

Europeanan sanomalehtiprojekti käynnistyi kesän aikana



Ryhmä eurooppalaisia kumppaniorganisaatioita on yhdistänyt voimansa “European Newspapers” – projektiin ja haluaa tarjota seuraavien kolmen vuoden aikana yli 18 miljoonaa sanomalehtisivua Europeana-palveluun. Europeanan kautta voi yhdestä palvelusta löytää miljoonia digitoituja kirjoja, maalauksia, elokuvia, museoesineitä sekä arkistoaineistoa ympäri Eurooppaa.

Europeanan sanomalehtiprojekti on rahoitettu Euroopan komission Kilpailukyvyyn ja innovoinnin puiteohjelmasta 2007–2013 (CIP). Se tähtää sanomalehtien viemiseen Europeanaan Eurooppalaisen kirjaston kautta ja sanomalehtien sisällön tekniseen jälkikäsitteilyyn digitoidun aineiston käytettävyyden parantamiseksi.

Projektiin osallistuvat kirjastot jakavat digitoituja sanomalehtiä ja kokotekstiaineistoa Europeanan kautta. Projekti saattaa sanomalehtisisällön käyttöön erityisen käyttöliittymän kautta. Käyttöliittymä liitetään Europeana-portaaliin ja mahdollistaa yksittäisten sanojen tai lauseiden hakemisen sanomalehtiteksteistä. Standardimaisten kirjastoluetteloiden hakumahdollisuudet rajoittuvat yleensä vain päivämäärä- tai nimekehakuihin.

Projekti osoittaa haasteet, jotka linkittyvät digitoituihin sanomalehtiin ja niiden jatkojalostamiseen, kuten niiden optiseen merkien tunnistukseen (OCR), artikkeleiksi segmentoimiseen tai niiden optiseen asetteluun (OLR), nimientiteettien (NER) ja sivutyypin tunnistamiseen. OCR tarkoittaa käsin, koneella tai tulostetusta tekstistä skannattujen kuvien muuntamista sähköiseksi koneluettavaksi tekstiksi. OLR on sähköisten artikkelien merkitsemistä skannatusta sanomalehden sivusta, mikä sisältää enemmän kuin yhden artikkelin. NER pyrkii paikantamaan ja luokittelemaan tekstistä määriteltäviä standardimaisia yksittäisiä osia kuten henkilöiden nimet, organisaatiot, sijainnit jne.

Projektissa arvioidaan myös digitoidun aineiston jalostusteknologioiden laatua ja paikallisten metatietojen siirtoa Europeanan tietomallistandardiin tiiviissä yhteistyössä julkisen sekä yksityisen sektorin sidosryhmien kanssa.

Europeanan sanomalehtiprojektia johtaa Berliinin valtionkirjasto (Berlin State Library).

Projektiikumppanit:

Berliinin valtionkirjasto
Viron kansalliskirjasto
Helsingin yliopisto, Kansalliskirjasto
Ranskan kansalliskirjasto
CCS Content Conversion Specialists GmbH
Latvian kansalliskirjasto
Belgradin yliopisto
Dr. Friedrich Tessmann -kirjasto
Salfordin yliopisto

Hollannin kansalliskirjasto
Itävallan kansalliskirjasto
Hampurin yliopiston kirjasto
Puolan kansalliskirjasto
LIBER-järjestö
Turkin kansalliskirjasto
Innsbruckin yliopisto
Iso-Britannian kansalliskirjasto
Eurooppalainen kirjasto (TEL)

Europeanan pähkinänkuoressa

Europeanan on eurooppalaisten museoiden, kirjastojen, arkistojen ja audiovisuaalisten kokoelmien ja miljoonien digitoitujen julkaisujen monikielinen online-kokoelma. Tällä hetkellä Europeanan tarjoaa integroidun pääsyn 23 miljoonaan kirjaan, elokuvaan, maalauksiin, museoesineisiin ja arkistoasiakirjoihin noin 2200 sisällöntuottajalta eri puolilta Eurooppaa.

www.europeana.eu

Sanomalehdet -projektin edistystä voi seurata osoitteessa: www.europeana-newspapers.eu. Lisätietoja saatavilla Hans-Jörg Liederiltä tai Thorsten Siegmannelta info@europeana-newspapers.eu, Suomessa Tiina Höltältä Kansalliskirjaston digitointi- ja konservointikeskuksesta: tiina.holtt@helsinki.fi

The European Newspapers project is partially funded under the ICT Policy Support Programme (ICT PSP, http://ec.europa.eu/ict_psp) as part of the Competitiveness and Innovation Framework Programme by the European Community. This document reflects only authors' views. The EC is not liable for any use that may be done of the information contained therein.

