

Technologies de la connaissance et recherche d'information

Luc Grivel

Maître de Conférences

Université Paris 1



2 juillet 2010 : Journée d'étude académique et professionnelle "Architectures Numériques d'Informations. Usages, Contenus, Technologies".

Co-organisée par "Document numérique & Usages" et le Service Commun de Documentation de l'Université d'Artois.

Déroulement de l'exposé

- Les limites des moteurs de recherche sur le Web
- Les fonctions d'assistance sémantique dans les moteurs de recherche évolués
- Les techniques employées
- Exemples de systèmes de recherche sémantique
- Conclusion

Limites des moteurs de recherche d'information sur le Web

Le schéma classique « formulation d'une requête sous forme de mots clés, retour d'une liste ordonnée de documents » est trop limité pour satisfaire toutes les pratiques de recherche.

Les moteurs de recherche :

- sont peu adaptés à la recherche d'une information précise, mais jamais vue (découverte de connaissances).

Impossible de faire une recherche sémantique, par exemple :

- *Trouver tous les fabricants de montres*
- *Trouver toutes les sociétés rachetées par la société X*
- *Trouver tous les effets secondaires dus au médicament X*
- Prennent mal en compte la polysémie (Exemple : requête « Orange » → documents consacrés à la couleur orange, à l'entreprise Orange, à la ville d'Orange, aux Comtes d'Orange, ...)

Fonctions sémantiques

Trois fonctions peuvent être remplies par un traitement sémantique

- Indiquer le type de quelque chose. Par exemple, *Pharmacia Corp.* est un type de *société*.
- Reconnaître que deux formes différentes en surface sont la même chose. (masc./fem. Sing./pluriel mais aussi *acquérir des parts* == *acheter des actions*;
- Indiquer la relation entre deux choses. Par exemple, *Pharmacia Corp.* est racheté par *Pfizer*.

Les traitements sur les documents

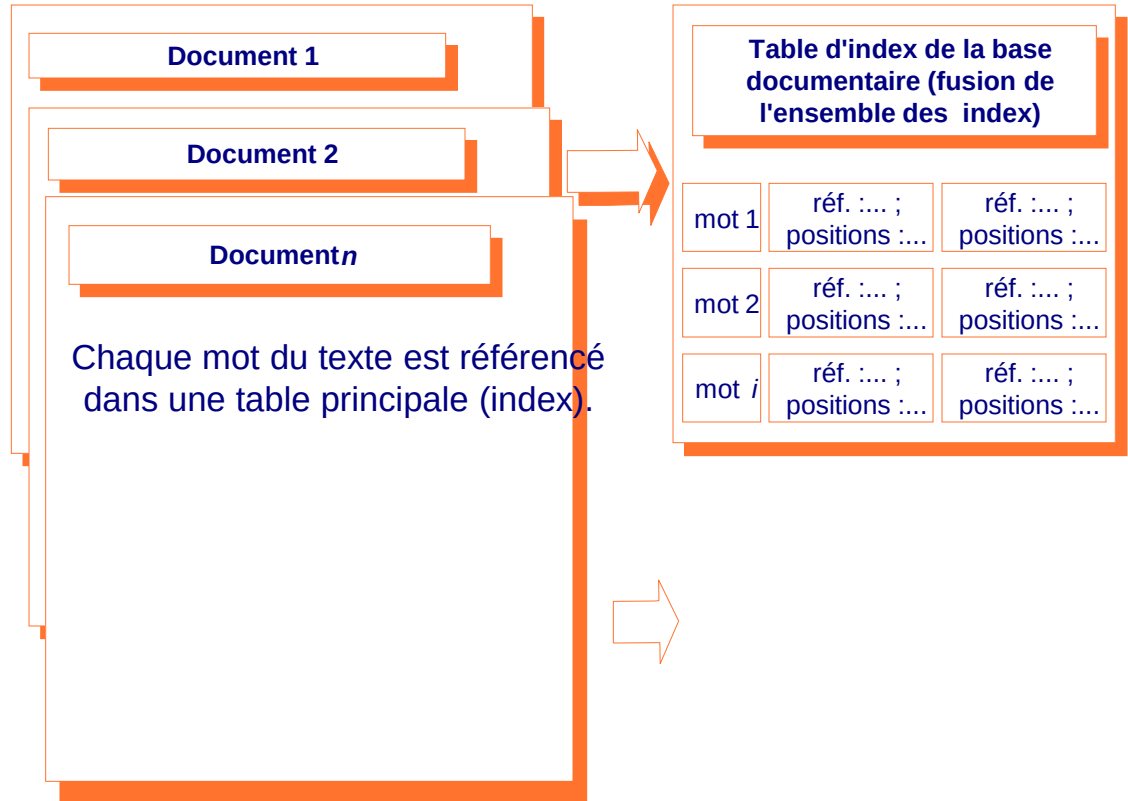
- La création de meta-données, c'est-à-dire des informations sur le contenu des documents, peut se faire **au niveau de chaque document pris isolément, et/ou en tenant compte qu'un document fait partie d'une collection de documents** constituée pour répondre à un besoin ou un objectif précis
- Techniques employées : mots vides, lemmatisation, phonétique, entités nommées, typage sémantique, mise en évidence de relations, catégorisation, clustering

Indexation

indexation full text,
éventuellement filtrée
par des critères
statistiques (fréquence,
position dans le texte,
pondération) et une
liste de mots vides
(mots fonctionnels
comme « de, en, ... »)

Indexation par extraction
d'éléments sur des critères

- Lexicaux
- syntaxiques
- sémantiques



Extraction d'éléments sur des critères lexicaux, syntaxiques , sémantiques

- Extraire des termes (descripteurs thématiques) relatifs à un domaine d'activité sans dictionnaire,
- Extraire des termes (descripteurs thématiques) relatifs à un domaine d'activité avec dictionnaire
 - Reconnaissance des termes de lexiques contrôlés sous leur forme d'enregistrement
 - Reconnaissance des termes sous des formes variantes de la formes d'origine :
 - **les variantes flexionnelles**
 - **les variantes syntaxiques**
 - **les variantes morpho-dérivationnelles**

Reconnaitre que deux formes différentes en surface sont la même chose

La **variation flexionnelle** :

method / methods acheter / acheté directeur / directrice

La **variation syntaxique** peut être :

- d'insertion : ajout de tout mot à l'intérieur du groupe nominal
competitive business intelligence → *competitive intelligence*

- de coordination : formes coordonnées de mots (adjectifs ou noms) à l'intérieur du groupe nominal

information and library sciences → *information sciences*

- de permutation : tout mot ou groupe de mots pouvant permuter autour d'un élément pivot (prépositions ou séquences verbales)

source for company information → *information source*

Reconnaitre que deux formes différentes en surface sont la même chose

La **variation morpho-dérivationnelle** : mots pouvant être dérivés en d'autres mots de catégories grammaticales différentes :

abdomen / abdominal dérivation nom → adjectif

modify / modification dérivation verbe → nom

rare species / rarely encountered species

dérivation de l'adjectif *rare* → adverbe *rarely*

insertion de *encountered*

changement de l'alimentation électrique / changer l'alimentation électrique

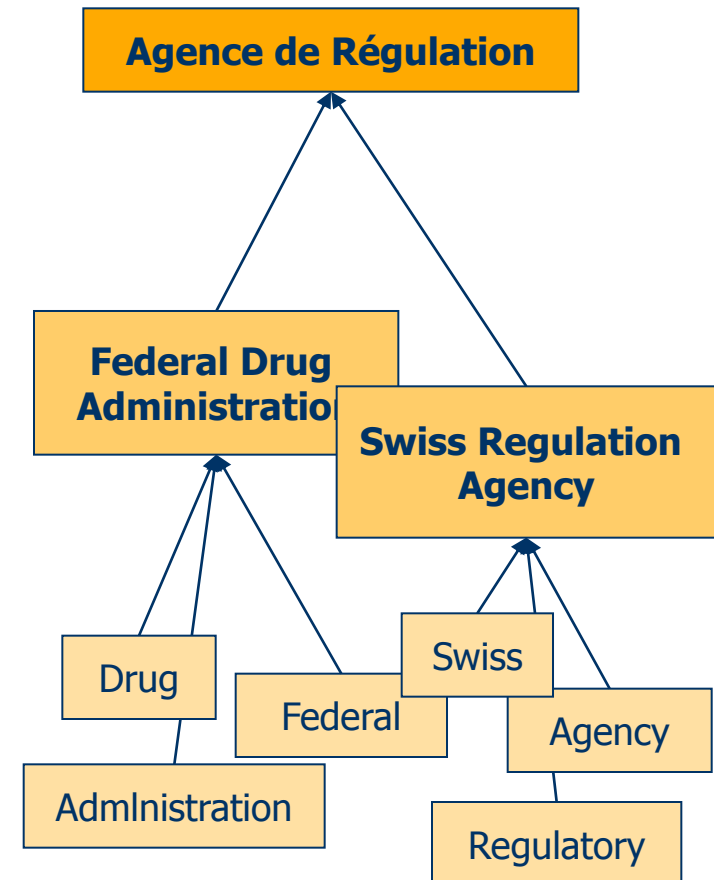
dérivation du nom *changement* → verbe *changer*

Typer

= sélectionner la signification des mots :

- noms d'organismes,
- noms de molécules,
- de traitements,
- de maladies, etc.

→ **extraction d'entités nommées**
(acteurs, lieux, dates, montants financiers, noms de molécules, noms de traitement, etc.) sur la base de listes et/ou de règles



Reconnaître des relations

Exemples :

- action d'une protéine sur un gène,
- achat de la société pharmaceutique X par la société Y en 2005,
- résolution d'un problème par une invention,
- fusion d'une société A avec une société B,
- effets secondaires dus au médicament X, ...)
- ...

nécessite l'identification des arguments de chaque prédicat et leur rôle sémantique

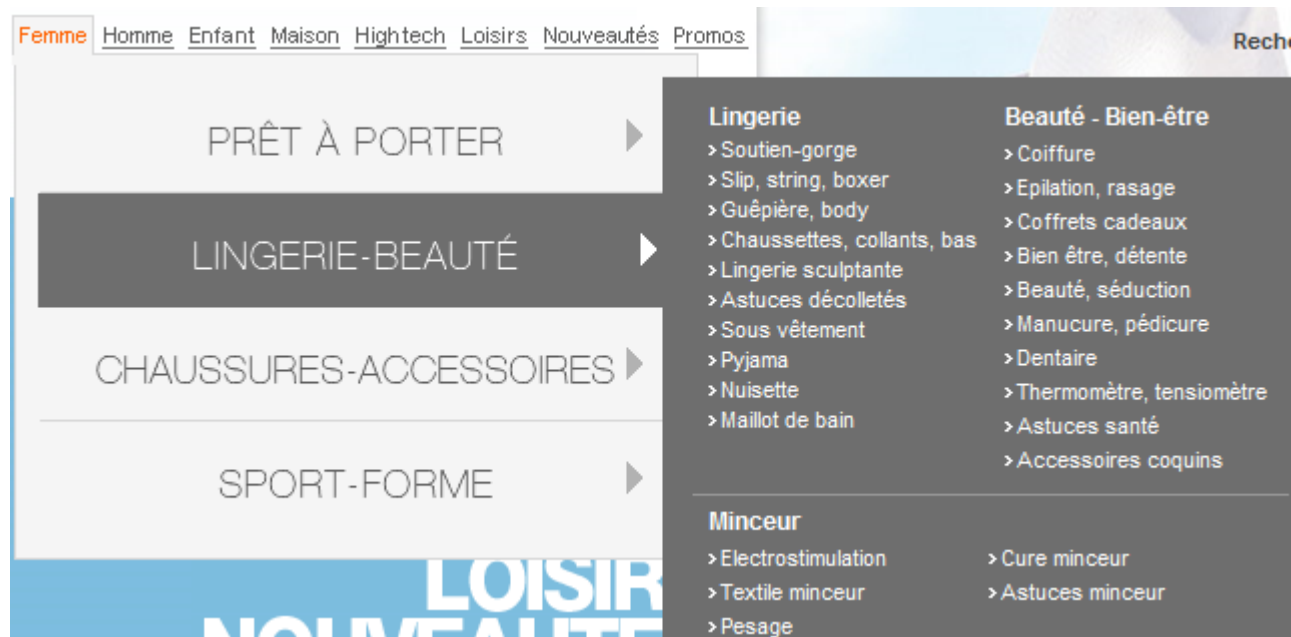
XXL Corp., éditeur de ..., a racheté XXX Inc., leader de ... pour 1 494 M € en 2000

<concept name="~Prise_de_participation">

~**BuyerCompany** / ANY / ~**AcquisitionVerb** / ANY / ~**TargetCompany** / ANY
/~**financial_operation** ...

</concept>

Recherche par facettes (catégories)



Chaque fois que l'on clique sur une facette, une nouvelle requête plus précise est exécutée.

Expansion sémantique

candidat du parti socialiste pour la présidentielle en France

candidat	
General (Anglais)	
candidate (N)	1
campaign (V)	5
campaigned (A)	5
campaigning (IG)	5
Politique et Gouvernement (Anglais)	
campaigner (N)	1
nominee (N)	1
potential presidential candidate (N)	5
running mate (N)	5
General (Français)	
colistière (N)	5
colistier (N)	5
Politique et Gouvernement (Français)	
présidentiable (N)	5
parti	
General (Anglais)	
party (N)	1
Politique et Gouvernement (Anglais)	
political party (N)	1
socialiste	
Sciences humaines (Anglais)	
socialist (A)	1
socialism (N)	5
Politique et Gouvernement (Anglais)	
socialist (N)	5

Pour le mot « candidat », ce dictionnaire distingue 2 domaines d'usage (en gras) et différents synonymes ou termes voisins, qui peuvent être des noms (N), adjectifs (A) ou des verbes (V).

La distance affichée dépend du nombre et du type de relation sémantique entre le terme recherché et le terme voisin. Ainsi entre le nom « vote » et le verbe « voter » le système attribue une distance de 5.



Thèmes

- [socialist](#) (5)
- [French](#) (4)
- [youth](#) (4)
- [premier tour](#) (4)
- [professional integration](#) (3)

>>

Personnes

- [Nicolas Sarkozy](#) (19)
- [Ségolène Royal](#) (15)
- [Dominique de Villepin](#) (13)
- [François Hollande](#) (13)
- [Jacques Chirac](#) (11)

>>

Organisations

- [TF1](#) (10)
- [Socialist Party](#) (9)
- [UMP](#) (8)
- [Agence France-Presse](#) (4)
- [Popular Movement Party](#) (4)

>>

Lieux

- [France](#) (29)
- [Paris](#) (21)
- [New York](#) (11)
- [Poitou-Charentes](#) (10)

> **lingway** | Retour - Graphe - Sources - Alertes

Filtre : sans - Sources: 2 / 2 - Documents du mois

Français

>chercher

[options](#)

Tri par : **score** [date](#) [format](#) [langue](#) [titre](#) [groupe](#)

[1] [2] [3] [4]

32 documents trouvés

 [Is France Ready to Elect a Woman?](#) New York Times 08/

 <http://www.nytimes.com/2006/04/07/world/europe/07france.html?ei=5087%0A&en=0a5bb3dcc22c7a87&ex=11445552>

beautiful image ... lack of experience ... recent months

socialist, socialists, french, presidential, party, candidate, presides, presidency, political, campaigning, potential, president, france

 [Is France Ready to Elect a Woman?](#) New York Times 09/

 <http://www.nytimes.com/2006/04/07/world/europe/07france.html?ei=5087%0A&en=1dee0e5cab258333&ex=1144641>

beautiful image ... lack of experience ... recent months

socialist, socialists, french, presidential, party, candidate, presides, presidency, political, campaigning, potential, president, france

 [Le meeting du MJS se transforme en lancement de la campagne présidentielle...](#) Le Monde 24/

 http://www.lemonde.fr/web/imprimer_element/0,40-0@2-3224,50-764510,0.html

Is **France** Ready to Elect a Woman? - [New York Times](#)

April 7, 2006

Is **France** Ready to Elect a Woman?

PARIS, April 6 -- Can this woman save **France**?

Can **Ségolène Royal**, the **politician** with the elegant profile and stratospheric poll ratings, lead the **Socialists** to victory in next year's **presidential** election?

In the confusion that has gripped **France** in recent months, with immigrant youth riots followed by huge protests turned violent, Ms. Royal, 52, is the only **politician** who looks good. On Thursday, she graced the cover of four French magazines. "The Mystery Royal," announced *Le Point*, while *Le Nouvel Observateur* explored "What Is in Her Head?" *VSD*, which covers entertainment and news, asked, "**President Ségolène**: Is She Ready?"

"For the first time, the French say they are ready to vote for a woman; this is a historic event," she told [Paris Match](#) in its cover story that proclaimed, "The Irresistible Ascension."

The media's interest is not accidental. Voters are disillusioned with President Jacques Chirac, who has held office since 1995, and less than enthusiastic about the gray-haired white men who have long run the opposition Socialist Party.

With the government in disarray over protests against a youth jobs law and the Socialists doing little more than scoring points, Ms. Royal -- a member of Parliament, regional president and former minister -- has moved quickly to fill the vacuum.

She is the most popular potential **Socialist presidential** candidate by far in poll after poll. She even edged past Interior Minister **Nicolas Sarkozy**, the front-runner on the right, in a hypothetical runoff for the **presidency** in two recent polls. She calls the jobs law, intended to encourage employers to hire young workers by making it easier to fire them, "a scandal" and "a form of violence." Asked in an interview late last month what she would do differently if she were in Mr. **Chirac's** shoes, she exclaimed: "I would be intelligent! Between the revolt in the suburbs last fall and the youth in the streets today, what a beautiful image of **France** we are giving to the world!"

Projecting a beautiful image is something that Ms. Royal does well. In addition to the magazine covers, she was the featured guest on **TF1's** television news program on Thursday. The first chapter of "Desires for the Future," her new online book intended to open a dialogue with the French people, appeared Thursday on her new Web site.

Luxid

demo Luxid®. Elle comprend 49000 dépêches AFP annotées avec CI.
<http://193.28.233.81/Luxid7>

Je suis analyste Lufthansa, chargé de la surveillance de la compagnie ALITALIA qui se trouve depuis plusieurs années en difficulté financières et pour laquelle une compagnie concurrente, Air France a fait une offre de rachat au 1er trimestre 2008 puis a retiré son offre le 21 avril 2008 suite à l'échec des négociations avec les syndicats de la compagnie.

Fin avril, les élections italiennes ont vu La victoire de Berlusconi qui soutient d'autres pistes de reprise (AEROFLOT, et surtout par un conglomérat italien CAI, Compagnie aérienne italienne (CAI) et dirigée par Roberto Colaninno

Pour Lufthansa, le suivi des évènements est vital, sur un plan politique, social et économique. En août 2008, après une modification de la loi sur les faillites, la compagnie est placée sous la tutelle de l'État, avec la nomination d'un « commissaire extraordinaire » (administration judiciaire) pour la diriger en la personne d'Augusto Fantozzi.

- Que s'est il passé concernant AIR France et ALITALIA durant la semaine du 11 au 20 septembre ?
- Recherche des informations relatives à la stratégie des deux compagnies et à la faillite d'Alitalia sur cette période

Rechercher web...

Sites suggérés Hotmail Plus de compléments...

Navigation sémantique

X-Match

Recherche rapide...

Bienvenue Igrivel (AFP) | Profil | Déconnexion | Aide

Général

Tableaux de bord

Recherche

Découvertes

Navigation sémantique

X-Match

Recherche rapide...

Navigation sémantique dans

Search containing Air France-KLM , containing Alitalia , 11 Sep 08 <= Date <= 20 Sep 08 (45 docs.)

Descripteurs

PLAT ARBRE

Entity

Company

Nombre d'éléments trop élevé pour affichage

Chercher

Afficher aperçu

Extension

Function

Location

Media

Organisation

Person

StockIndex

Relationships

PLAT ARBRE

Relationship

Asset

Board and Management Cha

Board Function

Management Change

Business Development

Corporate

Court Case

Financial Accounting and Pro

Human Resources

Marketshare and Products

Restructuring Business

Other descripteurs

PLAT ARBRE

Entity

Company

Nombre d'éléments trop élevé pour affichage

Chercher

Afficher aperçu

Extension

Function

Location

Media

Organisation

Person

StockIndex

Date

11/09/2008

Après

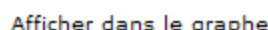
11/09 13/09 15/09 17/09 19/09

Avant

20/09/2008

Masque
context

Grands bouleversements en cours dans le ciel européenMariage, privatisation, disparition: la recomposition du ciel européen s'est accélérée cette semaine, la récente flambée de l'or noir et le ralentissement économique mondial sonnant l'hallali des plus faibles, telle Alitalia accrue à la faillite. L'allemande Lufthansa a ouvert le bal lundi en jetant son dévolu sur 45% du capital de la belge Brussels Airlines, héritière de Sabena, pour 65 millions d'euros, se réservant une option d'achat sur les 55% restants à partir de 2011. Mercredi, la Grèce présentait un plan de privatisation de son transporteur national, Olympic Airlines, le dernier de l'Union européenne à être 100% public.Des appels d'offres internationaux doivent être



Recherche sémantique

- Quelle est la situation financière d'ALITALIA?
- Que vont faire les concurrents, Berlusconi ?
- Quelle est la position des syndicats ?
- ...

Recherche d'information

« les pratiques de recherche d'information sont caractérisées par le passage d'une stratégie à une autre au sein d'une seule session de recherche »

Prendre en compte le caractère interactif de la recherche d'information. : la boucle de pertinence (*relevance feedback*) : retour des utilisateurs de l'outil de recherche sur la pertinence des réponses
(Exemple : annotation manuelle des réponses en vue d'une amélioration de cette liste.

Les pratiques de recherche et le besoin d'information de l'utilisateur doivent être analysés en fonction de ses besoins personnels (physiologiques, affectifs, cognitifs), son rôle social (travail, performance) et son environnement (de travail, socio-culturel, politico-économique et physique).

Environnement : l'espace dans lequel les acteurs socio-économiques évoluent Exemple- le cadre concurrentiel, le réseau des interdépendances ;

la dimension cognitive : c'est l'espace dans lequel le lecteur analyse, assimile, critique et finalement dégage le sens des discours tenus dans les collections de documents, bref, construit une nouvelle représentation de l'information : apprendre, sélectionner, choisir, décider :



TRAITÉ DES SCIENCES ET TECHNIQUES DE L'INFORMATION,

série *Environnements et services numériques d'information*, dirigée par Fabrice Papy

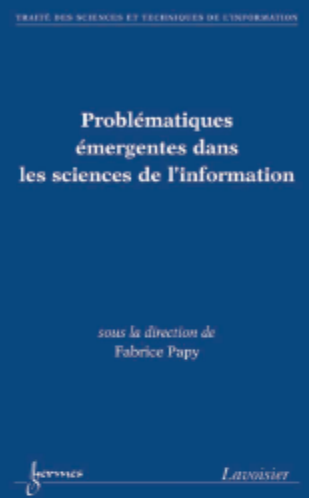
Problématiques émergentes dans les sciences de l'information

sous la direction de Fabrice PAPY

Malgré le discours politique et économique rôdé, qui accompagne la profusion technoinformationnelle d'où émerge la Société de l'Information, des situations inédites apparaissent, des interrogations se font jour, des complexités se révèlent, tant au niveau des individus censés maintenir un niveau d'activité satisfaisant dans cette société en mutation, qu'à celui de la macro-structure sociétale qui doit (re)trouver une nouvelle stabilité.

Au regard des problématiques identifiées par de nombreux chercheurs, relatives à l'archivage de données numériques structurées ou non, l'accessibilité et le partage, la propriété intellectuelle, le document numérique, la Recherche d'Information, les modalités d'interaction avec les systèmes techniques, les compétences numériques et informationnelles, la qualification et la pertinence de l'information, les profils informationnels, etc., les planifications politiques de la Sociétés de l'Information ne peuvent que susciter circonspection voire embarras d'un point de vue scientifique.

Cet ouvrage collectif, qui rassemble treize chapitres originaux de chercheurs en Sciences de l'Information et de la Communication, témoigne du caractère sensible de quelques problématiques scientifiques, présentées dans ce volume, qui sont pour certaines d'entre elles, anciennes, et pour d'autres, plus récemment formulées, et que la nouvelle ère numérique a mis en lumière.



Sommaire

Préface -F. PAPY

1. Quelques révisions du concept d'information -S. LELEU-MERVIEL,
P. USEILLE

2. De la communication scientifique à la médiation spécialisée :
communication des savoirs et formes d'hybridations -V. COUZINET

3. Temps, mémoire et document -J.-P. METZGER

4. Le document et le réseau -S. PÈNE

5. La culture informationnelle -A. SERRES

6. La pertinence en sciences de l'information : des modèles,
une théorie ? -B. SIMONNOT

7. L'étude des dispositifs d'accès à l'information électronique :
approches croisées -M. IHADJADENE, S. CHAUDIRON

8. L'information pertinente en intelligence économique -A. DAVID

9. Représentations des documents numériques dans le contexte
de l'analyse automatique de textes : une problématique de recherche
pluridisciplinaire -L. GRIVEL

10. Liberté d'information, démocratie participative et économie locale -
Y. THÉPAUT

11. L'internet muséal comme observatoire de l'interactivité -G. VIDAL

12. Internet à la bibliothèque publique -M. AMAR, A. CAMUS-VIGUÉ,
C. EVANS, F. GAUDET

13. Édition électronique pour mobile : les contenus originaux
de la télévision sur mobile, conditions techniques et socio-économiques
de leur émergence et culture d'écran de la génération Y -B. HÉNOUCQUE

« Technologies de la connaissance et recherche d'information »

Cet ouvrage vise à :

- dresser un panorama des principales avancées dans les systèmes de recherche d'information
- déterminer les impacts de l'introduction des technologies de la connaissance dans les systèmes de recherche d'information sur les usages et sur les pratiques de recherche d'information dans quelques secteurs clés de l'économie.

Il paraîtra chez **fin 2010**



14, rue de Provigny 94236 Cachan cedex, France • tél.: +33 (0)1 47 40 67 00 • fax: +33 (0)1 47 40 67 02 • e-mail : info@Lavoisier.fr

Lavoisier



Questions-Réponses

Luc Grivel

**Maître de Conférences
Université de Paris 1
(Panthéon-Sorbonne)**

luc.grivel@univ-paris1.fr

Merci de votre attention