
Les moteurs de recherche verticaux

Christophe ASSELIN

Recherche **verticale** ?

- Définition
 - Outils de recherche (seuls ou intégrés à un portail, un répertoire)
 - focalisé sur un secteur, une spécialité professionnelle (les télécoms, le juridique, la biotechnologie, la finance) ou contenu transverse dédié B2B
 - contenu dédié sélectionné manuellement ou automatiquement
 - couche de métadonnées pour les plus perfectionnées
 - Pas une spécialisation par **formats** : blog, forums, images, vidéos....
 - Vertical, sectoriel, spécialisé, B2B, portail industrie

Recherche **verticale** ?

- La recherche verticale grand public
- Le marché professionnel
- Typologie de la recherche verticale
- Exemples
- Nouvelles tendances
- Les Custom Search : GYM et autres services

Recherche "verticale" grand public

- Focalisé sur un secteur mais ciblé sur le consommateur final, avec très souvent une finalité marchande.
dérive sur la géolocalisation, l'annuaire, les comparateurs. Pour tous les goûts



properazzi

Yakaz

- Immobilier, Tourisme, Emploi. Exemple :
www.properazzi.com, www.comintoo.com, www.yakaz.com, www.gloobot.com
www.viteloge.com

COMINTOO

Le moteur de recherche d'annonce

- Jeux vidéos :
www.gameskoot.fr (techno Exalead),



- Tous les domaines sont couverts : restaurants, enfants, OVNI :
<http://www.sprice.com>,
www.boorah.com/restaurants "vertical search engines on restaurants"



Pourquoi ce développement de la recherche **verticale** **Digimind.**

?

- Des besoins, un marché
 - Seulement 43 % des professionnels trouvent ce qu'ils désirent via un moteur de recherche horizontal, après plusieurs tentatives,
 - La moitié de ceux qui ne trouvent pas se dirigent vers un moteur de recherche vertical
 - 38% estiment qu'il y a un manque de bons moteurs de recherche verticaux dans leur secteur
(e-consultancy, novembre 2007)
- Le marché des outils de recherche verticaux B 2 B : croissance de 15% par an
- Estimé à 1 milliard de dollars en 2009 (600 millions en 2005) (Outsell Inc. 2006)
- L'accès au marché des outils généralistes est devenu impossible : barrières technologiques et financières, prédominance de Google > développement de moteurs B2B
- Source d'augmentation de revenus pour les éditeurs de contenus spécialisés
- Publicité ciblée

Pourquoi ce développement de la recherche **verticale** **Digimind.** ?

- Des critères économiques

Les pertes de temps liées à la recherche d'information :

	Heures par semaine	Temps passé à la collecte %	Temps passé à l'analyse %	Impact sur la productivité
2001	8	44	56	NA
2004-2005	10,9	52	48	- 4,3 %
2005-2006	12,2	55	45	- 6,2%

(Outsell. Inc 2006)

- Plusieurs critères impactant la perte de temps :
 - La complexité du secteur
 - La formation des professionnels
 - La performance des outils (pertinence, adaptation au secteur)
 - La surabondance d'outils (multi interrogation)

Pourquoi ce développement de la recherche **verticale** **Digimind.** ?

- Un Google omniprésent : surtout en France (90 % parts de marché) fournissant :
 - des résultats généralistes,
 - en vrac
 - issus de sources hétérogènes
 - difficile à désambiguïser, même pour des professionnels de l'information



- ▶ trop d'informations, trop de pages indexées , pas de différence entre un blog qui publie un billet par an sur un domaine et un site qui traite en profondeur d'un sujet
- ▶ Pas de connaissance du contexte de votre recherche
- ▶ Seulement des fonctionnalités d'affinage

Pourquoi ce développement de la recherche **verticale** ?



- Les avantages (réels et attendus) de la recherche verticale

Professionnels équipés :

- Rapidité d'accès à l'information désirée 53%
- Focus sur le secteur d'activité 50%
- Gain de temps 48%
- Premiers résultats pertinents 47%

- Contextualisation (pas de recours obligé à l'affinage)

- Index limité en quantité, important en qualité :

- sélection par des spécialistes (sources ou informations)
- ligne éditoriale,
- enrichissement manuel ou par des algorithmes

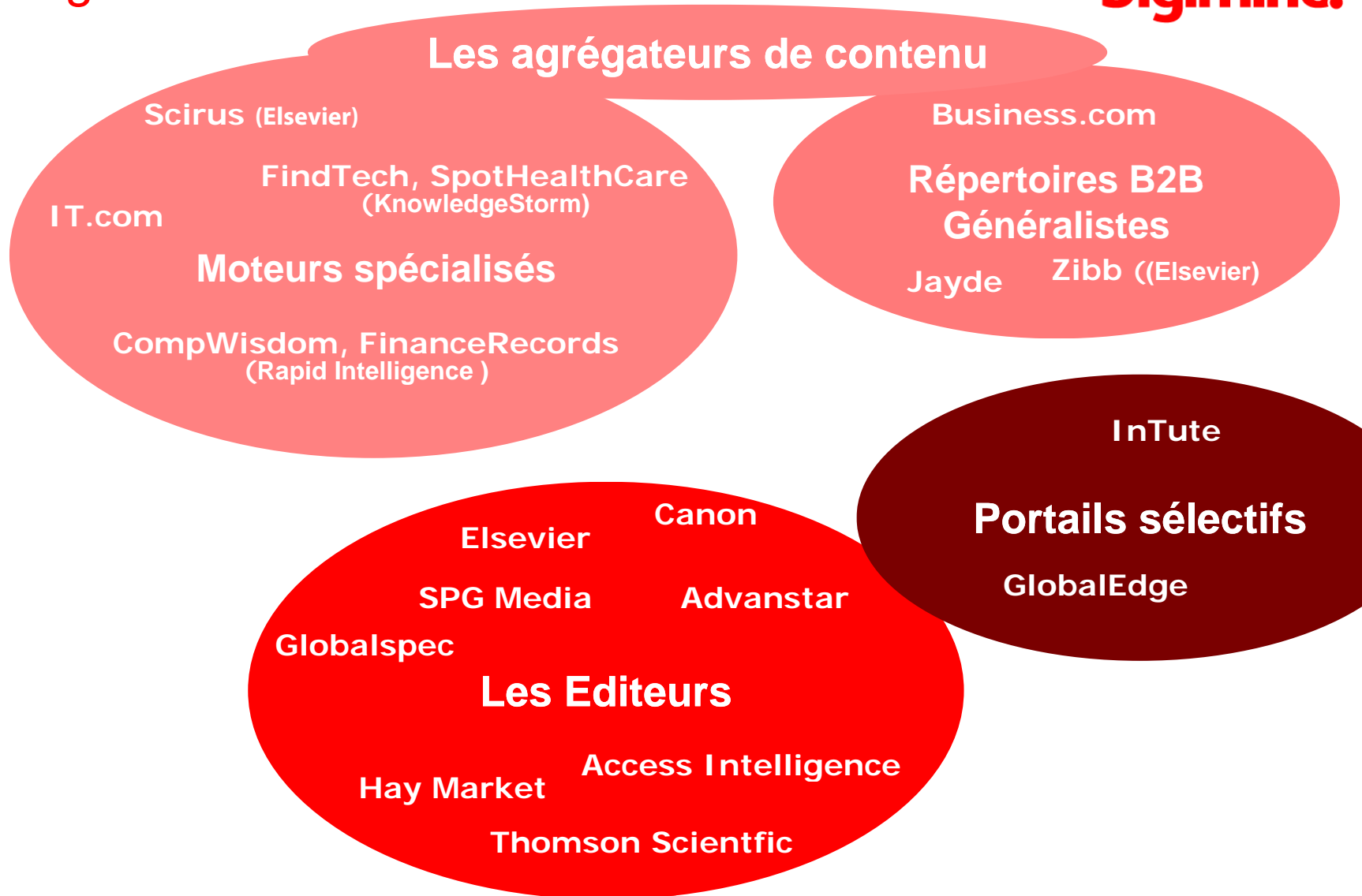
Typologie de la recherche verticale

- Recherche spécialisée ou généraliste B2B multi-secteurs

Acteurs :

- Start-up, SSII, Conseil : Outils agrégateurs de contenu, moteur de recherche (pure players Search, pas de contenu propre mais technologies + connaissances sectorielles)
 - Rapid Intelligence (financerecords.org, compwisdom.com)
 - IT.com
 - TechTarget : KnowledgeStorm (findtech.com, spothealthcare.com, spottelecom)
 - Secteurs porteurs : Finance, Santé, IT
- Les éditeurs : gèrent leur propre contenu et/ou le contenu de partenaires
- Universitaires : Portails académiques et répertoires sélectifs

Typologie de la recherche verticale



L' Exemple d'Elsevier

- Groupe Reed Elsevier
- Division Business >
 - Revues pro (*Computer Weekly, NewScientist*)
 - Un répertoire B2B généraliste : [Zibb](#) (janvier 2007), développé sur la technologie FAST (All The Web) + Catégorisation, extraction d'entités nommées, taxonomie Teragram

- ▶ Sélection de sources professionnelles recherche "[plastic](#)", = premiers résultats : liste de fabricants du secteur de la plasturgie.
- ▶ Différents types de résultats en sélectionnant des onglets :
 - All* (tous les résultats mixés),
 - News & Blogs* (les actualités),
 - Directory*
 - Web (filtré)*
- ▶ Affinage par sources, sociétés, secteurs
- ▶ Termes connexes (related searches)
- ▶ Génération de flux RSS à partir de votre requête

L' Exemple d'Elsevier (2)

- Groupe Reed Elsevier
- Division Science & Medical >
 - Revues Pro (*The Lancet*, *Gray's Anatomy*)
 - Un moteur scientifique généraliste
 - [Scirus](#) (2001), développé sur la technologie FAST (All The Web)

- ▶ 150 millions pages
- ▶ Mixe web (université, chercheurs, sociétés), et bases de données (ScienceDirect, MEDLINE, USPTO patents, ArXiv e-prints), contenu Elsevier

- ▶ Recherche avancée puissante
- ▶ Collaboration de spécialistes sur le contenu via un Wiki :
Topic Pages
- ▶ Affinage par sources (web ou revues), mots clés

L' Exemple d'Incisive Media

- Editeur britannique
 - Revues Pro (*en finance, IT, immobilier*) : VNUNet, IWR
 - Un moteur IT spécialisé White papers, information produits, rapports
 - [IThound](#) (2008), développé sur la technologie FAST (All The Web)
- ▶ Contenu IT présélectionné
- ▶ Recherche par moteur ou par catégorie
- ▶ Affinage par catégories significatives : Software, Hardware, Research, Services

L' Exemple de Search Finance

- Groupe Penton Media
 - Moteur de recherche spécialisé "Finance Executives" : niche des décideurs (2005)
 - Navigation par répertoire
- ▶ Large spectre de sources : sites corporate, actualités, blogs, podcasts, webcasts, emplois, événements, payant / gratuit
- ▶ Chaînes thématiques
- ▶ Veille : Fils RSS pour surveiller les requêtes, alertes

Multimédia, Niche, Contenu riche et sélectionné web + contenu de l'éditeur

<http://www.searchfinance.com>

L'exemple du secteur **pharmacie-santé**

- Très "riche"
- Multiples initiatives, avec des cibles variées :
Chercheurs, médecins, étudiants, industrie pharma (scientifique, marketing, finance),
puis professionnels de l'information, patients et grand public
- Des bases de données professionnelles (Pub Med) aux moteurs Web 2.0 (Medstory)
- 6 moteurs caractéristiques >>>

L'exemple du secteur **pharmacie-santé**

- [Kosmix](#) (2006):

- Positionnement : grand public et pro : agrégation de contenu web > santé
- Recherche "universelle" sur la santé : texte, vidéos, image.
- raffinage
- plus de 4 milliards de pages
- Interface grand public

- [SearchMedica](#) (2006)

- Positionnement : Pro : Généralistes, pharmaciens, infirmières, étudiants > Médecine
- initié par l'hebdomadaire spécialisé britannique Pulse, (enquête auprès de médecins généralistes, concluant qu'il n'y avait pas d'outils de recherche qui leur convenaient réellement)
En France, le moteur est avalisé par le Quotidien du Médecin (même groupe médias CMPMedica)
- élaboré en collaboration avec un groupe d'experts, médecins et cliniciens.
- rechercher parmi les sites web sélectionnés par les médecins ou sur tout le web.
- Affinage par sous-catégories
- recherches connexes
- Une recherche en français vous rapporte aussi des résultats en anglais. Exemple: Neurology/neurologie
- Résultats différents selon les cibles (onglets) : médecins, étudiants
- Version française sur enregistrement

L'exemple du secteur **pharmacie-santé**

- [Medstory](#) (2006):

- Positionnement : grand public et pro : agrégation de contenu web > santé, médecine avec séparation plus nette que Kosmix : Helath / Research
- Racheté par Microsoft en 2007
- Raffinage par mots clés : essais cliniques, nutrition, médicaments, experts...
- Indexation de contenu multimédias, pro ou grand public : Web ou News/Media, Audio/Video, ClinicalTrials, Research Articles.
- Génération de flux RSS de votre requête pour la mettre en surveillance.
- Ajout à Del.icio.us (bookmark social).
- octobre 2006 un partenariat avec le Wall treet Journal Online et BreastCancer.org
- Interface avancée visuelle

- [Medworm](#) (2006)

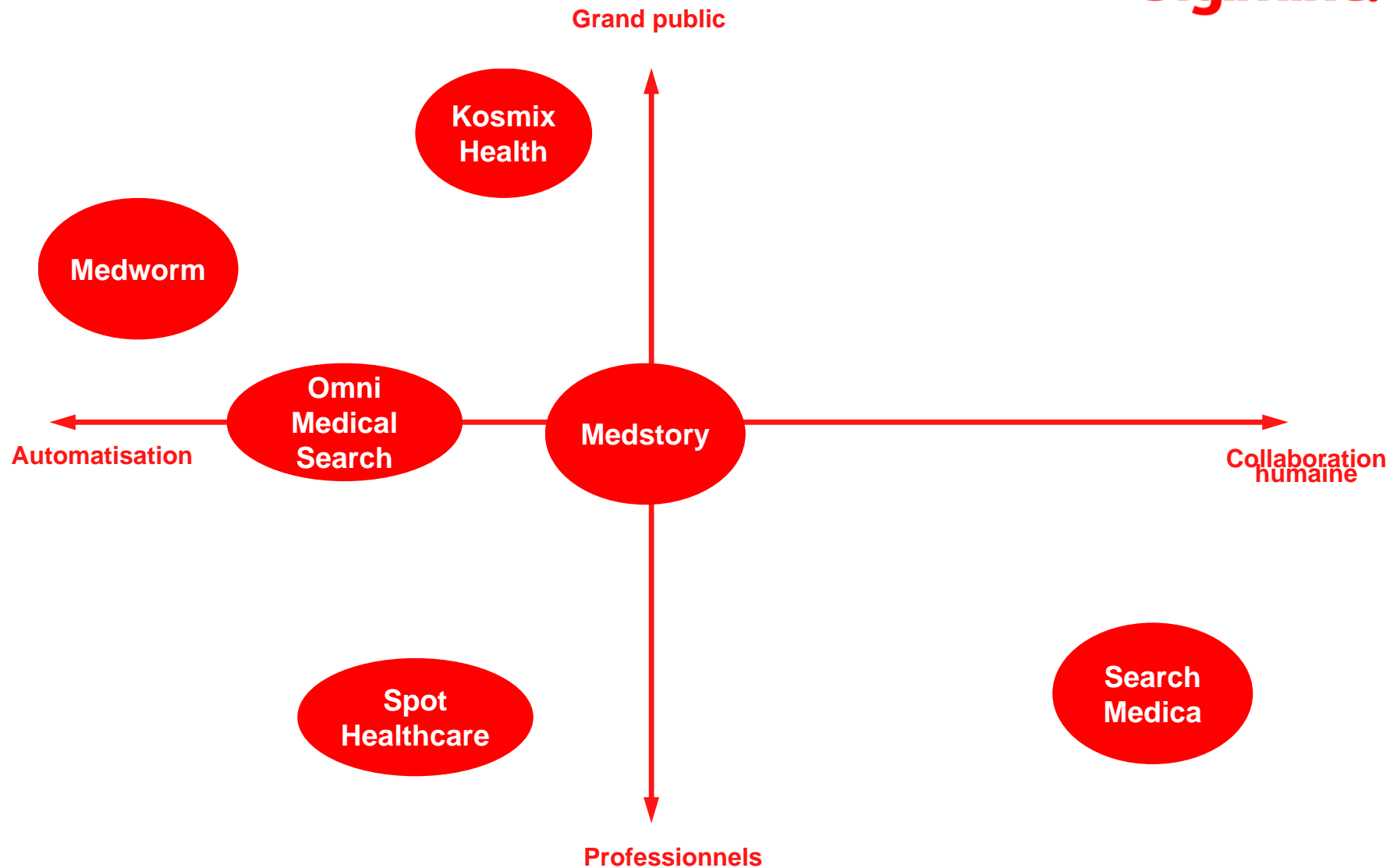
- agrégateur et fournisseur de flux RSS qui sélectionne plus de 5000 flux issus de sources sélectionnées.
- Génération de flux RSS à partir de votre requête.
- Affinage par sources
- (du même auteur : Libworm recherche parmi des flux de documentalistes, veilleurs, chercheurs et bibliothécaires).

L'exemple du secteur **pharmacie-santé**

- [SpotHealthCare](#) (2007):
 - Positionnement : pro : agrégation de contenu web > santé, médecine et pharmacie
 - résultats issus de grandes Bases de données privées ou publiques (Medscape) ou de revues spécialisées, de sociétés savantes.
 - Génération de fils RSS à partir de sa requête afin de surveiller les nouveautés.
 - VA ?

- [OmniMedical Search](#) (2004):
 - Positionnement : pro + grand public
 - résultats issus de 70 grandes Bases de données privées ou publiques
 - Web, news, images, forums, bases professionnelles
 - Interface Web 1.0

L'exemple du secteur pharmacie-santé



Les nouvelles **tendances** de la recherche **verticale** **Digimind.**

- Des acteurs de plus en plus nombreux (start-up Search spécialisées B2B / barrières à l'entrée d'un moteur généraliste)
- Outils intégrant de multiple sources et formats : texte , multimédia, web, bases de données
- Fonctionnalités "web 2.0" : rss, bookmarks sociaux, wiki (Scirus Topic Pages), personnalisation par l'utilisateur
- Interfaces plus "user friendly" : raffinement plus visuel, couches web 2.0 sur les moteurs web 1.0 :
ex : **PubMed** via **Hubmed** <http://www.hubmed.org>
et **PubMed Interact** <http://pmi.nlm.nih.gov/interact/index.php#top>

Les nouvelles **tendances** de la recherche **verticale** **Digimind.**

- L'approche Digimind : WebNews Factory
 - **Contenu sélectionné** par des experts veille. Plus de 450 000 sources au 25 mai.
 - **Bouquet de sources** spécialisées : Santé-Pharma, It-Télécom, Banque-Finance, Energie , etc..
 - **Tagage**
 - **Interface de recherche avancée** sur plus de 40 critères (type de sources, secteurs, type d'événement)
 - **Sémantique**
 - Extension sémantique des requêtes
 - Extraction/Reconnaissance automatique des entités nommées
 - Analyse de Tonalité
 - **Tableaux de bord visuels** d'aide à l'analyse :
 - algos et graphes pour analyser tendances et relations entre sociétés, personnes, produits, pays, thématiques,...
 - Analyse temps réel de forts volumes de données
- Voir : <http://www.digimind.fr/produits/digimind-evolution/modules/digimind-portal>

Les incursions des moteurs généralistes

- *Google, Yahoo! et MSN Live Search*

Custom Search : définition du spectre de recherche : périmètre des sites que vous souhaitez = ajout des sites de son choix

Faiblesses : **limité à l'index des moteurs** (grosse carence en sites professionnels)

- Et aussi :
Eurekster, Rollyo : utilisent l'index de Yahoo!

Merci !

Digimind.

Christophe ASSELIN

christophe.asselin@ digimind.com

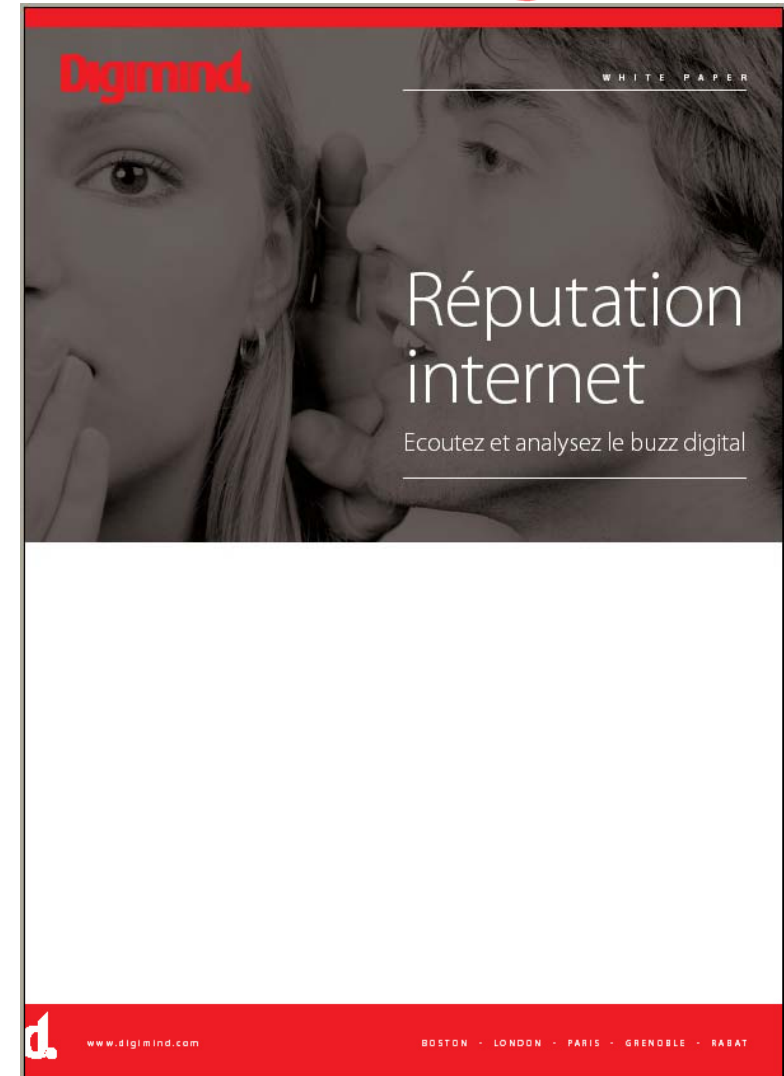
01 53 34 08 08

White paper

Réputation Internet

Ecoutez et analysez le buzz digital

<http://www.digimind.fr/actus/publications>



WWW.DIGIMIND.COM

I-Expo
28 mai 2008

Christophe ASSELIN
PAGE 23