

Celem projektu jest stworzenie wielojęzycznego systemu, który umożliwi dostawcom informacji przeszukiwanie i integrację zasobów wielkich kolekcji danych tekstowych oraz multimedialnych, takich jak biblioteki internetowe, zasoby wydawnictw i agencji prasowych oraz naukowe bazy danych.

System

Wielojęzyczny system agregacji informacji (M-CAST), poprzez agregację rozproszonych danych, dostępnych w różnych formatach, pozwoli na tworzenie nowoczesnych bibliotek cyfrowych. System zostanie przetestowany w dwóch bibliotekach, w których stworzone zostaną multimedialne portale agregujące informacje (M-CAP), bazujące na wcześniej stworzonych systemach i portalach. Będą one umożliwiały znalezienie w wielkich kolekcjach wielojęzycznych danych odpowiedzi na pytania zadawane w jednym z sześciu języków naturalnych. Warstwa prezentacji występująca w tych portalach będzie obsługiwała dane multimedialne i umożliwi prezentację cyfrowych obrazów starodruków, oryginałów dokumentów, zapisów muzycznych, fotografii, plików wideo itp. Oczywiście indeksowane będą mogły być tylko tekstowe opisy tych obiektów.

Techniki lingwistyczne

Projekt M-CAST będzie wykorzystywał wyniki projektu TRUST – Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (Wielojęzyczny mechanizm semantyczny i kognitywny do wyszukiwania tekstów z wykorzystaniem technik semantycznych – IST-1999-56416) dofinansowanego w ramach Piątego Programu Ramowego Badań, Rozwoju i Prezentacji Unii Europejskiej (<http://www.trustsemantics.com>). Wyszukiwarka TRUST, umożliwiająca wyszukiwanie w czterech językach (francuskim, polskim, portugalskim i włoskim), dostępna jest dotychczas w jednostronowej wersji na komputery

osobiste typu PC. W ramach projektu zostanie ona przeniesiona na serwery pracujące pod kontrolą systemu operacyjnego Unix lub Windows. Zasoby językowe zgromadzone w ramach projektu TRUST zostaną zaktualizowane i wzbogacone. Wykorzystywana dotychczas taksonomia języka zostanie dostosowana do uniwersalnej klasyfikacji dziesiętnej (UDC) używanej przez biblioteki na całym świecie do tworzenia katalogów tematycznych. System zostanie wzbogacony o dwa dodatkowe języki: angielski, dodany przez jednego z partnerów projektu TRUST, oraz czeski, w ramach projektu M-CAST.

Metodyka

System agregacji informacji M-CAST będzie centralnym elementem portalu agregującego M-CAP, który zostanie zbudowany zgodnie z metodyką budowy systemów zarządzania informacją opartych na wiedzy, stworzoną przez firmę Infovide S.A. w ramach innego projektu dofinansowanego z 5. Programu Ramowego ICONS – Intelligent Content Management System – IST-2001-32429 (<http://www.icons.rodan.pl>).

Wdrożenia

Portale zostaną wdrożone i przetestowane w dwóch bibliotekach publicznych: w Polskiej Bibliotece Internetowej (<http://www.pbi.edu.pl>), której operatorem jest Wojewódzka Biblioteka Publiczna – Książnica Kopernikańska w Toruniu oraz w Bibliotece Narodowej Republiki Czeskiej (<http://www.nkp.cz>). Informacje będą mogły być wyszukiwane w wielojęzycznych, cyfrowych zasobach bibliotek za pomocą zapytań sformułowanych w języku naturalnym. Odpowiedzią będą fragmenty tekstów zawierające odpowiedź na pytanie.

Zastosowania

Po zakończeniu projektu powstanie produkt ułatwiający zarządzanie wielojęzyczną wiedzą zgromadzoną w dużych zbiorach danych tekstowych. Przewiduje się, że znajdzie on zastosowanie w bibliotekach (wyszukiwanie informacji, katalogi tematyczne), zarządzaniu zbiorami (pozyskiwanie i agregacja danych, statystyki wykorzystania, poprawa jakości danych), bibliograficznych bazach danych, serwisach informacyjnych (selektywne upowszechnianie informacji, personalizacja, odkrywanie wiedzy) oraz w semantycznych sieciach danych.

Projekt programu eContent nr EDC 22249 M-CAST



Koordynator projektu:

Infovide S.A.
ul. Kolejowa 5/7
01-517 Warszawa
tel.: (22) 534 74 00, 534 74 01
fax: (22) 534 74 02

Osoba do kontaktów:

Borys Czerniejewski
Dyrektor Biura Współpracy
Międzynarodowej
tel.: (22) 534 74 50
bczerniejewski@infovide.pl

Partnerzy z Polski:

- Wojewódzka Biblioteka Publiczna – Książnica Kopernikańska, Toruń, Polska
- TIP sp. z o.o., Katowice, Polska

Pozostali partnerzy:

- Synapse Développement SARL, Tuluza, Francja
- Priberam Informática Lda., Lizbona, Portugalia
- Expert System S.p.A., Modena, Włochy
- Národní knihovna České republiky, Praga, Czechy
- Vysoká škola ekonomická v Praze, Praga, Czechy

Projekt M-CAST został dofinansowany przez Wspólnotę Europejską w ramach wieloletniego programu wspólnotowego mającego na celu stymulację rozwoju i wykorzystania europejskich treści cyfrowych w globalnych sieciach komputerowych oraz promocję różnicowania językowego w społeczeństwie informacyjnym (2001–2005).