

26.10.2018

PROJEKT FÖR AUTOMATISK ANONYMISERING AV DOKUMENT SOM INNEHÅLLER PERSONUPPGIFTER OCH BESKRIVNING AV INNEHÅLLET I DOKUMENTEN

Tillsättande

Justitieministeriet har i dag beslutat att tillsätta ett projekt för automatisk anonymisering av dokument som innehåller personuppgifter och beskrivning av innehållet i dokumenten.

Mandattid

1.10.2018–30.9.2020

Bakgrund till projektet

Inom den offentliga sektorn produceras mängder av informationsmaterial och data som med fördel kunde finnas tillgängligt för andra myndigheter, företag och medborgare, men som på grund av dataskyddsfrågor inte kan publiceras offentligt. Till exempel tingsrätterna meddelar årligen nära till en halv miljon avgöranden, men dessa publiceras enbart sällan på webben. Hovrätterna meddelar årligen ca 9 500 avgöranden, varav endast omkring 50 per år publiceras på hovrätternas webbplatser och i Finlex-tjänsten.

Ett av de största hindren för en mer omfattande publicering av materialet är personalresurserna som behövs för anonymisering/pseudonymisering av avgörandena och att det inte finns praktiska verktyg för detta arbete. Det begränsade antalet avgöranden som finns på webben gör att informationsmaterialet medborgarna och företagen har tillgång till inte utgör något representativt urval av domstolarnas avgöranden.

Ett annat problem när det gäller att publicera och utnyttja myndighetsmaterial och rättsfall på webben i form av öppna data, såsom i tjänsten data.finlex.fi, eller som applikationer avsedda för slutanvändare, är produktionen av innehållsbeskrivningar för dokumenten. På samma sätt som anonymisering är produktionen av innehållsbeskrivningar och annotering av datamaterial långsamt manuellt arbete. Genom att automatisera denna process skulle man kunna uppnå betydande besparingar.

Projektets uppgifter och mål

I projektet produceras och produktifieras två verktyg som bygger på artificiell intelligens, dvs. ett självlärande anonymiseringsverktyg (ANOPPI) och ett självlärande annoteringsverktyg (APPI). Med hjälp dessa ska anonymiseringen och produktionen av innehållsbeskrivningar för domstolsavgöranden och andra myndighetsbeslut kunna automatiseras. Båda verktygen ska kunna lära sig att hämta fram och dölja information på dokumenten.

ANOPPI ska automatiskt kunna identifiera och markera de uttryck som är väsentliga med tanke på anonymisering samt sambanden mellan dem, t.ex. olika hänvisningar till samma person.

Samma tekniska lösning och programvara kan användas för generering av automatiska innehållsbeskrivningar så att verktyget hämtar fram de mest centrala begreppen ur innehållet. APPI ska möjliggöra smart sökning och länkning av dokument till annat material, t.ex. förena rättsfall med andra motsvarande fall och med relevant lagstiftning.

Verktygen som byggs upp i projektet ska basera sig på öppen källkod och i fortsättningen kunna användas inom hela den offentliga förvaltningen. Med hjälp av verktygen ska material som har innehållit personliga uppgifter kunna göras digitalt tillgängligt, t.ex. för beslutsfattande och forskning.

Tidsplan

Projektet leds och samordnas av justitieministeriet.

Beredningen baserar sig på arbetet som gjorts i tidigare utvecklingsprojekt. Det tekniska genomförandet av verktygen sker i samarbete med Aalto-universitets och Helsingfors universitets HELDIG-central, som skapat de tekniska lösningarna för projektet och systemet för öppna data i Finlex (data.finlex.fi). Också Edita Publishing Ab, som svarar för de tekniska lösningarna och upprätthållandet av Finlex-databasen, deltar i projektet.

I projektet ingår också aktivt samarbete speciellt med domstolarna och deras ärendehanteringsprojekt.

Arbetspaket 6								
--------------------------	--	--	--	--	--	--	--	--

Arbetspaketens innehåll:

- Arbetspaket 1: produktifiering av ett självlärande anonymiseringsverktyg som bygger på webbt Teknologi (ANOPPI).
- Arbetspaket 2: produktifiering av en självlärande webbtjänst för generering av automatiska innehållsbeskrivningar (APPI).
- Arbetspaket 3: Testanvändning av verktygen i praktiken vid domstolarna, justitieministeriet och Edita Publishing Ab.
- Arbetspaket 4: Ibruktagande av anonymiseringsverktyget i domstolsavgöranden och administrativt beslutsmaterial.
- Arbetspaket 5: Generering av automatiska innehållsbeskrivningar för visst material i systemet för öppna data i Finlex (data.finlex.fi). I Finlex-systemet skapas dessutom en smart funktion som möjliggör semantiska sökningar och kan rekommendera lämpligt material samt en mekanism för sökning av liknande rättsfall.
- Arbetspaket 6: Utarbetande av anvisningar för ibruktagande av verktygen och ordnande av utbildning med hjälp av de elektroniska inlärningsmiljöerna.

Projektets kostnader

Projektet genomförs med särskild finansiering från finansministeriet som uppmuntrar statliga ämbetsverk att utnyttja artificiell intelligens, robotik och analys i utvecklingen av verksamheten.

Kanslichef

Pekka Timonen

Förvaltningsdirektör

Arja Apajalahti-Laine

DISTRIBUTION

Justitieministeriets avdelningar och fristående enheter
Projektgruppens medlemmar

Domstolarna
AIPA-projektbyrå
HAIPA-projektbyrå