



# LINGWAY Knowledge Extractor

## Structuration de documents et extraction de connaissances

Avec la généralisation d'Internet, la masse d'information disponible est en croissance exponentielle – plusieurs dizaines de milliards de pages Web et un doublement tous les 6 mois. Or plus de 90% de cette information électronique est non structurée. C'est l'enjeu du Web Sémantique, que de pouvoir tirer toute la valeur ajoutée contenue dans ces gisements de connaissances. Pour répondre à cette problématique et faciliter l'accès à la connaissance, LINGWAY a développé LINGWAY Knowledge Extractor (LKE), un composant technologique de structuration de documents et d'homogénéisation des connaissances, basée sur des technologies sémantiques.

### Les principales fonctionnalités de LINGWAY Knowledge Extractor

- Structuration automatisée des documents textuels au format XML
- Génération de métadonnées et possibilité de mise au format du Web Sémantique
- Identification des éléments clés d'un document (entités nommées, passages importants ...)
- Extraction automatisée des descripteurs ou sélection à partir de listes d'autorité ou de thésaurus
- Export des connaissances extraites vers un ERP, un ECM ou une application métier
- Aide à la construction de référentiels documentaires multilingues

Document au format PDF,  
Word, Html, etc.

Conversion au  
format XML

Base de connaissances  
multilingues



Structuration



Exploitation des  
connaissances



Extraction de  
métadonnées

Formulaires, Fiches de  
synthèse, Alimentation  
d'applications métiers...

**A propos de l'offre LINGWAY :** En complément de LKE, LINGWAY commercialise LINGWAY KM (LKM). Solution d'analyse et de recherche sémantique multilingue, LKM propose un ensemble de fonctionnalités d'indexation, de recherche et de valorisation de l'information. LKM s'appuie sur une ontologie de 150 000 concepts personnalisable et des dictionnaires spécialisés pour de nombreux secteurs d'activité. LINGWAY propose également plusieurs solutions métier adaptées aux domaines de la propriété intellectuelle, du commerce électronique, du médical, du juridique et fiscal et des Ressources Humaines.

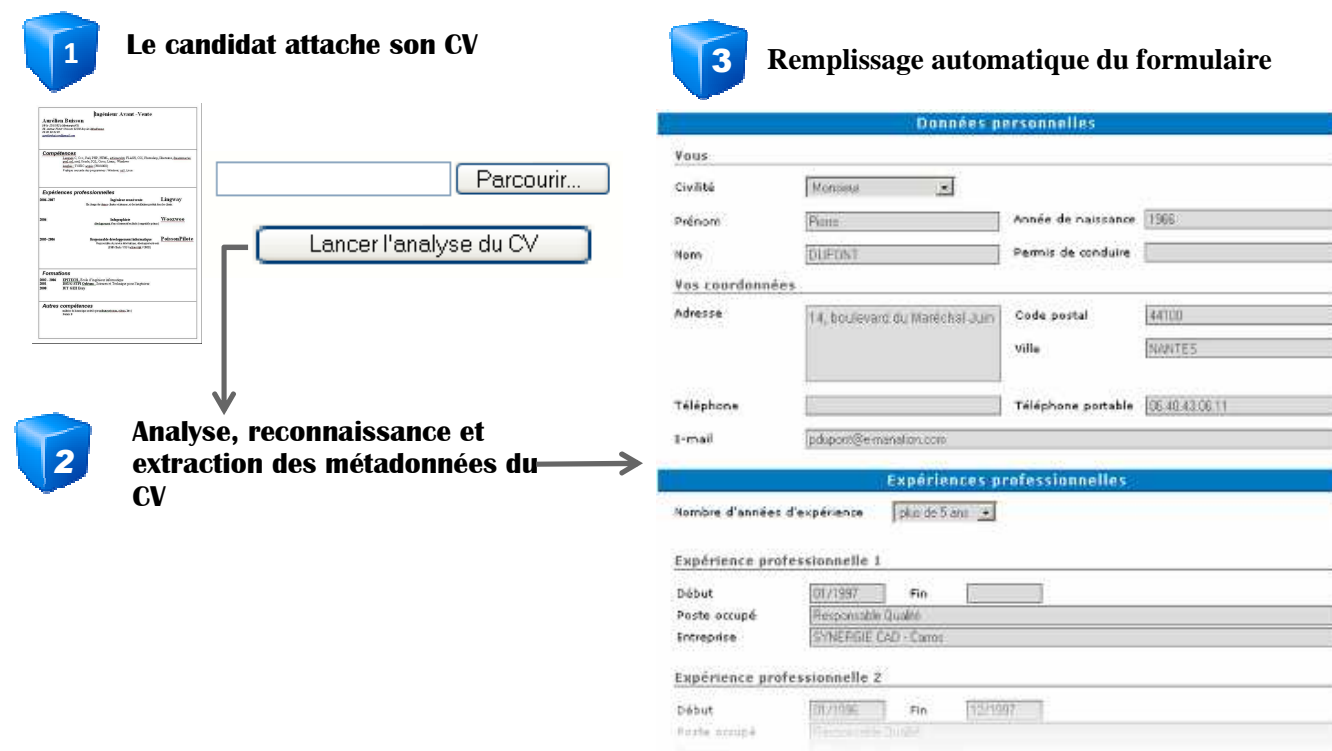




## Trois exemples d'applications

- Reprendre l'existant d'Intranet de première génération contenant divers formats de documents (HTML, PDF...) afin de structurer ces informations pour les importer dans un CMS (Content Management Systems) s'appuyant sur un format XML. Cette reprise de l'existant automatisée ou supervisée permet d'enrichir un logiciel de gestion de contenus. **Bénéfice** : *capter, réutiliser et valoriser les connaissances existantes.*
- Identifier et comprendre les éléments importants contenus dans certains types de documents (contrats, actes légaux ou juridiques ...) au format PDF texte pour les extraire dans un format normalisé (ex : la date qui peut être écrite de différentes manières suivant les documents, les lieux, les noms des contractants, le type de biens ...etc.). Les métadonnées ainsi extraites permettent de formaliser les connaissances clés qui vont venir alimenter les applications métiers. **Bénéfice** : *améliorer la qualité des référentiels documentaires et exploiter les connaissances.*
- Analyser et structurer les informations contenues dans un CV (nom, prénom, date de naissance expériences, formations, langues, noms des entreprises ...) afin de pouvoir pré-remplir automatiquement les questionnaires de dépôt de candidature. Cette technique permet également l'anonymisation des champs potentiellement discriminatoires d'un CV. **Bénéfice** : *un gain de temps de saisie pour le candidat et un format homogène de la CVthèque pour les entreprises pour faciliter la recherche de candidats.*

## Exemple d'extraction des informations contenues dans un CV



**Caractéristiques techniques** : LKE est disponible en Web services (SOAP/XML) sous Linux et Windows pour une configuration conseillée multi-cœur 2 GHz et 2 Go de RAM. Principaux formats supportés : TXT, XML, HTML, PDF, Formats Office (.doc, .xls et .ppt)... Langues disponibles : Français, Anglais et Espagnol.

Informations non contractuelles susceptibles de modifications sans préavis. Tous les noms de sociétés et/ou de produits appartiennent à leurs propriétaires respectifs. Tous droits réservés Copyright Lingway Mars 2009

