

26.10.2018

HENKILÖTIETOJA SISÄLTÄVIEN ASIAKIRJOJEN AUTOMAATTINEN ANONYMISOINTI JA SISÄLLÖNKUVAILU -HANKE

Asettaminen

Oikeusministeriö on tänään päättänyt asettaa hankkeen ”Henkilötietoja sisältävien asiakirjojen automaattinen anonymisointi ja sisällönkuvailu”.

Toimikausi

1.10.2018 – 30.9.2020

Hankkeen tausta

Julkisen sektorin eri toimijat tuottavat valtavia määriä tietoaineistoja ja dataa, jonka saaminen avoimesti muiden viranomaisten, yritysten ja kansalaisten käyttöön olisi hyödyllistä, mutta se ei ole mahdollista tietoon liittyvien tietosuojakysymysten takia. Esimerkiksi käräjäoikeudet antavat vuosittain lähes puoli miljoonaa ratkaisua, mutta niitä julkaistaan verkossa vain harvoin. Hovioikeudet antavat vuosittain noin 9 500 ratkaisua, mutta niistä julkaistaan verkossa hovioikeuksien kotisivuilla ja Finlex-palvelussa vain noin 50 ratkaisua vuodessa.

Keskeisenä esteenä ratkaisujen laajemmalle julkaisemiselle on anonymisoinnin/pseudonymisoinnin vaatimat henkilöresurssit ja käytännön työvälineiden puuttuminen. Verkossa julkaistavien ratkaisujen pieni määrä vaikuttaa siihen, että kansalaisten ja yritysten saatavilla olevat ratkaisuaineistot eivät ole kovin edustava otos tuomioistuinten ratkaisutoiminnasta.

Anonymisoinnin ohella toinen keskeinen haaste viranomaisaineistojen ja oikeustapausten julkaisemisessa ja hyödyntämisessä verkossa Semanttisen Finlexin (data.finlex.fi) kaltaisena datapalveluna ja toisaalta loppukäyttäjille tarkoitettuina

sovelluksina on aineiston sisällönkuvailu. Anonymisoinnin tapaan laajojen tietoaaineistojen sisällönkuvailu ja annotointi on hidasta, manuaalista työtä. Prosessin automatisoinnilla on saavutettavissa merkittäviä säästöjä.

Hankkeen tehtävät ja tavoitteet

Hankkeessa toteutetaan ja tuotteistetaan kaksi kieliteknologiseen tekoälyyn perustuvaa työkalua, oppiva anonymisointityökalu (ANOPPI) ja oppiva automaattinen annotointiväline (APPI), joiden avulla tuomioistuinratkaisujen ja muiden viranomaispäätösten anonymisointia ja sisällönkuvailua voidaan automatisoida. Molemmat työkalut voivat oppimalla tuoda esiin ja piilottaa tietoa asiakirjoista.

ANOPPI kykenee automaattisesti tunnistamaan ja merkkamaan anonymisoinnin kannalta keskeiset ilmaukset dokumentissa ja näiden väliset yhteydet, esimerkiksi eri tavoin tehdyt viittaukset samaan henkilöön.

Samaa teknistä ratkaisua ja ohjelmistoja voidaan käyttää myös aineistojen automaattiseen sisällönkuvailuun, jossa etsitään dokumentista sen sisällön kannalta keskeisiä käsitteitä. APPI mahdollistaa dokumenttien älykkään haun ja linkityksen muihin aineistoihin, esimerkiksi oikeustapausten yhdistämisen toisiin vastaaviin tapauksiin sekä lainsäädäntöön.

Toteutettavat sovellukset perustuvat avoimeen lähdekoodiin ja niitä voidaan jatkossa hyödyntää koko julkishallinnossa. Sovellusten ansiosta voidaan parantaa henkilötietoja sisältäneiden asiakirja-aineistojen sähköistä saatavuutta esimerkiksi päätöksenteossa ja tutkimuskäytössä.

Hankkeen organisointi

Hanketta johtaa ja koordinoi oikeusministeriö.

Hankkeen perusta on luotu aiemmissa kehittämishankkeissa. Nyt toteutettavien sovellusten teknisen toteutuksen osalta tehdään yhteistyötä Aalto-yliopiston ja Helsingin yliopiston HELDIG-keskuksen kanssa, jotka ovat toteuttaneet hankkeessa käytettävän teknologian ja Semanttinen Finlex -järjestelmän. Mukana hankkeessa on myös Edita Publishing Oy, joka on toteuttanut Finlexin tekniset ratkaisut ja ylläpitää järjestelmää.

Tiivistä yhteistyötä tehdään etenkin tuomioistuinten ja niiden asianhallintahankkeiden kanssa.

Ohjausryhmä

Hankkeen ohjausta varten kootaan eri hyödyntäjätahoja edustava ohjausryhmä, joka kokoontuu tarkastamaan hankkeen välitavoitteiden toteutumista alempana kuvatun aikataulun mukaisesti.

Projektiryhmä

Projektiryhmän tehtävänä on seurata hankkeen kokonaisuuden etenemistä sekä edistää ja koordinoita työpakettien toteuttamista. Projektiryhmä toimii tiiviissä yhteistyössä hankkeen eri yhteistyötahojen kanssa. Lisäksi projektiryhmä raportoi hankkeen etenemisestä ohjausryhmälle.

Projektiryhmän jäsenet ovat:

Aki Hietanen, kehittämisspäällikkö, OM
 Saara Packalén, projektipäällikkö, OM
 Tiina Husso, projektikoordinaattori, OM
 Oili Salminen, erityisasiantuntija, OM
 Mikko Aaltonen, erityisasiantuntija, OM
 Minna Tamper, tohtorikoulutettava, Aalto-yliopisto
 Eero Hyvönen, professori, Aalto-yliopisto ja HELDIG
 Jouni Tuominen, tutkija, HELDIG ja Aalto-yliopisto
 Sami Sarsa, tutkimusavustaja, Aalto-yliopisto ja HELDIG
 Risto Talo, ohjelmistokehityspäällikkö, Edita Publishing Oy
 Sari Korhonen, kehitysjohtaja, Edita Publishing Oy
 Arttu Oksanen, ohjelmistokehittäjä, Edita Publishing Oy

Hankkeen aikataulu ja työpaketit:

Hankkeen työ jakautuu kuuteen työpakettiin, joiden suunniteltu aikataulu on avattu alta löytyvästä taulukosta.

	Syksy 2018	Talvi 2019	Kevät 2019	Kesä 2019	Syksy 2019	Talvi 2020	Kevät 2020	Kesä 2020
Työpaketti 1								
Työpaketti 2								
Työpaketti 3								
Työpaketti 4								
Työpaketti 5								
Työpaketti 6								

työpaketit pitävät sisällään:

- Työpaketti 1: verkkoteknologiaan perustuvan oppivan anonymisointityökalun (ANOPPI) tuotteistaminen.
- Työpaketti 2: automaattisen sisällönkuvailun oppivan verkkopalvelun (APPI) tuotteistaminen.

- Työpaketti 3: Välineiden testaaminen todellisilla käyttäjillä tuomioistuimissa, oikeusministeriössä ja Edita Publishing Oy:ssä.
- Työpaketti 4: Anonymisointityökalun käyttöönotto tuomioistuinratkaisuisissa ja hallinnollisissa päätösaineistoissa.
- Työpaketti 5: Valittujen Semanttisen Finlexin aineistojen automaattinen sisällönkuvailu. Lisäksi Finlex-järjestelmään kehitetään älykäs semanttinen haku- ja suosittelutoiminnallisuus sekä mekanismi samankaltaisten oikeustapausten sisällöllistä hakua varten.
- Työpaketti 6: Työkalujen käyttöönottoa varten järjestelmistä laaditaan käyttöohjeet sekä järjestetään koulutustilaisuuksia sähköisiä oppimisympäristöjä hyödyntäen.

Hankkeen kustannukset

Hanke toteutetaan valtiovarainministeriön erillisrahoituksella, jolla kannustetaan valtion virastoja hyödyntämään tekoälyä, robotiikkaa ja analytiikkaa toimintansa uudistamisessa.

Kansliapäällikkö

Pekka Timonen

Hallintojohtaja

Arja Apajalahti-Laine

JAKELU

Oikeusministeriön osastot ja erillisyksiköt
Projektiryhmän jäsenet
Tuomioistuimet
AIPA-hanketoimisto
HAIPA-hanketoimisto