



Solutions in language processing

LINGWAY Terminology Builder for Systran : nouvelle solution pour le développement d'applications multilingues

Paris, le 31 mai 2005 - LINGWAY, éditeur de logiciels de « Text Mining » annonce à l'occasion de iExpo 2005 (1-2 juin 2005 au CNIT Paris La Défense), le lancement de **Lingway Terminology Builder for Systran**, une solution permettant le développement d'applications multilingues par la construction rapide de terminologies pour les logiciels de traduction automatique de SYSTRAN.

Cette solution, développée à partir du moteur de recherche sémantique multilingue, de catégorisation, d'extraction d'informations et d'aide à la lecture Lingway KM, réalise l'extraction automatique de terminologie bilingue, puis son importation dans le système de traduction automatique Systran.

Concrètement cette plateforme permet de construire très rapidement des terminologies en plusieurs langues qui seront utilisées pour la réalisation d'applications multilingues (Portails, Intranet, sites Web, applications de Knowledge Management, catalogues électroniques, ...), mais aussi pour la recherche et le « text mining » multilingues.

« Grâce à notre technologie, les temps de construction des terminologies bilingues, véritable goulot d'étranglement pour la réalisation d'applications multilingues vont être divisés par un facteur de 3 à 5 », indique Bernard Normier, Président Directeur Général de LINGWAY.

Une méthodologie en 4 étapes :

- 1- Acquérir deux corpus de documents traitant du même domaine technique, l'un en langue source et l'autre en langue cible, par exemple en aspirant des sites sur le Web qui traitent du sujet : textes de conférences, articles de presse spécialisés, brevets, sites commerciaux décrivant des produits, etc.
- 2- Extraire automatiquement les termes du corpus en langue source, sur des critères linguistiques et statistiques. On utilise pour cela l'extracteur terminologique de Lingway KM, proche de celui utilisé pour l'indexation automatique des textes, mais adapté à cette tâche particulière.
- 3- Utiliser le moteur de recherche « cross-language » de Lingway KM, qui va rechercher les différentes combinaisons dans la base de données constituée en langue cible. On choisit ensuite la traduction la plus fréquemment trouvée sur le corpus.
- 4- Importer cette terminologie bilingue, validée sur ce corpus, dans le système de traduction automatique SYSTRAN.

Avec cette approche, seuls les coûts de constitution des corpus servant de base à l'extraction, qui soit existent déjà dans l'entreprise, soit sont disponibles, gratuitement, sur le Web, et de révision manuelle éventuelle sont nécessaires. De plus le logiciel permet de consulter facilement les textes sources et cibles, et donc de resituer rapidement le terme dans son contexte, ce qui peut être utile dans le cas de termes peu fréquents.

« Faute de dictionnaires électroniques adaptés, la qualité des systèmes d'information multilingues mis en œuvre reste médiocre. Un travail long et fastidieux réalisé manuellement par des traducteurs spécialisés et dont le coût se situe entre 2 et 5 euros par terme et qui peut donc aujourd'hui être automatisé », commente Bernard Normier.