



Les solutions clé en main n'ont plus le vent en poupe.

CE QUI EST NÉCESSAIRE

- La capacité de faire de l'extraction d'information fine est un plus
- Le marché croît de façon importante
- Ce qui est important est de rester différent grâce à une expertise et la détermination d'un besoin client.

Pour les éditeurs de solutions de veille et de recherche d'information, ce qui compte avant tout c'est de s'adapter au besoin du client. Les solutions clé en main n'ont plus le vent en poupe. Il est nécessaire d'analyser les besoins des entreprises pour mettre en place les réseaux et les circuits de diffusion qui leur correspondent. Exalead, Sinequa mais aussi Infocodex, nous ont présenté des cas clients et ont particulièrement souligné cette nécessité.

La verticalisation permet à la fois, comme le souligne le cabinet Intellico, la structuration des informations et l'industrialisation du travail collaboratif comme leviers des projets d'innovation et de développement. De plus, l'information non structurée nécessite la mise en place de puissants outils d'analyse pour une diffusion et une exploitation fines de l'information retrouvée par les moteurs de recherche. La classification de cette information a pour but aussi bien la prise de décision stratégique, que la transmission de savoir et de compétences. L'explosion des sources d'information non structurées nous a fait passer progressivement d'une société de l'information à une société de la connaissance en action.

Moteurs de recherche entreprise

Search and Find

Enjeux et Perspectives

Search'2008, organisé le 21 février par Veille Magazine et Evalir, fut d'abord l'occasion d'une rencontre inédite entre utilisateurs, éditeurs et experts.

Pour cette première édition, le nombre d'inscrits et de participants à l'ensemble des conférences a prouvé la pertinence et l'actualité de la problématique proposée : la verticalisation métiers des moteurs de recherche en entreprise.

Nous vous livrons ici, librement et de façon parfaitement non-exhaustive puisque nous n'avons pas pu assister à l'ensemble des ateliers, nos impressions de la journée au fil de l'eau. Quelques tendances semblent émerger... La prochaine édition de Search, en février 2009, nous dira si nos analyses visaient juste. Par Marianne Dabbadie

Commençons tout d'abord par quelques données livrées par des experts comme Frédéric Lucard (Aestigia) ou Laurent Le Foll (Aproged), et confirmées par différents intervenants :

- Un cadre moyen passe 30% de son temps à rechercher une information qui souvent existe déjà au sein de son entreprise.
- Chaque collaborateur d'entreprise crée 850Mo de données chaque année, soit 9 mètres de livres en hauteur.
- la somme de connaissance créée par l'humanité depuis sa source a été générée dans les quatre dernières années. Les volumes dépassent la capacité de traitement de l'individu.
- le web représente aujourd'hui 30 milliards de pages

TROIS ÉTAPES DE LA PRISE DE DÉCISION

Plusieurs réalités incontournables sous-tendent la notion de recherche d'information, qui elle-même concourt activement au processus de prise de décision.

1. Il y a tout d'abord la nécessité de chercher ce qui existe et de trouver ce qui existe. Même si cela semble aller de soi, trouver ce qui existe ne correspond pas forcément à une réalité évidente. Trouver ce qui existe implique d'importants investissements technologiques. Comme nous venons de le voir, le réseau mondial compte aujourd'hui 30 milliards de pages. L'index de Google est à 20 milliards, celui d'Exalead à 8 milliards. Un tiers des ressources du web échapperait donc aux crawlers ?
2. L'information est une réalité duale : c'est un réducteur de savoir mais également un merveilleux instrument de construction de connaissance. L'élaboration de connaissance se fait par sérendipité : la pragmatique est au cœur de la démarche. * la sérendipité est une découverte, provoquée par une attitude d'esprit, qui consiste à rebondir sur les conséquences d'une aventure, d'une rencontre, d'une recherche ou d'une expérience. (wikipedia)
3. Le web 2.0 a vu s'effacer la notion de bases de données pour privilégier la notion de moteurs de recherche. Ce qui se profile derrière cette réalité liée au tagging de l'information c'est une évolution progressive vers d'autres modes de représentation que le mode linéaire : le mapping de l'information et la cartographie.

ACTUELLEMENT

Les acteurs économiques ont pris conscience du fait que s'équiper d'une solution de traitement de l'information non structurée est un enjeu important qui permet de se repérer au sein d'un océan d'information devenu mondial et de demeurer compétitif et pertinent grâce à la maîtrise de cette information.

POUR LES UTILISATEURS

Il est nécessaire de bénéficier d'une expertise pour s'y retrouver dans l'offre technologique. Les entreprises utilisatrices de solutions de veille et de Search ont de plus en plus besoin de s'y retrouver dans une offre de solutions aussi riche que foisonnante. De ce fait, s'affirme de façon croissante aujourd'hui la nécessité de recourir à une expertise extérieure à l'entreprise : le fameux tiers certificateur.

Il semble qu'aujourd'hui, il soit devenu absolument indispensable de dégager des tendances pour continuer à progresser dans un contexte devenu mondial. Du point de vue des utilisateurs la maîtrise de l'information d'un domaine professionnel est un facteur important de positionnement sur le marché mondial. L'information est stratégique lorsqu'elle concourt à la prise de décision. Elle l'est également lorsqu'elle participe à la transmission des connaissances et savoir-faire métier au sein d'un groupe.

POUR LES ÉDITEURS

Les éditeurs de solutions ont senti la nécessité de tester leurs produits par rapport à l'offre d'une concurrence de plus en plus fine et de plus en plus diverse. Les tests ne portent plus uniquement sur l'interopérabilité des modules entre eux, mais sur la pertinence et la compétitivité de l'outil. Il est important de tester la pertinence d'une solution non seulement pour elle-même mais également en la comparant aux offres équivalentes sur le marché. On voit donc se développer aujourd'hui des tests comparatifs ou benchmarks des solutions disponibles sur le marché.

LES TENDANCES

NOTION DE VALIDATION PAR LES PAIRS

- Le consensus valide l'information retrouvée et mise à disposition dans un cadre participatif.
- L'avenir de la recherche est dans la mise en concordance entre l'analyse, la transformation de l'information au sein de sa chaîne de valeur et la mise en intelligence d'un outil grâce au couplage des technologies.
- A titre de simple exemple, un outil comme Infocodex fonctionne sur la base du couplage de trois technologies : statistique, sémantique, réseaux de neurones

Les grands de la recherche d'information ne s'y sont pas trompés. Microsoft a misé sur le collaboratif avec SharePoint dont il existe une version gratuite. Et le géant de l'informatique a développé une boîte à idées qui réunit des PME innovantes, leur donnant les moyens de mettre en œuvre des projets qui donneront peut-être la tendance de demain.



ON CONSTATE UNE NÉCESSITÉ CROISSANTE DE

- Gérer le multilinguisme :
Des outils comme Sinequa, Infocodex, Exalead, gèrent le multilinguisme. Infocodex identifie automatiquement la langue utilisée.
- Indexer automatiquement des métadonnées structurées ET non structurées
- Catégoriser et filtrer l'information
- Gérer l'extraction d'entités nommées
- Disposer de la bonne information just-in-time et à la bonne personne
- Bien connaître les enjeux de la propriété intellectuelle pour le développement de la recherche et de l'innovation (ce que je fais a-t-il déjà été inventé, quel est l'état de l'art ?).
- Demeurer en conformité avec la réglementation en vigueur : La réglementation impose de pouvoir prouver l'origine des données et en développer la traçabilité.

CE QUI SE FAIT ET SE FERA DE PLUS EN PLUS

LES ACTEURS DEVIENNENT INTERACTIFS : ON NE RÉINVENTE PLUS LA ROUE.

De plus en plus souvent, les acteurs du Search collaborent pour allier leurs technologies au sein d'applications de verticalisation de plus en plus puissantes. Ainsi Temis collabore avec Sinequa pour enrichir sa solution d'une brique de data mining, Exalead collabore avec Mondeca afin de disposer d'ontologies pour le traitement de l'information spécifique.

UNIFICATION DES SOURCES ET DES CONTENUS DE FAÇON TRANSVERSALE

De plus en plus d'éditeurs sont capables de développer des connecteurs qui permettent non seulement de fédérer dans un seul moteur les différentes sources d'information, mais également de réinjecter l'information de façon verticale dans une chaîne de traitement et de valorisation.

Voici quelques exemples de ces connecteurs glanés au fil de la journée :

- Sinequa propose de nombreux connecteurs. Les développeurs de l'entreprise sont capables de connecter n'importe quelle source de données en quelques heures grâce à une technologie spécifique.
- Microsoft Search Server 2008 est capable de fédérer en temps réel les résultats de n'importe quel moteur de recherche
- Exalead a développé des outils de facilitation de la migration d'information entre bases de données
- Kartoo cartographie diverses sources d'information grâce à son meta- moteur.

LE GAP GÉNÉRATIONNEL

Le gap générationnel soulève deux problématiques, celle de l'appropriation des technologies par les utilisateurs d'une part, celle du départ d'un savoir faire d'autre part. Ce départ est couplé à l'arrivée sur le marché du travail d'une jeune génération qui maîtrise particulièrement les technologies du web 2.0.

L'appropriation concerne la tranche des 40/60 ans. L'implantation dans une entreprise d'une plateforme technologique qui implique une participation des acteurs, doit être accompagnée d'une action de formation. Sans formation et animation des réseaux, la mise en œuvre aboutit bien souvent à un échec par persistance des mauvaises habitudes et manque d'appropriation d'un nouvel outil. On a ainsi constaté que environ 50% des mises en place de solutions de recherche d'information se soldent par un échec.

NÉCESSITÉ D'ADOPTION DES TECHNOLOGIES PROPOSÉES PAR LES UTILISATEURS.

Deux générations posent question.

La montante

Ces enfants nés avec un ordinateur entre les mains commencent à intégrer les entreprises. Ils savent où trouver l'information et comment l'exploiter et possèdent un usage courant des technologies. Il s'agit cependant souvent d'un usage ludique, qu'une bonne pédagogie leur permet de transposer sur le plan professionnel.



WWW.SEARCH2008.ORG

La sortante

Le mouvement de départs massifs à la retraite souvent appelé « papy-boom » pose la question du départ d'un savoir faire métier qui ne doit pas mourir. Le défi consiste donc à conserver trace de ce savoir faire. La transmission des compétences se fait au travers de la constitution de bases de connaissances textuelles et multimédia accessibles aux successeurs des Anciens.

CE DONT ON N'A PAS PARLÉ...

L'INDEXATION DES PHOTOS ET DES VIDÉOS

Avec le développement du multimédia, ce thème représente actuellement l'un des enjeux majeurs pour les moteurs de recherche. L'indexation d'images et de contenus vidéo se fait essentiellement à partir de tags, plus ou moins bien renseignés lorsqu'il s'agit de technologies collaboratives grand public comme Daily motion ou Youtube. Les technologies de reconnaissance de formes et de Speech to text pour l'identification d'images et la description de contenus vidéo seront vraisemblablement amenées à se développer de façon très importante dans les années à venir.

LE RÔLE DU SEARCH DANS LES APPLICATIONS DE TÉLÉPHONES MOBILES

Ceci implique également une réflexion sur les parts de marché des consultations des moteurs de recherche via le téléphone portable aussi bien en France que dans le monde et notamment dans les pays émergents.

CE QUI SE DESSINE ...

PASSAGE DU WEB 2.0 À L'ENTREPRISE 2.0

L'entreprise 2.0 est également une thématique sensible, bien qu'encore émergente. Le participatif gagnera peu à peu le monde de l'entreprise au travers des applications collaboratives et les bases de données pour la transmission de la connaissance.

CARTOGRAPHIE DE L'INFORMATION

La visualisation permet d'être plus réactif et de se positionner dans un océan multimédia. Cette tendance se développera de plus en plus dans un monde qui voit une montée en puissance de la masse des contenus multimédia.

ÉVOLUTIONS DU WEB À VENIR

Si l'entreprise 2.0 a le vent en poupe, le web 2.0 bat aujourd'hui son plein. Aussi surprenant que cela puisse paraître, il est déjà presque dépassé. A été évoquée l'émergence du web 3.0. Encore mal défini, pour les acteurs des technologies, le Web 3.0 serait un web réellement sémantique, intelligent et capable d'interpréter automatiquement les données textuelles. Enfin derrière le web 3.0 se profile déjà le web 4.0 qui d'ici quelques années s'imposerait comme étant le web de la connaissance ou plutôt le web de la société de la connaissance en action. M.D.

Nos remerciements à nos

partenaires : Aestigia,

Inévidence, Microsoft

France, Exalead, Sinequa,

Questel, Infocodex,

Intelleco, Kartoo, KB Crawl,

Aproged, Orca Security.

QUESTIONS

Certaines informations, livrées au fil des commentaires tout au long de cette journée n'ont pas manqué de surprendre laissant l'auditeur un peu pantois, avec devant lui une porte ouverte sur un futur étonnant, parfois presque inconcevable. Ainsi par exemple, aujourd'hui, nous a-t-on dit, oui, aujourd'hui ... tout le savoir produit depuis 2004 sera égal en volumétrie au savoir de l'humanité depuis son origine.

De quel savoir s'agit-il donc et comment l'appréhender ? Des questions passent dans l'esprit et défilent, comme une ombre, rapidement oubliée face à ces technologies pleines de promesses. On se demande pourtant...

Devant cette masse mouvante de données, cet océan d'expression, qu'est-ce qui est du domaine de la connaissance, du savoir ou encore d'une banalité vouée naturellement à disparaître ? Dans un tel contexte, qu'est-ce donc que la pertinence ? Le volume de données créé est-il lié à la création de connaissance réelle et utile ou peut-il s'apparenter à une parole écrite, qui passe, étant naturellement destinée à l'oubli ?

Quelle est la valeur d'une information produite en masse ? Il est, bien entendu, d'autant plus indispensable de faire le tri que le volume de données disponibles ne peut être appréhendé par une seule intelligence humaine ou même une équipe. Aussi doit-on faire un réel effort de catégorisation, à l'aide d'outils suffisamment performants, afin de pour pouvoir appréhender rapidement l'information et se l'approprier

Une autre question revient, de façon lancinante mais persistante... face à cet océan de données, que devient donc la culture, celle qu'on nous a enseignée à l'école et à l'université, celle aussi que nous nous sommes appropriés par nos propres recherches ou vagabondages ? Ce socle commun qui est supposée constituer le lien social culturel entre les individus d'une société ? Doit-on envisager de revisiter le concept de culture ? Comment faut-il le définir ou le redéfinir ? Et si nous tentions d'apporter à ces questions quelques éléments de réponse ?

Pour cela www.search2008.org continuera à vous proposer des informations, des regards, des liens, des analyses...

Mais surtout vos questions, vos réactions et contributions nous intéressent.

mdabbadie@evalir.fr