

# Приказ завршног састанка конзорцијума пројекта „Europeana Newspapers“

Наташа Дакић

dakic@unilib.bg.ac.rs

Дејана Каваја Станишић

kavaja@unilib.bg.ac.rs

Универзитет у Београду  
Универзитетска библиотека  
„Светозар Марковић“

**ДАТУМ ПРИЈЕМА РАДА:**

**ДАТУМ ПРИХВАТАЊА РАДА:**

06. април 2015.

14. април 2015.

За разлику од књига, новине никада нису претендовале да буду дуговечне. Основни циљ, донети свакога дана свежу информацију, не водећи често рачуна о квалитету папира на ком су се штампале, резултирао је у великом броју крчих, ломљивих оригинала који се данас чувају у библиотечким фондовима. Многе библиотеке су, да би их сачувале и омогућиле њихово коришћење, првобитно микрофилмовале многе наслове, а са развојем технологија започеле и процес скенирања материјала. Како је и микрофилмован и скениран материјал често тешко читљив или пак нечитљив те пружа само ограничене могућности претраге неопходно је пречистити текст и додатно га обрадити користећи нове алате и технологије чиме би се омогућило оптичко препознавање текста (OCR-Optical Character Recognition), сегментација чланака (OLR-Optical Layout Recognition) и препознавање именованих ентитета (NER-Named Entity Recognition), а у циљу што делотворнијег претраживања слика, чланака или појединих термина у тексту.

Сагледавање ових потреба довело је до покретања пројекта „Europeana Newspapers“ који је финансиран од стране Европске комисије у оквиру петог позива Програма подршке развоју политике информационо-комуникационих технологија (CIP ICT-PSP Best Practice Network

5thCall), и који је ових дана успешно завршен. Првобитно, у пројекат је било укључено 18 пуноправних партнера,<sup>1</sup> међу којима су неке од најзначајнијих националних и универзитетских европских библиотека, као и Универзитетска библиотека „Светозар Марковић“. У међувремену, препознат је значај пројекта, те се он проширио и у његов рад су укључене бројне библиотеке, организације, као и други европски пројекти.

Анализа успешности реализованих циљева и постигнутих резултата пројекта, као и могућности даље сарадње међу партнерима на пројекту биле су неке од тема завршног састанка конзорцијума пројекта „Europeana Newspapers“. Овом дводневном састанку, који је одржан од 22. до 23. јануара ове године у Талину, у Националној библиотеци Естоније (Eesti Rahvusraamatukogu), присуствовали су само пуноправни пројектни партнери. Такође, састанку су присуствовали и лидери свих шест радних пакета који су предали извештаје за сваки радни пакет понаособ.

Клеменс Нојдекер (Clemens Neudecker), координатор пројекта, отварајући састанак, у свом излагању се осврнуо на задате циљеве и постигнуте резултате. У својој анализи, оценио је да су сви предвиђени циљеве и испуњени, док су неки чак и премашени. На пример, од предвиђених 10 000 000 дигитализованих страна које би биле

1 <http://www.europeana-newspapers.eu/consortium/project-partners/#project>

доступне и претраживе у пуном тексту, данас се на порталу Европске библиотеке (TEL-The European library)<sup>2</sup> налази 10 132 955 страна доступних у пуном тексту, док укупан број доступних дигитализованих страна износи 20 014 931, што укључује и стране које су уступили придружени партнери на пројекту. Такође, напоменуто је да је пројекат изнедрио и неке производе који нису били предвиђени пројектном документацијом а који су га обогатили. Наиме, Европеана је интегрисала апликацију за претрагу новинских колекција и тиме омогућила прегледање новина у пуном тексту и на порталу Европеана; пројектни партнери CCS (Content Conversion Specialists), Универзитетска библиотека у Хамбургу (SUBHN - Staats- und Universitätsbibliothek Hamburg) и Универзитет у Салфорду (USAL -University of Salford) развили су апликацију која омогућава прегледање дигитализованих новина на мобилним телефонима; док је Холандска национална библиотека (Koninklijke Bibliotheek) проширила свој рад везан за препознавање именских ентитета тиме што је развила алат који подржава вишезначност ентитета и њихово повезивање са идентификаторима у ДБпедији.

Током састанка истакнуто је да конзорцијум пројекта треба да обезбеди чврсту сарадњу са Европеаном, те да се приликом даље дисеминације пројекта пружају информације о обе приступне тачке, тј. и о Европеани и о Европској библиотеци; да се развије стратегија одржавања већ постојећих дигиталних колекција, као и методе за проширење новинских колекција и привлачење нових корисника. Даље, потребно је пратити квалитет достављених метаподатака и обезбедити да сви достављачи садржаја потпишу Споразум о размени података.

Алистер Данинг (Alastair Dunning) је испред TEL-а у свом извештају навео да је Европска библиотека прикупила, поред слика и страница у пуном тексту, и метаподатке, које је успешно конвертовала у Европски модел за метаподатке (EDM-Europeana Data Mode). Известно је присутне да је укупно у претраживач унето 3 342 наслова новина, што није коначан број будући да још треба унети завршни сет података који треба да доставе CCS и Универзитет у Инсбруку (UIBK -Universität Innsbruck), као и колекцију

новина из Турске. На крају свог излагања изнео је и одлуку борда Европеана и Европске библиотеке о спајању ове две организације, као и потенцијалне компликације које произилазе из ове одлуке, будући да је првобитно било планирано да ове две организације наставе свој рад као два одвојена ентитета.

Представници библиотеке партнера су у кратким петоминутним представљањима изложили своје резултате, запажања, очекивања и даље планове. Навели су изазове и потешкоће са којима су се сусрели радећи на овом пројекту, као и могуће начине интегрисања успешних резултата пројекта у свој будући рад. Изложили су своје планове на пољу дигитализације новинских колекција, што је најчешће укључивало проналажење начина за побољшавање квалитета већ прочишћеног текста, као и рад на неком новом пројекту.

Универзитетска библиотека је током реализације пројекта "Europeana Newspapers" дигитализовала 392.323 страница из 45 одабраних српских листова и новина, издатих у периоду од 1830. до 1944. године. За целокупан материјал урађен је комплетан OCR, чиме је омогућена претрага целокупног корпуса дигитализованог текста. Резултати оптичког препознавања карактера су били охрабрујући, имајући у виду да је целокупна новинска колекција (осим једног наслова) штампана на ћирилици, међутим како би омогућили бољу читљивост и претрагу дигитализоване новинске колекције, Универзитетска библиотека је успела да развије алат који би омогућио ручну корекцију METS/ALTO фајлова, а тиме и поправио квалитет страница доступних у пуном тексту. После првобитних потешкоћа, развијен је и алат за претрагу читаве колекције, који ће се користити у претрази и других дигиталних садржаја. Библиотека је у циљу промоције и што веће видљивости материјала доступних кроз овај пројекат, обезбедила промоцију путем јавног сервиса Радио Телевизије Србије у тридесетоминутној емисији, као и у многобројним чланцима објављиваним на интернет порталима. Ово је био уједно и садржај петоминутног представљања наших резултата на овом завршном састанку конзорцијума пројекта „Europeana Newspapers.“

Овај дводневни састанак који је протекао у

2 <http://www.theeuropeanlibrary.org/tel4/newspapers>

радној атмосфери, одавао је утисак да су партнери мотивисани да се фокусирају на долазеће задатке и активности у последњој фази пројекта. Тридесетак регистрованих учесника, сарадника

на пројекту, искористили су време да размене искуства, отворене су важне дискусије о могућим даљим активностима, као и сарадњи у оквиру неких других европских пројеката.