

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-101 Décembre 2005

Bibliothèque virtuelle: quantité ou qualité?

Le sujet est dans l'air, pas un jour ne se passe sans une annonce ambitieuse dans Internet.

Tantôt, Google et sa Bibliothèque virtuelle,

-tantôt Amazon qui va étendre son offre de librairie en ligne par la vente de tout ou partie des livres vendus sous forme numérique

-aujourd'hui Microsoft qui annonce la "numérisation de la British Library" et de 100.000 ouvrages qui seront disponibles dans Microsoft Book Search!

On peut comprendre que l'initiative de Google provoque des réactions mais on sent que celles-ci sont purement commerciales. D'ailleurs la course entre deux ou trois 'grands' acteurs actuels se passe à tous les niveaux: moteurs de recherche, barre de recherche personnelle, Google Earth (pour servir les terroristes?), Google Maps,...

Si l'un des buts est de faire parler de soi dans Internet, l'objectif est atteint, le nombre d'articles va croissant.

Le projet est ambitieux: 5 millions de livres! On peut lire sur le site de Google Print "La société Google s'est fixé pour mission d'organiser l'information mondiale"! Les premières réactions à l'annonce de Google furent d'ailleurs étonnantes puisqu'on a même vu aux États-Unis des associations d'auteurs se sentant lésés dans leurs droits attaquer Google en procès et plus tard aussi des Associations d'Éditeurs.. Lorsqu'on connaît par ailleurs l'esprit libéral des américains, cela surprend.

Google, dans un premier temps a donc été contraint au cours de l'été d'arrêter son expérience en fermant l'accès à 'Google Print'. Mais à l'occasion de la Foire du Livre de Frankfort, Google a annoncé un nouvel accord et une présence simultanée dans 8 pays européens. Annonçant également des accords avec de grands éditeurs et des adaptations aux accès qui résoudraient les problèmes de droits d'auteur: pour les livres soumis aux droits d'auteur, trois pages maximum sont proposées à l'utilisateur; des liens permettent d'acquérir l'ouvrage; l'utilisateur doit avoir un compte (gratuit) chez Google pour visualiser certaines pages (et être ainsi dans la Base de Données de Google!)

À l'annonce du projet par Google fin 2004, la France et 5 autres pays européens avaient lancé un appel aux instances communautaires pour la "création d'une bibliothèque numérique européenne" et Viviane Reding, commissaire européenne à la Société de l'information a annoncé que des fonds seraient débloqués pour développer un moteur de recherche grand public et permettre la numérisation des archives des bibliothèques européennes.

Le rythme de la vieille Europe n'est malheureusement pas celui de l'Internet, et entre-temps l'Angleterre a fait cavalier seul en prenant un accord avec Microsoft. La British Library était l'une des deux seules Bibliothèques à n'avoir pas signé la motion appelant à la création d'une bibliothèque numérique européenne!

Mais en fait de quoi s'agit-il? Beaucoup d'articles se préoccupent des enjeux économiques mais peu du service réellement rendu à l'utilisateur.

Nous avons fait l'exercice suivant: une simple recherche dans Google Print .fr (<http://print.google.fr>) du terme "Maredsous", la recherche ramène 4.370 références (quelle est

la base examinée? mystère. D'autant que dans les 8 pays européens la recherche donne le même résultat et qu'elle donne un résultat différent chaque jour!). Une liste des livres trouvés s'affiche alors dans le modèle habituel de Google. Et un clic sur une référence donne accès à la Table des Matières de l'ouvrage, ou à la page concernée, avec les restrictions mentionnées plus haut.

Une des références attire notre attention:

*Carnets conciliaires de l'évêque de Namur A.-M. Charue
de Decléck L. Soetens C - 2000
Page 134 - 10 G. Dayez (1903-1988), bénédictin **beige**, Abbé de Maredsous de 1950
it 1968, au temps du concile président de La Congrégation bénédictine **beige** de...*

Le contexte nous donne tout de suite à penser que les techniques de reconnaissance de caractères utilisées par Google le sont "à l'américaine", superficielles! Chacun sait que l'OCR peut produire des résultats intéressants mais que la qualité du résultat dépend de la source. Et que chaque résultat d'une reconnaissance automatique de caractères mérite une bonne relecture! Les américains ne le savent sans doute pas!

Mais le soupçon est là: quelle est la qualité de la recherche proposée par Google?

Une autre recherche sur "bénédictin beige" cette fois-ci, donne aujourd'hui 36 résultats ce qui montre qu'il ne s'agit pas d'un cas isolé mais bien de la procédure elle-même. La question est alors d'imaginer toutes les erreurs possibles de l'OCR pour avoir un résultat "exhaustif". Cela est bien sûr impossible. Essayez vous-même de chercher alors tout ce qui concerne la "beigique" et les "beiges" que nous sommes pour Google! Ceci n'est bien sûr qu'un exemple, à vous d'en chercher d'autres.

Feriez-vous confiance à Google Print sachant que le résultat ne dépend que de la qualité d'impression du livre scanné à toute allure sans relecture? A côté de combien de résultats pertinents passerez-vous si vous n'avez pas la curiosité d'imaginer que Google puisse faire des erreurs! Les éditeurs qui ont pris des accords avec Google Print se sont-ils au moins interrogés sur les procédures qui seront mises en œuvre pour mettre à disposition leurs ouvrages? Le lecteur ne se lassera-t-il pas très vite s'il ne peut limiter la recherche aux livres écrits dans la langue qu'il souhaite?

Nous qui, depuis 25 ans, passons d'innombrables heures à la relecture des textes saisis pour produire des textes de qualité ne pouvons que regretter de telles initiatives faisant appel à la crédulité des utilisateurs, à leur confiance aveugle dans 'Internet' et au pouvoir de l'informatique.

La spécialisation de Google dans un moteur de recherche rapide n'aurait-elle pas été avantageusement améliorée et perfectionnée lui permettant de rester le bon moteur de recherche (si pas le meilleur)? Quand on sait que seulement x% des pages sont prises en compte par les moteurs de recherche, les utilisateurs n'auraient-ils été mieux servis par un moteur de recherche qui améliore ses performances et la couverture du champs de recherche, plutôt que par une dispersion dans tout ce qui peut rapporter gros à travers Internet?

Nous proposons que quelqu'un crée un site avec les perles trouvées grâce à Google Print, et voici la première:

Pas de panic!
*Lettres européennes
de Annick Benoit, Guy Fontaine - 1992 - 1026 pages
Page 227 - ...La Bible en français **darts** une édition **panic** en 1535 It ... Quoi qu'il en
soin, cenne Bible...*

Google devrait fournir le dictionnaire français-français pour pouvoir utiliser sa Bibliothèque virtuelle!

Aucune recherche sérieuse ne peut se satisfaire de telles approximations.

La quantité ou la qualité? Notre choix est fait depuis longtemps, et le vôtre?

PS: Entre-temps Google a déjà annoncé un changement de nom pour ce projet qui est désormais "Google Book Search" pour "rendre clair le but de ce nouveau service car Google Print avait souvent été pris pour un service d'impression" !

