaan nopeasti muuttuviin olosuhteisiin.

Demon eri osia kehitetään nopeina ketterinä kokeiluina. Valittujen ratkaisujen hyötyjä ja ymmärrettävyyttä käsitellään Lapin hyvinvointi- ja turvallisuusosastojen koostuvassa työryhmässä hankkeen aikana säännöllisesti.

7. Demon komponentti 1 – Mobiilisovellus- laadullisen ilmiöpohjaisen tiedon kerääminen

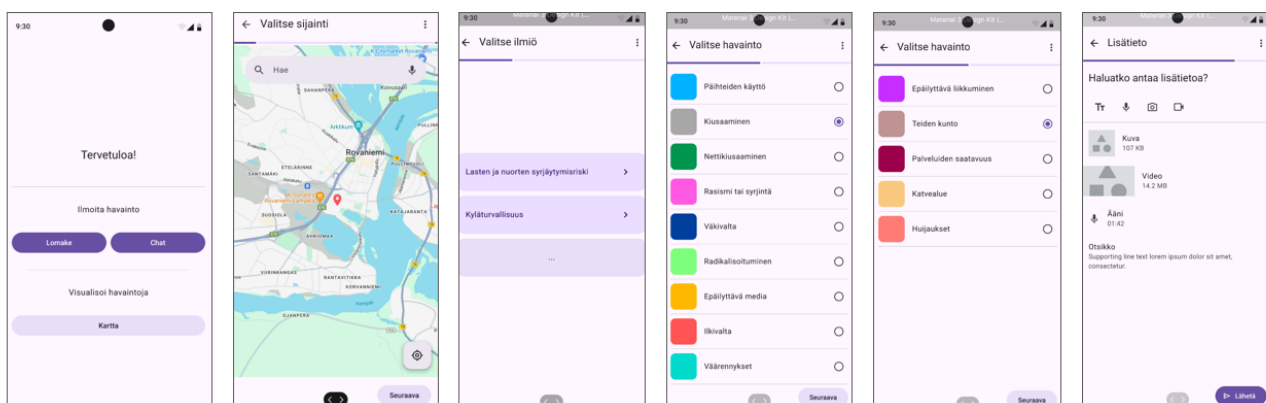
Kansalaisten ja ammattilaisten osallistuminen havaintotietojen keräämiseen mobiilisovellusten kautta on merkittävä askel kohti tiedolla johtamista tilanteissa, joissa tietoa on ollut perinteisesti hyvin vähän saatavilla tai sitä on ollut pirstaloituneena eri toimijoilla. Mobiilisovellukset tarjoavat helpon ja saavutettavan tavan kerätä arvokasta informaatiota reaaliajassa. Tämä voi nopeuttaa reagoitua ympäristön muutoksiin ja parantaa toimijoiden kykyä kohdistaa resurssejaan. Esimerkiksi sosiaali- ja terveydenhuollon alalla mobiilisovellukset voivat tukea kansalaisten omatoimisuutta terveyden ylläpidossa ja mahdollistaa ammattilaisten työn keventämisen, kun rutiininomaisia tehtäviä voidaan automatisoida ja tieto liikkuu sujuvasti eri toimijoiden välillä.

Tässä hankkeessa tullaan yhteiskehittämään mobiilisovellusta, jonka avulla kansalaiset ja ammattilaiset voivat kirjata havaintoja tarkastelun kohteeksi valituista ilmiöistä. Yhteiskehittäminen on nykyaikainen lähestymistapa, joka hyödyntää eri sidosryhmien asiantuntemusta ja kokemuksia sovelluksen kehitysprosessissa. Tämä prosessi mahdollistaa käyttäjälähtöisen suunnittelun, jossa loppukäyttäjien tarpeet ja toiveet otetaan huomioon jo varhaisessa vaiheessa. Yhteissuunnittelussa voidaan myös tunnistaa millaisia koulutus- ja opastustarpeita, jotta kaikilla kansalaisilla olisi mahdollisuus ja valmius käyttää sovellusta.

Keväällä järjestetyissä seudullisissa työpajoissa esiteltiin opiskelija Onni Courtin toteuttama ensimmäinen mobiilisovelluksen toiminnallinen malli osallistujille. Alla toteutetun mallin kuvat sekä lomake- että chat - pohjaisesta mallinnuksesta. Chat-pohjaisessa mallissa on käytetty pohjalla Rikksentorjuntaneuvoston turvallisuuskävelyopasta (2011).

Havaintotiedon kerääminen

1. mallinnus (lomake)

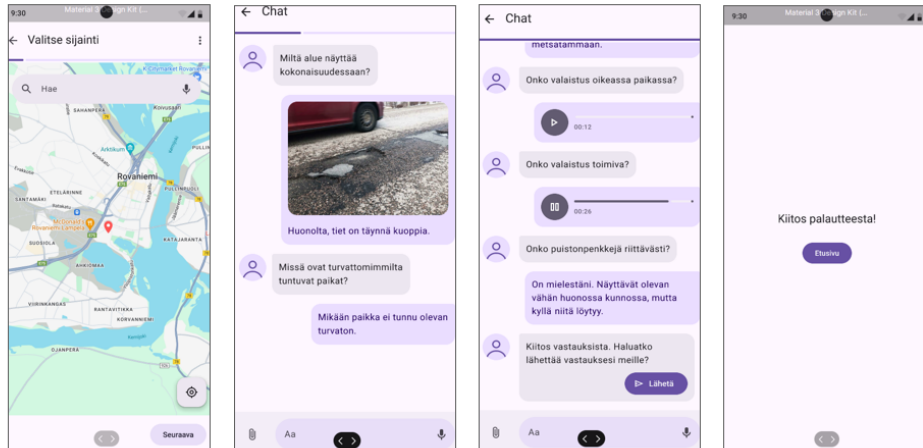


Mallinnus: Onni Court

Kuva 4: Mobiilisovelluksen 1. mallinnus, lomakepohjainen

Havaintotiedon kerääminen

1. mallinnus (chat)



Mallinnus:
Onni Court

Kuva 5: Mobiilisovelluksen 1. mallinnus, chat -pohjainen

8. Demon komponentti 2 - Tekoälyratkaisu - laadullisen ilmiöpohjaisen tiedon analysointi

Olemme kehittäneet tekoälyn tukeutuvaa tiedostojen analysointiratkaisua. Tekoäly on mullistanut laadullisen tiedon analysoinnin, tarjoten uusia näkökulmia ja mahdollisuuksia tiedon syvällisempään ymmärtämiseen. Laadullisen tiedon, kuten teksti- ja puheaineistojen, analysointi on perinteisesti aikaa vieviä tehtäviä. Tekoälyn avulla voidaan automatisoida ja nopeuttaa tätä prosessia ja tuottaa toimijoille tärkeää tietoa ilmiöistä ja niiden ilmentymistä paikallisesti.

Koneoppiminen ja luonnollisen kielen käsittely ovat keskeisiä tekoälyn osa-alueita, jotka mahdollistavat monimutkaisten tietokokonaisuuksien analysoinnin ja arvokkaiden oivallusten esiin tuomisen. Tämä ei ainoastaan nopeuta analyysiprosessia, vaan myös paljastaa piilotettuja yhteyksiä ja trendejä, jotka voivat jäädä huomaamatta ilman tekoälyn apua.

On kuitenkin tärkeää muistaa, että tekoälyn tehokas hyödyntäminen edellyttää laadukkaan ja relevantin datan saatavuutta, sekä ymmärrystä siitä, miten tekoälyä sovelletaan erilaisiin analyysitarpeisiin. Tekoäly voi tuoda merkittävää lisäarvoa laadullisen tiedon analysointiin, mutta sen käyttöönotto vaatii huolellista suunnittelua ja kriittistä suhtautumista datan laatuun ja analyysin tavoitteisiin.

8.1 Ketterä kokeilu tekoälyrajapinnan käytöstä dokumenttien käsittelyssä

Hankkeen yksi demon mahdollisuuksia havainnollistava ketterä kokeilu kevään 2024 aikana oli ohjelma, joka automatisoi dokumenttien tekstisisällön analysoinnin generatiivisen kielimallin avulla. Kokeilussa käytettiin OpenAI:n kehittämän ChatGPT-kielimallin kahta versiota: GPT-3.5 Turbo ja GPT-4 Turbo. Kokeilun tavoitteena oli määritellä työvaiheet pdf-muotoisten dokumenttien käsittelyssä, tekoälymallin käyttö tekstin analysoinnissa sekä tekoälyn antamien tulosten muotoilussa. Kokeilussa kirjoitettiin useampi ohjelmaskripti dokumenttien käsittelyyn. Alla on kuvaus ratkaisun toimintaperiaatteista ja siihen kuuluvien pythonohjelmaskriptien käyttötarkoituksesta.

Usean dokumentin käsittelyn periaatteet:

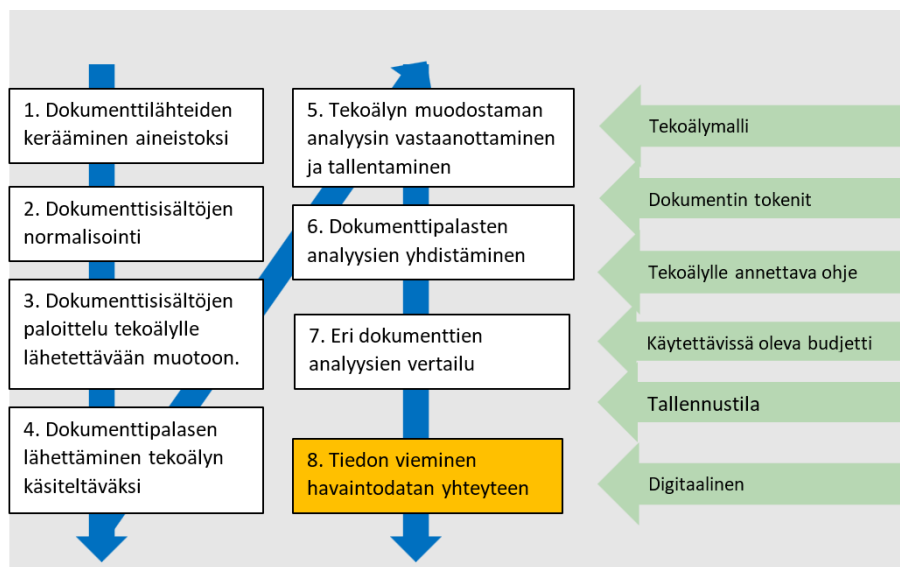
1. Yksi ohjelmaskripti käsittelee kaikki samantyyppiset dokumentit yhdellä kertaa. Lähdedokumenttia ei muokata, tulos tallennetaan uuteen tiedostoon ja nimitään.
2. Yksi ohjelmaskripti tekee yhden tai kaksi selkeästi rajattua toimenpidettä. Tarkoituksena on arvioida skriptin toimintaa ja sen antamia tuloksia selkeästi rajatussa tehtävässä.
3. Skriptien tuloksena olevat tiedostot ovat luettavissa muilla ohjelmilla (esim. tekstinkäsittely, taulukkolaskenta) ja muokattavissa. Tiedoston tekstisisällön muokkaus ei estä ohjelmaskriptin ajamista.

Ohjelmaskriptien käyttötarkoitus ja ajojärjestys:

1. *Ohjelmaskripti 1: Muuntaa pdf-dokumentin tekstitiedostoksi.* Skripti hakee pdf-muotoisesta dokumentista (.pdf) relevantin tekstisisällön ja tallentaa sen tekstitiedostona (.txt). Konversio tehdään kaikille skriptille osoitetuille pdf-dokumenteille.
2. *Ohjelmaskripti 2: Arvioi tekstitiedoston sisällön pituuden.* Ensin skripti arvioi kuinka monessa osassa teksti tulee lähettää ChatGPT:n käsiteltäväksi. Seuraavaksi se pilkkoo tekstisisällöt osiin tekstileikkeiksi, jotka tallennetaan omina tiedostoinaan.
3. *Ohjelmaskripti 3: Analysoi tekstitiedoston sisällön osissa.* Skripti lähettää tekstileiketiedoston yksi kerrallaan analysoitavaksi ChatGPT:lle. Tuloksena tulevan analyysin skripti tallentaa uuteen tiedostoon. Käsittely toistetaan kaikille skriptille osoitetuille tekstitiedostoille. Samaa alkuperäistä dokumenttia koskevat analyysit kootaan yhteen tiedostoon.
4. *Ohjelmaskripti 4: Muodostaa yhteenvetotiedoston tekstisisällön osista.* Skripti kokoaa jokaisen dokumentin osalta yksittäisistä analyyseistä yhteenvedon ja tallentaa ne taulukkomuotoiseen tiedostoon.

8.2 Kokeilun tuloksia

Jo yllä kuvattu verraten yksinkertaisen kokeilun avulla selvitettiin automaattiseen dokumenttien käsittelyyn liittyviä haasteita ja hytu-tietoa käsittelevän demoon määrittelyyn tarkennusta vaativia kohtia. Alla olevassa kuvassa kooste vaadituista työvaiheista vasemmalla ja oikealla digitaaliseen toimintaympäristöön liittyviä huomioita.



Kuva 6. Tekoälyn hyödyntämisen vaatimukset.

8.2.1 Dokumenttilähteiden kerääminen aineistoksi.

Kokeilua toteutettaessa hankkeen tietoperustan kerääminen oli vielä kesken. Vielä ei ollut tunnistettu laadullisen hytu-tiedon tietolähteitä riittävällä tarkkuudella, jotta niitä olisi voitu tuoda mukaan kokeiluun. Lisäksi kokeilu keskittyi dokumenttien analysoinnin teknisen toteutuksen mahdollisuuksiin. Tässä vaiheessa haluttiin välttää keskustelulta mitä käyttöehtoja tekoälypalveluja tuottavan kaupallisen toimijan OpenAI:n lisenssiehdot tarkkaan ottaen sisältävät ja millä tavalla Euroopan Union tekoälyasetuksen vaatimukset tulee huomioida toteutuksessa.

Kaupalliset toimijat kuten OpenAI voivat tarjota erilaisia palveluja itse tekoälymallien lisäksi. Tarjoajan lisenssiehdot ja kansalliset lait määräävät mitä palveluissa tarjotaan. Esimerkiksi ChatGPT:n palveluun on jo mahdollista tallentaa dokumentteja, joita käytetään tekoälyn ohjaamien kyselyjen tukena.

Kokeiluun valitut dokumenttilähteet olivat kokoelma pdf-muotoisia raportteja, joiden sisällöt, toteutustavat ja pituudet vaihtelivat suuresti. Vanhimmat dokumentit olivat vuodelta 2021. Dokumentit tallennettiin paikalliselle kovalevyille. Kokeilussa oli viisi dokumenttia.

Tekoälyohjelma käsittelee sille syötettyä hytu-tekstiä saamansa ohjeen mukaisesti. Mitä yksityiskohtaisempi ohje tekstisisällön analyysistä tekoälytyökalulle voidaan antaa, sitä parempi tulos tekoälyltä saadaan. Tästä syystä hytu-tietoa koskevien dokumenttien tekstin rakenteen olisi hyvä olla yhdenmukainen. Esimerkiksi tieteellisistä julkaisuista löytyy tieto artikkelin nimestä, kirjoittajista, jornaalista, tavoitteista, tutkimusmenetelmistä, tuloksista ja jatkotoimenpiteistä. Jos tietoa ei ole annettu dokumentissa, sitä ei voi sieltä löytää eikä tekoälymalli saa alkaa "hallusinoita" eli generoida keksittyä tietosisältöä.

Hytu-tietoa voi olla erityyppisissä dokumenteissa kuten raporteissa, muistioissa ja pöytäkirjoissa. Nämä dokumenttilähteet on tunnistettava. Dokumenttien rakenne ja sisältö on tarkistettava, jotta voidaan arvioida mitä tietoja niistä saadaan mukaan automaattiseen tekstin analysointiin. On arvioitava soveltuvatko ne käytettäväksi avoimena datana, mikä niiden säilytyspaikka voisi olla ja kuinka usein dokumenttilähteisiin päivitetään uutta tietoa. Jos dokumenttien tulee olla saatavilla useille organisaatioille, yhteinen tietovarasto voisi olla toimivin ratkaisu. Esimerkiksi avoindata.fi-palvelu on yksi mahdollinen avoimen datan tuottajien ja käyttäjien tietojärjestelmä.

8.2.2 Dokumenttisisältöjen normalisointi.

Dokumenttilähteiden sisältöjen muuttaminen samanmuotoiseen muotoon auttaa dokumenttien tietosisältöjen vertailussa. Tässä kokeilussa pdf-muotoiset dokumentit muunnettiin pelkkää tekstisisältöä käsittäviksi tiedostoiksi. Pdf-muotoiset dokumentit sisältävät monentyyppisiä elementtejä, joista osa määrittelee dokumentin metatiedot, osa dokumentin ulkoasun ja osa dokumentin tietosisällön. Tässä kokeilussa tarvittiin kustakin dokumenttilähteestä ainoastaan hytu-tietoa koskeva tekstimuotoinen tietosisältö. Kokeilun tavoitteena oli, että jokainen dokumentti analysoidaan vain kerran ChatGPT:llä. Kokeilussa oletettiin, että ChatGPT:n versio 3.5:lle voi lähettää vain tekstimuotoista sisältöä. Lisäksi haluttiin varmistaa, että tekoäly käsittelee vain hytu-tietoa koskevaa sisältöä.

Pdf-muotoiset tiedostot voi tallentaa toiseen tiedostomuotoon vaikkapa Adobe Acrobatissa. Tässä kokeilussa konversio haluttiin automatisoida. Paras ratkaisu oli hyödyntää ohjelmaskriptissä 1 avoimen lähdekoodin Xpdf pdftotext-komentorivityökalua. Pdftotext-työkalu valitsee pdf-dokumentin sivulta alueen, josta otetaan talteen tekstiä sisältävät elementit. Lisäksi tiedon voi hakea vain tietyiltä sivuilta käsiteltävästä dokumentista.

Pdf-dokumenteissa tekstimuotoisena on otsikoiden ja leipätekstin lisäksi mm sisällysluettelo, sivunumeroinnit, lähdeluettelo. Nämä tiedot eivät ole oleellisia tiedon automaattisen käsittelyn näkökulmasta. Kokeilussa käytetyt dokumenttilähteet olivat rakenteeltaan eri muotoisia, esimerkiksi sivunumeroinnit olivat eri dokumenteissa eri kohdilla sivua. Skriptin ajon jälkeen oli tarkistettava tekstitiedoston sisältö ja muutettava määrittelyä miltä alueelta teksti otetaan. PDF-dokumenteja oli kokeilussa viisi. Lyhyimmän dokumentin pdf-tiedostossa oli 25 sivua ja pisimmässä 71 sivua. Yksi teksteistä oli englanninkielinen, muut suomenkielisiä.

Tuloksena olevat tiedostot ovat muokattavissa tekstinkäsittelyohjelmalla kuten Windows-käyttöjärjestelmän mukana tulevalla Notepad (suom. Muistio) -ohjelmalla. Tekstin muunnos toimii hyvin uusilla pdf-dokumenteilla. Tekstisisällöt olivat oikeassa järjestyksessä ja suomenkielisen tekstin merkistö oli oikein. Vanhemmissa pdf-dokumenteissa tulisi tarkistaa dokumentin rakenne ennen konversiota. Notepad:ssa dokumentit olivat pituudeltaan 19981 (lyhin) merkkiä ja pisin 443041 merkkiä.

Jos tulevassa demossa halutaan käsitellä automaattisesti tekoälyä hyödyntävän ohjelman avulla esimerkiksi säännöllisissä organisaatioiden välisissä tapaamisissa tuotettuja dokumentteja kuten muistioita tai pöytäkirjoja, tulisi luoda dokumenttipohjia, jotka kuvaavat riittävän laadukkaan tietosisällön.

Dokumenttien tiedostomuodot tulisi tunnistaa. Dokumenttien ei tarvitse olla pdf-muotoisia. Rakenteeltaan (layout, otsikkotasot) erityyppiset dokumentit tulisi tunnistaa.

8.2.3 Dokumenttisisältöjen paloittelu tekoälylle lähetettävään muotoon

ChatGPT-kielimallilla on rajoitus kuinka pitkiä tekstejä se voi käsitellä. Kielimalli ei käsittele sille lähettyä tekstiä sanoina vaan yksittäisten sanojen osina eli *tokeneina*. Suomenkieliset tekstit kuluttavat tokeneita enemmän kuin esimerkiksi englanninkieliset tekstit. Lisäksi Open AI:n palvelu hinnoittelee tokeneiden mukaan jokaisen sille lähetettävän käsittelypyynnön. OpenAI:n rajapinnassa on ohjelmistokirjasto *tiktoken*, jonka avulla voidaan laskea ChatGPT:n käyttämien tokenien määrä kielimallille lähetettävälle tekstile.

Jotta pitkistä teksteistä voi tehdä analyysin tekoälyn kielimallin avulla, on teksti pätkittävä osiin ja lähetettävä jokainen tekstileike erikseen kielimallin käsiteltäväksi. Ohjelmaskripti 2 pilkkoo sille annetut dokumentit osiin.

Kokeilussa jokaisen tekstidokumentin pituus tokeneina laskettiin ja sen jälkeen teksti pilkottiin tekstileikkeiksi. Tokeneiden määrä laskettiin pelkästään analysoitavan dokumentin tekstin mukaan. Tokeneiden osuus analysoitavan tekstin merkkimäärästä oli suomenkielisillä teksteillä n. 39 %. Englanninkielisen tekstin tokeneiden osuus oli 27 %.

Skripti pilkkoi analysoitavan tekstin osiin n. 3000 tokenin määrän mukaisesti. Lyhyin dokumentti jaettiin kolmeen tekstileikkeeseen, pisin dokumentti 41:een tekstileikkeeseen. Analysoitavan tekstileikkeen koko yhdessä kielimallille lähetettävässä kutsussa oli siis keskimäärin 2900 tokenia.

Kielimalleille lähetettävä tekstisisältö olisi hyvä olla semanttisesti samankaltaista. Paloitellut teksti tulisi tarkistaa ja liittää yhteen samaa asiaa käsittelevät osat. Myös ns turhat osat kuten sisällysluettelot tai hakemistot voisi poistaa. Täysin automaattisena tekstien jakamista osiin ei siis välttämättä hyödytä toteuttaa. Sisällysluettelot ja hakemistot lisäävät tokeneiden kulutusta.

8.2.4 Dokumenttipalasen lähettäminen tekoälyn käsiteltäväksi

ChatGPT:lle lähetettävä kutsu käsitellä sille lähetettyä tietosisältöä sisältää *promptin*. Käytännössä *promti* on ChatGPT:lle annettu tekstimuotoinen ohje siitä mitä tekoälymallin halutaan tekävän sekä analysoitava tekstisisältö. Ohjeen kirjoittaminen, *promptaus*, on yksi tekoälyn käyttämisen perustaidoista. Ohje määrittelee, mitä kielimallin halutaan analysoivan sille annetusta tietosisällöstä ja missä muodossa kielimallin tulisi lähettää vastaus takaisin.

OpenAI:n hinnoittelu riippuu ChatGPT:n kielimallin versiosta. Sen lisäksi että kielimallille lähetettävillä prompteilla on pituusraja, myös kutsujen määrää rajoitetaan sekä sitä, kuinka usein kielimallille voi lähettää kutsuja.

Ohjeen muodostaminen oli tämän kokeilun vaikein tehtävä. Koska dokumentit olivat hyvin erimuotoisia ja niiden käyttötarkoitus oli eri, ohjeesta muodostui pyyntö analysoida tekstin rakenne.


```
"""
Tehtäväsi on auttaa hyvinvointi- ja turvallisuusalan asiantuntijaa muodostamaan tilannekuva hyvinvoinnin ja turvallisuuden
ilmiöistä verkkosivulle, johon yksilöidään turvallisuuspoikkeamia ja ilmiöitä ajankohtaisista havainnoista. Kirjoita yhteenveto
annetusta Hytu-tekstistä, joka on eroteltu kolminkertaisilla pienennys- ja suuremmuusmerkillä.

Saat ohjeen, jossa kuvataan, mitä sisältöjä Hytu-tekstistä halutaan yhteenvetoon. Ohje on [[[merkkien ja ]]]-merkkien
sisällä. Tee Hytu-tekstin analyysi seuraavien vaiheiden mukaan:

1. Analysoi Hytu-tekstin sisältö saamasi ohjeen perusteella. Arvioi jokaisen ohjeessa mainitun aiheen osalta, löytyykö Hytu-
tekstistä aihetta vastaavaa sisältöä.

2. Tee jokaisesta ohjeessa annetusta aiheesta yhteenveto Hytu-tekstistä. Jos et löytänyt ensimmäisessä vaiheessa sisältöä
aiheeseen Hytu-tekstistä, siirry seuraavaan aiheeseen.

3. Kirjoita jokaisesta analysoimastasi aiheesta yhteenveto sillä kielellä, jolla Hytu-teksti on kirjoitettu. Käytä asiallista
kirjakieltä. Käytä kokonaislauseita, virkkeitä ja kappaleita. Käytä Hytu-tekstissä käytettyjä sanoja yhteenvedossa. Käytä
ohjeessa annettuja aiheita otsikoina yhteenvedossa.

Ohje:
[[[

"Kirjoittajat":[KIRJOITTAJAT: Kirjoittajien kokonimet]
"Julkaisun otsikko":[JULKAISUN OTSIKKO:Julkaisun kaikki otsikot]
"Julkaisuvuosi":[JULKAISUN JULKAISUVUOSI]

"Tutkimuksen tavoite ja teoria":[TUTKIMUKSEN TAVOITE JA TEORIA: Tekstissä käsiteltyjä tutkimuksen tavoitteita ja
tutkimuksessa tarvittuja teorioita ja käsitteitä.]

"Tutkimuksen kohderyhmä ja tutkimusasetelma":[TUTKIMUKSEN KOHDERYHMÄ JA TUTKIMUSASETELMA: Teksissä mainitut tutkimukseen
osallistuneet henkilöt. Kohderyhmä, jolle on hyötyä tutkimuksen tuloksista.]

"Tutkimusote ja tutkimusmenetelmät":[TUTKIMUSOTE JA TUTKIMUSMENETELMÄT: Tutkimuksessa käytetty metodologia aineiston keruun,
käsittelyn ja tulosten esittämisen osalta.]

"Keskeiset tulokset":[KESKEISET TULOKSET: Tutkimuksen tärkeimmät tulokset ja arvio niiden reliabiliteetista, validiteetista
ja hyödynnettävyydestä. ]

"Keskeiset päätelmät ja suositukset tulevista tutkimuksista":[KESKEISET PÄÄTELMÄT JA SUOSITUKSET TULEVASTA TUTKIMUKSESTA:
Arvio, mitä tuloksista seuraa ja mihin asioihin seuraavan tutkimuksen tulisi keskittyä.]

]]]

Hytu-teksti: <<<{text_for_analysis}>>>
"""
```

Kuva 7. Promptin sisältö.

Tässä kokeilussa neljän dokumentin tekstisisältöjen käsittely ChatGPT 4-kielimallilla vaati yhteensä 86 kutsua. Jokainen kutsu sisälsi yllä olevan ohjeen dokumentin tekstin analysoinnista ja sekä tekstileikkeen dokumentin tekstisisällöstä. Ohjeen lisääminen kielimallille lähetettävään promtiin lisää kulutettujen tokenien määrää. Tämän ohjeen osuus kaikkien lähetettyjen kutsujen tokeneista oli n. 26 %.

OpenAI:n hinnoittelu riippuu kielimallin versiosta. GPT 4 on huomattavasti kalliimpi kuin GPT 3.5. Tässä kokeilussa GPT 4 oli n. 20 kertaa kalliimpi kuin GPT 3.5.

Dokumenttilähteiden määrittelyn lisäksi kielimallille lähetettävän ohjeen määrittely on tärkeässä roolissa, jotta tuleva demo voisi tuottaa laadukasta tietoa päätöksenteon tueksi. Vähintäänkin eri dokumenttityypeille tulisi kehittää oma käsittelyohje eli prompti.

Koska kielimallien käyttöä rajoitetaan, saadut resurssit tulisi kohdentaa hyödyllisesti. Kielimallin käsiteltäväksi tulisi lähettää semanttisesti samankaltaista sisältöä, jotta se voi muodostaa järkeviä koosteita käsittelemästään tekstistä.

Pitkien tekstien jakaminen osiin tarkoittaa, ettei kaikkia ohjeessa määriteltyjä analyysikohtia ole mahdollista löytää yhden tekstileikkeen sisältä. Kielimallin tulee osata käsitellä oikein puuttuva sisältö, se ei saa hallusinoita. Kielimallin version tulee siten olla riittävän hyvätasoinen, ettei se ala hallusinoita eli tuottaa satunnaista sisältöä.

8.2.5 Tekoälyn muodostaman analyysin vastaanottaminen ja tallentaminen

ChatGPT-kielimalli tuottaa vastauksen sille annettua ohjetta soveltamalla. Vastauksen generointi kuluttaa tokeneita. Kielimallin kuluttamat tokenit koostuvat kutsussa lähetetyistä tokeneista (Context tokens) ja vastauksessa generoiduista tokeneista (Generated tokens).

Tässä kokeilussa ei ylitetty sallittujen tokenien määrää tai kielimallille sallittujen kutsujen rajoitusta. Kaikkiin ChatGPT:lle lähetettyihin pyyntöihin saatiin vastaus. Vastausten saaminen neljän dokumentin analysoinnista kesti n. 1 h 51 minuuttia. Kokeilussa ChatGPT tuotti yhteenvetoja neljän dokumentin tekstileikkeistä keskimäärin n. yhden ja puolen minuutin välein. Osa kutsuista käsiteltiin erittäin hitaasti. Hitain vastaus saapui 46 minuutissa.

Tekoälymallin muodostaman analyysin sisällön laatu riippuu vahvasti käytetyn tekoälymallin versiosta. Tässä kokeilussa GPT 3.5 muodostama analyysi oli käyttökelpoton. GPT 4 seurasi annettua ohjetta ja palautti suurinpiirtein halutussa muodossa vastauksen. Vastausten otsikointi oli tärkeä osa analyysin rakennetta ja se jäi usein pois ilmaisen version käsitellessä tekstileikkeitä.

Tekoälyn muodostaman vastauksen tallentaminen on ongelmallista koska vastaus on tekstimuotoista. Samaa dokumenttia koskevat vastaukset tallennettiin yhteen tiedostoon. Koska dokumentti käsiteltiin osissa jokaiselle vastaukselle, merkittiin otsikko, joka kertoo kutsujärjestyksen ja ajankohdan.

8.2.6 Dokumentti palasten analyysien yhdistäminen

Tässä vaiheessa haluttiin muodostaa lopullinen tiivistelmä yhden dokumentin sisällöstä, jossa on otsikoiden alla seuraavat tiedot: Kirjoittajat, Julkaisun otsikko, Julkaisuvuosi, Tutkimuksen tavoite ja teoria, Tutkimuksen kohderyhmä ja tutkimusasetelma, Tutkimusote ja tutkimusmenetelmät, Keskeiset tulokset, Keskeiset päätelmät ja suositukset tulevasta tutkimuksesta. Yhtä dokumenttia koskevien tekoälymallin muodostamien analyysien yhdistäminen toteutettiin myös Chat GPT:n avulla. Tekoälymallille lähetettiin kaikki yhtä dokumenttia koskevat vastaukset ja pyydettiin tekoälymallia yhdistämään ne.

Myös tässä tehtävässä näkyy ero tekoälymallien versioissa. Käyttötarkoitukseltaan laadukkaampi versio tekstien analyysistä saatiin maksulliselta GPT 4 -mallilta. Vastaus tuli myös muotoilla taulukoksi. Taulukon muodostaminen oli ongelmallista, koska analyysien yhdistäminen ei vastannut pyydettyä muotoa. Tässä tehtävässä tekoälyn tuli muodostaa taulukon solut sille annettuja otsikointeja käyttäen.

8.2.7 Tiedon vieminen havaintodatan yhteyteen

Hankkeessa toteutettavassa demossa on ajatuksena, että kentältä eri toimijoiden tuottamaa havaintoaineiston yhteyteen voidaan tuoda muuta tietoa. Tämä muu tieto voi olla esimerkiksi tärkeistä dokumenttilähteistä poimittuja suosituksia, ohjeita, huomioita. Tässä kokeilussa ei vielä suunniteltu, miten havaintotiedon ja eri dokumenttilähteistä kerättävän tiedon välinen linkki muodostettaisi.

Tulevassa demon suunnittelussa ja toteutuksessa tulee arvioida havaintoina kerätyn tiedon laatua ja määritellä ilmiö, johon se kuuluu. Tietolähteenä olevat dokumentit tulee tunnistaa ja sekä havainnosta saatuun tietoon ja ilmiöön perustuen. Demon ajatellessa tuottavan ajankohtaista tietoa matalalla kynnyksessä päätöksentekoon.

Kerättyjen havaintojen, olemassa olevien tietolähteiden ja päätöksenteon välinen prosessi ja digitaalisten työkalujen hyödyntäminen on yksi osa tulevista toimenpiteistä.

8.3 Kokeilun tekniset tiedot

Ohjelmointikieli: Python 3.12.3

Open AI ChatGPT tekoälymallien versiot: gpt-3.5-turbo-0125 ja gpt-4-turbo-2024-04-09

PDF-konversiotyökalu: Xpdf pdftotext versio win 4.05

9. Jatkosuunnitelma

Seuraavassa vaiheessa jatketaan demon yhteiskehittämistä pienemmän monitoimijaisen ryhmän kanssa. Heidän kanssa Järjestetään yhteensä kuusi (6) kahden tunnin mittaista työpajaa, joissa jatkamme digitaalisen työkalun demon kehittämistä. Alustavasti sovitut työpajapäivät ovat 21.8., 10.9., 1.10., 30.10., 13.11. ja 27.11.

Tämän työvaiheen aikana valitaan tarkemmat ilmiöt, joiden ympärille digitaalisen työkalun demo tullaan rakentamaan. Samalla määritellään tarkemmin, mitkä työvaiheet ovat tällä hetkellä ongelmallisia ja miten demo voisi auttaa niissä.

10. Liitteet

Liite 1. Listaus digitaalisista sovelluksista

Liite 2. Listaus työkaluista

11. Lähteet

Hirvonen, M., Itsenäisyyden juhluvuoden lastensäätiö (Itla), Teams -palaveri 25.1.2024

Itsenäisyyden juhluvuoden lastensäätiö (Itla), Näkymä-sovellus, viitattu 12.6.2024,
<https://itla.fi/blogi/havainnoista-tiedolla-johtamiseen-sisaltapain-lahtevan-muutoksen-tukeminen/>

Lakkala, H., yksityinen sähköpostikeskustelu 20.3.2024, viestin saaja Eeva Helameri.

Pohjois-Suomen osaamiskeskus (Poske): Tarina-auttamisen chat-palvelu. Viitattu 12.6.2024
<https://ekollega.fi/-/ta-marraskuu2023>

Pyykkönen, J., 2021. Havainnoista tiedolla johtamiseen – Tilannehuone-prototyypin esittely, Lapsuuden rakentajat alueellisen oppimisverkoston päivät 17.–18.11.2021, PTC Services Oy. Viitattu 4.1.2024,
<https://www.youtube.com/watch?v=oDsC2xYbzwQ&t=28s>

Rikoksentorjuntaneuvosto, 2011. Turvallisuuskävelyt – Opas järjestäjille ja osallistujille. Erweko Oy 2011.
<https://rikoksentorjunta.fi/documents/5235988/5514012/Turvallisuusk%C3%A4velyopas+2011/7745871f-e74f-403e-ba11-fef620c72506/Turvallisuusk%C3%A4velyopas+2011.pdf?t=1509720380000>

Suomen itsenäisyyden juhlarahasto (Sitra), 2024. Datatalouden kiinnostavimmat. Viitattu 12.6.2024
<https://www.sitra.fi/hankkeet/datatalouden-kiinnostavimmat/#valitut>

Weforum.org 2024. Strategic Intelligence. World Economic Forum. Viitattu 1.8.2024 [Strategic Intelligence \(weforum.org\)](https://weforum.org)