

Communiqué de presse

Publication Électronique en Format Original

Une Technologie Innovante de Publication Électronique

Principes de fonctionnement :

- **Un traitement automatique d'analyse d'images** et de reconnaissance automatique de caractères (OCR) effectue la lecture des pages originales et reconstitue le contenu textuel.
- Cette lecture est précédée par un réglage spécifique à la mise en page du fond traité. Ce réglage, effectué une fois, est utilisé pour tous les documents ayant la même mise en page.
- Les éléments de la **structure** (titres, chapitres, entêtes etc...) sont intégrés au corps du texte au sein d'un document balisé (**XML**). Ces éléments sont reconnus automatiquement d'après leurs attributs typographiques (par exemple : la taille de la police de caractères). Le balisage reflète le découpage de l'ouvrage en **unités documentaires**.
- **La richesse du document XML** qui résulte du traitement automatique, permet d'effectuer des recherches avancées y compris des requêtes multi-critères. **Ce XML est adaptable aux besoins**, cela est particulièrement utile dans un contexte de sources hétérogènes.
- Le contenu textuel est **entièrement recherchable** : chaque mot peut être identifié et marqué.
- Les unités documentaires étant parfaitement délimitées, la recherche prend en compte la vraie structure logique du contenu.
- Le contenu affiché étant toujours une copie exacte de l'original, la lisibilité est complète, notamment pour **les formules et les éléments graphiques**.
- La recherche - multicritères ou en texte intégral - restitue au lecteur les documents originaux dans lesquels les mots correspondant à sa recherche sont affichés en **surbrillance**.
- Le support peut être un **CD/DVD ou/et Internet**. L'un ne conditionne pas l'autre. Il est important de noter qu'une partie importante du travail - la préparation des données - est commune.
- La technologie est **indépendante du moteur** de recherche. Elle s'intègre facilement dans une infrastructure existante.

Précisions techniques - formats standards utilisés:

- **Formats d'entrée** : TIFF après numérisation (scan) ou bien le format PDF, s'il est déjà disponible sans passer par la numérisation. De nombreux autres formats d'images sont possibles.
- **Formats d'indexation** : XML ou bien une répartition en deux : le texte intégral d'un côté et des éléments de la structure indexés parallèlement de l'autre.
- **Formats d'affichage** : TIFF sur CD/DVD ou le format PDF/GIF/JPG sur Internet. D'autres formats d'images sont également possibles.

Utiliser le PDF en entrée représente un intérêt tout particulier, car les éditeurs ont souvent une version PDF de leurs ouvrages depuis plusieurs années. Ainsi, il est possible de réaliser une publication électronique performante et conviviale sans changer ni alourdir les procédures éditoriales existantes.

Références

La société ISAKO est spécialisée dans l'informatique documentaire et l'édition électronique. ISAKO a réalisé de nombreux services documentaires sur support CD/DVD et pour des sites Internet. Notre équipe a une riche expérience dans l'édition électronique : logiciels de rédaction, retro-conversion, systèmes éditoriaux, moteurs de recherche et logiciels documentaires.

Le 1er tome du **Grand Dictionnaire Universel de Pierre Larousse** (1866), édité sur DVD par les éditions Redon et commercialisé par les éditions Le Robert, est accessible via notre site au www.isako.com, puis sur la page "**Technologies**", cliquez sur le lien "**démonstration**".

Contact :

Shalev VAYNESS Directeur du développement ISAKO.
tél : 01 5830 6855 e-mail : shalev@isako.com