



START PROJEKTU EUROPEANA NEWSPAPERS

17 europejskich instytucji, wśród nich Biblioteka Narodowa, rozpoczęło realizację projektu *Europeana Newspapers*. W ciągu najbliższych trzech lat instytucje te udostępnią za pośrednictwem serwisów internetowych portalu Europeana ponad 18 milionów stron gazet, głównie z pierwszego dwudziestolecia XX w. Portal Europeana umożliwi dostęp do milionów zdigitalizowanych książek, obrazów, filmów, obiektów muzealnych i archiwalnych dokumentów znajdujących się w kolekcjach europejskich instytucji kultury.

Projekt *Europeana Newspapers* finansowany jest ze środków Komisji Europejskiej w ramach Programu Ramowego na rzecz Konkurencyjności i Innowacji 2007-2013 (CIP-ICT PSP), a jego celem jest agregacja wysokiej jakości kopii cyfrowych gazet za pośrednictwem The European Library.

Każda z bibliotek uczestniczących w projekcie udostępni za pośrednictwem portalu Europeana pełnotekstowe wersje zdigitalizowanych gazet. Celem projektu jest umożliwienie bezpośredniego dostępu do treści gazet za pomocą specjalistycznego interfejsu portalu Europeana, pozwalającego na wyszukiwanie całych fraz lub poszczególnych słów w zdigitalizowanym tekście.

Projekt stanowi odpowiedź na wyzwania związane z digitalizacją prasy i dotyczy przede wszystkim udoskonalenia metod Optycznego Rozpoznawania Pisma (OCR - Optical Character Recognition), segmentacji artykułów i Optycznego Rozpoznawania Układu Strony (OLR - Optical Layout Recognition), rozpoznawania nazw (NER - Name Entity Recognition) oraz rozpoznawania typu strony. OCR umożliwia komputerową konwersję rastrowych plików graficznych zawierających teksty drukowane, pisane maszynowo i odręcznie na tekst możliwy do przetwarzania w edytorach tekstu. OLR pozwala na cyfrowe wydzielenie artykułów z zeskanowanej strony zawierającej więcej niż jeden artykuł. NER jest narzędziem służącym do wykrywania i klasyfikacji pojedynczych elementów tekstu w ramach określonych kategorii, takich jak nazwiska, nazwy organizacji, nazwy geograficzne i in.

W ramach projektu dokonana zostanie również ocena jakościowa zastosowanych technologii oczyszczania zdigitalizowanego tekstu oraz transformacja lokalnych formatów metadanych, które zostaną dostosowane do standardu Europeana Data Model (EDM) w ścisłej współpracy partnerów z sektora publicznego i prywatnego.

Koordynatorem projektu jest Biblioteka Państwowa w Berlinie.

The Europeana 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.



Partnerzy projektu:

1. Biblioteka Państwowa w Berlinie (koordynator projektu)
2. Biblioteka Narodowa Holandii
3. Biblioteka Narodowa Estonii
4. Biblioteka Narodowa Austrii
5. Biblioteka Narodowa Finlandii
6. Biblioteka Państwowa i Uniwersytecka w Hamburgu
7. Biblioteka Narodowa Francji
8. Biblioteka Narodowa Polski
9. Uniwersytet w Salford
10. CCS Content Conversion Specialists GmbH
11. LIBER
12. Biblioteka Narodowa Łotwy
13. Biblioteka Narodowa Turcji
14. Biblioteka Uniwersytecka w Belgradzie
15. Uniwersytet w Innsbrucku
16. Dr. Friedrich Tessmann Library
17. The British Library
18. The European Library

O portalu Europeana

Europeana to wielojęzyczna wirtualna kolekcja, obejmująca miliony zdigitalizowanych obiektów z europejskich muzeów, bibliotek, archiwów i zbiorów audiowizualnych. Dziś Europeana umożliwia zintegrowany dostęp do ponad 23 milionów książek, filmów, obrazów, eksponatów muzealnych i archiwalnych dokumentów z ponad 2200 kolekcji w całej Europie.

www.europeana.eu

The Europeana 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.

