

BASES

Connaître et bien utiliser
les banques de données

S O M M A I R E

Actualités

- De Roger Summit à Tom Cruise, l'adoption des nouvelles technologies par les serveurs traditionnels, pp.1-6,
- Migration prochaine des utilisateurs de DataStar vers Proquest Dialog, p.9

A Lire

- Du bon usage de la bibliométrie pour l'évaluation des chercheurs, p.8
- L'usager numérique, p.10

Web invisible

- Gallica vs Google Books : le match !, p.11

Agenda

- Search 2011, p.9

De Roger Summit à Tom Cruise : l'adoption des nouvelles technologies par les serveurs traditionnels

Mary Ellen Bates

Nous remercions Mary Ellen Bates – l'auteure – et Marydee Ojala, l'éditrice de la revue américaine Online (www.onlinemag.net), de nous avoir autorisés à publier une traduction de l'article « From Roger Summit to Tom Cruise », paru dans le numéro 5 vol. 34 (sept.-oct. 2010) de la revue.

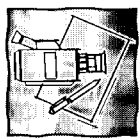
Il y a peu de temps, alors que j'étais un soir au volant de ma voiture sur une petite route du Wyoming, j'ai eu soudain très envie d'un café, de préférence dans un Starbucks. J'ai alors ouvert mon smartphone, téléchargé l'application Starbucks-Finder et obtenu une carte avec l'itinéraire me permettant de trouver un Starbucks encore ouvert à proximité. Cela ne m'a pris que 30 secondes au total. En fait, j'ai été très désagréablement surprise que le site de Starbucks lui-même n'offre pas une telle application. En effet, pour attirer les consommateurs potentiels, il n'y a rien de tel que de les aider à vous trouver facilement.

Tout ce qu'offre le site de Starbucks, c'est une page statique avec la localisation des différents établissements. Ce qui ne présente aucun intérêt quand je ne sais même pas de quelle ville je m'approche. Starbucks n'a manifestement pas encore appris que les pages statiques datent du siècle dernier.

A l'heure actuelle, les sociétés mettent en avant leurs pages Facebook, leurs comptes Twitter, leur présence sur LinkedIn et leurs chaînes YouTube avec autant d'énergie que leur site web ou un numéro de téléphone surtaxé.

Les consommateurs sont désormais encouragés à interagir avec la société plutôt que de simplement recevoir, avec passivité, de l'information. Qui aurait cru en 2004, quand Burger King lançait son site web «*Subservient Chicken*», que se développerait une véritable interactivité en ligne à peine quelques années plus tard ?

Tout cela me rappelle d'ailleurs quelque chose qui n'est pas sans lien avec l'interactivité, mais remonte à 1972 : la naissance de Lockheed Dialog, le «grand-père» de l'expérience d'interactivité en ligne que nous a apporté l'incalculable Roger Summit, qui le baptisa Dialog pour une raison bien particulière.



De Roger Summit à Tom Cruise *suite*

A cette époque, la plupart des interrogations d'une banque de données étaient réalisées en «batch», c'est-à-dire en différé, sans aucune possibilité d'ajouter quoi que ce soit à la formulation initiale. C'était un état d'esprit lié à l'utilisation des cartes perforées.

De façon étonnante, Dialog permettait d'exécuter une recherche, de voir le résultat, d'ajouter quelque chose, de combiner ces éléments avec un troisième concept, de faire exécuter la nouvelle stratégie, puis de parcourir les titres des résultats obtenus, tout cela de manière interactive. Ce qui, en y repensant, était révolutionnaire.

C'est à peu près à la même époque que l'université américaine Dartmouth College inventa le concept de «time sharing» – permettant d'interroger un ordinateur à distance – et le déploya sur son campus (ndt).

Passons rapidement à l'année 2002, lorsque les professionnels de l'information furent confrontés au IAOTWFF (It's All On The Web For Free), et durent se justifier en montrant que de multiples contenus n'étaient pas accessibles librement sur le Web et en mettant en évidence les nombreuses possibilités de recherche et d'édition perfectionnées qu'ils mettaient à disposition de leur public.

Bien sûr, utiliser les services en ligne payants n'était pas à la portée des cœurs pusillanimes, car la syntaxe de recherche était différente pour chaque service et, dans la plupart des cas, le système ne pardonnait aucune erreur de l'utilisateur.

Peut-être vous souvenez-vous que 2001 fut l'année où l'on pu voir Tom Cruise dans *Minority Report*, manipuler à travers une interface de réalité virtuelle des données d'un simple geste de la main. Je suis sûr que, comme moi, tous les professionnels de l'information ont mémorisé ces scènes et s'imaginent construire facilement des stratégies de recherche complexes dans l'espace.

Malheureusement, la reconnaissance des gestes ne s'applique aujourd'hui qu'aux commandes de télévision et aux jeux vidéo.

Cependant, les services en ligne haut de gamme ont commencé à adopter la philosophie et l'approche des outils de recherche du Web et du Web social.

Bien que personne ne puisse confondre Dialog et Google, il y a des développements prometteurs pour ceux qui font des recherches en ligne.

Quel est le nouveau standard ?

L'un des plus grands défis des serveurs professionnels est de créer une interface qui soit à la fois simple à utiliser et puissante. Libby Trudell, Vice président marketing chez Dialog, a dit un jour qu'une bonne interface utilisateur devait ressembler à celle d'Excel. Les utilisateurs occasionnels d'Excel, en effet, peuvent facilement créer une feuille de calcul basique, tandis que les utilisateurs expérimentés créent avec la même interface des équations complexes.

De la même façon, les services en ligne haut de gamme veulent développer une expérience de la recherche qui soit à la fois intuitive pour les chercheurs novices et suffisamment puissante pour les chercheurs experts.

A mesure que nous nous familiarisons avec les nouveaux designs des sites web – et notamment avec la présence d'un plan du site et de nombreux liens –, nous avons développé certaines attentes concernant les outils de recherche, dans un environnement web et en particulier :

■ des suggestions pour des orthographes possibles ;

■ un classement des résultats par pertinence ;

■ des extraits de texte avec les mots de recherche en contexte ;

■ un panneau sur la gauche, offrant des possibilités de navigation ou de filtre (pensez aux options de recherche de Google).

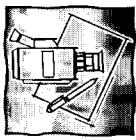
Ajoutons-y ce que nous attendons des sites institutionnels et des sites online, financés par les pouvoirs publics.

Les serveurs/agrégateurs doivent relever le défi de proposer des outils puissants pour les professionnels de l'information et, dans le même temps, une recherche intuitive aux usagers des bibliothèques.

Si l'on examine les écrans de recherches d'EBSCOhost Academic Search Premier, OCLC First Search, Gales's General OneFile, ISI's Web of Science ou PubMed de l'US National Library of Medicine, on se rend compte qu'ils possèdent tous :

■ plusieurs zones de saisie, chacune permettant de limiter la requête à un champ particulier, grâce à un menu déroulant ;

■ un menu déroulant pour relier les boîtes de recherche avec des opérateurs ET, OU SAUF ;



■ des boîtes pour limiter la recherche selon différents critères : format, domaine de la recherche, type de documents, champ spécifique à une banque de données particulière

■ troncature, recherche par phrase, logique emboîtée et autres syntaxes de recherche avancées.

L'écran présentant les résultats inclut toujours :

■ un panneau sur la gauche permettant de limiter les résultats selon différents critères ;

■ des hyperliens vers des termes de recherche et d'autres champs à vocabulaire contrôlé ;

■ la possibilité de sauvegarder, imprimer et envoyer par mail un ou plusieurs documents.

Qu'en est-il des services destinés à l'utilisateur final

Pour beaucoup de services d'information présents sur le Web, prendre en compte les spécificités du Web était assez

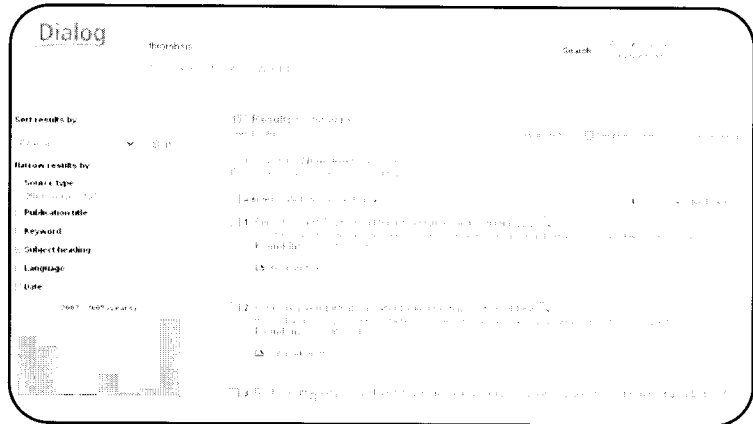
simple ; leurs fonctionnalités étaient relativement limitées, par conséquent, créer une interface qui offrait simplicité et facilité d'usage était aisément réalisable.

Quand vous partez d'une banque de données qui utilise majoritairement du vocabulaire contrôlé et des champs bien définis, vous pouvez mettre beaucoup d'intelligence dans un écran de recherche avancée.

Cependant, la question se pose de savoir comment bien répondre aux attentes pour un service en ligne proposé sur le Web, lorsque les données ne sont pas structurées ou pas uniformes.

Dialog (www.dialog.com) fournit un exemple classique de ce problème. Les documents dans Dialog peuvent être, entre autres, des brevets, des formules chimiques, des données financières d'entreprises, des articles en texte intégral, des documents émanant des pouvoirs publics ou des comptes-rendus de conférences.

Comment la même interface peut-elle permettre aux utilisateurs de trouver de l'art antérieur pour un brevet, d'identifier les auteurs les plus prolifiques sur un sujet, d'analyser la concurrence ou de surveiller les appels d'offres émanant du secteur public ?



Proquest Dialog : page de résultats (prototype)

Proquest Dialog

Proquest – le nouveau propriétaire de Dialog et DataStar – a prévu de fusionner les deux plateformes en lançant un nouveau produit baptisé Proquest Dialog.

Ce produit sera lancé par étapes. Le contenu lié à la pharmacie et à la médecine a été mis en ligne à la fin 2010 et sera suivi en 2011 par les sciences et techniques, la propriété intellectuelle et, pour finir, le business et l'actualité.

Alors que le produit est toujours en phase de test à l'heure où j'écris ces lignes, les fonctionnalités dévoilées lors des premières étapes seront très vraisemblablement maintenues dans la version finale.

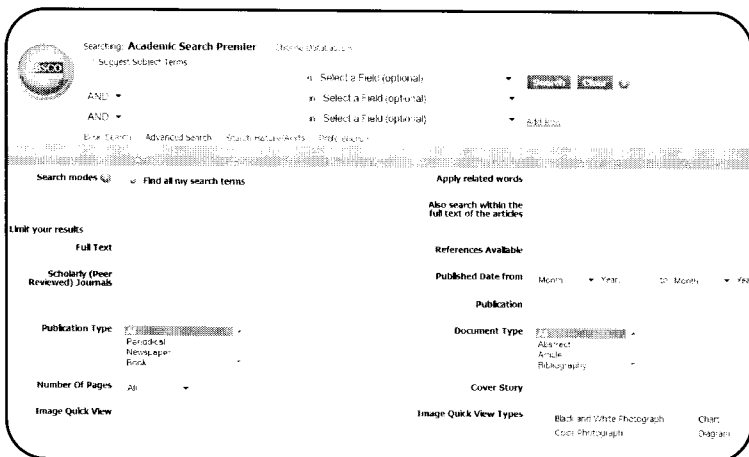
Ces fonctionnalités, que je trouve particulièrement utiles, sont les suivantes :

■ la page de résultats présente sur la gauche une colonne proposant des options pour filtrer les réponses par type de source, titre, mots-clés, grandes thématiques, langue et date.

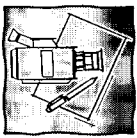
Je vois là des ressemblances avec ce que l'on trouve sur Google, Factiva ou LexisNexis. Comme chez eux, il est possible d'utiliser un menu déroulant pour modifier la période sur laquelle doit être effectuée la recherche ;

■ les utilisateurs peuvent taguer des documents, ce qui permet de retrouver facilement l'article que vous aviez vu il y a quelques semaines, mais dont vous n'arrivez pas à retrouver le titre ;

■ les résumés peuvent être traduits à la volée, de l'une des cinq langues sources vers l'une des dix langues cibles. Il s'agit de traductions machine qui manquent parfois de précision, mais qui s'en soucie ?



Academic Search Premier : grille de recherche avancée



La traduction automatique des pages web par les moteurs de recherche résulte d'un développement relativement récent, ce qui rend particulièrement intéressant son inclusion dans Proquest Dialog ;

■ le lien «cite this» conduit vers une liste de citations des documents que vous avez sélectionnés, dans des formats tels que MLA, APA ou d'autres. On trouve aussi un lien «*Export citations/abstracts*» dans la page de résultats, qui permet d'exporter une sélection de documents à un gestionnaire de citations comme RefWorks ;

■ la page des résultats contient une boîte dont le titre est «*Other items you might like*», qui propose des réponses obtenues avec une partie seulement des critères utilisés dans la recherche. C'est un moyen très efficace d'être sûr que le chercheur trouve des résultats intéressants, même s'ils ne correspondent pas exactement à la recherche telle qu'elle a été formulée initialement.

■ placer le curseur sur un document dans la page de résultats fait apparaître des informations complémentaires telles qu'une citation, un résumé ou des mots-clés. J'apprécie cette fonctionnalité dans Bing.com, et il est intéressant de la retrouver dans Dialog.

La tarification de cette nouvelle fonctionnalité n'a pas encore été établie.

En tout état de cause, les prototypes actuellement disponibles ne sont accessibles qu'à ceux qui ont souscrit un abonnement annuel ;

■ un nouvel espace «My Research» sert d'espace de travail où il est possible d'établir des stratégies de recherche, d'enregistrer des documents, de rechercher des documents préalablement tagués, de lire des flux RSS et de voir la liste de ses alertes. Je pense utiliser cet espace fréquemment, pour pouvoir organiser et documenter mes projets de recherche.

Une autre possibilité assez curieuse du nouveau Dialog est celle qui consiste à ajouter ou enlever facilement des mots ou des phrases à la formulation originale de la question.

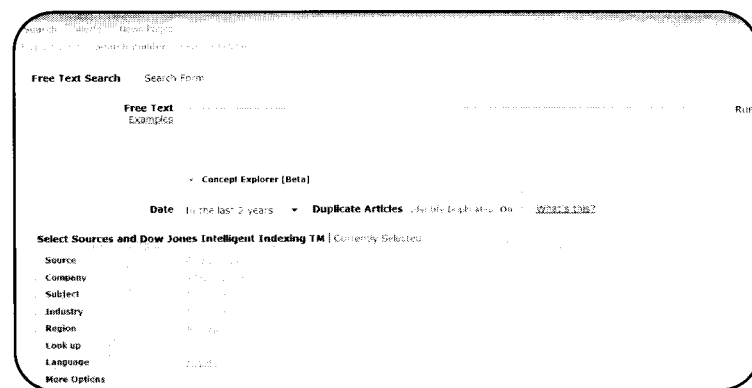
Quand vous utilisez les panneaux de navigation à gauche de l'écran pour affiner votre recherche, chacun de vos filtres apparaît dans une boîte au-dessus de l'écran de recherche. Il est alors facile d'ajouter ou d'enlever des filtres et de comparer les résultats.

Cette possibilité nous mène au plus près de l'environnement de «Tom Cruise», je peux passer facilement d'un aspect à un autre de ma recherche d'un simple clic et je peux imaginer ce clic devenir un simple glissement de mon doigt sur un écran....

Dow Jones Factiva

Dow Jones a dévoilé son interface web conviviale en 2006, à une époque où il ne semblait pas possible d'utiliser la même interface à la fois pour les professionnels de l'information et les utilisateurs moins expérimentés.

Sur l'écran de recherche, les utilisateurs peuvent choisir entre la «recherche simple» et la «création de recherche» (qui en dépit de son nom, permet aussi aux utilisateurs de formuler leurs recherches en utilisant les opérateurs booléens). Bien que cette approche soit différente du modèle d'Excel, le choix entre les deux modes de recherche est transparent.



Factiva : grille de recherche en langage libre

La page dédiée à la recherche simple comprend une zone de saisie accompagnée de menus déroulants pour limiter la requête par type de publication et par période.

La page «création de recherche» offre deux choix supplémentaires. Le formulaire de recherche, comme beaucoup de grilles de recherches avancées sur le Web, offre des zones de saisie que l'on peut combiner avec les opérateurs AND, OR et NOT.

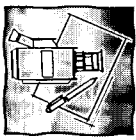
Il suffit de les remplir et les mots sont insérés dans la boîte de recherche. Dans le formulaire en langage libre, la requête doit être écrite avec la bonne syntaxe.

C'est un outil destiné spécifiquement aux chercheurs, à l'aise avec cette démarche.

De nombreuses options, permettant de limiter la requête par type de publication, sujet large, industrie, etc., sont proposées.

Factiva propose depuis quelque temps une nouvelle fonctionnalité en bêta, le *Context Explorer*, disponible dans le formulaire de recherche en langage libre.

Si vous cliquez sur n'importe quel terme de votre recherche, apparaît alors une liste ressemblant à un thésaurus de termes en relation, qui peuvent être ajoutés à la requête d'un simple clic.



Industries	Date	Companies
Energy 191	1925	Federal Energy R... 12
Electric Power G... 117	1900	The Commonwealth... 10
Industrials 99	1975	Gannett Science... 9
Renewable Energy 81	1950	United Nations 9
Solar Energy 74	1925	Pristine Power I... 7
Electricity-Gas 62	1900	Alternate Energy 6
Wind Energy 28	1 January 2009	Vestas Wind Syst... 6
Electric Power G... 28	31 December 2009	Federal Deposit... 6
Fuel Cells 22		

Factiva Discovery Pane

La page de résultats est la même pour tous les utilisateurs, à savoir une liste classique d'articles avec des petites boîtes à cocher pour sélectionner ceux que l'on souhaite visualiser, télécharger ou envoyer par mail.

trouvez une communauté interactive où Factiva vous encourage à partager vos réflexions, idées, réclamations et suggestions, avec un objectif global d'amélioration.

Nexis

Sur la partie gauche de l'écran, on trouve dans un format agréable une analyse dans les résultats de l'importance relative de sujets particuliers, de noms, de sociétés, etc. Un clic sur un des termes l'introduit dans la recherche avec un opérateur AND.

Nexis (www.lexisnexis.com) a lancé en juillet 2010 sa nouvelle interface web pour son produit «academic».

Pour plus de précision, on pourra se reporter au *News break* écrit par Marydee Ojala (<http://goo.gl/px228>).

Une autre possibilité très intéressante, qui existe aussi sur Proquest Dialog comme nous l'avons vu, est l'option de traduction automatique.

Nexis ne passe – pas plus que Factiva – le test de simplicité d'Excel, dans la mesure où il a deux interfaces de recherche différentes *Easy Search* et *Power Search*. Cela dit, Nexis a réalisé le maximum avec chacune des interfaces.

Si l'on active cette option, on peut faire traduire un ou deux paragraphes à la fois ou attendre la traduction du document dans sa totalité. Actuellement, neuf langues peuvent être traduites en anglais et les documents en anglais peuvent être traduits dans l'une des neuf langues.

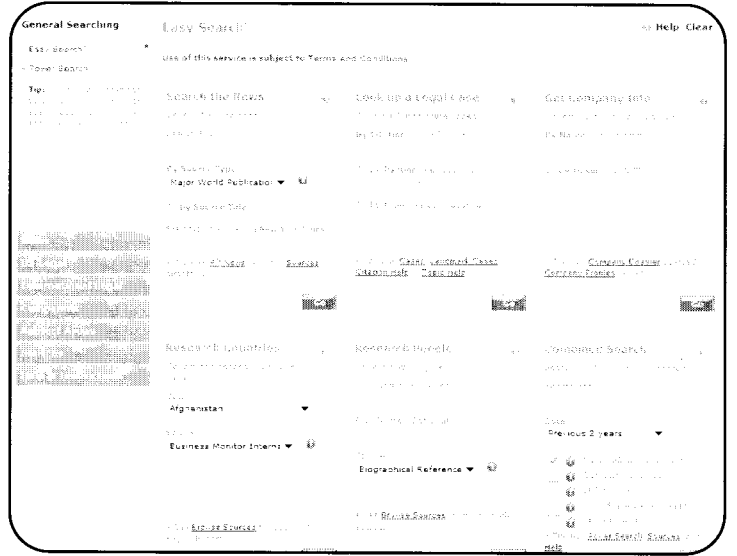
On notera que les deux formats proposent un panneau de navigation à gauche de l'écran, avec des liens et de grandes catégories, rendant intuitive la navigation dans les grandes catégories de documents.

La transparence et la collaboration sont des aspects importants du paradigme actuel du Web.

L'interface *Easy Search* est une agréable collection de modèles de recherche pour des types de requêtes fréquentes, avec également une option de recherche en texte libre.

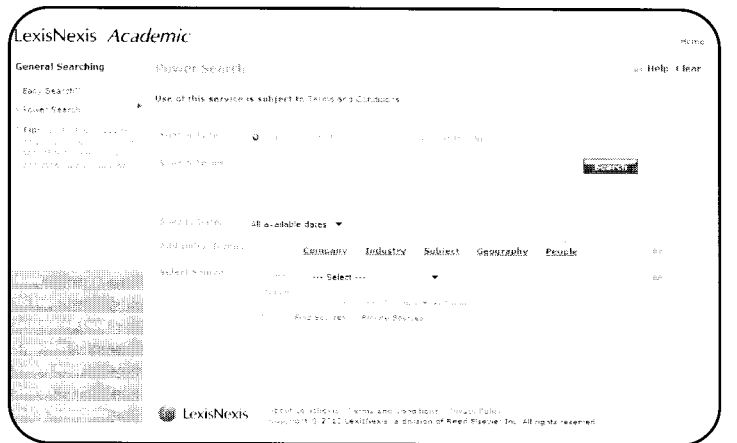
Si vous cliquez sur le bouton bleu *Dow Jones Idea Share*, vous

Malheureusement, cette dernière possibilité (sans doute la plus utile) est enfouie en bas de l'écran, la rendant difficile à trouver. L'écran de recherche «simple» de Nexis fait davantage penser à la période pré-Web que ceux de ses concurrents.

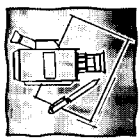


Nexis Easy Search

La «Power Search» de Nexis offre la possibilité d'utiliser le langage naturel ou des opérateurs booléens, une fonctionnalité utile lorsque vous avez des difficultés à préciser votre sujet. L'accès à l'indexation en profondeur se trouve dans des menus déroulants, rendant facile l'accès à de nombreuses possibilités.



Nexis Power Search



Orbit.com

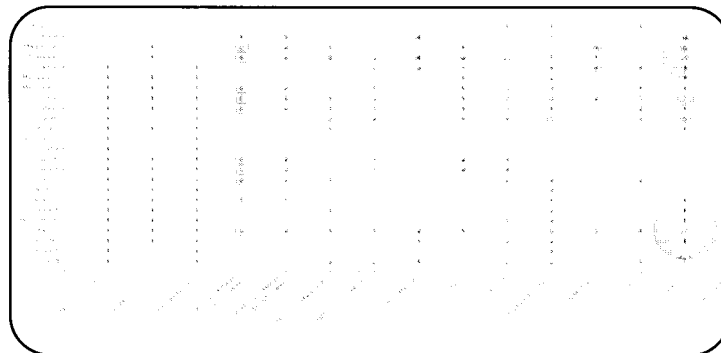
Bien que ce serveur ne soit pas habituellement considéré comme un des «Big Three» des serveurs/agrégateurs, l'ergonomie d'Orbit.com en fait un site incontournable pour les «fondus» de recherche de l'information. Même pour ceux d'entre nous qui ne sont pas spécialisés dans la propriété intellectuelle, les possibilités de visualisation des données offertes par le portail de Questel sont à remarquer.

Dans la mesure où la plupart des documents brevet comportent de nombreux champs séparés, cela favorise le data mining.

Orbit.com a simplifié les choses en préparant le format des résultats dans ce but.

C'est en effet un service en ligne qui anticipe la manière dont vous allez pouvoir extraire de l'intelligence à partir de ses services et vous faciliter la tâche.

A la vue de ces exemples, imaginez comment vous pourriez utiliser ce type d'outils pour faire du data mining dans une collection d'articles, un ensemble de statistiques ou une banque de données, avec des membres de conseils d'administration.



Orbit.com table

Un format d'édition particulièrement intéressant est un tableau qui fait apparaître les corrélations entre le contenu des brevets et les déposants. Cela peut être un indicateur très utile pour détecter si un concurrent est en train de s'orienter vers une nouvelle technologie, avec une réponse dans un format visuellement très convaincant.



Orbit.com relationship map

Une recherche réputée difficile consiste à identifier les collaborations professionnelles des chercheurs.

Orbit.com peut générer une carte dynamique faisant apparaître les relations entre inventeurs à partir de l'étude des co-inventeurs. Là encore, il s'agit d'un outil fascinant qui pourrait être appliqué dans un grand nombre de contextes.

Dans la mesure où Orbit.com est limité à la recherche dans le domaine de la propriété intellectuelle, on peut espérer que ces outils de data mining migreront vers les «Big Three».

Ce type de possibilités nous permet à nous, professionnels de l'information, de nous rapprocher du schéma de l'écran tactile à la «Minority Report» pour jouer avec l'information.

Quand Tom rencontre Roger

Après avoir examiné les interfaces web proposées aujourd'hui, mon idéal se situe au-delà d'un mélange entre Orbit.com et les tentatives intéressantes de Tom Cruise.

Mon rêve d'un monde parfait est un mélange de la pensée innovante de Roger Summit («Comment ? on laisserait les utilisateurs interagir avec les données ? êtes-vous fou ?») et de la suprême décontraction de Tom Cruise, interagissant avec l'information d'un simple geste de la main.

Compte tenu de l'aisance de Roger au piano, je peux parfaitement l'imaginer déplacer des résultats de recherche d'un simple mouvement de poignet.

Je rêve du jour où je pourrai tactilement récupérer l'intelligence contenue dans un ensemble d'articles ou dans la banque de données de la Securities and Exchange Commission !

Mary Ellen Bates
mbates@batesinfo.com