



VALTIOVARAINMINISTERIÖ

Lausuntojen esianalysointi ja koostaminen tekoälyn keinoin -kokeilun loppuraportti

Loppuraportin malli

Oikeusministeriö

VM/2359/02.02.03.09/2018

Versio 0.2

25.6.2019

Sisällys

Sisällys	2
Dokumentin versiohistoria	2
1. Yhteenveto	3
2. Kokeilun toteutuminen	3
2.1. Kokeilun tiedot	3
2.2. Kokeilun rahoitus, kustannukset ja henkilötyöpäivät	5
2.3. Hankintakäytännöt	5
2.4. Riskienhallinta	5
2.5. Kokeilun tavoitellut hyödyt ja niiden toteutuminen	6
3. Kokeilun päättäminen	7
3.1. Kokeilun opit	7
3.2. Kokeilun kokemusten jakaminen	8
3.3. Kokeilun hyödyntäminen	8

Dokumentin versiohistoria

Versio	Päiväys	Laatija	Muutoksen kuvaus
0.1	20.6.2019	SD	Ensimmäinen versio
0.2	25.6.2019	SD	Viimeistely versio

1. Yhteenveto

Tämä dokumentti on uuden toimintamallin tai teknologiaratkaisun toiminnan todentamiseen tähtäävän [Lausuntojen esianalysointi ja koostaminen tekoälyn keinoin -kokeilun](#) loppuraportti.

2. Kokeilun toteutuminen

2.1. Kokeilun tiedot

Kokeilun tavoitteena oli arvioida luonnollisen kielen käsittelymenetelmien (NLP, natural language processing) ja tekoälyn hyödyntämismahdollisuuksia lausuntomenettelyn yhteydessä erityisesti lausuntopalautteiden käsittelyssä, mutta samalla arvioida koko lausuntomenettelyprosessia tukevien järjestelmien, työkalujen ja menetelmien kehittämismahdollisuuksia.

Kokeilu toteutettiin oikeusministeriön ja Lingsoftin yhteisinä työpajoina, johon osallistuivat Lingsoftin asiantuntijat sekä oikeusministeriöt asiantuntijat ja lainvalmistelijat. Kokeilu sisälsi seuraavat vaiheet:

- lausuntoprosessin ja siihen liittyvien käytötapausten läpikäynnin,
- lausuntopalvelusta saadun lausuntodatan analysoinnin ja soveltuvien käsittelymenetelmien tunnistamisen,
- sekä uusien toimintamallien ja työkalujen ideoinnin.

Kokeilun tuloksena saatiin näkemys sekä lausuntopalvelun datan soveltuvuudesta että erilaisten menetelmien hyödyistä lausuntomenettelyssä. Saatuja tuloksia voidaan hyödyntää myös yleisemminkin tekoälyn hyödyntämistä pohdittaessa.

Keskeisimpiä havaintoja lausuntopalvelun datan osalta:

- nykyinen palvelu on sellaisenaan parantanut vastausten sisältöä ja keräämistä sekä hyödyntämistä valmistelutyössä
- tallennettu data on kuitenkin vaihtelevaa sisällöllisesti ja rakenteellisesti - osin kirjavien käyttötapojen vuoksi - mikä hankaloittaa koneellista analysointia ja datan jatkokäsittelyä sellaisenaan. Datan laadun parantamiseksi löytyi ratkaisuja, ja
- tallennettua tekstidataa voidaan tässä käyttötapauksessa pienin lisäyksin täydentää sekä yhdenmukaistaa ja rakenteistaa automaattisin menetelmin niin, että prosessointityökalujen kehittäminen ja koneellinen analysointi helpottuvat.

Keskeisimpiä tuloksia datan käsittelymenetelmien, tekstianalytiikan ja koneoppimisen soveltuvuuden osalta:

- yhdenmukaistettu ja rakenteistettu data tehostaa sekä manuaalista tiedon käsittelyä että koneellista analysointia, mikä mahdollistaa edelleen työtehtäviin liittyvän datan esikäsittelymenetelmien kehittämisen
-

- soveltuva ratkaisu on yleensä kokoelma toisiaan täydentäviä menetelmiä, joiden soveltuvuus määräytyy niin käytettävissä olevan datan ominaisuuksien kuin tehtäväsäällön ja saavutettavien etujen perusteella
- tässä käyttötapauksessa tehokkaimmiksi menetelmiksi osoittautuivat tekstidatan rakenneanalyysi ja perusmuotoistaminen esimerkiksi datan analysoitavuuden, hakuindeksoinnin, asiasanoittamisen, tilastollisen sisältöanalyysin, tekstin koneluettavuuden, merkityksellisten termien poiminnan, kopioitujen tekstien löytämisen ja epäoleellisen tietosisällön tiivistämisen kehittämisessä. Myös manuaalinen avainsanoitus tukee aineiston luokittelulla tulkintojen tekemistä
- hyötyjä on saatavilla automatisoimalla lausuntodatan rikastamista esimerkiksi asiasanoituksella, luokitteluilla, ontologioilla ja sanastoilla. Varausina tuetaan myös sentimenttianalyysin, asiasanoituksen ja epäoleellisen osoittamisen toteutusta koneoppimismenetelmien avulla, sillä näissä on huomioitava tässä käyttötapauksessa myös tarvittavan kehityspanoksen suuruus verrattuna mahdollisiin hyötyihin
- menetelmien hyöty syntyy kokemusperäisessä - tulkintaa ja laajan taustamateriaalin hallintaa vaativassa - asiantuntijatehtävässä ennemminkin tulkinnan tukemista avustavien prosessointityökalujen ja tiedon esikäsitelyn kautta kuin pyrkimällä koneellisen tulkinnan tai päättelyn toteuttamiseen
- vähän hyötyjä antaviksi menetelmiksi tässä käyttötapauksessa osoittautuivat klusterointi, kaikki lausunnot kattava analysointi sekä koneellisesti suoritettavat oleellisen osoittaminen ja tiivistelmän tuottaminen. Kaikissa näissä keskeisinä haasteina ovat lausuntodatan sisällön heterogeenisyys ja datan käsittelyssä inhimillisen tulkinnan keskeisyys, jossa korostuu usean erilaisen kokemusperäiseen tietoon perustuvan asiayhteyden samanaikainen hallinta
- tekstianalytiikan ja koneoppimisen menetelmät ovat tehokkaita tekstimassojen käsittelyyn ja hahmottamiseen tavoilla, jotka ovat mahdottomia tai ainakin tehottomia inhimillisenä työnä. Menetelmät edellyttävät käsiteltävältä datalta riittävää määrää sekä säännönmukaisuutta sisällöllisesti ja rakenteellisesti, mikä tulisi huomioida jo järjestelmien suunnitteluvaiheessa

Kokeilu osoittaa, että tekstianalysointi- ja prosessointityökaluilla on mahdollista tukea lausuntopalautteen käsittelyyn liittyvää asiantuntijatyötä, mutta varsinaisen lausuntokoosteen automatisointi ei nykyteknologian avulla vielä onnistu. Tulosten pohjalta ideoitiin jatkokehittämissideoita lausuntopalveluun ja sähköiseen lausuntomenettelyyn:

- lausuntopalvelun käyttöliittymän lisätoiminnot valmistelijoille
 - kenttien ja taulujen täydennykset tietomalliin ja vastaavasti tietovarastoon uusien ja täydennettävien metatietojen osalta sekä uusien toimintojen tarvitsemille välitalennuksille
 - uudet automatisoidut esi- ja jälkikäsitteilyprosessit lausuntopyynnön tekemiseen, lausuntojen ja tiivistelmien tuottamiseen
 - koneoppimisen mallinnusdatan taustaprosessit
 - koneoppimisen algoritmit ja menetelmät
-

- APIn laajentaminen (ja korjaaminen) tarvittavin osin

2.2. Kokeilun rahoitus, kustannukset ja henkilötyöpäivät

Kokeilun suunnitellut (käyttö- ja kirjausoikeuspäätöksen mukaiset) ja toteutuneet kustannukset euroina ovat eriteltynä omaan ja ostettuun työhön sekä muihin kustannuksiin seuraavat:

Kustannus	Suunniteltu €	Toteutunut €
Oma työ (nykyresursseilla tehtävä työ)	-	-
Oma työ (kokeiluun erikseen palkattavien resurssien työ)	-	-
Palvelujen ostot	50 000	49 000
Muut kustannukset	-	-
Kokonaiskustannus	50 000	49 000

Kustannukset eriteltynä rahoituslähteittäin euroina ovat seuraavat:

Rahoituslähde	Suunniteltu €	Toteutunut €
28.70.22 Hallinnon palveluiden digitalisoinnin tuki	50 000	49 000
Kokonaiskustannus	50 000	49 000

Kokeilun kustannusten ylitys/alitus johtuu pääosin seuraavista tekijöistä:

- Lingsoftin tarjous oli 49 000 euroa, eikä kokeiluun kohdistunut muita kustannuksia. Oikeusministeriön virkamiehet hoitivat kokeiluun osallistumisen virkatyönä.

Oman, kokeiluun erikseen palkatun henkilöstön toteutunut kustannus euroina ja henkilötyöpäivinä:

€	htp
-	-

2.3. Hankintakäytännöt

Hankinta toteutettiin valtion hankintakäsikirjan periaatteiden mukaisesti markkinakartoituksella, jossa kartoitettiin mahdollisia tarjoajia, joilla on ollut kokemusta luonnollisen kielen analysoinnista.

Markkinakartoitukseen osallistuneille yrityksille esiteltiin hanke ja pyydettiin tekemään ehdotus hankkeen toteutumisesta. Markkinakartoituksessa ja ehdotusten arvioinnissa kiinnitettiin huomiota toimittajan osaamiseen, kokemukseen, projektin toteutusmalliin ja hankinnan reunaehtoihin (aika- ja budjetti).

2.4. Riskienhallinta

Kokeilun riskien tilanne kokeilun päättyessä:

Riski	Lopullinen tila	Toimenpiteet	Toimenpiteiden vaikutus
-------	-----------------	--------------	-------------------------

Lausuntojen analysointiin liittyvien laatuvaatimusten täyttyminen ei onnistu	Lausuntojen analysointiin löydettiin riittävästi laadukkaita menetelmiä, joissa on hyötypotentiaalia.	Riittävän laava toimeksianto ja riittävän ajan varaaminen suunnitteluun.	Mahdollisesti toimittajan riittävän perehtymisen aihealueeseen sekä laajasti erilaisten menetelmien tutkimisen.
Aikatauluun liittyvät riskit, mm. lainvalmistelijoiden osallistuminen (muut hankkeet ja EU:n puheenjohtajuusvalmistelut aiheuttivat haasteita).	Aikatauluun liittyvät riskit eivät toteutuneet ja hanke saatiin aikataulussa toteutettua	Riittävä suunnittelu ja mahdollisuus reagoida muuttuvaan tilanteeseen nopeasti, esim. vaihtamalla työpajojen aikatauluja.	Toimenpiteillä onnistuttiin pysymään aikataulussa siten, että oikeusministeriön lainvalmistelijat pystyivät osallistumaan ja Lingsoft pystyi työstämään aineistoa riittävästi myös työpajojen välissä

2.5. Kokeilun tavoitellut hyödyt ja niiden toteutuminen

Kuvaa alla olevaan taulukkoon kehitettävän prosessin vaikuttavuus- ja asiakas-hyötypotentiaali hakemuksen mukaan ja arvioi sen toteutumista kokeilun jälkeen:

Arvio kehitettävän prosessin vaikuttavuus- ja asiakashyötypotentiaalista		
Tavoiteltava yhteiskunnallinen vaikuttavuus	Hyötyjen realisoituminen hakemuksen mukaan (keskeiset toimenpiteet ja vastuutahot vaikuttavuushyötyjen realisoimiseksi)	Arvio hyötyjen realisoitumisen toteutumisesta, jos kokeilussa rakennettu muutos otetaan tuotantoon
Lausuntojen käsittelyn tehostaminen, mikä vähentää lausuntokierrosten aiheuttamaa kuormaa ja edesauttaa hyvän hallintotavan mukaisen lausuntomenettelyn toteutumista mm. säädösvalmistelussa. Yhteisten raportointitapojen luominen, mikä lisää lausuntokierrosten läpinäkyvyyttä ja lausuntodatan jatkohyödyntämistä.	<p>Kehitettävillä analysointimenetelmillä valmistelijoiden tulisi kyetä analysoimaan ja raporttoimaan lausuntopalautetta nykyistä nopeammin ja tehokkaammin, ja siten keventää lausuntokierrosten kuormittavuutta.</p> <p>Samalla lausuntojen sisältämä tietosisältö voitaisiin saada nykyistä paremmin esille, jolloin myös niiden vaikuttavuutta voitaisiin paremmin seurata/raportoida.</p> <p>Kokeilun onnistumista ja yhteiskunnallisten vaikutusten realisoitumista arvioidaan seuraavasti:</p> <p>a) arvio esianalysoidun aineiston laadusta ja hyötypotentiaalista</p> <p>b) arvio automaation asteesta, eli kuinka lähelle julkaistavaa lausuntotiivistelmää päästään</p> <p>c) arvio tehokkuudesta, eli säästetystä ajasta</p> <p>d) arvio valmiin lausuntoraportin laadusta</p>	Kokeilu osoittaa, että tekstianalysointi- ja prosessointityökaluilla on mahdollista tukea lausuntopalautteen käsittelyyn liittyvää asiantuntijatyötä, mutta varsinaisen lausuntokoosteen automatisointi ei nykyteknologian avulla vielä onnistu.

Kuvaa alla olevaan taulukkoon kehitettävän prosessin vaikuttavuus- ja asiakas-hyötypotentialiaali hakemuksen mukaan ja arvioi sen toteutumista kokeilun jäl-keen:

Arvio kehitettävän prosessin tuottavuuspotentiaalista		
Taloudelliset hyödyt	Hyötyjen realisoituminen (keskeiset toimenpiteet ja vastuutahot vaikuttavuushyötyjen realisoimiseksi)	Arvio hyötyjen realisoitumisen toteutumisesta, jos kokeilussa rakennettu muutos otetaan tuotantoon
Taloudelliset hyödyt liittyvät analysoinnista säästettyyn asiantuntijatyöhön.	Alkuvaiheessa hyödyt realisoituva selkeimmin yksittäisissä hankkeissa, joissa säästetty aika on suoraan kohdennettavissa muihin saman hankkeen vaativampiin/tuottavampiin valmistelutehtäviin.	Taloudellisten hyötyjen realisoitumista on vaikea arvioida kokeilun tuloksilla, sillä nyky-menetelmillä ei ole mahdollista poistaa kokonaan lausunnotiivistelmään tekemiseen vaadittua asiantuntijatyötä, vaan ainostaan tukea sitä. Menetelmien käyttöönotto voisi lisätä alussa jonkin verran työtä, mikä johtuu uusien työmenetelmien opettelusta, mutta varsinkin suuremmissa lausuntokierroksissa, jossa lausuntoja annetaan sadoittain, ja pidemmällä aikavälillä on kuitenkin mahdollista saada ajallistakin hyötyä uusien analysointimenetelmien käytöstä.

3. Kokeilun päättäminen

3.1. Kokeilun opit

Kokeilu mahdollisti laaja-alaisen ja perusteellisen tutustumisen erilaisiin tekoälymenetelmiin, mikä edesauttaa jatkossa tunnistamaan paremmin tekoälyyn liittyviä kehittämiskohteita (esim. säädösvalmistelussa), eri menetelmistä saatavaa hyötypotentialiaalia ja tekoälyyn liittyviä rajoituksia. Tärkeää oli myös huomata, että tekoälymenetelmien hyödyntämisessä parhaan lopputuloksen tarjoaa useampien eri menetelmien yhdistelmäkäyttö.

Kokeilussa onnistuttiin löytämään lausunnotmenettelyyn sopivia analysointimenetelmiä, joilla on mahdollista tehostaa lausuntojen läpikäymistä. Näiden menetelmien hyödyntämistä saatiin myös konkretisoitua työkaluehdotuksilla, jotka olisi mahdollista ottaa käyttöön lausuntopalvelussa. Lisäksi opittiin tunnistamaan lausunnotmenettelyn vaiheet, joita kehittämällä tekoälyn hyödyntämismahdollisuuksia voidaan tulevaisuudessa entisestään lisätä (mm. metadatan parantaminen).

Kaiken kaikkiaan kokeilu lisäsi merkittävästi ymmärrystä tekoälystä ja sen hyödyntämismahdollisuuksista, mikä on erittäin tärkeää, kun lähdetään miettimään tekoälyä hyödyntävien sovellusten kehittämistä ja hankkimista. Kokeilu osoitti

myös, miten tärkeää on saada substanssiasiantuntijat osalliseksi kehittämistyöhön. Tekoälyhankkeissa substanssiasiantuntijoiden rooli on jopa vieläkin merkittävämpi kuin muissa it-hankkeissa. Tämä on tärkeä havainto, kun mietitään esimerkiksi säädösvalmisteluun liittyvien tekoälyhankkeiden toteuttamista. Lainvalmistelijaresurssit ovat sidottuna säädösvalmisteluhankkeisiin, jolloin heidän osaamistaan voi olla vaikea saada tekoälyhankkeisiin.

3.2. Kokeilun kokemusten jakaminen

Kokeilua ja siitä saatuja kokemuksia on jo esitelty laajasti niin oikeusministeriön hallinnonalan kuin muiden ministeriöiden toimijoille (mm. säädösvalmistelun kehittäjille, oikeusministeriön hallinnonalan tietohallinnon asiantuntijoille). Kiinnostus tämän tyyppistä kokeilua kohtaan on ollut suurta, koska kokeilussa käydään huomattavasti laaja-alaisemmin läpi tekstianalytiikkaan liittyviä menetelmiä kuin muissa hankkeissa. Kokeilua on myös esitelty lyhyesti Lainvalmistelun kehittämisen yhteistyöryhmän loppuraportissa (Oikeusministeriön julkaisuja, Mietintöjä ja lausuntoja 2019:29) ja lainsäädäntötyön kehittämistä pohtineen valmisteluryhmän loppuraportissa (Valtioneuvoston kanslian julkaisuja 2019:17).

Kokeilusta julkaistaan laajempi raportti, jossa käydään yksityiskohtaisesti läpi kokeilun toteuttamista, hyödynnettyjä menetelmiä ja kehitysehdotuksia. Näin kokemuksia ja saatuja tuloksia on mahdollista jakaa laajasti hallinnossa ja siten hyödyntää eri organisaatioiden tekoälyyn liittyvissä hankkeissa tai niiden suunnitelmassa. Raportti julkaistaan [Hankeikkunassa](#) kesän 2019 aikana.

Kokeilusta saatuja tuloksia tullaan myös esittelemään kokeilun kannalta keskeisille tahoille (mm. erilaiset tekoälyverkostot), jotta kokemuksia saadaan levitettyä mahdollisimman tehokkaasti.

3.3. Kokeilun hyödyntäminen

Kokeilun tavoitteena oli kehittää valmistelijoiden kanssa yhteistyössä sähköiseen lausuntomenettelyyn tekstianalysointi- ja prosessointityökaluja, joilla lausuntokierrosten aiheuttamaa kuormitusta valmisteluhankkeissa voitaisiin keventää. Lopputuloksena haluttiin saada aikaiseksi POC analysointimenetelmistä, joilla voitaisiin esianalysoitu lausuntokooste valmistelijoiden käyttöön.

Vaikka lausuntokoosteen luominen automaattisesti ei vielä nykyteknologian kanssa onnistu tai sellaisen toteuttaminen ei ole kustannustehokasta lausuntodatan luonteesta johtuen, kokeilussa löydettiin mielekkäät analysointikeinot, joilla on mahdollista tehostaa lausuntojen käsittelyä ja siten helpottaa valmistelijoiden työtä. Samalla kokeilussa sivuttiin laajemminkin laajojen tekstimassojen analysointiin liittyviä haasteita ja mahdollisuuksia, jolloin kokeilusta saadut hyödyt eivät vain rajoitu lausuntomenettelyyn.

Näin ollen kokeilun tuloksia ja kokemuksia on mahdollista hyödyntää sekä sähköisen lausuntomenettelyn edelleen kehittämiseen että laajemminkin lainvalmistelutyön digitalisoinnissa. Näissä kummassakin asiassa oikeusministeriöllä on koordinoiva rooli valtioneuvostossa, mikä edesauttaa kokeilun tulosten jatkojalostamista ja hyödyntämistä.