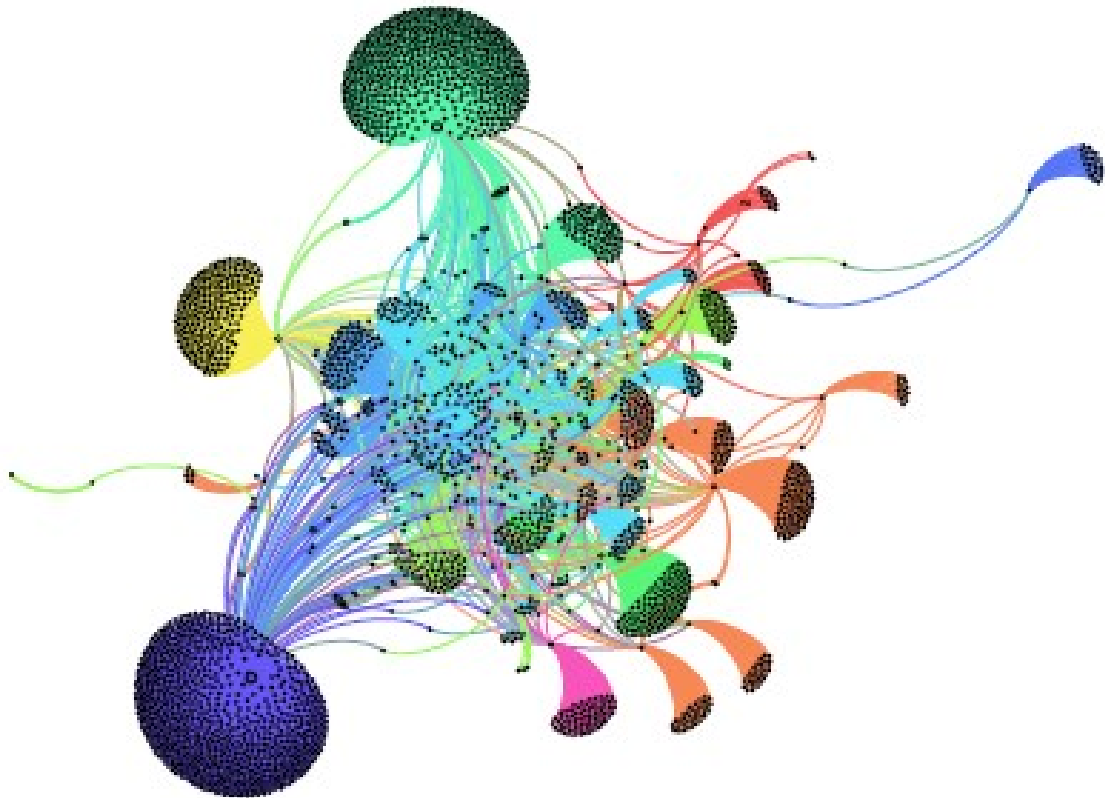


Humanistyczne projekty cyfrowe w Polsce

Marcin Werla, Maciej Maryl

Humanistyczne projekty cyfrowe w Polsce



Opracowanie:

Marcin Werla, Poznańskie Centrum Superkomputerowo-Sieciowe – IChB PAN

Maciej Maryl, Centrum Humanistyki Cyfrowej IBL PAN

Wersja 1.1 (26 VI 2014)

Poznań-Warszawa 2014

Licencja: CC BY 3.0 (<http://creativecommons.org/licenses/by/3.0/pl/>).



<http://rcin.org.pl>

SPIS TREŚCI

Wstęp	6
Podsumowanie otrzymanych ankiet	7
Aleph Polska (1).....	7
Biblioteka Narodowa (2)	7
Instytut Badań Literackich PAN (12).....	7
Instytut Języka Polskiego PAN (1)	8
Instytut Podstaw Informatyki PAN (13).....	8
Instytut Sławistyki PAN (3)	9
Instytutu Chemii Bioorganicznej PAN - Poznańskie Centrum Superkomputerowo-Sieciowe (3)	9
Muzeum i Instytut Zoologii PAN (1)	9
Politechnika Wroclawska (8).....	9
Polski Instytut Antropologii (1)	10
Polskie Towarzystwo Bibliologiczne (1).....	10
Uniwersytet Marii Curie-Skłodowskiej (2).....	10
Uniwersytet Pedagogiczny w Krakowie, Uniwersytet Jagielloński (1)	10
Uniwersytet Śląski (2).....	11
Uniwersytet Warszawski (27).....	11
Uniwersytet Wroclawski (2).....	13
Statystyki wybranych elementów ankiety	14
Rodzaj podmiotu realizującego projekt	14
Status projektu.....	14
Źródło finansowania.....	14
Słowa kluczowe	15
Typ projektu	16
Efekt projektu.....	16
Opisy projektów	17
Academica.....	17
Adam Mickiewicz 150 rocznica śmierci	20
Adaptowalny zespół specjalizowanych narzędzi do przetwarzania morfosyntaktycznego polskich tekstów z dziedziny określonej przez użytkownika	22
Aktualizacja kształcenia akademickiego bibliotekarzy pracujących w małych bibliotekach gminnych.	24
Analiza doświadczeń wybranych instytucji w dziedzinie długookresowej ochrony zasobów cyfrowych z wykorzystaniem systemu Rosetta firmy Ex Libris	26
Archiwum kobiet: Piszące	28
ATLAS	31
Automatyczna ekstrakcja informacji.....	33

Bibliograficzna baza danych światowego językoznawstwa slawistycznego iSybislaw	36
Bibliologiczna Biblioteka Cyfrowa	39
Biblioteka Cyfrowa KLF UW	41
Biuletyn Polonistyczny	43
Blog jako nowa forma piśmiennictwa multimedialnego	45
CESAR	47
Computational Stylistics Group	50
CORE	52
CorpoGrabber	54
Cyfrowa Biblioteka Narodowa Polona	56
Cyfrowa Etnografia - Bibliotek Cyfrowa Pilskiego Instytutu Antropologii (BC PIA)	59
Digitalizacja i udostępnienie w sieci kolekcji starodruków Biblioteki IH UW	64
Digitalizacja wydawnictw Stowarzyszenia Bibliotekarzy Polskich	66
Dygitalizacja słownika Lindego	68
Dziedzictwo narodowe i kulturowe, polskie ślady za granicą	70
Dziedzictwo narodowe, badania i upowszechnianie wiedzy o tradycji i historii	72
E-edukacja bez barier cyfrowych i społecznych	74
Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do roku 1772)	76
Federacja Bibliotek Cyfrowych	78
GeTClasS - system do anotacji semantycznej tekstu	81
IMPACT-UWAR	83
Inforex	86
Interaktywna mapa literacka	88
Internetowe repozytorium tekstów rozpraw doktorskich i preprintów innych prac z zakresu światowej slawistyki i dziedzin pokrewnych iReteslaw	91
Komputerowa Analiza Nieprojektowanego Odbioru	94
Komputerowe Wspomaganie Analizy Wywiadów Etnograficznych	97
Konstrukcja elektronicznych słowników (baz danych leksykalnych) i dezambiguacja znaczeniowa dla celów tłumaczenia automatycznego.	99
Korpus Politechniki Wrocławskiej (KPWr)	103
Korpus polszczyzny XVI wieku. Etap I: Digitalizacja źródeł oraz stworzenie narzędzi informatycznych i udostępnienie materiałów testowych korpusu	106
Korpusy dwuwarstwowe tekstów skanowanych	108
Laboratorium Technologii Informacyjnych (LaTel)	111
Listy pożegnalne samobójców – lingwistyczne metody ustalania autentyczności tekstu	113
LUNA - spoken Language UNderstanding in multilinguAl communication systems	116
Magik	118

MedialabUMCS	120
Metronom	122
Narodowy Korpus Języka Polskiego	124
Narzędzia dygitalizacji tekstów na potrzeby badań filologicznych	126
Naukowe edycje cyfrowe	129
NEKST	131
Okres Wędrówek Ludów między Odrą a Wisłą	134
Opracowanie narzędzi do ekstrakcji informacji z tekstów w języku polskim	136
Otwarte Repozytorium Nauk Historycznych LECTORIUM	138
Otwórz Książkę	141
PARSEME	143
Pogromy. Przemoc kolektywna wobec Żydów na Ziemiach Polskich w XIX i XX wieku i jej wpływ na relacje polsko - żydowskiej. Historia, pamięć, tożsamość	145
Polonijna Biblioteka Cyfrowa	147
Polscy pisarze i badacze literatury przełomu XX i XXI wieku. Słownik biobibliograficzny	149
Polska Bibliografia Literacka - laboratorium wiedzy o współczesnej kulturze polskiej	151
Polskie czasopisma slawistyczne online. Wdrożenie innowacyjnego systemu pracy i oprogramowania Open Journal Systems	154
Prezbiterzy na późnoantycznym Zachodzie	156
RCIN	158
Rzutowanie Słowosieci na Princeton WordNet	161
semantyczno-składniowy słownik zelźności subkategoryzacyjnych leksemów polskich	163
Sensualność w kulturze polskiej. Przedstawienia zmysłów człowieka w języku, piśmiennictwie i sztuce od średniowiecza do współczesności	166
Słownik polszczyzny XVI wieku i Internetowy słownik polszczyzny XVI wieku - tomy 36-43	168
Słowosieć	171
System do zintegrowanej, głębokiej, składniowo-semantycznej analizy tekstów w języku polskim	173
System przeszukiwania dużych bogato znakowanych korpusów językowych	175
System wydobywania wielowyrzowych lematów języka polskiego	177
The Central European Journal of Social Sciences and Humanities CEJSH	179
TRENDMINER	181
Trudne marki – definicja i propozycja zastosowania metod językoznawczych do badania zjawiska	183
Usługa Powszechnej Archiwizacji	185
Wirtualne Laboratorium Transkrypcji	187
Wizualizacja sztuki	189
Wyszukiwanie w dokumentacyjnych systemach informacyjnych w świetle potrzeb środowiska akademickiego	191

Wyszukiwanie w dokumentacyjnych systemach informacyjnych w świetle potrzeb środowiska akademickiego	193
Wyszukiwarka dokumentów sumeryjskich z III dynastii z UR	195
Zestawienie zborów protestanckich w Rzeczypospolitej XVI-XVIII stuleciu	197
Zintegrowana Baza Bibliograficzna e-ZBB IBL PAN	199
Znaleziska monet rzymskich z ziem Polski i historycznie z Polską związanych	201

WSTĘP

Oddajemy w Państwa ręce opracowanie wyników ankiety, której celem było zebranie danych o humanistycznych projektach cyfrowych realizowanych przez polskie instytucje pragnące uczestniczyć w pracach powstającego konsorcjum DARIAH-PL. Przygotowaniem i opracowaniem ankiety zajęły się Poznańskie Centrum Superkomputerowo Sieciowe przy IChB PAN oraz Centrum Humanistyki Cyfrowej IBL PAN. Ankiety w dniach 9-20 VI można było wypełniać na stronie internetowej ankieta.dariah.pl. Wszystkie odpowiedzi zamieściliśmy w pisowni oryginalnej z minimalnymi ingerencjami korektorskimi.

Zebrane tu projekty mają pomóc przyszłym członkom konsorcjum w projektowaniu badań i szukaniu partnerów. Jednocześnie ankieta stanowi zapis aktualnego stanu badań w polskiej humanistyce cyfrowej, co czyni ją cennym źródłem dla wszystkich zainteresowanych tą problematyką.

Trudno tu silić się na analizy i podsumowania, ponieważ z pewnością jest to niepełny jeszcze obraz pola podlegającego dynamicznym przeobrażeniom. Czytelnicy tego tomu będą mogli samodzielnie wyrobić sobie zdanie o kierunkach badań w interesujących ich obszarach. Warto jednak wyciągnąć z tej ankiety dwa wnioski.

Po pierwsze, uderza wielość i różnorodność projektów. W ankiecie zgłoszono prawie 80 projektów obejmujących zadania badawcze, budowę infrastruktury, bazy danych, digitalizację i edukację. Dzięki tej ankiecie badacze cyfrowi mogą się wreszcie poznać i policzyć. Ale to nie jedyna zaleta – różnorodność projektów, którą obserwujemy na gruncie polskim doskonale koresponduje z rozproszeniem charakterystycznym dla projektów z zakresu humanistyki cyfrowej na całym świecie. Podkreśla to tylko fakt, iż cyfrowość nie jest wcale nowym paradygmatem badań, tylko narzędziem do realizacji przeróżnych celów badawczych często odmiennych dyscyplin.

Po drugie, tak duży odzew środowiska pokazuje iż tego typu baza wiedzy o projektach z zakresu humanistyki cyfrowej jest potrzebna i należy ją rozwijać. Podstawowy wniosek z ankiety nasuwa się sam – należy ją rozszerzać i cyklicznie powtarzać. Warto pewnie pomyśleć o przekształceniu jej w uaktualnianą na bieżąco bazę danych. Być może jest to jedno z pierwszych zadań stojących przed powstającym właśnie konsorcjum DARIAH-PL.

dr Maciej Maryl

Centrum Humanistyki Cyfrowej IBL PAN

Dane źródłowe zebrane w ramach ankiety:

- Format OpenDocument (ODS):

http://lib.psync.pl/Content/655/Humanistyczne_projekty_cyfrowe_w_Polsce.ods

- Forma MS Excel (XLSX):

http://lib.psync.pl/Content/655/Humanistyczne_projekty_cyfrowe_w_Polsce.xlsx

PODSUMOWANIE OTRZYMANYCH ANKIET

Nazwa projektu	Typ projektu	Status	Typ efektu
ALEPH POLSKA (1)			
Analiza doświadczeń wybranych instytucji w dziedzinie długookresowej ochrony zasobów cyfrowych z wykorzystaniem systemu Rosetta firmy Ex Libris	badawczy	zakończony	analiza systemu
BIBLIOTEKA NARODOWA (2)			
Academica	infrastrukturalny, bazodanowy, digitalizacyjny	realizowany	oprogramowanie, baza danych, repozytorium
Cyfrowa Biblioteka Narodowa Polona	infrastrukturalny, bazodanowy, digitalizacyjny	realizowany	oprogramowanie, baza danych, repozytorium
INSTYTUT BADAŃ LITERACKICH PAN (12)			
Adam Mickiewicz 150 rocznica śmierci	edukacyjny	zakończony	multimedialne archiwum materiałów, skupionych wokół śmierci Mickiewicza i jej legendy i
Archiwum kobiet: Piszące	badawczy	realizowany	baza danych
Biuletyn Polonistyczny	bazodanowy	realizowany	baza danych
Blog jako nowa forma piśmiennictwa multimedialnego	badawczy	realizowany	publikacja, metodologia
Interaktywna mapa literacka	badawczy, infrastrukturalny	realizowany	narzędzie badawcze
Korpus polszczyzny XVI wieku. Etap I: Digitalizacja źródeł oraz stworzenie narzędzi informatycznych i udostępnienie materiałów testowych korpusu	digitalizacyjny	realizowany	baza danych, repozytorium
Naukowe edycje cyfrowe	badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny	planowany	narzędzie badawcze, baza danych
Polscy pisarze i badacze literatury przełomu XX i XXI wieku. Słownik biobibliograficzny	badawczy	realizowany	baza danych, publikacja
Polska Bibliografia Literacka - laboratorium wiedzy o współczesnej kulturze polskiej	bazodanowy, digitalizacyjny	planowany	narzędzie badawcze, baza danych
Sensualność w kulturze polskiej. Przedstawienia zmysłów człowieka w języku, piśmiennictwie i sztuce od średniowiecza do współczesności.	badawczy	zakończony	narzędzie badawcze, publikacja

Nazwa projektu	Typ projektu	Status	Typ efektu
Słownik polszczyzny XVI wieku i Internetowy słownik polszczyzny XVI wieku - tomy 36-43	bazodanowy, digitalizacyjny	realizowany	narzędzie badawcze, baza danych
Zintegrowana Baza Bibliograficzna e-ZBB IBL PAN	badawczy, bazodanowy, digitalizacyjny	realizowany	baza danych, standardy opisu i przetwarzania danych bibliograficznych o charakterze zgodnym ze specyfiką projektu
INSTYTUT JĘZYKA POLSKIEGO PAN (1)			
Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do roku 1772)	badawczy	realizowany	narzędzie badawcze, oprogramowanie, korpus tekstów średniopolskich
INSTYTUT PODSTAW INFORMATYKI PAN (13)			
Adaptowalny zespół specjalizowanych narzędzi do przetwarzania morfosyntaktycznego polskich tekstów z dziedziny określonej przez użytkownika	badawczy, infrastrukturalny	realizowany	narzędzie badawcze, oprogramowanie
ATLAS	infrastrukturalny	zakończony	narzędzie badawcze, oprogramowanie
CESAR	infrastrukturalny	zakończony	narzędzie badawcze, oprogramowanie, repozytorium
CORE	badawczy	realizowany	narzędzie badawcze, oprogramowanie, publikacja, ręcznie anotowany korpus tekstów
LUNA - spoken Language UNderstanding in multilinguAl communication systems	badawczy	zakończony	narzędzie badawcze, oprogramowanie, publikacja
Narodowy Korpus Języka Polskiego	badawczy, digitalizacyjny	zakończony	narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja
NEKST	badawczy, infrastrukturalny	realizowany	narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja
Opracowanie narzędzi do ekstrakcji informacji z tekstów w języku polskim	badawczy	zakończony	narzędzie badawcze, publikacja

Nazwa projektu	Typ projektu	Status	Typ efektu
PARSEME	badawczy, infrastrukturalny	realizowany	publikacja
semantyczno-składniowy słownik zależności subkategoryzacyjnych leksemów polskich	infrastrukturalny	realizowany	narzędzie badawcze, oprogramowanie, baza danych
System do zintegrowanej, głębokiej, składniowo-semantycznej analizy tekstów w języku polskim	badawczy, infrastrukturalny	realizowany	narzędzie badawcze, oprogramowanie
System przeszukiwania dużych bogato znakowanych korpusów językowych	badawczy, infrastrukturalny	realizowany	narzędzie badawcze, oprogramowanie
TRENDMINER	badawczy	realizowany	narzędzie badawcze, oprogramowanie
INSTYTUT SŁAWISTYKI PAN (3)			
Bibliograficzna baza danych światowego językoznawstwa sławistycznego iSybisław	badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny	realizowany	baza danych
Internetowe repozytorium tekstów rozpraw doktorskich i preprintów innych prac z zakresu światowej sławistyki i dziedzin pokrewnych iRetesław	infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny	realizowany	repozytorium
Polskie czasopisma sławistyczne online. Wdrożenie innowacyjnego systemu pracy i oprogramowania Open Journal Systems	infrastrukturalny, digitalizacyjny	realizowany	narzędzie badawcze, repozytorium
INSTYTUTU CHEMII BIOORGANICZNEJ PAN - POZNAŃSKIE CENTRUM SUPERKOMPUTEROWO- SIECIOWE (3)			
Federacja Bibliotek Cyfrowych	infrastrukturalny, bazodanowy	realizowany	oprogramowanie, baza danych
Usługa Powszechnej Archiwizacji	infrastrukturalny	zakończony	repozytorium
Wirtualne Laboratorium Transkrypcji	badawczy	realizowany	narzędzie badawcze, oprogramowanie
MUZEUM I INSTYTUT ZOOLOGII PAN (1)			
RCIN	digitalizacyjny	realizowany	repozytorium
POLITECHNIKA WROCŁAWSKA (8)			
Automatyczna ekstrakcja informacji	badawczy, infrastrukturalny	realizowany	narzędzie badawcze, oprogramowanie
CorpoGrabber	badawczy, infrastrukturalny, bazodanowy	realizowany	narzędzie badawcze, oprogramowanie
GeTClasS - system do anotacji semantycznej tekstu	badawczy, infrastrukturalny	realizowany	narzędzie badawcze, oprogramowanie

Nazwa projektu	Typ projektu	Status	Typ efektu
Inforex	infrastrukturalny, bazodanowy	realizowany	narzędzie badawcze, oprogramowanie, baza danych
Korpus Politechniki Wrocławskiej (KPWr)	badawczy, infrastrukturalny, bazodanowy	realizowany	narzędzie badawcze, baza danych, publikacja
Rzutowanie Słowosieci na Princeton WordNet	badawczy, infrastrukturalny, bazodanowy	realizowany	narzędzie badawcze, oprogramowanie, baza danych, publikacja
Słowosieć	badawczy, infrastrukturalny, bazodanowy	realizowany	narzędzie badawcze, oprogramowanie, baza danych, publikacja
System wydobywania wielowyrzowych lematów języka polskiego	badawczy, infrastrukturalny, bazodanowy	realizowany	narzędzie badawcze, oprogramowanie, baza danych, publikacja
POLSKI INSTYTUT ANTROPOLOGII (1)			
Cyfrowa Etnografia - Bibliotek Cyfrowa Piłskiego Instytutu Antropologii (BC PIA)	infrastrukturalny, bazodanowy, digitalizacyjny	realizowany	narzędzie badawcze, baza danych, repozytorium, publikacja
POLSKIE TOWARZYSTWO BIBLIOLOGICZNE (1)			
Aktualizacja kształcenia akademickiego bibliotekarzy pracujących w małych bibliotekach gminnych.	badawczy, edukacyjny, wdrożeniowy	zakończony	narzędzie badawcze, publikacja, wdrożenia zaprojektowanych kursów
UNIwersytet Marii Curie-Skłodowskiej (2)			
MedialabUMCS	badawczy, edukacyjny	realizowany	narzędzie badawcze, oprogramowanie, publikacja
Wizualizacja sztuki	badawczy, digitalizacyjny, edukacyjny	realizowany	repozytorium, publikacja, kurs
UNIwersytet Pedagogiczny w Krakowie, UNIwersytet Jagielloński (1)			
Computational Stylistics Group	badawczy	realizowany	narzędzie badawcze, repozytorium, publikacja, kurs

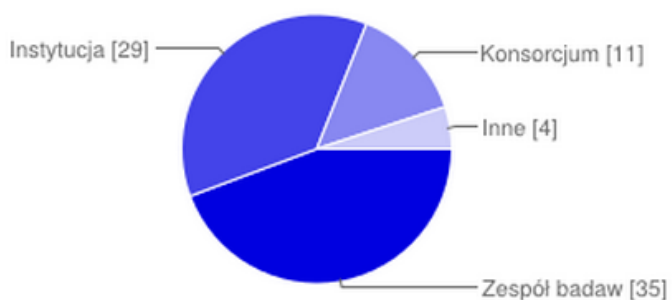
Nazwa projektu	Typ projektu	Status	Typ efektu
UNIwersytet Śląski (2)			
E-edukacja bez barier cyfrowych i społecznych	edukacyjny	planowany	narzędzie badawcze, baza danych, repozytorium, kurs
Konstrukcja elektronicznych słowników (baz danych leksykalnych) i dezambiguizacja znaczeniowa dla celów tłumaczenia automatycznego.	badawczy	realizowany	baza danych
UNIwersytet Warszawski (27)			
Bibliologiczna Biblioteka Cyfrowa	infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny	realizowany	baza danych, repozytorium
Biblioteka Cyfrowa KLF UW	infrastrukturalny, bazodanowy	realizowany	baza danych, repozytorium
Digitalizacja i udostępnienie w sieci kolekcji starodruków Biblioteki IH UW	badawczy, digitalizacyjny	zakończony	oprogramowanie, baza danych, repozytorium, publikacja
Digitalizacja wydawnictw Stowarzyszenia Bibliotekarzy Polskich	digitalizacyjny	zakończony	repozytorium
Dygitalizacja słownika Lindego	badawczy, infrastrukturalny, bazodanowy, digitalizacyjny	realizowany	baza danych, repozytorium, publikacja
Dziedzictwo narodowe i kulturowe, polskie ślady za granicą.	badawczy, digitalizacyjny	zakończony	repozytorium
Dziedzictwo narodowe, badania i upowszechnianie wiedzy o tradycji i historii	badawczy, digitalizacyjny	realizowany	repozytorium
IMPACT-UWAR	badawczy, infrastrukturalny, bazodanowy, digitalizacyjny	zakończony	narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja
Komputerowa Analiza Nieprojektowanego Odbioru	badawczy, infrastrukturalny	planowany	narzędzie badawcze, oprogramowanie, baza danych, publikacja
Komputerowe Wspomaganie Analizy Wywiadów Etnograficznych	bazodanowy, digitalizacyjny, edukacyjny	zakończony	oprogramowanie, repozytorium

Nazwa projektu	Typ projektu	Status	Typ efektu
Korpusy dwuwarstwowe tekstów skanowanych	badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, korpusowy	realizowany	narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja, przeglądarki korpusów dwuwarstwowych
Laboratorium Technologii Informacyjnych (LaTel)	badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny	realizowany	narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja, kurs
Magik	edukacyjny, narzędziowy	zakończony	oprogramowanie, 2 prace magisterskie
Metronom	badawczy	zakończony	oprogramowanie
Narzędzia dygitalizacji tekstów na potrzeby badań filologicznych	badawczy, infrastrukturalny, bazodanowy, digitalizacyjny	zakończony	narzędzie badawcze, oprogramowanie, publikacja
Okres Wędrówek Ludów między Odrą a Wisłą	badawczy	realizowany	publikacja, nowa perspektywa badawcza przeszłości ziem polskich, liczne opracowania i wystawy
Otwarte Repozytorium Nauk Historycznych LECTORIUM	bazodanowy, digitalizacyjny	realizowany	repozytorium
Otwórz Książkę	digitalizacyjny	realizowany	kolekcja książek naukowych
Pogromy. Przemoc kolektywna wobec Żydów na Ziemiach Polskich w XIX i XX wieku i jej wpływ na relacje polsko - żydowskiej. Historia, pamięć, tożsamość	badawczy	realizowany	narzędzie badawcze, baza danych, publikacja
Polonijna Biblioteka Cyfrowa	infrastrukturalny, bazodanowy, digitalizacyjny	realizowany	baza danych, repozytorium, publikacja
Prezbiterzy na późnoantycznym Zachodzie	badawczy	realizowany	narzędzie badawcze
The Central European Journal of Social Sciences and Humanities CEJSH	bazodanowy	realizowany	baza danych, kolekcja czasopism naukowych w otwartym dostępie
Wyszukiwanie w dokumentacyjnych systemach informacyjnych w świetle potrzeb środowiska akademickiego	badawczy	zakończony	publikacja
Wyszukiwarka dokumentów sumeryjskich z III dynastii z UR	badawczy	zakończony	oprogramowanie, baza danych
Zestawienie zbiorów protestanckich w Rzeczypospolitej XVI-XVIII stuleciu	badawczy, bazodanowy	zakończony	narzędzie badawcze, baza danych

Nazwa projektu	Typ projektu	Status	Typ efektu
Znaleziska monet rzymskich z ziem Polski i historycznie z Polską związanych	badawczy, bazodanowy	realizowany	narzędzie badawcze, baza danych, publikacja
UNIwersytet Wrocławski (2)			
Listy pożegnalne samobójców – lingwistyczne metody ustalania autentyczności tekstu	badawczy	zakończony	narzędzie badawcze, oprogramowanie, baza danych, publikacja
Trudne marki – definicja i propozycja zastosowania metod językoznawczych do badania zjawiska	badawczy	realizowany	publikacja

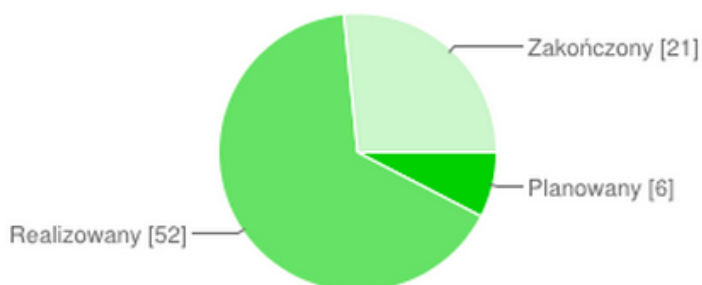
STATYSTYKI WYBRANYCH ELEMENTÓW ANKIETY

RODZAJ PODMIOTU REALIZUJĄCEGO PROJEKT



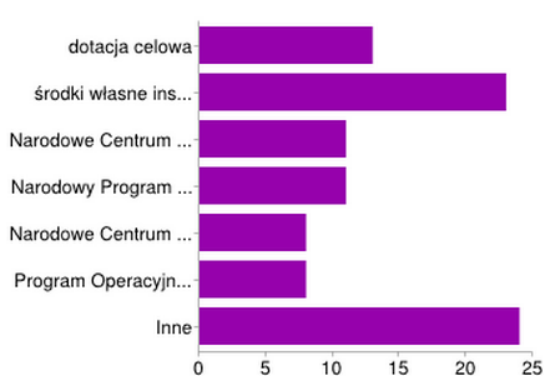
Zespół badawczy	35	44%
Instytucja	29	37%
Konsorcjum	11	14%
Inne	4	5%

STATUS PROJEKTU



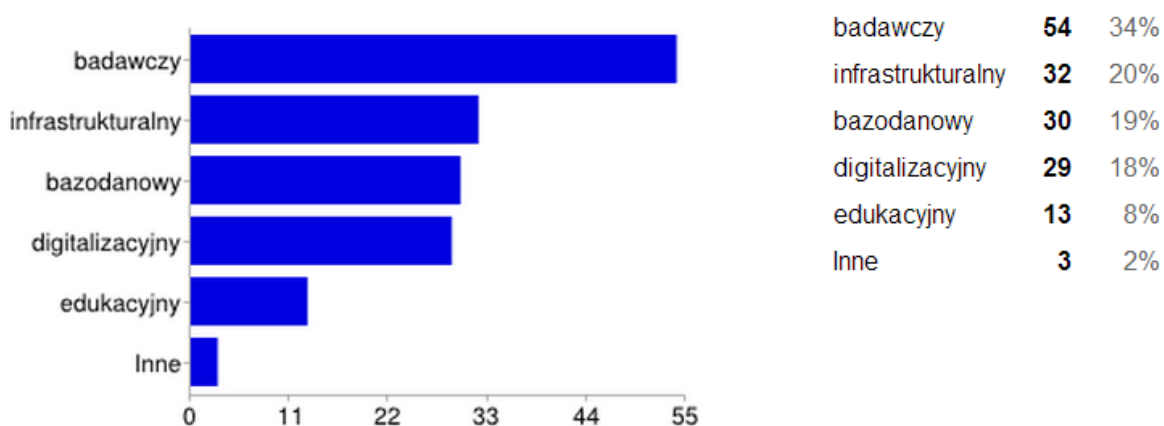
Planowany	6	8%
Realizowany	52	66%
Zakończony	21	27%

ŹRÓDŁO FINANSOWANIA

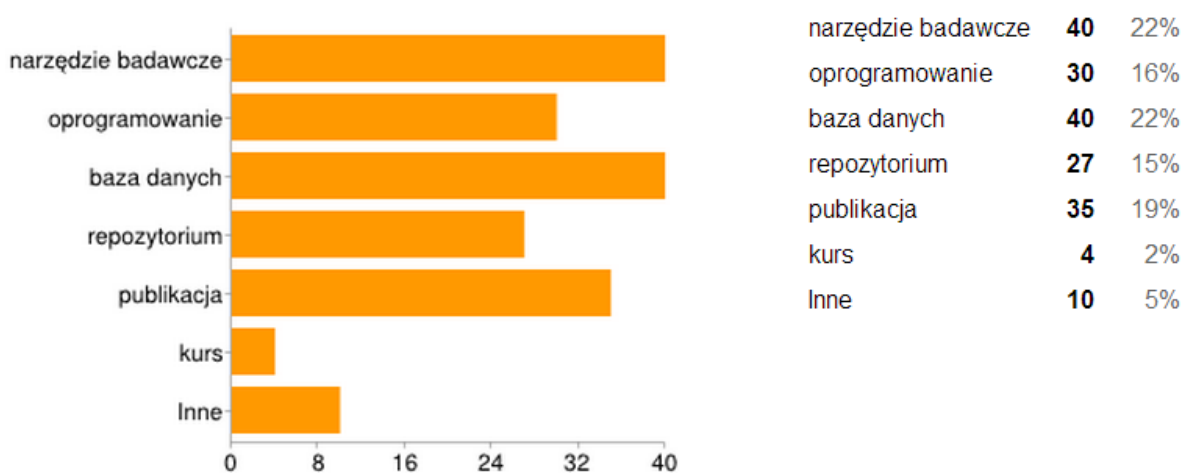


dotacja celowa	13	13%
środki własne instytucji	23	23%
Narodowe Centrum Nauki	11	11%
Narodowy Program Rozwoju Humanistyki	11	11%
Narodowe Centrum Badań i Rozwoju	8	8%
Program Operacyjny "Innowacyjna Gospodarka"	8	8%
Inne	24	24%

TYP PROJEKTU



EFEKT PROJEKTU



OPISY PROJEKTÓW

ACADEMICA

Osoba wypełniająca ankietę: Osoba kontaktowa

academica.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Biblioteka Narodowa

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Naukowa i Akademicka Sieć Komputerowa

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Jakub Czułba, j.czulba@bn.org.pl

Osoba kontaktowa: Katarzyna Ślaska, k.slaska@bn.org.pl

Pozostałe osoby realizujące projekt:

Katarzyna Ślaska, BN; Dariusz Paradowski, BN; Paweł Mazur, BN; Hanna Wróblewska, BN; Agnieszka Leszyńska, BN; Anna Wasilewska, BN; Dorota Fortuna, BN; Robert Michalak, NASK; Michał Matusiak, NASK; Piotr Martyniak, NASK

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 1/1/2010 - 10/31/2014

Źródło finansowania: Program Operacyjny "Innowacyjna Gospodarka" POIG 02.03.02-00-021/09

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy, digitalizacyjny

Efekt projektu: oprogramowanie, baza danych, repozytorium

Słowa kluczowe: publikacja naukowa; reprodukcja cyfrowa; biblioteka; digitalizacja; udostępnianie

Skrócony opis projektu:

Projekt „Cyfrowa Wypożyczalnia Publikacji Naukowych Academica” polega na utworzeniu cyfrowej bazy danych publikacji naukowych ze zbiorów Biblioteki Naukowej i ich szerokim udostępnieniu środowisku akademickiemu. Udostępnianie publikacji naukowych odbywa się zgodnie z zapisami Ustawy o prawie autorskim, a dzieła podlegają ograniczeniom kopiowania lub drukowania ich fragmentów. Wyselekcjonowane publikacje są digitalizowane za pomocą skanerów, zakupionych w ramach projektu. W wyniku cyfryzacji powstaje obecnie baza danych publikacji naukowych zawierająca 250 tys. reprodukcji cyfrowych publikacji naukowych. Cyfrowe

wersje publikacji są umieszczone w repozytorium cyfrowym w celu ich długotrwałego przechowywania. Zebrane dane będą udostępniane w formie elektronicznej za pośrednictwem Internetu oraz terminali w systemie Academica w październiku 2014 r.

Skrócony opis części informatycznej:

W ramach projektu zakupiono wyposażenie Repozytorium Cyfrowego - archiwum wieczystego, które zapewnia ciągłość przechowywania informacji. Zastosowane przy jego projektowaniu technologie odznaczają się elastycznością, otwartością, a sama baza jest skalowalna i daje możliwość włączania nowych zasobów bez konieczności wyłączenia dostępności zasobów wcześniej wytworzonych oraz migracji danych w ramach repozytorium pomiędzy kolejnymi generacjami sprzętu lub oprogramowania. Na infrastrukturę informatyczną składa się: farma serwerów wieloprocesorowych, zwirtualizowane środowisko macierzowe o surowej pojemności 430 TB, zrobotyzowana biblioteka taśmowa o pojemności startowej ponad 600TB i automatyczna biblioteka BLURAY. Ponadto w ramach projektu powstało oprogramowanie Academica służące obsłudze wypożyczalni publikacji naukowych.

Opis efektów projektu:

Efektom projektu jest udostępnienie liczącego ponad 250 tys. jednostek zbioru współczesnych publikacji naukowych (książek i czasopism) oraz wprowadzenie nowej jakości w systemie wypożyczeń międzybibliotecznych poprzez zastąpienie ich tradycyjnej formy polegającej na przesyłaniu pocztą egzemplarzy papierowych wypożyczalnią publikacji w postaci cyfrowej. Użytkownicy będą mogli korzystać z elektronicznych publikacji naukowych za pomocą dedykowanych terminali zlokalizowanych w bibliotekach naukowych i publicznych w całej Polsce. Dostęp do zasobów Akademiki będzie bezpłatny. Ponadto powstała w Bibliotece Narodowej infrastruktura do masowej digitalizacji zbiorów bibliotecznych oraz Repozytorium Cyfrowe do ich bezpiecznej archiwizacji. Udostępnienie zasobu dla użytkowników jest przewidziane na październik 2014 r.

Powiązania z innymi polskimi i zagranicznymi projektami:

Informacja o zasobach Cyfrowej Wypożyczalni Publikacji Naukowych Academica chronionych prawem autorskim dostępna będzie w portalu Cyfrowej Biblioteki Narodowej Polona.

Projekt jest unikatowy w skali europejskiej.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.bn.org.pl/programy-i-uslugi/cyfrowa-biblioteka-publicacji-naukowych-academica>

<http://academica.edu.pl/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

W ramach kontynuacji projektu przewidziane jest m.in.:

1. rozbudowa zawartości Akademiki
2. udostępnienie reprodukcji cyfrowych wszystkich czasopism naukowych z listy Ministerstwa Nauki i Szkolnictwa Wyższego
3. rozbudowa funkcjonalności serwisu

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Reprodukcje cyfrowe udostępnione w systemie Academika mogą być wykorzystywane w pracy naukowej oraz dydaktycznej na poziomie akademickim. Stworzona w ramach projektu pełnotekstowa baza wiedzy może być wykorzystywana w badaniach z zakresu lingwistyki komputerowej, przetwarzania tekstu i obrazu, badań statystycznych oraz porównawczych. Baza danych bibliograficznych może być również przedmiotem zainteresowania badaczy. System Academica może być bazą do utworzenia ogólnopolskiego systemu antyplagiatowego.

Profil potencjalnego partnera:

Badacze zajmujący się humanistyką cyfrową mogą wskazywać tytuły bądź typy publikacji, które zostaną poddane digitalizacji i udostępnione w systemie Academica, jak również opracowywać nowe funkcjonalności niezbędne w pracy badawczej.

ADAM MICKIEWICZ 150 ROCZNICA ŚMIERCI

Osoba wypełniająca ankietę: Kierownik projektu

<http://www.am.ibl.waw.pl/flash.html>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Badań Literackich PAN, Pracownia Romantyczna

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Marta Zielińska, ciocia5211@wp.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr hab. Marta Zielińska (IBL PAN), dr Monika Rudaś-Grodzka (IBL PAN), mgr Katarzyna Czeczot (IBL PAN), Rafał Benedek (grafika, umowa o dzieło), Marcin Śmiałek (opracowanie informatyczne, umowa o dzieło)

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 3/10/2005 - 6/30/2006

Źródło finansowania: środki własne instytucji nie dotyczy

OPIS PROJEKTU

Typ projektu: edukacyjny

Efekt projektu: multimedialne archiwum materiałów, skupionych wokół śmierci Mickiewicza i jej legendy i

Słowa kluczowe: śmierć Mickiewicza; pogrzeby Mickiewicza; otrucie Mickiewicza; pomniki Mickiewicza; 150-letnia rocznica śmierci Mickiewicza

Skrócony opis projektu:

Prezentacja materiałów tekstowych i ikonograficznych związanych ze śmiercią Mickiewicza (z okazji 150-tej rocznicy), w działach problemowych: Ostatnie dni, Tajemnica śmierci, Zwłoki wieszczą, Pogrzeby i skandale, Pomniki, Upiory Mickiewicza. Stronka zawiera pełne dossier tego wydarzenia (od momentu śmierci po pochówki na Wawelu) zawarte w źródłach, artykułach prasowych, opracowaniach naukowych, zdjęciach, pocztówkach, ilustracjach, pomnikach. Ten archiwalny zestaw, zawierający całkowite wersje trudno dostępnych dziś tekstów publicystycznych i obrazów umożliwia czytelnikowi zapoznanie się ze źródłami, uporządkowanymi w sposób problemowy: okoliczności śmierci, dyskusje o jej przyczynach, pogrzeby (Konstantynopol, Paryż, Kraków), wreszcie zagospodarowanie legendy w sztuce wysokiej i niskiej.

Skrócony opis części informatycznej:

Projekt został zrealizowany w technologii flash.html, pozwalającej na atrakcyjną prezentację treści w wartościowej formie graficznej ruchomych tekstów, pokazów slajdów, a nawet animowanej, interaktywnej mapki. Nawigacja na stronie prowadzi włąb, do kolejnych podstron, coraz dokładniej problematyzujących temat, wreszcie do konkretnych tekstów i ilustracji. Możliwe jest również przemieszczanie się poziomo poprzez menu główne dostępne z każdej podstrony.

Opis efektów projektu:

Projekt ma charakter edukacyjno-materiałowy i stanowi pomoc dla studentów i w pracy naukowej oferując cyfrową antologię trudną do realizacji tradycyjnym sposobem.

Powiązania z innymi polskimi i zagranicznymi projektami:

nie dotyczy

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

nie dotyczy

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Sam projekt jest zamknięty ze względu na jego merytoryczny/okolicznościowy charakter i technologię, w jakiej został zrobiony, może natomiast wchodzić w większe całości internetowej historii literatury, jeśli taka powstanie.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

-

Profil potencjalnego partnera:

-

ADAPTOWALNY ZESPÓŁ SPECJALIZOWANYCH NARZĘDZI DO PRZETWARZANIA
MORFOSYNTAKTYCZNEGO POLSKICH TEKSTÓW Z DZIEDZINY OKREŚLONEJ PRZEZ UŻYTKOWNIKA

Osoba wypełniająca ankietę: Agnieszka Mykowiecka, agn@ipipan.waw.pl

<http://clarin-pl.eu/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Instytut Podstaw Informatyki PAN

Pozostałe instytucje zaangażowane w projekt: Cały projekt CLARIN-PL jest realizowany przez konsorcjum koordynowane przez Politechnikę Wrocławską (kierownik: Maciej Piasecki) – zob. ankietę dot. całego CLARIN-PL. Tu opisane jest konkretne zadanie realizowane w IPI PAN. Poniżej „kierownik” itd. rozumiane jest jako „kierownik danego zadania” itd.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Małgorzata Marciniak, mm@ipipan.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Małgorzata Marciniak, IPI PAN;

Agnieszka Mykowiecka, IPI PAN;

Piotr Rychlik, IPI PAN;

Michał Lenart, IPI PAN;

Dominika Rogozińska, IPI PAN;

Michał Woźniak;

Łukasz Kobyliński, IPI PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 2/1/2013 - 12/31/2015

Źródło finansowania: dotacja na dużą infrastrukturę badawczą

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: analiza morfologiczna, rozpoznawanie nazw własnych, słowniki dziedzinowe

Skrócony opis projektu:

Celem zadania A19 projektu CLARIN-PL jest opracowanie zestawu narzędzi służących do analizy tekstów dziedzinowych. Budowane narzędzia mają umożliwić użytkownikowi automatyczne lub częściowo automatyczne anotowanie lub rozpoznawanie w tekstach różnego typu informacji zarówno o charakterze lingwistycznym jak i dziedzinowym. W ramach prowadzonych prac zostaną opracowane programy do segmentacji i analizy morfologicznej tekstów dziedzinowych, program rozpoznający frazy będące nazwami utworów czy dzieł sztuki, program wspomagający narzędzia korekty ortograficznej oraz program wspomagający automatyczne tworzenie słownika terminów wielowyrazowych.

Skrócony opis części informatycznej:

j.w.

Opis efektów projektu:

j.w.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Opracowanie narzędzi do rozpoznawania w tekstach różnych typów pojęć oraz o rozpoznawanie relacji między tymi pojęciami.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

AKTUALIZACJA KSZTAŁCENIA AKADEMICKIEGO BIBLIOTEKARZY PRACUJĄCYCH W MAŁYCH BIBLIOTEKACH GMINNYCH.

Osoba wypełniająca ankietę: Kierownik projektu

<http://www.projekt.ptbow.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Instytut Informacji Naukowej i Studiów Bibliologicznych UW, Instytut Informacji Naukowej i Bibliotekoznawstwa UP, Instytut Informacji Naukowej i Bibliotekoznawstwa UJ, Instytut Informacji Naukowej i Bibliotekoznawstwa UW, Instytut Informacji Naukowej i Bibliologii UMK, Instytut Informacji Naukowej i Dziennikarstwa, UJK, Instytut Bibliotekoznawstwa i Informacji Naukowej UŚ, Instytut Bibliotekoznawstwa i Informacji Naukowej UMCS, Katedra Bibliotekoznawstwa i Informacji Naukowej UŁ

Instytucja przy której projekt jest afiliowany: Polskie Towarzystwo Bibliologiczne Oddział Warszawski

Pozostałe instytucje zaangażowane w projekt: Fundacja Rozwoju Społeczeństwa Informacyjnego

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Mikołaj Ochmański (m.ochmanski@uw.edu.pl), Michał Zajac (michal.zajac@uw.edu.pl)

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 10/1/2009 - 12/31/2011

Źródło finansowania: dotacja celowa

OPIS PROJEKTU

Typ projektu: badawczy, edukacyjny, wdrożeniowy

Efekt projektu: narzędzie badawcze, publikacja, wdrożenia zaprojektowanych kursów

Słowa kluczowe: bibliotekoznawstwo; dydaktyka; biblioteki publiczne; biblioteki gminne

Skrócony opis projektu:

Istotą prezentowanego Projektu było zwrócenie uwagi na potrzeby bibliotekarzy z małych bibliotek publicznych i wprowadzenie w dydaktykę na poziomie akademickim większej liczby elementów, pomagających w rozwiązywaniu bieżących problemów gminnych bibliotek i wspomagających ich rozwój. Projekt zakłada korektę kształcenia uniwersyteckiego pracowników gminnych bibliotek publicznych dokonaną bez umniejszania roli zagadnień ogólnych i teoretycznych w treściach programowych.

Celem zmodyfikowanego kształcenia ma być przygotowanie absolwentów do kreowania nowej roli i nowego obrazu lokalnych bibliotek publicznych. Współczesne kształcenie powinno kształcić umiejętności, które dadzą bibliotekarzom możliwość stworzenia z bibliotek centrów życia kulturalnego i społecznego lokalnych społeczności, ośrodków integrujących, służących wzmocnieniu więzi i poczucia tożsamości, otwartych placówek medialnych bogatych w zasoby informacji różnego typu.

Skrócony opis części informatycznej:

Opracowanie programów zajęć, praktyk zawodowych i programów studiów II stopnia z wykorzystaniem ICT. Szczegółowe programy, które powstały w ramach Projektu dostępne w publikacjach na stronie www:

<http://www.projekt.ptbow.pl/badania/publikacja/>

<http://www.projekt.ptbow.pl/badania/publikacja-zamykajaca/>

Opis efektów projektu:

Publikacje dostępne w wersji tradycyjnej i elektronicznej dające analizy dydaktyki bibliotekoznawczej na uniwersytetach w Polsce oraz podsumowanie Projektu i propozycje zajęć dydaktycznych, praktyk i projekty studiów II stopnia.

Realizacje zaprojektowanych zajęć oraz praktyk zawodowych.

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekt był sfinansowany z grantu Fundacji Rozwoju Społeczeństwa Informacyjnego w ramach Programu Rozwoju Bibliotek (<http://www.biblioteki.org/pl>).

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.projekt.ptbow.pl/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

ANALIZA DOŚWIADCZEŃ WYBRANYCH INSTYTUCJI W DZIEDZINIE DŁUGOOKRESOWEJ OCHRONY ZASOBÓW CYFROWYCH Z WYKORZYSTANIEM SYSTEMU ROSETTA FIRMY EX LIBRIS

Osoba wypełniająca ankietę: Osoba kontaktowa

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Projekt realizowany przez dr. Grzegorz Gmiterka (IINiSB UW) we współpracy z firmą ALPEPH Polska

Institucja przy której projekt jest afiliowany: Aleph Polska

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Dziubecki, m.dziubecki@aleph.pl

Osoba kontaktowa: Grzegorz Gmiterek, ggmiterek@gmail.com

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 6/1/2012 - 10/31/2012

Źródło finansowania: Program Operacyjny "Kapitał Ludzki"

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: analiza systemu

Słowa kluczowe: długoterminowa archiwizacja zasobów cyfrowych; digitalizacja, Rosetta, Ex Libris; kolekcje cyfrowe

Skrócony opis projektu:

Przygotowanie analizy strategii i praktyki wykorzystania nieznanego szerzej w polskim środowisku bibliotecznym, oprogramowania Rosetta firmy Ex Libris, służącego długookresowej ochronie zasobów cyfrowych przez instytucje kultury i nauki (zwłaszcza biblioteki). Analiza przeprowadzona na podstawie badań możliwości wykorzystania oprogramowania Rosetta, jak i raportów czy branżowych publikacji przygotowanych przez pracowników zagranicznych instytucji, które zaimplementowały to oprogramowanie (np. Biblioteka Narodowa w Nowej Zelandii). Celem jest zaprezentowanie sprawdzonych już rozwiązań technologicznych

umożliwiających stworzenie systemu organizacyjnego zapewniającego bezpieczeństwo oraz autentyczność przechowywanych zasobów cyfrowych bez względu na ich typ czy format, w jakim zostały zapisane.

Skrócony opis części informatycznej:

Analiza możliwości wykorzystania systemu Rosetta

Opis efektów projektu:

Przygotowanie analizy strategii i praktyki wykorzystania nieznanego szerzej w polskim środowisku bibliotecznym oprogramowania Rosetta firmy Ex Libris, służącego długookresowej ochronie zasobów cyfrowych.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.euro-konsult.pl/szkolenia/woj-mazowieckie/21-z-duchem-czasu-transf-innow-na-mazow-woj-mazowieckie.html>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

ARCHIWUM KOBIET: PISZĄCE

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://archiwum.archiwumkobiet.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Literatura i gender Instytutu Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: -

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Dr Monika Rudaś-Grodzka, mail: monikarudasgrodzka@gmail.com

Osoba kontaktowa: Dr Katarzyna Nadana-Sokołowska, mail: katarzyna.nadana-sokolowska@ibl.waw.pl

Pozostałe osoby realizujące projekt:

Dr Monika Rudaś-Grodzka, IBL PAN;

dr Katarzyna Nadana-Sokołowska, IBL PAN;

Dr Agnieszka Mrozik, IBL PAN;

Dr Katarzyna Czeczot, IBL PAN;

prof. dr hab. Anna Nasiłowska, IBL PAN;

Kazia Szczuka, IBL PAN;

Agnieszka Wróbel, IBL PAN;

Dr Iwona Wiśniewska, IBL PAN;

Dr Katarzyna Sierakowska, UW;

dr Agnieszka Janiak-Jasińska, UW;

Dominik Purchała, UW;

Weronika Grzebalska, IBL PAN;

Dr Iwona Misiak, IBL PAN;

Monika Rynkowska, IBL PAN;

Barbara Grodzka, IBL PAN;

Ewa Serafin, IS PAN;

Współpracownicy

Dr hab. prof. IS PAN Joanna Nowak, IS PAN;

Dr hab. Katarzyna Wrzesińska, IS PAN;

Dr Monika Gabryś, UMCS;

Dr Beata Dejneka, Uniwersytet Przyrodniczo-Humanistyczny w Siedlcach;

Dr Natalia Yakubova, IBL PAN;

Dr Joanna Partyka, IBL PAN;

Dr Ania Sobieska, IBL PAN;

Dr Lucyna Marzec, UAM

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 3/1/2013 - 2/28/2018

Źródło finansowania: Narodowy Program Rozwoju Humanistyki 555/NPRH0087/2013

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: baza danych

Słowa kluczowe: dokumenty osobiste, archiwa, doświadczenie kobiet, biografie, autobiografie

Skrócony opis projektu:

Projekt „Archiwum kobiet: Piszące” zakłada stworzenie bazy danych oraz zainicjowanie digitalnego archiwum niewydanych rękopisów kobiet mieszkających na obszarze historycznej Polski od XVI wieku do współczesności. Baza danych obejmować będzie rękopisy odnajdywane w trakcie kwerend w poszczególnych archiwach na terenie współczesnej Polski i krajów europejskich, ze szczególnym uwzględnieniem obszarów powiązanych kulturowo z Polską poprzez fakt dawniejszej łączności z jej terytoriami (Litwa, Ukraina, Białoruś, itp.). Zgromadzone w bazie rękopisy zostaną zindeksowane według kryteriów zasadniczych dla możliwości przyszłego naukowego opracowania ich zawartości, takich jak: autorstwo, miejsce i czas pochodzenia, gatunek piśmienniczy, słowa. Część najciekawszych znalezisk zostanie w ramach projektu umieszczona w formie skanów w bazie, stając się zaczątkiem digitalnego archiwum kobiet.

Skrócony opis części informatycznej:

Platforma informatyczna została specjalnie zaprojektowana dla realizowanego projektu. Serwis wykorzystuje CMS /DRUPAL/ o otwartym kodzie źródłowym, bazy danych oparte są na PHP i MySQL. Podczas tworzenia serwisu i baz danych nie zostało użyte zamknięte oprogramowanie komercyjne. Obecnie serwis składa się z publicznej witryny internetowej projektu oraz baz danych dostępnych obecnie jedynie dla zalogowanych członków zespołu. Relacyjne bazy danych umożliwiają zbieranie informacji za pomocą formularzy. Wbudowane relacje pomagają w szybki sposób przedstawiać powiązania i wzajemne zależności pomiędzy wprowadzanymi

rekordami bibliograficznymi oraz rozbudowanymi polami opisowymi dotyczącymi danych osobowych. Możliwy jest eksport danych w formacie csv.

Opis efektów projektu:

Efektom projektu będzie baza danych złożona z rekordów bibliograficznych niewydanych rękopisów kobiet. Baza będzie dostępna dla wszystkich, którzy wejdą na stronę archiwum. Do rekordów dołączone będą skany dokumentów itp., dostęp do nich będzie otwarty tylko jeśli realizatorzy projektu pozyskają zgodę na ich użycie. Planowane jest przygotowanie dokumentu pozwalającego na użycie i udostępnienie zeskanowanych materiałów. W pierwszym roku trwania projektu stworzono bazę danych. Obecnie prace polegają na zbieraniu i opracowywaniu materiałów do bazy.

Powiązania z innymi polskimi i zagranicznymi projektami:

-

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

-

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Zespół chce kontynuować realizację projektu również po zakończeniu grantu. Chcemy rozszerzyć bazę danych o dokumenty z archiwów Europy Środkowo-Wschodniej. Mamy zamiar rozpocząć współpracę z Instytutem Sławistyki w Poczdamie.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

W pierwszej fazie projekt jest zarządzany i prowadzony przez zespół badawczy pod kierownictwem dr Moniki Rudaś-Grodzkiej, w dalszej perspektywie planowane jest otwarcie projektu dla wszystkich zainteresowanych instytucji, zespołów badawczych oraz pracowników naukowych. Partycypacja w projekcie będzie możliwa na zasadzie odiorcy oraz zaangażowania się w jego dalszy rozwój. Przewidziana jest możliwość współpracy pojedynczych badaczy jak i jednostek naukowych. Dla osób/jednostek przystępujących do współpracy zostaną stworzone konta w serwisie umożliwiające redakcję oraz tworzenie nowych rekordów. Forma współpracy będzie ustalana z zainteresowanymi podmiotami. Wyniki kwerend i rekordy będą dostępne publicznie i nieodpłatnie dla wszystkich zainteresowanych.

Profil potencjalnego partnera:

- Jednostki akademickie, instytucje naukowe, zespoły badawcze zajmujące się w szczególności literaturoznawstwem, historią, historią kobiet, bibliotekoznawstwem i informacją naukową.
- Archiwa państwowe, instytucjonalne, prywatne.
- Biblioteki naukowe.

ATLAS

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://www.atlasproject.eu/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Tetracom Interactive Solutions, Bułgaria; DFKI, Niemcy; Atlantis Consulting SA, Grecja; Institute for Bulgarian Language, Bułgaria; Universitaet Hamburg, Niemcy; Universitatea Alexandru Ioan Cuza,

Rumunia; Sveučilište u Zadru, Chorwacja; Institute of Technologies and Development Foundation, Serbia

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Adam Przepiórkowski

Osoba kontaktowa: Maciej Ogrodniczuk, maciej.ogrodniczuk@gmail.com

Pozostałe osoby realizujące projekt:

Maciej Ogrodniczuk, IPI PAN; Adam Przepiórkowski, IPI PAN; Mateusz Kopeć, IPI PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 3/1/2010 - 2/8/2013

Źródło finansowania: CIP ICT-PSP 250467

OPIS PROJEKTU

Typ projektu: infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: potoki przetwarzania lingwistycznego; lematyzacja; CMS; system zarządzania dokumentami; automatyczne streszczanie

Skrócony opis projektu:

Celem projektu ATLAS było dostarczenie nowoczesnych metod i narzędzi tworzenia i rozpowszechniania treści wykorzystujących zaawansowane mechanizmy lingwistyczne i realizujących za ich pomocą ważne funkcje użytkowe (automatyczne streszczanie i klasyfikację tekstów, automatyczną identyfikację nazw własnych, liczb i dat, komputerowe wspomaganie tłumaczenia) w celu ułatwienia zarządzania przetwarzaniem tekstów wielojęzycznych.

Skrócony opis części informatycznej:

Część informatyczna obejmowała opracowanie narzędzi informatyczno-lingwistycznych dla języków projektu, w tym polskiego, złożonych w potok przetwarzania tekstów realizujący m.in. identyfikację jednostek nazwniczych, ważnych tematów w postaci fraz rzeczownikowych, streszczanie i maszynowe tłumaczenie tekstów oraz identyfikację tekstów podobnych.

Opis efektów projektu:

Wynikiem projektu są narzędzia do przetwarzania tekstów zapisanych w językach projektowych zestawione w łańcuch realizujący podstawową analizę tekstu oraz serwisy demonstracyjne i-Publisher, i-Librarian oraz EUDocLib demonstrujące ich działanie na przykładzie różnych zbiorów dokumentów.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://zil.ipipan.waw.pl/ATLAS>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Ulepszanie lingwistyczno-informatycznych narzędzi tworzących potok, integracja nowych narzędzi analizy tekstu.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Partnerzy zainteresowani zaawansowaną analizą tekstów w języku polskim.

AUTOMATYCZNA EKSTRAKCYJA INFORMACJI

Osoba wypełniająca ankietę: Kierownik projektu

<http://nlp.pwr.wroc.pl/synat>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Grupa Naukowa G4.19, Instytut Informatyki Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Projekt jest realizowany wyłącznie na Politechnice Wrocławskiej jako etap A13 projektu strategicznego SyNaT, częściowo bazuje na rezultatach z projektu NEKST POIG; prace są kontynuowane w ramach etapu A25, którego celem jest skonstruowanie systemu do wdrożenia.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr inż. Michał Marcińczuk, dr inż. Maciej Piasecki (PWr.), dr inż. Adam Radziszewski, mgr inż. Paweł Kędzia, mgr inż. Jan Kocoń

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 8/16/2011 - 7/31/2014

Źródło finansowania: Narodowe Centrum Badań i Rozwoju SyNaT projekt strategiczny

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: podstawowe narzędzia językowe dla języka polskiego, ujednoznacznianie morfo-syntaktyczne, lematyzacja,

Skrócony opis projektu:

Celem projektu było opracowanie zespołu programów do podstawowej analizy tekstów w języku polskim na poziomie:

- lematyzacji (czyli przypisania wyrazom w tekście form hasłowych)

- morfo-syntaktycznym (czyli rozpoznaniu własności wyrazów w tekście jako elementów struktur składniowych),
- płytkiego opisu składniowego (rozpoznanie uproszczonych struktur składniowych)
- rozpoznania wystąpień nazw własnych (wraz z ich klasyfikacją semantyczną)
- rozpoznania wybranych relacji semantycznych pomiędzy wystąpieniami nazw własnych
- rozpoznanie do czego odnoszą się zaimki
- ujednoznacznienie znaczeń wyrazów w tekście poprzez ich przypisanie do znaczeń ze Słownosieci.

Powyższy ciąg przetwarzania opiera się na opracowanym wcześniej narzędziu do analizy morfologicznej.

Opracowane narzędzia mogą być połączone w ciąg przetwarzania tekstu i mogą być stosowane w bardzo wielu aplikacjach analizy tekstu jako pierwszy etap przetwarzania, np. przy klasyfikacji tematycznej tekstu, analizie leksykalnej, porównywaniu tekstów itd.

Skrócony opis części informatycznej:

Projekt ze swojej natury jest informatyczny, ale dostarcza narzędzi podstawowych do pracy z kolekcjami dokumentów tekstowych.

Większość opracowanych programów-narzędzi jest dostępnych do testów na stronie:

<http://nlp.pwr.wroc.pl/synat/>

Opis efektów projektu:

Opracowane narzędzia są opublikowane na otwartych licencjach (głównie GNU LGPL). Są dostępne zarówno w postaci programów działających pod systemem operacyjnym Linux, jak i usług sieciowych do wykorzystania zdalnego. Narzędzia można testować na stronie: <http://nlp.pwr.wroc.pl/synat/>

Powiązania z innymi polskimi i zagranicznymi projektami:

Podobne zestawy narzędzi językowych są konstruowane dla wielu języków europejskich. Naszą główną ideą było zbudowanie zestawu podstawowych narzędzi językowych, tzn. w pewnym sensie zestawu standardowego.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://nlp.pwr.wroc.pl/synat/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Zbudowane narzędzia mają charakter ogólny, mogą być zastosowane do dowolnego dokumentu tekstowego w języku polskim i mogą być podstawą do konstruowania różnorodnych aplikacji badawczych do pracy z dużymi zbiorami tekstu, np. do inteligentnego wyszukiwania informacji w tekście, porównywania tekstów, automatycznego zbierania danych badawczych z dużych kolekcji tekstowych itd. Dla wielu konkretnych aplikacji i zastosowań badawczych konieczne może być dostosowanie narzędzi i charakterystyki ich działania.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Institucja prowadząca badania w obszarze nauk humanistycznych i społecznych, w ramach których konieczna jest praca z dokumentami tekstowymi lub transkrybowanymi zapisami mowy.

BIBLIOGRAFICZNA BAZA DANYCH ŚWIATOWEGO JĘZYKOZNAWSTWA SŁAWISTYCZNEGO
ISYBISLAW

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://www.isybislaw.ispan.waw.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Centrum Sławistycznej Informacji Naukowej, Instytut Sławistyki Polskiej Akademii Nauk, Polska Akademia Nauk

Instytucja przy której projekt jest afiliowany: Instytut Sławistyki PAN

Pozostałe instytucje zaangażowane w projekt: Instytucje naukowe współpracujące w ramach Protokołów o współpracy między Instytutem Sławistyki PAN i zagranicznymi Akademiemi oraz innymi placówkami naukowymi:

Białoruś

Цэнтр даследаванняў беларускай культуры, мовы і літаратуры Нацыянальнай акадэміі навук Беларусі (wcześniej: Інстытут мовы і літаратуры ім. Я. Коласа і Я. Купалы Нацыянальнай акадэміі навук Беларусі) (Мінск)

Czechy

Ústav pro jazyk český Akademie věd České republiky (Praha)

Macedonia

Истражувачки центар за ареална лингвистика при Македонската академија на науките и уметностите (Скопје)

Niemcy

Serbiski institut (Budyšin) (Sorbisches Institut) (Bautzen)

Rosja

Институт славяноведения Российской академии наук (Москва)

Instytucje naukowe współpracujące bez zawartego porozumienia:

Chorwacja

Ured Predstavnik poljske manjine grada Zagreba (Zagreb)

Estonia

Slaavi filoloogiasakond Tartu Ülikool (Tartu)

Polska

Biblioteka Sejmowa (Warszawa)

Instytut Filologii Polskiej Uniwersytetu im. Adama Mickiewicza (Poznań)

Instytut Filologii Polskiej Uniwersytetu Marii Curie-Skłodowskiej (Lublin)

Instytut Filologii Słowiańskiej Uniwersytetu Wrocławskiego (Wrocław)

Instytut Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego (Warszawa)

Towarzystwo Naukowe Warszawskie (Warszawa)

Rosja

Исторический факультет Московского государственного университета им. М.В. Ломоносова (Москва)

Факультет иностранных языков и регионоведения Московского государственного университета им. М.В. Ломоносова (Москва)

Филологический факультет Московского государственного университета им. М.В. Ломоносова (Москва)

Ukraina

Інститут української мови Національної академії наук України (Київ)

Organizacje międzynarodowe

Komisja Bibliografii Lingwistycznej przy Międzynarodowym Komitecie Słowistów

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: dr Zofia Rudnik-Karwatowa, z.karwatowa@gmail.com

Osoba kontaktowa: dr Marcin Fastyn, avtandil@gazeta.pl

Pozostałe osoby realizujące projekt:

Zofia Rudnik-Karwatowa (IS PAN); Marcin Fastyn (IS PAN); Paweł Kowalski (IS PAN); Jakub Banasiak (IS PAN); Zenon Mikos (Biblioteka Sejmowa, IS PAN); Iwona Łuczków (IFS UW); Beata Maksymiuk-Pacek (IFP UMCS); Stanisława Niebrzegowska-Bartmińska (IFP-UMCS); Błażej Osowski (IFP UAM); Maria Przystek-Samokowa (IR UW; IINiSB UW); Sebastian Wasiuta (IFP UMCS) oraz z zagranicy: Henadz Cychun (Інстытут мовы і літаратуры ім. Я. Коласа і Я. Купалы Нацыянальнай акадэміі навук Беларусі); Aleksandr D. Duličenko (Slaavi filoologia osakond Tartu Ülikool (Tartu)); Walientina G. Kulpiņa (Факультет иностранных языков и регионоведения Московского государственного университета им. М.В. Ломоносова (Москва)); Oksana Ostarczuk (Институт славяноведения Российской академии наук (Москва)); Филологический факультет Московского государственного университета им. М.В. Ломоносова (Москва)); Jana Pařcunová (Ústav pro jazyk český Akademie věd České republiky (Praga)); Anja Pohončowa (Serbski instytut (Budziszyn)); Julija Romanjuk (Інститут української мови Національної академії наук України (Київ)); Marek Stanojević (Ured Predstavnik poljske manjine grada Zagreba (Zagrzeb)); Wiktor A. Tatarinow (Исторический факультет Московского государственного университета им. М.В. Ломоносова (Москва)); Jauhienija Wołkawa (Інстытут мовы і літаратуры ім. Я. Коласа і Я. Купалы Нацыянальнай акадэміі навук Беларусі (Мі́нск))

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 6/1/2006 - 12/31/2064

Źródło finansowania: środki własne instytucji, dofinansowanie z grantów 213291

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny

Efekt projektu: baza danych

Słowa kluczowe: bibliografia naukowa, językonawstwo slawistyczne, baza danych, system informacyjno-wyszukiwawczy, iSybislaw

Skrócony opis projektu:

Bibliograficzna baza danych prezentuje nowoczesny system informacyjno-wyszukiwawczy iSybislaw z zakresu językoznawstwa slawistycznego. Projekt – instytucjonalnie związany z Instytutem Slawistyki Polskiej Akademii Nauk – jest rozwijany we współpracy z polskimi i zagranicznymi placówkami slawistycznymi, zawiera obecnie zbiór informacyjny dotyczący ponad 20 tysięcy dokumentów. Zakres tematyczny systemu iSybislaw obejmuje nie tylko językoznawstwo slawistyczne, językoznawstwo konfrontatywne i językoznawstwo kontrastywne słowiańsko-niesłowiańskie, lecz także w mniejszym stopniu językoznawstwo teoretyczne i ogólne. System informacyjno-wyszukiwawczy iSybislaw prezentuje dorobek naukowy językoznawstwa slawistycznego w okresie od 1992 r. do dnia dzisiejszego. W systemie iSybislaw rozbudowana jest struktura informacyjna rekordów (formalna i rzeczowa), uwzględniająca m.in. słowa kluczowe i klasyfikację, a także abstrakty i adnotacje (czasem w dwóch językach: języku obcym i polskim). Korzystanie z zasobów informacyjnych systemu iSybislaw ułatwia słownik słów kluczowych z zakresu językoznawstwa slawistycznego (w języku polskim).

Skrócony opis części informatycznej:

System iSybislaw został wykonany jako aplikacja webowa (użytkownik na swoim komputerze musi mieć zainstalowaną jedynie przeglądarkę internetową). Część serwerowa została wykonana w technologiach dostępnych bezpłatnie na licencjach Open Source. System składa się z serwera baz danych PostgreSQL oraz serwera aplikacyjnego JBoss. Aplikacja została napisana w języku Java, a do jej uruchomienia zastosowano system operacyjny Linux. Dane tekstowe w systemie przechowywane są w kodowaniu Unicode (UTF-8).

Opis efektów projektu:

Internetowa baza danych udostępniająca bezpłatnie i na wolnej licencji informacje bibliograficzne dotyczące dorobku naukowego językoznawstwa naukowego od roku 1992 do dziś (obecnie ponad 20 tysięcy dokumentów). Planowane jest dalsze poszerzanie dostępnego zbioru informacji, a także obsługa przeszukiwania pełnych tekstów dokumentów udostępnionych na licencji open access oraz umożliwienie użytkownikom dodawania tagów.

Powiązania z innymi polskimi i zagranicznymi projektami:

Baza danych iSybislaw jest merytorycznie powiązana z repozytorium cyfrowym tekstów slawistycznych iReteslaw, opracowywanym w Instytucie Slawistyki PAN.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

http://www.ciniba.edu.pl/index.php?option=com_content&view=article&id=852%3Aisybislaw-bibliograficzna-baza-danych-wiatowego-jzykoznawstwa-slawistycznego&catid=15%3Abazy-danych&Itemid=36

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Dalsza rozbudowa zasobów bibliograficznych i powiększanie liczby dokumentów zawartych w systemie.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Wszystkie slawistyczne instytucje naukowe.

BIBLIOLOGICZNA BIBLIOTEKA CYFROWA

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Stowarzyszenie Bibliotekarzy Polskich

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Zdzisław Dobrowolski, z.dobrowolski@uw.edu.pl

Osoba kontaktowa: Zdzisław Dobrowolski, z.dobrowolski@uw.edu.pl

Pozostałe osoby realizujące projekt:

Zdzisław Dobrowolski, UW

Michał Krupa, UW

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 12/12/2009 - 12/31/2064

Źródło finansowania: środki własne instytucji

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny

Efekt projektu: baza danych, repozytorium

Słowa kluczowe: biblioteka cyfrowa ; digitalizacja ; bibliologia ; informacja naukowa

Skrócony opis projektu:

Unikalna biblioteka cyfrowa gromadząca podstawowy korpus publikacji z zakresu bibliotekarstwa, historii książki i bibliotek oraz informacji naukowej. Zamieszcza również książki i czasopisma najnowsze.

Skrócony opis części informatycznej:

Bibliologiczna Biblioteka Cyfrowa oparta jest na oprogramowaniu dLibra 4.0

Opis efektów projektu:

Bibliologiczna Biblioteka Cyfrowa jest unikalnym przedsięwzięciem na skalę europejską, bowiem podstawowy wydawca książek z zakresu bibliologii i informatologii w Polsce, Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, zdecydowało się na digitalizację swojego zasobu i udostępnienia go w postaci otwartej biblioteki cyfrowej. Wykonawcą tego projektu jest Instytut Informacji Naukowej UW

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Digitalizacja kolejnych książek i czasopism wydawanych przez Stowarzyszenie Bibliotekarzy Polskich.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

BIBLIOTEKA CYFROWA KLF UW

Osoba wypełniająca ankietę: Joanna Bilińska, j.bilinska@uw.edu.pl

<http://bc.klf.uw.edu.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Katedra Lingwistyki Formalnej

Wydział Neofilologii

Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Janusz S. Bień, jsbien@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 8/13/2008 -

Źródło finansowania: środki własne instytucji

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy

Efekt projektu: baza danych, repozytorium

Słowa kluczowe: biblioteka cyfrowa, repozytorium, Katedra Lingwistyki Formalnej, Uniwersytet Warszawski, OAI-PMH

Skrócony opis projektu:

Biblioteka Cyfrowa KLF UW stanowi repozytorium i bibliografia publikacji oraz innych opracowań pracowników i współpracowników KLF.

Skrócony opis części informatycznej:

System EPrints (<http://www.eprints.org/>) odpowiednio skonfigurowany na serwerze Katedry (Debian GNU/Linux).

Opis efektów projektu:

Pierwsze, jak się wydaje, repozytorium instytucjonalne w Polsce funkcjonujące nieprzerwanie od 2008 r.

Podobne biblioteki cyfrowe są liczne, ale ta oferuje pewne pożyteczne funkcje niedostępne gdzie indziej, takie jak np. brak ostrej granicy pomiędzy repozytorium a bibliografią, funkcja prośby o udostępnienie egzemplarza elektronicznego niedostępnego w publicznej części biblioteki (np. ze względu na ograniczenia nałożone przez wydawcę).

Powiązania z innymi polskimi i zagranicznymi projektami:

WorldCat i Federacja Bibliotek Cyfrowych oraz niektóre inne agregatory pobierają z biblioteki za pomocą protokołu OAI-PMH metadane do swoich katalogów.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://bc.klf.uw.edu.pl/information.html>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Polonizacja systemu EPrints w formie tzw. modułu Bazaar, poprawienie eksportu metadanych do BibTeXa, dodanie dodatkowych typów eksportowanych danych, np. biblatex, różne inne ulepszenia.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Udostępnianie metadanych, wymiana doświadczeń i ulepszeń systemu.

Profil potencjalnego partnera:

Inni administratorzy oraz użytkownicy bibliotek i repozytoriów cyfrowych, przede wszystkim systemu EPrints.

BIULETYN POLONISTYCZNY

Osoba wypełniająca ankietę: Kierownik projektu

www.biuletynpolonistyczny.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Centrum Humanistyki Cyfrowej Instytutu Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: polskie i zagraniczne instytucje polonistyczne

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Mariola Wilczak, mariola.wilczak@ibl.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Przemysław Górecki, UAM; Sylwia Pikula, KUL; Marlena Sęczek, IBL PAN; Mariola Wilczak, IBL PAN; Olga Zakolska, IBL PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 10/1/2013 - 6/30/2015

Źródło finansowania: środki własne instytucji

OPIS PROJEKTU

Typ projektu: bazodanowy

Efekt projektu: baza danych

Słowa kluczowe: polonistyka; portal specjalistyczny; personalizacja; model pracy rozproszonej; narzędzia pracy naukowej

Skrócony opis projektu:

Celem projektu jest utworzenie portalu specjalistycznego, współtworzonego przez polonistów i kulturoznawców w kraju i zagranicą, integrującego informacje o bieżącym życiu naukowym i prezentującego je w przystępnej formie. Przyjęty model prac redakcyjnych (model pracy rozproszonej) i współtworzenie „Biuletynu” przede wszystkim przez instytucje zintegrują środowisko polonistyczne i nadadzą „Biuletynowi” charakter instytucjonalny. Otwarty dostęp do portalu, połączony ze zróżnicowaniem poziomów jego użytkowania uczynią z niego źródło informacji zarówno dla odbiorców okazjonalnie zainteresowanych osiągnięciami polonistyki, jak i badaczami chcącymi opublikować informacje o swoich osiągnięciach.

„Biuletyn Polonistyczny” będzie także narzędziem do nawiązania współpracy interdyscyplinarnej i autopromocji środowiska polonistycznego i przyczyni się do wypełnienia luki w kształtującym się na gruncie polskim paradygmacie cyfrowej humanistyki.

Skrócony opis części informatycznej:

„Biuletyn” będzie przestrzenią agregowania treści bieżących z dziedziny polonistyki i kulturoznawstwa; w mniejszym stopniu będzie on pełnił rolę archiwizującą. Posiadać będzie: wyszukiwarkę z właściwymi dla danego modułu, zróżnicowanymi kryteriami filtrowania; newsletter; zestaw tagów (słów kluczowych); narzędzia umożliwiające personalizację newslettera i strony głównej; profilowane konta użytkownika (czytelnika, redaktora oraz administratora systemu).; kalendarz – z możliwością oznaczenia i zapamiętania wybranych wydarzeń; funkcję przypominania o upływających terminach zgłoszeń, procesów rekrutacyjnych itp.; narzędzia umożliwiające organizowanie wydarzeń naukowych.

Portal posiadać będzie mechanizm umożliwiający stworzenie anglojęzycznej wersji portalu, o strukturze identycznej do wersji polskojęzycznej.

Opis efektów projektu:

Efektom będzie budowa portalu specjalistycznego, współtworzonego przez polonistów i kulturoznawców w kraju i za granicą, integrującego w postaci bazy informacje o bieżącym życiu naukowym (na początku o wydarzeniach naukowych, pracach badawczych i rozprawach doktorskich) i prezentującego je w przystępnej formie. „Biuletyn” skierowany będzie przede wszystkim do szeroko pojętego środowiska polonistycznego, tj. instytucji, badaczy, dydaktyków i studentów z obszarów nauk o literaturze i kulturze polskiej. Docelowo stanie się on wyspecjalizowanym narzędziem w pracy polonistów, aktywnie przez nich współtworzonym i rozbudowywanym o kolejne przydatne funkcjonalności.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Budowa portalu ma umożliwić w dalszej perspektywie dodanie nowych funkcjonalności, m.in.: implementację narzędzi pomocnych w prowadzeniu badań i projektów oraz narzędzi do analizy statystycznej i mapowania treści, pozwalających uzyskać obraz tendencji w badaniach naukowych i rozwoju kierunków badawczych; implementację innych najnowocześniejszych narzędzi umożliwiających retrospektywną analizę badań polonistycznych i kulturoznawczych, pozwalającą na studia nad historią polonistyki i kulturoznawstwa; publikowanie danych poprzez API (interfejsu programowania aplikacji).

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Możliwość informowania o realizowanych pracach badawczych i projektach na kilka sposobów: poprzez tworzenie podstron wydarzeń związanych z projektami, publikowanie tekstów informujących o projektach czy też odsyłanie do istniejących stron www projektów.

Profil potencjalnego partnera:

Podmioty prowadzące badania na polu nauk o literaturze, kulturze i języku.

BLOG JAKO NOWA FORMA PIŚMIENICTWA MULTIMEDIALNEGO

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Centrum Humanistyki Cyfrowej, Instytut Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Politechnika Wrocławska (CLARIN-PL) - współpraca
Computational Stylistics Group - współpraca

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Maryl, maciej.maryl@ibl.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr Maciej Maryl, IBL PAN; dr Anna Buchner, ISNS UW; mgr Krzysztof Niewiadomski, IH PAN; mgr Maciej Kidawa, UW;

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 8/24/2012 - 8/24/2014

Źródło finansowania: Narodowe Centrum Nauki 2011/03/N/HS2/06232

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: publikacja, metodologia

Słowa kluczowe: Piśmiennictwo multimedialne; blogi; analiza dyskursu; genologia; computer assisted data analysis

Skrócony opis projektu:

Celem projektu jest analiza blogu jako nowej formy piśmiennictwa multimedialnego, w której przecinają się wpływy tradycyjnych gatunków wypowiedzi i nowych form komunikacji elektronicznej.

Blog, czyli prosta strona internetowa, na której wpisy ukazują się w antychronologicznej kolejności i mogą być komentowane przez odbiorców, stanowi specyficzne połączenie tradycyjnych form wypowiedzi (m.in. dokumentu osobistego, felietonu czy poradnika) i dyskursu elektronicznego (wtórna oralność; interakcja z

odbiorcą), choć jego własny status gatunkowy pozostaje niejasny i zbliżony do funkcji kanału przekazywania innych form gatunkowych.

Analiza wpływu innych form komunikacji multimedialnej i tradycyjnych form piśmiennictwa na blog, uzupełniona badaniami przebiegu komunikacji z odbiorcami (np. w formie komentarzy pod wpisami), pozwoli (1) usytuować blog na mapie współczesnych form wypowiedzi, (2) określić status gatunkowy blogu i wypracować typologię gatunków blogowych; (3) opracować poetykę tekstu blogowego. Metodologia wypracowana w toku prac nad projektem będzie stanowić podstawę do dalszych badań nad piśmiennictwem elektronicznym

Skrócony opis części informatycznej:

Blogi wybrane do analizy zostaną szczegółowo opisane (otagowane) przy pomocy Atlas.ti – specjalistycznego oprogramowania do analizy danych (Computer Assisted Data Analysis). Dane tekstowe zostaną ponadto przeanalizowane przy pomocy oprogramowania do komputerowej analizy dyskursu (np. WordSmith, stylo), które pozwoli porównać formy komunikacji językowej we wszystkich badanych blogach.

Opis efektów projektu:

Wyniki badań będą stanowić wkład zarówno w badania literackie, jak i w refleksję kulturoznawczą. Zasadnicze nowatorstwo projektu widzimy w proponowanej metodzie badawczej (połączenie metod literaturoznawczych, komputerowej analizy tekstu i socjologicznych metod zbierania danych), która może posłużyć za podstawę do dalszych badań na polu piśmiennictwa elektronicznego, prowadzonych ze świadomością literaturoznawczą. Za osobny, choć równie istotny, efekt projektu uznajemy popularyzację problematyki piśmiennictwa elektronicznego wśród badaczy literatury.

Powiązania z innymi polskimi i zagranicznymi projektami:

Politechnika Wroclawska (CLARIN-PL) - współpraca przy opracowaniu materiału (strony blogowe) i przeprowadzeniu analiz językowych. W efekcie powstaje narzędzie do przeprowadzania podobnych analiz.

Computational Stylistics Group - współpraca przy stylometri

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://maryl.org/publikacje/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Wykorzystanie wypracowanych narzędzi i metodologii do badań nad innymi formami piśmiennictwa multimedialnego.

Rozwinięcie, dopracowanie i udostępnienie narzędzi analitycznych, np. w formie infrastruktury badawczej.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

CLARIN

Profil potencjalnego partnera:

Instytucje badawcze (kultura i piśmiennictwo cyfrowe).

Instytucje informatyczne (narzędzia do analizy dyskursu i przetwarzania tekstów cyfrowych)

CESAR

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://www.meta-net.eu/projects/cesar>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Research Institute for Linguistics, Hungarian Academy of Sciences;

Budapest University of Technology and Economics;

Institute for Bulgarian Language, Bulgarian Academy of Sciences;

Faculty of Humanities and Social Sciences, University of Zagreb;

Uniwersytet Łódzki;

Faculty of Mathematics, University of Belgrade;

Mihajlo Pupin Institute;

Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Adam Przepiórkowski

Osoba kontaktowa: Maciej Ogrodniczuk, maciej.ogrodniczuk@gmail.com

Pozostałe osoby realizujące projekt:

Maciej Ogrodniczuk, IPI PAN; Adam Przepiórkowski, IPI PAN; Łukasz Szałkiewicz, UW; Marcin Miłkowski, IFiS PAN (ogółem ok. 50 wykonawców)

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 2/1/2011 - 1/31/2013

Źródło finansowania: CIP ICT-PSP 271022

OPIS PROJEKTU

Typ projektu: infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie, repozytorium

Słowa kluczowe: META-NET; META-SHARE; cyfrowe narzędzia i zasoby językowe; udostępnianie narzędzi i zasobów lingwistycznych; usługi sieciowe

Skrócony opis projektu:

Celem projektu było opracowanie zintegrowanego repozytorium zasobów i narzędzi badawczych dla języków: węgierskiego, polskiego, słowackiego, serbskiego oraz chorwackiego. W ramach projektu zostały zbudowane, rozszerzone i zrównoleżone liczne zasoby oraz narzędzia językowe, m.in. korpusy referencyjne oraz równoległe, leksykony oraz elektroniczne słowniki morfologiczne, anotowane korpusy językowe, banki drzew składniowych, multimedialne bazy danych mówionych, systemy identyfikacji nazw własnych, systemy anotacji lingwistycznej itp. Ważnym aspektem upowszechnienia wyników projektu było uregulowanie sytuacji prawnej udostępnianych zasobów.

Skrócony opis części informatycznej:

Rozwój i zrównoleżenie tak dużego i różnorodnego zbioru zasobów i narzędzi dla pięciu języków wymagał wypracowania standardów informatycznych, które mogły zapewnić odpowiedni poziom interoperacyjności między scalanymi narzędziami, co ma umożliwić ich wykorzystanie w tworzeniu zaawansowanych, wielojęzycznych systemów przetwarzania lingwistycznego. Po opracowaniu uniwersalnych standardów anotacji i wymiany danych, zasoby i narzędzia opracowane i zintegrowane w ramach projektu zostały udostępnione szerokiej ogółowi badaczy w całej Europie oraz podmiotom komercyjnym w postaci źródłowej oraz poprzez interfejs programistyczny (usługi sieciowe).

Opis efektów projektu:

Efektom działań konsorcjum było opracowanie i udostępnienie w infrastrukturze META-SHARE 251 narzędzi i zasobów językowych, w tym 59 dla języka polskiego (zasoby korpusowe, leksykalne, formalne opisy języka polskiego, narzędzia i usługi do przetwarzania języka polskiego). Udostępnione zasoby zostały przygotowane zgodnie z najnowszymi standardami lingwistycznymi oraz informatycznymi; w przypadku wszystkich została także uregulowana ich sytuacja prawna. Ważnym wynikiem projektu było także utworzenie portalu informacyjnego „Lingwistyka komputerowa w Polsce” (CLIP – Computational Linguistics in Poland).

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekty współfinansowane przez Komisję Europejską w ramach programu ICT-PSP: zbiorczy projekt META-NET/T4ME (nr kontraktu 249119), METANET4U (nr kontraktu 270893) i META-NORD (nr kontraktu 270899).

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://clip.ipipan.waw.pl/CESAR>

<http://glass.ipipan.waw.pl/metashare/>

<http://www.meta-net.eu/>

<http://www.meta-net.eu/meta-share/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Digitalizacja, standaryzacja, rozszerzenie, uzupełnienie, rozbudowa, zrównoleżenie narzędzi i zasobów lingwistycznych dla języka polskiego.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Instytucje dysponujące zasobami i narzędziami do przetwarzania języka polskiego, które chciałyby włączyć je do dużej infrastruktury udostępniania zasobów.

COMPUTATIONAL STYLISTICS GROUP

Osoba wypełniająca ankietę: Jan Rybicki, jkrybicki@gmail.com

<https://sites.google.com/site/computationalstylistics/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Computational Stylistics Group

Instytucja przy której projekt jest afiliowany: UP,UJ,Uniwersytet w Antwerpii

Pozostałe instytucje zaangażowane w projekt: Uniwersytet w Lipsku

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Eder, maciejeder@gmail.com; Mike Kestemont, mike.kestemont@gmail.com; Jan Rybicki, jkrybicki@gmail.com

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Maciej Eder, UP;

Mike Kestemont, Uniw. w Antwerpii;

Jan Rybicki, UJ

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 6/14/2011 - 6/14/2020

Źródło finansowania: środki własne uczestników projektu nie dotyczy

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, repozytorium, publikacja, kurs

Słowa kluczowe:

Skrócony opis projektu:

Zespół zajmuje się produkcją narzędzi komputerowych do badań nad tekstem, ze szczególnym uwzględnieniem atrybucji autorskiej i stylometrii. Główny nacisk kładziony jest na stosowanie metod komputerowo-statystycznych do badań literackich i przekładoznawczych.

Skrócony opis części informatycznej:

Obecnie trwają prace nad rozszerzeniem istniejącego pakietu stylometrycznego "stylo" w środowisku programowania statystycznego R; dodawane są nowe funkcje, możliwości wyboru dodatkowych parametrów; trwa również dostosowywanie pakietu do współpracy z innymi programami i środowiskami dostosowanymi w szeroko rozumianej analizie tekstowej.

Opis efektów projektu:

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

CORE

Osoba wypełniająca ankietę: Kierownik projektu

<http://zil.ipipan.waw.pl/CORE>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Instytut Podstaw Informatyki PAN

Pozostałe instytucje zaangażowane w projekt: Brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Ogrodniczuk, maciej.ogrodniczuk@gmail.com

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Maciej Ogrodniczuk, IPI PAN; Mateusz Kopeć, IPI PAN; Magdalena Zawistawska, UW; Katarzyna Głowińska, Lingventa; Agata Savary, François Rabelais University Tours

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 4/18/2011 - 7/17/2014

Źródło finansowania: Narodowe Centrum Nauki 6505/B/T02/2011/40

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, oprogramowanie, publikacja, ręcznie anotowany korpus tekstów

Słowa kluczowe: koreferencja; anafora; Polski Korpus Koreferencyjny; identyfikacja nawiązań;

Skrócony opis projektu:

Celem projektu jest stworzenie nowatorskich metod automatycznej identyfikacji nawiązań i narzędzi informatycznych do tego celu dla języka polskiego. Nawiązania, zwane inaczej konstrukcjami koreferencyjnymi, są strukturami lingwistycznymi służącymi do reprezentacji wzajemnych zależności między częściami tekstu odwołującymi się do tych samych jednostek pojęciowych. Automatyczna identyfikacja nawiązań oznacza np. że w tekście «Wyrok nie jest prawomocny. Matka jednego z zabitych policjantów wyszła z sali już podczas jego ogłaszania.

program powinien rozpoznać, że niepełnoznaczny zaimek «jego» odnosi się do wyroku, a nie do zabitego policjanta. Samo użycie nawiązań służy podtrzymaniu spójności tematycznej tekstu (poprzez odwołanie w

kolejnych zdaniach tekstu do informacji występującej w zdaniach poprzedzających), przy jednoczesnym zachowaniu jego poprawności gramatycznej i stylistycznej (poprzez przywołanie wyrażen zależnych, jednak nie identycznych z użytymi uprzednio).

Skrócony opis części informatycznej:

Celem informatycznej części projektu projektu jest opracowanie komputerowych technik, narzędzi oraz zasobów realizujących zadanie automatycznej identyfikacji nawiązań w języku polskim. Metody te powinny pozwolić na otrzymanie wyników o jakości porównywalnej z uzyskiwaną obecnie dla innych języków.

Opis efektów projektu:

Wynikiem projektu są dostępne na zasadach licencji CC-BY ręcznie anotowany korpus nawiązań (Polski Korpus Koreferencyjny) opracowany na bazie Narodowego Korpusu Języka Polskiego oraz narzędzia do automatycznej identyfikacji nawiązań.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Kontynuację projektu, pomocną w identyfikacji nawiązań w języku polskim, mogłoby stanowić opracowanie lingwistyczno-informatycznych metod konstrukcji, normalizacji, weryfikacji i aktualizacji otwartej i ogólnodostępnej referencyjnej bazy utrwalonych omownych wyrażen nominalnych (tj. skonwencjonalizowanych fraz rzeczownikowych), które stanowią odzwierciedlenie elementów wiedzy ogólnej w języku wraz z ich tekstowo częstymi wariantami. Na przykład, dla hasła „Neil Armstrong” baza zawierałaby frazę "pierwszy człowiek na Księżycu", a dla hasła „pielęgniarka” – „biały personel”, o ile odpowiedni stopień utrwalenia obu tych wyrażen uda się zweryfikować w referencyjnych korpusach języka polskiego.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Lingwiści mogący wziąć udział w opracowaniu bazy referencyjnej, dysponujący wiedzą i/lub materiałami pomocnymi w jej tworzeniu.

CORPOGRABBER

Osoba wypełniająca ankietę: Jan Kocoń, jan.kocon@pwr.edu.pl

<http://nlp.pwr.wroc.pl/redmine/projects/corpora/wiki/CorpoGrabber>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Grupa Naukowa G4.19, Instytut Informatyki Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Projekt jest realizowany wyłącznie na Politechnice Wrocławskiej

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

mgr inż. Jan Kocoń, PWr

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 1/2/2014 - 12/19/2014

Źródło finansowania: dotacja celowa CLARIN-PL dotacja na inwestycję MNiSW

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: korpus, pozyskiwanie tekstu, narzędzie językowe, deduplikacja tekstu, wydobywanie informacji

Skrócony opis projektu:

Celem projektu było opracowanie narzędzia, które służy do pozyskiwania danych z Internetu na potrzeby tworzenia dużych korpusów tekstu. Narzędzie umożliwia wydobywanie tekstu ze stron internetowych, deduplikację pobranego tekstu na poziomie akapitów oraz odfiltrowanie tekstów obcojęzycznych.

Skrócony opis części informatycznej:

Narzędzie CorpoGrabber umożliwia pobranie wszystkich podstron dla każdego adresu wejściowego do głębokości określonej przez użytkownika (wykorzystano pomocniczo narzędzie WGET). Możliwe jest zdefiniowanie maksymalnej wielkości pobieranych plików oraz dodatkowe filtry, dzięki którym pobierane są jedynie pliki HTML. Z pobranych dokumentów w formacie HTML wydobywana jest treść istotna z punktu

widzenia użytkownika (np. tytuł i treść artykułu w postaci czystego tekstu, bez dodatkowych elementów na stronie - wykorzystano pomocniczo narzędzie JusText). Otrzymany tekst jest poddawany procesowi deduplikacji na poziomie akapitów. Następnie dokumenty są poddawane analizie morfologicznej i odfiltrowywane wszystkie te, które zawierają mniej niż określoną liczbę znaków lub zawierających więcej niż określoną część słów spoza słownika analizatora (dokumenty obcojęzyczne lub niebędące poprawnym dokumentem tekstowym - wykorzystano pomocniczo narzędzie MACA). Ostatnim etapem jest ujednoznacznianie morfosyntaktyczne pozostałych tekstów przy pomocy narzędzia WCRFT.

Opis efektów projektu:

Projekt realizowany był na potrzeby systemu do klasyfikacji tekstów o kulturze, dla którego wejściem były adresy stron WWW wskazane przez użytkownika, a wyjściem strony (lub ich fragmenty) przypisane automatycznie do określonych klas semantycznych. W ramach zadania zleconego przez użytkownika, przy pomocy narzędzia CorpoGrabber pozyskano ok. 30000 dokumentów zawierających prawie 7 000 000 tokenów, które posłużyły jako dane do dalszej analizy.

Powiązania z innymi polskimi i zagranicznymi projektami:

Narzędzie utworzone w ramach projektu może posłużyć do pozyskiwania dokumentów tekstowych z Internetu na potrzeby przygotowania danych dla innych narzędzi inżynierii języka naturalnego (np. wspomniany system do klasyfikacji tekstów o kulturze wykorzystuje narzędzie jako źródło danych do klasyfikacji).

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://nlp.pwr.wroc.pl/redmine/projects/corpora/wiki/CorpoGrabber>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Planowane jest przekształcenie narzędzia do postaci usługi sieciowej oraz aplikacji webowej, dzięki której użytkownik będzie mógł zdefiniować źródła danych w postaci adresów internetowych i pozyskać z nich próbki oczyszczonego tekstu do dalszych badań. Planowana jest także rozbudowa narzędzia o możliwość wydobywania określonych informacji z dokumentu częściowo ustrukturalizowanego (HTML), np. tytuł artykułu, autor, data utworzenia, treść i autorzy komentarzy, itd.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Narzędzie utworzone w ramach projektu może posłużyć do pozyskiwania dokumentów tekstowych z Internetu na potrzeby przygotowania danych dla innych narzędzi inżynierii języka naturalnego.

Profil potencjalnego partnera:

Instytucje, które tworzą oraz wykorzystują systemy przetwarzające język naturalny.

CYFROWA BIBLIOTEKA NARODOWA POLONA

Osoba wypełniająca ankietę: Katarzyna Ślaska, k.slaska@bn.org.pl

www.polona.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Biblioteka Narodowa

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Laboratorium EE

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Mikołaj Baliszewski, m.baliszewski@bn.org.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Mikołaj Baliszewski, BN; Jan Mejor, BN; Agnieszka Leszyńska, BN; Paweł Mazur, BN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: -

Źródło finansowania: dotacja celowa, środki własne instytucji

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy, digitalizacyjny

Efekt projektu: oprogramowanie, baza danych, repozytorium

Słowa kluczowe: biblioteka cyfrowa; digitalizacja; udostępnianie; zbiory biblioteczne

Skrócony opis projektu:

Cyfrowa Biblioteka Narodowa Polona została uruchomiona w 2006 roku, następnie poddana gruntownej modernizacji i oddana do użytku w nowej postaci w 2013 roku. Jest to usługa internetowa umożliwiająca udostępnianie w sieci zdigitalizowanych tradycyjnych publikacji papierowych oraz publikacji elektronicznych. Do zgromadzonego zasobu dostęp jest darmowy i nieobjęty żadnymi ograniczeniami formalnymi. Ważnym założeniem jest także ułatwienie użytkownikom poznawania jej obszernego zasobu. W tym celu tworzone są kolekcje, w których zgromadzono obiekty związane z różnymi tematami. Poznawanie zasobu ma ułatwić też założony w tym samym celu blog (<http://blog.polona.pl/>). Bardzo istotną częścią misji serwisu jest próba jak najszerszego udostępnienia zasobu Biblioteki Narodowej. Między innymi w tym celu istotną częścią promocji jej zasobów jest duża aktywność na serwisach społecznościowych (facebook, google+, twitter) oraz narzędzia

ułatwiający użytkownikom udostępnianie na tego typu serwisach treści z zasobu Polony. Zasoby Polony w czerwcu 2014 r. liczyły ponad 290 tys. obiektów.

Skrócony opis części informatycznej:

W CBN Polona skany prezentowane są w formacie jpg. Udostępniane są w wysokiej jakości, bez znaków wodnych, nie jest też potrzebna instalacja dodatkowych aplikacji do ich przeglądania. W postaciach zbiorczych udostępniane są w formacie PDF (częściowo wraz z możliwością wyszukiwania w tekście dzięki warstwie OCR). Skany na stronie można znacznie powiększyć dzięki narzędziu "zoom", ponadto pobrać w najwyższej dostępnej rozdzielczości (jako pojedynczy plik lub w postaci zbiorczej PDF). Strona oparta jest na technologii one-page application (do działania nie wymaga przeładowania) oraz RWB (Responsive Web Design) dzięki czemu dopasowuje się płynnie do rozmiaru okna przeglądarki. Polona umożliwia korzystanie z narzędzi ułatwiających pracę z czytany publikacjami (możliwość tworzenia własnych kolekcji, notatek, zakładki itp.) oraz udostępniania ich za pomocą portali społecznościowych. Jako część większego systemu jest powiązana bezpośrednio z oprogramowaniem digitalizacyjnym Biblioteki Narodowej, co pozwala skanować i udostępniać powyżej 100 tysięcy skanów dziennie.

Opis efektów projektu:

Efektom prac jest serwis, z którego można korzystać bezpłatnie rejestracji bądź po zalogowaniu się, podzielony na trzy główne działy:

1. Strona główna jest określana jako „cyfrowa pinakoteka” i zawiera prezentację subiektywnie wybranych obiektów, fragmentów skanów, będących równocześnie odnośnikiem do publikacji, z których pochodzą.
2. Biblioteka dająca możliwość swobodnego poruszania się pośród zgromadzonych w serwisie publikacji. Posiada wyszukiwarkę (wersja uproszczona i zaawansowana) oraz szereg filtrów pozwalających zawęzić wyniki (typ, źródło, czas, autor, miejsce, wydawca, język, częstotliwość, gatunek, słowa kluczowe) oraz funkcję przeglądania zbiorów poprzez przygotowane przez twórców portalu kolekcje. Istnieje możliwość zmiany sposobu przeglądania obiektów z "kafelków" na "listę".
3. Moje zbiory - zakładka dostępna po wcześniejszym zarejestrowaniu/zalogowaniu. Przy każdym obiekcie można dokonać działań pozwalających w przyszłości powrócić do interesujących użytkownika treści. Istnieje możliwość tworzenia własnych kolekcji obiektów (Moje obiekty) oraz systemu notatek i zaznaczeń (punktowych i obszarowych), a także zakładek.

Po wybraniu interesującego użytkownika obiektu trafia on na stronę obiektu, gdzie posiada możliwość przeglądania skanów, powiększania ich za pomocą narzędzia "zoom", podglądu miniatur u dołu strony bądź po wybraniu odpowiedniej opcji na stronie głównej obiektu, oraz dostęp do panelu pobierania pojedynczych skanów i postaci zbiorczych. Każdy obiekt posiada zakładkę "info" w której znajduje się opis bibliograficzny publikacji pochodzący z katalogów Biblioteki Narodowej. W obszarze prezentacji dokumentu można też przeszukać treść publikacji (jeśli istnieje jej wersja tekstowa), dodać publikację do swojej kolekcji, utworzyć swoje notatki, dodać zakładkę do tekstu, udostępnić publikację w serwisach społecznościowych lub poprzez wiadomość e-mail, skomentować publikację na serwisach społecznościowych.

Powiązania z innymi polskimi i zagranicznymi projektami:

Cyfrowa Biblioteka Narodowa Polona współpracuje z Biblioteką Europejską oraz serwisem Europeana. Jej zasoby są dostępne poprzez wyszukiwarkę Federacji Bibliotek Cyfrowych.

Polona będzie zawierała przekierowania do dokumentów cyfrowych chronionych prawem autorskim, udostępnianych w systemie Academia.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.polona.pl>

<http://blog.polona.pl/>

<https://www.facebook.com/bnpolona?ref=ts&fref=ts>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Planowane jest stałe powiększanie zasobów Polony oraz rozwój jej funkcjonalności.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Polona może dostarczać metadanych oraz zasobów cyfrowych dla większości polskich projektów z zakresu humanistyki cyfrowej. Zasoby Polony zawierają reprodukcje książek, gazet i czasopism, map, nut, druków ulotnych, grafik i fotografii, pocztówek, rękopisów, albumów i starych druków.

Profil potencjalnego partnera:

Wszystkie instytucje realizujące projekty z zakresu humanistyki cyfrowej

CYFROWA ETNOGRAFIA - BIBLIOTEK CYFROWA PILSKIEGO INSTYTUTU ANTROPOLOGII (BC PIA)

Osoba wypełniająca ankietę: Kierownik projektu

www.cyfrowaetnografia.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Polski Instytut Antropologii (fundacja)

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Instytut Etnologii i Antropologii Kulturowej UW,

Informatyczne Centrum Matematyczne UW,

Instytut Sztuki PAN,

Instytut Archeologii i Etnologii PAN,

Polskie Towarzystwo Ludoznawcze,

Komitet Nauk Etnologicznych

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Joanna Koźmińska, j.m.kozminska@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

mgr Joanna Koźmińska - Polski Instytut Antropologii oraz Instytut Etnologii i Antropologii Kulturowej UW;

mgr Joanna Zamorska - doktorantka w Instytucie Etnologii i Antropologii Kulturowej UW oraz Polski Instytut Antropologii;

Opieka prawna: dr Sybilla Kloc-Stanisławska, Instytut Prawa Własności Intelektualnej UJ;

Informatyk, obsługa systemu i serwera: mgr Paweł Kołodziej,

Pracownia Interakcji WUCH (<http://wuch.pl>);

Prawa autorskie: mgr Krzysztof Braun – emerytowany pracownik Archiwum Instytutu Etnologii i Antropologii Kulturowej UW;

Konsultant merytoryczny projektu: dr Paweł Krzyworzeka - współtwórca projektu Cyfrowa Etnografia, adiunkt w Akademii Leona Koźmińskiego w Warszawie, Katedra Zarządzania Międzynarodowego, stypendysta Fundacji Fullbrighta

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 12/10/2009 -

Źródło finansowania: dotacja celowa, środki własne instytucji, Działalność Upowszechniająca Naukę MNiSW (2008-2009,2011-12; program Partnerzy finansowany przez Fundację na rzecz Nauki Polskiej 2009-2011

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy, digitalizacyjny

Efekt projektu: narzędzie badawcze, baza danych, repozytorium, publikacja

Słowa kluczowe: biblioteka cyfrowa;

antropologia kulturowa;

etnologia

Skrócony opis projektu:

Biblioteka Cyfrowa PIA (www.cyfrowaetnografia.pl) – jedyna polska biblioteka cyfrowa o tematyce etnologiczno/antropologicznej, zintegrowana z menedżerem bibliografii: programem Zotero. Tworzona projektowo, bez etatów, powstała w Polskim Instytucie Antropologii w 2008 roku, we współpracy z Instytutem Etnologii i Antropologii Kulturowej UW i instytucjami partnerskimi. Od 2011 r. zrzeszona w Federacji Bibliotek Cyfrowych, indeksowana w Google Scholar. Od uruchomienia w roku 2010 zanotowaliśmy prawie 5,5 miliona wejść, w bazie znajduje 5033 publikacji, z czego pełnotekstowo udostępnionych jest 2774 artykułów. Dzięki wprowadzeniu metadanych, wszystkie artykuły niedostępne pełnotekstowo stanowią bazę bibliograficzną. Innowacyjność BC PIA polega na implementacji programu Zotero oraz publikowaniu pojedynczych artykułów, z osobnymi metadanymi. Zawartość BC PIA to zasób złożony z 3 czasopism: Etnografia Polska, Lud i Konteksty - Polska Sztuka Ludowa, rozwinięty o kolekcje: tomów pokonferencyjnych, prac zbiorowych, tekstów opisujących dziedzictwo kulturowe, podręczników akademickich, tekstów PIA. Możliwość indeksowania treści zawartych w BC PIA, niezbędna w działaniach związanych z parametryzacją nauki, pozwala naukowcom publikującym w czasopismach zamieszczonych w BC PIA rejestrować cytowania oraz uzyskiwać korzyści z opublikowania własnych tekstów w internecie.

BC PIA jako jedna z niewielu polskich bibliotek cyfrowych udostępnia współczesne teksty naukowe, wiąże się to z pozyskiwaniem praw autorskich. Do tej pory uzyskaliśmy prawa do opublikowania 2774 tekstów od 428 autorów. Do opublikowania pozostało 2259 tekstów.

Skrócony opis części informatycznej:

Tworzenie i obsługa BC PIA - platforma dLibra

OCR i obróbka techniczna skanów - Abby Fine Reader i Acrobat Reader

Automatyczna bibliografia i przypisy - implementacja programu Zotero

Zakładka BC na stronie PIA - drupal

Bazy danych o prawach autorskich w postaci edytowalnej - edytory tekstów i programy do tworzenia arkuszy kalkulacyjnych (MS Office)

Magazynewanie i przekazywanie danych - serwer UW, dyski zewnętrzne, Dropbox

Komunikacja elektroniczna - serwer pocztowy (?)

Opis efektów projektu:

Otwarty dostęp do współczesnych tekstów naukowych z zakresu nauk antropologiczno/etnologicznych; 5,5 miliona odwiedzających od 2010;

Baza bibliograficzna polskiej etnologii i antropologii kulturowej;

Prawa autorskie: podwyższenie świadomości w środowisku nauk antropologiczno/etnologicznych i dyskusja na ten temat;

Upowszechnianie nauk antropologiczno/etnologicznych;

Edukacja w zakresie wiedzy o kulturach regionów Polski, Europy, świata, grup etnicznych, religijnych, mniejszości, grup marginalnych itd. w Polsce i zagranicą;

Publikacje na temat BC PIA popularyzujące ideę tworzenia środowiskowej biblioteki cyfrowej (wybór) :

Joanna Koźmińska, Cyfrowa Etnografia: o projekcie antropologicznej biblioteki cyfrowej, w: Homo interneticus: etnograficzne wędrówki w głąb sieci,

pod red. Ewy Jagiełło i Pawła Schmidta, Lublin, Wydawnictwo Portalu Wiedza i Edukacja, 2010, s.113-119 (e-book)

<http://www.cyfrowaetnografia.pl/dlibra/doccontent?id=4664&dirids=1>

Joanna Koźmińska, Niematerialne dziedzictwo kulturowe na platformie cyfrowej Polskiego Instytutu Antropologii, w: Niematerialne dziedzictwo

kulturowe: źródła- wartości-ochrona, red. Jan Adamowski i Katarzyna Smyk, Lublin, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2013 s.

289-299

Powiązania z innymi polskimi i zagranicznymi projektami:

uczestnictwo w Federacji Bibliotek Cyfrowych;

Polskie Towarzystwo Ludoznawcze - udział w projekcie kontynuacji wydawnictwa serii Atlas Polskich Strojów Ludowych - cyfryzacja i publikacja atlasu online;

Uniwersytet Śląski : uczestnictwo w granicie NPRH Polski Atlas Etnograficzny – opracowanie naukowe, elektroniczny katalog

danych, publikacja zasobów w sieci Internet;

Pracownia Etnograficzna im. W. Dynowskiego: projekty skanseny.net, strojeludowe.net - udział w tworzeniu czytelnicy online;

Muzeum Etnograficzne w Krakowie - drzeworyty.eu - pomoc w tworzeniu czytelnicy online;

Barbarzyńca - czasopismo antropologiczne - publikacja online

Kurpiowski Park Kultury - umieszczanie regionalnych publikacji online

Archiwum Cyfrowe im. Józefa Burszty - współpraca

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

www.cyfrowaetnografia.pl

www.pia.org.pl

oraz

<http://cyfrowaetnografia.pl/dlibra/doccontent?id=4664&dirids=1>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Aktualizacja oprogramowania i migracja systemu dLibra do najnowszej postaci (z 4.0 do 5.6)

W latach 2014-2015 planujemy aktualizację i rozwój bazy czasopism, rozwój kolekcji książkowych i kolekcji atlasów etnograficznych, a także stworzenie kolekcji klasyków antropologii polskiej i obcej.

W najbliższym czasie planujemy wprowadzenie do kolekcji 2 czasopism najmłodszego pokolenia etnologów: (op.cit...) - maszyna interpretacyjna, <http://opcit.pl/>. (wydawany nieregularnie w Warszawie przez absolwentów i doktorantów UW) i Barbarzyńca (periodyk wydawany w Krakowie przez studentów i absolwentów etnologii UJ).

Budowa kolekcji kanonu dzieł antropologicznych stała się koniecznością. Dotyczy to zarówno polskich utworów (S. Ciszewski, J. S. Bystroń, C. Baudouin de Courtenay Ehrenkreutz Jędrzejewiczowa, S. Poniatowski), jak i tłumaczeń podstawowych dzieł nauk antropologicznych (E. B. Tylor, L. S. Morgan, H. Spencer).

Obok kolekcji czasopism i książek budowane jest repozytorium tekstów naukowych. Zamierzamy także wprowadzić szerszą ofertę artykułów w języku angielskim - przetłumaczone prace polskich autorów, w tym angielską wersję antologii tekstów z czasopisma Konteksty-PSL.

Ważnym dla środowiska działaniem, jest cyfryzacja Atlasu Polskich Strojów Ludowych i umieszczenie wszystkich zeszytów APSL w BC PIA. Do tej pory brak atlasów etnograficznych w polskich zasobach cyfrowych. W planach jest też cyfryzacja Polskiego Atlasu Etnograficznego we współpracy z Uniwersytetem Śląskim.

Planowane są działania zmierzające do poszerzenia oferty BC PIA o materiały audiowizualne (projekt historii mówionej etnografii warszawskiej, filmy i nagrania z projektu SAiD) oraz utworzenia platformy Cyfrowa Etnografia skupiającej cyfrowe projekty z zakresu etnologii i antropologii.

Po uzyskaniu dofinansowania projektujemy dostosowanie BC PIA do możliwości korzystania najnowszych rozwiązań technologicznych - czytników, tabletów etc oraz zintegrowanie BC PIA z oprogramowaniem GIS (zwłaszcza atlasy etnograficzne).

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

współpraca z instytutami naukowymi w zakresie tworzenia wspólnego repozytorium naukowego;

współpraca z muzeami etnograficznymi;

Ośrodek Dokumentacji i Informacji Etnograficznej PTL - wspólne olinkowanie Bibliografii Etnografii Polskiej i artykułów w BC PIA;

współpraca z organizacjami promującymi otwarty dostęp do nauki;

Profil potencjalnego partnera:

instytucje kultury i nauki, instytucje edukacyjne, instytucje wspierające rozwój regionalny, instytucje ochrony dziedzictwa kulturowego

otwarta nauka, creative commons, środowiska zajmujące się problematyką praw autorskich i digitalizacji

biblioteki cyfrowe, archiwa naukowe, badacze instytucjonalni i niezależni, studenci i doktoranci, archiwa audiowizualne - cyfrowe i tradycyjne.

DIGITALIZACJA I UDOSTĘPNIENIE W SIECI KOLEKCJI STARODRUKÓW BIBLIOTEKI IH UW

Osoba wypełniająca ankietę: Kierownik projektu

www.staredruki.ihuw.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Fundacja Przyjaciół IH UW Klio

Instytucja przy której projekt jest afiliowany: IH UW ; MHP

Pozostałe instytucje zaangażowane w projekt: Instytut Historyczny Uniwersytetu Warszawskiego

Muzeum Historii Polski

Zespół BAZHUM

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Dominik Purchała, d.purchala@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Zespół BAZHUM

Zespół IH UW

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: -

Źródło finansowania: dotacja celowa, środki własne instytucji

OPIS PROJEKTU

Typ projektu: badawczy, digitalizacyjny

Efekt projektu: oprogramowanie, baza danych, repozytorium, publikacja

Słowa kluczowe: digitalizacja; starodruki; domena publiczna; ochrona kultury; dziedzictwo kulturowe

Skrócony opis projektu:

Głównym celem projektu „Digitalizacja i rozbudowa cyfrowej kolekcji starodruków Biblioteki Instytutu Historycznego UW” realizowanego w ramach programu Dziedzictwo kulturowe - Ochrona i cyfryzacja dziedzictwa kulturowego była digitalizacja części kolekcji starodruków należącej do Instytutu Historycznego UW. Z powodu złego stanu zachowania i przedstawianej sobą wartości kolekcja oddana została do przechowania w skarbcu Biblioteki Uniwersytetu Warszawskiego. Tym samym stała się praktycznie niedostępna

dla czytelników. Główną ideą, która przyświecała nam podczas realizacji zadania była chęć przywrócenia i udostępnienia tekstów wszystkim zainteresowanym, przy jednoczesnym utrwaleniu i zabezpieczeniu samych tekstów przed bezpowrotnym zniszczeniem.

Skrócony opis części informatycznej:

Stworzony w jego ramach system do prezentacji dostępny jest na wolnej licencji GNU. System jest możliwy do pobrania i wykorzystania do podobnych projektów (<https://github.com/bazhum/starodruki>)

Metadane opisowe powstały w oparciu o rozbudowane pola formatu MARC-21.

Opis efektów projektu:

Kolekcja starych druków licząca 1003 dzieła w 1105 woluminach z XVI-XVIII w. stanowi najcenniejszą część zbiorów BIH. Obok unikatowych egzemplarzy dzieł związanych ściśle z dziejami dawnej Rzeczypospolitej i jej wybitnych osobistości, takich jak król Stanisław August Poniatowski i Jan Heweliusz, zawiera w znacznej mierze cenne pomoranika i prussica pochodzące z dawnych historycznych kolekcji m.in. dawnej Biblioteki Senatu Gdańskiego i dawnej biblioteki zamku w Malborku. Oprócz pomników głównie dawnej historiografii i prawoznawstwa obejmuje także szereg rzadkich druków kartograficznych: atlasów, map i planów.

Do tej pory zdigitalizowano i udostępniono w serwisie 609 druków co przekłada się na ok. 145512 skanów.

W ramach projektu zostało wytworzone i opublikowane na licencji GNU oprogramowanie do prezentacji skanów.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

www.staredruki.ihuw.pl

<https://github.com/bazhum/starodruki>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Dalsze plany zakładają rozbudowę kolekcji oraz wprowadzenie dodatkowych funkcjonalności do serwisu prezentacyjnego. Chcemy też rozbudować opisy i streszczenia prezentowanych dzieł. Zakładamy powstanie zespołu badawczego mającego na celu pracę nad treścią i standardami digitalizacji i udostępniania starodruków w sieci oraz opracowaniem merytorycznym starodruków, które nie zostały jeszcze udostępnione w internecie, a pozostają w zbiorach polskich instytucji kultury i nauki.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Instytucje naukowe, zespoły badawcze, archiwa, biblioteki naukowe, muzea.

DIGITALIZACJA WYDAWNICTW STOWARZYSZENIA BIBLIOTEKARZY POLSKICH

Osoba wypełniająca ankietę: Grzegorz Gmiterek

<http://bbc.uw.edu.pl/dlibra>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Stowarzyszenie Bibliotekarzy Polskich; IINiSB UW

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Zdzisław Dobrowolski, z.dobrowolski@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 8/15/2011 - 12/15/2011

Źródło finansowania: program Dziedzictwo Cyfrowe Narodowego Instytutu Audiowizualnego

OPIS PROJEKTU

Typ projektu: digitalizacyjny

Efekt projektu: repozytorium

Słowa kluczowe: biblioteka cyfrowa, SBP, digitalizacja, Przegląd Biblioteczny, Zagadnienia Informacji Naukowej

Skrócony opis projektu:

W roku 2011 została zdigitalizowana i opublikowana na stronach Bibliologicznej Biblioteki Cyfrowej seria Nauka, Dydaktyka, Praktyka za lata 1993-2007 (96 tytułów). Wszystkie zdigitalizowane tytuły są w wolnym dostępie w sieci pod adresem www.bbc.uw.edu.pl. Projekt objął również kontynuację digitalizacji Przeglądu Bibliotecznego i Zagadnień Informacji Naukowej za lata 2007-2008 (12 zeszytów).

Skrócony opis części informatycznej:

j.w.

Opis efektów projektu:

j.w.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

DYGITALIZACJA SŁOWNIKA LINDEGO

Osoba wypełniająca ankietę: Joanna Bilińska, j.bilinska@uw.edu.pl

<http://korpusy.klf.uw.edu.pl/pl/sownik-lindego/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Katedra Lingwistyki Formalnej

Wydział Neofilologii

Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Janusz S. Bień, jsbien@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 2/1/2009 -

Źródło finansowania: środki własne instytucji, początkowo finansowany przez grant MNiSZW (do 12.05.2012)

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy, digitalizacyjny

Efekt projektu: baza danych, repozytorium, publikacja

Słowa kluczowe: słownik Lindego, dygitalizacja, Samuel Bogumił Linde, korpus, OCR

Skrócony opis projektu:

W formie tzw. korpusów dwuwarstwowych udostępniane są 4 słowniki i

jeden korpus dawnych tekstów polskich.

Skrócony opis części informatycznej:

Aktualnie wykorzystywane jest oprogramowanie przygotowane w ramach opisanego w osobnej ankiecie grantu "Narzędzia dygitalizacji tekstów na potrzeby badań filologicznych" (13.05.2009 - 12.05.2012).

Opis efektów projektu:

Wstępna dygitalizacja drugiego wydania słownika Lindego udostępniona jako korpus dwuwarstwowy wykorzystujący tzw. brudny OCR.

Szczegółowa koncepcja projektu została przedstawiona we wniosku na konkurs NPRH, który niestety nie otrzymał finansowania.

Planowane jest m.in. uzyskanie wysokiej jakości OCR-u, co ze względu na wielojęzyczność słownika wymaga zastosowania nowatorskich metod.

Powiązania z innymi polskimi i zagranicznymi projektami:

Prace nad dygitalizacją słowników historycznych były prowadzone w ramach projektu SYNAT (<https://github.com/kawu/hist-pl>), ale ze względu na odmienną metodologię nie są one przydatne dla opisywanego projektu.

Wykonawcy projektu uczestniczą w przedsięwzięciu:

ISCH COST Action IS1305: Individuals, Societies, Cultures and Health European Cooperation in Science and Technology European Network of e-Lexicography (<http://www.elexicography.eu/>)

Working Group 2:

Retro-digitized dictionaries:

Do weryfikacji wyników OCR planuje się w miarę możliwości wykorzystywać wstępne wyniki projektu "Automatyczna analiza fleksyjna polszczyzny XIX wieku" (<http://www.f19.uw.edu.pl>) realizowanego w latach 2013-2016.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Bień, Janusz S. and Bilińska, Joanna A. (2012) Słownik Lindego jako

korpus. In: Bibliotheca Lindiana. Samuel Bogumił Linde (1771 – 1847), pierwszy dyrektor Biblioteki Uniwersyteckiej w Warszawie. W 165 rocznicę śmierci., 19 – 20 listopada 2012 roku, Warszawa. <http://bc.klf.uw.edu.pl/324/>.

Bilińska, Joanna A. (2012) Составление перечня сокращенных названий языков в рамках проекта дигитализации «Словаря польского языка» С.Б.Линде. In: Информационные технологии и письменное наследие. El'Manuscript-2012. IV научная конференция. Петрозаводский государственный университет, Ижевский государственный технический университет, Petrozawodsk, Iżewsk, pp. 34-43. ISBN 978-5-8021-1402-5 <http://bc.klf.uw.edu.pl/301/>.

Bilińska, Joanna A. (2011) Describing Linde's Dictionary of Polish for Digitalisation Purposes. In: Electronic lexicography in the 21st century: new applications for new users (eLex 2011). Trojina, Institute for Applied Slovene Studies, pp. 43-51. ISBN 978-961-92983-3-6 <http://bc.klf.uw.edu.pl/240/>.

Bilińska, Joanna A. (2013) Analiza i leksykograficzny opis struktury słownika Lindego na potrzeby dygitalizacji. Rozprawa doktorska, Uniwersytet Warszawski. <http://bc.klf.uw.edu.pl/347/>.

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

DZIEDZICTWO NARODOWE I KULTUROWE, POLSKIE ŚLADY ZA GRANICĄ.

Osoba wypełniająca ankietę: Mikołaj Ochmański

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Informacji Naukowej i Studiów Bibliologicznych UW

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Ministerstwo Spraw Zagranicznych

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Dariusz Kuźmina

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

4. Koordynator projektu: Dariusz Kuźmina UW; wykonawcy: , Anna Kamler UW; Robert Brzóska UW.

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 5/1/2012 - 12/31/2012

Źródło finansowania: dotacja celowa

OPIS PROJEKTU

Typ projektu: badawczy, digitalizacyjny

Efekt projektu: repozytorium

Słowa kluczowe: digitalizacja; polonia; biblioteka cyfrowa; dziedzictwo kulturowe

Skrócony opis projektu:

Polonijna Biblioteka Cyfrowa zamieszcza cyfrowe kopie unikalnych poloników z wybranych kolekcji zagranicznych. Polonijna Biblioteka Cyfrowa została utworzona przez Instytut Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego i jest współfinansowana ze środków finansowych otrzymanych od Ministerstwa Spraw Zagranicznych w ramach konkursu na realizację zadania "Współpraca z Polonią i

Polakami za granicą". W badaniach i digitalizacji brali udział przedstawiciele kilku ośrodków naukowych z Polski i Europy. Projekt był realizowany na terenie: Anglii, Francji, Peru i USA.

Skrócony opis części informatycznej:

W ramach Projektu powstała Polonijna Biblioteka Cyfrowa dostępna: <http://www.pbc.uw.edu.pl/>

Opis efektów projektu:

W efekcie realizacji projektu powstało repozytorium cyfrowe dostępne online, w którym zgromadzono kopie dokumentów rejestrujących życie i kulturę Polaków żyjących poza granicami kraju. Na dzień dzisiejszy (czerwiec 2014) Polonijna Biblioteka Cyfrowa zawiera skany publikacji polonijnych wydawanych w USA, Wielkiej Brytanii, Rosji i w Związku Radzieckim oraz w byłych republikach ZSRR, w Niemczech oraz w Ameryce Południowej.

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekt był kontynuowany w następnych latach.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Projekt jest kontynuowany, Polonijna Biblioteka Cyfrowa jest uzupełniana i rozwijana na bieżąco.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

DZIEDZICTWO NARODOWE, BADANIA I UPOWSZECHNIANIE WIEDZY O TRADYCJI I HISTORII

Osoba wypełniająca ankietę: Mikołaj Ochmański

<http://www.pbc.uw.edu.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Informacji Naukowej i Studiów Bibliologicznych UW

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Ministerstwo Spraw Zagranicznych

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Dariusz Kuźmina

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Koordynator projektu – Dariusz Kuźmina UW;

wykonawcy - Barbara Zybert UW; Elżbieta Maruszak UW; Robert Brzóska UW.

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 3/1/2014 - 12/31/2014

Źródło finansowania: dotacja celowa 37/2014

OPIS PROJEKTU

Typ projektu: badawczy, digitalizacyjny

Efekt projektu: repozytorium

Słowa kluczowe: digitalizacja; polonia; biblioteka cyfrowa; dziedzictwo narodowe

Skrócony opis projektu:

Projekt realizowany w celu rozbudowania Polonijnej Biblioteki Cyfrowej. W badaniach i digitalizacji brali udział przedstawiciele kilku ośrodków naukowych z Polski i Europy. Projekt będzie realizowany na terenie: Brazylii, Chile, Izraelu, Francji i Kazachstanie.

Skrócony opis części informatycznej:

W ramach Projektu powstała Polonijna Biblioteka Cyfrowa dostępna: <http://www.pbc.uw.edu.pl/>

Opis efektów projektu:

W efekcie realizacji projektu powstało repozytorium cyfrowe dostępne online, w którym zgromadzono kopie dokumentów rejestrujących życie i kulturę Polaków żyjących poza granicami kraju. Na dzień dzisiejszy (czerwiec 2014) Polonijna Biblioteka Cyfrowa zawiera skany publikacji polonijnych wydawanych w USA, Wielkiej Brytanii, Rosji i w Związku Radzieckim oraz w byłych republikach ZSRR, w Niemczech oraz w Ameryce Południowej.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

E-EDUKACJA BEZ BARIER CYFROWYCH I SPOŁECZNYCH

Osoba wypełniająca ankietę: Kierownik projektu

przewidywany

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Instytut Języków Romańskich i Translatoryki, Wydział Filologiczny, Uniwersytet Śląski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Fundacja Widzialni

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: dr Izabela Mrochen, izabela.mrochen@us.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr Izabela Mrochen, UŚ,

Wojciech Kulesza, Fundacja Widzialni

INFORMACJE ORGANIZACYJNE

Status projektu: Planowany

Okres realizacji projektu: 10/1/2015 - 12/30/2018

Źródło finansowania: Narodowe Centrum Nauki

OPIS PROJEKTU

Typ projektu: edukacyjny

Efekt projektu: narzędzie badawcze, baza danych, repozytorium, kurs

Słowa kluczowe: włączenie cyfrowe; zasady WCAG; opis alternatywny; audio deskrypcja; technologie asystujące

Skrócony opis projektu:

Polska jako członek Unii Europejskiej zobowiązana jest do wspierania kompetencji cyfrowych, do rozwijania e-usług i włączenia cyfrowego osób niepełnosprawnych. Celem projektu jest stworzenie platformy edukacyjnej, której główne treści będą obejmowały podstawowe zasady przygotowania stron internetowych, szczególnie instytucji o charakterze edukacyjnym oraz administracji publicznej do potrzeb osób z niepełnosprawnościami, ze szczególnym uwzględnieniem osób z dysfunkcją wzroku. Przygotowanie repozytorium z bazą narzędzi badawczych oraz aplikacji, które będą dostępne i wykorzystane w poprawnym przygotowaniu stron internetowych, e-podręczników, materiałów multimedialnych, itp. zgodnie z międzynarodowymi standardami

dostępności WCAG (ang. Web Content Accessibility Guidelines). Informacje zamieszczone na platformie, w ramach projektu E-edukacja bez barier cyfrowych i społecznych umożliwią otwarty dostęp do bazy danych.

Skrócony opis części informatycznej:

W celu udostępniania treści projektu, konieczne jest przygotowanie platformy edukacyjnej, która ułatwi otwarty dostęp do przykładów wdrażania dobrych praktyk w ramach włączenia cyfrowego i społecznego osób zagrożonych wykluczeniem.

Opis efektów projektu:

Założenie projektu to przede wszystkim udostępnienie zdigitalizowanej bazy danych, przykładów dobrze przygotowanych materiałów edukacyjnych w formie cyfrowej. Ponadto stworzenie otwartego i ogólnodostępnego repozytorium z ciekawymi publikacjami, odnośnikami do ważnych organizacji, fundacji, itp. które mają zasięg ogólnonarodowy i międzynarodowy. Efekty projektu to głównie wpieranie instytucji administracji publicznej, uczelni wyższych, placówek oświatowych, pracowników, wykładowców i nauczycieli, którzy są zobowiązani do przygotowania materiałów informacyjnych w postaci zdigitalizowanej zgodnie z wymogami WCAG 2.0 i założeniami Ministerstwa Administracji i Cyfryzacji.

Powiązania z innymi polskimi i zagranicznymi projektami:

W ramach projektu E-edukacja bez barier cyfrowych i społecznych planowana jest współpraca merytoryczna z organizacją AbilityNet oraz Fundacją Widzialni

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

w przygotowaniu

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

W związku z ciągłym rozwojem nowych technologii i potrzebą nabywania nowych kompetencji potrzebnych na rynku pracy, projekt będzie punktem wyjścia do ciągłego doskonalenia prezentacji tematyki o charakterze edukacyjnym zgodnie z międzynarodowymi standardami dostępności. Wizja tzw 'trójkątów wiedzy', czyli współpracy między uczelniami wyższymi, otoczeniem i gospodarką, wymusza na instytucjach publicznych przekazywanie informacji w odpowiednio przygotowany sposób, ze szczególnym uwzględnieniem grup zagrożonych wykluczeniem cyfrowym i społecznym. Ogólnonarodowy charakter projektu przyczyni się do polepszenia stanu wiedzy wśród osób z dysfunkcjami, do większego ich zaangażowania w pracę zawodową a przede wszystkim do realizacji celów polityki społecznej w nadchodzących latach.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

W ramach przygotowania treści merytorycznych, planowane jest nawiązanie współpracy z organizacją AbilityNet, Audio Description Associates, LLC (Joel Snyder) w USA, the Royal National Institute of Blind People w Londynie, Fundacją Widzialni, i szereg innych, które w swoich celach wspierają grupy osób zagrożonych wykluczeniem cyfrowym i społecznym.

Profil potencjalnego partnera:

Organizacje pozarządowe, fundacje, stowarzyszenia

ELEKTRONICZNY KORPUS TEKSTÓW POLSKICH Z XVII I XVIII W. (DO ROKU 1772)

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://clip.ipipan.waw.pl/KORBA>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Pracownia Historii Języka Polskiego XVII i XVIII wieku, Instytut Języka Polskiego PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Zespół Inżynierii Lingwistycznej, Instytut Podstaw Informatyki PAN

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Włodzimierz Gruszczyński

Osoba kontaktowa: Maciej Ogrodniczuk, maciej.ogrodniczuk@gmail.com

Pozostałe osoby realizujące projekt:

Włodzimierz Gruszczyński, IJP PAN; Małgorzata Majewska, IJP PAN; Magdalena Majdak, IJP PAN; Maciej Ogrodniczuk, IPI PAN; Renata Bronikowska, IJP PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 1/1/2013 - 1/1/2018

Źródło finansowania: Narodowy Program Rozwoju Humanistyki 0036/NPRH2/H11/81/2012

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, oprogramowanie, korpus tekstów średniopolskich

Słowa kluczowe: korpus tekstów średniopolskich; korpus tekstów dawnych; anotacja tekstów dawnych

Skrócony opis projektu:

Celem projektu jest zbudowanie i udostępnienie korpusu tekstów polskich z okresu XVII i XVIII wieku (do roku 1772 r.) w określonym formacie, zgromadzonych na nośniku elektronicznym wraz z oprogramowaniem umożliwiającym przeszukiwanie tego zbioru, tworzenie konkordancji żądanych form, tworzenie indeksów występujących w nim słów oraz otrzymywanie informacji o frekwencji słów tekstowych i leksemów, a także o dokładnej lokalizacji poszczególnych słów (rozumianej jako lokalizacja w tekście oryginalnym, z dokładnością do strony).

Skrócony opis części informatycznej:

Informatyczna część projektu jest związana z obsługą korpusu i procesu anotacji oraz rozwojem narzędzi informatycznych do przetwarzania tekstów średniopolskich – tagera morfoskładniowego, lematyzatora, wyszukiwarki uwzględniającej różnicę między językiem współczesnym a dawnym itp.

Opis efektów projektu:

Wynikiem projektu będzie systematyczne przebadanie słownictwa i gramatyki języka polskiego z XVII i XVIII w., stworzenie zasobu anotowanego strukturalnie, pod względem cech autorskich, stylistyczno-genologicznych, edytorskich, językowych i typograficznych oraz stworzenie narzędzi do lematyzacji (co umożliwi wyszukiwanie nie tylko form, ale wszystkich wystąpień form danego leksemu) i udostępnienie korpusu w Internecie na zasadzie wolnego dostępu.

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekty powiązane: Korpus polskiej łaciny średniowiecznej, Elektroniczny korpus tekstów staropolskich (do 1500 roku), Korpus polszczyzny XVI wieku, Korpus polszczyzny barokowej, Automatyczna analiza fleksyjna tekstów polskich z lat 1830–1918 z uwzględnieniem zmian w odmianie i pisowni 2013–2016, Korpus Gwar Polskich, Automatyczna lematyzacja tekstu gwarowego w zapisie fonetycznym – program FonOrt.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://clip.ipipan.waw.pl/KORBA>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Opracowanie narzędzi informatycznych do szerszej analizy polszczyzny XVII- i XVIII-wiecznej (identyfikacja grup składniowych, jednostek nazewniczych, analiza semantyczna itp.)

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Jak w pkt 21.

Profil potencjalnego partnera:

Grupy badawcze zajmujące się polszczyzną tego okresu.

FEDERACJA BIBLIOTEK CYFROWYCH

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://fbc.pionier.net.pl>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Poznańskie Centrum Superkomputerowo-Sieciowe

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Kilkaset instytucji kultury z całej Polski współtworzących kilkadziesiąt bibliotek cyfrowych - więcej informacji: <http://fbc.pionier.net.pl/owoc/list-libs>

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: dr inż. Maciej Stroiński, stroins@man.poznan.pl

Osoba kontaktowa: Marcin Werla, mwerla@man.poznan.pl

Pozostałe osoby realizujące projekt:

dr inż. Maciej Stroiński, stroins@man.poznan.pl; dr inż. Cezary Mazurek, mazurek@man.poznan.pl; mgr inż. Marcin Werla, mwerla@man.poznan.pl; mgr inż. Aleksandra Nowak, anowak@man.poznan.pl

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 6/22/2007 -

Źródło finansowania: środki własne instytucji

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy

Efekt projektu: oprogramowanie, baza danych

Słowa kluczowe: biblioteki cyfrowe, metadane, Europeana, agregacja danych, digitalizacja

Skrócony opis projektu:

Federacja Bibliotek Cyfrowych (FBC) jest kolejnym etapem budowy sieci rozproszonych bibliotek i repozytoriów cyfrowych w Polsce. Nazwa serwisu odzwierciedla jego charakter – serwis FBC jest zbiorem zaawansowanych usług sieciowych opartych na zasobach cyfrowych dostępnych w polskich bibliotekach i repozytoriach cyfrowych uruchomionych w sieci PIONIER. Zasoby te współtworzone są przez wiele instytucji naukowych i publicznych, takich jak wyższe uczelnie, biblioteki, archiwa, muzea czy ośrodki badawcze. Federacja Bibliotek Cyfrowych prowadzona jest przez Poznańskie Centrum Superkomputerowo-Sieciowe afiliowane przy Instytucji Chemii Bioorganicznej PAN, a za rozwój FBC odpowiada Zespół Bibliotek Cyfrowych PCSS.

Misją Federacji Bibliotek Cyfrowych jest:

- ułatwienie wykorzystania zasobów polskich bibliotek cyfrowych i repozytoriów,
- zwiększenie widoczności i popularyzacja zasobów polskich bibliotek cyfrowych i repozytoriów w internecie,
- udostępnienie użytkownikom internetu i twórcom bibliotek cyfrowych nowych, zaawansowanych usług sieciowych opartych na zasobach polskich bibliotek cyfrowych i repozytoriów.

Misja ta realizowana jest poprzez tworzenie i udostępnianie użytkownikom i dostawcom danych FBC kolejnych funkcji.

Skrócony opis części informatycznej:

Podstawą części informatycznej projektu jest portal internetowy dostępny pod adresem <http://fbc.pionier.net.pl/>, który umożliwia przeszukiwanie dostępnych on-line zbiorów kilkuset polskich instytucji kultury. Dane do portalu agregowane i przetwarzane są za pomocą systemu Clepsydra (<http://fbc.pionier.net.pl/pro/clepsydra>). FBC jest również głównym dostawcą danych z polskich instytucji kultury do Europeany.

Opis efektów projektu:

Funkcje serwisu FBC bazują na opisach (metadanych) publikacji cyfrowych dostępnych w bibliotekach cyfrowych i repozytoriach współpracujących z FBC. Informacje te aktualizowane są obecnie w serwisie FBC raz na dobę – w nocy. Technicznie wymiana danych odbywa się poprzez otwarty protokół OAI-PMH. Dzięki temu serwis FBC może współpracować z dowolnymi innymi systemami – jedynym warunkiem jest wsparcie dla wspomnianego protokołu.

Portal FBC daje użytkownikom możliwość przeszukiwania opisów cyfrowych publikacji dostępnych w polskich bibliotekach cyfrowych i repozytoriach współpracujących z FBC. Każdy z takich opisów skojarzony jest z odnośnikiem prowadzącym do strony WWW danej biblioteki cyfrowej, na której wyszukany obiekt jest prezentowany użytkownikowi. Pełną listę naszych dostawców można znaleźć w zakładce „Biblioteki cyfrowe,..” Możliwe jest przeszukiwanie całego opisu bądź też jego wybranego elementu (np. pola tytuł czy autor). Zestaw elementów zgodny jest ze schematem Dublin Core. Ekran prezentacji wyników wyszukiwania umożliwia zawężanie wyników do danych z jednej lub kilku bibliotek cyfrowych. Z wyników wyszukiwania użytkownik przekierowywany jest bezpośrednio do odpowiednich stron serwisu, z którego pochodzi przechowywany obiekt.

Baza Bibliotek Cyfrowych pozwala na zapoznanie się z informacjami na temat dostawców danych Federacji Bibliotek Cyfrowych. Dla każdej z bibliotek cyfrowych dostępne są:

Informacje ogólne, takiej jak adres WWW, e-mail czy opis, lista instytucji współtworzących bibliotekę cyfrową, zbiór kilku polecanych obiektów z tej biblioteki, informacje techniczne.

Powiązania z innymi polskimi i zagranicznymi projektami:

FBC ściśle współpracuje ze swoim europejskim odpowiednikiem czyli portalem Europeana.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Serwis informacyjny na temat FBC: <http://fbc.pionier.net.pl/pro/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

- Możliwość przeszukiwania pełnotekstowego zbiorów polskich bibliotek cyfrowych
- Zwiększanie ilości zaindeksowanych w bazie bibliotek cyfrowych
- Integracja z nowymi narzędziami informatycznymi dla cyfrowej humanistyki

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Dalsza współpraca z polskimi projektami może przebiegać dwutorowo: z jednej strony w indeks wyszukiwawczy FBC mogą być włączana kolejne biblioteki cyfrowe i repozytoria; z drugiej zaś dane zgromadzone w FBC mogą być udostępniane w sposób maksymalizujący możliwość ich automatycznego wykorzystania w powstających narzędziach badawczych cyfrowej humanistyki.

Dalsza współpraca z projektami zagranicznymi będzie oparta przede wszystkim na wynikach projektu Europeana Cloud, którego celem jest wytworzenie jednej współdzielonej infrastruktury dla gromadzenia i udostępniania danych o dziedzictwie kulturowym Europy. PCSS jest jednym z głównych partnerów technicznych tego projektu.

Profil potencjalnego partnera:

- operatorzy bibliotek cyfrowych i repozytoriów
- twórcy narzędzi informatycznych zainteresowani automatycznym dostępem do danych zebranych w FBC

GETCLASS - SYSTEM DO ANOTACJI SEMANTYCZNEJ TEKSTU

Osoba wypełniająca ankietę: Kierownik projektu

www.clarin-pl.eu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Centrum Technologii Językowych CLARIN-PL, Instytut Informatyki Politechnika Wroclawska,

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr inż. Maciej Piasecki (PWr.), mgr inż. Błażej Ryśnik, mgr inż. Marcin Pol, mgr inż. Paweł Kędzia

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 11/2/2013 - 12/31/2014

Źródło finansowania: dotacja celowa CLARIN-PL

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: automatyczna analiza treści materiałów źródłowych; klasyfikacja semantyczno-pragmatyczna tekstu; analiza korpusu tekstów; tagowanie semantyczne tekstu

Skrócony opis projektu:

Celem projektu jest konstrukcja systemu wspomagającego ilościową analizę tekstowych materiałów źródłowych na poziomie semantycznym i pragmatycznym. Zakładamy, że analiza materiału jest realizowana poprzez przypisanie zdefiniowanych uprzednio kategorii semantyczno-pragmatycznych do poszczególnych tekstów (zapisów) lub ich fragmentów.

Skrócony opis części informatycznej:

System opiera się na automatycznej analizie lingwistycznej materiałów źródłowych i maszynowym uczeniu się na podstawie ręcznie oznaczonych przykładów kategoryzacji. System wspomaga również użytkownika w

wyborze próbek do ręcznej anotacji w oparciu o metody automatycznej analizy podobieństwa pomiędzy próbkami tekstu, tak aby dobrać automatycznie reprezentatywne próbki do anotacji ręcznej. Opcjonalnie, w procesie wstępnym, możliwe jest pobranie zawartości stron internetowych z zadanego wykazu.

Opis efektów projektu:

Zbudowany zostanie system wspomagający ilościową analizę tekstowych materiałów źródłowych na poziomie semantycznym i pragmatycznym. Zakładamy, że analiza materiału jest realizowana poprzez przypisanie zdefiniowanych uprzednio kategorii semantyczno-pragmatycznych do poszczególnych tekstów (zapisów) lub ich fragmentów. Tekst może mieć dowolną postać, nie muszą być w nim wyodrębnione części składowe. Kategorie semantyczno-pragmatyczne są definiowane przez badacza używającego systemu poprzez podanie przykładów wyodrębnionych z analizowanego materiału źródłowego. System uczy się na podstawie przykładów i dokonuje automatycznej analizy całości materiału. System ma zapewnić również różnego rodzaju metody analizy statystycznej wyników klasyfikacji.

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekt jest realizowany w ramach projektu CLARIN-PL we współpracy z użytkownikami infrastruktury CLARIN.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

www.clarin-pl.eu

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

System może zostać rozszerzony na analizę transkrybowanych nagrań mowy oraz analizę tekstów nieformalnych, np. blogów. W przypadku tych ostatnich konieczne będzie wprowadzenie etapu wstępnej automatycznej korekty tekstu. System może też zostać poszerzony o moduł wydobywania informacji jako dodatkowe źródło informacji do końcowej, statystycznej analizy materiału źródłowego. System może też zapewnić w pełni automatyczną analizę materiału opartą na automatycznym grupowaniu próbek tekstu, które są bliskie sobie pod względem treści.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Inne projekty w ramach CLARIN Europa, inne projekty z dziedziny nauk humanistycznych, system może być dostosowany do innych języków niż polski.

Profil potencjalnego partnera:

Wszystkie grupy badawcze, które potrzebują narzędzia wspomagające analizę bardzo dużej ilości tekstowych materiałów źródłowych.

IMPACT-UWAR

Osoba wypełniająca ankietę: Joanna Bilińska, j.bilinska@uw.edu.pl

<http://www.impact-project.eu>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół pracowników i współpracowników Katedry Lingwistyki Formalnej (Wydział Neofilologii, Uniwersytet Warszawski) powołany specjalnie do realizacji zadań projektu

Institucja przy której projekt jest afiliowany: Uniwersytet Warszawski

Pozostałe instytucje zaangażowane w projekt: Konsorcjum projektu IMPACT (www.impact-project.eu)

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Janusz S. Bień, jsbien@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 2/1/2010 - 12/31/2011

Źródło finansowania: 7. program ramowy

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy, digitalizacyjny

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja

Słowa kluczowe: polszczyzna, dygitalizacja, ground-truth, OCR, słowniki

Skrócony opis projektu:

Zespół brał udział w drugim etapie projektu w pracach podprojektu SP3 Enhancement & Enrichment (SP-EE) – Budowa słowników i leksykonów dla tekstów historycznych oraz innych narzędzi wspomagających automatyczne rozpoznawanie tekstów historycznych i ich udostępnianie w ramach realizacji następujących zadań:

EE2 – Lexicon structures and tools (Struktury leksykonów i narzędzia do ich obsługi)

EE3 – Lexicon content (Treści leksykonów),

których wynik (deliverable) stanowi DEE3.13 Proof-of-concept historical lexica for French, Spanish, Polish, Bulgarian, Slovene and Czech (Historyczne prototypowe leksykony dla języka francuskiego, hiszpańskiego, polskiego, czeskiego, słoweńskiego i bułgarskiego).

Skrócony opis części informatycznej:

Wykorzystywano przede wszystkim oprogramowanie stworzone przez partnera projektu - Instituut voor Nederlandse Lexicologie. Dodatkowo wykorzystywano również System Analizy Morfologicznej (<https://www.mimuw.edu.pl/~kszafran/publikacje/tr226.pdf>) oraz inne narzędzia własne.

Opis efektów projektu:

Zasadnicze wyniki są przygotowane do udostępnienia, ale wymaga to zgody Instytutu Języka Polskiego PAN, który dostarczył dane wejściowe w postaci bazy danych internetowego Słownika języka polskiego XVII i pierwszej połowy XVIII wieku. Na razie są one dostępne tylko dla członków centrum kompetencji stanowiącego kontynuację projektu IMPACT (<http://www.digitisation.eu>).

Niektóre zasoby i narzędzia są jednak już dostępne publicznie - np. zdigitalizowany i poddany korekcie indeks a tergo do słownika Lindego (<http://wbl.klf.uw.edu.pl/4/>), proste narzędzie do uwspółcześniania pisowni opracowane w ramach projektu (<https://bitbucket.org/jsbien/pol>).

Dodatkowo udostępniono jako korpus dwuwarstwowy zasoby opracowane przez drugiego polskiego partnera projektu - Zespół Bibliotek Cyfrowych Poznańskiego Centrum Superkomputerowo-Sieciowego (<http://korpusy.klf.uw.edu.pl>).

Powiązania z innymi polskimi i zagranicznymi projektami:

Słownik polszczyzny XVII wieku nadal realizowany w Instytucie Języka Polskiego Polskiej Akademii Nauk.

Podobne narzędzie do uwspółcześniania transkrypcji zostało opracowane w ramach projektu SYNAT (<https://github.com/kawu/hist-pl>).

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Bień, Janusz S. (2011) Polish language resources in IMPACT project. In: IMPACT Final Conference 2011. Digitisation & OCR: Better, faster, cheaper, 24-25 October 2011, London: <http://bc.klf.uw.edu.pl/215/>;

B Bień, Janusz S. (2011) Polskie zasoby językowe w projekcie IMPACT. In: Konferencja "Polskie biblioteki cyfrowe", 12 października 2011, Poznań: <http://bc.klf.uw.edu.pl/213/>;

Bień, Janusz S. (2012) Proof of Concept Historical Lexicon for Polish. Project Report. IMPACT. (Unpublished): <http://bc.klf.uw.edu.pl/316/>;

Bień, Janusz S. (2012) Changes to the IMPACT project Polish Ground-Truth texts. Technical Report. Katedra Lingwistyki Formalnej UW. (Unpublished): <http://bc.klf.uw.edu.pl/288/>;

Bień, Janusz S. (2012) Delivering the IMPACT project Polish Ground-Truth texts with Poliqarp for DjVu. Technical Report. Katedra Lingwistyki Formalnej UW. (Unpublished): <http://bc.klf.uw.edu.pl/289/> ;

de Does, Jesse and Depuyd, Katrien and Schulz, Klaus and Gotscharek, Annette and Ringlstetter, Christoph and Bień, Janusz S. and Erjavec, Tomaž and Kučera, Karel and Martinez, Isabel and Mihov, Stoyan and Souvay, Gilles (2012) Cross-language Perspective on Lexicon Building and Deployment in IMPACT. Project Report. IMPACT: <http://bc.klf.uw.edu.pl/280/>;

Kresa, Monika and Szafran, Krzysztof (2011) Nowe zastosowania słowników historycznych. In: Konferencja "Polskie biblioteki cyfrowe", 12 października 2011, Poznań: <http://bc.klf.uw.edu.pl/211/>;

Kresa, Monika and Szafran, Krzysztof (2011) Nowe zastosowania słowników historycznych. Prace Filologiczne (LXIV). pp. 159-171. ISSN 0138-0567: <http://bc.klf.uw.edu.pl/284/>.

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Projekt jako taki uważamy za definitywnie zakończony, natomiast niektóre z opracowanych narzędzi i zasobów można rozwijać w ramach innych projektów / przedsięwzięć.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

INFOREX

Osoba wypełniająca ankietę: Michał Marcińczuk, michal.marcinczuk@pwr.wroc.pl

<http://nlp.pwr.wroc.pl/infores>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Grupa Naukowa G4.19, Instytut Informatyki Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Projekt jest realizowany wyłącznie na Politechnice Wrocławskiej

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr inż. Michał Marcińczuk, PWr; mgr inż. Jan Kocoń, PWr; mgr inż. Adam Kaczmarek, PWr

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 8/10/2010 -

Źródło finansowania: Narodowe Centrum Nauki, Narodowe Centrum Badań i Rozwoju, Program Operacyjny "Innowacyjna Gospodarka" NEKST POIG, SyNaT projekt strategiczny NCBiR, CLARIN-PL

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych

Słowa kluczowe: korpusy tekstowe, system webowy, zarządzanie korpusami tekstowymi, anotacja

Skrócony opis projektu:

Infores to webowy system do tworzenia i zarządzania korpusami tekstowymi. Wspiera proces konstrukcji korpusów tekstowych na etapie transkrypcji skanów, czyszczenia i selekcji dokumentów oraz opisywania ich na kilku poziomach, m.in. metadanych, anotacji składniowych (frazy składniowe) i semantycznych (nazwy, wyrażenia temporalne, przestrzenne) oraz relacji pomiędzy anotacjami (relacje składniowe, relacje semantyczne).

Skrócony opis części informatycznej:

System umożliwił równoległy i zdalny dostęp do zgromadzonych i opracowanych danych, pozwala na śledzenie i kontrolowanie realizowanych prac oraz współdzielenie zasobów między użytkownikami. Infores został

zrealizowany jako strona internetowa, dzięki czemu nie wymaga instalacji dodatkowego oprogramowania na komputerze użytkownika. Korzystanie z systemu jest możliwe za pośrednictwem przeglądarki internetowej obsługującej JavaScript.

Opis efektów projektu:

Efektom projektu są kody źródłowe systemu umożliwiające instalację narzędzia na dowolnym serwerze oraz instalacja na serwerze Politechniki Wrocławskiej umożliwiająca autoryzowany dostęp do istniejących korpusów tekstowych, w tym korpusu KPWr, korpusu listów pożegnalnych (<http://www.pcsn.uni.wroc.pl/>) oraz korpusu wiadomości gospodarczych CEN. Instalacja jest dostępna pod adresem <http://www.nlp.pwr.wroc.pl/inforex/>. Kody systemu są dostępne na licencji GPL (<http://www.nlp.pwr.wroc.pl/narzedzia-i-zasoby/inforex>).

Powiązania z innymi polskimi i zagranicznymi projektami:

System Inforex został wykorzystany m.in. w projekcie pt. "Polski korpus listów pożegnalnych samobójców" do stworzenia korpusu listów pożegnalnych. Jest także wykorzystywany w projekcie Clarin-PL do rozbudowy korpusu KPWr.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.nlp.pwr.wroc.pl/inforex>

<http://www.nlp.pwr.wroc.pl/narzedzia-i-zasoby/inforex>

http://www.lrec-conf.org/proceedings/lrec2012/pdf/446_Paper.pdf

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Rozbudowa systemu o nowe funkcjonalności, w tym: liczenie zgodności na różnych poziomach anotacji, integracja z usługami sieciowymi umożliwiającymi tagowanie i znakowanie tekstów na życzenie użytkownika, poprawa użyteczności poprzez dodanie skrótów przyspieszając wykonywanie częstych czynności, ocena użyteczności interfejsu użytkownika.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Instytucja prowadząca badania w obszarze nauk humanistycznych i społecznych, w ramach których konieczna jest praca z kolekcjami dokumentów ręcznie znakowanymi zgodnie z określonymi schematami anotacji.

INTERAKTYWNA MAPA LITERACKA

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Instytut Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Centrum Technologii Językowych CLARIN-PL, Instytut Informatyki, Politechnika Wrocławska

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Aleksandra Wójtowicz, aleksandra.wojtowicz@ibl.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Maciej Maryl (IBL PAN), Jacek Paulinek (IBL PAN), Katarzyna Rdzanek (IBL PAN), Magdalena Górecka (UMCS), Aleksandra Wójtowicz (IBL PAN), Maciej Piasecki (PWr.), Michał Marcińczuk (PWr.), Marcin Pol (PWr.)

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 10/1/2013 - 10/1/2018

Źródło finansowania: środki własne instytucji

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze

Słowa kluczowe: Mapa; literatura; interaktywność, geolokacja, badania kulturowe, automatyczne rozpoznawanie jednostek identyfikacyjnych, wydobywanie informacji, rozpoznawanie relacji semantycznych, wizualizacja danych oparta na mapie

Skrócony opis projektu:

Projekt ma charakter interdyscyplinarny opiera się na badaniach literaturoznawczych i bibliologicznych wykorzystujących narzędzia geokrytyczne i analizy prowadzone na polu humanistyki cyfrowej. Autorzy projektu skupiają się na: literackiej redefinicji kartografii (proponując nowe ujęć badawczych), badaniach nad interaktywną kartografią oraz nad wytworzeniem narzędzia badawczego pozwalającego na przestrzenne ujęcie relacji między podmiotami życia literackiego. Projekt realizuje zalecenia European Science Foundation, sformułowane w raporcie Research Infrastructures in the Digital Humanities (2011), dotyczące ściśle

współpracy interdyscyplinarnej między informatykami a humanistami oraz tworzenia cyfrowych infrastruktur badawczych służących innym badaczom. Wykorzystane są geokodowanie oraz indeksowanie danych przestrzennych. Bardzo istotną częścią projektu jest opracowanie metody automatycznego rozpoznawania odniesień do przestrzeni i czasu występujących w tekście oraz relacji semantycznych wiążących byty wspomniane w tekście z rozpoznanymi miejscami.

Skrócony opis części informatycznej:

Projekt opiera się na metodach wydobywania informacji. Na potrzeby projektu opracowywane są rozszerzone metody rozpoznawania i klasyfikacji semantycznej wystąpień w tekście nazw własnych i innych jednostek identyfikacyjnych. Ponadto konstruowana jest rozszerzona metoda rozpoznawania relacji semantycznych pomiędzy obiektami wymienionymi w tekście oraz metoda rozpoznawania i klasyfikacji wyrażen temporalnych. Na potrzeby projektu opracowana została metoda konstrukcji wyrażen lokalizacyjnych na podstawie wydobytych z tekstu odniesień do obiektów geograficznych. Prowadzone są prace na metodami wizualizacji opartymi na mapie.

Opis efektów projektu:

Wybrane aspekty geografii literackiej i krytyki kartografii zostaną użyte jako narzędzia otwierające nowe ścieżki interpretacji dzieł literackich i relacji kulturowych. Mapa interaktywna stanowić będzie narzędzie badań z zakresu humanistyki cyfrowej, w obrębie której możliwe jest połączenie perspektywy filologicznej z nową geopolityką. Wytworzone narzędzie pozwoli przetwarzać i wykorzystywać opracowaną wiedzę w dalszych pracach. Skonstruowany zostanie elastyczny system do odnoszenia tekstu do mapy i czasu w oparciu o jego treść.

Powiązania z innymi polskimi i zagranicznymi projektami:

Joyce Ways (Dublin Joyce'a) <http://joyceways.com/>

HESTIA (interaktywne mapy Herodota) <http://hestia.open.ac.uk/>

A literary Atlas of Europe <http://www.literaturatlas.eu/>

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Zastosowanie technologii GIS i systemu połączeń geolokacyjnych do badań kulturowych i literaturoznawczych oraz wypracowania nowej ścieżki badań na polu geokrytyki. Opracowywanie sieci relacji pisarzy i bazy miejsc istotnych z perspektywy badań filologicznych.

Opracowanie dla języka polskiego rozszerzonej metody częściowej automatycznej formalnej interpretacji tekstu opartej na rzutowaniu tekstu na mapę geograficzną połączoną z wymiarem temporalnym.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

CLARIN-PL

Joyce Ways (Dublin Joyce'a) <http://joyceways.com/>

HESTIA (interaktywne mapy Herodota) <http://hestia.open.ac.uk/>

Profil potencjalnego partnera:

Grupy badawcze zajmujące się mapowaniem przestrzeni i badaniem relacji przestrzennych w literaturze.

INTERNETOWE REPOZYTORIUM TEKSTÓW ROZPRAW DOKTORSKICH I PREPRINTÓW INNYCH PRAC Z ZAKRESU ŚWIATOWEJ SLAWISTYKI I DZIEDZIN POKREWNYCH IRETESLAW

Osoba wypełniająca ankietę: Kierownik projektu

<http://www.ireteslaw.ispan.waw.pl>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Centrum Sławistycznej Informacji Naukowej, Instytut Sławistyki Polskiej Akademii Nauk, Polska Akademia Nauk

Instytucja przy której projekt jest afiliowany: Instytut Sławistyki PAN

Pozostałe instytucje zaangażowane w projekt: Fundacja Sławistyczna

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: dr Paweł Kowalski, kowalion@wp.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Paweł Kowalski, IS PAN; Marcin Fastyn, IS PAN; Jakub Banasiak, IS PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 3/15/2013 -

Źródło finansowania: środki własne instytucji, dofinansowanie z grantów 195342

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny

Efekt projektu: repozytorium

Słowa kluczowe: Repozytorium; sławistyka; językoznawstwo sławistyczne; baza danych, iReteslaw

Skrócony opis projektu:

Nowatorskie repozytorium rozpraw doktorskich, autoreferatów i recenzji prac doktorskich oraz preprintów innych typów prac naukowych z zakresu światowej sławistyki językoznawczej i dziedzin pokrewnych iReteslaw ma na celu upowszechnianie i popularyzowanie prac naukowych, a tym samym usprawnianie procesu wymiany myśli naukowej pomiędzy badaczami z różnych dziedzin w różnych ośrodkach i krajach.

Ponadto celem projektu jest upowszechnianie wśród niespecjalistów i osób spoza środowiska naukowego i akademickiego wiedzy z zakresu językoznawstwa sławistycznego i dziedzin pokrewnych.

Baza pomyślana została jako nowatorskie, komparatystyczne i interdyscyplinarne źródło tekstów opatrzonych rozbudowanymi opisami bibliograficznymi wraz ze słowami kluczowymi i abstraktami, umożliwiającymi użytkownikowi wyszukiwanie informacji o tekstach odpowiednio do potrzeb informacyjnych.

Skrócony opis części informatycznej:

System repozytorium został wykonany w

oparciu o technologie dostępne na licencjach OpenSource:

- system operacyjny Linux Debian,
- język programowania Java (maszyna wirtualna OpenJDK),
- system zarządzania bazą danych PostgreSQL,
- serwer aplikacji JBoss Application Server,
- serwer HTTP Apache,
- system indeksowania pełnotekstowego Apache Lucene,
- biblioteki narzędziowe Hibernate, Spring, Struts i inne.

Główne części systemu:

- baza danych zawierająca metadane o przechowywanych dokumentach,
- repozytorium plikowe zawierające pełne teksty przechowywanych dokumentów w formacie PDF (ewentualnie też w innych formatach),
- indeks pełnotekstowy umożliwiający wygodne przeszukiwanie metadanych oraz treści dokumentów,
- interfejs WWW użytkownika,
- interfejs WWW redaktora,
- interfejs WWW administratora,
- podsystem integrujący pozostałe części (odpowiedzialny między innymi za synchronizację danych między bazą danych a indeksem pełnotekstowym)

Opis efektów projektu:

Efektom projektu jest stworzenie od podstaw repozytorium iReteslaw, w którym zamieszczane są pełne teksty rozpraw doktorskich obronionych na uczelniach wyższych oraz w placówkach naukowych akademii nauk w Polsce i za granicą, teksty autoreferatów i recenzji prac doktorskich i habilitacyjnych, a także preprinty innych typów tekstów slawistycznych wraz z ich pełnym opisem bibliograficznym (formalnym i merytorycznym). Korzystanie z repozytorium jest w pełni bezpłatne, zarówno dla użytkowników publikujących swoje teksty, jak też dla czytelników. Teksty w repozytorium są dostępne w formacie PDF (możliwe są także formaty DOC i HTML) na licencji CC BY-NC-SA 3.0. Repozytorium jest w trakcie rozbudowy. Planowane jest rozszerzenie zawartości repozytorium oraz usprawnienie jego języka informacyjno-wyszukiwawczego.

Powiązania z innymi polskimi i zagranicznymi projektami:

Repozytorium iReteslaw jest merytorycznie związane z projektem iSybislaw – Bibliograficzna baza danych światowego językoznawstwa slawistycznego opracowywanym w Instytucie Slawistyki PAN.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Repozytorium jest w trakcie rozbudowy. Do najpilniejszych zadań należą:

- dalsze gromadzenie tekstów slawistycznych,
- usprawnienie języka informacyjno-wyszukiwawczego stosowanego w repozytorium.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Wszystkie slawistyczne instytucje naukowe.

KOMPUTEROWA ANALIZA NIEPROJEKTOWANEGO ODBIORU

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Badawczy KANO, Instytut Dziennikarstwa, Wydział Dziennikarstwa i Nauk Politycznych, Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Społem Warszawska Spółdzielnia Spożywców Śródmieście; Net Constructor

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Dorota Lewandowska-Jaros, lewdorota@gmail.com

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Dorota Lewandowska-Jaros, UW; Tomasz Jaros (Net Constructor)

INFORMACJE ORGANIZACYJNE

Status projektu: Planowany

Okres realizacji projektu: 9/15/2014 - 9/15/2016

Źródło finansowania: własne

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, publikacja

Słowa kluczowe: percepcja; przekaz; openCV; OOH; reklama

Skrócony opis projektu:

Projekt obejmuje badania oparte na analizie wyników automatycznej rejestracji czasu trwania i liczby spojrzeń odbiorców (zidentyfikowanych pod względem wieku i płci) na emitowany materiał audiowizualny (m.in. percepcja reklamy w systemie OOH).

Próbę badawczą stanowią wszystkie osoby znajdujące się w zasięgu elektronicznego punktu pomiarowego, umieszczonego w miejscu publicznym. Odpowiednie umiejscowienie terminala daje możliwość rejestracji każdego ze spojrzeń skierowanych bezpośrednio na ekran, na którym emitowany jest materiał. Osoby biorące udział w badaniu nie są o nim uprzedzone, co pozwala zakładać, iż badany jest odbiór nieprojektowany

(nieuwarunkowany jakimkolwiek nastawieniem). Zastosowanie komputerowej rejestracji wskaźników nieprojektowanego odbioru do analizy materiałów tekstowych i audiowizualnych w badaniach medioznawczych ma dać odpowiedź m.in. na pytania:

1. Jaką strukturę ma grupa osób stanowiąca próbę badawczą?
2. Które (i w jakim stopniu) materiały i ich sekwencje (jednostki tekstu i obrazy) przyciągają uwagę odbiorców?
3. Czy sekwencje (jednostki tekstu i obrazy) zakładane jako kluczowe dla danego przekazu w największym stopniu przyciągają uwagę odbiorców?
4. Jaki jest wpływ czasu i miejsca emisji na percepcję danego przekazu?
5. Jaki jest związek pomiędzy strukturą grupy stanowiącej próbę badawczą a percepcją określonego typu przekazu, zatem np. która z grup jest najwłaściwsza jako docelowa dla danego przekazu lub sekwencji?

Sposób rejestracji jest zgodny z wymogami Ustawy o ochronie danych osobowych (system działa w czasie rzeczywistym - nie rejestruje obrazu).

Skrócony opis części informatycznej:

Rejestracja materiału badawczego prowadzona jest za pomocą terminala komputerowego i kamery (z wykorzystaniem bibliotek openCV i własnej), umieszczonych przy ekranie, na którym emituje się materiał audiowizualny. Zebrane dane (liczba i czas spojrzeń oraz wiek i płeć odbiorców) zestawiane są z emitowanymi sekwencjami wykorzystanego do badania przekazu. Dane z rozproszonych geograficznie terminali eksportowane są do serwera centralnego, gdzie podlegają agregacji w systemie baz danych. Szczegółowe analizy ilościowe i jakościowe próbek danych przeprowadzane są na odrębnych stacjach roboczych.

Sposób rejestracji jest zgodny z wymogami Ustawy o ochronie danych osobowych (system działa w czasie rzeczywistym - nie rejestruje obrazu).

Prace informatyczne obejmują w założeniu także stworzenie i prowadzenie strony internetowej projektu.

Opis efektów projektu:

Efektom ma być:

1. stworzenie metody i innowacyjnego narzędzia badania percepcji przekazów tekstowych i audiowizualnych, opartej na założeniu zminimalizowania wpływu świadomości bycia uczestnikiem badania na reakcje osób badanych w miejscach publicznych, przy zachowaniu wymogów Ustawy o ochronie danych osobowych i maksymalnym (potencjalnie nieograniczonym) zwiększeniu próby badawczej;
2. weryfikacja hipotez, m.in.:
 - Istnieje związek pomiędzy płcią i wiekiem odbiorców a długością czasu, częstotliwością i momentem kierowania spojrzeń na materiał audiowizualny o danej tematyce czy konstrukcji.
 - Istnieją takie przekazy (ich sekwencje) audiowizualne, które przyciągają w równym stopniu uwagę odbiorców, bez względu na strukturę grupy stanowiącej próbę badawczą.
 - Czas (pora) emisji danego materiału audiowizualnego w określonym miejscu, przy maksymalnie zbliżonej pod względem struktury próbie badawczej, ma wpływ na wskaźniki percepcji tego materiału.

Ze względu na specyfikę metody i narzędzia badawczego oraz wymogów Ustawy o ochronie danych osobowych, dostęp do wykorzystywanych systemów informatycznych musi być ograniczony wyłącznie do uprawnionych osób z zespołu badawczego. Dane uzyskane w efekcie badań zleconych będą udostępniane odpłatnie - po kosztach ich pozyskania (koszt pracy urzędów, dzierżawa powierzchni pod terminale badawcze, obsługa).

Powiązania z innymi polskimi i zagranicznymi projektami:

Współpraca z działami IT firm Net Constructor oraz Społem WSS Śródmieście, prowadzących projekty w dziedzinie określania profili demograficznych klientów, oraz z firmą Lovo, realizująca projekt elektronicznej platformy analiz systemów ogłoszeniowych EPASO

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

www.netconstructor.pl; www.wss.spolem.org.pl; www.lovo.pl

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

1. Badania porównawcze oparte na zestawieniu wyników uzyskanych metodami klasycznymi (m.in. ankiety, eye-tracking, badania fokusowe) i metodą komputerowej analizy nieprojektowanego odbioru (z wykorzystaniem tych samych materiałów tekstowych i audiowizualnych);
2. Rozszerzenie badań o pogłębioną analizę lingwistyczną, socjologiczną, psychologiczną i kulturoznawczą.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

psychologowie, socjologowie, językoznawcy, badacze kultury

KOMPUTEROWE WSPOMAGANIE ANALIZY WYWIADÓW ETNOGRAFICZNYCH

Osoba wypełniająca ankietę: Osoba kontaktowa

www.etnologia.uw.edu.pl/edet

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: UW, Wydział Historyczny, Instytut Etnologii i Antropologii Kulturowej

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Univeristy Autonoma de Barcelona

MIM UW (konsultacje prof. Jerzy Tyszkiewicz)

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: prof. Lech Mróz

Osoba kontaktowa: Iwona Kaliszewska iwona.kaliszewska@gmail.com

Pozostałe osoby realizujące projekt:

IWona Kaliszewska, IEIAK UW,

konsultacja: prof. Jerzy Tyszkiewicz (MIM UW)

W dalszym rozwoju programu pomagali pracownicy Univeristy Autonoma de Barcelona.

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 1/1/2014 - 12/31/2014

Źródło finansowania: Fundusz Innowacji Dydaktycznych UW

OPIS PROJEKTU

Typ projektu: bazodanowy, digitalizacyjny, edukacyjny

Efekt projektu: oprogramowanie, repozytorium

Słowa kluczowe: wywiad etnograficzny; strukturyzacja tekstu, analiza wywiadów, analiza tekstów

Skrócony opis projektu:

Celem projektu było stworzenie programu wspomagającego analizę wywiadów etnograficznych oraz innych tekstów przydatnych w pracy antropologa.

Luźna struktura wywiadów etnograficzny sprawia, że niewielki sens ma używanie do ich analizy takich programów jak Atlas.ti lub Maxqda, przydatnych w analizie wywiadów ustrukturyzowanych. Program EdEt

ułatwia porządkowanie nieustrukturyzowanych informacji, sprawiając, że ich analiza staje się łatwiejsza i bardziej przejrzysta.

Skrócony opis części informatycznej:

Program napisany w Pythonie, używający XML-a.

Opis efektów projektu:

Program wspomagający analizę wywiadów etnograficznych oraz innych tekstów przydatnych w pracy antropologa.

Luźna struktura wywiadów etnograficzny sprawia, że niewielki sens ma używanie do ich analizy takich programów jak Atlas.ti lub Maxqda, przydatnych w analizie wywiadów ustrukturyzowanych. Program EdEt ułatwia porządkowanie nieustrukturyzowanych informacji, sprawiając, że ich analiza staje się łatwiejsza i bardziej przejrzysta.

Powiązania z innymi polskimi i zagranicznymi projektami:

Univeristy Autonomia de Barcelona współpracował z IEIAK UW w zakresie stworzenia hiszpańskojęzycznego podręcznika do programu EDET. Pracowniczka IEIAK prowadziła dwukrotnie szkolenia dla socjologów z Univeristy Autonomia de Barcelona

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

www.etnologia.uw.edu.pl/edet

oraz podręczniki użytkownika dostępne po zainstalowaniu programu EDet

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Przystosowanie programu do platform MAC i Linux. Obecnie program działa tylko na WIndows.

Usprawnienie szybkości działania programu przy większych danych,

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Kontunujaca przerwaniej w ostatnich latach (z braku funduszy) współpracy z Univeristy Autonomia de Barcelona

Profil potencjalnego partnera:

Uniwersytety, placówki badawcze

KONSTRUKCJA ELEKTRONICZNYCH SŁOWNIKÓW (BAZ DANYCH LEKSYKALNYCH) I
DEZAMBIGUIZACJA ZNACZENIOWA DLA CELÓW TŁUMACZENIA AUTOMATYCZNEGO.

Osoba wypełniająca ankietę: Osoba kontaktowa

strona w przygotowaniu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zakład Językoznawstwa Stosowanego i Translatoryki

(Uniwersytet Śląski, Wydział Filologiczny, Instytut Języków Romańskich i Translatoryki)

Institucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Prof. zw. dr hab. Wiesław Banyś wbanys@gmail.com; wieslaw.banys@us.edu.pl

Osoba kontaktowa: kсения.galuskina@us.edu.pl

aleksandra.zlobinska-nowak@us.edu.pl

Pozostałe osoby realizujące projekt:

dr Chrupała Aleksandra, UŚ, IJRIT

dr Ciszewska-Jankowska Ewa, UŚ, IJRIT

dr Czekał Anna, UŚ, IJRIT

dr Drzazga Anna, WSB Dąbrowa Górnicza

dr Gabrysiak Katarzyna, UŚ, IJRIT

mgr Gałuska Ksenia, UŚ, IJRIT

mgr Hrabia Michał, UŚ, IJRIT

dr Mrochen Izabela, UŚ, IJRIT

dr Nykiel Jerzy, UŚ, IJRIT

mgr Palion-Musioł Agnieszka, ATH, Bielsko-Biała

dr Perz Magdalena, UŚ, IJRIT

dr Słapek Daniel, UŚ, IJRIT

dr Szramek-Karcz Sonia, UŚ, IJRIT

dr Śmigielska Beata, UŚ, IJRIT

dr Żłobińska-Nowak Aleksandra, UŚ, IJRIT

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 1/1/2002 - 12/31/2020

Źródło finansowania: środki własne instytucji nie dotyczy

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: baza danych

Słowa kluczowe: ujęcie zorientowane obiektowo;

klasy obiektowe;

dezambiguizacja;

polisemia;

tłumaczenie

Skrócony opis projektu:

Projekt koncentruje się na analizie słownictwa w ramach ujęcia obiektowego. Głównym jego założeniem jest dezambiguizacja (odwieloznacznianie) wyrażeń polisemicznych języka francuskiego, hiszpańskiego, włoskiego i angielskiego dla celów tłumaczenia automatycznego oraz konstrukcja słowników elektronicznych - elektronicznych baz danych leksykalnych. Słownictwo opisywane jest w kategoriach klas obiektów. Ujęcie zorientowane obiektowo jest dopełnione poprzez powracający uogólniający cykl klasyfikowania operatorów i atrybutów w klasy, których klamrą jest dana forma leksykalna i stanowi jednocześnie metodę wskazywania w kontekście, w której występuje dana forma leksykalna, tych elementów, należących do jakiejś klasy obiektowej, które pozwalają na nowy sposób wydobywania ilości jej znaczeń, która jest tutaj wypadkową różnych klas obiektowych realizowanych na pozycjach argumentowych i różnych odpowiedników w języku, na który dokonuje się tłumaczenia.

Skrócony opis części informatycznej:

Trwają prace nad informatycznym opracowaniem baz danych leksykalnych oraz ich scaleniem.

Opis efektów projektu:

Zastosowania ujęcia zorientowanego obiektowo do kolejnych części słownictwa:

zrealizowane: francusko-polski opis rzeczowników wyrażających pojęcie komunikacji; planowane: francusko-polski opis wybranych, najbardziej polisemicznych rzeczowników o największej frekwencji użycia z naciskiem na szczegółowy opis kadrów, w których zostały one użyte w celu poprawnej interpretacji ich znaczenia. (dr B. Śmigielska).

zrealizowane: opis terminologii specjalistycznej z dziedziny prawa dla potrzeb przekładu automatycznego. Umieszczenie prowadzonych badań w nurcie opisowych badań nad terminologią (leksykografią terminologiczną), językoznawstwa komputerowego oraz językoznawstwa korpusowego. Analiza aspektów praktycznych zastosowania metod językoznawczych oraz baz leksykograficznych w środowisku pracy tłumacza, w tym środowisku informatycznym narzędzi wspomagających tłumaczenie (mgr K. Gałuskińska).

zrealizowane: francusko-polski opis odwieloznaczający czasowników ruchu monter i sortir; planowane: opis wybranych czasowników ruchu (j. francuski, j. hiszpański i j. polski) w perspektywie kontrastywnej. Analiza z uwzględnieniem klas obiektowych występujących w otoczeniu badanych czasowników i wyróżnieniem podstawowych ich struktur syntaktyczno-semantycznych. Tłumaczenie wyróżnionych struktur i znaczeń na język polski. (dr A. Żłobińska-Nowak).

zrealizowane: francusko-polski opis klasy obiektów: <parties du corps>; planowane: konstrukcje metonimiczne w tłumaczeniu automatycznym wg ujęcia obiektowego (dr A. Czekaj).

zrealizowane: francusko-polski opis klasy obiektów: <phénomènes naturels> w ujęciu zorientowanym obiektowo planowane: odwieloznaczanie wybranej klasy przymiotników polisemicznych (dr M. Perz).

opis konstrukcji ze słowem femme w języku francuskim, badanie stopnia skostnienia językowego w celu tłumaczenia automatycznego (dr A. Chrupała).

zrealizowane: francusko-polski opis klasy obiektów: <professions> w ujęciu zorientowanym obiektowo (dr S. Szramek-Karcz).

zrealizowane: francusko-polski opis odwieloznaczający przymiotnika farouche; planowane: opis francuskich czasowników ruchu tłumaczonych na język polski przez jechać/jeździć związanych z wyborem właściwego ekwiwalentu tłumaczeniowego. (mgr M. Hrabia).

zrealizowane: opracowanie narzędzi komputerowych oraz ram przygotowywanej włosko-polskiej bazy danych leksykalnych (dr D. Słapek)

zrealizowane: angielsko-polski opis odwieloznaczający czasowników send i open (dr A. Drzazga).

zrealizowane: francusko-polski opis odwieloznaczający czasowników déterminer i produire (dr K. Gabrysiak)

planowane: hiszpańsko-polski opis odwieloznaczający czasowników salir i entrar (mgr A. Palion-Musioł)

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekt MATCHPAD V Programu Ramowego Unii Europejskiej, tworzonego we współpracy z:

- Laboratoire de Linguistique Informatique (LLI; obecnie LDI: Lexiques, Dictionnaires, Informatique) Uniwersytetu Paris 13, szczególnie z profesorem Gastonem Grossem,

- Instytutem Romanistyki Uniwersytetu Warszawskiego z prof. zw. dr. hab. Krzysztofem Bogackim oraz

- Instytutem Filologii Romańskiej Uniwersytetu Śląskiego (obecnie Instytutem Języków Romańskich i Translatoryki), pod kierunkiem prof. zw. dr. hab. Wiesława Banysia.

Projekt miał na celu opracowywanie komputerowych baz danych do automatycznego wspomaganie tłumaczenia tekstów.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

www.ifr.us.edu.pl (strona Instytutu Języków Romańskich i Translatoryki oraz podstrona Zakładu Języków Romańskich i Translatoryki:

<http://www.ifr.us.edu.pl/o-instytucie/struktura-instytutu/zaklad-jezykoznawstwa-stosowanego-i-translatoryki.html>)

<http://www.sbc.org.pl/publication/3284> (kolejne wydania czasopisma Neophilologica od numeru 15 z licznymi publikacjami na temat ujęcia obiektowego i analiz przeprowadzanych w ramach tejże koncepcji)

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Konsekwentna analiza słownictwa pochodzącego z różnych dziedzin, zarówno ze szczególnym naciskiem na słownictwo ogólne, w obrębie którego mamy do czynienia z szeroko pojętą wieloznacznością oraz słownictwo specjalistyczne, ma w założeniu stanowić spójny, modułarny opis języka dla potrzeb tłumaczenia automatycznego. Podjęte i stale rozwijane prace zmierzają do bardzo szczegółowej analizy leksyki stanowiąc odpowiedź na lukę i niedomagania wśród dostępnych na rynku słowników dwujęzycznych i stale rosnące zapotrzebowanie na tłumaczenie automatyczne.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Współpraca z instytutami, laboratoriami zajmującymi się konstruowaniem językowych baz danych, słowników elektronicznych, badających zagadnienia polisemii i odwieloznaczniania oraz możliwości jego rozwiązywania po stronie językowej dla celów tłumaczenia automatycznego z wykorzystaniem narzędzi informatycznych (metody symboliczne i statystyczne).

Profil potencjalnego partnera:

Instytucja lub grupa badawcza mogąca dostarczyć opracowanie informatyczne dla rozwijanego projektu polegającą na abstrakcji danych z zastosowaniem określonego języka programowania (np. język C++) z wykorzystaniem jego paradygmatów.

Nadrzędną cechą realizowanych w ramach projektu opisów językoznawczych i ich aplikacji informatycznych jest ich modularność, rozumiana tutaj w sposób rozszerzony, nie tylko jako możliwość wydzielenia w opracowywanym materiale językoznawczym czy też realizowanym dlań software części składowych i oddzielne ich badanie czy też konstruowanie (w sensie dekompozycji zatem), ale także jako takie ich opracowywanie, aby opis / konstrukcja proponowany/a mógł być wykorzystywany/a w innych kompleksach opisowych i softwarowych, niż aktualnie prowadzony (w sensie kompozycji zatem, co pozwala na maksymalną "rentowność" i "mobilność" realizowanych prac) a także takie, które dąży do jak największej autonomizacji modułów opisowych/aplikacyjnych, co gwarantuje konieczność mniejszych zmian w modułach połączonych tak, by jego zawartość nie musiała ewentualnie podlegać modyfikacji. Takie podejście do opisu językoznawczego i konstrukcji software określa szczegóły metodologiczne i konstrukcyjne realizowanych zadań.

KORPUS POLITECHNIKI WROCŁAWSKIEJ (KPWR)

Osoba wypełniająca ankietę: Marcin Oleksy, marcin.oleksy@pwr.wroc.pl

<http://www.clarin-pl.eu>, <http://www.nlp.pwr.wroc.pl/projekty/clarin>,
<http://www.nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/kpwr>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Grupa Naukowa G4.19, Instytut Informatyki Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Projekt jest realizowany wyłącznie na Politechnice Wrocławskiej

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr Marek Maziarz (PWr.), dr inż. Maciej Piasecki (PWr.), dr Marcin Oleksy (PWr), mgr Jan Wieczorek (PWr), dr inż. Adam Radziszewski (PWr), dr inż. Michał Marcińczuk, mgr inż. Paweł Kędzia, mgr inż. Jan Kocoń oraz lingwiści-anotatorzy

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 8/16/2010 - 12/31/2015

Źródło finansowania: dotacja celowa, Narodowe Centrum Badań i Rozwoju, Program Operacyjny "Innowacyjna Gospodarka" NEKST POIG, SyNaT projekt strategiczny NCBiR, CLARIN-PL dotacja na inwestycję MNISW

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy

Efekt projektu: narzędzie badawcze, baza danych, publikacja

Słowa kluczowe: korpus językowy, zasób językowy, ręczna anotacja korpusu, uczenie maszynowe, metadane

Skrócony opis projektu:

Korpus Politechniki Wrocławskiej (KPWr) to korpus pisanych i mówionych tekstów gromadzonych na licencji Creative Commons podzielonych na 14 kategorii (blogi, nauka, stenogramy parlamentarne, proza dawna, proza współczesna, krótkie wiadomości prasowe, długie wiadomości prasowe, dialog, teksty popularno-naukowe i podręczniki, teksty religijne, techniczne, urzędowe, ustawy, wikipedia). Jest on w znacznej części ręcznie anotowany na kilku poziomach (fraz składniowych oraz wybranych relacji predykatowo-argumentowych,

hierarchicznym schematem nazw własnych, sensów wybranych słów, anaforą i koreferencją między frazami rzeczownikowymi oraz wybranymi relacjami semantycznymi między nazwami własnymi)

Skrócony opis części informatycznej:

Projekt ze swojej natury jest w dużej mierze lingwistyczny, ale opracowanie korpusu wymagało rozwiązania szeregu problemów informatycznych. W szczególności opracowano oparty na XML-u format danych (format CCL) oraz narzędzia wspomagające przetwarzanie tych danych.

Poza tym dane z korpusu służą za cenny materiał umożliwiający opracowanie heurystyk doboru danych uczących i testowych na potrzeby realizacji zadań z zakresu przetwarzania języka naturalnego, m.in. rozpoznawania zdarzeń, wyrażeń temporalnych i przestrzennych, nazw własnych czy struktury dokumentu i relacji między tekstami.

Opis efektów projektu:

Korpus Języka Polskiego Politechniki Wrocławskiej (KPWr) jest obecnie rozwijany w ramach projektu CLARIN. Dodano kolejne poziomy anotacji oraz rozwinięto już istniejące (korpus jest ręcznie anotowany m.in. rozszerzonym hierarchicznym schematem nazw własnych zawierającym 100 kategorii nazw własnych, wyrażeniami temporalnymi i przestrzennymi, sytuacjami, metadanymi, informacjami o strukturze oraz płytką strukturą semantyczną na potrzeby uczenia i testowania parsera). Zwiększane jest pokrycie anotacjami także na pozostałych poziomach (m.in. anafory i koreferencji między frazami rzeczownikowymi).

Powiązania z innymi polskimi i zagranicznymi projektami:

W ramach projektu badawczo-rozwojowego realizowanego w latach 2007–2012 przez konsorcjum kierowane przez Instytut Podstaw Informatyki PAN zbudowano Narodowy Korpus Języka Polskiego. Korpus narodowy niż KPWr jest znacznie większy, natomiast nie zawiera wielu z warstw anotacji, które zawarto w KPWr. Licencja korpusu narodowego nie jest też tak otwarta i swobodna jak KPWr.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.clarin-pl.eu>

<http://www.nlp.pwr.wroc.pl/projekty/clarin>, <http://www.nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/kpwr>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Rozbudowa Korpusu Języka Polskiego Politechniki Wrocławskiej (KPWr). Jej celem będzie po pierwsze zrównoważenie korpusu w ramach stylów funkcjonalnych języka polskiego (naukowego, artystycznego, dziennikarskiego, urzędowego i potocznego), co nada korpusowi bardziej uniwersalny charakter. Po drugie zwiększone zostanie pokrycie ręcznym znakowaniem na istniejących poziomach oraz zostaną dodane kolejne poziomy anotacji zwiększające zakres analizowanych zachowań komunikacyjnych. Powstały w ten sposób zasób językowy może stanowić bardzo cenny materiał do konstruowania narzędzi z zakresu automatycznego przetwarzania języka naturalnego, m.in. inteligentnego wyszukiwania informacji w tekście, porównywania tekstów, automatycznego zbierania danych badawczych z dużych kolekcji tekstowych itd.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Możliwe jest połączenie Korpusu Politechniki Wrocławskiej z wordnetem języka polskiego Słowosieć w celu testowania nowych metod ekstrakcji informacji, klasyfikowania tematycznego tekstów i zarządzania nimi oraz z innymi korpusami języka polskiego.

Profil potencjalnego partnera:

różne instytucje prowadzące badania w obszarze nauk humanistycznych i społecznych, w ramach których wykorzystywany jest dostęp do dużych zbiorów dokumentów tekstowych. KPWr dzięki otwartej licencji stanowi otwarte, publiczne archiwum zasobów językowych i materiałów źródłowych.

KORPUS POLSZCZYNY XVI WIEKU. ETAP I: DIGITALIZACJA ŹRÓDEŁ ORAZ STWORZENIE NARZĘDZI INFORMATYCZNYCH I UDOSTĘPNIENIE MATERIAŁÓW TESTOWYCH KORPUSU

Osoba wypełniająca ankietę: Osoba kontaktowa

www.spxvi.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Patrycja Potoniec, spxvi@poczta.onet.pl

Osoba kontaktowa: Krzysztof Opaliński, k.opalinski.ibl@gmail.com

Pozostałe osoby realizujące projekt:

Patrycja Potoniec, IBL PAN; Krzysztof Opaliński, IBL PAN; Roman Spandowski, UKW; Paweł Bohuszewicz, UMK; Katarzyna Galińska; Zofia Chyła-Bełkot; Eliza Biesaga; Jan Dokurno; Bartosz Biały

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 3/1/2012 - 2/28/2017

Źródło finansowania: Narodowy Program Rozwoju Humanistyki 0138/FNiTP/H11/80/2011

OPIS PROJEKTU

Typ projektu: digitalizacyjny

Efekt projektu: baza danych, repozytorium

Słowa kluczowe: korpus tekstów; korpus historyczny; językoznawstwo polskie; polszczyzna dawna; polska literatura renesansu

Skrócony opis projektu:

Projekt ma na celu utworzenie elektronicznej bazy tekstów polskich powstałych w XVI wieku. Nadrzędnym celem projektu jest ocalenie bazy materiałowej „Słownika polszczyzny XVI wieku”, która ze względu na wiek i intensywność użytkowania coraz bardziej ulega zniszczeniu. Baza tekstów ma służyć dalszym pracom redakcyjnym nad słownikiem. Projekt ze względu na ogrom materiału przewidziano jako przedsięwzięcie etapowe. W pierwszym etapie dygitalizacji poddane zostaną teksty z tzw. kanonu źródeł „Słownika polszczyzny XVI wieku”, które po zakończeniu prac będą dostępne jako repozytorium. Próbką tekstów tak utworzonego

zbioru zostanie poddana tagowaniu słowoform i jako próbka korpusu posłuży dyskusji nad jego dalszym kształtem.

Skrócony opis części informatycznej:

Przyjęto, że korpus będzie spełniał ogólnie stosowane standardy dla tego typu zbiorów tekstów. W związku z tym opracowany został wyciąg ze specyfikacji TEI (schemat DTD), który zakłada format wynikowy XML. Wyciąg ten zastosowano w specjalnie utworzonym edytorze, który pozwala edytować teksty transliterowane i jednocześnie kontroluje poprawność tworzonego pliku. Gotowe teksty po zakończeniu prac będą dostępne wyłącznie w formie elektronicznej (po przekształceniach XSLT na format HTML lub PDF) na stronie www.spxvi.edu.pl. Ponadto zostanie opracowany program ułatwiający tagowanie słowoform w korpusie.

Opis efektów projektu:

Niemal cały kanon źródeł „Słownika” został już przepisany, obecnie trwają żmudne korekty. Prace nad korpusem zaowocują powstaniem największej bazy polskich tekstów XVI-wiecznych w formacie tekstowym. Format wynikowy stworzy możliwość zastosowania narzędzi wyszukiwawczych i budowania wyrafinowanych kwerend. Korpus da również możliwość zintegrowania go ze „Słownikiem polszczyzny XVI wieku”, a to z kolei pozwoli na weryfikację bazy materiałowej słownika, która powstała w latach pięćdziesiątych ubiegłego stulecia przy pomocy dostępnych wówczas niezbyt doskonałych narzędzi.

Powiązania z innymi polskimi i zagranicznymi projektami:

„Korpus polszczyzny XVI wieku” w założeniu pretenduje do bycia częścią działu diachronicznego Narodowego Korpusu Języka Polskiego. Poza formalnymi związkami z istniejącymi już korpusami współczesnej polszczyzny otwiera się szerokie pole współpracy z innymi korpusami diachronicznymi, wśród których najbardziej zaawansowany jest elektroniczny korpus tekstów polskich XVII i XVIII wieku.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.spxvi.edu.pl/korpus/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Projekt jest przewidziany jako przedsięwzięcie etapowe. W kolejnych etapach baza tekstów będzie uzupełniona o brakujące (nierozpisane) fragmenty tekstów z kanonu źródeł „Słownika polszczyzny XVI wieku” oraz o kolejne teksty (spoza kanonu źródeł słownika). Planowane jest też kontynuowanie tagowania słowoform w całym korpusie, a także jego integracja zarówno ze słownikiem, jak i z innymi korpusami historycznymi języka polskiego.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Korpus polszczyzny XVI wieku może stanowić podstawę wydawniczą dla cyfrowych edycji dzieł staropolskich.

Profil potencjalnego partnera:

zespoły edytorskie, zespoły korpusowe

KORPUSY DWUWARSTWOWE TEKSTÓW SKANOWANYCH

Osoba wypełniająca ankietę: Joanna Bilińska, j.bilinska@uw.edu.pl

korpusy.klf.uw.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Katedra Lingwistyki Formalnej,

Wydział Neofilologii,

Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Janusz S. Bień, jsbien@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 12/12/2009 -

Źródło finansowania: środki własne instytucji, do 12.05.2012 grant MNiSW grant nr N N519 384036

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, korpusowy

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja, przeglądarki korpusów dwuwarstwowych

Słowa kluczowe: korpusy, korpusy historyczne, słowniki historyczne, DjVu

Skrócony opis projektu:

Pierwotna koncepcja zakładała ułatwienie przeszukiwania wybranych słowników dostępnych w innych bibliotekach cyfrowych (stąd robocza nazwa Wirtualna Biblioteka Leksykograficzna), ale najbardziej interesujące nas drugie wydanie słownika Lindego zostało specjalnie zdigitalizowane na potrzeby tego projektu (wydaje się, że jest to najlepsza jakościowo dygitalizacja tego wydania dostępna w Internecie).

Skrócony opis części informatycznej:

W projekcie wykorzystywane jest oprogramowanie przygotowane w ramach opisanego w osobnej ankiecie grantu "Narzędzia dygitalizacji tekstów na potrzeby badań filologicznych" (13.05.2009 - 12.05.2012). Istotnie wykorzystywany jest też format DjVu.

Opis efektów projektu:

W formie tzw. korpusów dwuwarstwowych udostępniane są 4 słowniki i

jeden korpus dawnych tekstów polskich.

Słownik warszawski liczy ok. 8 tysięcy stron,

Słownik polszczyzny XVI wieku ok. 20 tysięcy,

słownik Lindego ok. 5 tysięcy,

Słownik geograficzny Królestwa Polskiego 15 tysięcy stron,

korpus projektu IMPACT w wersji jedno- i dwuwymiarowej – ok. 1,5 mln tzw. segmentów.

Powiązania z innymi polskimi i zagranicznymi projektami:

Wszystkie udostępniane w tym projekcie słowniki dostępne są też w krajowych bibliotekach cyfrowych, jednak możliwości ich przeszukiwania są bardzo ubogie. Udostępniany za pomocą wyszukiwarki ""Słownik polszczyzny XVI wieku"" został udostępniony przez IBL PAN za pośrednictwem KPBC. ""Słownik warszawski"" pochodzi z e-BUW, a korpusy tekstów dawnych z projektu IMPACT. Konkurencyjną wyszukiwarkę dla ""Słownika geograficznego"" udostępnia natomiast ICM UW.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Bień, Janusz S. (2012) The IMPACT project Polish Ground-Truth texts as a DjVu corpus.

<http://bc.klf.uw.edu.pl/296/>.

Bień, Janusz S. (2012) Skanowane teksty jako korpusy. Prace

Filologiczne, LXIII. pp. 25-36. ISSN 0138-0567

<http://bc.klf.uw.edu.pl/322/>.

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Katedra Lingwistyki Formalnej nie planuje dodawania nowych zasobów do wyszukiwarki, lecz jedynie ulepszanie wybranych z nich.

Wskazany jest rozwój przeglądarek DjVu, zarówno ogólnych (djview4), jak i specjalizowanych (djview4poliqarp), w tym przygotowanie przystępnych dokumentacji i tutoriali.

Bardziej radykalna koncepcja rozwoju bazująca na wykorzystaniu Poliqarpa 2 została przedstawiona w innej ankiecie.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Możliwe jest tworzenie replik korpusów na serwerach zainteresowanych instytucji, a także tworzenie nowych korpusów.

Wskazane jest wsparcie projektu DjVuLibre, w ramach którego rozwijany jest djview4.

Profil potencjalnego partnera:

Biblioteki cyfrowe zainteresowane ułatwieniem przeszukiwania swoich zasobów. Zespoły programistyczne z biegłą znajomością QT i/lub języka Python oraz z umiejętnością przygotowywania pakietów i instalatorów dla podstawowych platform (w tym Debian GNU/Linux i Ubuntu), najlepiej posiadające także uprawnienia deweloperskie dla Macintosha.

LABORATORIUM TECHNOLOGII INFORMACYJNYCH (LATEI)

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Informacji Naukowej i Studiów Bibliologicznych, Wydział Historyczny, Uniwersytet Warszawski

Institucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Grzegorz Gmiterek, ggmiterek@gmail.com ; mikolaj.ochmanski@gmail.com

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Dariusz Kuźmina, IINiSB UW

Grzegorz Gmiterek, IINiSB UW

Mikołaj Ochmański, IINiSB UW

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 6/1/2014 - 12/31/2015

Źródło finansowania: MNiSW

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja, kurs

Słowa kluczowe: nowe technologie; laboratorium; digitalizacja; restrukturyzacja; informatologia

Skrócony opis projektu:

Projekt restrukturyzacji Instytutu Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego (IINiSB) w latach 2014-2015 obejmuje utworzenie zaawansowanej technologicznie międzyzakładowej pracowni naukowo-badawczej pod nazwą: Laboratorium Technologii Informacyjnych (LaTel), w ramach której realizowane będą zarówno prace tematycznie pozostające w zainteresowaniach ścisłej bibliologii i

informatologii, jak i z rozległego, interdyscyplinarnego pogranicza, np. na styku z naukoznawstwem, historią, medioznawstwem, socjologią, pedagogiką, czy teorią literatury.

Skrócony opis części informatycznej:

Zadaniem Laboratorium będzie dostarczenie i rozbudowa narzędzi badawczych umożliwiających wykorzystanie nowych technologii w bibliologii i informatologii, a jednocześnie badanie nowoczesnych technologii informacyjnych. Dzięki projektowanym rozwiązaniom pracownia przyczyni się do odkrywania i rozwijania różnych pól badawczych bibliologii i informatologii oraz wprowadzania innowacji w badaniach naukowych, a także do włączenia do narzędzi nauk społecznych rozwiązań zaawansowanych technologicznie.

Opis efektów projektu:

W efekcie utworzenia pracowni nastąpi znaczne zwiększenie liczby badań stosowanych w IINiSB UW, wzrost potencjału badawczego Instytutu, wprowadzenie elementu innowacyjności badań, nawiązanie/wzmocnienie współpracy z partnerami naukowymi, pozarządowymi i biznesowymi z kraju i zagranicy. Ponadto ułatwiony będzie dostęp do zebranych i opracowanych informacji oraz wyników badań.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

LISTY POŻEGNALNE SAMOBÓJCÓW – LINGWISTYCZNE METODY USTALANIA AUTENTYCZNOŚCI TEKSTU

Osoba wypełniająca ankietę: Kierownik projektu

www.pcsn.uni.wroc.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Uniwersytet Wrocławski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Instytut Informatyki Politechniki Wrocławskiej

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Monika Zaśko-Zielińska, monik@uni.wroc.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Monika Zaśko-Zielińska UWr; Maciej Piasecki PWr; Michał Marcińczuk PWr; Ewelina Moroń UWr; Magdalena Plinta UWr

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 10/28/2010 - 10/27/2013

Źródło finansowania: Narodowe Centrum Nauki N N104 058439

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, publikacja

Słowa kluczowe: list pożegnalny samobójcy; językoznawstwo sądowe, korpus specjalistyczny, potwierdzanie autentyczności tekstu, idiolekt

Skrócony opis projektu:

Prowadzony projekt objął prace nad:

- 1) Opisem tła społeczno-kulturowego gatunku (w tym: tabu związanego z samobójstwem, uwarunkowań religijnych dawne i współczesne, opisem materialnego wymiaru listu, dane statystyczne na temat samobójstw oraz listów pożegnalnych w Polsce z lat 1998-2008 (roczniki GUS, baza danych Temida), kontekstem prawnym funkcjonowania listu pożegnalnego, miejscem listu pożegnalnego w dyskursie presuicydalnym oraz sytuacją komunikacyjną powiązaną z listem pożegnalnym).

- 2) Polskim Korpusem Listów Pożegnalnych, który składa się z korpusu listów prawdziwych oraz dwóch subkorpusów (Subkorpusu Listów Symulowanych i Subkorpusu Listów Sfałszowanych).
- 3) Opisem wzorca gatunkowego listu pożegnalnego.
- 4) Analizą jakościowo-ilościową prowadzoną w odniesieniu to trzech korpusów. W badaniach ilościowych źródłem do porównań stały się również zasoby Narodowego Korpusu Języka Polskiego. Uzyskane wyniki sprawdzone zostały testem istotności statystycznej.

Skrócony opis części informatycznej:

W ramach projektu opracowano internetową aplikację do edycji i anotacji korpusu listów pożegnalnych, którą wykonano na podstawie systemu Inforex umożliwiającego zarządzanie korpusami tekstowymi poprzez przeglądarkę internetową. Opracowano też projekt i implementację modułu do ręcznej transkrypcji skanów listów pożegnalnych. Transkrypcję dokumentów poprzedziło opracowanie schematu zapisu listów, który został zapisany w standardzie XSD. Schemat listu oparto na powszechnie dostępnym standardzie kodowania dokumentów - TEI P5. Perspektywę transkrypcji wyposażono w mechanizmy wspomagające proces ręcznej transkrypcji listów (przybornik z dopuszczalnymi elementami struktury listu oraz mechanizm automatycznej weryfikacji zgodności transkrypcji ze schematem XSD).

Przygotowana aplikacja umożliwia automatyczną analizę morfologiczną oraz obejmuje anotację semantyczną i pragmatyczną listów.

Opis efektów projektu:

- Przygotowanie Polskiego Korpusu Listów Pożegnalnych (614 autentycznych listów z lat 1999-2008) oraz dwóch subkorpusów z materiałów zebranych eksperymentalnie: Subkorpusu Listów Symulowanych (79 tekstów) i Subkorpusu Listów Sfałszowanych (117 tekstów) - dostęp do anonimizowanej wersji materiałów korpusowych po uzyskaniu zgody;
- Opracowanie zagadnienia dotyczącego badania autentyczności listów pożegnalnych na podstawie wskaźników lingwistycznych - Listy pożegnalne: w poszukiwaniu lingwistycznych wyznaczników autentyczności tekstu, / Monika Zaśko-Zielińska, Wrocław: Quaestio, 2013. - 285 s.

Powiązania z innymi polskimi i zagranicznymi projektami:

Korpusy listów pożegnalnych:

Birmingham Corpus of suicide Notes – opr. Jess Shapiro;

- Suicide Notes Database – opr. J. Pestian, P. Matykiewicz i in. (Division of Biomedical Informatics (BMI) at Cincinnati Children's Research Foundation)

- Vienna Corpus of Suicide Notes - Brigitte Eisenwort

Korpusy listów prywatnych:

- Digital Archive of Letters by Flemish Authors and Composers from the 19th & 20th century

- Projects CARDS (Cartas Desconhecidas – Unknown Letters) and FLY (Forgotten Letters, Years 1900-1974)

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

- Strona projektu: – pcsn.uni.wroc.pl

Structure annotation in the polish corpus of suicide notes / Michał Marcińczuk, Monika Zaśko-Zielińska, Maciej Piasecki // W: Text, speech and dialogue : 14th International Conference, TSD 2011, Pilsen, Czech Republik

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Prace nad projektem pokazały, jak istotne jest udostępnianie do badań tekstów prywatnych pisanych przez autorów różnego wieku, wykształcenia i pochodzenia, które są realizacją języka pisanego w odmianie nieoficjalnej. Utrwalenie w wersji elektronicznej tego typu dokumentów i ich anotacja przy użyciu jednolitego oprogramowania zapewni dostęp do badań przedstawicielom różnych specjalności, a także umożliwi prowadzenie analiz porównawczych.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Współpraca z projektami dotyczącymi wykorzystania narzędzi informatycznych do badań w zakresie humanistyki.

Współpraca z ośrodkami zajmującymi się przygotowaniem korpusów listów prywatnych i innych dokumentów osobistych.

Współpraca z ośrodkami dokumentacji życia codziennego.

Profil potencjalnego partnera:

Partnerem projektu może być grupa badawcza, która ma doświadczenie w wykorzystaniu narzędzi informatycznych do przygotowania korpusów specjalistycznych umożliwiających analizę jakościowo-ilościową tekstów.

LUNA - SPOKEN LANGUAGE UNDERSTANDING IN MULTILINGUAL COMMUNICATION SYSTEMS

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Instytut Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: LOQUENDO S.P.A., Torino, Italy ;

RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN, , Aachen, Germany;

UNIVERSITÀ DEGLI STUDI DI TRENTO, Trento, Italy;

UNIVERSITE D'AVIGNON ET DU PAIS-VAUCLUSE, Avignon, France;

FRANCE TELECOM S.A., Paris, France;

CONSORZIO PER IL SISTEMA INFORMATIVO (CSI-PIEMONTE), Torino, Italy;

POLSKO JAPONSKA WYZSZA SZKOŁA TECHNIK KOMPUTEROWYCH, Warszawa, Poland

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Agnieszka Mykowiecka, agn@ipipan.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Agnieszka Mykowiecka, IPI PAN;

Małgorzata Marciniak, IPI PAN;

Joanna Rabiega-Wisniewska, IPI PAN;

Katarzyna Głowińska, IPI PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 9/4/2006 - 9/3/2009

Źródło finansowania: FP6 Contract no. 033549

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, oprogramowanie, publikacja

Słowa kluczowe: systemy dialogowe, analiza mowy, analiza semantyczna

Skrócony opis projektu:

Projekt LUNA dotyczył problemu rozumienia mowy spontanicznej w nowoczesnych serwisach telekomunikacyjnych. Zasadniczym celem projektu było stworzenie niezawodnego systemu rozumienia mowy przeznaczonego dla wielojęzycznych systemów dialogowych w centrach informacyjnych, zapewniającego dostatecznie wysoką akceptowalność ze strony użytkowników dwustronnego systemu komunikacji werbalnej człowiek – maszyna.

Zadania projektu LUNA obejmowały między innymi następujące zagadnienia:

- Modelowanie języka w systemie rozumienia mowy;
- Modelowanie semantyczne w systemie rozumienia mowy;
- Przenoszalność elementów systemu rozumienia mowy na inne języki.

Skrócony opis części informatycznej:

Wykorzystywanymi w projekcie metodami były głównie metody maszynowego uczenia się (dynamiczne sieci bayesowskie, CRF), wspomagane przez procedury 'active learning'.

Opis efektów projektu:

W trakcie realizacji projektu powstały prototypy rozpoznawania informacji istotnych z punktu widzenia wybranej aplikacji w telefonicznych rozmowach w języku polskim przy wykorzystaniu metod zbliżonych do tych zastosowanych dla języków francuskiego i włoskiego

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Doświadczenia uzyskane w wyniku realizacji projektu LUNA mogą być przydatne przy przetwarzaniu wszelkiego rodzaju tekstów o niezbyt standardowej strukturze - transkrypcje mowy, komentarze na forach, teksty niepoprawne gramatycznie.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

MAGIK

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Instytut Informatyki UW

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Brak.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Jerzy Tyszkiewicz jty@mimuw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Martyna Barczuk

Mateusz Konikowski

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: -

Źródło finansowania:

OPIS PROJEKTU

Typ projektu: edukacyjny, narzędziowy

Efekt projektu: oprogramowanie, 2 prace magisterskie

Słowa kluczowe: kolacja;

warianty tekstu

Skrócony opis projektu:

Powstało narzędzie do porównywania wielu wersji tego samego tekstu, wyszukujące występujące w nich różnice, czyli wykonujące kolację.

Skrócony opis części informatycznej:

Powstało narzędzie do porównywania wielu wersji tego samego tekstu, wyszukujące występujące w nich różnice, czyli wykonujące kolację.

Opis efektów projektu:

Powstało narzędzie do porównywania wielu wersji tego samego tekstu, wyszukujące występujące w nich różnice, czyli wykonujące kolację.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Dostosowanie do potrzeb innych grup użytkowników, ulepszenia samego narzędzia, wdrażanie do zastosowania badawczych i edukacyjnych.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

MEDIALABUMCS

Osoba wypełniająca ankietę: Kierownik projektu

<http://medialab.umcs.lublin.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Kulturoznawstwa UMCS

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Radosław Bomba, radoslawbomba@gmail.com

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Studenci specjalizacji cyberkultura UMCS

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 6/1/2014 -

Źródło finansowania:

OPIS PROJEKTU

Typ projektu: badawczy, edukacyjny

Efekt projektu: narzędzie badawcze, oprogramowanie, publikacja

Słowa kluczowe: humanistyka cyfrowa, analityka kulturowa, medialab cybekultura, nowe media

Skrócony opis projektu:

Pomysł na MediaLab UMCS zrodził się na zajęciach z przedmiotu Multimedia realizowanych na między kierunkowej specjalizacji Cyberkultura (utworzonej przy Instytucie Kulturoznawstwa UMCS). Istotną inspiracją dla naszych działań jest wizualizacja danych, dziennikarstwo społeczne, dziennikarstwo wykorzystujące duże ilości danych (Big data journalism), kulturomika i humanistyka cyfrowa. Zasadniczym celem przedsięwzięcia jest stworzenie otwartej przestrzeni twórczej wykorzystującej nowe media do badania, wizualizowania, analizowania i opowiadania o kulturze.

Skrócony opis części informatycznej:

MediaLab UMCS jest inicjatywą, która została zainspirowana zjawiskiem medialabów i projektów realizowanych w interdyscyplinarnych zespołach z wykorzystaniem cyfrowych technologii i sieci. Przykładem takich

przedsięwzięć może być MIT Medialab, medialab KiBU i Prado Medialab. W Polsce nie ma instytucji odpowiadającej formule medialabu, ale rokrocznie odbywają się warsztaty medialabowe organizowane przez fundację Ortus. Oprócz tego w naszym kraju powstaje coraz więcej projektów z pogranicza nauki, sztuki i technologii takich jak: THATCamp Polska, Warsztaty Kultury 2.0, Akademi Cyfrowych Kompetencji. W sieci natomiast znaleźć można liczne projekty, które inspirowane są ideami otwartej kultury i bezpłatnie kreują usieciowaną przestrzeń poświęconą sztuce, badaniom, nowym mediom, historii. Przykładem mogą tu być projekty Lublin 2.0 czy Wizualizacja sztuki.

Opis efektów projektu:

Stworzenie otwartego projektu Internetowego

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

METRONOM

Osoba wypełniająca ankietę: Kierownik projektu

<http://metronom.jtomaszewski.com/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Magistrant

Nazwa podmiotu realizującego projekt: Jacek Tomaszewski

Instytut Informatyki

Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Jerzy Tyszkiewicz jty@mimuw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Jacek Tomaszewski UW

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: -

Źródło finansowania: środki własne instytucji

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: oprogramowanie

Słowa kluczowe: łacina; metrum; poezja

Skrócony opis projektu:

Aplikacja komputerowa dopasowująca metrum do łacińskiego utworu poetyckiego

Skrócony opis części informatycznej:

Aplikacja komputerowa dopasowująca metrum do łacińskiego utworu poetyckiego

Opis efektów projektu:

Aplikacja komputerowa dopasowująca metrum do łacińskiego utworu poetyckiego

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://metronom.jtomaszewski.com/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Badania stylometryczne nad poetami łacińskimi

Zastosowania edukacyjne

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

NARODOWY KORPUS JĘZYKA POLSKIEGO

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Instytut Podstaw Informatyki PAN (lider)

Instytut Języka Polskiego PAN

Uniwersytet Łódzki

Wydawnictwo Naukowe PWN

Instytucja przy której projekt jest afiliowany: Instytut Podstaw Informatyki PAN

Pozostałe instytucje zaangażowane w projekt: j.w.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Adam Przepiórkowski, adamp@ipipan.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Zob. <http://nkjp.pl/>.

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 12/13/2007 - 6/12/2011

Źródło finansowania: Narodowe Centrum Badań i Rozwoju R17 003 03

OPIS PROJEKTU

Typ projektu: badawczy, digitalizacyjny

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja

Słowa kluczowe: korpus; narzędzia przetwarzania języka; tager; parser; znajdowanie jednostek identyfikacyjnych

Skrócony opis projektu:

Zob. <http://zil.ipipan.waw.pl/NKJP>.

Skrócony opis części informatycznej:

Zob. <http://zil.ipipan.waw.pl/NKJP>.

Opis efektów projektu:

Zob. <http://nkjp.pl/>.

Powiązania z innymi polskimi i zagranicznymi projektami:

Inne korpusy narodowe (szczególnie czeski i brytyjski).

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://nkjp.pl/>

<http://zil.ipipan.waw.pl/NKJP>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Zbieranie tekstów do NKJP zakończyło się w roku 2010. Coraz częściej pojawiają się postulaty użytkowników, by projekt kontynuować, a z NKJP uczynić korpus monitorujący bieżący rozwój polszczyzny, szczególnie w zakresie leksykalnym.

Inne naturalne kierunki rozwoju to rozszerzenie korpusu w przeszłość, czyli uczynienie z NKJP korpusu diachronicznego. Planowane jest także dodanie tekstów gwarowych.

Należałoby także zadbać o rozwój anotacji w NKJP – zarówno lingwistycznej, jak i metadanymi.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

NKJP wykorzystywany jest w wielu projektach związanych z językiem polskim. Wersja rozszerzona dalej zwiększy swoją użyteczność.

Projekt jest związany z realizowaną w IPI PAN częścią infrastruktury CLARIN-PL dotyczącą konstrukcji nowego narzędzia przeszukiwania dużych bogato znakowanych korpusów tekstów (A17).

Profil potencjalnego partnera:

- jednostki zajmujące się zbieraniem i znakowaniem tekstów historycznych
- jednostki zajmujące się zbieraniem i znakowaniem tekstów gwarowych
- partnerzy zainteresowani akwizycją i digitalizacją tekstów istniejących wyłącznie w postaci papierowej
- dostawcy technologii językowych

NARZĘDZIA DYGITALIZACJI TEKSTÓW NA POTRZEBY BADAŃ FILOLOGICZNYCH

Osoba wypełniająca ankietę: Joanna Bilińska, j.bilinska@uw.edu.pl

<https://bitbucket.org/jsbien/ndt/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Katedra Lingwistyki Formalnej,

Wydział Neofilologii,

Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Janusz S. Bień, jsbien@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 5/13/2009 - 5/12/2012

Źródło finansowania: Projekt badawczy Ministerstwa Nauki i Szkolnictwa Wyższego N N519 384036

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy, digitalizacyjny

Efekt projektu: narzędzie badawcze, oprogramowanie, publikacja

Słowa kluczowe: dygitalizacja, narzędzia, korpusy, fiszki, DjVu

Skrócony opis projektu:

Zadania projektu polegały na rozwoju narzędzi do obsługi formatu DjVu, który dominuje w polskich bibliotekach cyfrowych. Niektóre z programów zostały już wcześniej zaprojektowane i wstępnie zaimplementowane przez Jakuba Wilka pod kierunkiem Janusza S. Bienia, a niektóre zostały zaprojektowane specjalnie na potrzeby projektu. Część z nich stanowi alternatywne rozwiązania dla typowych zadań, ale niektóre mają całkowicie oryginalny charakter.

Skrócony opis części informatycznej:

Projekt miał charakter czysto informatyczny, jednak w ramach testowania stworzono zasoby (dygitalizacje i korpusy) opisane w osobnej ankiecie.

Opis efektów projektu:

Na wyniki projektu składają się następujące programy dostępne na licencji GNU GPL:

serwer korpusów dwuwarstwowych (Poliqarp for DjVu),

klient www dla serwera korpusów dwuwarstwowych (marasca),

zdalny klient graficzny dla tego serwera (DjView for Poliarp)

przeglądarka materiałów leksykograficznych (zdygitalizowanych

kartotek fiskalnych),

konwerter PDF na DjVu (pdf2djvu),

konwerter plików graficznych na DjVu (didjvu),

narzędzie do optycznego rozpoznawania znaków w dokumentach DjVu (ocrodjvu),

inne programy pomocnicze.

Powiązania z innymi polskimi i zagranicznymi projektami:

Dosyć ściśle związki z projektem Nowego Korpusu Języka Polskiego - projekt NDT bazował na serwerze NKJP, a korpus zaczął używać opracowanego przez nas klienta www.

W programie didjvu wykorzystano algorytmy opracowane w ramach projektu Gamera (<http://gamera.informatik.hsnr.de/>). W programie ocrodjvu wykorzystano m.in. wyniki projektu Tesseract (<https://code.google.com/p/tesseract-ocr/>).

Istnieją krajowe i zagraniczne o podobnych celach i zbliżonych metodach, ale ze względu na stosowanie niekompatybilnych formatów danych, nie była dotąd możliwa bezpośrednia współpraca. Najbliższy ideologicznie wydaje się projekt Paradiit (<https://sites.google.com/site/paradiitproject/>), który jednak jest zorientowany wyłącznie na użytkowników systemu Windows.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Bień, Janusz S. (2012) The IMPACT project Polish Ground-Truth texts as a DjVu corpus.

<http://bc.klf.uw.edu.pl/296/>

Bień, Janusz S. (2012) Skanowane teksty jako korpusy. Prace Filologiczne, LXIII. pp. 25-36. ISSN 0138-0567

<http://bc.klf.uw.edu.pl/322/>

Bień, Janusz S. (2010) Dygitalizacja i komputeryzacja słowników na przykładzie Słownika polszczyzny XVI wieku. In: Język polski - wczoraj, dziś, jutro. Instytut Języka Polskiego PAN i Wydawnictwo LEXIS, Kraków, pp. 131-138. ISBN 978-83-89425-61-4 <http://bc.klf.uw.edu.pl/165/>.

Bień, Janusz S. (2012) Changes to the IMPACT project Polish Ground-Truth texts. Technical Report. Katedra Lingwistyki Formalnej UW. <http://bc.klf.uw.edu.pl/288/>.

Bień, Janusz S. (2012) Delivering the IMPACT project Polish Ground-Truth texts with Poliarp for DjVu. Technical Report. Katedra Lingwistyki Formalnej UW. <http://bc.klf.uw.edu.pl/289/>.

Bień, Janusz S. (2012) Eksperymentalne korpusy dawnych tekstów polskich. In: Seminarium "Przetwarzanie języka naturalnego", Warszawa, Instytut Podstaw Informatyki PAN. <http://bc.klf.uw.edu.pl/287/>.

Bień, Janusz S. and Bilińska, Joanna A. (2012) Słownik Lindego jako korpus. In: Bibliotheca Lindiana. Samuel Bogumił Linde (1771 – 1847), pierwszy dyrektor Biblioteki Uniwersyteckiej w Warszawie. W 165 rocznicę śmierci, 19 – 20 listopada 2012 roku, Warszawa. <http://bc.klf.uw.edu.pl/324/>.

Bień, Janusz S. (2011) Dygitalizacja słowników historycznych w Katedrze Lingwistyki Formalnej UW. In: Seminarium "Przetwarzanie języka naturalnego", 10 stycznia 2011, Warszawa, Instytut Podstaw Informatyki PAN. <http://bc.klf.uw.edu.pl/176/>.

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Wszystkie narzędzia można ulepszać, ale najważniejsze byłoby umożliwienie obsługi korpusów dwuwarstwowych przez Poliqarpa 2 opracowywanego w ramach CLARIN-PL.

Wskazane byłoby rozwiązanie przynajmniej niektórych problemów niekompatybilności formatów danych, np. <https://transkrypcja.uservoice.com/forums/202346-og%C3%B3lne/suggestions/5288371-wynikowy-hocr-nie-waliduje-si%C4%99>

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Zgodnie z nieformalnymi ustaleniami Poliqarp 2 opracowywany w ramach projektu CLARIN-PL ma mieć wbudowaną obsługę korpusów dwuwarstwowych aktualnie wymagającą dodatkowego serwera (Poliqarp for DjVu). Ze względu na znacznie bogatsze możliwości Poliqarpa 2 realizacja tego ustalenia, która wymaga modyfikacji lub opracowania nowych klientów serwera oraz intensywnego testowania na już dostępnych korpusach dwuwarstwowych, jest najważniejszym kierunkiem dalszego rozwoju projektu.

Niewykluczone są powiązania z ewentualnym rozwojem niektórych narzędzi przygotowanych w ramach projektu SYNAT, takich jak: Wirtualne Laboratorium Transkrypcji (<http://wlt.synat.pcss.pl/wlt-web/index.xhtml>) oraz Wycinanki (<http://wlt.synat.pcss.pl/cutouts/>).

Profil potencjalnego partnera:

Zespoły informatyczne zainteresowane rozwojem narzędzi dygitalizacyjnych.

Humaniści będący zaawansowanymi użytkownikami korpusów zainteresowani pełnieniem roli testerów nowych narzędzi.

NAUKOWE EDYCJE CYFROWE

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Sekcja edytorska Centrum Humanistyki Cyfrowej Instytutu Badań Literackich.

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Planowany jest m.in. udział KUL.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Paweł Bem, pawel.bem@ibl.waw.pl

Osoba kontaktowa: Łukasz Cybulski, lukcybulski@gmail.com

Pozostałe osoby realizujące projekt:

Paweł Bem, IBL PAN;

Łukasz Cybulski, UKSW

INFORMACJE ORGANIZACYJNE

Status projektu: Planowany

Okres realizacji projektu: 12/15/2014 - 12/15/2017

Źródło finansowania: środki własne instytucji, Narodowe Centrum Nauki, Narodowy Program Rozwoju Humanistyki

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy, digitalizacyjny, edukacyjny

Efekt projektu: narzędzie badawcze, baza danych

Słowa kluczowe: edycja krytyczna; naukowa edycja cyfrowa; krytyka tekstu; tekstologia; edytorstwo cyfrowe

Skrócony opis projektu:

Do najbliższych celów projektu należy opracowanie teoretycznych podstaw cyfrowej edycji naukowej, które w kolejnych latach dadzą możliwość zaprojektowania modelu innowacyjnej edycji polskiego dzieła literackiego przeznaczonej do celów naukowych, wykorzystującej elektroniczne narzędzia badawcze. W pierwszym etapie prac mieścić się będzie udział w szkoleniach z zakresu wykorzystania narzędzi cyfrowych oraz organizacja spotkań roboczych i seminariów zmierzających do integracji środowiska edytorów oraz wymiany doświadczeń

związanych z pracą nad naukowymi edycjami elektronicznymi. Drugi etap stanowić będzie stworzenie edycji modelowej. Dalekosiężnym celem projektu jest budowa infrastruktury do tworzenia edycji cyfrowych.

Skrócony opis części informatycznej:

Tekst dzieła będzie zakodowany wg standardów TEI; ponadto część analityczna edycji opierać się będzie na programach open-source spełniających konkretne funkcje: kolacjonowania, porównywania wersji w ich postaci transliterowanej z postacią dokumentu, analiz stylometrycznych.

Opis efektów projektu:

Projekt zakłada udostępnienie użytkownikom edycji dokumentacji materiału przedpublikacyjnego danych tekstów oraz kolejne ich przekazy, co umożliwi bezpośredni kontakt z odmianami utworów, ukazującymi ich dynamikę i tekstologiczne komplikacje. Przekazy udostępnione będą w formie dokumentów oraz transkrypcji. Porównywaniu kolejnych przekazów służyć będą programy do kolacjonowania tekstu oraz śledzenia występujących między nimi zmian. Tak przygotowana edycja nie będzie musiała ograniczać się – jak edycje drukowane – do faworyzowania któregoś z przekazów: jej funkcją podstawową jest ukazanie wagi wszystkich przekazów tekstu oraz prezentacja wszelakich kontekstów, które mogą mieć wpływ na interpretację każdego z nich.

Powiązania z innymi polskimi i zagranicznymi projektami:

Przykładowe projekty zagraniczne: <http://cather.unl.edu>; <http://mel.hofstra.edu/index.html>;
<http://www.marktwainproject.org/homepage.html>

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Ważnym elementem edycji będzie jej otwartość na działania odbiorcy. Po udostępnieniu wydania w podstawowym kształcie planowane jest udostępnienie czytelnikom funkcji komentowania edycji, zapisywania wyników wykonanych w jej ramach prac, dzielenia się nimi itp. poprzez umożliwienie mu interakcji z wieloma elementami edycji oraz (kontrolowanego) wpływu na jej kształt.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Planowane są powiązania z projektami Ośrodka Badań Filologicznych i Edytorstwa Naukowego IBL PAN oraz CHC IBL PAN.

Profil potencjalnego partnera:

Zespoły edytorskie, repozytoria, instytucje informatyczne (budowa infrastruktury)

NEKST

Osoba wypełniająca ankietę: Osoba kontaktowa

<https://nekst.ipipan.waw.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Instytut Podstaw Informatyki PAN

Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Instytut Podstaw Informatyki PAN

Pozostałe instytucje zaangażowane w projekt: j.w.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Jacek Koronacki, korona@ipipan.waw.pl

Osoba kontaktowa: Adam Przepiórkowski, adamp@ipipan.waw.pl

Pozostałe osoby realizujące projekt:

Duże zespoły w Instytucie Podstaw Informatyki PAN (kierowane przez Mieczysława Kłopotka i Adama Przepiórkowskiego) i na Politechnice Wrocławskiej (kierowane przez Macieja Piaseckiego i Halinę Kwaśnicką).

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 4/1/2009 - 8/10/2014

Źródło finansowania: Program Operacyjny "Innowacyjna Gospodarka"

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, repozytorium, publikacja

Słowa kluczowe: przeszukiwanie polskiego Internetu; odpowiadanie na pytania; technologie językowe, podstawowe zasoby i narzędzia językowe dla języka polskiego

Skrócony opis projektu:

Zob. <http://zil.ipipan.waw.pl/NEKST>

Skrócony opis części inżynierskiej:

Zob. <http://zil.ipipan.waw.pl/NEKST>

W ramach projektu opracowano również szereg podstawowych narzędzi (programów) do analizy tekstu w języku polskim, które mają szerokie zastosowanie w różnorodnych aplikacjach badawczych.

Opis efektów projektu:

IPI PAN: narzędzia przeszukiwania polskiego internetu oraz wielorakie narzędzia i zasoby lingwistyczne: słownik walencyjny, gramatyki formalne i oparte na nich parsery, narzędzia analizy wydźwięku, narzędzia związane z jednostkami wielowyrzowymi, metody przenoszenia informacji gramatycznych pomiędzy językami i in.

Grupa Naukowa G4.19, Inst. Informatyki PWr.: podstawowe zasoby dla języka polskiego (Słowosieć - wielka leksykalna sieć semantyczna, słownik 1,5 mln. polskich nazw własnych, Korpus PWr. języka polskiego opisany meta-danymi na kilku poziomach analizy języka), podstawowe narzędzia dla języka polskiego (segmentacja tekstu, tager - program do analizy morfo-syntaktycznej, program do rozpoznawania i klasyfikacji nazw własnych, systemy do wydobywania z tekstu relacji semantycznych), SuperMatrix - system do płytkiej, statystycznej analizy semantycznej korpusu tekstów, WordnetWeaver - system do półautomatycznej konstrukcji leksykalnych sieci semantycznych, Borsuk - system do odpowiadania na pytania w języku polskim, treningowy korpus pytań i odpowiedzi w języku polskim.

Zakład Sztucznej Inteligencji, Inst. Inf. PWr.: system do wyszukiwania obrazów podobnych do zadanego przez użytkownika, system do automatycznej, semantycznej klasyfikacji obrazów i ich anotowania za pomocą słów kluczowych, program do grupowania obrazów na podstawie ich podobieństwa, program do wydobywania wybranych informacji z obrazów sztucznych (np. map lub wykresów).

Wszystkie narzędzia, zasoby, programy i systemy opracowane dla przetwarzania tekstów są dostępne na otwartych licencjach.

Powiązania z innymi polskimi i zagranicznymi projektami:

Prace oparte na wynikach wcześniejszych projektów realizowanych w IPI PAN i na Politechnice Wrocławskiej, w szczególności na Narodowym Korpusie Języka Polskiego i na Słowosieci.

Prace częściowo kontynuowane w ramach projektów SyNaT i CLARIN-PL.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://zil.ipipan.waw.pl/Walenty>

<http://zil.ipipan.waw.pl/Skladnica>

<http://zil.ipipan.waw.pl/LFG>

<http://zil.ipipan.waw.pl/SloownikWydzwieku>

<http://nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/narzedzia-przetwarzania-morfosyntaktycznego>

<http://nlp.pwr.wroc.pl/narzedzia-i-zasoby/maca>

<http://nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/wcrft-tagger>

<http://nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/lexcsd>

<http://nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/liner2>

<http://plwordnet.pwr.wroc.pl>

<http://nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/kpwr>

<http://nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/nelexicon>

<http://nlp.pwr.wroc.pl/pl/narzedzia-i-zasoby/wordnetloom>

<http://nlp.pwr.wroc.pl/narzedzia-i-zasoby/supermatrix>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Projekt częściowo rozwijany w ramach CLARIN-PL.

Opracowane narzędzia i zasoby zostały już wykorzystane w aplikacjach badawczych z dziedziny nauk humanistycznych. Mogą znaleźć szereg dalszych zastosowań.

System do odpowiadania na pytania w języku polskim może stać się podstawą do budowy inteligentnego systemu semantycznego wyszukiwania informacji w korpusach i zbiorach materiałów źródłowych. System SuperMatrix może być wykorzystany do analizy słownictwa i terminologii, wydobywania dziedzinowych leksykalnych sieci semantycznych oraz analizy podobieństwa semantycznego dokumentów i fragmentów tekstu.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Współpraca jest już realizowana w ramach CLARIN ERIC.

Profil potencjalnego partnera:

Opracowane narzędzia, zasoby i systemy stanowią bardzo szeroką ofertę dla wielu instytucji badawczych z dziedziny nauk humanistycznych i sztuki.

OKRES WĘDRÓWEK LUDÓW MIĘDZY ODRĄ A WISŁĄ

Osoba wypełniająca ankietę: Kierownik projektu

www.mpov.uw.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Instytut Archeologii, Wydział Historyczny, Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Uniwersytet Jagielloński, Uniwersytet Rzeszowski, Uniwersytet Łódzki, Uniwersytet Gdański, Muzeum Narodowe w Szczecinie, Państwowe Muzeum Archeologiczne w Warszawie

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Aleksander Bursche abursche@yahoo.com

Osoba kontaktowa: Anna Zapolska anna.zapolska@gmail.com

Pozostałe osoby realizujące projekt:

cf <http://www.mpov.uw.edu.pl/pl/wykonawcy>

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 5/7/2012 - 5/7/2017

Źródło finansowania: Narodowe Centrum Nauki 2011/02/A/HS3/00389

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: publikacja, nowa perspektywa badawcza przeszłości ziem polskich, liczne opracowania i wystawy

Słowa kluczowe: Germanie, Słowianie, migracje, kontakty interkulturowe, paleobotanika

Skrócony opis projektu:

<http://www.mpov.uw.edu.pl/pl/projekt>

Skrócony opis części informatycznej:

dokumentacja badań terenowych przy wykorzystaniu skanerów 3-D i fotogrametrii, ich wizualizacja w formacie 3-D, dokumentacja i wizualizacja 3-D zabytków, dokumentacja, analiza porównawcza i kartografowanie w google-maps profili palynologicznych, komputerowa analiza wyników spektrograficznych analiz składu chemicznego zabytków

Opis efektów projektu:

<http://www.mpov.uw.edu.pl/pl/projekt/rezultaty-naukowe>

<http://www.mpov.uw.edu.pl/pl/projekt/oddziaływanie-spoeczne>

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekty palynologiczne realizowane przez kilka uniwersyteckich wydziałów biologii, projekty metalograficzne realizowane przez kilka wydziałów chemii i inne instytucje np. IBJ

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.mpov.uw.edu.pl/pl/badania>

<http://www.mpov.uw.edu.pl/pl/publikacje>

<http://www.mpov.uw.edu.pl/pl/konferencje>

<http://www.mpov.uw.edu.pl/pl/aktualnosc>

<http://www.mpov.uw.edu.pl/pl/sterujacy>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Kontynuacja portalu projektu,

Dalsze wizualizacje 3-D na potrzeby portalu, publikacji i wystaw

Uwzględnienie kolejnych profili palynologicznych w bazie danych i na mapach

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

De facto ze wszystkimi projektami realizującymi dokumentacje i wizualizacje terenową i materiałową w formacie 3-D

Profil potencjalnego partnera:

Instytucje posługujące się zaawansowanymi

OPRACOWANIE NARZĘDZI DO EKSTRAKCJI INFORMACJI Z TEKSTÓW W JĘZYKU POLSKIM

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: pracownicy Zespołu Inżynierii Lingwistycznej IPI PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Agnieszka Mykowiecka, agn@ipipan.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Agnieszka Mykowiecka, IPI PAN; Małgorzata Marciniak, IPI PAN; Anna Kupść, IPI PAN, Jakub Piskorski, DFKI, Saarbruecken;

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 10/20/2004 - 10/19/2007

Źródło finansowania: Komitet Badań Naukowych 3 T11C 007 27

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, publikacja

Słowa kluczowe: ekstrakcja informacji; systemy regułowe

Skrócony opis projektu:

Celem projektu było stworzenie narzędzi pozwalających na ekstrakcję informacji z nieustrukturalizowanych tekstów w języku polskim z wykorzystaniem metod analizy powierzchniowej, wzbogaconej o pewne elementy analizy głębokiej budowy i znaczenia zdań. Projekt łączył w sobie dwa cele. Pierwszym z nich było stworzenie prototypu systemu ekstrakcji informacji z tekstów dla pewnego konkretnego zastosowania. Drugi cel, to włączenie do budowanego systemu elementów dokładnej, formalnej analizy znaczenia przetwarzanych zdań. Wybrane na potrzeby stworzenia prototypu zastosowanie to analiza tekstów medycznych gromadzonych w archiwum szpitalnym i zawierających najważniejsze informacje o przebiegu choroby i leczenia pacjentów oraz opis słowny wyników badań.

Skrócony opis części informatycznej:

Systemy do ekstrakcji informacji dziedzinowej oraz ogólne narzędzia do rozpoznawania nazw własnych osób i instytucji opracowywano wykorzystując platformę SProUT (<http://sprout.dfki.de/>).

Opis efektów projektu:

Efektom projektu były dwa prototypy systemów ekstrakcji informacji z tekstów medycznych oraz program do rozpoznawania nazw własnych osób i instytucji wraz z listą takich nazw.

Narzędzia do rozpoznawania nazw własnych wykorzystane były później w projekcie NKJP.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Prace wykonywane w tym projekcie, kontynuowane potem w projekcie Nekst, mogą stanowić punkt wyjścia do realizacji projektów, których celem będzie opracowanie metod wspomagania budowy ontologii dziedzinowych i ekstrakcji z tekstu opisów zdarzeń czy innego typu zależności złożonych.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

OTWARTE REPOZYTORIUM NAUK HISTORYCZNYCH LECTORIUM

Osoba wypełniająca ankietę: Kierownik projektu

www.lectorium.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Wydział Historyczny Uniwersytetu Warszawskiego

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Centrum Modelowania Matematycznego i Komputerowego UW (ICM)

CEON (ICM UW)

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Dominik Purchała, d.purchala@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Zespół WH UW oraz ICM UW (CEON)

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 9/26/2013 -

Źródło finansowania: dotacja celowa

OPIS PROJEKTU

Typ projektu: bazodanowy, digitalizacyjny

Efekt projektu: repozytorium

Słowa kluczowe: otwarta nauka; repozytorium; Open Access; publikacje;

Skrócony opis projektu:

Otwarte Repozytorium Nauk Historycznych „Lectorium” to dziedzinowe repozytorium naukowe, umożliwiające przechowywanie i publiczne udostępnianie dorobku polskich uczonych zajmujących się naukami historycznymi i pokrewnymi. Materiały w Lectorium prezentowane są w sześciu głównych działach tematycznych: historia, archeologia, historia sztuki, bibliotekoznawstwo i informacja naukowa, etnologia i antropologia kulturowa oraz muzykologia.

Lectorium powstało na wzór najbardziej znanych tego typu repozytoriów funkcjonujących w naukach ścisłych. Jego celem jest zwiększenie widoczności prac polskich humanistów oraz wzrost efektywności komunikacji naukowej poprzez umożliwienie łatwiejszego i szybszego zapoznawania się z najnowszymi wynikami badań.

Repozytorium działa w ramach klasycznego modelu opartego o samoarchiwizację (self-archiving) materiałów przez autorów, ale umożliwia także dostarczanie treści przez instytucje.

Skrócony opis części informatycznej:

Repozytorium korzysta z rozwiązań informatycznych przygotowanych przez Centrum Otwartej Nauki ICM (CeON) oraz Wydział Historyczny UW w opraciu o otwarte oprogramowanie repozytoryjne dSpace oraz CMS o otwartym kodzie źródłowym – Wordpress. Metadane oparte są na standardzie Dublin Core. Implementowany jest otwarty standard wymiany metadanych OAI-PMH.

Opis efektów projektu:

Upowszechnienie wyników prowadzonych badań poprzez dziedzinowe repozytorium powinno przełożyć się na wzrost cytowalności i oddziaływania opublikowanych prac. Lectorium zapewnia widoczność w największej wyszukiwarce tekstów naukowych - Google Scholar, która jest źródłem danych m. in. dla baz takich jak Web of Science czy Scopus oraz programów typu "Publish or Perish", będących źródłem informacji o indeksie Hirscha.

Repozytorium umożliwia gromadzenie i prezentowanie następujących zasobów:

- artykuły opublikowane w czasopismach naukowych oraz popularno-naukowych;
- preprinty i postprinty artykułów naukowych;
- rozprawy i dysertacje stanowiące podstawę uzyskania tytułów zawodowych lub stopni naukowych;
- monografie, prace zbiorowe lub rozdziały książek naukowych;
- materiały konferencyjne (w tym prezentacje) oraz materiały typu working papers;
- materiały będące spuścizną wybitnych naukowców (materiały robocze, zapiski, niedokończone teksty, konspekty, kwerendy, itp.);
- raporty i cząstkowe monografie, powstające w wyniku prowadzonych projektów/grantów naukowych;
- bazy danych w formatach możliwych do otwarcia na komputerze czytelnika;
- materiały graficzne (np. dokumentacja zdjęciowa wykopalisk);
- materiały multimedialne (dokumentacja video, nagrania wykładów, itp.).

Materiały można deponować w najpopularniejszych formatach tekstowych, graficznych i multimedialnych. Wszystkie materiały w repozytorium można deponować na jednej z sześciu licencji Creative Commons (zalecamy C-BY lub CC-BY-SA) oraz na zasadzie dozwolonego użytku. Korzystanie z repozytorium jest całkowicie darmowe.

Powiązania z innymi polskimi i zagranicznymi projektami:

Otwarte Repozytorium Nauk Historycznych „Lectorium” może współpracować na zasadzie wymiany metadanych z dowolnym repozytorium instytucjonalnym obsługującym protokół wymiany metadanych OAI-PMH. Repozytorium współpracuje m.in. z serwisem Otwórz Książkę oraz repozytoriami instytucjonalnymi polskich uczelni wyższych.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Nawiązanie współpracy z polskimi uczelniami wyższymi oraz istniejącymi repozytoriami instytucjonalnymi. Rozwijanie projektu w kierunku ułatwiania deponowania materiałów, zarządzania kolekcjami oraz zabezpieczania długoterminowego przechowywania i udostępniania materiałów.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Głównym założeniem serwisu jest otwarcie na inne instytucje naukowe i badawcze. Wszystkie jednostki, które przyłączą się do repozytorium będą reprezentowane na takich samych prawach i warunkach.

Oferta współpracy skierowana jest przede wszystkim do jednostek uniwersyteckich oraz instytucji naukowych. Nowy członek Lectorium uzyska możliwość lokalnego administrowania i redagowania zasobami swojej jednostki, łącznie z możliwością umieszczania materiałów swoich pracowników w ich imieniu. Jest to korzystne rozwiązanie dla instytucji nieposiadających repozytoriów instytucjonalnych.

Dla ośrodków prowadzących repozytoria instytucjonalne istnieje możliwość nawiązania współpracy dzięki implementowanemu protokołowi OAI-PMH umożliwiającemu wymianę metadanych z innymi - polskimi i zagranicznymi - zasobami sieciowymi.

W przypadku współpracy z istniejącym repozytorium pliki będą pobierane przez czytelników Lectorium z oryginalnej lokalizacji. Pomoże to zwiększyć statystyki repozytorium instytucjonalnego przy jednoczesnym uzyskaniu reprezentacji opisów bibliograficznych i metadanych także w repozytorium dziedzinowym.

Profil potencjalnego partnera:

Uczelnie wyższe – wydziały humanistyczne, instytucje naukowe, zespoły badawcze, istniejące repozytoria naukowe, biblioteki cyfrowe.

OTWÓRZ KSIĄŻKĘ

Osoba wypełniająca ankietę: Kierownik projektu

<http://otworzksiazke.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego, Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: nie dotyczy

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Jakub Szprot

Osoba kontaktowa: Joanna Raczkowska, j.raczkowska@icm.edu.pl

Pozostałe osoby realizujące projekt:

Joanna Raczkowska, ICM UW

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: -

Źródło finansowania:

OPIS PROJEKTU

Typ projektu: digitalizacyjny

Efekt projektu: kolekcja książek naukowych

Słowa kluczowe: książki; monografie; digitalizacja; otwarty dostęp

Skrócony opis projektu:

Otwórz Książkę to cyfrowa kolekcja książek naukowych polskich uczonych, przede wszystkim z obszaru nauk humanistycznych i społecznych (m.in. bibliologia i informatologia, filozofia, historia, historia sztuki, językoznawstwo, kulturoznawstwo, literaturoznawstwo, religioznawstwo, nauki o rodzinie, nauki o zarządzaniu, nauki o polityce, nauki o polityce publicznej, nauki o poznaniu i komunikacji społecznej, pedagogika, psychologia, socjologia, ekonomia, finanse, nauki o administracji, prawo). Książki można przeszukiwać, czytać online oraz pobierać w postaci plików PDF. Serwis działa w ramach Platformy Otwartej Nauki (pon.edu.pl). Patronat honorowy nad serwisem objęła Polska Akademia Nauk.

Skrócony opis części informatycznej:

Serwis Otwórz Książkę wykorzystuje oprogramowanie przygotowane na jego potrzeby w ICM UW.

Opis efektów projektu:

Efektem projektu jest kolekcja książek naukowych w postaci cyfrowej. Wszystkie książki są udostępniane w sposób otwarty (na licencjach Creative Commons albo do wykorzystania na zasadzie dozwolonego użytku).

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Planowana jest rozbudowa kolekcji książek o nowe tytuły.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Współpraca z innymi projektami może polegać na bezpłatnym i otwartym udostępnianiu w serwisie Otwórz Książkę publikacji książkowych powstałych w ramach tych projektów.

Profil potencjalnego partnera:

PARSEME

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://parseme.eu/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Instytut Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: W projekt (a właściwie coś w rodzaju sieć doskonałości) zaangażowane są oficjalnie nie instytucje, lecz kraje (30):

http://www.cost.eu/domains_actions/ict/Actions/IC1207?parties

i poszczególne osoby (ponad 100).

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Agata Savary

Osoba kontaktowa: Adam Przepiórkowski, adamp@ipipan.waw.pl

Pozostałe osoby realizujące projekt:

- Agata Savary (U. Tour, Francja) – lider
 - Adam Przepiórkowski (IPI PAN) – wicelider i przedstawiciel „grant holdera” (IPI PAN)
-

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 3/8/2013 - 3/7/2017

Źródło finansowania: COST Action (European Science Foundation) IC1207

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: publikacja

Słowa kluczowe: jednostki wielowyrazowe; korpusy; parsery; gramatyki

Skrócony opis projektu:

Projekt dotyczy roli jednostek wielowyrazowych (nazw własnych, jednostek frazeologicznych itp.) w gramatykach, w korpusach i w systemach przetwarzania języka naturalnego.

Zob. <http://typo.uni-konstanz.de/parseme/index.php/the-action>

Skrócony opis części informatycznej:

Sieć doskonałości – w zasadzie brak „części informatycznej”, choć partnerzy uczestniczą w projekcie często w związku z rozwojem swoich narzędzi informatycznych (np. parserów składniowych).

Opis efektów projektu:

Wymiana doświadczeń, raporty opisujące „state of the art” i kierunki rozwoju itp.

Powiązania z innymi polskimi i zagranicznymi projektami:

PARSEME ma bezpośredni związek z prowadzonymi w IPI PAN pracami nad systemami przetwarzania składniowo-semantycznego.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://parseme.eu/> =

<http://typo.uni-konstanz.de/parseme/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

http://www.cost.eu/domains_actions/ict/Actions/IC1207

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Liczne powiązania już istnieją, projekt ma jednak charakter otwarty.

Profil potencjalnego partnera:

Naukowcy zainteresowani formalnym opisem jednostek wielowrazowych i wykorzystaniem takiego opisu w systemach przetwarzania tekstów.

POGROMY. PRZEMOC KOLEKTYWNA WOBEC ŻYDÓW NA ZIEMIACH POLSKICH W XIX I XX WIEKU I JEJ WPŁYW NA RELACJE POLSKO - ŻYDOWSKIEJ. HISTORIA, PAMIĘĆ, TOŻSAMOŚĆ

Osoba wypełniająca ankietę: Kierownik projektu

<http://www.ihuw.pl/instytut/dzialalnosc/projekty/pogromy-przemoc-kolektywna-wobec-zydow-na-ziemiach-polskich-w-xix-xx-wieku-i-jej-wplyw>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Historyczny UW przy współpracy z TAU i UCL

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: TAU, UCL

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Artur Markowski, amarkowski@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

blisko 60 osób z 7 krajów (trudno wymieniać wszystkich)

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 3/4/2013 - 3/4/2016

Źródło finansowania: Narodowy Program Rozwoju Humanistyki

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, baza danych, publikacja

Słowa kluczowe: pogromy, przemoc, Żydzi, ofiary, antysemityzm

Skrócony opis projektu:

Celem projektu jest zbadanie i przedstawienie, w formie trzyltomowego wydawnictwa, przejawów kolektywnej przemocy wymierzonej w Żydów zamieszkujących ziemię w XIX i XX wieku, i jej wpływu na kształtowanie się nowoczesnej wspólnoty narodowej (zarówno polskiej jak i żydowskiej) oraz na wzajemne relacje tych dwóch grup etnicznych. Zagadnienie pogromów nie ma bowiem tylko charakteru historycznego jest przecież ciągle obecne w sferze wyobrażeń o Polsce i Polakach, i staje się przedmiotem rozliczeń przy okazji kolejnych debat nad przeszłością. Nasz projekt winien rozwiązać wielu mitów i usprawnić dementowanie fałszywych informacji funkcjonujących w obiegu naukowym.

Projekt ma charakter interdyscyplinarny (historycy, politolodzy, socjologowie, literaturoznawcy, historycy sztuki) oraz międzynarodowy (skupia badaczy z kilkunastu ośrodków na świecie). To ogromne przedsięwzięcie nie mające swego odpowiednika w polskiej literaturze przedmiotu. Instytut Historyczny realizuje ten projekt we współpracy z Department of Hebrew and Jewish Studies University College London i The Institute for the History of Polish Jewry and Israel – Poland Relations, The Goldstein – Goren Diaspora Research Center, Tel Aviv University.

Skrócony opis części informatycznej:

Baza danych stanowiąca zarówno encyklopedię pogromów jak i analityczne narzędzie badawcze do ustaleń kwantytatywnych rozmaitych problemów natury ogólnej (np. profilu ofiar). Zebranie wskazówek bibliograficznych w powiązaniu z konkretnymi przypadkami i kumulacja danych umożliwia tworzenie map.

Opis efektów projektu:

trzytomowe wydawnictwo naukowe kompleksowo analizujące problematykę pogromów Żydów na ziemiach polskich w XIX i XX wieku, baza danych (encyklopedia) pogromów i archiwum pogromowe w bibliotece IH UW

Powiązania z innymi polskimi i zagranicznymi projektami:

projekt ma charakter unikatowy

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

brak

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

rozszerzenie działalności w kierunku edycji źródeł i poszerzenie obszaru badawczego na inne państwa europejskie

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

z projektami realizowanymi przez prof. Tokarską - Bakir

Profil potencjalnego partnera:

Instytucja naukowo-badawcza z dziedziny historii lub nauk społecznych posiadająca wykwalifikowanych pracowników naukowych zajmujących się problematyką przemocy, antysemityzmu lub historią Żydów.

POLONIJNA BIBLIOTEKA CYFROWA

Osoba wypełniająca ankietę: Kierownik projektu

pbc.uw.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Informacji Naukowej i Studiów Bibliologicznych UW

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Ministerstwo Spraw Zagranicznych RP

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Zdzisław Dobrowolski, z.dobrowolski@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Zdzisław Dobrowolski, UW

Robert Brzóska, UW

Michał Krupa, UW

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 8/1/2012 - 12/31/2064

Źródło finansowania: środki własne instytucji, Ministerstwo Spraw Zagranicznych

OPIS PROJEKTU

Typ projektu: infrastrukturalny, bazodanowy, digitalizacyjny

Efekt projektu: baza danych, repozytorium, publikacja

Słowa kluczowe: Polonia ; archiwa rodzinne ; czasopisma polonijne ; polonika

Skrócony opis projektu:

Digitalizacja archiwów rodzinnych Polaków za granicą i wydawnictw polonijnych

Skrócony opis części informatycznej:

Polonijna Biblioteka Cyfrowa oparta jest na programie EPRINTS 3.3

Opis efektów projektu:

Planujemy systematyczne uzupełnianie zasobu o nowe materiały pozyskiwane w czasie badań terenowych i kontaktów zagranicznych z Polonią

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

POLSCY PISARZE I BADACZE LITERATURY PRZEŁOMU XX I XXI WIEKU. SŁOWNIK
BIOBIBLIOGRAFICZNY

Osoba wypełniająca ankietę: Osoba kontaktowa

polscpisarze.ibl.waw.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Pracownia Dokumentacji Literatury Współczesnej

Instytut Badań Literackich Polskiej Akademii Nauk

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Ministerstwo Nauki i Szkolnictwa Wyższego

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Alicja Szałagan, alicja.szalagan@ibl.waw.pl

Osoba kontaktowa: Marlena Sęczek, marlena.seczek@ibl.waw.pl

Pozostałe osoby realizujące projekt:

Beata Dorosz IBL PAN; Marlena Sęczek IBL PAN; Ewa Głębińska IBL PAN; Barbara Tyszkiewicz IBL PAN; Maria Kotowska-Kachel IBL PAN; Barbara Marzęcka IBL PAN; Katarzyna Batora IBL PAN; Alicja Szałagan IBL PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 2/1/2010 - 2/1/2017

Źródło finansowania: Narodowy Program Rozwoju Humanistyki 0012/FNiTP/H11/80/2011

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: baza danych, publikacja

Słowa kluczowe: biografia; bibliografia; współczesność; słownik; biobibliografia

Skrócony opis projektu:

Celem projektu jest dostarczenie podstawowej wiedzy faktograficznej o nowym pokoleniu twórców – poetów, prozaików, dramatopisarzy, krytyków literackich, tłumaczy oraz literaturoznawców.

Trójczłonowe hasła osobowe składają się z biogramu w formie zwięzłego szkicu, podającego najistotniejsze dane o życiu autora, z bibliografii twórczości oraz wykazu ważniejszych opracowań. Przygotowywane tomy ukazują się najpierw drukiem, a następnie w formie elektronicznej.

Projekt jest tworzony dla środowiska polonistycznego jako uzupełnienie warsztatu do studiów nad literaturą przełomu XX i XXI wieku. Podstawą jego koncepcji, stosowaną już we wcześniejszych seriach słowników pisarzy współczesnych przygotowywanych w IBL PAN jest operowanie tylko faktami, które nie podlegają interpretacji. Podstawę opracowaniu biogramów stanowią odpowiedzi na rozesłaną ankietę. Część bibliograficzna sporządzana na podstawie źródeł pośrednich oraz autopsji autorek haseł zachowuje najwyższe standardy rzetelności.

Skrócony opis części informatycznej:

polskypisarze.ibl.waw.pl to strona internetowa słownika, na której sukcesywnie zamieszczone są hasła opublikowane najpierw tradycyjnie w postaci tomu drukowanego. Wersja internetowa słownika stworzona jest na softwarze Mediawiki, zawiera multiwyszukiwarkę i narzędzia ułatwiające nawigację po poszczególnych hasłach. Co pół roku wszystkie dane (biograficzne i bibliograficzne) na stronie są aktualizowane.

Opis efektów projektu:

Jak dotąd wyszedł drukiem tom 1. słownika zawierający 90 haseł osobowych pisarzy i badaczy literatury. Ten tom w ramach projektu został zamieszczony na stronie internetowej. Dostęp do strony jest darmowy i nie wymaga logowania.

Tom 2. słownika wyjdzie drukiem pod koniec 2014, a na początku 2015 będzie dostępny online.

Tom 3. planowany jest na 2016 - druk, początek 2017 - online.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Po zakończeniu projektu planowane są prace prowadzące do udostępnienia online 10-tomowego słownika Współcześni polscy pisarze i badacze literatury (wyd. 1993-2007), zawierającego ponad 2000 haseł osobowych pisarzy i badaczy literatury, rozwijających twórczość w latach 1918-1989. Konieczne jest uaktualnienie haseł i opracowanie funkcjonalnej platformy internetowej do ich zamieszczania. Przysłuży się to nie tylko do popularyzacji ważnych dla współczesności nazwisk, lecz także do upowszechniania rzetelnych, sprawdzonych źródeł wiedzy.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

POLSKA BIBLIOGRAFIA LITERACKA - LABORATORIUM WIEDZY O WSPÓŁCZESNEJ KULTURZE
POLSKIEJ

Osoba wypełniająca ankietę: Kierownik projektu

<http://pbl.ibl.poznan.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Pracownia Bibliografii Bieżącej i Centrum Humanistyki Cyfrowej,
Instytut Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Poznańskie Centrum Superkomputerowo-Sieciowe, afiliowane
przy Instytucie Chemii Bioorganicznej PAN

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Maryl, maciej.maryl@ibl.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Maciej Maryl, IBL PAN; Jerzy Kandziora, IBL PAN; Piotr Wciślik, CHC IBL PAN / OSA; Jacek Biesiada, IBL PAN;
Beata Domosławska, IBL PAN; Małgorzata Szkudlarska, IBL PAN; Beata Latosińska, IBL PAN; Zyta Szymańska, IBL
PAN; Aleksandra Włoszczyńska, IBL PAN; Stanisław Wróbel, IBL PAN; Izabela Hojdis, IBL PAN; Anna Ziomek-
Miarkowska, IBL PAN.

Marcin Werla, PCSS; Cezary Mazurek, PCSS;

INFORMACJE ORGANIZACYJNE

Status projektu: Planowany

Okres realizacji projektu: 9/1/2014 - 9/1/2019

Źródło finansowania: Narodowy Program Rozwoju Humanistyki (Numer jeszcze nie wyznaczony)

OPIS PROJEKTU

Typ projektu: bazodanowy, digitalizacyjny

Efekt projektu: narzędzie badawcze, baza danych

Słowa kluczowe: bibliografia; baza danych; infrastruktura badawcza; życie literackie; historia literatury i kultury
polskiej

Skrócony opis projektu:

Celem projektu jest modernizacja i rozwój Polskiej Bibliografii Literackiej (PBL), rozumiane jako przeniesienie wypróbowanej metodologii PBL w realia nowych technologii: usprawnienie sposobu pozyskiwania i agregowania danych, digitalizacja i aktualizacja bazy bibliograficznej oraz budowa interfejsu badawczego wspomagającego pracę z bazą (narzędzia do eksploracji, wizualizacji i eksportowania danych). W toku projektu powstanie nowoczesne narzędzie naukowe, które pozwoli badaczom literatury i kultury polskiej na pogłębiony i kompleksowy dostęp do kompendium kultury za lata 1939-2009. Służyć temu będzie zarówno digitalizacja, opracowanie i włączenie w obręb bazy archiwalnych roczników PBL (1944-1988) i bibliografii sporządzonych według zbliżonej metodologii (Drugi Obieg, bibliografia literatury czasu II wojny światowej), jak i aktualizacja danych współczesnych (za okres 2000-2009). Konieczna jest także ewaluacja i rewizja metodologii zarówno w zakresie pozyskiwania danych, jak i ich klasyfikacji, która powinna odpowiadać na potrzeby współczesnych badań literackich i, szerzej, humanistycznych.

Skrócony opis części informatycznej:

Badacze literatury i kultury polskiej będą mogli skorzystać z kompletnej bazy bibliograficznej, obejmującej literaturę i sztuki pokrewne za lata 1939-2009 (ok. 3,5 miliona rekordów), która łączy w unikalny sposób różnorodne typy materiałów: utwory literackie, książki, czasopisma, dokumenty życia literackiego, przedstawienia teatralne, audycje radiowe, filmy, programy telewizyjne, a docelowo także internetowe formy życia literackiego (portale, czasopisma internetowe, fora, blogi, itp.), zaś interfejs badawczy pozwoli na eksplorację materiału i odkrywanie powiązań między danymi.

Cechą szczególną projektu jest jego otwartość, którą należy rozumieć dwojako. Po pierwsze, zaktualizowane dane będą udostępniane online w formacie „Linked Open Data” – czyli w formacie umożliwiającym relacyjne łączenie danych w sposób zrozumiały dla maszyn, co znacznie ułatwia ich ponowne użycie, a zwłaszcza poszerza ich potencjał badawczy (podobne narzędzia stosuje np. Europeana, Biblioteka Kongresu USA czy też British Library). Po drugie, projekt zakłada budowę narzędzi przyjaznych dla użytkowników, uwzględniających ich punkt widzenia już na etapie prac nad projektem.

Opis efektów projektu:

Efektom projektu będzie laboratorium wiedzy o współczesnej kulturze polskiej – narzędzie badawcze wykorzystujące o dane bibliograficzne PBL i opracowań pokrewnych (Drugi Obieg, II wojna światowa) za lata 1939-2009, obejmująca ponad 3 miliony rekordów. Użytkownicy (badacze literatury i kultury współczesnej) będą mogli zarówno w sposób tradycyjny docierać do konkretnych dzieł i ich treści (linki do repozytoriów cyfrowych) przez wyszukiwarkę, jak i prowadzić własne badania przy użyciu specjalistycznych narzędzi do zestawiania danych z bazy i wizualizacji tychże połączeń.

Powiązania z innymi polskimi i zagranicznymi projektami:

PBL, jako specjalistyczna, literacka baza bibliograficzna pełni podobne funkcje jak inne narodowe projekty tego typu, np. ABELL (Annual Bibliography of English Language and Literature), BDSL (Bibliographie der deutschen Sprach- und Literaturwissenschaft), czy BLE (Bibliografía de la Literatura Española), które jednak korzystają z prostych interfejsów bibliotecznych.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Narzędzie opracowane w projekcie można znacznie rozwinąć, czyniąc z niego uniwersalną infrastrukturę do wprowadzania różnorodnych treści bibliograficznych (z innych bibliografii) lub innych, podobnych materiałów.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Współpraca może dotyczyć zarówno dostawców treści (zespoły bibliograficzne i archiwalne), jak i partnerów informatycznych, którzy mogą pomóc rozwinąć bazę.

Profil potencjalnego partnera:

CLARIN - w zakresie automatycznego przetwarzania zdigitalizowanych rekordów.

Zespoły dokumentacyjne z zakresu badań literackich i kulturowych.

Biblioteka Narodowa (Bibliografie).

POLSKIE CZASOPISMA SLAWISTYCZNE ONLINE. WDROŻENIE INNOWACYJNEGO SYSTEMU PRACY I OPROGRAMOWANIA OPEN JOURNAL SYSTEMS

Osoba wypełniająca ankietę: Osoba kontaktowa

<https://ispan.waw.pl/journals/index.php>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Sławistyki PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: dr hab. Anna Engelking, prof. IS PAN, engelking@ispan.waw.pl

Osoba kontaktowa: Ewa Wróblewska-Trochimiuk, wroblewskae@gmail.com

Pozostałe osoby realizujące projekt:

dr hab. Anna Engelking, prof IS PAN; dr Dorota Leśniewska, IS PAN; dr Wojciech Sosnowski, IS PAN; mgr Piotr Otręba, Polskie Towarzystwo Botaniczne, dr Dorota Pazio, IS PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 12/13/2012 - 12/13/2014

Źródło finansowania: Ministerstwo Nauki i Szkolnictwa Wyższego IxP

OPIS PROJEKTU

Typ projektu: infrastrukturalny, digitalizacyjny

Efekt projektu: narzędzie badawcze, repozytorium

Słowa kluczowe: platforma, czasopisma, sławistyczne,

Skrócony opis projektu:

Projekt polega na wprowadzeniu czasopism wydawanych w IS PAN na platformę Open Journal Systems (OJS). Projekt „Polskie czasopisma sławistyczne online. Wdrożenie innowacyjnego systemu pracy i oprogramowania Open Journal Systems” ma na celu transformację ośmiu czasopism Instytutu Sławistyki PAN do postaci elektronicznej i udostępnienie ich online w modelu otwartym (open access).

Skrócony opis części informatycznej:

Czasopisma zamieszczane są na platformie Open Journal Systems (OJS), która jest systemem publikacji i zarządzania czasopismami naukowymi oraz artykułami. Powstała w ramach projektu Public Knowledge Project

jako open source. System ten umożliwia nie tylko odpowiednie zarządzanie procesem redakcyjnym, ale także łatwy i szeroki dostęp do publikacji naukowych online.

Opis efektów projektu:

Wprowadzenie na platformę i usprawnienie recenzowania następujących czasopism:

Acta Baltico-Slavica, Adeptus, Cognitive Studies | Études Cognitives, Colloquia Humanistica, Slavia Meridionalis, Sprawy Narodowościowe, Studia Litteraria et Historica, Studia z Filologii Polskiej i Słowiańskiej.

- otwarte, bezkosztowy dostęp

Powiązania z innymi polskimi i zagranicznymi projektami:

xxx

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

xxx

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Dzięki platformie OJS czasopisma Instytutu Slawistyki PAN zostaną zgłoszone do indeksów cytowań (Scopus, ERIH, Google Scholar, Web of Knowledge) i internetowych katalogów czasopism (DOAJ, Lista Regensburska, Open J-Gate).

- tłumaczenia artykułów na j. angielski
- digitalizacja czasopism archiwalnych
- tłumaczenia na język angielski artykułów z archiwalnych numerów czasopism
- otwarcie czasopism na dyskusję online

W przyszłości zostaną wprowadzone na platformę wszystkie czasopisma slawistyczne.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

- zagraniczne platformy czasopism slawistycznych,
- powiązanie z repozytoriami wizualnymi

Profil potencjalnego partnera:

slawiści, językoznawcy, etnografowie/etnologowie/antropologowie, folklorysty, a także w dużym stopniu informatycy (rozbudowywanie narzędzi do przeszukiwania)

PREZBITERZY NA PÓŻNOANTYCZNYM ZACHODZIE

Osoba wypełniająca ankietę: Kierownik projektu

<http://ihuw.pl/instytut/dzialalnosc/projekty/prezbiterzy-na-poznoantycznym-zachodzie>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Historyczny, Wydział Historyczny, Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: ---

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Robert Wiśniewski, r.wisniewski@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Stanisław Adamiak, IH UW

Dwoje doktorantów (zostaną wyłonieni w konkursie)

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 5/22/2014 - 5/21/2019

Źródło finansowania: Narodowe Centrum Nauki 2013/10/E/HS3/00202

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze

Słowa kluczowe: Późna starożytność - Historia Kościoła - Prezbiterzy - Zachód

Skrócony opis projektu:

Celem projektu jest zrozumienie i opisanie roli prezbiterów w kościelnym i społecznym życiu późnoantycznego Zachodu. Badania będą prowadzone na czterech polach: instytucjonalnym, społecznym, ekonomicznym i w dziedzinie mentalności. Na każdym z nich obserwowana będzie ewolucja zachodząca w ciągu kolejnych wieków i różnice między poszczególnymi częściami świata chrześcijańskiego Zachodu.

Siłą projektu skupienie badań nie na wyróżniających się jednostkach, ale na tłumie szeregowych prezbiterów. Całość materiału wprowadzana jest do dostępnej on-line bazy danych, która pozwoli wyszukiwać i sortować informacje dotyczące pochodzenia, kariery, powiązań społecznych, majątku i aktywności ekonomicznej

prezbiterów. Zgromadzony materiał zostanie wykorzystany na trzy sposoby: przez ostrożne zastosowanie metod kwantytatywnych w odniesieniu do jego całości, przez dogłębną analizę dobrze udokumentowanych case studies oraz przez systematyczną konfrontację danych prozopograficznych z tekstami normatywnymi.

Skrócony opis części informatycznej:

Podstawowym instrumentem wytworzonym na potrzeby projektu będzie dostępna on-line baza danych. Każdy jej rekord zawierać będzie zgromadzone z różnych źródeł informacje dotyczące jednego prezbitera, odnotowując m. in. jego pochodzenie społeczne i geograficzne, sytuację rodzinną, ścieżkę kariery, funkcje kościelne, sytuację majątkową, aktywność ekonomiczną etc. Każda z tych kategorii będzie przeszukiwalna, możliwe więc będzie np. wyszukanie wszystkich żonatych prezbiterów trudniących się handlem w szóstowiecznej Galii.

Opis efektów projektu:

Angielskojęzyczna baza danych, wolnodostępna on-line, służyć będzie odpowiedzi na pytania postawione w projekcie, ale stanowić będzie też narzędzie dla wszystkich badaczy zainteresowanych historią społeczną i dziejami Kościoła późnoantycznego.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

RCIN

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://rcin.org.pl/dlibra>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Muzeum i Instytut Zoologii PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Instytut Archeologii i Etnologii PAN

Instytut Badań Literackich PAN

Instytut Biologii Doświadczalnej im. Marcelego Nenckiego PAN

Instytut Biologii Ssaków PAN

Instytut Chemii Fizycznej PAN

Instytut Chemii Organicznej PAN

Instytut Filozofii i Socjologii PAN

Instytut Geografii i Przestrzennego Zagospodarowania PAN

Instytut Historii PAN

Instytut Języka Polskiego PAN

Instytut Matematyczny PAN

Instytut Medycyny Doświadczalnej i Klinicznej PAN

Instytut Podstawowych Problemów Techniki PAN

Instytut Sławistyki PAN

Instytut Technologii Materiałów Elektronicznych

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Kierownik projektu RCIN w IS PAN: mgr Zbigniew Trzciniński

Osoba kontaktowa: Ewa Wróblewska-Trochimiuk, wroblewskae@gmail.com

Pozostałe osoby realizujące projekt:

dr Zofia Rudnik-Karwatowa, IS PAN; dr Marcin Fastyn, IS PAN; mgr Zenon Mikos, IS PAN; mgr Bożena Itoya, IS PAN; dr Paweł Kowalski, IS PAN; mgr Jarosław Bojar (projekt iSybislaw), mgr Krystyna Budek, IS PAN; mgr Paweł Płociński, IS PAN; mgr Katarzyna Sosnowska-Gizińska, IS PAN; mgr Zbigniew Trzciniński, IS PAN; mgr Marta Wojewódzka, IS PAN.

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 12/2/2010 - 6/30/2014

Źródło finansowania: Program Operacyjny "Innowacyjna Gospodarka" POIG.02.03.02-00-043/10

OPIS PROJEKTU

Typ projektu: digitalizacyjny

Efekt projektu: repozytorium

Słowa kluczowe: nagrania audio; badania terenowe; slawistyka, folklor i kultura Słowian, żywy język mówiony

Skrócony opis projektu:

Głównym założeniem projektu w IS PAN jest utworzenie ponadregionalnego i multidyscyplinarnego Repozytorium Cyfrowego złożonego ze zdigitalizowanych materiałów archiwalnych, publikacji naukowych, dokumentacji badań oraz piśmienniczego dziedzictwa kulturowego wyselekcjonowanych ze zbiorów Instytutu Sławistyki PAN

Skrócony opis części informatycznej:

Zdigitalizowane zbiory są udostępniane w Internecie na platformie Repozytorium utworzonego za pomocą systemu dLibra. Materiały znajdujące się w repozytorium są w formacie PDF i DjVu oraz mp3, w pełni zindeksowane, z dodatkowymi ułatwieniami nawigacyjnymi.

Opis efektów projektu:

Zdigitalizowano 903 obiekty, w tym 251 nagrań z badań terenowych, ponad 147 tys. stron.

W kolejnych latach planujemy poszerzyć repozytorium o kolejne zapisy z badań terenowych oraz archiwa pracowników IS PAN.

Obiekty są częściowo udostępniane zgodnie z idą "open access", ale z pełnym poszanowaniem praw autorskich. Cały czas trwa pozyskiwanie praw do publikacji w open access. Podpisano ok. 150 licencji niewyłącznych. Na podpisanie czeka jeszcze około 1000 umów.

Dostęp do repozytorium jest bezpłatny.

Powiązania z innymi polskimi i zagranicznymi projektami:

W ramach realizacji projektu Instytut Sławistyki PAN digitalizuje i udostępnia ponad 100 tysięcy stron publikacji książkowych ze zbiorów Biblioteki IS PAN i starszych publikacji Sławistycznego Ośrodka Wydawniczego, cyfryzuje audio-archiwalne zapisy badań terenowych prowadzonych przez pracowników IS PAN oraz przyspiesza budowę bibliograficznej bazy światowego językoznawstwa sławistycznego iSybisław.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

- poszerzenie o materiały audio pochodzące z badań terenowych
- tworzenie narzędzi do przeszukiwania materiałów audio

- utworzenie archiwum Michała Chęcińskiego
- połączenie i zlinkowanie z planowanym projektem "Cyfrowa slawistyka"

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

- możliwość współpracy ze slawistycznymi repozytoriami zagranicznymi

Profil potencjalnego partnera:

slawistyka, etnografia/ etnologia/antropologia, dialektologia, historia mówiona, informatyka w zakresie tworzenia narzędzi do przeszukiwania nagrań audio

RZUTOWANIE SŁOWOSIECI NA PRINCETON WORDNET

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://plwordnet.pwr.wroc.pl>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Grupa Naukowa G4.19, Instytut Informatyki Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Projekt jest realizowany wyłącznie na Politechnice Wrocławskiej

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Ewa Rudnicka, ewa.rudnicka@pwr.wroc.pl

Pozostałe osoby realizujące projekt:

dr Ewa Rudnicka (PWr.), mgr Wojciech Witkowski (UWr. w ramach projektu pracuje na rzecz PWr.), dr inż. Maciej Piasecki (PWr.),

oraz kilku lingwistów

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 1/1/2012 - 12/31/2015

Źródło finansowania: Program Operacyjny "Innowacyjna Gospodarka" NEKST POIG, CLARIN-PL dotacja na inwestycję MNiSW

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, publikacja

Słowa kluczowe: rzutowanie wordnetów; polsko-angielska sieć relacji leksykalno-semantycznych; dwujęzyczny zasób językowy; relacje międzyjęzykowe; podpowiedzi automatyczne

Skrócony opis projektu:

W ramach rzutowania Słownosieci (ang. plWordNet) na Princeton WordNet powstaje polsko-angielska sieć relacji leksykalno-semantycznych, która zawiera w najnowszej wersji 2.2 ponad 90 tysięcy powiązań międzyjęzykowych pomiędzy synsetami (zbiorami synonimów) Słownosieci i Princeton WordNet.

Skrócony opis części informatycznej:

W ramach części informatycznej opracowane zostały dwa systemy automatycznych podpowiedzi dla rzutujących leksykografów, osobno dla rzeczowników i przymiotników; dla rzeczowników w oparciu o algorytm Relaxation Labeling; dla przymiotników w oparciu o reguły zdefiniowane na podstawie analizy struktury i składu sieci relacji w obu wordnetach, oba systemy w powiązaniu z wiedzą o semantyce leksykalnej wydobytą z dostępnych słowników.

Opis efektów projektu:

bardzo duży polsko-angielski wordnet - rodzaj dwujęzycznej sieci relacji leksykalno-semantycznych; jeden większych słowników polsko-angielskich;

docelowa wielkość: ponad 200 000 nawiązanych relacji

międzyjęzycznych między znaczeniami Słowosieci i Princeton WordNet

Powiązania z innymi polskimi i zagranicznymi projektami:

Współpracujemy z Global WordNet Association, Princeton WordNet, GermaNet, CzechWordNet, SloNet (słoweński wordnet), Open Multilingual WordNet, DanNet (duński wordnet)

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://plwordnet.pwr.wroc.pl>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Rozbudowa rzutowania Słowosieci na Princeton WordNet (nowe kategorie leksykalne, nowe relacje, kompletny zestaw relacji) do stanu, który zapewni pełne powiązanie obu zasobów relacjami międzyjęzycznymi.

Utworzony dwujęzyczny zasób leksykalno-semantyczny będzie oferował wiele możliwości zastosowań, takich jak np. tłumaczenie maszynowe, dwujęzyczne przeszukiwanie tekstów i wydobywanie informacji.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Słowosieć jest łączona stopniowo z innymi wordnetami świata w tzw. Global WordNet Grid

Profil potencjalnego partnera:

różnorodne instytucje: wykorzystanie połączonych zasobów Słowosieci i Princeton WordNet w badaniach, wykorzystanie w systemach indeksujących dwujęzyczne kolekcje i bazy danych, wykorzystanie w systemach przetwarzających język naturalny oraz w systemach maszynowego tłumaczenia.

SEMANTYCZNO-SKŁADNIOWY SŁOWNIK ZELŻNOŚCI SUBKATEGORYZACYJNYCH LEKSEMÓW POLSKICH

Osoba wypełniająca ankietę: Kierownik projektu

<http://clarin-pl.eu/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Instytut Podstaw Informatyki PAN, ul. Jana Kazimierza 5, Warszawa

Pozostałe instytucje zaangażowane w projekt: Projekt stanowi pojedyncze zadania realizowane w ramach infrastruktury badawczej CLARIN-PL. Jest ona realizowana przez konsorcjum koordynowane przez Politechnikę Wrocławską (kierownik: Maciej Piasecki) – zob. ankietę dot. całego CLARIN-PL.

Poniżej „kierownik” itd. rozumiane jest jako „kierownik opisywanego zadania” itd.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Elżbieta Hajnicz, Elzbieta.Hajnicz@ipipan.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Tomasz Bartosiak, IPI PAN;

Anna Grzeszak, UW;

Elżbieta Hajnicz, IPI PAN;

Natalia Kocyba, IPI PAN;

Bartłomiej Nitoń, IPI PAN i SWPS;

Jakub Szymczak;

Sebastian Żurowski, UMK

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 4/1/2013 - 12/31/2015

Źródło finansowania: dotacja na dużą infrastrukturę badawczą

OPIS PROJEKTU

Typ projektu: infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych

Słowa kluczowe: słownik walencyjny; składnia; role tematyczne; preferencje selekcyjne; język polski

Skrócony opis projektu:

Celem zadania A13 w CLARIN-PL jest stworzenie syntaktyczno-semantycznego słownika walencyjnego leksemów polskich, w tym 12000 czasowników, 2000 rzeczowników oraz 1000 przymiotników i przysłówków. Hasła czasownikowe dzielą się na podhasła ze względu na aspekt, zanegowanie i zwrotność. Na poziomie składniowym (pod)hasło ma przypisywane schematy składniowe będące listami pozycji, z których każda jest listą mogących się koordynować wymagań. Wyróżniamy pozycję podmiotu i dopełnienia (ze względu na pasywowizalność). Reprezentujemy zjawiska kontroli składniowej, niestandardowej koordynacji.

Słownik zawiera bogaty opis frazeologizmów.

Na poziomie semantycznym hasłom przypisywane są ramy semantyczne stanowiące listę argumentów opisywanych za pomocą roli tematycznej i listy preferencji selekcyjnych. Każdej ramie przypisywana jest lista realizujących ją schematów składniowych, a każdemu argumentowi --- pozycja realizująca go w tym schemacie.

Skrócony opis części informatycznej:

Tworzenie słownika odbywa się przy użyciu dedykowanego narzędzia za pośrednictwem przeglądarki internetowej. Program dostarcza leksykografom opracowującym hasła wspomaganie dotyczące poprawności tworzonych haseł (na poziomie poprawności pojedynczego wymagania, zasad koordynacji, przykładów przypisywanych hasłom). Program generuje słownik w formacie tekstowym (na potrzeby parserów), graficznym (LaTeX, na potrzeby czytelnika) oraz XML.

Opis efektów projektu:

słownik walencyjny zgodny z powyższym opisem do wykorzystania przez parsery językowe, do ekstrakcji informacji, tłumaczenia maszynowego itp.

Powiązania z innymi polskimi i zagranicznymi projektami:

NKJP

Słowosieć

NEKST

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Informacja o walencji, zwłaszcza na poziomie semantycznym, stanowić ma m.in. podstawę do opracowania głębokiej struktury semantycznej tekstu wykraczającej poza pojedyncze zdanie.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Słowosieć

CORE

inne zadania CLARIN

Profil potencjalnego partnera:

SENSUALNOŚĆ W KULTURZE POLSKIEJ. PRZEDSTAWIENIA ZMYŚŁÓW CZŁOWIEKA W JĘZYKU, PIŚMIENICTWIE I SZTUCE OD ŚREDNIOWIECZA DO WSPÓŁCZESNOŚCI.

Osoba wypełniająca ankietę: Kierownik projektu

sensualnosc.ibl.waw.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Pracownia Poetyki Historycznej w Instytucie Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Biblioteka Narodowa; Muzeum Narodowe w Warszawie, Instytut Sztuki PAN, Muzeum Sztuki w Łodzi, Muzeum Etnograficzne w Toruniu.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Włodzimierz Bolecki; wbol@wp.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Magdalena Rembowska-Płuciennik; Beata Śniecikowska, Tamara Brzostowska-Tereszkiewicz, Agnieszka Kluba, Andrzej Karcz, Piotr Sobolczyk, Maciej Kidawa.

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 1/2/2010 - 12/31/2012

Źródło finansowania: Narodowe Centrum Badań i Rozwoju N R17 0005 06/2009

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, publikacja

Słowa kluczowe: Sensualność; kultura polska, multimedialność, interdyscyplinarność

Skrócony opis projektu:

Projekt poświęcony jest problematyce sensualności rozumianej jako historycznie zmienny zbiór form reprezentacji ludzkich zmysłów (np. wzrok, słuch, smak, dotyk, węch, ciało). Celem badawczym projektu jest przedstawienie sensualności w szerokim kontekście kulturowym (np. literatura, język, sztuka, filozofia, kultura ludowa, historia, teatr) oraz różnorodnych zagadnień związanych z poszczególnymi zmysłami.

Więcej: <http://sensualnosc.ibl.waw.pl/pl/informacje-o-projekcie/>

Skrócony opis części informatycznej:

Projekt jest realizowany on-line w formie witryny pełniącej funkcję multimedialnej internetowej encyklopedii tematycznej. Ze względu na planowany rozwój treści w witrynie, a także możliwość nieograniczonego łączenia haseł z innymi zasobami internetowymi projekt jest typowym work in progress. Każde hasło może zostać uzupełnione nowymi materiałami multimedialnymi (ilustracje, obrazy, wizualizacje, animacje, symulacje, nagrania audio i wideo itp.). Użytkownik tej witryny ma do dyspozycji indeksy i wyszukiwarki umożliwiające wyszukiwanie interesujących go haseł i tematów.

Więcej: <http://sensualnosc.ibl.waw.pl/pl/informacje-o-projekcie/>

Opis efektów projektu:

Open Access

Powiązania z innymi polskimi i zagranicznymi projektami:

brak danych

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

www.sensualnosc.ibl.waw.pl

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Warunkiem rozwoju projektu jest otrzymanie finansowania pozwalającego na jego rozbudowanie i kontynuację.

Projekt został złożony do trzeciego konkursu NPRH w module 1.2. Otrzymał średnią ocen 97 punktów (100 + 94), ale nie został skierowany do finansowania.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Otrzymano zapewnienia współpracy od grup badawczych z UK, Francji, Norwegii, Węgier.

Profil potencjalnego partnera:

grupa badawcza w zakresie literaturoznawstwa i interdyscyplinarnych studiów kulturowych.

SŁOWNIK POLSZCZYNY XVI WIEKU I INTERNETOWY SŁOWNIK POLSZCZYNY XVI WIEKU - TOMY 36-43

Osoba wypełniająca ankietę: Osoba kontaktowa

www.spxvi.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Badań Literackich PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Patrycja Potoniec, spxvi@poczta.onet.pl

Osoba kontaktowa: Krzysztof Opaliński, k.opalinski.ibl@gmail.com

Pozostałe osoby realizujące projekt:

Patrycja Potoniec, IBL PAN; Krystyna Wilczewska, IBL PAN; Ewa Cybulska-Bohuszewicz, IBL PAN; Anetta Luto-Kamińska, IBL PAN; Aleksandra Ćwiklińska, IBL PAN; Anna Nath-Dokurno, IBL PAN; Krzysztof Opaliński, IBL PAN; Małgorzata Pierzgalska, IBL PAN; Maria Popielarska, IBL PAN; Lucyna Wilczewska, IBL PAN; Lucyna Woronczakowa, IBL PAN; Małgorzata Nobis, IBL PAN; Jadwiga Bugajska-Rutyna, IBL PAN; Marta Cegielska, IBL PAN; Danuta Dziadowicz-Jezioro

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 1/1/2013 - 12/31/2018

Źródło finansowania: Narodowy Program Rozwoju Humanistyki 0015/NPRH2/H11/81/2013

OPIS PROJEKTU

Typ projektu: bazodanowy, digitalizacyjny

Efekt projektu: narzędzie badawcze, baza danych

Słowa kluczowe: słownik; językoznawstwo polskie; leksykografia polska; polska literatura renesansu; polszczyzna dawna

Skrócony opis projektu:

Projekt ma na celu wspieranie kontynuacji prac redakcyjnych nad „Słownikiem polszczyzny XVI wieku” w wersji drukowanej oraz systematyczne uzupełnianie wersji elektronicznej słownika dostępnej publicznie w internecie. Zakres prac przewidziany w projekcie obejmuje planowane tomy XXXVI-XLIII (hasła na litery R-S). W ramach projektu realizowane są prace redakcyjne nad kolejnymi hasłami, korekty, konsultacje ze specjalistami i

wreszcie prace wydawnicze oraz druk. Równoległe z pracami wydawniczymi gotowe hasła znajdują się w bazie danych elektronicznej wersji słownika.

Skrócony opis części informatycznej:

W ramach przygotowań do realizacji projektu powstało zaplecze sprzętowo-programowe do realizacji wersji elektronicznej. Stworzono serwis internetowy, w ramach którego działa aplikacja obsługująca bazę danych. Powstał i jest systematycznie uzupełniany indeks haseł, który stanowi podstawę dla umieszczania w bazie kolejnych artykułów hasłowych. Przyjęto założenie, że artykuły hasłowe będą miały format XML, który podlega regułom określonym w schemacie DTD opracowanym specjalnie na potrzeby słownika. Utworzono specjalny edytor, pod którego kontrolą można przygotować artykuły hasłowe w formacie XML w oparciu o schemat DTD dla słownika.

Opis efektów projektu:

Powstaje kompletny indeks haseł „Słownika polszczyzny XVI wieku” obejmujący materiał opublikowany (hasła na A-R), nieopublikowany (hasła na R-Ż) oraz materiały uzupełniające (ekscerpty z tekstów spoza kanonu źródeł słownika). Indeks jest już częściowo dostępny w serwisie słownika (hasła A-J oraz R-Ż).

Ponadto przeniesiono wstępnie do bazy danych XXXV tom słownika (dostępny online: www.spxvi.edu.pl). Kolejne tomy będą pojawiały się wraz z postępem prac redakcyjnych i wydawniczych. Wydano drukiem XXXVI tom słownika i trwają prace nad jego elektroniczną wersją. Prowadzone są prace redakcyjne nad kolejnymi hasłami, które wejdą do kolejnych tomów z zakresu przewidzianego przez projekt.

Wersja drukowana „Słownika” będzie rozprowadzana na dotychczasowych zasadach przez Wydawnictwo IBL, natomiast wersja elektroniczna będzie dostępna publicznie na stronie www.spxvi.edu.pl.

Powiązania z innymi polskimi i zagranicznymi projektami:

„Słownik polszczyzny XVI wieku” wpisuje się w ideę ujęcia w leksykograficzne ramy całości piśmiennictwa polskiego od jego zarania. Naturalny jest zatem ścisły związek z dokonaniem pozostałych słowników historycznych języka polskiego, zwłaszcza ze „Słownikiem staropolskim” oraz „Słownikiem języka polskiego XVII i 1. poł. XVIII w.”. Wspólne pola zainteresowań można również wskazać dla „Słownika łaciny średniowiecznej w Polsce”, a zwłaszcza jego wersji elektronicznej.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.spxvi.edu.pl>

<http://kpbc.umk.pl/dlibra/publication?id=17781>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Pracownia nadal będzie aplikowała o kolejne środki na kontynuację prac nad „Słownikiem polszczyzny XVI wieku”. Celem tych działań będzie dokończenie tego monumentalnego dzieła polskiej leksykografii. Dodatkowo należy uzupełnić dotychczas wydany materiał o liczne dodatkowe ekscerpty i poprawki. Równoległe z pracami redakcyjnymi będzie się poszerzać baza danych elektronicznej wersji słownika. W planach jest też rozbudowa narzędzia do wyszukiwania informacji w e-słowniku. Ponadto przewidziana jest ścisła integracja słownika elektronicznego z powstającym „Korpusem polszczyzny XVI wieku”.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Elektroniczny słownik może stanowić bazę leksykograficzną dla cyfrowych edycji dzieł staropolskich.

Profil potencjalnego partnera:

zespoły leksykograficzne

SŁOWOSIEĆ

Osoba wypełniająca ankietę: Kierownik projektu

<http://plwordnet.pwr.wroc.pl>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Grupa Naukowa G4.19, Instytut Informatyki Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Projekt jest realizowany wyłącznie na Politechnice Wrocławskiej

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr Marek Maziarz (PWr.), dr inż. Maciej Piasecki (PWr.), prof. Stanisław Szpakowicz (University of Ottawa, w ramach projektu pracuje na rzecz PWr.), mgr Marta Dobrowolska (PWr.), mgr Agnieszka Dziob (PWr.), mgr Amelia Kiełbawska (PWr.), Justyna Ławniczak (PWr.) oraz kilkunastu lingwistów

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 10/30/2005 - 12/31/2015

Źródło finansowania: Narodowe Centrum Nauki, Narodowe Centrum Badań i Rozwoju, Program Operacyjny "Innowacyjna Gospodarka" Nr 3 T11C 018 29, N N516 068637, NEKST POIG, SyNaT projekt strategiczny NCBiR, CLARIN-PL dotacja na inwestycję MNiSW

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, publikacja

Słowa kluczowe: wordnet, leksykalna sieć semantyczna, tezaurus, zasób językowy, wordnet języka polskiego, edytor wordnetu, metody wydobywania relacji semantycznych z korpusu

Skrócony opis projektu:

Słowosieć (ang. plWordNet) to lingwistyczna sieć semantycznych relacji leksykalnych, która opisuje w najnowszej wersji 2.2 ponad 140 000 lematów (form hasłowych), 200 000 znaczeń leksykalnych (jednostek leksykalnych) za pomocą ponad 500 000 relacji leksykalno-semantycznych i ponad 150 000 synsetów.

Skrócony opis części informatycznej:

W ramach części informatycznej zdefiniowany został formalizm do zapisu Słowosieci, skonstruowany został sieciowy system WordnetLoom do współpracy grupy leksykografów nad konstrukcją słownika oraz system WordnetWeaver do półautomatycznego rozszerzania sieci w oparciu o wiedzę o semantyce leksykalnej wydobytą automatycznie z wielkich korpusów tekstu.

Opis efektów projektu:

bardzo duży wordnet języka polskiego - rodzaj lingwistycznej leksykalnej sieci semantycznej; jeden z największych słowników języka polskiego;

docelowa wielkość: około 200 000 lematów (form hasłowych) oraz 260-280 tys. znaczeń leksykalnych

Powiązania z innymi polskimi i zagranicznymi projektami:

Współpracujemy z Global WordNet Association, Princeton WordNet, GermaNet, CzechWordNet, SloNet (słoweński wordnet), Open Multilingual WordNet, DanNet (duński wordnet)

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://plwordnet.pwr.wroc.pl>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Rozbudowa Słowosieci (lematy, jednostki leksykalne i relacje) do stanu, w którym zapewni wyczerpujący opis polskiego systemu leksykalnego.

Słowosieć jest stopniowo łączona z innymi zasobami leksykalnymi i zasobami wiedzy, w celu zbudowania wielkiego zasobu semantycznego o szerokim spektrum zastosowań.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Słowosieć jest łączona stopniowo z innymi wordnetami świata w tzw. Global WordNet Grid

Profil potencjalnego partnera:

różnorodne instytucje: wykorzystanie Słowosieci w badaniach, rozszerzanie Słowosieci o dodatkowe informacje (struktury semantyczne, polaryzacja emocjonalna, dziedziny semantyczne), wykorzystanie w systemach indeksujących kolekcje i bazy danych, wykorzystanie w systemach przetwarzających język naturalny.

SYSTEM DO ZINTEGROWANEJ, GŁĘBOKIEJ, SKŁADNIOWO-SEMANTYCZNEJ ANALIZY TEKSTÓW W JĘZYKU POLSKIM

Osoba wypełniająca ankietę: Kierownik projektu

<http://clarin-pl.eu/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Instytut Podstaw Informatyki PAN

Pozostałe instytucje zaangażowane w projekt: Cały projekt CLARIN-PL jest realizowany przez konsorcjum koordynowane przez Politechnikę Wrocławską (kierownik: Maciej Piasecki) – zob. ankietę dot. całego CLARIN-PL. Tu opisane jest konkretne zadanie realizowane w IPI PAN. Poniżej „kierownik” itd. rozumiane jest jako „kierownik danego zadania” itd.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Adam Przepiórkowski, adamp@ipipan.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Adam Przepiórkowski (IPI PAN)

Agnieszka Patejuk (IPI PAN)

Marcin Woliński (IPI PAN)

Wojciech Jaworski (IPI PAN, UW, PJWSTK)

Szymon Acedański (IPI PAN, UW)

i in.

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 2/1/2013 - 12/31/2015

Źródło finansowania: dotacja na dużą infrastrukturę badawczą

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: gramatyka; parser; składnia; semantyka; przetwarzanie tekstów

Skrócony opis projektu:

W ramach projektu rozwijane są powstałe w IPI PAN narzędzia analizy składniowej i semantycznej tekstów polskich.

Skrócony opis części informatycznej:

Głównym wynikiem projektu będą rozbudowane wersje dwóch parserów języka polskiego: Świgr (http://zil.ipipan.waw.pl/%C5%9Awigra) i POLFIE (http://zil.ipipan.waw.pl/LFG).

Opis efektów projektu:

j.w.

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekt jest kontynuacją wcześniejszych prac dotyczących przetwarzania składniowego w ramach projektu NEKST.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

j.w.

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Oba parsery zwiększają pokrycie empiryczne i jakość analiz wynikowych. Wyniki analizy zdań parserem POLFIE to także struktura semantyczna tych zdań – główny nacisk powinien zostać położony na zwiększenie pokrycia i jakości tego modułu semantycznego. Dla obu parserów powstają też wersje probabilistyczne, pozwalające szeregować wyniki od najbardziej prawdopodobnego. Dalsze prace powinny przede wszystkim zwiększyć odporność parserów na błędy w tekstach wejściowych (i na konstrukcje nie objęte opisem).

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

(zob. 25 poniżej)

Profil potencjalnego partnera:

Użytkownicy technologii głębokiego przetwarzania tekstów, np. do celów odpowiadania na pytania (ang. Question Answering), ustalania, czy dwa teksty pozostają w relacji wynikania (ang. Recognising Textual Entailment) itp.

Mile widziana współpraca z semantykami i pragmatykami.

SYSTEM PRZESZUKIWANIA DUŻYCH BOGATO ZNAKOWANYCH KORPUSÓW JĘZYKOWYCH

Osoba wypełniająca ankietę: Kierownik projektu

<http://clarin-pl.eu/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Instytut Podstaw Informatyki PAN

Pozostałe instytucje zaangażowane w projekt: Cały projekt CLARIN-PL jest realizowany przez konsorcjum koordynowane przez Politechnikę Wrocławską (kierownik: Maciej Piasecki) – zob. ankietę dot. całego CLARIN-PL. Tu opisane jest konkretne zadanie realizowane w IPI PAN. Poniżej „kierownik” itd. rozumiane jest jako „kierownik danego zadania” itd.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Adam Przepiórkowski, adamp@ipipan.waw.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Adam Przepiórkowski (IPI PAN)

Bartosz Zaborowski (IPI PAN)

Aleksander Zabłocki (IPI PAN i UW)

Dominika Pawlik (IPI PAN i UW)

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 2/1/2013 - 12/31/2015

Źródło finansowania: dotacja na dużą infrastrukturę badawczą

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: korpusy; wyszukiwanie; języka zapytań; znakowanie lingwistyczne

Skrócony opis projektu:

Celem zadania A20 w CLARIN-PL jest stworzenie efektywnej przeszukiwarki korpusowej dla wielkich korpusów tekstowych znakowanych na wielu poziomach lingwistycznych.

Skrócony opis części informatycznej:

j.w.

Opis efektów projektu:

j.w.

Powiązania z innymi polskimi i zagranicznymi projektami:

NKJP i inne projekty wymagające przeszukiwania korpusów.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Dostosowanie przeszukiwarki do specyfiki tekstów historycznych i gwarowych, oraz do nowych typów informacji lingwistycznych.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Użytkownicy przeszukiwarki korpusowej.

SYSTEM WYDOBYWANIA WIELOWYRAZOWYCH LEMATÓW JĘZYKA POLSKIEGO

Osoba wypełniająca ankietę: Adam Radziszewski, adam.radziszewski@pwr.edu.pl

<http://www.clarin-pl.eu>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Grupa Naukowa G4.19, Instytut Informatyki Politechnika Wrocławska

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Projekt jest realizowany wyłącznie na Politechnice Wrocławskiej jako moduł A14 systemu wytwarzanego w ramach projektu CLARIN-PL, którego Politechnika Wrocławska jest koordynatorem.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Piasecki, maciej.piasecki@pwr.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

dr inż. Maciej Piasecki, PWr;

dr Marek Maziarz, Pwr;

dr inż. Mariusz Paradowski, Pwr;

dr inż. Adam Radziszewski, PWr;

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 2/15/2013 - 12/31/2015

Źródło finansowania: dotacja celowa CLARIN-PL dotacja na inwestycję MNiSW

OPIS PROJEKTU

Typ projektu: badawczy, infrastrukturalny, bazodanowy

Efekt projektu: narzędzie badawcze, oprogramowanie, baza danych, publikacja

Słowa kluczowe: wyrażenia wielowyrzowe; język polski; kolokacje; statystyka; słownik

Skrócony opis projektu:

Wynikiem realizacji projektu będzie baza danych wielowyrzowych jednostek leksykalnych języka polskiego. Baza zawierać będzie przynajmniej 60 tys. wielowyrzowych jednostek leksykalnych. Każdej jednostce przypisany zostanie wzorzec odmiany, forma podstawowa oraz opis semantyczny. Drugim produktem będzie

system pozwalający na rozszerzanie takiej bazy danych na podstawie statystycznej analizy dużych zbiorów tekstu.

Skrócony opis części informatycznej:

Wytwarzany jest system informatyczny, które umożliwi efektywne przetwarzanie ogromnych zbiorów tekstu (kilku miliardów słów). Wymaga to opracowania wydajnych algorytmów, struktur danych i metod przechowywania danych. System korzysta z technik płytkiej analizy składniowej, m.in. analizuje związki składniowe przy pomocy wyrażeń funkcyjnych napisanych w języku dziedzinowym WCCL. System implementuje heurystyki znajdowania form podstawowych jednostek.

Opis efektów projektu:

Efektom projektu jest algorytm rozpoznawania jednostek wielowyrazowych na podstawie analizy statystycznej dużych korpusów tekstowych i informacji semantycznej ze Słownosieci. Obecnie gotowa jest jedynie prototypowa implementacja. Implementacja produkcyjna jest realizowana. Będzie ona dostępna na jednej z otwartych licencji typu open source. Uzyskana baza danych zostanie również udostępniona na otwartej licencji, np. z rodziny Creative Commons.

Powiązania z innymi polskimi i zagranicznymi projektami:

Jednostki wielowyrazowe kilku języków są tematem przewodnim sieci naukowej PARSEME (http://www.cost.eu/domains_actions/ict/Actions/IC1207), z tym że działania tam podejmowane skupiają się na analizie składniowej zdań przy pomocy informacji o takich jednostkach. Opisywane tutaj przedsięwzięcie skupia się bardziej na opisie (również semantycznym) samych w sobie jednostek jako elementów leksyki języka.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Obecne badania skupiają się na rozpoznawaniu jednostek wielowyrazowych należących do języka ogólnego. Nie prowadzone są badania nad rozpoznaniem jednostek, które charakteryzują dany tekst, typ tekstu czy dziedzinę. Celowe jest rozszerzenie systemu o możliwość analizy porównawczej wielu źródeł i ekstrakcji zbiorów jednostek charakterystycznych dla danych źródeł. Umożliwi to także odfiltrowanie pseudojednostek, które są używane jedynie przez danego autora bądź bardzo wąskie grono (np. autora oraz osoby komentujące wypowiedzi tego autora).

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Możliwa jest współpraca z instytucjami biorącymi udział w przedsięwzięciu PARSEME.

Profil potencjalnego partnera:

Instytucje zajmujące się naukami humanistycznymi, w szczególności: leksykografią, stylometrią, składnią języka polskiego i semantyką leksykalną.

Instytucje zajmujące się inżynierią lingwistyczną.

THE CENTRAL EUROPEAN JOURNAL OF SOCIAL SCIENCES AND HUMANITIES CEJSH

Osoba wypełniająca ankietę: Jakub Szprot, jszprot@icm.edu.pl

<http://cejsh.icm.edu.pl>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Polska Akademia Nauk oraz Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego, Uniwersytet Warszawski

Institucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Akademia Nauk Republiki Czeskiej, Słowacka Akademia Nauk, Węgierska Akademia Nauk

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Miłosz Sosnowski, Milosz.Sosnowski@amu.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: -

Źródło finansowania:

OPIS PROJEKTU

Typ projektu: bazodanowy

Efekt projektu: baza danych, kolekcja czasopism naukowych w otwartym dostępie

Słowa kluczowe: czasopisma; otwarty dostęp; baza; współpraca międzynarodowa

Skrócony opis projektu:

CEJSH jest bazą bibliograficzno-abstraktową oraz pełnotekstową, rejestrującą artykuły z czasopism z obszaru nauk humanistycznych i społecznych, ukazujących się w Europie Środkowej i Wschodniej (przede wszystkim w Polsce, Republice Czeskiej, Słowacji i na Węgrzech, ale także w Estonii, Rumunii, Słowenii oraz na Łotwie, Litwie i Ukrainie). Ma na celu ułatwienie komunikacji naukowej poprzez zapewnienie badaczom z całego świata dostępu do wyników badań humanistycznych i społecznych prowadzonych w tej części Europy.

Baza została utworzona w 2004 r. przez Prezesów Akademii Nauk państw Grupy Wyszehradzkiej: Akademii Nauk Republiki Czeskiej, Polskiej Akademii Nauk, Słowackiej Akademii Nauk oraz Węgierskiej Akademii Nauk.

Obecnie jest rozbudowywana oraz udostępniana we współpracy PAN (wraz z pozostałymi Akademiemi) i ICM UW. ICM UW odpowiada za platformę internetową bazy oraz zawiera z wydawcami czasopism indeksowanych w CEJSH umowy, na podstawie których w bazie udostępniane są pełne teksty artykułów publikowanych w tych czasopismach.

Skrócony opis części informatycznej:

Baza CEJSH jest tworzona, przechowywana i udostępniana przy użyciu stworzonego w ICM UW oprogramowania YADDA.

Opis efektów projektu:

W bazie indeksowane są 723 czasopisma, w czego 137 dostępnych jest w sposób otwarty w wersji pełnotekstowej (stan na 20.06.2014).

Powiązania z innymi polskimi i zagranicznymi projektami:

Baza powstaje we współpracy z Akademią Nauk Republiki Czeskiej, Słowacką Akademią Nauk i Węgierską Akademią Nauk. Indeksowane czasopisma wydawane są w różnych krajach Europy Środkowej i Wschodniej.

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Planowane jest zwiększenie liczby czasopism dostępnych w wersji pełnotekstowej.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Możliwości współpracy dotyczą z jednej strony indeksowania i otwartego udostępniania w CEJSH czasopism z obszaru nauk humanistycznych i społecznych, z drugiej zaś wykorzystania treści naukowych w postaci cyfrowej dostępnych w bazie.

Profil potencjalnego partnera:

Wydawcy, redakcje czasopism i badacze z obszaru nauk humanistycznych i społecznych.

TRENDMINER

Osoba wypełniająca ankietę: Kierownik projektu

<http://www.trendminer-project.eu/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Zespół Inżynierii Lingwistycznej Instytutu Podstaw Informatyki PAN

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: DFKI, Niemcy; Uniwersytet Sheffield, Wielka Brytania; Ontotext, Bułgaria, Uniwersytet Southampton, Wielka Brytania; Eurokleis, Włochy; Sora Ogris & Hofinger, Austria; Hardik Fintrade, Indie; Internet Memory Research, Francja; Uniwersytet Carlosa III w Madrycie, Hiszpania; Węgierska Akademia Nauk; Daedalus-Data, Hiszpania.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Ogrodniczuk, maciej.ogrodniczuk@gmail.com

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Maciej Ogrodniczuk, IPI PAN; Mateusz Kopeć, IPI PAN; Aleksander Wawer, IPI PAN; Dariusz Czerski, IPI PAN; Paweł Łoziński, IPI PAN; Paweł Ziniewicz, IPI PAN

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 11/1/2011 - 10/31/2013

Źródło finansowania: FP7-ICT-2011-7 287863

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: opinion mining, analiza emocjonalna, ontologia polskiej sceny politycznej; korpus sejmowy; Twitter

Skrócony opis projektu:

Celem projektu jest opracowanie innowacyjnych, replikowalnych, dostępnych na zasadzie open-source metod wielojęzycznego wydobywania informacji i agregacji dużych ilości danych z mediów strumieniowych. Środkiem do realizacji tego celu jest podejście łączące zaawansowane metody lingwistyczne z dziedziny przetwarzania tekstu, wnioskowania, uczenia maszynowego, ekonomii i nauk politycznych. W przeciwieństwie do innych projektów tego rodzaju nie jest planowana kosztowna anotacja danych.

Skrócony opis części informatycznej:

W części informatycznej projektu zamierzamy wykorzystać metody uczenia maszynowego ze słabym nadzorem do automatycznego wykrywania nowych trendów i zachodzących korelacji.

Skalowalność i przystępność tego rodzaju przetwarzania danych planujemy osiągnąć za pomocą infrastruktury chmury obliczeniowej udostępniającej narzędzia analityczne działające w czasie rzeczywistym. Wyniki prac zostaną poddane ewaluacji w studium przypadku polegającym na udostępnieniu narzędzi analizy

politycznej i monitorowania strumienia danych społecznościowych w języku polskim pochodzących z wpisów Twittera.

Opis efektów projektu:

Docelowo powstałe metody i narzędzia będą mogły zostać użyte komercyjnie w analizie biznesowej, zarządzaniu relacją z klientem czy kontekstach społecznościowych. Korzyści z realizacji projektu odniesie również ogół społeczeństwa i pojedynczy obywatele poprzez uzyskanie dostępu do bieżącej informacji politycznej i ogólnej w postaci podsumowań trendów w strumieniach danych Twittera.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://clip.ipipan.waw.pl/TrendMiner>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Analiza semantyczna polskich tekstów zasumionych w rodzaju wpisów Twittera, blogów, wpisów z Facebooka itp.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Lingwiści dysponujący zasobami wspomagającymi przetwarzanie tekstów zasumionych lub gotowi do współpracy przy ich tworzeniu.

TRUDNE MARKI – DEFINICJA I PROPOZYCJA ZASTOSOWANIA METOD JĘZYKOZNAWCZYCH DO BADANIA ZJAWISKA

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Doktorant w ramach studiów doktoranckich

Nazwa podmiotu realizującego projekt: Doktorat realizowany w Instytucie Filologii Polskiej, Wydział Filologiczny Uniwersytetu Wrocławskiego

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Politechnika Wrocławska

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Agnieszka Dziob, agnieszka_kaczor@wp.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 10/1/2009 - 12/31/2014

Źródło finansowania: dotacja celowa, środki własne doktoranta

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: publikacja

Słowa kluczowe: wizerunek; kolokacje; językoznawstwo; Inforex; Vector Extractor

Skrócony opis projektu:

Celem projektu jest wypracowanie językoznawczej metody badania złożonych zjawisk, mających swój wizerunek (cechujących się jakąś opinią, renomą). Badam marki, ale pojęcie marki rozszerzam na wszystkie podmioty czy też przedmioty, wobec których stosuje się działania marketingowe lub z zakresu PR, mające wpłynąć na ich postrzeganie. Stosowane obecnie działania i metody badania zjawiska marki (zaczepnięte głównie z psychologii, statystyki, socjologii) są nieadekwatne i nie oddają efemeryczności zjawiska. W swoich badaniach wychodzę od tego, jak informacja publiczna wpływa na postrzeganie tych przedmiotów i podmiotów.

Skrócony opis części informatycznej:

W trakcie realizacji projektu korzystam z aplikacji do anotacji korpusów Inforex, stworzonej przez Politechnikę Wrocławską (<http://nlp.pwr.wroc.pl/narzedzia-i-zasoby/inforex>) oraz narzędzia Vector Extractor (<http://www.clarin-pl.eu/vector-extractor/>) służącego do analizy konotacji leksykalno-składniowych (narzędzie udostępnione i rozwijane w ramach projektu CLARIN). Projekt dąży do algorytmizacji procesu ręcznej analizy tekstów tak, by w przyszłości była możliwa częściowa automatyzacja procesu.

Opis efektów projektu:

Zrealizowane:

artykuł monografii: "Zjawisko trudnej marki na przykładzie marki Polska" w: "Marka i promocja jednostek terytorialnych. Dylematy i wyzwania współczesności", red. K. Malicki, str. 387-406.

Planowane:

publikacja książkowa po zakończeniu projektu (doktoratu)

Powiązania z innymi polskimi i zagranicznymi projektami:

Współpraca z Grupą Technologii Językowych G4.19 Politechniki Wrocławskiej w ramach projektu CLARIN (Vector-Extractor, Inorex)

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Dalszym etapem badań mogłoby być sformalizowanie postępowania badawczego tak, by zautomatyzować żmudny proces ręcznej analizy materiału przez badacza i móc go powierzyć nieistniejącej jeszcze aplikacji, która potrafiłaby wg wyznaczonych przez badacza kryteriów analizować tak złożone zjawiska, pozostawiając badaczowi jedynie weryfikację wyników. Wyobrażam sobie rozwój mojego projektu jako współpracę z programistami nad taką właśnie aplikacją. Taka aplikacja miałaby duży potencjał komercyjny.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Interesuje mnie dalsza współpraca z Grupą Technologii Językowych G4.19 Politechniki Wrocławskiej

USŁUGA POWSZECHNEJ ARCHIWIZACJI

Osoba wypełniająca ankietę: Marcin Werla, mwerla@man.poznan.pl

<http://storage.pionier.net.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Konsorcjum

Nazwa podmiotu realizującego projekt: Poznańskie Centrum Superkomputerowo-Sieciowe

Institucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Wykonwacy projektu PLATON - uczestnicy konsorcjum PIONIER (22 jednostki MAN): <http://www.platon.pionier.net.pl/online/konsorcjum.php>

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: dr inż. Norbert Meyer, meyer@man.poznan.pl

Osoba kontaktowa: Maciej Brzeźniak, maciekb@man.poznan.pl

Pozostałe osoby realizujące projekt:

Szczegółowych informacji udziela osoba kontaktowa.

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 7/1/2008 - 7/25/2012

Źródło finansowania: Program Operacyjny "Innowacyjna Gospodarka" POIG.02.03.00-00-028/08

OPIS PROJEKTU

Typ projektu: infrastrukturalny

Efekt projektu: repozytorium

Słowa kluczowe: archiwizacja; zabezpieczenie danych; kopie zapasowe; dane badawcze

Skrócony opis projektu:

Usługi powszechnej archiwizacji polegać będą na udostępnieniu w skali kraju funkcjonalności zdalnej archiwizacji i backupu danych jako wartości dodanej do ogólnopolskiej akademickiej sieci naukowej PIONIER. Usługa ta jest jednym z elementów koniecznych do zwiększenia niezawodności funkcjonowania każdej jednostki, a skierowana została do środowiska akademickiego, w tym uczelni wyższych, jednostek badawczo-rozwojowych oraz szpitali zależnych od uniwersytetów oraz akademii medycznych. Zapewnienie odpowiedniej jakości usługi zrealizowane będzie przez przedstawicieli jednostek MAN oraz KDM w największych środowiskach akademickich. W 10 lokalizacjach miejskich rozmieszczone zostaną węzły przechowywania danych oraz węzły usługowe, rozproszone geograficznie połączone z optycznymi sieciami miejskimi (zapewniając odpowiednią przepustowość tzw. 'ostatniej mili') oraz siecią krajową (zapewniając odpowiedniej

jakości połączenia szkieletowe). Wybór technologii oraz konfiguracja sprzętowa dokonana została z uwzględnieniem parametrów funkcjonalnych oraz ekonomicznych.

Usługi powszechnej archiwizacji umożliwiać będą archiwizację danych (tworzenie kopii zapasowej danych, zlokalizowanej w odległym geograficznie miejscu) wraz z możliwością dostępu on-line oraz odtworzeniem kopii zapasowej w przypadku utraty danych podstawowych w trybie 24h x 7 dni w tygodniu. Opisana funkcja jest przykładem outsourcingu usług i dostępna będzie na terenie całego kraju w ramach sieci miejskich oraz w niedalekiej przyszłości w szerokopasmowych sieciach regionalnych.

Skrócony opis części informatycznej:

Patrz opis projektu.

Opis efektów projektu:

Efektom projektu jest Usługa Powszechnej Archiwizacji umożliwiająca przechowywanie danych, tworzenie kopii zapasowych i archiwalnych w sieci PIONIER. Usługa skierowana jest do środowiska akademickiego tzn. do jednostek prowadzących działalność naukową lub edukacyjną, do których stosują się przepisy ustawy z dnia 30 kwietnia 2010 r. o zasadach finansowania nauki (dz. U. Z 2010 r. Nr 96, poz. 615, Art. 2, lit. 9), a także instytucji edukacyjnych i oświatowych, jednostek administracji rządowej lub samorządowej, instytucji kulturalnych, bibliotek, jednostek służby zdrowia, instytucji wyższej użyteczności i pożytku publicznego, mająca zawartą umowę na realizację usług telekomunikacyjnych w sieci MAN, tj. będąca abonentem sieci MAN.

Powiązania z innymi polskimi i zagranicznymi projektami:

Projekt EUDAT

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Wiki projektu: <https://www.storage.pionier.net.pl/wiki>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Usługa Powszechnej Archiwizacji jest dalej rozwijana m.in. w ramach projektu KMD2 (<http://kmd.pcss.pl/>).

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

- instytucje zainteresowane archiwizacją danych naukowych

WIRTUALNE LABORATORIUM TRANSKRYPCJI

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://wlt.synat.pcss.pl/>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Poznańskie Centrum Superkomputerowo-Sieciowe

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: WLT realizowane jest przez PCSS w ramach projektu SYNAT.

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: dr inż. Cezary Mazurek, mazurek@man.poznan.pl

Osoba kontaktowa: Marcin Werla, mwerla@man.poznan.pl

Pozostałe osoby realizujące projekt:

dr inż. Cezary Mazurek, mazurek@man.poznan.pl; mgr inż. Adam Dudczak, maneo@man.poznan.pl; mgr inż. Marcin Werla, mwerla@man.poznan.pl

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 8/16/2010 - 7/31/2014

Źródło finansowania: Narodowe Centrum Badań i Rozwoju SP/I/1/77065/10

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: narzędzie badawcze, oprogramowanie

Słowa kluczowe: OCR, transkrypcja, biblioteki cyfrowe, digitalizacja, korekta

Skrócony opis projektu:

Wirtualne Laboratorium Transkrypcji to narzędzie, które pozwala tworzyć przeszukiwalny, cyfrowy tekst na podstawie zeskanowanych dokumentów historycznych. Narzędzie powstało w ramach projektu SYNAT.

Skrócony opis części informatycznej:

Część informatyczna WLT to portal, poprzez który użytkownicy mogą przysyłać skany publikacji, poddawać je procesowi automatycznego rozpoznawania znaków (OCR), a następnie prowadzić zespołową transkrypcję i korektę wyników przetwarzania OCR. Uzyskane wyniki można eksportować do kilku popularnych formatów tekstowych.

Opis efektów projektu:

Głównym efektem projektu jest portal, poprzez który użytkownicy mogą przysyłać skany publikacji, poddawać je procesowi automatycznego rozpoznawania znaków (OCR), a następnie prowadzić zespołową transkrypcję i korektę wyników przetwarzania OCR. Uzyskane wyniki można eksportować do kilku popularnych formatów tekstowych. Możliwy jest również import zbiorów bezpośrednio z wybranych bibliotek cyfrowych. Dodatkowo można też poprawiać wyniki procesu OCR poprzez przygotowywanie profili rozpoznawania tekstu dopasowanych do specyfiki konkretnych dokumentów.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Wprowadzenie do WLT:

<https://confluence.man.poznan.pl/community/display/WLT/Wprowadzenie+do+Wirtualnego+Laboratorium+Transkrypcji>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

- poprawa jakości wyników OCR, m.in. poprzez integrację z narzędziami lingwistycznymi
- wsparcie dla nowych formatów danych wejściowych

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

WIZUALIZACJA SZTUKI

Osoba wypełniająca ankietę: Kierownik projektu

<http://wizualizacjasztuki.umcs.lublin.pl>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Kulturoznawstwa UMCS

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Radosław Bomba, radoslawbomba@gmail.com

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Studenci kulturoznawstwa UMCS

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 3/23/2012 - 6/9/2020

Źródło finansowania: środki własne instytucji

OPIS PROJEKTU

Typ projektu: badawczy, digitalizacyjny, edukacyjny

Efekt projektu: repozytorium, publikacja, kurs

Słowa kluczowe: wizualizacja, humanistyka cyfrowa, wizualizacja mediów, analityka kulturowa

Skrócony opis projektu:

Projekt "Wizualizacja sztuki" zainspirowany został badaniami z pogranicza nauk humanistycznych i nowych mediów określanych jako humanistyka cyfrowa, analityka kulturowa lub kulturonomika. Zasadniczym założeniem tych koncepcji jest analiza dużych ilości danych kulturowych z użyciem metod z pogranicza badań statystycznych i jakościowych oraz oprogramowania komputerowego i aplikacji sieciowych. Istotnym aspektem tego typu analiz jest wykorzystanie wizualizacji danych jako elementu metody badawczej.

Celem przedsięwzięcia jest pokazanie nowych możliwości jakie daje naukom humanistycznym wykorzystanie w badaniach nad sztuką i kulturą narzędzi cyfrowych i sieciowych aplikacji służących do wizualizowania danych.

Zgromadzone w ramach projektu badania opierają się na analizie twórczości artystów w wybranych przez zespoły badawcze aspektach. Dzięki wizualizacji problematyki związanej m.in. z kwestiami: stylu, zmiany tematyki i treści dzieł, częstotliwości tworzenia dzieł przez artystę, badający mogli zaobserwować istotne prawidłowości i tendencje w twórczości danego artysty i powiązać je z szerszym kontekstem kulturowym.

Projekt powstał na zajęciach z przedmiotu "Multimedia" prowadzonych przez dr Radosława Bombę. Zajęcia zrealizowane zostały w Instytucie Kulturoznawstwa UMCS w Lublinie

Skrócony opis części informatycznej:

Studenci podzieleni na zespoły robocze w pierwszej kolejności opracowali tematy badań. Po uściśleniu tematyki przystępowali do gromadzenia odpowiednich danych. Przygotowany w ten sposób korpus danych był wizualizowany i interpretowany. Bazując na zwizualizowanych danych studenci wyszukiwali charakterystyczne trendy obecne w twórczości danego artysty.

Dopełnienie przygotowanych wizualizacji stanowiła kulturoznawcza analiza sformułowanych tematów oraz metoda biograficzna, tłumacząca szersze tendencje zaobserwowane w skali całej twórczości danego artysty faktami z jego życia.

Stworzone przez studentów grupy badawcze wykorzystały do swoich analiz: linie czasu przedstawiając z jej pomocą twórczość danego artysty lub jego sylwetkę, aplikacje do wizualizacji danych (Meny Eyes), aplikacje do wizualizacji tekstu (Wordle), aplikacje do wizualizacji trendów kulturowych (Ngram Viewer), mapy, narzędzia do analiz formalnych i badania stylu artysty (ImagePlot). Dopełnieniem były wizualizacje stworzone samodzielnie przez studentów. Każdy zespół dodatkowo stworzył również bloga poświęconego badanemu artyście.

Opis efektów projektu:

Projekty wizualizacyjne poświęcone różnym artystom i ich twórczości

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://wizualizacjasztuki.umcs.lublin.pl/>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Instytucja zainteresowana wizualnymi badaniami nad sztuką

WYSZUKIWANIE W DOKUMENTACYJNYCH SYSTEMACH INFORMACYJNYCH W ŚWIETLE POTRZEB ŚRODOWISKA AKADEMICKIEGO

Osoba wypełniająca ankietę: Grzegorz Gmiterek

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Grant promotorski

Nazwa podmiotu realizującego projekt: IINiSB UW

Wykonawca: Anna Mierzecka-Szczepańska, anna.mierzecka@uw.edu.pl

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: prof. dr hab. Barbara Sosińska-Kalata, b.sosinska@uw.edu.pl

Osoba kontaktowa: Anna Mierzecka-Szczepańska

Pozostałe osoby realizujące projekt:

Anna Mierzecka-Szczepańska

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 11/6/2009 - 3/25/2011

Źródło finansowania: MNiSW N N103 021137

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: publikacja

Słowa kluczowe: dokumentacyjne systemy informacyjne; nauki humanistyczne; bibliologia; informatologia; wyszukiwanie informacji

Skrócony opis projektu:

Prowadzone badania mają na celu identyfikację i charakterystykę metod podnoszenia satysfakcji użytkowników końcowych dokumentacyjnych systemów informacyjnych. W niniejszym badaniu analizowane są potrzeby i oczekiwania przedstawicieli nauk humanistycznych polskiego środowiska naukowego. W tym celu do udziału w badaniach zaproszono pracowników i doktorantów sześciu instytutów Uniwersytetu Warszawskiego: Instytutu Archeologii, Instytutu Etnologii i Antropologii Kulturowej, Instytutu Historii Sztuki, Instytutu Historycznego, Instytutu Informacji Naukowej i Studiów Bibliologicznych oraz Instytutu Muzykologii.

Skrócony opis części informatycznej:

j.w.

Opis efektów projektu:

j.w.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

WYSZUKIWANIE W DOKUMENTACYJNYCH SYSTEMACH INFORMACYJNYCH W ŚWIETLE POTRZEB ŚRODOWISKA AKADEMICKIEGO

Osoba wypełniająca ankietę: Osoba kontaktowa

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Instytut Informacji Naukowej i Studiów Bibliologicznych, Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: prof. dr hab. Barbara Irena Sosińska-Kalata tel.: (0-22) 649-15-71 e-mail: b.sosinska@uw.edu.pl

Osoba kontaktowa: Anna Mierzecka-Szczepanska, mail: anna.mierzecka@uw.ed.pl

Pozostałe osoby realizujące projekt:

Barbara Irena Sosińska-Kalata; Anna Mierzecka- Szczepańska UW

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 11/9/2009 - 6/30/2012

Źródło finansowania: Narodowe Centrum Nauki N N103 021137

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: publikacja

Słowa kluczowe: zachowania informacyjne, wyszukiwanie informacji, bazy literatury naukowej, relewancja

Skrócony opis projektu:

Prowadzone badania miały na celu identyfikację i charakterystykę metod podnoszenia satysfakcji użytkowników końcowych dokumentacyjnych systemów informacyjnych. W pierwszym etapie badań na podstawie analizy literatury przedmiotu, ustalono w jaki sposób mierzono dotąd mierzony był poziom satysfakcji użytkowników oraz które ze stosowanych metod mogłyby zostać wykorzystane w badaniach autorki. W prowadzonych badaniach empirycznych przeanalizowano potrzeby i oczekiwania przedstawicieli nauk humanistycznych polskiego środowiska naukowego. W tym celu do udziału w badaniach empirycznych

zaproszono pracowników i doktorantów sześciu instytutów Wydziału Historycznego Uniwersytetu Warszawskiego.

Skrócony opis części informatycznej:

brak

Opis efektów projektu:

Analiza zebranego materiału empirycznego stanowiła podstawę do wyciągnięcia wniosków na temat sposobów wyszukiwania informacji przez przedstawicieli środowiska akademickiego, ich potrzeb i oczekiwań wobec oferty informacyjnej niezbędnej do realizacji zadań badawczych i dydaktycznych.

Powiązania z innymi polskimi i zagranicznymi projektami:

Brak

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

1. Anna Mierzecka-Szczepańska (2013): Badania zachowań informacyjnych. Warszawa: Wydawnictwo SBP, 257 s. ISBN:978-83-64203-04-6.

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Na razie nie planuje się zadań mających bezpośredni związek z projektem

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Brak

Profil potencjalnego partnera:

WYSZUKIWARKA DOKUMENTÓW SUMERYJSKICH Z III DYNASTII Z UR

Osoba wypełniająca ankietę: Kierownik projektu

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Instytut Informatyki

Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Instytut Historyczny

Uniwersytet Warszawski

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Jerzy Tyszkiewicz jty@mimuw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Wojciech Jaworski II UW;

Marek Stępień IH UW

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: -

Źródło finansowania: dawny grant ministerialny

OPIS PROJEKTU

Typ projektu: badawczy

Efekt projektu: oprogramowanie, baza danych

Słowa kluczowe: wyszukiwarka tekstów; Sumer; III Dynastia z Ur

Skrócony opis projektu:

Powstała wyszukiwarka tekstów sumeryjskich, głównie z okresu II Dynastii z Ur, uwzględniająca wszystkie warianty zapisu wyszukiwanych terminów.

Skrócony opis części informatycznej:

Powstała wyszukiwarka tekstów sumeryjskich, głównie z okresu II Dynastii z Ur, uwzględniająca wszystkie warianty zapisu wyszukiwanych terminów.

Opis efektów projektu:

Powstała wyszukiwarka tekstów sumeryjskich, głównie z okresu II Dynastii z Ur, uwzględniająca wszystkie warianty zapisu wyszukiwanych terminów. Aktualnie oprogramowanie jest nieczynne.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Projekt aktualnie nieczynny. Wymagałby ponownego uruchomienia oprogramowania i poważnej aktualizacji bazy tekstów, które indeksował. Ze względu na brak środków i osób, aktualnie nie planujemy tego działania.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

ZESTAWIENIE ZBORÓW PROTESTANCKICH W RZECZYPOSPOLITEJ XVI-XVIII STULECIU

Osoba wypełniająca ankietę: Kierownik projektu

<http://ihuw.pl/instytut/o-instytucie/zespoly/zespol-historii-kultury-staropolskiej/zestawienie-zborow>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Zespół Historii Kultury Staropolskiej Instytutu Historycznego Uniwersytetu Warszawskiego, Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: -

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Maciej Ptaszynski, m.ptaszynski@uw.edu.pl

Osoba kontaktowa: Kierownik projektu

Pozostałe osoby realizujące projekt:

Marek Janicki (IHUW); Agnieszka Kuś; Rafał Jaworski; Anna Wilanek-Kińska; Krzysztof Kubik; Maciej Próba; Anna Szacoń.

INFORMACJE ORGANIZACYJNE

Status projektu: Zakończony

Okres realizacji projektu: 6/15/2002 - 6/15/2012

Źródło finansowania: dotacja celowa, środki własne instytucji

OPIS PROJEKTU

Typ projektu: badawczy, bazodanowy

Efekt projektu: narzędzie badawcze, baza danych

Słowa kluczowe: Protestanci; reformacja; zbory

Skrócony opis projektu:

Zestawienie obejmuje zbory protestanckie czynne w Rzeczypospolitej w XVI-XVIII stuleciu.

Skrócony opis części informatycznej:

Baza danych ma charakter relacyjny i została wykonana w programie Access. Informacje zawarte w bazie danych zostały przedstawione w formie elektronicznej mapy, opracowanej przez dr. hab. Bogumiła Szadego (KUL).

Link do mapy:

http://hgis.kul.lublin.pl/pmapper-4.0/map_zbory.phtml?winsize=large&language=pl&config=zbory

Opis efektów projektu:

W opublikowanej bazie znajdują się wszystkie zbory, o których wzmianki można znaleźć w literaturze przedmiotu. W niedostępnej części bazy danych uwzględniono również ministrów.

W przygotowaniu opublikowanej części wykorzystano: Henryk Merczyng, Zbory i senatorowie protestancy w dawnej Rzeczypospolitej. Z mapą Dawnej Polski pod względem kościelnym ewangelickim, Warszawa 1905; Continuatio Historiae Ecclesiasticae Slavoniae Reformata ab Anno MDCL ad Annom MDCXCVI, , [w:] H. Barycz, Obraz historyczny kalwinizmu na Litwie 1650-1696, "Reformacja w Polsce", IV, 1926, 13-16, s. 213-214; Rejestr zborów, szkół i księży in Magno Ducatu Lithuaniae anno 1748, [w:] S. Tworek, Z zagadnień liczebności zborów kalwińskich na Litwie, „Odrodzenie i Reformacja w Polsce”, XVII, 1972, s. 213-214; Józef Łukaszewicz, Dzieje kościołów wyznania helweckiego w Litwie, t. 1, Poznań 1842; Józef Łukaszewicz, Dzieje kościołów wyznania helweckiego w Litwie, t.2, Poznań 1843; Józef Łukaszewicz, O kościołach Braci Czeskich w dawnej Wielkiej Polsce, Poznań 1835; Eduard Kneifel, Die Pastoren der Evangelisch-Augsburgischen Kirche in Polen, Neuendettelsau b.d.; Jolanta Dworzaczkowa, Bracia czescy w Wielkopolsce w XVI i XVII wieku, Warszawa 1997; Wojciech Kriegseisen, Ewangelicy polscy i litewscy w epoce saskiej (1696-1763) Sytuacja prawna, organizacja i stosunki międzywyznaniowe, Warszawa 1996; Protokoły obrad synodów prowincjonalnych wileńskich z lat 1611-1686, AGAD, Archiwum Radziwiłłów, dz. VIII, nr 713.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Projekt powinien zostać rozwinięty przy wykorzystaniu źródeł archiwalnych. Powinien ponadto zostać uzupełniony o dane biograficzne zatrudnionych ministrów.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Projekt Luise Schorn-Schütte Geistlichkeit in der Frühen Neuzeit. Ein Vergleich der Sozialbiographien evangelischer Pfarrer und des katholischen Seelsorgeklerus im Alten Reich, in der polnischen Adels-republik und in der schweizerischen Eidgenossenschaft (Frankfurt-M.)

Profil potencjalnego partnera:

Samodzielna instytucja badawcza; katedra uniwersytecka lub network badawczy (np. finansowany przez DFG)

ZINTEGROWANA BAZA BIBLIOGRAFICZNA E-ZBB IBL PAN

Osoba wypełniająca ankietę: Osoba kontaktowa

<http://bar.ibl.waw.pl>

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Instytucja

Nazwa podmiotu realizującego projekt: Instytut Badań Literackich Polskiej Akademii Nauk

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: brak

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Joanna Krauze-Karpińska, joanna.krauze-karpinska@ibl.waw.pl

Osoba kontaktowa: Maciej Pieczyński, maciej.pieczynski@ibl.waw.pl

Pozostałe osoby realizujące projekt:

Zespół pod kierunkiem dr Joanny Krauze-Karpińskiej

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 1/30/2012 - 1/30/2017

Źródło finansowania: Narodowy Program Rozwoju Humanistyki

OPIS PROJEKTU

Typ projektu: badawczy, bazodanowy, digitalizacyjny

Efekt projektu: baza danych, standardy opisu i przetwarzania danych bibliograficznych o charakterze zgodnym ze specyfiką projektu

Słowa kluczowe: bibliografia zawartości czasopism - XIX wiek; bibliografia zawartości czasopism - do 1945; czasopisma literackie - XIX wiek; czasopisma literackie - do 1945; bibliografia staropolskich literatów rękopiśmiennych; literatura staropolska - bibliografia

Skrócony opis projektu:

"Zintegrowana Baza Bibliograficzna IBL PAN" jest długofalowym projektem mającym na celu digitalizację oraz udostępnienie w internecie zawartości bibliograficznych baz danych wytworzonych przez pracowników IBL. W pierwszym etapie realizacji zakłada się:

1. opracowanie zasad tworzenia rozbudowanej elektronicznej bazy danych e-ZBB na podstawie bibliografii, kartotek oraz katalogów istniejących w postaci papierowej, gromadzonych latami w IBL PAN przez zespoły wykwalifikowanych bibliografów, współpracujących z historykami, historykami literatury, kultury i nauki;

2. opracowanie schematów elektronicznej rejestracji danych przechowywanych do tej pory w postaci rękopiśmiennych fiszek;

3. opracowanie w wersji cyfrowej i udostępnienie dwóch zespołów danych bibliograficznych: Indeksu Rękopiśmiennych Literariów Staropolskich oraz części Retrospektywnej Bibliografii Zawartości Czasopism XIX i XX w. (tzw. kartoteki Bara).

Skrócony opis części informatycznej:

W ramach projektu tworzone są bazy danych w programie MAK, zawierające opisy bibliograficzne w formacie MARC21. Bazy te są udostępniane w sieci w postaci strony WWW. Na potrzeby prac digitalizacyjnych zorganizowana została stała pracownia komputerowa wyposażona w 20 stanowisk roboczych.

Opis efektów projektu:

Efektom projektu będzie wytworzenie i udostępnienie w sieci wszystkim zainteresowanym w postaci łatwo przeszukiwalnych baz danych: bibliografii literackiej zawartości czasopism do roku 1945 oraz indeksu staropolskich literariów rękopiśmiennych. Obie bazy przedstawiają znaczną wartość naukową dla historyków, literaturoznawców, antropologów, teatrologów czy historyków muzyki.

Powiązania z innymi polskimi i zagranicznymi projektami:

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

Opis projektu:

<http://bar.ibl.waw.pl/oproj.htm>

Opis kartoteki Bara:

Jadwiga Czachowska, Rozwój bibliografii literackiej w Polsce, Wrocław 1979, s. 152-164.

<http://bar.ibl.waw.pl/bar-opis.htm>

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Wykorzystanie wypracowanych w 1. etapie metod i standardów pracy: digitalizacja i udostępnienie na stronie internetowej kolejnych zespołów danych bibliograficznych.

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

Profil potencjalnego partnera:

Instytucja zainteresowana digitalizacją posiadanych kartotek zawierających dane o charakterze bibliograficznym, zwłaszcza dotyczących literatury, dziennikarstwa, życia społecznego, kulturalnego bądź naukowego, krytyki artystycznej, literackiej lub muzycznej, bądź dziedzin pokrewnych.

ZNALEZISKA MONET RZYMSKICH Z ZIEM POLSKI I HISTORYCZNIE Z POLSKĄ ZWIĄZANYCH

Osoba wypełniająca ankietę: Kierownik projektu

www.frcpl.uw.edu.pl

INSTYTUCJE ZAANGAŻOWANE W PROJEKT

Rodzaj podmiotu realizującego projekt: Zespół badawczy

Nazwa podmiotu realizującego projekt: Instytut Archeologii, Wydział Historyczny, Uniwersytet Warszawski

Instytucja przy której projekt jest afiliowany: Jak wyżej

Pozostałe instytucje zaangażowane w projekt: Uniwersytet Jagielloński; Uniwersytet Mikołaja Kopernika w Toruniu; Zakład Narodowy im. Ossolińskich; Uniwersytet Karazina w Charkowie (Ukraina), Muzeum Narodowe w Warszawie

OSOBY ZAANGAŻOWANE W PROJEKT

Kierownik projektu: Aleksander Bursche abursche@yahoo.com

Osoba kontaktowa: Anna Zapolska anna.zapolska@gmail.com

Pozostałe osoby realizujące projekt:

Aleksander Bursche, UW, Anna Zapolska UW, Marcin Rudnicki UW; Jarek Bodzek UJ; Kamil Kopij, UJ; Katarzyna Lach, UJ; Emilia Smagur, UJ; Jacek Rakoczy UMK; Arkadiusz Dymowski, ABW; Adam Degler, Ossolineum, Andrzej Romanowski, MNW; Kyrillo Myzgin, Uniwersytet Karazina,

INFORMACJE ORGANIZACYJNE

Status projektu: Realizowany

Okres realizacji projektu: 12/20/2012 - 12/19/2017

Źródło finansowania: Narodowy Program Rozwoju Humanistyki 0047/NPRH2/H11/81/2012

OPIS PROJEKTU

Typ projektu: badawczy, bazodanowy

Efekt projektu: narzędzie badawcze, baza danych, publikacja

Słowa kluczowe: Monety rzymskie, skarby, importy rzymskie, inwentaryzacja

Skrócony opis projektu:

Głównym celem projektu jest stworzenie w j. angielskim komputerowej bazy danych on-line wszystkich znanych znalezisk monet rzymskich (III w. przed Chr. – VI w. po Chr.) z ziem Polski i historycznie z Polską związanych, bazy zintegrowanej z innymi ośrodkami europejskimi, umożliwiające sukcesywne jej uzupełnianie o nowe odkrycia oraz dokonywanie automatycznych przeszukiwań, kwerend, analiz statystycznych i zestawień kartograficznych. Istotnym efektem ma być opublikowanie, zgodnych z obowiązującymi standardami, inwentarzy znalezisk monet.

Skrócony opis części informatycznej:

Własna komputerowa baza danych on-line umieszczona na serwerze UW, kompatybilna z innymi analogicznymi europejskimi projektami, możliwości przeszukiwania wedle dowolnego klucza, edytowania w formacie World, analiz statystycznych i kartografowania, pełny dostęp dla uprawnionych użytkowników, ograniczony dla zewnętrznych. Możliwości komputerowej analizy porównawczej wyników spektrograficznych analiz metalograficznych monet.

Opis efektów projektu:

Oprogramowanie dla komputerowej bazy danych on-line znalezisk monet rzymskich już istnieje na serwerze UW. Obecnie zespół pracuje nad wprowadzaniem do bazy odpowiednich danych. Baza zostanie udostępniona on-line dla wszystkich zainteresowanych użytkowników, ponadto na jej podstawie zostaną opublikowane odpowiednie katalogi z mapami i opracowaniami statystycznymi. Całość będzie stanowić docelowo ważne narzędzie badawcze dla pokoleń archeologów, historyków starożytnych i numizmatyków.

Powiązania z innymi polskimi i zagranicznymi projektami:

W Europie funkcjonuje kilkanaście dużych zespołów badawczych realizujących analogiczne projekty skupionych w sieci European Coin Find Network por. <http://www.ecfn.fundmuenzen.eu/>, poszczególne projekty: <http://www.ecfn.fundmuenzen.eu/index.php/partners>; najbliższa współpraca z Fundmuenze der Antike we Frankfurcie: http://141.2.2.167:8080/AFE_RGK_1, Oxford Roman Economy Project: http://oxrep.classics.ox.ac.uk/coin_hoards_of_the_roman_empire_project/ oraz Numisma.org w American Numismatic Society <http://nomisma.org>

Referencje do dodatkowych materiałów informacyjnych na temat projektu bądź jego wyników:

<http://www.ecfn.fundmuenzen.eu/index.php/news/carnuntum-2013>

http://www.ecfn.fundmuenzen.eu/images/Western_Ukraine.pdf

PLANY ROZWOJU PROJEKTU

Wizja rozwoju projektu:

Kontynuacja bazy danych znalezisk monet on-line, optymalizacja statystycznego i kartograficznego przetwarzania danych

Możliwe powiązania z innymi polskimi i zagranicznymi projektami:

większość skupionych jest w światowej sieci European Coin Find Network por. <http://www.ecfn.fundmuenzen.eu/>, z którą mamy stałą współpracę.

Profil potencjalnego partnera:

środowiska numizmatyków amatorów i poszukiwaczy skupionych w towarzystwach i na portalach internetowych