



O objectivo do projecto é desenvolver uma infra-estrutura multilíngue que permita aos produtores de conteúdos pesquisar, consultar e integrar recursos de vastas colecções multilíngues de textos (e multimédia), tais como bibliotecas digitais, recursos de editoras, agências de notícias e bases de dados de publicações científicas.

serão utilizados e actualizados. A ontologia (taxonomia) utilizada no TRUST será convertida para ser compatível com a Classificação Decimal Universal (CDU), o *standard* utilizado nos sistemas de catalogação das bibliotecas em todo o mundo. Dois outros módulos de língua serão adicionados: o inglês – desenvolvido por um dos parceiros do TRUST – e o checo – que será desenvolvido durante o projecto M-CAST.

Projecto eContent n.º 22 249

Sistema

O Multilingual Content Aggregation System (M-CAST) permitirá o desenvolvimento de Bibliotecas Digitais agregando dados digitais disponíveis em vários formatos e locais. O sistema vai ser testado por duas bibliotecas, para as quais serão criados Multimedia Content Aggregation Portals (M-CAP), utilizando os seus portais e as infra-estruturas existentes. Os portais permitirão encontrar respostas a perguntas feitas em língua natural em grandes colecções digitais de dados multilíngues. A interface de apresentação dos portais terá capacidades multimédia, permitindo a visualização de digitalizações de documentos antigos, documentos legais, partituras musicais, imagens e vídeos, apesar de apenas as suas descrições textuais serem indexadas.

Metodologia

O sistema de agregação M-CAST será o elemento central do portal de agregação M-CAP e será desenvolvido seguindo a metodologia de desenho de aplicações de gestão de conteúdos preparada pela Infovide S.A. num outro projecto do 5.º Programa Quadro – o ICONS – Intelligent Content Management System – IST-2001-32429 (<http://www.icons.rodan.pl>).

Implementação

Os portais serão implementados e testados em duas bibliotecas públicas: a Biblioteca Polaca na Internet (<http://www.pbi.edu.pl>) e a Biblioteca Nacional da República Checa (<http://www.nkp.cz>), como forma de disponibilizar *on-line* os seus recursos digitais para permitir encontrar respostas em colecções multilíngues digitais a perguntas em língua natural.

Tecnologias Linguísticas

O M-CAST basear-se-á nos resultados do projecto TRUST – Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (IST-1999-56416) – co-financiado pelo 5.º Programa Quadro da União Europeia (<http://www.trustsemantics.com>). O motor de pesquisa TRUST, desenvolvido para 4 línguas (francês, italiano, polaco e português), será convertido de um programa isolado para PC numa aplicação para servidores. Os recursos linguísticos TRUST

Aplicação

No fim do projecto estará disponível para comercialização um produto para gestão de conhecimento multilíngue. Está prevista a exploração comercial dos resultados em bibliotecas (recuperação de informação, classificação e catalogação), gestão de colecções (aquisição e agregação de dados, estatísticas de circulação e filtragem), bases de dados bibliográficas, serviços de informação (disseminação selectiva de informação, personalização, descoberta de conhecimento) e redes semânticas de dados.

Parceiro português:

Priberam Informática Lda.
Av. Defensores de Chaves, 32 - 3.º Esq.
1000-119 Lisboa
Tel: 21 781 72 60
Fax: 21 781 72 79
www.priberam.pt

Coordenador do projecto:

• Infovide S.A., Varsóvia, Polónia

Outros parceiros:

- Polska Biblioteka Internetowa, Polónia
- TiP sp. z o.o., Katowice, Polónia
- Expert System S.p.A., Modena, Itália
- Synapse Développement SARL, Toulouse, França
- Národní knihovna České republiky, Praga, República Checa
- Vysoká škola ekonomická v Praze, Praga, República Checa

Contacto:

Carlos Amaral
Tel: 21 781 72 60
carlos@priberam.pt

O projecto M-CAST é desenvolvido com o apoio financeiro da Comunidade Europeia no âmbito do programa comunitário multianual para estimular o desenvolvimento e utilização de conteúdos digitais europeus nas redes globais e para promover a diversidade linguística na sociedade de informação (2001-2005).