



eContent project № EDC 22249 M-CAST



The aim of the project is to develop a multilingual infrastructure enabling content producers to access, search and integrate the assets of large multilingual text (and multimedia) collections, such as internet libraries, resources of publishing houses, press agencies and scientific databases.

and updated. The language ontology (taxonomy) used in the engine will be converted to be compliant with the standard Universal Decimal Classification (UDC) used in library cataloguing systems worldwide. Two other languages modules will be added: English - developed by one of the TRUST partners and Czech - developed during the M-CAST project.

System

The Multilingual Content Aggregation System (M-CAST) will allow for the development of Digital Libraries by aggregating digital data available in different formats and locations. The system will be tested by two libraries, for which Multimedia Content Aggregation Portals (M-CAPs) will be created, using their existing portals and infrastructures. The portals will allow to find answers to natural language queries in large digital collections of multilingual data. The presentation layer of the portals will be multimedia capable, allowing for presentation of digitalized copies of old printings, legal documents, music scores, pictures, video etc., while only their textual descriptions will be indexed

Methodology

The M-CAST aggregation system will be the central element of the M-CAP aggregation portal and will be developed following the Knowledge-based Content Management Application Design Methodology prepared by Infovide S.A. in another 5th Framework Programme project - ICONS - Intelligent Content Management System - IST-2001-32429 (<http://www.icons.rodan.pl>).

Implementation

The portals will be deployed and tested in two public libraries: the Polish Internet Library (<http://www.pbi.edu.pl>) and the National Library of the Czech Republic (<http://www.nkp.cz>) to make their digital resources available online for finding answers to natural language queries in multilingual digital collections.

Linguistic technologies

M-CAST will be based on the outcome of the TRUST - Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (IST-1999-56416) project financed by the 5th European Union Framework Programme for RTD (<http://www.trustsemantics.com>). The TRUST search engine, developed for 4 languages (French, Italian, Polish and Portuguese) will be transformed from a stand-alone, PC-based program into a server application. The TRUST language resources will be used

Application

At the end of the project a marketable product for multilingual knowledge management purposes will be available. Commercial exploitation of the results in libraries (information retrieval, classified catalogues), collection management (data acquisition and aggregation, circulation statistics, weeding), bibliographic databases, information services (selective dissemination of information, personalisation, knowledge discovery) and in semantic data networks is foreseen.

Project Coordinator:

Infovide S.A.
ul. Kolejowa 5/7
01-517 Warszawa
POLAND
ph: +(48-22) 534 74 00, 534 74 01
fax: +(48-22) 534 74 02

Consortium Partners:

- Polish Internet Library, Toruń, Poland
- TIP sp. z o.o., Katowice, Poland
- Synapse Développement SARL, Toulouse, France
- Priberam Informática Lda., Lisbon, Portugal
- Expert System S.p.A., Modena, Italy
- Národní knihovna České republiky, Prague, Czech Republic
- Vysoká škola ekonomická v Praze, Prague, Czech Republic

Contact Person:

Borys Czerniejewski
Director, International Liaison Office
ph: +(48-22) 534 74 50
bczerniejewski@infovide.pl

The M-CAST project is carried out with the financial support of the European Community in the framework of the multiannual community programme to stimulate the development and the use of European digital content on the global networks and to promote linguistic diversity in the information society (2001-2005).