



Cílem projektu je vyvinout vícejazyčný systém, který bude umožňovat tvůrcům obsahu prohledávat a integrovat rozsáhlé soubory textů (a multimédií), jako jsou internetové knihovny, informační zdroje nakladatelství a tiskových agentur nebo databáze vědeckých informací.

Systém

Vícejazyčný systém agregace informací (M-CAST, Multilingual Content Aggregation System) umožní tvorbu digitálních knihoven prostřednictvím agregace dat dostupných v různých formátech a z různých zdrojů. Systém bude testován dvěma knihovnami, pro které budou vytvořeny portály agregující multimediální informace (M-CAP, Multimedia Content Aggregation Portals) na základě existujících portálů a systémů. Tyto portály budou schopny hledat odpovědi na dotazy v přirozeném jazyce v rozsáhlých souborech vícejazyčných dat. Prezentací vrstva portálu bude multimediální, díky čemuž bude možné zpřístupnit digitalizované kopie starých tisků, právních dokumentů, notových záznamů, obrázků a videa, i když indexovány budou pouze jejich textové popisy.

Lingvistické technologie

Systém M-CAST bude založen na výsledcích projektu TRUST - Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (Vícejazyčný sémantický a kognitivní mechanismus vyhledávání textů využívající sémantické technologie, IST-1999-56416), který byl financován z Pátého rámcového programu EU pro rozvoj vědy a výzkumu (<http://www.trustsemantics.com>). Vyhledávací stroj TRUST umí v současnosti vyhledávat ve čtyřech jazycích (francouzštině, italštině, polštině a portugalštině). V rámci projektu bude transformován z jednorázové aplikace pro

PC na serverovou aplikaci pro operační systém UNIX nebo Windows. Jazykové zdroje systému TRUST budou aktualizovány a v systému nadále využity. Dosud využívaná jazyková ontologie (taxonomie) bude převedena do standardního mezinárodního desetinného třídění (MDT), celosvětově používaného v knihovních systémech. Systém bude obohacen o dva další jazyky: angličtinu - bude dodána partnerem projektu TRUST a o češtinu - bude doplněna v rámci projektu M-CAST.

Metodologie

Agregační systém M-CAST bude ústředním prvkem agregačního portálu M-CAP a bude vyvinut podle zásad metodologie návrhu systémů pro správu obsahu založených na znalostech (Knowledge-based Content Management Application Design Methodology). Tuto metodologii vytvořila firma Infovide, S.A. v rámci jiného projektu financovaného z Pátého rámcového programu EU - ICONS - inteligentní systém pro správu obsahu (Intelligent Content Management System, IST-2001-32429, (<http://www.icons.rodan.pl>)).

Implementace

Portály budou nasazeny a testovány ve dvou veřejných knihovnách: v Polské internetové knihovně (<http://www.pbi.edu.pl>), jejímž provozovatelem je Kopernikova knihovna v Toruni a v Národní knihovně České republiky v Praze (<http://www.nkp.cz>). Cílem je zpřístupnit vícejazyčné digitální informační zdroje pro zodpovídání dotazů v přirozeném jazyce. Výsledkem hledání budou fragmenty textu, které obsahují odpovědi na dotaz.

Aplikace

Výsledkem projektu bude komerční produkt pro management vícejazyčných znalostí. Očekává se komerční využití výsledků v knihovnách (vyhledávání informací, předmětové katalogy), správě sbírek (akvizice a agregace dat, statistiky výpůjček, vyřazování z fondu), bibliografických databázích, informačních službách (selektivní šíření informací, personalizace, objevování znalostí) a v sémantických datových sítích.

Projekt programu eContent č. EDC 22249 M-CAST



Čeští spoluřešitelé:

Vysoká škola ekonomická v Praze
Nám. W. Churchilla 4
130 67 Praha 3
tel.: 224 095 407
fax.: 224 095 400
Národní knihovna České republiky
Klementinum 190
110 01 Praha 1
tel.: 221 663 111
fax.: 222 220 370

Hlavní řešitel:

Infovide S.A., Varšava, Polsko

Kontaktní osoba:

Vilém Sklenák
vedoucí katedry informačního a znalostního inženýrství
Vysoká škola ekonomická v Praze
tel.: 224 095 407
sklenak@vse.cz

Ostatní spoluřešitelé:

- Polská internetová knihovna, Toruń, Polsko
- TIP sp. z o.o., Katowice, Polsko
- Synapse Développement SARL, Toulouse, Francie
- Priberam Informática Lda., Lisabon, Portugalsko
- Expert System S.p.A., Modena, Itálie

Projekt M-CAST je financován Evropskou unií v rámci víceletého komunitárního programu na podporu rozvoje a užívání evropského digitálního obsahu v globálních sítích a současně překonávání jazykové různorodosti v informační společnosti (2001-2005).

