



Solutions in language processing

## Documation 2004 : lancement de Lingway KM 3.0

Paris, le 8 mars 2004 - Lingway, éditeur de logiciels documentaires fondés sur les technologies linguistiques, annonce à l'occasion de Documation 2004 (16-17 mars 2004 au CNIT Paris La Défense) la nouvelle version 3.0 de Lingway KM, moteur sémantique multilingue pour la recherche, la catégorisation et l'aide à la lecture, dédié à la gestion de connaissances (Knowledge Management) et aux applications de fouille textuelle (Text Mining) et de veille.

**Lingway KM version 3.0** qui sera présenté en avant-première sur le **stand Lingway (C5)**, s'enrichit de 5 modules et fonctionnalités nouveaux :

- 1- Un module de catégorisation sémantique permettant de structurer automatiquement en XML l'information contenue dans des documents non structurés à partir de règles d'extraction. Cette structuration est faite à un niveau suffisamment fin pour qu'on puisse ensuite d'extraire uniquement les parties intéressantes du point de vue de l'indexation, du résumé et de la catégorisation.
- 2- Un module d'aide à la lecture articulé autour de 4 fonctionnalités complémentaires : le résumé automatique, la navigation entre la synthèse et le document source, la colorisation des éléments importants suite à la requête pour immédiatement attirer l'attention du lecteur, la cartographie et le clustering pour avoir une vue synthétique sous forme de graphe du contenu d'un ensemble de documents.
- 3- Un module de correction ortho-phonétique qui permet de proposer des corrections en cas de mot ou de nom inconnu pour faciliter l'interrogation.
- 4- Une interface basée sur SOAP/XML pour faciliter l'intégration entre applications.
- 5- Le support de 3 nouvelles langues, en plus de l'anglais et le français, soit l'allemand, l'espagnol et le néerlandais.

*« Le volume sans cesse croissant d'information disponible (notamment sur Internet) nécessite d'avoir recours à des outils de plus en plus performants. Cette nouvelle version de Lingway KM va permettre de retrouver, organiser et lire plus facilement l'information interne et externe à l'entreprise de manière multilingue, et ceci grâce à une approche sémantique du document. Ces trois besoins sont importants et doivent être traités simultanément. Il ne sert à rien de trouver plus d'information, si l'on ne sait pas l'organiser et si l'on n'a pas le temps de l'analyser »* indique Bernard Normier, Président de Lingway.

### Caractéristiques techniques de Lingway KM version 3 :

- Windows NT 2000, Solaris 2.6-8 et Linux
- Configuration conseillée : 1 GHz, 512 Mo de RAM, 60 Mo d'espace disque
- Format supporté en natif : XML, HTML et texte
- Web services SOAP/XML
- API C++ et Java
- Connecteurs (en cours de développement) pour Verity, Hummingbird, Oracle ...
- SGBR et indexation en open-source (Lucene, MySql)

Lingway KM est vendu en mode licence à partir de 30 000 euros HT pour 1 serveur et en mode ASP à partir de 1 000 euros HT par mois en fonction du volume de page indexé.

## A propos de Lingway KM

**Lingway KM** est une suite d'outils dédiés à la gestion de connaissances (Knowledge Management) et aux applications de fouille textuelle (Text Mining) et de veille.

Lingway KM comporte 3 modules principaux :

\* **Un moteur de recherche sémantique** permet de combiner l'interrogation des bases de texte intégral en langage naturel et par taxonomie (plan de classement) en plusieurs langues.

\* **Un moteur d'indexation et la catégorisation** s'appuyant sur un module de structuration XML permettant le marquage du texte à un niveau sémantique.

\* **Un moteur d'aide à la lecture** (résumé automatique, colorisation du texte et cartographie) permettant d'appréhender le contenu d'un document ou d'un ensemble de documents.

Basé sur une ontologie de 150 000 concepts mis en correspondances avec 5 langues (allemand, anglais, espagnol, français et néerlandais), Lingway KM permet de réaliser des recherches « cross-language »).

De plus, des applications sectorielles développées autour de Lingway KM sont disponibles pour le domaine médical et celui de la propriété intellectuelle :

**Lingway Patent Suite** : permet un accès multilingue à des bases de brevets, utilisant simultanément la nomenclature internationale des brevets et l'accès full-text. Des outils additionnels d'aide à la lecture permettent également une synthèse orientée des textes.

**Lingway Medical Dictionary Encoder** : permet un codage automatique des phrases décrivant les effets secondaires des médicaments, les antécédents et les pathologies associés dans la nomenclature internationale MedDRA (Medical Dictionary for Drug Regulatory Activities).