

Informatique & Bible, asbl - Belgique
Rue de Maredsous, 11 B5537 Denée - Belgique
Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69
cib@cibmaredsous.be



Interface n° e-86 Juin 2002

KnowHowSphere.net

C'est un site Web consacré à des recherches sérieuses et enrichies sur d'importantes bases de données textuelles ou documentaires à haute valeur ajoutée.

[Pourquoi Knowhowsphere?](#)

I&B applique depuis 25 ans son savoir-faire à l'“informatisation” de Corpus littéraires. Cela va de la saisie des documents pour les transformer en écriture électronique, jusqu'à la création d'outils de recherche sur tous les éléments significatifs ainsi enregistrés, en passant, souvent, par la photocomposition électronique programmée pour produire l'image des pages d'un livre imprimé.

Dans les premières années de ces travaux, la démarche fut souvent pionnière, tout était occasion de découverte et de “premières” (premiers caractères hébreux électronique, première photocomposition automatisée en colonnes avec notes de bas de pages, premier livre religieux présenté simultanément en imprimé et dans le Minitel, premiers textes bibliques français sur disquette, etc...). Avec les micro-ordinateurs et la banalisation d'un travail de saisie de textes (traitements de textes) associés aux facilités de systèmes de copie (disquettes, CD-ROM, CD-R), le CIB s'est trouvé quelque peu submergé avec des procédures héritées de l'informatique lourde de ses débuts (datant de l'époque des “main-frames” de type IBM-360!).

L'extension de l'Internet a créé une nouvelle situation. La multiplication des applications a tiré l'informatique et, notamment les travaux sur les textes, vers une qualité moindre qui visait l'usage domestique du plus grand nombre d'utilisateurs. De nouveaux standards d'échange de données électroniques devenaient nécessaires (mais faisaient aussi partie des stratégies des grands constructeurs pour s'assurer des marchés captifs).

Ces travaux banalisés, réalisés souvent par des gens sans compétence ni formation spécifique dans le domaine littéraire, aboutissent à des accumulations de textes dont la saisie à la base est parfois honteuse et dont le caractère souvent “piraté” ne peut inspirer confiance.

Face à cette situation, le CIB n'a pas baissé les bras. Il a maintenu ses exigences de qualité pour porter son savoir-faire (know-how) au sein de cette grande mêlée où le bon et moins bon se côtoient sans qu'il soit toujours facile de juger critiquement de l'information disponible.

D'où les décisions prises en Mars 2000 de faire un nouveau bond technologique qui serait aussi une occasion de mettre à jour et de rationaliser la chaîne de travail sur les données textuelles et documentaires au CIB.

Pour cela, on a adopté l'un des Standards qui risquent de s'imposer pour les années à venir dans la transmission de données électroniques: l'XML (Extended Markup Language) qui remplace les autres modes de balisage pratiqués, selon la nature des données et selon les protocoles propres au CIB depuis le début de ses travaux.

C'était l'occasion également de commencer à travailler franchement avec l'UNICODE (en passe de remplacer le code ASCII) pour représenter les caractères.

Une occasion également de résoudre, au niveau de la distribution de produits

électroniques, le problème de la compatibilité avec des plates-formes diverses (Windows de différentes versions, Mac, Linux et autres). En effet, la réception multiplateforme est largement gérée à travers l'Internet et les browsers (même s'il reste des variantes dans la qualité ou l'intégrité de la transmission de l'image-écran originelle des données).

Qu'offre knowhowsphere?

Knowhowsphere n'est pas une offre supplémentaire de "e-book" (livre électronique) en format PDF. Ce genre d'offre peut avoir son utilité en permettant l'affichage d'un document, mais il ne permet aucune recherche sérieuse sur les contenus du texte affiché.

Knowhowsphere n'a pas comme objectif d'accumuler le plus de textes possibles sans en garantir toujours la fiabilité quant aux sources, aux méthodes d'enregistrement. Knowhowsphere voudrait apporter des textes et une recherche à haute valeur ajoutée sur des corpus littéraires ou des ensembles documentaires contrôlés et enrichis.

Parmi les originalités du site, on notera les suivantes:

1. L'affichage permanent et dynamique d'une Table des Matières détaillée et en expansion jusqu'au niveau de la plus petite unité documentaire du corpus traité.
2. Une fenêtre dans laquelle le document en cours est visible à tout moment dans une graphie agréable à la lecture.
3. La totalité du "cherchable" pour un corpus documentaire ou littéraire donné est visible à tout moment. Il n'y a pas de mots-creux ou stop-words sur lesquels on ne pourrait pas effectuer de recherche.
4. Toutes les recherches booléennes structurées sont possibles et combinables (AND, OR, NOT).
5. Les formes des Noms Propres sont regroupées sous un lemme et les Noms sont désambigués.
6. Dates et Nombres ou autres données particulières reçoivent un traitement spécifique et logique, hors de la série alphabétique.
7. Il y a possibilité de donner des bornes pour limiter la recherche à une partie du Corpus présenté.
8. À tout moment, plusieurs techniques permettent de saisir un passage ou un document sur son propre ordinateur, notamment en l'envoyant par e-mail, en format HTML, sur sa propre machine.
9. Suivant les décisions du détenteur des droits intellectuels sur les données, celles-ci peuvent voir leur accès protégé à différents degrés, voir demander éventuellement une rétribution pour l'accès.

Cela explique les principaux éléments de la présentation des Bases de données:

- Le texte ou les documents sont "référénciés" jusqu'à un niveau qui constitue une unité logique pour une Base de donnée déterminée (par exemple, le "verset" pour une Bible) – cette référence permet d'individualiser des phénomènes et de faire des liens avec d'éventuelles sources imprimées;
- On présente en permanence et de façon "synoptique": une Table des matières en arborescence, les textes dans lesquels se trouvent l'argument de recherche (souligné en surbrillance), l'Index complet des éléments recherchables, soit par ordre alphabétique des formes et/ou des lemmes, soit dans l'ordre logique de certaines données non-alphabétiques comme les Dates ou les Nombres.

Autour de cet ensemble de base, on trouve les "boutons" correspondant aux différentes fonctions de recherche proposées et les autres outils (aide, impressions, informations générales sur le site ou la Base de données) qui permettent l'usage simple d'un ensemble très complexe.

Nous souhaitons que chaque utilisateur fasse un instant l'effort de se demander ce que

cela pourrait impliquer comme réalisation cachée quand il vient d'effectuer une recherche multicritère comportant deux ensembles complexes: d'une part, une série de mots, et l'autre, une série de dates – par exemple; et qu'ayant trouvé un résultat, il s'est envoyé immédiatement par la fonction "e-mail" le résultat de sa recherche (plusieurs pages de textes parfois) sur son ordinateur en format HTML !! (voir à ce sujet notre Éditorial).

Avec ce nouvel outil et la chaîne de travail modifiée et améliorée qu'il entraîne, le CIB est prêt pour affronter quelques nouveaux chantiers importants qui lui sont confiés.

Le tout a pris le nom de knowhowsphere par allusion, d'une part, à la vision du grand jésuite, le P. Teilhard de Chardin, d'une socialisation progressive de la connaissance et de l'esprit humain dans un règne qui nous ferait passer de la biosphère à la "noosphère" (la sphère ou le règne de l'esprit); et, d'autre part, pour réserver un nom de domaine libre dans l'Internet, une allusion au savoir-faire (know-how) accumulé au long des années par l'équipe du CIB et qui lui permet, aujourd'hui, sous la direction prévoyante de Yolande Juste, de franchir une étape importante dans l'usage de l'écriture électronique appliquée aux textes et documents.

Plusieurs Bases de données, à accès plus ou moins public (les couleurs vert, orange ou rouge indiquent immédiatement ce statut d'une Base et s'il faut ou non des procédures spéciales pour y accéder), seront disponibles dans les mois qui viennent sur knowhowsphere.net.

La première Base de données importante présentée dans ce cadre a été celle des Oeuvres du P. Pio Bruno Lanteri, fondateur des Oblats de la Vierge Marie. D'autres vont suivre, à mesure des rythmes possibles de mise en forme et des travaux qui arrivent au seuil de publication électronique.

Fr. R.F. POSWICK, osb , Directeur

Y. JUSTE, Responsable du projet et de la réalisation Knowhowsphere

