



**C**e projet a pour objet de développer une infrastructure multilingue autorisant des producteurs de contenus à accéder, à rechercher et à extraire l'information de grands corpus de textes multilingues, provenant du Web, de maisons d'éditions, d'agences de presse ou de bases de données scientifiques.

monoposte sous Windows en application serveur. Les ressources linguistiques de TRUST seront utilisées et mises à jour. L'ontologie linguistique (taxonomie) utilisée par le moteur sera convertie afin d'être compatible avec la Classification Décimale Universelle (CDU) utilisée par de nombreuses bibliothèques dans le monde. Deux autres langues seront prises en compte: l'anglais - développé par un des partenaires de TRUST, et le tchèque - développé durant le projet M-CAST.

Projet e-Contenu n° EDC 22249 M-CAST

### Système

Le système M-CAST prendra en compte le développement des bibliothèques numériques en rassemblant les données numériques disponibles dans différents formats et différents lieux. Le système sera testé sur deux bibliothèques, pour lesquelles seront créés des Portails d'unification de Contenu Multimedia (M-CAPs), en utilisant leurs portails actuels et leurs infrastructures. Les portails permettront de fournir des réponses aux questions en langage naturel dans de grands corpus de données multilingues. L'interface de présentation des portails sera multimedia afin d'autoriser la reproduction de copies numérisées de vieux documents, les documents administratifs, les partitions musicales, les vidéos, les images, etc., même si le texte seul sera indexé.

### Méthodologie

Le système M-CAST sera l'élément central du portail M-CAP et sera développé en conformité avec la méthodologie "Knowledge-based Content Management Application Design Methodology" conçue par Infovide S.A. dans un autre projet du 5e Programme-Cadre de la Communauté Européenne - ICONS - Intelligent Content Management System - IST-2001-32429 (<http://www.icons.rodan.pl>).

### Implémentation

Les portails seront déployés et testés dans deux bibliothèques publiques: La Bibliothèque Internet Polonaise (<http://www.pbi.edu.pl>) et la Bibliothèque Nationale Tchèque (<http://www.nkp.cz>) afin de permettre, sur l'ensemble de leurs ressources numériques disponibles en ligne, de poser des questions en langage naturel et d'obtenir des réponses issues des corpus multilingues.

### Technologies linguistiques

M-CAST sera basé sur les résultats de TRUST - Moteur de Recherche Sémantique et Cognitive pour retrouver de l'information par des méthodes sémantiques (Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (IST-1999-56416) projet financé par le 5e Programme-Cadre de la Communauté Européenne pour la Recherche et le Développement Technologique (<http://www.trustsemantics.com>). Le moteur de recherche TRUST, développé pour 4 langues (Français, Italien, Polonais et Portugais) sera converti d'un programme

### Réalisation

À la fin du projet un produit commercialisable pour la gestion de connaissances multilingues sera disponible. L'exploitation commerciale des résultats dans les bibliothèques (recherche d'information, classification de catalogues), la gestion de contenu (acquisition et unification de contenus, statistiques de diffusion, débroussaillage), les bases de données bibliographiques, les informations de service dissémination sélective de l'information, personnalisation, découverte du savoir) et dans les réseaux de données sémantiques est prévu.

#### Partenaire français:

Synapse Développement SARL  
33 rue Maynard  
31000 Toulouse  
tél.: (561) 63 69 90  
fax.: (561) 630 287

#### Coordinateur du projet:

• Infovide S.A., Varsovie, Pologne

#### Autres partenaires:

- Polska Biblioteka Internetowa, Pologne
- TIP sp. z o.o., Katowice, Pologne
- Expert System S.p.A., Modène, Italie
- Priberam Informática Lda., Lisbonne, Portugal
- Národní knihovna České republiky, Praha, République tchèque
- Vysoká škola ekonomická v Praze, Praha, République tchèque

#### Personne à contacter:

Dominique Grégoire  
Gérant  
tél.: (561) 621 570  
dlaurent@synapse-fr.com

Le projet M-CAST est développé avec le soutien financier de la Communauté Européenne dans le cadre du programme communautaire pluriannuel visant à encourager le développement et l'utilisation du contenu numérique européen sur les réseaux mondiaux ainsi qu'à promouvoir la diversité linguistique dans la société de l'information (2001-2005).

