



**HAL**  
open science

## Le moteur de recherche d'entreprise : quels enjeux organisationnels et technologiques ?

Jean-Yves Vuillequez

### ► To cite this version:

Jean-Yves Vuillequez. Le moteur de recherche d'entreprise : quels enjeux organisationnels et technologiques ?. domain\_shs.info.docu. 2013. mem\_00945629

**HAL Id: mem\_00945629**

**[https://memic.ccsd.cnrs.fr/mem\\_00945629](https://memic.ccsd.cnrs.fr/mem_00945629)**

Submitted on 12 Feb 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CONSERVATOIRE NATIONAL DES ARTS ET METIERS

Ecole Management et Société-Département CITS

INTD

MEMOIRE pour obtenir le  
Titre professionnel "Chef de projet en ingénierie documentaire" INTD  
RNCP niveau I

Présenté et soutenu par

*Jean-Yves Vuillequez*

le 14 novembre 2013

Le moteur de recherche d'entreprise : quels enjeux  
organisationnels et technologiques ?

**Jury :**

Maryse CARMES, Directrice de mémoire, CNAM-INTD  
Véronique PETER, Co-jury, DOCAPOST

**Promotion 43**

*À mon père*

# Remerciements

Je remercie tout particulièrement la responsable du pôle documentation de Coe-Rexecode, Murielle Prevost, pour sa disponibilité, ses conseils et la confiance qu'elle m'a témoigné dans la conduite de ma mission. Je tiens également à remercier Fabienne Besson-Lhoste et Sylvie Foutrier ainsi que tous les collaborateurs de Coe-Rexecode qui m'ont permis, par leur disponibilité et leur intérêt pour mon travail, de mener la mission m'a été confiée dans des conditions idéales.

Je remercie l'équipe pédagogique de l'INTD ainsi que les élèves de la promotion 2012/2013.

Je tiens également à témoigner ma gratitude à l'association AVARAP et plus spécialement à Simone, Béatrice, Carole, Grégoire, Nicolas, Christine, Caroline, Alain, Guillaume et Jamila pour leur bienveillance et leurs conseils avisés.

Un grand merci à ma mère sans qui rien n'aurait été possible, pour sa confiance et son indéfectible soutien.

Enfin, mes plus vifs remerciements vont à Maryse Carnes qui a su, par ses conseils de lecture et ses conseils méthodologiques, nourrir ma réflexion autour du thème passionnant de ce travail de fin d'études.

# Notice

VUILLEQUEZ Jean-Yves. Le moteur de recherche d'entreprise : quels enjeux organisationnels et technologiques ?. 2013. 125 p. Mémoire professionnel INTD, Titre I, Ingénierie documentaire. CNAM - INTD, 2013

Face à l'accroissement du volume de données au sein des organisations et dans un contexte de surcharge informationnelle accru, la recherche efficace d'information devient un facteur clé dans la maîtrise et la valorisation des flux de données en entreprise. Ce mémoire propose une réflexion sur les enjeux et les usages des moteurs de recherche d'entreprise souvent présentés comme des plateformes d'accès transversal à une information hétérogène, pléthorique et très inégalement structurée. Après avoir cerné dans un premier temps les spécificités de l'information professionnelle, ce mémoire décrit dans un second temps les principes fonctionnels d'un moteur de recherche d'entreprise. La troisième partie s'attache à montrer en quoi le moteur de recherche peut être considéré comme un projet managérial en explorant deux facettes complémentaires : l'aspect organisationnel et l'aspect technologique. La quatrième partie s'appuie sur une étude de cas dans le cadre du déploiement de la solution de gestion de contenu d'entreprise Sharepoint Server 2010 au sein de l'institut d'analyse macro-économique et de veille conjoncturelle, Coe-Rexecode.

accès à l'information ; capital immatériel ; catégorisation automatique ; classification automatique ; indexation automatique ; information non structurée ; information professionnelle ; information structurée ; maîtrise de l'information ; management de l'information ; moteur de recherche ; moteur de recherche d'entreprise ; recherche d'information ; recherche fédérée ; stratégie de recherche ; traitement linguistique

*In order to cope with the increase in volume of data within organizations and in a context of information overload, efficient information retrieval is becoming a key factor in the management of information data flow. This paper discusses the issues of Enterprise search engines, which are often presented as unified platforms in order to access heterogeneous, oversized and unevenly structured information. After initially identifying the specificities of professional information, this paper describes the functional principles of Enterprise search engines. The third part of this work attempts to show how the search engine can be regarded as a managerial and technological project. The fourth part is based on a case study concerning the deployment of a content management system, Sharepoint Server 2010, within the macro-economic analysis institute, Coe-Rexecode.*

# Table des matières

Remerciements .....	3
Notice.....	4
Table des matières.....	5
Liste des tableaux .....	9
Liste des figures.....	10
Introduction.....	11
<b>Première partie - Un moteur de recherche d'entreprise : Au service de l'information professionnelle.....</b>	<b>16</b>
1 La recherche d'information en entreprise : enjeux et usages .....	17
1.1 Pourquoi mettre en place une application de recherche d'entreprise ?.....	17
1.2 Spécificités de l'information en entreprise .....	18
1.2.1 Ses principales caractéristiques.....	18
1.2.2 Les contenus structurés et les contenus non-structurés.....	18
1.2.3 La part massive du « non-structuré ».....	19
1.3 L'information comme actif du capital immatériel .....	20
1.3.1 La maîtrise de ce capital : des enjeux pour les organisations .....	21
1.4 La réduction et la maîtrise de la surcharge informationnelle .....	23
1.4.1 La notion « d'urgence informationnelle ».....	23
1.4.2 La notion de « brouillard informationnel » .....	24
2 Les stratégies de recherche de l'information en entreprise .....	25
2.1 Une recherche d'information bicéphale.....	25
2.1.1 Répondre aux besoins des « Knowledge workers » .....	25
2.1.2 Répondre aux objectifs stratégiques de l'organisation .....	26
2.2 Les modes de recherche d'informations .....	27
2.2.1 Une question de vocabulaire : recherche et recherche de l'information .....	27
2.2.2 La recherche d'informations : une question de pertinence .....	27
2.2.3 De la simple consultation à la découverte d'information .....	30
<b>Deuxième partie - Les fonctionnalités du moteur de recherche d'entreprise .....</b>	<b>31</b>
3 Les différences fonctionnelles entre un moteur d'entreprise et un moteur web .....	32
3.1 Un critère de différenciation marqué : l'exhaustivité.....	32
3.2 Deux modes de mise en valeur des résultats .....	33
3.3 Périmètre d'interrogation et qualité des sources .....	34
3.4 Gestion des droits d'accès .....	35
3.5 Pourquoi une logique « à la Google » ne peut pas être appliquée en entreprise ?..	35

4	Le fonctionnement du moteur de recherche d'entreprise.....	37
4.1	Le cœur fonctionnel du moteur .....	37
4.2	Processus global : de l'acquisition à la présentation des résultats .....	38
4.2.1	Acquisition des contenus .....	38
4.2.2	Indexation des contenus .....	39
4.2.3	Prétraitements linguistiques et statistiques .....	40
4.3	Traitement des requêtes utilisateurs .....	42
4.3.1	L'apport des traitements linguistiques .....	42
4.4	La performance du moteur : un dialogue permanent avec les utilisateurs.....	43
4.4.1	Classification et catégorisation : modes de représentation de l'information .....	43
4.4.2	L'utilisation de facettes et de filtres : usages et limites .....	44
4.4.3	La gestion des métadonnées : usages et limites .....	44
<b>Troisième partie - Le moteur de recherche comme projet d'entreprise : enjeux organisationnels, acteurs du marché et choix de la solution .....</b>		<b>46</b>
5	Un projet managérial .....	47
5.1	Un outil de gouvernance de l'information .....	47
5.2	Opportunités et risques pour les entreprises.....	48
5.2.1	Une évaluation efficace du capital informationnel.....	48
5.2.2	L'impact sur la visibilité des informations.....	49
5.2.3	Un retour sur investissement difficilement quantifiable.....	50
6	Les éditeurs de solutions.....	52
6.1	Caractéristiques du marché .....	52
6.1.1	Un mouvement de concentration et de consolidation .....	52
6.1.2	Classification des éditeurs de solutions.....	52
6.2	Stratégies et limites de la communication des éditeurs.....	53
6.2.1	La cartographie des connaissances et des expertises .....	54
6.2.2	Les pièges de la recherche fédérée .....	55
6.3	Les positionnements innovants .....	56
6.3.1	Les plateformes d'accès unifié à l'information .....	56
6.3.2	Les Search Base Application ou Applications orientées recherche .....	57
7	Choix de la solution .....	58
7.1	Les critères de vigilance pour l'organisation.....	58
7.1.1	L'étude des besoins utilisateurs .....	58
7.1.2	L'audit documentaire .....	59
7.1.3	Un moteur spécifique pour un type projet spécifique.....	59
7.2	Une équipe projet dédiée .....	61
7.2.1	Pour mettre en place la solution .....	61
7.2.2	Pour maintenir les performances du moteur dans le temps.....	61

<b>Quatrième partie - Le déploiement de Sharepoint 2010 et du moteur Sharepoint Search Server chez Coe-Rexecode.....</b>	<b>63</b>
8 Contexte, périmètre et objectifs de la mission .....	64
8.1 Coe-Rexecode, un institut d'analyse macro-économique .....	64
8.1.1 La problématique informationnelle chez Coe-Rexecode .....	65
8.2 Un système d'information segmenté .....	66
8.2.1 Des besoins d'informations différenciés .....	66
8.3 Le déploiement de Sharepoint 2010.....	67
8.3.1 Une réponse à des besoins métiers spécifiques.....	67
8.4 Un bilan en demi teinte .....	67
8.5 Une mission menée en mode projet.....	68
8.5.1 Des outils de suivi et de capitalisation .....	68
8.5.2 Les livrables attendus et leurs objectifs .....	69
9 Recueil des besoins utilisateurs .....	70
9.1 Méthodologie de l'analyse des besoins .....	70
9.2 Résultats des entretiens utilisateurs .....	71
9.2.1 Des tendances fortes et marquées .....	72
9.2.2 La perception des fonctionnalités de recherche .....	73
9.3 Analyse des risques et opportunités .....	75
9.3.1 Utilisation de la matrice SWOT .....	75
9.3.2 Choix de la stratégie à mettre en œuvre .....	77
10 Benchmark de solutions alternatives .....	78
10.1 Descriptif des solutions retenues .....	78
10.1.1 Gains et limites des solutions retenues .....	79
10.2 Comparatif des versions de Sharepoint 2010 .....	81
10.3 Bilan du comparatif.....	83
11 Optimisation du moteur de recherche .....	84
11.1 Spécificités fonctionnelles.....	84
11.1.1 Modes d'interrogation de filtres et présentation des résultats .....	85
11.1.2 Limites et écueils du moteur .....	87
11.2 Actions correctives menées .....	87
11.2.1 Gestion des « meilleurs résultats ».....	87
11.2.2 Actions complémentaires.....	90
<b>Conclusion.....</b>	<b>91</b>
Sigles .....	94
Glossaire .....	96
Bibliographie.....	99
Annexes .....	115

Annexe 1 : Organigramme de Coe-Rexecode .....	116
Annexe 2 : Macro-planning et phasage .....	117
Annexe 3 : Guide d'entretien.....	118
Annexe 4 : Listing des « meilleurs résultats » par mots clés.....	121
Annexe 5 : Extrait du référentiel de mots clés pour les « meilleurs résultats ».....	122

## Liste des tableaux

Tableau 1 : Modes de recherche et de découverte [51, Ertzscheid, Gallezot] .....	27
Tableau 2 : modes de recherche et de découverte en entreprise [49, Russell-Rose] .....	30
Tableau 3 : Bénéfices tangibles et intangibles d'un moteur de recherche d'entreprise.....	50
Tableau 4 : Listing des livrables attendus .....	69
Tableau 5 : Perception des fonctionnalités de recherche par pôles .....	74
Tableau 6 : Analyse SWOT du projet de déploiement Sharepoint .....	76
Tableau 7 : Solutions par typologies et domaines de spécialité.....	78
Tableau 8 : Gains et limites des solutions retenues .....	80
Tableau 9 : Fonctionnalités de recherche des différentes versions de Sharepoint 2010 .....	82
Tableau 10 : métadonnées exploitées par Sharepoint Server Search .....	85

## Liste des figures

Figure 1 : Processus de recherche d'information .....	28
Figure 2 : L'architecture fonctionnelle d'un moteur de recherche [36, Balmissé].....	37
Figure 3 : Les étapes de l'indexation automatique pilotées par TAL .....	40
Figure 4 : Pipeline de traitement des requêtes [36, Balmissé] .....	42
Figure 5 : Les deux modèles de la recherche fédérée [42, Search Technologies].....	55
Figure 6 : Infrastructure de l'accès unifié à l'information [39, Freyermuth].....	56
Figure 7 : Typologie des moteurs de recherche [39, Freyermuth].....	59

# **Introduction**

## **La gestion de l'information : un défi constant pour les organisations**

Un des défis actuels pour les organisations, quelle que soit leur taille, est d'exploiter de manière optimale la valeur de leurs contenus informationnels, facteurs différenciants en termes de performance individuelle et collective [20, Hugonie]. Ces informations, qu'elles soient structurées (dans les bases de données) ou non structurées (dans les documents d'activité), sont transversales. C'est la capacité d'action par rapport à ce flux qui contribue en permanence à l'alignement de l'entreprise avec son environnement

La recherche performante et maîtrisée d'informations en interne, mais aussi l'échange avec les partenaires externes sont devenus des enjeux majeurs. Les organisations s'articulent dorénavant en réseaux plutôt qu'en pyramides de pouvoirs et en systèmes informationnels plutôt qu'en silos aux frontières bien délimitées. Comme le souligne le rapport du Cigref en 2006, la création de valeur par le système d'information oblige à (re)penser l'information en adoptant une vision englobante pour en optimiser la circulation par la mise en place de dispositifs dédiés et transverses [03, Cigref].

## **L'accroissement du volume des données et l'apport des TIC**

Pour un utilisateur, il est toujours plus simple de créer de l'information mais toujours plus difficile de la traiter et de la rechercher. La multiplication des sources (structurées et non structurées) et des canaux de diffusion ainsi que la forte augmentation des volumes d'information et de données, ou encore l'intensification des rythmes d'échange conduisent les utilisateurs à consacrer une part importante de leur temps à la «manutention de l'information» [61, Vacher].

La généralisation et l'appropriation de l'informatique par la quasi-totalité des métiers a radicalement transformé le rapport à l'information, celle-ci est devenue à la fois surabondante (on parle «d'infobésité»<sup>1</sup>) et omniprésente grâce à la démultiplication des supports de diffusion et de stockage. En 2010, un changement de cycle (et probablement d'époque) s'est opéré. En effet, le volume de données de l'écosystème numérique a changé d'échelle avec un basculement d'unité: le zettaoctet (ZO). Celui-ci passait de 0,8 à 1,2 Zo soit un 1,2 milliard de giga-octets [30, Lefassy].

Concrètement, dans le monde de l'entreprise, cela se traduit par une élévation des coûts liés à la gestion, à la production et à la conservation des données ainsi qu'une explosion des volumes de données et des contenus. La maîtrise des volumes d'informations devient LA préoccupation majeure des organisations car la conservation des données structurées et non

---

<sup>1</sup> Terme apparu au Québec il y a 6 ans et faisant référence à l'abondance d'information numérique. Forme contractée du terme « information » et du terme « obésité », traduction directe d'« information overload ».

structurées dans différentes solutions déployées au fil des développements internes empêche toute vue globale de l'information [38, Bennett].

Pour remédier à ces dysfonctionnements, la quasi-totalité des entreprises concernées ont investi dans un outil de GED<sup>2</sup> ou un ECM<sup>3</sup> lors des 5 dernières années. Mais en termes de trouvabilité de l'information, ces outils restent très limités et reposent sur un système par arborescence (dossiers, sous-dossiers). L'utilisateur doit en permanence anticiper sa recherche d'information et savoir à l'avance où elle se trouve (dans quel répertoire ? quel espace projet ?...)

Il est intéressant de noter que le regain d'intérêt concernant les applications de recherche d'entreprise, amorcé dès 2011, passe par une prise de conscience face à l'accroissement frénétique du volume de données et à l'incapacité des applications de type GED ou ECM de donner du sens et de la pertinence à ce magma informationnel.

Lors de la conférence Enterprise Search Europe 2012, les raisons invoquées pour ce renouveau sont dans l'ordre : le Big Data, la recherche sur terminaux mobiles et le Cloud Computing

### **Le moteur de recherche d'entreprise : une approche singulière**

La promesse de tout moteur de recherche d'entreprise et ce, tous éditeurs confondus, est de retrouver rapidement une information depuis un point d'accès unique, sans savoir à l'avance où elle se trouve, dans quelle base, dans quel applicatif ou sur quel lecteur réseau.

Le moteur de recherche d'entreprise propose une approche stratégique de la gestion de l'information, pilotée par les usages et non par l'organisation ou les outils. Cette approche englobante de l'information est particulièrement visible dans les évolutions récentes des moteurs : Les Search Base Applications (les « applications orientées recherche ») ont pour objectif de permettre une amélioration de la productivité des utilisateurs en agrégeant dans un espace fédéré, contenus pertinents, outils décisionnels performants et pratiques métiers/usages.

L'optimisation des processus de recherche d'information et la gestion intelligente des flux informationnels sont deux leviers sur lesquels devrait s'appuyer le management de l'entreprise.

L'implémentation d'un moteur de recherche d'entreprise implique d'importants changements humains, organisationnels et technologiques. Il influence les pratiques métiers par une

---

<sup>2</sup> Gestion Electronique de Documents

<sup>3</sup> Enterprise Content Management

remise en cause des modes de management de l'information et par le passage d'un mode de fonctionnement cloisonné à un mode de fonctionnement plus transversal. Ces impacts et ces changements doivent être accompagnés pour palier à des dysfonctionnements majeurs : une perte de repères, une résistance au changement et un attrait du nouveau qui peut être déçu si la solution mise en œuvre n'est pas à la hauteur des espérances. En outre, l'accès facilité à l'information et le partage de connaissances peuvent être perçus comme une source de remise en cause des positionnements dans l'organisation, la détention de l'information étant classiquement considérée comme un attribut du pouvoir.

### **Une application de recherche, gage de succès pour trouver l'information pertinente ?**

Comme le souligne Martin White<sup>4</sup>, de nombreuses organisations sont déçues par les performances de leur solution de recherche d'entreprise. Selon l'étude 2011 du cabinet NetStrategy/JMC [46, Intranet Focus] sur les tendances du «Digital Workplace»<sup>5</sup>, 45% des répondants étaient satisfaits de la performance de l'application de recherche, contre 50% en 2008.

L'installation d'un moteur de recherche est beaucoup plus facile à appréhender que n'importe quel autre logiciel d'entreprise (ECM, GED) car il n'est pas nécessaire (en théorie) de former les utilisateurs à son mode de fonctionnement. Tout ce qu'ils ont à faire est d'entrer un terme de requête et le moteur renvoie presque immédiatement des résultats pertinents (toujours en théorie).

Cette idée est largement reprise par les éditeurs qui packagent leurs produits comme des solutions Plug&Play (on branche et ça fonctionne). Mais qu'en est-il réellement ?

### **Le moteur de recherche d'entreprise en question(s)**

La mise en place d'un moteur de recherche d'entreprise n'est pas un projet anodin et répond à des impératifs d'efficacité afin de contribuer au mieux aux exigences de l'organisation, du système d'information et de ses utilisateurs. La gestion d'un projet de déploiement d'un moteur de recherche d'entreprise (cela vaut tout autant pour une GED ou un ECM) où la question technique est pensée d'abord par les outils, puis par la valeur d'usage est-elle vraiment efficace ? Les réflexions technologiques ou limitées aux R.O.I. ne doivent-elles pas être dépassées pour considérer le moteur de recherche dans sa globalité et ainsi saisir les opportunités à tous les niveaux de l'organisation ? Quelle est l'étendue de la création de

---

<sup>4</sup> Fondateur et directeur général de Intranet focus Ltd et consultant en stratégie de gestion de l'information et spécialiste des intranets. <http://www.intranetfocus.com/>

<sup>5</sup> Le concept de digital workplace (espace de travail numérique) est habituellement attribué à Jeffery Bier, qui fonda Instinctive Technologies en 1996 et développa le projet collaboratif eRoom.

valeur et de la satisfaction client ? Comment agir face à l'explosion du volume d'informations numériques et comment valoriser les contenus d'entreprise ?

Quelles sont également les opportunités pour les professionnels et gestionnaires de l'information ? D'un point de vue management, permet-il aux professionnels de l'information d'avoir un rôle de conseil sur les phases amont de développement ? D'un point de vue opérationnel, donne-t-il les moyens de mieux structurer et préparer l'information afin de maintenir la pertinence des réponses fournies ? Enfin, permet-il d'anticiper les besoins des utilisateurs en s'associant aux pratiques métiers et aux outils de ceux-ci ?

### **Structure générale du mémoire**

La première partie définit les principaux termes du sujet. Elle décrit les caractéristiques et les enjeux de l'information en entreprise en abordant certaines de ses spécificités et composantes telles que l'information structurée et non structurée ou la maîtrise de la surcharge informationnelle. Elle permet d'appréhender la problématique de la recherche d'informations en entreprise à travers les différentes stratégies répondant aux besoins spécifiques des organisations et de leurs employés.

La deuxième partie décrit les principes fonctionnels d'un moteur de recherche d'entreprise et ses particularités face aux moteurs de recherche web, tel que Google. Elle permet d'appréhender les technologies mises en œuvre ainsi que les outils statistiques et linguistiques permettant de structurer l'information grâce à un moteur de recherche en vue de sa consultation par les utilisateurs.

La troisième partie s'attache à montrer en quoi l'implémentation d'une application de recherche influe sur la visibilité des informations et sur ses modes de production, de diffusion, de capitalisation et de management. Elle dresse également un panorama des éditeurs de solutions grâce à un état des lieux en matière d'offre et de positionnements innovants. Elle permet d'envisager la question sur ses deux versants : organisationnels et technologiques.

Enfin, la quatrième partie présente de façon concrète comment la question du moteur de recherche d'entreprise est abordée sur le terrain avec un projet de déploiement d'outil collaboratif et de gestion de contenu, Microsoft Sharepoint Server 2010. Cette partie illustre les enjeux liés à la gestion documentaire et à l'appropriation par les utilisateurs du nouveau système dans le contexte d'une structure de taille modeste : Coe-Rexecode, un institut d'analyse macro-économique.

Première partie  
**Un moteur de recherche  
d'entreprise :**  
Au service de l'information professionnelle

# 1 La recherche d'information en entreprise : enjeux et usages

---

Le volume croissant d'informations numériques disponibles, aussi bien sur internet que sur les intranets d'entreprises, ainsi que l'importante hétérogénéité des gisements d'information rend l'identification de l'information pertinente et son accès de plus en plus complexes.

Ces données peuvent être issues des systèmes structurés ou non structurés avec une granularité documentaire variable et dans une extrême multiplicité de formats d'enregistrement, de codage et de multilinguisme [10, Chaudiron].

## 1.1 Pourquoi mettre en place une application de recherche d'entreprise ?

Les systèmes d'exploitation, les outils de collaboration, les clients de messagerie et les solutions de gestion de contenus d'entreprise (ECM ou ERP) ont tous des capacités de recherche. Mais leurs possibilités d'interrogation sont extrêmement limitées et ne permettent de rechercher des informations que dans leur propre périmètre. [32, Owens].

Le moteur de recherche d'entreprise est un élément clé d'une stratégie de l'information. En effet, il unifie la recherche dans divers environnements d'informations et solutions informatiques déployées dans les organisations au gré des développements successifs. Son déploiement vise à éliminer les silos d'information et le risque que l'information importante ne soit pas trouvée ou utilisée [34, white].

Les raisons habituellement invoquées pour la mise en place d'une solution de recherche sont diverses et reflètent cette volonté d'exploitation stratégique des gisements d'information disponibles au sein des organisations. Les raisons majeures sont :

- Trouver des informations dans des sources hétérogènes structurées ou non structurées, dans des formats d'écriture différents et favoriser leur exploitation optimale.
- Permettre aux utilisateurs de trouver toutes les informations via un point d'accès unique sans avoir besoin de savoir où l'information est stockée.
- Mettre en valeur le capital immatériel de l'entreprise : favoriser et améliorer la réutilisation des informations, des connaissances et des expertises en interne.
- Résoudre les problèmes récurrents et critiques de la surcharge informationnelle et de l'accroissement du volume des données : favoriser la prise de décision sur la foi d'informations exhaustives et fiables

Pour autant, mettre en place une application de recherche ne garantit pas automatiquement l'exploitation systématique et optimale des gisements informationnels [48, Mindmetre]. Selon une enquête menée auprès de plus de 2.000 directeurs et responsables de systèmes d'informations aux États-Unis, au Royaume-Uni, en Allemagne et en France, la recherche d'entreprise reste bien en-deçà des bénéfices espérés. Plus de la moitié (52%) des répondants estiment qu'ils ne peuvent pas trouver l'information qu'ils recherchent en utilisant leur application de recherche d'entreprise. Cet aspect sera plus spécifiquement couvert en section [3.5](#).

## **1.2 Spécificités de l'information en entreprise**

### **1.2.1 Ses principales caractéristiques**

Les gisements informationnels au sein des organisations deviennent de plus en plus difficiles à exploiter et à valoriser. L'information y est pléthorique, évolutive, dispersée, multilingue et hétérogène. [23, Hérigault].

Le périmètre informationnel est vaste et ne cesse de croître. Il couvre des ressources internes (intranet, extranets, bases de données documentaires, qualité ou KM<sup>6</sup>, serveurs mails, serveurs de fichiers...) et des ressources externes (sites Web, bases de données gratuites et payantes, données envoyées par les clients...).

Ces informations se divisent en trois types : les informations structurées, non structurées et semi-structurées. Les deux derniers cas, qui constituent un des aspects central de ce mémoire, représentent un problème plus vaste et complexe appelé « la recherche d'informations » ou *Information Retrieval* en anglais. Cet aspect est traité plus spécifiquement en section [2.2.1](#).

### **1.2.2 Les contenus structurés et les contenus non-structurés**

Les informations structurées se trouvent dans les bases de données et les langages informatiques. Ces informations sont dites structurées parce qu'elles sont disposées de façon à être traitées automatiquement et efficacement par un agent informatique, et pas nécessairement par un humain. Elles ne représentent qu'une infime partie des informations disponibles au sein des systèmes d'information.

Les informations non structurées et semi-structurées se trouvent dans les documents d'activité (rapports, factures...), les mails, les wikis, les blogs. Une information est dite non structurée « quant elle renvoie à un usage non-inscrit dans un modèle et se construit à partir d'un sens dérivé du langage ou de la pensée » [17, Garnier, p.23]. Elles sont

---

<sup>6</sup> Knowledge management ou capitalisation des connaissances

directement compréhensibles et interprétables par un humain et se définissent donc par opposition à l'information dite structurée, interprétable et utilisable directement par un ordinateur à des fins de calcul et de traitement. Elles sont en sur-représentativité dans les organisations.

Cependant, comme le souligne Eric Debonne<sup>7</sup>, ces données ne sont pas sans structure. En fait, cette information comporte bien une structure : il y a généralement un titre, un auteur, une date, et éventuellement des titres de section et des tableaux. Les données de l'entreprise sont beaucoup plus structurées : on sait de quel site, service ou base de données vient le document et bien souvent le réceptacle de stockage donne une indication de son utilisation fonctionnelle. Une typologie de document peut souvent en être déduite : page Web, rapport, facture, document d'activité ...

De plus la date est très souvent visible et donne une indication pertinente. Certaines bases d'informations contiennent également des données structurées : localisation géographique, auteur, descripteurs et mots clés, langues. [42, Debonne]

Ce mix de données semi-structurées et non-structurées pourra être directement exploité par le moteur de recherche. Ce premier niveau de structuration repose sur des « étiquettes de données » ou métadonnées. L'importance de métadonnées dans la qualité des réponses fournies aux utilisateurs par le moteur de recherche est primordiale. (voir section [4.4.3](#))

### **1.2.3 La part massive du « non-structuré »**

L'information non-structurée manipulée par les entreprises croît de manière significative à un rythme soutenu [17, Le Foll] : selon Goldman Sachs, 90% des contenus à haute valeur pour l'entreprise sont non-structurés (alors que l'on parlait jusqu'à présent de 80%)<sup>8</sup>. Selon une étude complémentaire du cabinet Gartner, le volume de ces données double tous les mois [25, Autonomy].

En dehors du monde structuré des bases de données, la volumétrie de l'information produite et manipulée par les utilisateurs des systèmes d'informations est très élevée. Ces données sont évolutives, « en permanente métamorphose » et génératrices de versions transitoires [15, Cotte]. Le contenu est souvent du travail en cours, qui change rapidement et avec une valeur de court-terme uniquement pour son créateur ou un petit cercle de collaborateurs. Certains documents pourront revêtir un caractère confidentiel, en accès restreint. Ce point sera plus spécifiquement traité en section [3.4](#).

---

<sup>7</sup> Consultant et fondateur du cabinet de conseil Solaci spécialisé en veille, réseaux Sociaux d'Entreprise et accès à l'information

<sup>8</sup> Selon le consultant Martin White, aucune preuve n'a été présentée pour appuyer cette affirmation qui semble être uniquement fondée sur le principe du « 80-20 » de la [loi de Pareto](#)

Un des points essentiels à prendre en compte dans un contexte de déploiement d'une application de recherche, c'est l'hétérogénéité de l'information et des documents selon deux critères : leur granularité et leur niveau de structuration (absence ou présence d'une structuration). L'information peut avoir plusieurs niveaux de profondeur et de précision : une valeur dans une base de données, un mot-clé descripteur dans un formulaire de fiche descriptive dans une GED, un paragraphe dans une présentation Powerpoint ou un titre dans un document Word.

Cette extrême variabilité de granularité aura une influence sur le paramétrage du moteur. Son efficacité et la pertinence des réponses aux requêtes des utilisateurs dépendront de la part de structuré et semi-structuré d'un côté et de non structuré de l'autre.

### **1.3 L'information comme actif du capital immatériel**

Le capital immatériel est « un constituant de l'entreprise, identifiable séparément, qui participe aux opérations génératrices de rentabilité présente ou future, mais dont la valeur ne figure pas au bilan. » [04, Fustec]. Le capital informationnel est donc une composante de ces actifs immatériels à prendre en compte lors d'une démarche d'évaluation de la valeur informationnelle d'une entreprise. Il s'agit de la somme des informations détenues par l'entreprise sur :

- Son environnement externe, ses clients, son marché, ses concurrents, ses actionnaires.
- Son environnement interne, son personnel, ses savoir-faire, ses processus, ses méthodes de fabrication.

C'est deux environnements, indissociables, contribuent à l'efficacité, à la continuité de l'activité de l'entreprise et à l'efficacité de ses employés [20, Hugonie]. Alan Fustec<sup>9</sup> et Bernard Marois<sup>10</sup> définissent neufs actifs immatériels dont trois sont directement liés à l'information :

- La connaissance ou « capital savoir » produit par les salariés d'une organisation : données R&D, publication et supports de communication, brevets, formules, systèmes et données liées à la capitalisation des connaissances internes.
- Le système d'information comme fonction support fournissant les outils informatiques opérationnels afin de stocker cette connaissance et ces données.
- L'organisation ou « capital organisationnel » garant d'une bonne répartition des rôles et une bonne circulation de l'information.

---

<sup>9</sup> Alan Fustec est enseignant à HEC et président de l'Observatoire de l'Immatériel.

<sup>10</sup> Bernard Marois est diplômé d'HEC et actuellement professeur emeritus au Groupe HEC.

Même s'il existe dorénavant une préoccupation croissante exprimée, entre autre, par les dirigeants d'entreprise au sujet de l'ampleur des ressources d'information disponibles au sein des organisations (notamment avec l'émergence du concept de « Big Data »), la perception de la valeur des contenus informationnels (notamment non-structurées) reste grandement négligée [34, white].

Une étude menée par Marklogic<sup>11</sup> en 2011 auprès de 446 professionnels de l'information montre que la question de la valorisation reste un problème au sein de leurs organisations respectives. En effet, 40% des répondants affirment que leurs directions sont soit "partiellement" conscientes ou ignorent l'existence de gisements informationnels dans les données non-structurées qui résident dans les différents applicatifs de l'entreprise.

Ce manque de sensibilisation à la gestion informationnelle souligne la difficulté des gestionnaires de l'information à assurer le financement et l'affectation des ressources nécessaires pour tirer pleinement parti des grands volumes de données non-structurées, notamment avec la mise en place d'une application transversale comme un moteur de recherche d'entreprise.

### **1.3.1 La maîtrise de ce capital : des enjeux pour les organisations**

Il existe donc un contraste très marqué entre, d'une part, les données structurées en provenance des ERP<sup>12</sup> (les données RH financières et commerciales) ou des outils de Business Intelligence<sup>13</sup> (données stratégiques) qui sont généralement correctement valorisées et, d'autre part, les données non-structurées issues des outils bureautiques, des messageries électroniques, des applicatifs métier, mais aussi des documents manuscrits qui sont généralement peu ou pas valorisées.

Mais alors, les unes ont-elles plus de valeur que les autres ? Si l'on se place du point de vue du décideur, les données en provenance des ERP ont une structure connue, maîtrisée et maîtrisable. Tandis que les données non-structurées en provenance des outils bureautiques paraissent beaucoup plus anarchiques, faute d'outils adéquats afin d'exploiter pleinement leur valeur.

Cette gestion partielle ou défaillante des données non-structurées engendre une rupture dans la transmission ainsi que le partage de la connaissance, et accentue le sentiment d'incertitude autour des Technologies de l'information dans les organisations [03, Cigref].

---

<sup>11</sup> <http://www.marklogic.com/press-releases/marklogic-survey-reveals-unstructured-information-is-growing-rapidly-will-soon-surpass-relational-data/>

<sup>12</sup> «Enterprise Resource Planning» ou « progiciel de gestion intégré »

<sup>13</sup> Informatique décisionnelle à l'usage des décideurs et des dirigeants des entreprises.

Sur la problématique de la maîtrise de l'information, les enjeux sont doubles pour les organisations : Il sont à la fois opérationnels et stratégiques.

La mise en place opérationnelle d'un moteur de recherche, dans un premier temps, pose la question de la transformation de la richesse informationnelle inexploitée en connaissances utiles et, dans un deuxième temps, l'amélioration de la performance de l'outil de façon continue, tout en augmentant sa rentabilité. La question du maintien de la performance de l'outil par une équipe dédiée est abordée en section [7.2](#).

Les entreprises, « engagées dans une course à la compétitivité dans laquelle les leviers essentiels de différenciation sont l'innovation, l'agilité et la performance commerciale »<sup>14</sup>, sont confrontées à des enjeux stratégiques quand il s'agit de maîtriser les flux et les processus informationnels :

- En interne, c'est la question de l'innovation (développements de nouveaux produits, veille technologique et concurrentielle) qui est directement liée à la performance de la gestion des connaissances métier, à leur partage, leur réutilisation et leur enrichissement.
- En externe, c'est la question des processus métiers documentaires : la maîtrise des contenus émis vers les partenaires commerciaux et les clients en garantissant la sécurité et le contrôle des informations sortantes.

La maîtrise de l'information apparaît donc comme un avantage concurrentiel et durable dans un contexte globalisé. En effet, l'intelligence d'une organisation repose en grande partie sur l'identification, la collecte, l'analyse, la validation, la valorisation, le stockage et la recherche d'informations nécessaires à son fonctionnement et à son développement [02, Bourret].

Mais comme le fait remarquer Didier Lambert, ancien président du Cigref<sup>15</sup> et actuel président d'Essilor, ce n'est plus seulement « l'accès à l'information qui constitue le ressort essentiel de la compétitivité, mais bien l'aptitude de toutes les « parties prenantes » à la transformer, la comprendre, l'interpréter et à l'utiliser qui constitue le défi majeur pour toute Entreprise ».

Tirillés entre surcharge et déficit informationnel, les utilisateurs des systèmes d'information et les dirigeants de ces entreprises se voient contraints de développer « des stratégies individuelles empiriques de tri et de recherche. Ce qui ne fait que renforcer leur incertitude face à la prise de décision. » [18, Le Foll, p.7]

---

<sup>14</sup> Jenny de Montaigne (Analyste KM et gestion de contenus) dans « Valorisation de l'information non structurée ».Aproged, 2007.

<sup>15</sup> Réseau des DSI des grands groupes et grandes entreprises françaises. <http://www.cigref.fr/>

## **1.4 La réduction et la maîtrise de la surcharge informationnelle**

L'introduction des technologies de l'information et leur appropriation par les utilisateurs, aussi bien dans le cadre professionnel que dans le cadre privé, produit souvent des dysfonctionnements majeurs, au premier rang desquels figure la surcharge informationnelle. Ce problème récurrent est souvent perçu comme une situation d'urgence et d'incertitude pour les utilisateurs des systèmes d'information [60, Isaac]. La question n'est pas nouvelle mais elle prend une autre dimension compte tenu du volume croissant d'information numérique disponible et de son hétérogénéité.

Même si cette problématique déborde largement du cadre purement fonctionnel du moteur de recherche, la réduction de cette surcharge est souvent un des arguments mis en avant pour la mise en place d'une application de recherche transversale [32, Owens].

Dans ce contexte, les moteurs de recherche ne seraient « qu'un palliatif à une mauvaise gestion des documents dématérialisés dans une entreprise » [47, Jung] ou plus précisément un moyen de réduire l'impact de la surcharge informationnelle à travers ses 3 composantes :

- Une volumétrie d'information trop abondante à traiter par l'organisation et ses collaborateurs
- Une capacité cognitive limitée des individus à traiter cette volumétrie d'informations
- Une surcharge de communication (notamment à travers la sur-utilisation du mail)

### **1.4.1 La notion « d'urgence informationnelle »**

Certes, les TIC<sup>16</sup> ont permis de réduire les temps de traitement des processus opérationnels ainsi que les processus de décision en autorisant une disponibilité permanente de l'information pour les salariés où qu'ils se situent.

Mais cette disponibilité s'accompagne d'une saturation cognitive liée à la notion de maîtrise du temps et au fait que les TIC contribuent fortement à l'augmentation du temps de traitement de l'information au détriment d'activités métiers à plus forte valeur [60, Isaac].

Selon l'étude conduite par l'éditeur SmartLogic<sup>17</sup> en 2011, 1 utilisateur sur 2 considère qu'un bon moteur de recherche devrait permettre de retrouver en interne une information en moins de 2 minutes. Or seulement 41 % trouvent effectivement en moins de 2 minutes [48, Mindmetre].

---

<sup>16</sup> Technologies de l'information et de la communication

<sup>17</sup> <http://www.smartlogic.com/>

Il existe un paradoxe entre l'état d'urgence informationnelle plus ou moins ressenti par les utilisateurs et leur souhait de trouver l'information pertinente de plus en plus rapidement. Dans un contexte professionnel, ne pas être en mesure de trouver l'information pertinente rapidement peut difficilement être supportable voire frustrant.

En outre, l'urgence informationnelle peut venir d'une faible qualité des informations et des données ainsi qu'une complexité croissante des sources reçues amenant à se frayer un chemin vers une information qui n'a pas été repérée ni classée et qui est conservée dans un certain désordre. On parle de chaos informationnel ou « Content Chaos ». Pourtant, un des critères d'efficacité pour les organisations est de pouvoir offrir à ses employés les moyens de retrouver rapidement un élément pour vérifier ou infirmer des idées dans une masse d'information importante [59, Guyot].

### **1.4.2 La notion de « brouillard informationnel »<sup>18</sup>**

Avec des centaines d'applications utilisées chaque jour dans les entreprises, quelque soit leur taille, la quantité de données et d'informations générées est souvent peu ou mal mise en valeur. Pire encore, en raison du faible coût de stockage on assiste le plus souvent à une croissance informationnelle qui résulte de la combinaison de données anciennes et de données récentes [34, white]. Les DSI<sup>19</sup> parient sur la valeur de ces informations sans pouvoir réellement l'évaluer. Le défi pour les utilisateurs est de rechercher de l'information dans cette masse de données.

Selon une étude de l'AIIM<sup>20</sup> publiée en 2011, le temps moyen consacré par un employé à la recherche d'information est de 45,2 minutes en moyenne soit environ 8% d'une journée de travail [52, AIIM].

Comme le souligne Stéphane Chaudiron<sup>21</sup> : « La croissance de la quantité d'informations dans ses multiples sphères engendre un sentiment d'incertitude qui est devenu l'état cognitif normal ». L'accès à l'information est d'abord un processus d'élimination de l'information superflue dans un contexte de surinformation et de désorientation cognitive. Il consiste à rendre pertinente une information par rapport aux autres [10, Chaudiron].

Ce processus de discrimination a pour objectif de réduire l'écart entre un trop grand nombre de messages périphériques ou externes (mails, informations parasites, trop grand choix dans une liste de réponses) à un seul choix qui fera sens pour l'utilisateur.

---

<sup>18</sup> David Shenk dans « Data Smog : Surviving the Information Glut » en 1998

<sup>19</sup> Direction des Services d'Information

<sup>20</sup> The Global Community of Information Professionals (<http://www.aiim.org/>)

<sup>21</sup> Directeur du Laboratoire GERIICO - Université Lille III

## 2 Les stratégies de recherche de l'information en entreprise

---

Comme le soulignent Jean-Max Noyer<sup>22</sup> et Maryse Carmes<sup>23</sup>, la recherche « d'informations recouvre un ensemble de pratiques, d'usages extrêmement ouverts ». Ces pratiques sont liées à la nature de ce que l'on recherche, à l'existence supposée de ce que l'on cherche et aux raisons de cette recherche [18, Noyer, Carmes].

D'un point de vue pratique, les outils de recherche fournissent aux utilisateurs les moyens de rechercher, trouver et accéder aux informations dont ils ont besoin dans un ensemble de référentiels via un point d'accès unique et de manière fédérée [14, Antidot].

En interne tout d'abord : bases de données documentaires, données de Business Intelligence. Mais également en externe : bases de données payantes, données d'organisations tiers, sites internet. Le principal intérêt des applications de recherche consiste à ne présenter aux utilisateurs que les informations pertinentes ou perçues comme telles, tout en occultant la complexité des sources d'informations afin d'éviter le syndrome « Trop de réponses à votre demande - veuillez affiner votre recherche » [01, Atos Origin].

Malgré l'apparente simplicité de ces outils, et revendiquée comme telle par les éditeurs (voir section 6.2), les utilisateurs comme les organisations sont contraintes de mettre en place des stratégies de recherche appropriées pour faire face à l'accroissement du volume d'informations disponibles, à leur hétérogénéité ainsi qu'à leur manque de pertinence. Au delà du moteur de recherche c'est bien la problématique de la recherche d'information qui pose question.

### 2.1 Une recherche d'information bicéphale

#### 2.1.1 Répondre aux besoins des « Knowledge workers »

Le « knowledge worker » ou « travailleur du savoir » se caractérise par des objectifs tels que la maîtrise des sources d'information, la gestion et l'exploitation des connaissances [2, Bourret]. Son besoin va s'exprimer dans diverses situations. André Tricot<sup>24</sup> propose de les regrouper en 5 catégories, associées à des besoins fondamentaux [23, Hérigault] :

---

<sup>22</sup> Maître de conférences en Sciences de l'Information et de la Communication, Paris 7

<sup>23</sup> Maître de Conférences en Sciences de l'Information et de la Communication, Cnam-INTD

<sup>24</sup> Professeur d'université en psychologie à l'IUFM Midi-Pyrénées et auteur de « *Apprentissages et documents numériques* ». Belin, 2007

- **Amélioration** : besoin d'une connaissance que l'on n'a pas ; besoin de confirmer une connaissance que l'on a,
- **Complétude** : besoin d'une connaissance plus complète que celle que l'on a,
- **Conformité** : besoin d'être conforme aux buts, contraintes ou attentes de la situation,
- **Exactitude** : besoin d'information sur la forme de la connaissance à utiliser dans la situation,
- **Détection** : besoin de détecter un marqueur de pertinence dans la situation.

La recherche vise à trouver une réponse à une question précise, à découvrir de nouvelles connaissances, à réduire l'incertitude liée à un déficit ou à une surcharge d'informations, à palier à l'état de connaissance transitoire ou insatisfaisant, ou encore donner du sens.

### 2.1.2 Répondre aux objectifs stratégiques de l'organisation

Pour l'organisation, la mise en œuvre d'une stratégie de recherche répond à plusieurs objectifs organisationnels et opérationnels :

D'un point de vue global, c'est adopter une vision d'ensemble de la problématique de recherche d'information afin d'accroître la visibilité de cette dernière ainsi que des gains de productivité. C'est également initier un meilleur partage des connaissances garant d'une plus grande capacité d'innovation.

C'est aussi l'opportunité de prendre conscience du rôle moteur de l'information dans l'amélioration des performances collectives et individuelles de ses collaborateurs et de ses processus de décisions.

D'un point de vue opérationnel, c'est mettre en place les conditions nécessaires pour identifier plus clairement les besoins d'information des collaborateurs et leur offrir plus de pertinence : décloisonner l'information contenue dans les multiples applications, apporter plus de transversalité, valoriser l'information structurée ou non (voir section [6.3.1](#)).

L'adoption de cette stratégie permet également de dissocier d'un côté la recherche d'information et de l'autre l'accès à l'information [35, Balmisse] : Cette stratégie s'articule et s'agrège souvent autour d'autres projets d'entreprise liés (portails, intranets, amélioration des processus métiers ou documentaires).

D'un point de vue financier, l'enjeu se situe autour de rentabilité de l'entreprise : réduire les coûts du « not found » littéralement le contenu « non trouvé » et l'impact d'une recherche peu ou pas efficace qui engendrent des retards projet ou des créations de contenus.

Étant donné les avantages potentiels ainsi que les défis croissants de la recherche d'informations d'entreprise, il est surprenant de constater le faible engagement des organisations vers la réalisation de cette stratégie. Selon l'étude Findwise<sup>25</sup> « Enterprise Search and Findability – 2012 », seulement 14% des répondants avaient une stratégie de recherche et 30% seulement avaient l'intention d'élaborer une stratégie en 2012/2013 [45, Intranet Focus].

## 2.2 Les modes de recherche d'informations

### 2.2.1 Une question de vocabulaire : recherche et recherche de l'information

Comme le soulignent Olivier Ertzscheid<sup>26</sup> et Gabriel Gallezot<sup>27</sup>, il existe une distinction entre « rechercher » de « la recherche d'information » (Information Retrieval en anglais, IR) et « rechercher » de « la recherche » (Research en anglais). En français, ce terme unique introduit un signifié commun « trouver de l'information » avec une distinction claire entre les moyens et les objectifs de la recherche d'information :

	<i>Information Retrieval</i>	<i>Research</i>
<b>Objectifs</b>	Repérer et de ramener des informations pertinentes	Découvrir et produire de nouvelles connaissances
<b>Moyens</b>	Outils du traitement de l'information sur un corpus documentaire.	Empirisme ou outils de traitement de l'information dont l'objet change (devient « naturel »).

Tableau 1 : Modes de recherche et de découverte [51, Ertzscheid, Gallezot]

Dans la suite de ce mémoire, la recherche d'information se référera au terme anglais Information Retrieval (IR).

### 2.2.2 La recherche d'informations : une question de pertinence

Il existe deux types de pertinence : la pertinence « système », obtenue par l'utilisation d'algorithmes sophistiqués et brevetés et la pertinence « utilisateur » complètement subjective et personnelle. L'enjeu est de réduire la distance qui peut exister entre l'approche système, qui montre vite ses limites et l'approche utilisateur, souvent imprécise et mal cadrée.

<sup>25</sup> <http://www.findwise.com/>

<sup>26</sup> Enseignant-chercheur (Maître de Conférences) en Sciences de l'information et de la communication et blogueur (<http://affordance.typepad.com/>)

<sup>27</sup> Maître de conférence) l'Université de Nice Sophia Antipolis

### 2.2.2.1 L'utilisateur au centre du processus de recherche

Ce schéma décrit le processus type d'une recherche d'informations et met en évidence l'importance du rôle de l'utilisateur qui, en procédant par itérations, interprète et en évalue en permanence l'adéquation entre les documents renvoyés par le moteur et le besoin d'information initial.

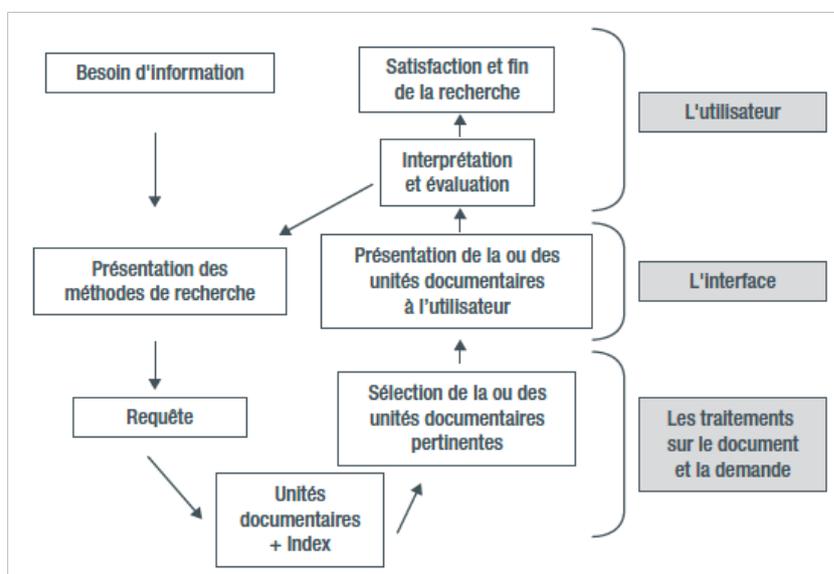


Figure 1 : Processus de recherche d'information<sup>28</sup>

La consultation des résultats amène l'utilisateur à leur donner du sens en fonction de ses besoins et de ses attentes. Et c'est bien parce que « l'individu est au cœur de la notion de pertinence que celle-ci est particulièrement complexe » [36, Balmisse] .

### 2.2.2.2 La pertinence « système »

Afin d'être en capacité de fournir aux utilisateurs des informations correspondant à leurs besoins, l'application de recherche va mettre en œuvre un modèle de pertinence qui calcule à chaque requête et pour chaque information le « meilleur » score de pertinence. Un classement par ordre décroissant sera alors proposé à l'utilisateur. Plusieurs critères entrent en jeu dans ce calcul. Ils sont plus spécifiquement abordés en section [3.2](#).

Ce calcul de pertinence a également pour but d'améliorer les taux rappel et de précision, notions de base en information retrieval :

- Le rappel est la mesure du nombre de documents pertinents retrouvés en fonction du nombre de documents **pertinents** que possède la base de données. L'utilisateur souhaite voir apparaître tous les documents qui pourraient répondre à son besoin d'information.

<sup>28</sup> Issu de « La recherche d'informations assistée par ordinateur : quelle représentation des connaissances ? » Maria Caterina Manes-Gallo et Céline Paganelli, 2004

- La précision est la mesure du nombre de documents pertinents retrouvés rapportés au nombre de documents **total** proposé par le moteur de recherche pour une requête donnée. Plus l'indice de précision est élevé, moins le moteur aura renvoyé de documents inutiles et non pertinents.

Les notions de rappel et de précision sont indissociables de la notion de pertinence. Le problème est que la pertinence est définie en termes absolus : soit un document est pertinent soit il ne l'est pas. Elle dépend du contexte dans lequel évolue l'organisation et de la capacité du moteur à correctement interpréter et contextualiser la requête de l'utilisateur [34, white].

### **2.2.2.3 La pertinence « utilisateur »**

Pour Gilles Balmisse<sup>29</sup> : « La pertinence n'est pas du tout un problème de correspondance entre une information et une requête. Elle est au contraire fortement dépendante du contexte de jugement ». La pertinence est donc un jugement personnel basé sur la valeur qu'un utilisateur accorde à un élément d'information. Elle se substitue donc au caractère computationnel (logico-algébrique) de la pertinence système.

En pratique, ce qui est pertinent pour un utilisateur ne le sera pas pour un autre, à fortiori s'ils évoluent dans un contexte métier différent et même si ces derniers ont effectué une requête identique. Chaque utilisateur possède sa propre vision des contenus, son propre système de valeurs et interprète une information en fonction de ses connaissances personnelles, de son contexte personnel ou de son contexte d'affaires.

Pour l'utilisateur, trois critères subjectifs se confrontent : la relativité, l'utilité et l'utilisabilité<sup>30</sup>. La relativité s'appuie sur la pertinence de l'information et son rapport avec le besoin initial ou en cours de formulation. L'utilité repose sur l'apport d'un nouvel élément de connaissance jugé pertinent par l'utilisateur. L'utilisabilité est jugée optimale si l'information peut être facilement traitée par l'utilisateur.

---

<sup>29</sup> Gilles Balmisse est dirigeant et fondateur du cabinet Knowledge Consult

<sup>30</sup> Source : Dictionnaire des concepts info-documentaires  
([http://www.cndp.fr/savoircsdi/fileadmin/fichiers\\_auteurs/PDF\\_manuels/dicoduplessis.pdf](http://www.cndp.fr/savoircsdi/fileadmin/fichiers_auteurs/PDF_manuels/dicoduplessis.pdf))

### 2.2.3 De la simple consultation à la découverte d'informations

Pour l'utilisateur, la recherche (plus ou moins ouverte à la découverte) correspond toujours à un ou plusieurs objectifs. Le tableau ci-dessous, inspiré du modèle proposé par Tony Russel-Rose<sup>31</sup>, classe du plus fermé au plus ouvert les différents modes de recherche possibles : consulter, enquêter et apprendre<sup>32</sup> [49, Russell-Rose].

Modes de recherche	Objectifs informationnels
Consulter	Repérage, récupération
	Vérification
	Contrôle, surveillance
Enquêter	Analyse
	Evaluation
	Synthèse
Apprendre	Comparaison
	Compréhension
	Exploration

Tableau 2 : modes de recherche et de découverte en entreprise  
[49, Russell-Rose]

Chaque mode de recherche mobilise plusieurs niveaux d'expertise :

- Une expertise du domaine : critère qui repose sur l'expertise métier de la personne et la profondeur de connaissance liée aux acteurs et aux enjeux du domaine
- Une expertise technique des systèmes : critère qui a tendance à devenir négligeable étant donné la volonté des éditeurs de rendre « User Friendly »<sup>33</sup> les interfaces d'interrogation/consultation
- Une expertise en recherche : critère qui repose sur la connaissance des sources et des stratégies de recherche.

L'utilisation d'un moteur est censée couvrir tous les modes de recherche et répondre à tous les objectifs associés : accéder simplement à des documents ou à des unités documentaires jugées pertinentes, délimiter les contours du problème en vue de son analyse et identifier des informations existantes dans un processus de découverte et d'apprentissage.

---

<sup>31</sup> Fondateur UXlabs et consultant en recherche et conception d'applications d'accès aux informations (<http://uxlabs.co.uk/>)

<sup>32</sup> Traduction libre de « Lookup », « Investigate », « Learn »

<sup>33</sup> Peut être traduit en Français par « facile à utiliser »

Deuxième partie

**Les fonctionnalités du moteur de  
recherche d'entreprise**

## 3 Les différences fonctionnelles entre un moteur d'entreprise et un moteur web

---

Par moteur généraliste web, on comprendra évidemment Google. En effet, en Juin 2013, Google concentre en France 91,1% des visites effectuées depuis un moteur et enregistrées par les sites web<sup>34</sup>.

Comme base de comparaison, Eric Debonne<sup>35</sup> propose 5 critères de différenciation entre un moteur de recherche web et un moteur de recherche d'entreprise : La volumétrie des sources, la gestion de l'information hétérogène, la prise en charge des informations structurées, la pertinence de la recherche et la sécurité des accès aux données. On peut cependant rajouter deux autres critères : la polysémie des sources et la qualité des informations.

### 3.1 Un critère de différenciation marqué : l'exhaustivité

Les deux types de moteurs ne répondent pas aux mêmes besoins. Le moteur de recherche web répond à une démarche d'exploration dans une logique de découverte pour l'apprentissage tandis que le moteur de recherche d'entreprise répond plutôt à une démarche de consultation dans une logique de vérification, de précision et d'exhaustivité.

L'exhaustivité n'existe pas sur Internet. Le volume d'information est tel qu'il n'est pas simple de savoir si une page a été indexée et sous quelle temporalité. L'exhaustivité n'est pas l'objectif du moteur de recherche web, mais la constitution d'une liste primaire de résultats : les 10 premiers [40, Debonne].

Dans un contexte d'entreprise en revanche, « Il ne s'agit pas de retrouver "un document parlant de", mais le document précis traitant du sujet » [16, Cotte, p. 220].

D'un point de vue quantitatif, le moteur de recherche d'entreprise satisfait avant tout les besoins de consultation des utilisateurs. Il répond à une démarche de « récupération » (retrieval) d'une information. D'un point de vue qualitatif, il peut aussi répondre à une démarche de découverte ou de réutilisation des connaissances par la plasticité de l'information qu'il manipule.

En effet, l'environnement informationnel des organisations s'est profondément modifié et l'information importante peut ainsi se retrouver aussi bien dans une base de données que

---

<sup>34</sup> <http://www.atinternet.com/documents/barometre-des-moteurs-de-recherche-juin-2013/>

<sup>35</sup> <http://solaci.fr/accueil/>

dans un email. Grâce au paramétrage du moteur et à l'interrogation croisée des sources structurées et non-structurées, on obtient une quasi-exhaustivité.

Cette recherche hybride a pour objectif de faire en sorte que « l'utilisateur ne loupe aucun document » [40, Debonne]. Les moteurs d'entreprise, via l'utilisation de connecteurs (voir section [4.2.1](#)), permettent d'interroger de manière fédérée des sources hétérogènes et disparates dans une même recherche (GED, ERP, BI...).

### **3.2 Deux modes de mise en valeur des résultats**

Les moteurs de recherche Web appliquent un classement ou « ranking » qui s'appuie sur un tri par pertinence et un indice de popularité qui se mesure au nombre de liens entrants ou « backlinks » qui pointent vers la page d'un site.

Le tri par pertinence est obtenu par un score attribué aux mots selon plusieurs facteurs [37, Lardy] :

- Le poids du mot dans une page déterminé par sa position dans la page (titre, début du texte, etc...). Le critère de poids est maximal pour le titre
- La densité basée sur le nombre d'occurrences par rapport à la taille de la page
- Le poids du mot dans la page basé sur le nombre d'occurrences dans toute la page
- La correspondance d'expression basée sur la similarité entre la question (la requête) et l'expression dans la page
- La relation de proximité basée sur la proximité des termes de la question entre eux dans la page

Si le tri par pertinence reste viable pour un moteur de recherche d'entreprise, l'indice de popularité, qui fonctionne très bien pour le Web n'est pas forcément valable dans un contexte d'affaires [39, Freyermuth].

Dans une entreprise, les informations importantes ne sont pas les plus populaires et reposent plutôt sur des critères de qualité, spécifiques à un métier ou à l'activité de l'organisation. L'algorithme de Google classant les résultats par popularité n'est donc pas le plus approprié pour l'entreprise.

En revanche, la pertinence du moteur d'entreprise est fondée sur l'analyse statistique ou linguistique du contenu et repose sur un calcul de similarité entre les termes de la requête et les termes de l'index représentant le document. Ce calcul de similarité s'appuie sur quatre critères principaux [36, Balmisse] :

- la présence ou non de l'ensemble des termes de la requête,

- le nombre d'occurrences des termes de la requête par rapport à la longueur des documents, leur localisation dans les documents (titre, corps, métadonnées),
- la proximité des termes trouvés entre eux : plus les mots de la requête sont proches dans le texte, meilleur sera le classement,
- la rareté relative des termes de la requête pour privilégier les documents contenant les termes rares.

### **3.3 Périmètre d'interrogation et qualité des sources**

Contrairement au moteur Web qui traite essentiellement des pages html<sup>36</sup> liées entre elles et sur un vaste périmètre, le moteur d'entreprise traite des informations enregistrées dans différents formats (structurés et non structurés), dans différents entrepôts de données et dans différentes applications mais sur un périmètre à priori maîtrisé.

Par son caractère restreint et fermé, le système d'information de l'entreprise permet de se concentrer sur des documents et des données dont on connaît à priori les caractéristiques et la structure (ou l'absence de structure parfois). Le vocabulaire employé en entreprise est également moins ambigu, plus ciblé et moins polysémique [26, Poibeau].

Le moteur de recherche d'entreprise permet donc, en théorie, d'utiliser de façon optimale les ressources informationnelles disponibles au sein de l'organisation et ainsi garantir au maximum une recherche d'information pertinente. La limite de cette approche est qu'un moteur de recherche reste un outil automatisé qui nécessite, pour qu'il soit utilisé correctement, d'adapter son paramétrage à chaque type de documents et données en fonction de leurs propriétés. « On doit donc à la fois adapter l'outil à son environnement et l'environnement à l'outil » et adapter chaque type de données aux exigences de l'organisation [22, Helmlinger].

C'est donc la qualité intrinsèque des contenus qui pose question. Si le moteur d'entreprise ne retrouve pas un document précis, l'utilisateur estimera qu'il ne fonctionne pas correctement tandis que si le moteur web remonte plusieurs milliers de résultats très éloignés de sa recherche, l'internaute ne lui en tiendra, à posteriori, pas rigueur.

La qualité, comme la pertinence, est relative, subjective et personnelle : un nombre important de liens pointant vers la page d'un site web ne signifie pas que l'information est de bonne qualité [34, white].

---

<sup>36</sup> L'HyperText Markup Language est format de données ainsi qu'un langage de balisage conçu pour représenter les pages web et intelligible pour les navigateurs web.

### 3.4 Gestion des droits d'accès

Sur le Web, La sécurité n'est pas une priorité. Les informations moissonnées et indexées par les moteurs web sont publiques (en théorie). En revanche, la gestion de la sécurité par le moteur d'entreprise est une ossature technologique essentielle [40, Debonne].

Dans la pratique, les enjeux d'une bonne gestion de la sécurité sont : la confidentialité des données et la performance du moteur [47, Jung]. Lors d'une requête, l'application de recherche ne doit restituer à l'utilisateur que les documents auxquels il a accès et au minimum les documents sur lesquels il a des droits en lecture.

Le moteur de recherche d'entreprise répond à un double enjeu : « ouvrir à tous le plus d'informations possible tout en protégeant l'information qui doit l'être » [16, Garnier]. Pour ce faire, le moteur doit s'appuyer sur plusieurs annuaires utilisateurs et plusieurs référentiels déjà présents au sein de l'organisation qui donnent lieu à des droits différenciés par utilisateur ou groupes d'utilisateurs. La gestion de la sécurité peut devenir complexe car le moteur doit prendre en compte deux niveaux de droits d'accès [36, Balmisse] :

- Le niveau d'autorisation : l'utilisateur est autorisé ou non à accéder directement à certaines sources d'information
- Le niveau d'accréditation : l'utilisateur est accrédité ou non à rechercher et/ou à visualiser l'information pour laquelle il a une autorisation

### 3.5 Pourquoi une logique « à la Google » ne peut pas être appliquée en entreprise ?

Exhaustivité, fiabilité, qualité, pertinence, spécificité des formats et sécurité, toutes ces contraintes font que la recherche en entreprise est un domaine différent de la recherche sur Internet. Le système d'information de l'entreprise n'est pas « un Internet à échelle réduite<sup>37</sup> » [38, Bennett]. Pourtant, cette vision d'un web transposé au monde de l'entreprise, très compréhensible et extrêmement répandue chez les utilisateurs, vient du fait qu'ils ont souvent comme unique point de référence Google.

D'un point de vue organisationnel et en termes de performances collectives, le modèle introduit par Google n'est pas le plus pertinent. Même si Google a le mérite de refléter d'une manière très grossière l'intelligence collective des utilisateurs du Web, les algorithmes de recherche sont très génériques et peu personnalisés [58, Levy].

Or, dans un contexte professionnel, la possibilité de pouvoir initier des modes de filtrage plus fins fondés sur un tri personnalisé en fonction du contexte métier d'un utilisateur ou

---

<sup>37</sup> Traduction libre de : « The Enterprise is not just a small Internet ».

d'un groupe d'utilisateurs est un vrai facteur différenciant pour les entreprises ainsi que pour leurs employés. Comme le précise Pierre Levy<sup>38</sup>, « la valeur de l'information dépend d'un grand nombre de paramètres contextuels ou de « points de vue » et elle devrait pouvoir être calculée selon des jeux d'intelligence collective extrêmement variés ».

Mais, largement influencés par la rapidité et la simplicité de Google, les utilisateurs établissent naturellement un parallèle entre leur expérience sur le web et leur expérience au sein de leurs organisations respectives, posant ainsi la question du manque d'efficacité des solutions de recherche déployées en entreprise.

Selon le rapport "Global Intranet Trends 2010"<sup>39</sup>, dans le chapitre consacré à la recherche d'information, seulement 14% des répondants étaient très satisfaits de leur application de recherche alors que 47% étaient moyennement satisfaits. Comparativement, 6% ont été très satisfaits du contenu qui a été retourné à partir d'une recherche et 40% des répondants ne sont «pas vraiment satisfaits» ou «pas du tout satisfaits» de leur application de recherche [44, White]. Le fossé ne cesse de se creuser entre les utilisateurs et leur application de recherche, malgré les sommes importantes investies par les entreprises [48, Mindmetre]. Plusieurs raisons à cela :

Les critères qui définissent "une bonne recherche" dépendent du contexte. Sur le web, elle est synonyme de facilité et de rapidité apparente, rendant invisible pour l'utilisateur l'énorme travail d'optimisation réalisé par les équipes de R&D de Google (40% des effectifs de la société [34, white]). La demande d'information reste très en surface et peu précise. Cette vision très optimiste a établi un point de référence assez faible chez les utilisateurs.

Quand il s'agit de la recherche d'information en entreprise, les utilisateurs adoptent une logique héritée de leur pratique sur le web et sont forcément déçus par ce qu'ils trouvent (ou ce qu'ils ne trouvent pas). En entreprise, la demande d'information est synonyme de précision afin de trouver des documents spécifiques et pertinents. L'optimisation des contenus informationnels et l'amélioration de la pertinence du moteur impliquent la constitution en interne d'une équipe dédiée. Or, actuellement, la plupart des organisations ont moins d'une personne à temps plein pour assurer cette fonction. [27, Reamy]

D'autre part, les sources d'information utilisées dans l'entreprise pour éclairer les processus de décisions et les processus d'affaires sont beaucoup plus hétérogènes (mix de structuré et de non-structuré) et sont moins bien reliées entre elles que des pages Web et des sites Internet [11, Microsoft Corporation].

---

<sup>38</sup> Docteur en sociologie et en sciences de l'information et de la communication, il occupe la chaire de recherche sur l'intelligence collective à l'université d'Ottawa.

<sup>39</sup> <http://www.digital-workplace-trends.com/>

## 4 Le fonctionnement du moteur de recherche d'entreprise

---

On l'a vu, les interfaces de recherche sont au centre de la « différenciation des réseaux informationnels et communicationnels » et notamment dans le cadre de la recherche d'informations en entreprise [18, Noyer, Carmes]. Mais, dans le même temps, l'utilisation des moteurs de recherche web dans la sphère privée a défini chez les utilisateurs des attentes fortes en termes de rapidité et de facilité d'utilisation.

Le défi de la recherche d'informations, c'est de considérer chaque utilisateur de manière singulière, avec sa propre perception individuelle de ce qu'est un contenu pertinent ou pas. Dans l'environnement informationnel de l'entreprise, l'objectif consiste donc à trouver un moyen de répondre aux attentes individuelles à l'aide d'un outil « générique ». Quels sont les mécanismes fonctionnels qui permettent de répondre le plus efficacement possible aux attentes des utilisateurs et de leurs organisations en conciliant rapidité, exhaustivité et qualité de l'information ?

### 4.1 Le cœur fonctionnel du moteur

Un moteur de recherche, à un niveau extrêmement basique et quel que soit sa sophistication, repose sur trois grandes fonctionnalités qui se répartissent en trois phases distinctes mais complémentaires :

- La phase d'acquisition et d'indexation de l'information
- La phase de traitement des requêtes utilisateurs
- La phase de présentation des résultats.

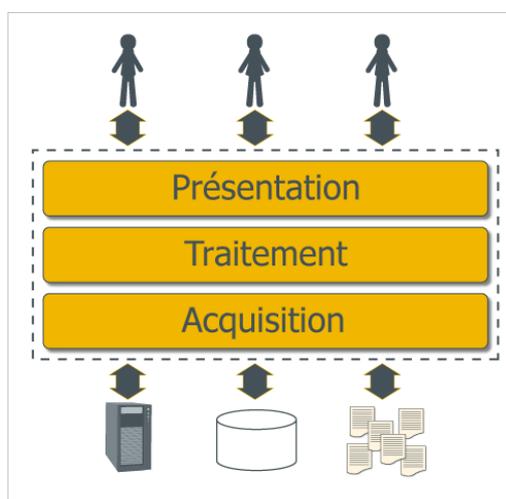


Figure 2 : L'architecture fonctionnelle d'un moteur de recherche [36, Balmisse]

Les phases d'acquisition et de traitement des données, c'est-à-dire leur analyse et leur interrogation sont totalement invisibles pour les utilisateurs. La phase de présentation représente l'unique couche visible par ces derniers. Le moteur sera jugé par ses utilisateurs sur la seule perception (bonne ou mauvaise) de cette couche de présentation.

## **4.2 Processus global : de l'acquisition à la présentation des résultats**

### **4.2.1 Acquisition des contenus**

Afin de rendre accessible une information par le biais d'une recherche, l'application de recherche doit acquérir les informations (données ou documents) sollicitées lors de cette recherche, autrement dit les rendre disponibles pour elle-même [24, Imbert]. Cette étape d'extraction et de sauvegarde de données constitue la base du futur index (voir section [4.2.2](#)). Pour faire l'acquisition des données deux approches sont envisageables :

La première approche repose sur l'utilisation de logiciels, des « crawlers », ou agents de recherche, programmés pour la détection de nouvelles informations dans une source identifiée. Cette approche, très répandue sur le web, ne permet cependant pas une indexation en temps réel de l'information.

En effet, les crawlers sont programmés pour scruter les sources de données suivant une temporalité fixe. Une recherche d'informations pourra potentiellement aboutir à une version antérieure non mise à jour ou obsolète des sources d'informations concernées.

La seconde approche repose sur l'analyse temps réel des modifications sur les contenus avec l'utilisation de connecteurs spécifiques. Un connecteur est « un logiciel qui permet de faire le lien entre l'API<sup>40</sup> du moteur de recherche et l'API d'une source d'information » [36, Balmisse]. Cette solution offre beaucoup plus de souplesse car, contrairement au crawlers qui procèdent de la même manière d'une source à l'autre, les connecteurs sont capables de communiquer d'une manière très fluide avec la source d'information en prenant en compte sa structure, son format et ses paramètres de sécurité.

Cette approche présente toutefois deux inconvénients : le coût du connecteur (plusieurs milliers d'euros dans certains cas) et sa spécificité par rapport à la source d'information concernée. En effet, il faut procéder au changement du connecteur ou à son paramétrage à chaque modification de la source d'information, comme l'installation d'une nouvelle version par exemple. A l'issue de cette phase, l'application de recherche stocke dans un index les données générées par l'analyse des sources.

---

<sup>40</sup> Application Programming Interface ou interface de programmation : ensemble normalisé de fonctions qui sert d'accès par lequel un logiciel offre des services à d'autres logiciels.

## 4.2.2 Indexation des contenus

Les applications de recherche reposent sur l'utilisation de deux types d'index pour stocker les données générées par l'analyse des sources : un index principal ou « maître »<sup>41</sup> et un index inversé qui associe les termes d'accès susceptibles d'être sollicités pendant une recherche à tous les documents présents dans l'index maître [24, Imbert].

En principe, un index inversé est similaire à l'index d'un ouvrage, mais au lieu de simplement donner un numéro de page, l'index du moteur est capable de donner la position du mot dans le document mais également sa position dans la page.

L'index inversé repose donc sur deux composantes fondamentales : L'index au niveau des mots et l'index au niveau du document. L'index des mots est défini par la position du mot dans le document tandis que l'index du document est fondé sur un comptage du nombre d'occurrences du terme dans le document [34, white].

### 4.2.2.1 Particularités de l'indexation automatique des contenus

La norme NF Z 47-102 décrit l'indexation comme « une représentation condensée d'un document en s'appuyant sur les mots ou les concepts qu'il contient", c'est-à-dire une transcription en langage documentaire des informations après les avoir extraites du document par une analyse [12, Afnor].

En effet, l'étape d'indexation a pour objectif de faciliter et/ou d'améliorer la recherche en disposant de l'information sous un format optimisé. Pour ce faire, l'application de recherche utilise une indexation automatique ou algorithmique. On parle aussi d'indexation « plein texte » ou d'indexation libre. Ce type d'indexation présente le principal inconvénient de générer un bruit<sup>42</sup> important.

Pour limiter ce bruit et ne retenir que l'information utile, il est donc nécessaire de mettre en œuvre des prétraitements afin que l'indexation et la recherche d'informations soient aussi performantes que possible. Cette indexation libre combine deux approches complémentaires mais très différentes dans leur mise en œuvre :

La première approche repose sur des technologies linguistiques proposant « des solutions qui s'attachent à traiter le contenu informationnel des documents en prenant en compte les niveaux morphologiques, syntaxiques et sémantiques de la langue » [10, Chaudiron]. Le défi pour ces techniques est de devoir composer avec les ambiguïtés du langage naturel.

---

<sup>41</sup> L'index principal peut être comparé à « une mini base de données »

<sup>42</sup> Le bruit est caractérisé par des réponses non-pertinentes proposées à l'utilisateur et augmente le risque que les informations pertinentes soient noyées dans un nombre trop important d'informations parasites.

La seconde approche repose sur l'extraction automatique des unités linguistiques ou mots du document. Par un traitement algorithmique, le système d'indexation choisit ensuite les plus significatifs et les plus représentatifs par un calcul de fréquence [57, Zackald]. L'inconvénient de cette approche, fondée essentiellement sur des techniques statistiques, est qu'elle n'est capable que d'identifier des chaînes de caractères et non des concepts qui font sens.

### 4.2.3 Prétraitements linguistiques et statistiques

Ces phases sont pilotées par Traitement Automatique du Langage ou TAL<sup>43</sup>. Les technologies de TAL « visent à faire manipuler, interpréter ou générer par les machines le langage naturel écrit ou parlé par les humains » [08, Amar]. Il s'agit notamment pour le TAL d'analyser les contenus afin de faire émerger l'intégralité des concepts pertinents porteurs de sens par l'analyse de la langue ou de proposer un classement plus pertinent des informations à la présentation des résultats [36, Balmisse].

On distingue plusieurs niveaux d'analyse et de traitements [22, Helmlinger] :

- La segmentation qui consiste à découper le texte en « unités élémentaires »
- L'analyse morphologique s'intéresse aux variations de formes entre mots d'une même famille
- L'analyse syntaxique qui détermine la structure des phrases (relations entre les mots et relations entre les groupes de mots) en fonction d'une grammaire de référence
- L'analyse sémantique qui traite du sens des mots et des phrases [10, Chaudiron]

#### 4.2.3.1 Les étapes de l'analyse linguistique

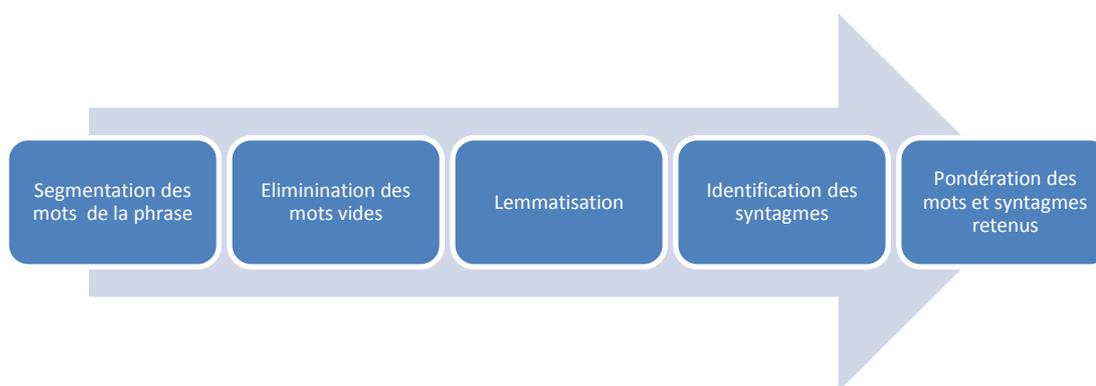


Figure 3 : Les étapes de l'indexation automatique pilotées par TAL

<sup>43</sup> En anglais NLP ou Natural Language Processing

Le schéma précédent décrit les différents modules qui constituent le pipeline<sup>44</sup> de prétraitements nécessaires à l'enrichissement de l'indexation automatique.

L'étape de segmentation des mots de la phrase (parsing et tokenising en anglais) vise à identifier les mots puis les phrases constitutifs du texte et à former des unités lexicales élémentaires. Cette étape d'analyse consiste à reconnaître puis regrouper les chaînes de caractères alphabétiques, numériques et typographiques en identifiant au préalable les signes séparateurs entre ces unités lexicales [14, Chaudiron]. Cette étape peut poser problème pour les sigles ainsi que pour les mots composés où les points et les tirets ne sont pas des « marqueurs de fin de phrase » [23, Hérigault].

L'élimination des mots vides, ou Stop Words en anglais, vise à écarter de l'indexation certains mots ou termes considérés comme trop génériques, d'un usage trop courant (les auxiliaires *être* ou *avoir* par exemple) ou à faible valeur sémantique (*le, de, pour* par exemple). L'objectif de cette étape est de constituer une liste générique dépendante de la langue d'origine ainsi qu'une liste moins discriminante et spécifique au domaine d'activité de l'entreprise.

L'étape de lemmatisation repose sur la mise en œuvre d'un analyseur morphologique ou lemmatiseur. Ce traitement consiste à mettre sous leur forme canonique les unités lexicales élémentaires issues de la segmentation et de reconnaître les mots simples ou composés. La forme canonique d'un mot, le lemme, est l'infinitif pour un verbe et le masculin-singulier pour les autres formes [36, Balmisse]. En sortie de module, le texte apparaît dans l'index sous la forme d'une liste de lemmes avec leur catégorie syntaxique (ou grammaticale) et les informations morphologiques nécessaires.

Les étapes d'identification et de pondération des syntagmes sont réalisées par la mise en œuvre d'un analyseur syntaxique, ou « parser » en anglais, dont le but est d'abord d'identifier les différents éléments constitutifs de la phrase, appelés syntagmes. Les syntagmes sont des groupes de mots qui se suivent avec un sens.

Cette extraction de syntagmes ou groupes nominaux est importante car de nombreux systèmes (automatisés ou non) utilisent ces derniers comme candidats descripteurs, des termes représentatifs du contenu informationnel d'un document. Dans cette liste de termes, qui possèdent les propriétés pour devenir des descripteurs, le système détermine grâce à un calcul de fréquence ou de pondération leur « représentativité » [57, Zackald]. Ce calcul de représentativité dépend de l'emplacement du mot dans le texte, de son nombre

---

<sup>44</sup> La notion de pipeline de traitement est issue de l'ingénierie logicielle où la sortie (output) d'un processus devient l'entrée (input) du processus suivant.

d'occurrences ou de sa présence éventuelle dans un vocabulaire contrôlé, un référentiel linguistique prédéfini : une liste d'autorités, un thésaurus ou une taxonomie<sup>45</sup>.

Ce traitement optionnel consiste à remplacer les termes et les syntagmes extraits des documents par leurs équivalents choisis dans un référentiel. Cette fonction permet donc de rapporter la question de l'indexation libre à celle de l'indexation contrôlée [10, Chaudiron].

### 4.3 Traitement des requêtes utilisateurs

Afin d'être en mesure de comprendre la requête utilisateur, différents modules exécutent une grande diversité de calculs (ou instructions) algorithmiques, statistiques ou linguistiques, qui ont pour but d'analyser les requêtes afin de proposer les résultats les plus pertinents possible.



Figure 4 : Pipeline de traitement des requêtes [36, Balmisse]

Comme l'illustre la figure précédente, le pipeline de traitement de la requête s'effectue en plusieurs étapes :

- Le moteur de génération de requête applique des traitements linguistiques : normalisation de la casse<sup>46</sup> et de l'accentuation, transformation des mots en lemmes
- Le module analyse la requête, la transforme dans une optique d'amélioration et d'enrichissement puis effectue un appariement. Cette étape consiste à récupérer (Retrieve) l'ensemble des informations répondant le mieux à la requête initiale.
- Une opération de fusion des résultats est effectuée par « collection, dédoublement et ordonnancement » [36, Balmisse]. Enfin, le composant en charge de la présentation des résultats procède à une mise en forme avant affichage.

#### 4.3.1 L'apport des traitements linguistiques

On l'a vu, les traitements linguistiques permettent d'améliorer la qualité de l'indexation. Ces traitements, invisibles pour l'utilisateur, permettent également d'améliorer de façon significative la qualité des réponses qui lui sont proposées.

<sup>45</sup> Une taxonomie est un outil d'accès à l'information qui sera mis en œuvre dans un système d'indexation manuel ou automatisé. C'est une « carte conceptuelle » représentative du contenu informationnel et organisationnel d'une organisation [09, Bennett].

<sup>46</sup> On parle de casse pour désigner l'alternative entre capitale (ou majuscule) et minuscule.

#### **4.3.1.1 Aide au traitement de la requête**

Grâce à l'index enrichi par les traitements linguistiques et morfo-syntaxiques, le moteur de recherche va être en capacité de gérer les extensions de requêtes : reformulation de la requête avec les lemmes, extension de la requête aux singuliers et pluriels, masculins et féminins, à des mots proches orthographiquement et phonétiquement. Grâce à la lemmatisation, l'utilisateur va pouvoir rechercher une information sans avoir à se préoccuper des déclinaisons des termes employés, par exemple singulier ou pluriel. La recherche de « perspectives économiques » va permettre tout naturellement de rechercher « perspective économique » de manière totalement transparente pour l'utilisateur. De même, l'extraction de syntagmes permet d'améliorer le taux de précision car les mots composés et les expressions sont moins ambiguës que les mots simples.

### **4.4 La performance du moteur : un dialogue permanent avec les utilisateurs**

Quels sont les mécanismes qui permettent de répondre au mieux aux besoins informationnels des utilisateurs dans le contexte d'une recherche d'informations en entreprise ? : Classification et catégorisation des données, aide au tri et à l'affinage des réponses, aide à la navigation, exploitation des métadonnées descriptives attachées aux documents.

#### **4.4.1 Classification et catégorisation : modes de représentation de l'information**

Outre la construction d'un index, l'extraction des termes des documents ou des syntagmes issus d'un texte ou d'un flux d'information permet la constitution automatique et « libre » d'agrégats de termes ou de concepts, aussi appelés « clusters » en anglais. Cette construction automatique s'effectue via l'analyse des occurrences lexicales contenues dans ces données et non en fonction d'un plan de classement préexistant [10, Chaudiron].

Ce mode de représentation de l'organisation ascendante ou « bottom up » de l'information contribue à l'élaboration de représentations diverses. Une utilisation fréquente des clusters se concrétise sous la forme d'une visualisation de l'information dont l'objectif est d'aider à appréhender rapidement un important volume d'informations textuelles (analyses textuelles, cartographies décisionnelles).

Contrairement à la classification, la catégorisation consiste à classer les informations textuelles en fonction d'un ensemble préexistant de catégories structurées, organisées et éventuellement hiérarchisées comme les thésaurus, les taxonomies ou encore les ontologies.

Ce mode de représentation déductive ou « bottom down » s'appuie un traitement principalement fondé sur une identification des termes du document, et vise à assigner

automatiquement ce document ou ce flux entrant d'informations textuelles dans le plan de classement préexistant [55, Bourdoncle]. Cette fonctionnalité, constitue une aide efficace au classement automatique des flux.

#### **4.4.2 L'utilisation de facettes et de filtres : usages et limites**

Traditionnellement, après avoir lancé une recherche, les utilisateurs sont souvent laissés à eux-mêmes avec trois options possibles :

- trier à nouveau la liste de résultats (par exemple par date, par auteur ou par ordre alphabétique)
- aller de documents en documents au risque de se perdre et d'être saturés d'informations
- tenter de reformuler la requête originale.

Compte tenu du modèle actuel de recherche qui repose sur "un ou deux mots-clés", il est important de donner les moyens aux utilisateurs d'affiner le résultat d'une indexation en texte intégral renvoyé par une requête en langage naturel grâce à des outils supplémentaires leur permettant de naviguer dans un grand nombre de sources d'informations avec une grande fluidité. Ces outils de « raffinement » sont généralement présentés en complément ou à côté d'une liste de résultats. Les utilisateurs sont invités à sélectionner des entrées de manière interactive pour contraindre ou élargir le champ de leur recherche [11, Microsoft Corporation].

Ces raffineurs (ou facettes) sont très répandus sur le web. Ils sont même devenus la norme, notamment sur les sites de e-commerce et les sites médias. Pour autant, leur implémentation en entreprise est loin d'être acquise en raison, entre autre, de leur coût. Mais, même lorsque ces outils complémentaires de recherche sont disponibles leur efficacité dépend largement de la qualité des métadonnées.

#### **4.4.3 La gestion des métadonnées : usages et limites**

La norme FD X50-185<sup>47</sup> les définit comme un « ensemble structuré de données servant à localiser et à décrire une ressource informationnelle consignée sur un support documentaire en vue de faciliter et d'améliorer son repérage, sa gestion, son usage ou sa préservation ». En d'autres termes, les métadonnées sont des « données sur les données ». Elles représentent un ensemble de propriétés qui décrivent un document ou tout autre élément de contenu. Le but des métadonnées est double : d'une part, l'organisation du contenu et d'autre part, la localisation de ce contenu.

---

<sup>47</sup> <http://www.boutique.afnor.org/norme/fd-x50-185/outils-de-management-management-de-l-information/article/632618/fa122236>

Des métadonnées de faible qualité, absentes ou incohérentes dégradent les performances de l'application de recherche, érodent la confiance des utilisateurs dans l'application et contribue à une faible adoption de cette dernière [11, Microsoft Corporation].

Il faut toutefois distinguer deux types d'informations qui sont impactées négativement à la recherche par des métadonnées mal ou peu structurées :

- l'information structurée où les métadonnées sont présentes. Mais elles sont faiblement exploitées par le moteur faute de stratégie globale de gestion de l'information et de normalisation.
- l'information non structurée qui représente la majorité des informations circulant dans l'entreprise. Leur exploitation optimale dépend des métadonnées descriptives saisies manuellement (quand elles sont saisies).

Leur variabilité est extrêmement importante tant d'un point de vue qualitatif que quantitatif. Elles sont soit inexistantes (pas d'entrée ou valeurs manquantes dans certains champs) soit incohérentes (variations orthographiques, abréviations, synonymes qui provoquent des interprétations différentes pour une même entrée) ou bien encore inexactes ou fausses.

Or, le développement des métadonnées, qu'elles soient gérées manuellement (par une équipe dédiée) ou bien gérées automatiquement (enrichissement par intégration de référentiels tels que les taxonomies ou les thesaurus, classification et catégorisation automatique) exige un engagement fort aussi bien humain que financier [29, Reamy].

Le coût relatif des métadonnées doit être mis en balance avec le coût de l'inaction pour les organisations concernées. Selon une estimation du cabinet IDC<sup>48</sup>, le coût d'une politique de recherche d'informations défaillante aurait un coût financier qui s'élèverait à 4,5 millions d'euros par an pour une entreprise de 1000 personnes. Toutefois, selon Tom Reamy<sup>49</sup>, il est peu probable que l'ajout de métadonnées permette de résoudre tous les problèmes liés à la recherche, mais l'investissement reste modéré comparé aux sommes perdues par les entreprises [13, Reamy].

La mise en œuvre d'un moteur de recherche est donc, au-delà de l'enjeu technologique évoqué dans ce chapitre, un véritable projet d'entreprise.

---

<sup>48</sup> <http://www.idc.fr/>

<sup>49</sup> Fondateur de KAPS Group et consultant en gestion des connaissances, spécialiste des taxonomies

Troisième partie

**Le moteur de recherche comme  
projet d'entreprise :**  
enjeux organisationnels, acteurs du marché et  
choix de la solution

## 5 Un projet managérial

---

La mise en place d'une solution de recherche d'entreprise questionne l'organisation bien au delà de ses aspects purement techniques ou fonctionnels. Comment sa mise en œuvre participe-t-elle à une recomposition possible de la gouvernance informationnelle, à la visibilité des informations? Quelles sont également les opportunités pour les organisations ? Car comme le souligne Gilles Balmisse : « il ne peut pas y avoir de bonne recherche d'information sans une bonne gestion voire une gouvernance de l'information » [35, Balmisse].

### 5.1 Un outil de gouvernance de l'information

Selon une étude menée par Mindjet<sup>50</sup> en Angleterre, un collaborateur perdrait quotidiennement en moyenne 21 minutes de son temps de travail à rechercher des informations dont il connaît l'existence mais qu'il ne parvient pas à retrouver dans divers applicatifs de l'entreprise. Cette étude révèle également que tout ceci a un coup pour les organisations : 1500€ par an et par salariés [21, Poulet]. Ces deux chiffres soulignent l'importance de la recherche dans la vie des salariés ainsi qu'un manque de rationalité dans la gestion de l'information.

La gouvernance de l'information est liée à la stratégie des entreprises et des organisations. Cette démarche consiste à recenser, à évaluer, à organiser l'information afin de permettre sa valorisation. Cette action stratégique doit se faire au bénéfice des 4 principaux objectifs de l'entreprise : le développement commercial, l'innovation, la maîtrise des coûts et l'exigence de conformité [20, Hugonie].

Un moteur de recherche reste un outil, mais sa grande plasticité et sa capacité à indexer et traiter de grands volumes de données contenus dans différents applicatifs et entrepôts de données dans l'entreprise permet, au moins d'un point de vue opérationnel, d'envisager une telle gouvernance.

Dans ce contexte, Christian Fauré<sup>51</sup> met en lumière deux usages « managériaux » des moteurs de recherche d'entreprise :

- Un instrument d'audit pour vérifier les lacunes dans la politique de confidentialité, pour repérer des dysfonctionnements dans le versioning<sup>52</sup> des documents, pour

---

<sup>50</sup> Enquête Mindjet : « L'information 2.0 submerge l'entreprise et ses collaborateurs » [http://www.mindjet.com/fr/press/releases/fr\\_2012-07-18\\_enquete/](http://www.mindjet.com/fr/press/releases/fr_2012-07-18_enquete/)

<sup>51</sup> Enseignant à l'Université Technologique de Compiègne et consultant

<sup>52</sup> Gestion des versions de l'ensemble des versions d'un ou plusieurs fichiers

vérifier des informations équivoques, des informations qui se contredisent ou encore pour fournir des informations sur les pratiques grâce à la consultation des logs de recherche

- Un instrument de gouvernance, en jouant sur les sources indexées et l'algorithme de classement, qui permettrait de mettre en avant des documents, d'écartier des pratiques déviantes.

## **5.2 Opportunités et risques pour les entreprises**

### **5.2.1 Une évaluation efficace du capital informationnel**

Une étude du SerdaLAB<sup>53</sup> parue en 2011 révèle que seulement 10% des répondants estiment que la gouvernance documentaire est un moyen de valoriser les informations en tant que capital immatériel de l'organisation [07, Serdalab].

Un des bénéfices de la mise en œuvre d'un moteur de recherche d'entreprise est de pouvoir évaluer le capital informationnel qui est une des composantes des actifs immatériels détenus par les organisations (voir section [1.3](#))

En effet, différents métiers ont l'obligation de l'évaluer : Les projets d'archivage et de records management, de Master Data Management<sup>54</sup> ou encore d'eDiscovery<sup>55</sup>. Cette évaluation du patrimoine informationnel répond à plusieurs objectifs :

- faire le tri entre l'information qui a de la valeur de celle à éliminer
- gérer les documents, leurs versions successives, leur maintenance, leur mise à jour, leur stockage et enfin leur sauvegarde
- optimiser les coûts, accroître la rentabilité et les gains de productivité, et gagner en compétitivité

L'explosion du volume des données et de leur coût de traitement devient également un enjeu majeur pour les organisations (voir section [1.4](#)). Conserver l'intégralité des informations pour se prémunir du risque de destruction par erreur est un non sens puisque, non seulement les coûts deviennent trop importants mais également le temps de recherche des informations augmente [20, Hugonie].

---

<sup>53</sup> <http://www.serdalab.com/>

<sup>54</sup> Gestion des données de référence : définit un ensemble de processus et méthodes visant à définir, stocker, maintenir à jour des données référentielles (données « clients/fournisseurs », données « produits » et données « financières » ) au sein d'un système d'information.

<sup>55</sup> Découverte de données juridiques notamment dans le cadre de litiges.

De ce fait, la mise en œuvre d'un moteur de recherche devient pertinente afin de valoriser le capital informationnel de l'entreprise et d'en évaluer le contenu selon plusieurs critères qualitatifs:

- l'évaluation de la valeur juridique de contenus (obligation de conformité, gestion des risques)
- l'évaluation de la valeur « métier » des contenus (pertinence de la conservation ou de la destruction de l'information dans un contexte d'affaires)
- en marge, l'évaluation de la valeur historique ou patrimoniale des contenus

### **5.2.2 L'impact sur la visibilité des informations**

Pour autant la mise en place d'un moteur de recherche d'entreprise ne résout pas tous les problèmes, il peut éventuellement aussi les accentuer. Comme le constate Christian Fauré : « Devoir chercher de l'information est le symptôme d'une défaillance dans les procédures de l'organisation et devient difficilement acceptable ». Cela peut être le reflet d'un manque de visibilité et d'homogénéité de l'information mais aussi un « marqueur » hiérarchique ou de compétences.

Ainsi, dans certains contextes organisationnels, le moteur de recherche est chargé de régler « le degré de visibilité » des informations disponibles [18, Noyer, Carnes]. La recherche d'information en entreprise ne serait utile que pour une petite partie des fonctions considérés comme à « haute valeur » pour l'entreprise, tels que les chercheurs ou commerciaux, qui doivent nécessairement être très bien informés avec les dernières informations disponibles. Le moteur de recherche d'entreprise peut apparaître comme une interface discriminante uniquement accessible aux cadres (règles d'accès restreintes à certaines sources de données, gestion de la mobilité, gestion de la sécurité).

D'autre part, l'enjeu de la visibilité des informations peut aussi se jouer au niveau de la gestion des métadonnées. La production de métadonnées coûte chère quelle que soit la solution retenue : automatique, semi automatique et manuelle [13, Reamy]. Ce coût, parfois élevé, peut pousser les organisations à focaliser leurs efforts uniquement sur un contenu considéré comme « à haute valeur » ou stratégique et à délaisser un contenu considéré comme « secondaire ». Cela pose la question de la valeur « supposée » des informations (qui décide de leur valeur? sur quels critères?) et de leur visibilité dans les réponses fournies par le moteur. Un contenu riche en métadonnées a plus de chance d'être retrouvé facilement et valorisé par le moteur.

### 5.2.3 Un retour sur investissement difficilement quantifiable

On peut définir le ROI comme un ratio financier qui permet de calculer la perte ou le gain par rapport à un investissement initial. Il permet d'estimer l'écart entre une valeur initiale et une valeur finale, toutes deux estimées en termes monétaires [54, Lesperance, Ulmer].

Or, la mise en œuvre d'une application de recherche, qui repose sur des changements humains, organisationnels et opérationnels (gain en temps de recherche, exhaustivité, et précision de l'information retrouvée), est un projet dont les retours se mesurent à long terme et qui comporte une part importante de bénéfices intangibles et donc plus difficilement quantifiables.

<b>Bénéfices tangibles</b>	<b>Bénéfices intangibles</b>
Gains de temps et de productivité	Meilleure gestion des connaissances
Meilleure performance individuelle	Satisfaction et fidélisation des employés
Rapidité d'accès à l'information	Valorisation des expertises
Moins de recréation de contenus	Meilleure capacité d'innovation
Qualité accrue de l'information	Compétitivité renforcée
Réduction de nombre d'emails	Relation client améliorée
Réduction des coûts d'assistance	Veille économique
Stratégies de marketing plus efficace	Meilleure intégration et communication
Développement de produits plus rapide	Renforcement de marque

Tableau 3 : Bénéfices tangibles et intangibles d'un moteur de recherche d'entreprise

Le calcul d'un ROI, à posteriori, cherche surtout à faire émerger des bénéfices tangibles et mesurables de façon monétaire. Mais, comme le montre le tableau ci-dessus, les bénéfices d'une application de recherche ne se limitent pas aux seuls gains financiers et reposent souvent sur des critères plus « subjectifs » comme une meilleure gestion des connaissances ou la satisfaction des utilisateurs [53, Poncier].

La question du ROI peut être également envisagée dans les phases amont du projet. L'étude initiale du besoin doit permettre de révéler le ou les services répondant aux plus fortes valeurs ajoutées dans le contexte d'affaires de l'entreprise. Dans ce cas, les bénéfices attendus par la mise en œuvre du moteur sont donc plus facilement mesurables et quantifiables [50, Couillault, Debonne].

Enfin, pour qu'il y ait création de ROI, cela doit s'envisager sur le long terme : « la performance d'un système d'organisation et d'information doit avoir un caractère régulier et

durable » [54, Lesperance, Ulmer]. Ceci implique un engagement et un soutien non court-termiste du management qui espère trop souvent des gains rapides ou « quick wins ».

### **5.2.3.1 La notion ROC<sup>56</sup> : la collaboration trans-organisationnelle et inter-silos**

Comme le souligne Christian Bourret<sup>57</sup>, les investissements consacrés « à la recherche de l'information ont été considérés comme un mal nécessaire. Aujourd'hui, on reconnaît que ces fonctions sont aptes à créer de la valeur » [02, Bourret]. Encore faut-il que cette création de valeur soit calculée avec des indicateurs valables.

La notion de ROC ou « retour sur collaboration » mesure l'amélioration inhérente à un investissement « dans l'intangible » contrairement au ROI dont le calcul se base initialement sur des sommes investies, gagnées ou perdues. On l'a vu, les bénéfices intangibles lors de la mise en place d'un moteur de recherche (co-création, réutilisation de documents et de connaissances) sont assez difficiles à quantifier.

Dans ce contexte, la notion de ROC est particulièrement adaptée pour mesurer l'apport d'une application de recherche, puisque c'est une méthode de « calcul » des retours qui ne se base pas uniquement sur des données financières. Ce modèle a le mérite de souligner le rôle majeur du partage de l'information dans le succès à long terme d'un investissement comme un moteur de recherche [21, Poulet].

---

<sup>56</sup> Return On Collaboration

<sup>57</sup> Maître de conférences en sciences de l'information et de la communication - Université Paris-Est Marne-la-Vallée

## 6 Les éditeurs de solutions

---

Comparativement aux autres acteurs du marché proposant des solutions pour les entreprises (ECM, CMS, GED, ERP...), le marché des éditeurs de moteurs de recherche reste un marché de niche, très compétitif, concurrentiel, et de plus en plus mature. Aujourd'hui, le marché se concentre autour de 80 éditeurs qui se partagent un marché estimé à 2,3 milliards<sup>58</sup> d'euros par an [29, Intranet Focus]. Trois acteurs dominent le marché : Microsoft Fast, Google et HP Autonomy.

### 6.1 Caractéristiques du marché

#### 6.1.1 Un mouvement de concentration et de consolidation

Les acquisitions sont une réelle tendance sur le marché de la recherche notamment depuis 2008, avec une série de rachats d'éditeurs par des acteurs majeurs. Cette tendance s'inscrit clairement dans une stratégie de diversification :

- Rachat par Microsoft du norvégien Fast Search and Development en 2008
- Rachat par Dassault Systèmes du français Exalead en 2010
- Rachat par Hewlett-Packard de Autonomy en 2011
- Rachat par Oracle d'Endeca en 2011
- Rachat de Vivisimo par IBM en mai 2012

Ces phénomènes de concentration et de consolidation, qui concernent des acteurs majeurs, présentent plusieurs avantages : une pérennité des offres, une qualité support technique accrue et une vision stratégique à long terme. En parallèle, il existe une myriade de petites entreprises, plus fragiles financièrement mais aussi plus agiles, qui fournissent des outils de plus en plus spécifiques ou qui développent en partenariat avec des leaders du marché des outils de recherche [33, Whit].

#### 6.1.2 Classification des éditeurs de solutions

Le marché des solutions de recherche en entreprise est un marché relativement restreint et segmenté répondant à des besoins très spécifiques. On peut classer les éditeurs en trois catégories : les éditeurs de recherche spécialisés, les éditeurs « universels » et enfin les éditeurs de recherche intégrés<sup>59</sup> [32, Owens].

---

<sup>58</sup> L'auteur de la source cite un chiffre de 3 milliards de dollars

<sup>59</sup> Traduction de « specialized search vendors », « detached search vendors » et « integrated search vendors »

Dans le groupe d'éditeurs de recherche spécialisés, on trouve Attivio, Coveo, Endeca, Exalead, Sinequa, PolySpot et Vivisimo. Ils commercialisent de manière autonome leurs solutions et perçoivent clairement la recherche comme une stratégie en réponse à une problématique d'entreprise spécifique : CRM<sup>60</sup> dans une optique de support consommateur pour Coveo et Visisimo, applications métier de BI et SBA<sup>61</sup> pour Polyspot et Exalead et enfin fonctionnalités sémantiques pour Sinequa.

Le groupe d'éditeurs « universels » inclut Google, ISYS, et Fabasoft. Ils misent sur la simplicité de mise en œuvre et de déploiement, ce qui constitue souvent pour les décideurs un argument majeur. Dans le cas de Google, avec sa plateforme « Google Search Appliance », il y a clairement une volonté de capitaliser sur la marque Google qui domine la recherche sur le web<sup>62</sup>.

La catégorie des éditeurs de recherche intégrés inclut des éditeurs tels que Microsoft, IBM et Autonomy. Les fonctions de recherche sont associées à des fonctionnalités de gestion de l'information, par exemple une plateforme de gestion de contenus ou de travail collaboratif [31, Martin]. C'est notamment le cas pour Microsoft avec sa solution Sharepoint. Les spécificités et les fonctionnalités de recherche de Sharepoint, dans sa version 2010, seront couvertes dans la quatrième partie de ce mémoire.

En marge des solutions propriétaires citées ci-dessus, les solutions open source telle que Lucene SolR, supportée par la fondation Apache<sup>63</sup>, sont davantage des « boîtes à outils » que des solutions à part entière. Toutefois, des éditeurs comme LucidWorks, Polyspot ou Antidot intègrent cette technologie de façon partielle à leur moteur [30, Lefassy].

## **6.2 Stratégies et limites de la communication des éditeurs**

Il existe deux types d'approches chez les éditeurs : l'approche techniciste et l'approche simplificatrice. Ces deux approches entretiennent une certaine ambiguïté autour des fonctionnalités réelles ou supposées de leurs solutions.

Comme le souligne Leslie Owens<sup>64</sup>, les éditeurs ont tendance à sur-intellectualiser leurs offres en proposant des « tableaux de bord décisionnels, conceptuels et optimisés » alors que les organisations et leurs employés souhaitent tout simplement une application de recherche [32, Owens]. Dans le même temps, ils présentent leurs applications comme des

---

<sup>60</sup> Customer Relationship Management : en français « gestion de la relation client »

<sup>61</sup> Search Base Application

<sup>62</sup> En décembre 2012, 65 % des parts de marché mondiales pour Google, 8% pour Baidu et 5% pour Yahoo - Source : Le journal du net, 2012

<sup>63</sup> <http://www.apache.org/>

<sup>64</sup> Directrice de recherche et analyste au cabinet Forrester

solutions "clé en main" nécessitant très peu de paramétrage, facile à appréhender et entretiennent le mythe du « plug & play<sup>65</sup> » ou du « easy-to-use<sup>66</sup> ».

Il est vrai que le déploiement d'un moteur de recherche s'avère moins coûteux comparativement à d'autres projets informatiques (ECM, ERP,...), à la fois en temps et en investissement financier. La durée de déploiement est raisonnable et l'impact sur l'infrastructure informatique existante est modéré [47, Jung]. Toutefois, certains aspects complexes qui doivent retenir l'attention ne sont quasiment pas abordés par les éditeurs : l'analyse préalable du volume et de la structure des informations aussi bien externes qu'internes à interroger, ainsi que le paramétrage et la gestion de la sécurité qui peut être un sujet problématique notamment dans le cas d'une recherche dite « fédérée ».

Cette approche paradoxale, à la fois simplificatrice et sur-intellectualisée conduit les différents éditeurs à surestimer certains bénéfices des moteurs de recherche d'entreprise (par exemple la cartographie des connaissances) et à minorer certains aspects techniques et potentiellement problématiques de leur mise en œuvre (par exemple la recherche fédérée).

### **6.2.1 La cartographie des connaissances et des expertises**

Comme le souligne Eric Debonne, il existe trois grands cas d'utilisation des moteurs de recherche d'entreprise : l'utilisation classique dans le cadre de la recherche (repérer et restituer documents et informations), l'utilisation dans le cadre de l'aide à la décision (produire des indicateurs pertinents de décision) et enfin l'utilisation dans le cadre de la cartographie des connaissances et de la mise en valeur des expertises. Cette dernière utilisation a tendance à être survalorisée par les éditeurs, notamment l'éditeur Antidot<sup>67</sup> qui fonde toute une partie de sa communication sur cet axe.

Or, pour François Bourdoncle<sup>68</sup>, la masse des informations produites en entreprise ne sont pas aujourd'hui réellement exploitables pour faire émerger une « intelligence collective » fondée sur des technologies qui permettent, à partir de corpus plus ou moins vastes et hétérogènes de données textuelles non structurées, d'extraire du « sens », ou en tout cas, une « certaine connaissance » [55, Bourdoncle].

D'autre part, la connaissance (comme la pertinence) est excessivement personnelle [56, Balmissé]. Elle repose en grande partie sur la capacité d'un utilisateur ou d'un groupe d'utilisateurs à donner un sens aux données pour les transformer en connaissances. Quant à

---

<sup>65</sup> Peut être traduit par : « On branche et ça fonctionne »

<sup>66</sup> Peut être traduit par : « Facile à utiliser »

<sup>67</sup> <http://blog.antidot.net/wp-content/uploads/2012/10/LinkedEnterpriseData-WP-fr-v2.2.pdf>

<sup>68</sup> Cofondateur et actuel Directeur Technique d'Exalead (aujourd'hui filiale de Dassault Systèmes)

la mise en valeur des expertises, elle est fortement tributaire du mode d'organisation de l'entreprise, c'est-à-dire plus ou moins ouverte au partage en interne du savoir. Dans ce contexte, le moteur peut être considéré comme une « passerelle » technique vers des données, qu'il regroupe et contextualise avec plus ou moins de pertinence [23, Hérigault].

## 6.2.2 Les pièges de la recherche fédérée

La recherche fédérée est une fonctionnalité proposée par de multiples éditeurs. Mais les enjeux techniques et les difficultés qu'elle engendre, tant au niveau de l'interrogation que de la présentation des résultats ne sont pratiquement jamais évoqués dans les supports de communication des éditeurs.

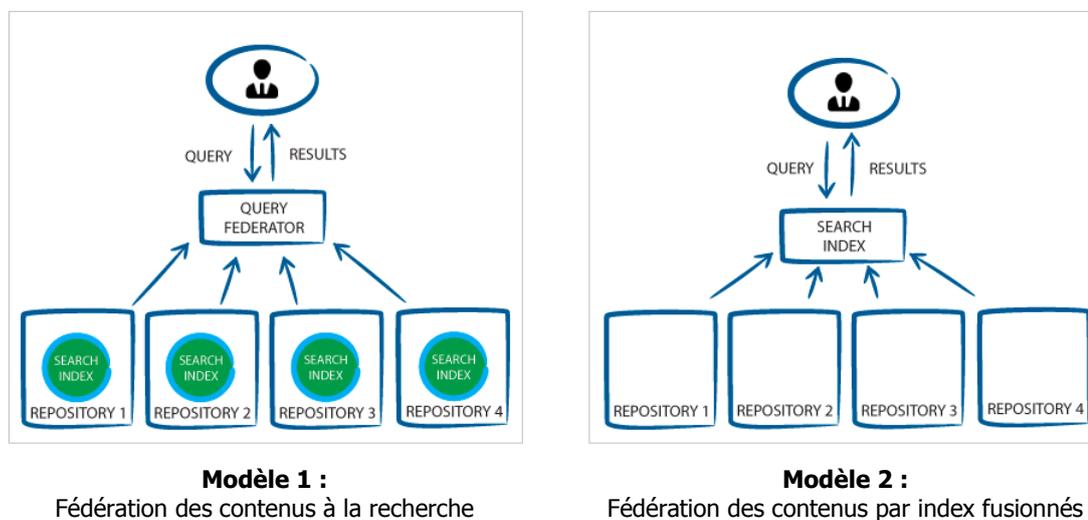


Figure 5 : Les deux modèles de la recherche fédérée [42, Search Technologies]

Le premier modèle repose sur une fédération à la recherche. Le module de fédération fusionne les réponses des différents index et les présente dans une seule liste unifiée. Ce modèle présente l'inconvénient d'être tributaire d'index externes et potentiellement lents à l'interrogation.

Le second modèle repose sur la fédération lors de l'indexation. Un index central facilite l'agrégation des résultats et leur classement mais il est coûteux en acquisition des données et en bande passante.

Cette fédération de contenus comporte plusieurs aspects problématiques qui doivent être anticipés : les risques de doublons, le développement de connecteurs spécifiques, la gestion des éventuels droits d'accès différenciés, la gestion des différents algorithmes de pertinences. D'autre part, cette approche de fédération de contenu se heurte à un problème de taille : l'utilisateur se retrouve souvent devant une longue liste hétérogène de documents complètement décontextualisée sans grand rapport avec son besoin informationnel.

## 6.3 Les positionnements innovants

De plus en plus, les éditeurs présentent leurs solutions comme des plateformes d'intermédiation centrées sur les besoins informationnels et décisionnels des utilisateurs. Deux tendances illustrent ces nouveaux positionnements.

### 6.3.1 Les plateformes d'accès unifié à l'information

Ce nouveau paradigme d'accès à l'information constitue un positionnement fort de certains éditeurs notamment les éditeurs français Polyspot<sup>69</sup> et Exalead<sup>70</sup>. Cette approche tente de résoudre, du moins en partie, les problèmes créés par la recherche fédérée (voir section [6.2.2](#)).

L'accès unifié consiste à fournir une plateforme de recherche dite « ubiquitaire » dans le sens où elle agrège les données des bases de données situées dans des entrepôts de données (« data warehouse ») et le contenu non structuré et semi-structuré qui se situe dans les systèmes de gestion de contenu de l'entreprise ou les serveurs d'email par exemple. L'accès unifié à l'information est rendu possible par une homogénéisation de ces données dans un seul index et une même interface de recherche.

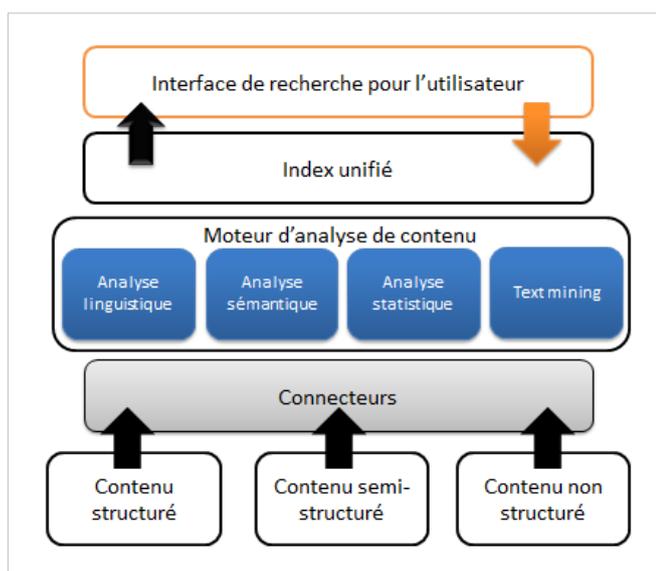


Figure 6 : Infrastructure de l'accès unifié à l'information [39, Freyermuth]

Le schéma ci-dessus présente l'architecture type utilisée par les éditeurs de solution proposant un accès unifié à l'information : Des connecteurs (voir section [4.2.1](#)) permettent de capturer le contenu de sources de données hétérogènes. Le moteur d'analyse de contenu permet un enrichissement des données en ajoutant des métadonnées, en créant des

<sup>69</sup> <http://www.polyspot.com/>

<sup>70</sup> <http://www.3ds.com/fr/products-services/exalead/>

catégories et des classifications grâce notamment aux technologies de TAL (voir section [4.4.1](#)).

Cette logique d'agrégation de contenu permet la création d'applications spécifiques qui représentent « la cristallisation de la logique de l'accès unifié à l'information » : les « Search Base Applications » ou SBA [39, Freyermuth].

### **6.3.2 Les Search Base Application ou Applications orientées recherche**

Dans l'informatique décisionnelle traditionnelle, on construit des applications spécifiques à partir de « datamarts » (littéralement des magasins de données) permettant de constituer des bases de données spécifiques à un contexte métier. Ces applications interrogent des données structurées pour retrouver une information ou afficher un jeu de données spécifiques. Ce fonctionnement implique que l'utilisateur connaisse par avance la source, la façon dont l'information y est structurée et la syntaxe à utiliser pour formuler des requêtes [41, Formosa].

Au croisement des technologies d'indexation et des technologies Web, les SBA représentent la convergence de la recherche et de l'informatique décisionnelle [39, Freyermuth]. Contrairement aux applications qui sont développées sur des bases de données structurées, les SBA sont créées à partir d'un index de moteur de recherche. Celui-ci agrège dans un index unifié les données structurées contenues dans les bases métiers de l'entreprise aux données non structurées dans des applications de gestion de contenu (GED par exemple). Le moteur de recherche devient le cœur sur lequel repose le développement d'applications personnalisables en fonction de contextes métiers aux besoins variés et orchestrées par des requêtes en langage naturel [23, Hérigault].

Les applications développées sur la base de cet index utilisent la technologie du « mashup » qui agrège potentiellement plusieurs applications et plusieurs jeux de données (structurés et non structurés) pour créer une visualisation combinée sous la forme de tableaux de bords personnalisés [41, Formosa]. L'agilité dont font preuve les SBA par rapport aux applications de décisionnel classique permet :

- d'accéder aux informations de l'entreprise que tout le monde utilise sans connaissance préalable de la source
- de retrouver toute l'information pertinente en une seule requête en combinant données structurées et non structurées
- d'augmenter la compréhension des chiffres en les contextualisant à l'aide de tableaux de bord

## 7 Choix de la solution

---

Lors de la sélection de l'outil, quelques paramètres nécessitent d'être correctement appréhendés. Leur choix dépend, entre autres, des applications utilisées en interne, de la volumétrie et de la structure des données, des fonctionnalités requises, de la gestion de la sécurité et du type de projet envisagé.

La recherche d'information en entreprise est une problématique complexe qui repose sur trois éléments fondamentaux : les individus et leurs besoins informationnels, l'information au centre de ces besoins, et les technologies qui vont permettre cette recherche d'information [35, Balmisse].

### 7.1 Les critères de vigilance pour l'organisation

#### 7.1.1 L'étude des besoins utilisateurs

L'analyse des besoins consiste à bien cerner les besoins informationnels des individus par segment de population et en fonction de leurs attentes respectives : nombre, répartition géographique, besoins fonctionnels, ergonomie et multilinguisme. Cette étape peut également questionner l'organisation sur l'usage potentiel que feront les utilisateurs de la future application au travers de trois aspects complémentaires mais distincts :

- **Catégorisation** : cibler une ou plusieurs catégories d'information de manière spécifique : mise en avant de caractéristiques informationnelles particulières (fraîcheur de l'information, criticité, valeur stratégique, popularité).
- **Anticipation** : Envisager le futur modèle de pertinence<sup>71</sup> utilisé par le moteur de recherche en ayant une vision globale des flux d'information transversaux dans l'entreprise
- **Personnalisation** : répondre à des besoins individuels clairement identifiés (valorisation de contenus métiers spécifiques, valorisation des expertises, valorisation de la connaissance et transferts des savoirs)

Or, selon l'enquête annuelle "Global Intranet Trends 2010", seulement 10% des personnes interrogées ont été consultées dans la phase d'analyse et de définition des besoins et seulement 13% ont été sollicités pendant la phase de déploiement pour conduire des tests d'utilisabilité [44, White].

---

<sup>71</sup> La grande majorité des moteurs exploitent un modèle de pertinence (basé sur un calcul de similarité entre la requête et les documents) qui exprime en pourcentage la distance entre la requête et les termes d'index représentant le document et non le document lui-même [36, Balmisse] .

### 7.1.2 L'audit documentaire

Cette phase d'audit consiste à cerner le périmètre et les caractéristiques des documents et des informations à indexer en vue de leur future interrogation : volumétrie, format, nature structurée ou non structurée, gestion homogène et transversale des droits d'accès, temporalité d'indexation des sources (actualisation, validation), rythme de croissance informationnelle des sources internes.

En effet, la difficulté de trouver de l'information pertinente est souvent proportionnelle à la croissance des sources de données [43, Search Technologies]. Il est également important d'identifier et de recenser les applications qui hébergent ces données, et notamment les applications à grande couverture fonctionnelle (comme SAP et Oracle) car leur indexation est plus complexe [23, Hérigault].

### 7.1.3 Un moteur spécifique pour un type projet spécifique

Les critères à prendre en compte dans le choix la solution sont, soit indépendants de l'outil (secteur d'activité de l'entreprise, profil des utilisateurs, problématiques métiers pointues), soit spécifiques à l'outil (modèle commercial de l'éditeur, socle technologique, qualité des connecteurs) soit spécifiques au type de projet mis en œuvre (accès unifié à l'information, partage des connaissances, recherche basique, Business Intelligence et décisionnel).

#### 7.1.3.1 Segmentation fonctionnelle des solutions

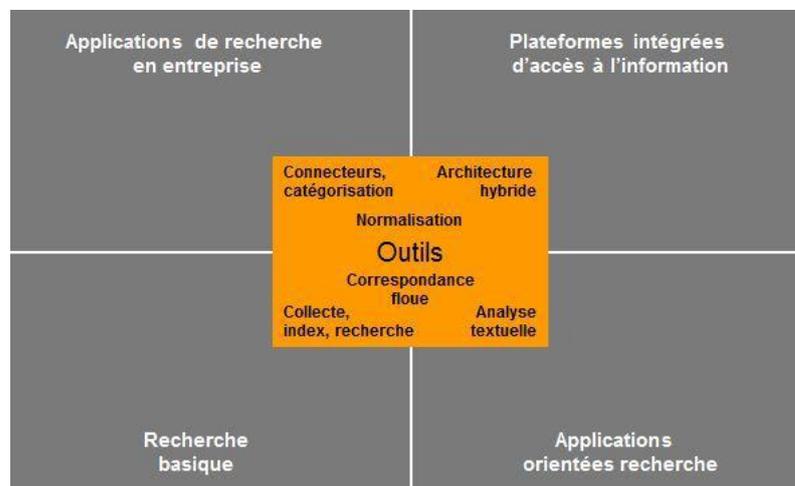


Figure 7 : Typologie des moteurs de recherche [39, Freyermuth]

Ce schéma présente une matrice issue d'une classification très opérationnelle élaborée par le cabinet IDC<sup>72</sup> en 2008. Elle permet de distinguer quatre pôles de différenciation majeurs

<sup>72</sup> Cette classification présente l'avantage d'être moins généraliste que celle proposée par le cabinet Forrester (voir section 6.1.2). Elle permet d'appréhender les différentes offres en fonction de leur segmentation fonctionnelle et non pas marketing. <http://www.idc.fr/>

entre les différentes solutions du marché et peuvent servir de critères de sélection en fonction du projet envisagé et des fonctionnalités attendues. La zone de recouvrement centrale recense les applications complexes et les technologies additionnelles comme la catégorisation, l'analyse de texte, l'accès multilingue, les outils décisionnels, les workflows<sup>73</sup> ou encore les outils d'analyse et de data visualisation [39, Freyermuth]. [45]

#### **7.1.3.2 Les moteurs de recherche basiques**

D'un coût financier modéré, ces solutions proposent un faible niveau d'intégration avec les autres applications de l'entreprise ainsi qu'un nombre limité de documents indexés et de formats pris en charge. Ces solutions présentent l'avantage d'être peu complexes à installer et d'être peu intrusives pour les utilisateurs (pas ou peu d'apprentissage)

#### **7.1.3.3 Les applications de recherche en entreprise**

Comparativement aux moteurs basiques, ces solutions prennent en charge un périmètre d'indexation beaucoup plus large en fédérant de multiples sources d'information et constituent un moyen rapide et non invasif de rechercher des informations dans plusieurs applications et répertoires au sein de l'entreprise. Elles offrent des niveaux de paramétrage et d'intégration élevés (catégorisation, texte mining, aide à la navigation, algorithmes de pertinence paramétrables).

#### **7.1.3.4 Les applications orientées recherche (« Search-based applications »)**

Ces outils décisionnels (voir section [6.3.2](#)) s'intègrent dans la tendance actuelle « qui montre une certaine maturité du marché et une perception par les éditeurs que les tâches impliquant un grand volume d'information ont besoin de meilleurs outils pour être productif et minimiser le risque de mauvaises décisions » [39, Freyermuth].

#### **7.1.3.5 Les plateformes intégrées d'accès à l'information (« Integrated information access and management platforms »)**

Cette catégorie réunit les solutions d'accès unifié à l'information (voir section [6.3.1](#)). Ces outils intégrés incluent une large gamme de connecteurs, gèrent un volume important de données et intègrent des outils d'informatique décisionnelle, des outils d'enrichissement automatique de contenu, des outils d'analyse linguistiques et sémantiques. Ces outils nécessitent toutefois une expertise en interne pour assurer leur déploiement et leur gestion.

#### **7.1.3.6 Les outils d'accès à l'information (« Information access tools »)**

Ces outils exploitent, homogénéisent et enrichissent les données en leur donnant du « sens » en fonction d'un contexte particulier pour en faciliter leur recherche et leur exploitation. Ces outils intègrent des connecteurs « intelligents », des outils de construction

---

<sup>73</sup> Flux de travail : représentation d'une suite de tâches ou d'opérations effectuées par une personne, un groupe de personnes ou une organisation

de taxonomie, des outils de data visualisation, des outils d'extraction automatique de métadonnées et des outils d'analyse linguistique. Un bon exemple de ces technologies est l'offre proposée par Temis<sup>74</sup>.

## **7.2 Une équipe projet dédiée**

### **7.2.1 Pour mettre en place la solution**

L'équipe chargée de la mise en place est à la fois assez resserrée et multi disciplinaire, reflétant les deux dominantes organisationnelles et technologiques représentatives de la mise en œuvre d'une solution de recherche [46, Intranet Focus] :

- Un chef de projet ou « Search manager »,
- Un gestionnaire informatique ou « Search Technology Manager »
- Un spécialiste de l'information ou « Information Specialist » (autrement dit un documentaliste).
- Un analyste en charge des logs de consultation ou « Search Analytics Manager »
- Un responsable du support utilisateur ou « Search Support Manager »

Il est bien évident que cette équipe représente « un idéal » que beaucoup d'organisations ne pourront, faute de moyens financiers suffisants et de ressources internes, mettre en place. Il est fort probable que dans la plupart des cas, ce soit une personne qui agrège tous ces rôles. Mais cette classification proposée par Martin White<sup>75</sup> a le mérite de définir de façon très opérationnelle les contours d'une équipe pour gérer de manière transversale et efficace un projet d'envergure.

Elle joue un rôle de médiateur entre les utilisateurs et l'éditeur (ou l'intégrateur de la solution) et veille à la mise en place de solutions de recherche d'information permettant de faciliter à la fois la mise en contexte de l'information et les interactions qui ne devront pas se faire au détriment de la satisfaction utilisateur [35, Balmissé].

### **7.2.2 Pour maintenir les performances du moteur dans le temps**

En phase d'exploitation, l'équipe doit veiller à l'accompagnement des utilisateurs ainsi qu'à l'enrichissement des référentiels et à l'évolution du périmètre d'indexation des sources.

Dans cette phase, le rôle du gestionnaire de l'information est essentiel, notamment par sa connaissance des métadonnées et des référentiels qui régissent la classification et la catégorisation automatique des contenus au sein de l'organisation. Il pilote également

---

<sup>74</sup> <http://www.temis.com/fr>

<sup>75</sup> <http://www.intranetfocus.com/about/martin-white>

l'enrichissement des référentiels linguistiques afin de répondre au mieux à des objectifs clairement identifiés, et ainsi mettre en place des processus d'indexation valables en alignant le management du moteur avec les besoins de l'organisation et de ses utilisateurs.

Cette problématique est à mettre en perspective avec la qualité des métadonnées et au coût éventuel de leur gestion et de leur enrichissement. Comme évoqué précédemment (voir section [4.4.3](#)), des métadonnées faibles, incohérentes ou carrément inexistantes rendent le moteur inefficace.

Quatrième partie

**Le déploiement de Sharepoint  
2010 et du moteur Sharepoint  
Search Server chez Coe-Rexecode**

## 8 Contexte, périmètre et objectifs de la mission

---

### **8.1 Coe-Rexecode, un institut d'analyse macro-économique**

Coe-Rexecode<sup>76</sup> est né de la fusion en octobre 2006 de deux instituts d'études économiques cinquantenaires : Rexecode et le Centre d'Observation Economique de la Chambre de Commerce et d'Industrie de Paris Ile-de-France<sup>77</sup>. Coe-Rexecode est également représenté dans la plupart des instances françaises de politique économique (Conseil d'analyse économique, Commission économique de la nation, Conseil Économique et Social, etc.) ou professionnelles (CCI Ile-de-France, Groupe des Fédérations Industrielles). La mission principale de Coe-Rexecode est d'assurer l'information économique conjoncturelle et statistique des entreprises et associations professionnelles qui la financent via des réunions mensuelles, des publications régulières et des services exclusifs. Cette offre de service passe notamment par la production d'un ensemble de publications économiques périodiques, ainsi que d'indicateurs originaux, tant macroéconomiques que sectoriels ainsi que la fourniture de données statistiques sous forme de tableaux de bord et de graphiques.

Constitué en association, Coe-Rexecode compte environ 80 organisations adhérentes ou clients (entreprises, institutions financières, organisations professionnelles ou administrations publiques). Coe-Rexecode a deux activités complémentaires :

- la veille conjoncturelle et les prévisions portant sur l'économie mondiale, les principales zones développées et émergentes, les marchés de matières premières et les marchés financiers.
- la participation au débat de politique économique, notamment sur toute question économique concernant le système productif et les entreprises.

Cette politique de diffusion et production régulière de documents économiques s'illustre par la mise en ligne hebdomadaire sur un extranet de publications en accès restreints (adhérents) ou en accès public (journalistes, autres instituts d'analyse économique) :

- une synthèse conjoncturelle (bilan de la semaine) : Document de 4 pages uniquement accessibles aux adhérents
- un cahier de graphiques détaillés et spécifiques (commentés par un économiste) : une synthèse en 10 graphiques statistiques conjoncturels accessibles en profil adhérents et public

---

<sup>76</sup> <http://www.coe-rexecode.fr/public>

<sup>77</sup> <http://www.cci-paris-idf.fr/>

- une veille documentaire et conjoncturelle sous la forme de 2 produits : un bulletin de veille hebdomadaire pour les adhérents et une synthèse sur un document important paru dans la semaine, accessible en profil adhérents et public.

Des publications plus événementielles ou plus spécifiquement liées à une demande client peuvent être également produites.

### **8.1.1 La problématique informationnelle chez Coe-Rexecode**

Fin 2012, l'extinction de l'ancien Intranet et le basculement total vers un ECM a marqué une évolution majeure dans la politique de gestion de l'information au sein de la structure. Après avoir comparé différentes solutions, Nuxeo<sup>78</sup> et Alfresco<sup>79</sup> notamment, le choix s'est porté sur un outil propriétaire : Sharepoint 2010 de Microsoft.

Envisagée depuis la fin de l'année 2011, cette transition d'une logique de serveur classique vers un ECM qui centralise toutes les ressources informationnelles de l'organisation répond à plusieurs impératifs organisationnels et techniques :

- Une migration nécessaire de la base de données documentaire sous le logiciel AIRS, devenue obsolète, vers Microsoft ACCESS afin d'assurer une meilleure intégration avec Sharepoint
- « Désenclaver » la base de données documentaire en la rendant accessible à tous les économistes et statisticiens
- Améliorer la circulation des informations et des connaissances piégées dans des silos et éviter les contenus doublonnés et redondants
- Augmenter la visibilité du pôle documentaire et renforcer son positionnement (tant à l'interne qu'à l'externe).
- Offrir de nouvelles prestations en diversifiant l'offre documentaire
- Accroître la portée de la base documentaire et améliorer la gestion des flux de données

Le pilotage du projet a été confié au pôle documentation et plus spécifiquement à la responsable des systèmes d'information et de la documentation. Le projet est soutenu activement en interne par la direction.

---

<sup>78</sup> <http://www.nuxeo.com/fr/produits>

<sup>79</sup> <http://www.alfresco.com/fr>

## 8.2 Un système d'information segmenté

Coe-Rexecode comprend trois directions outre les services généraux. Afin de répondre aux sollicitations des organisations adhérentes, l'activité d'analyse économique et statistique est structurée autour de 4 pôles complémentaires :

- le pôle conjoncture
- le pôle études
- le pôle statistique (fonction support)
- le pôle documentation (fonction support)

Chaque économiste (au total 17 collaborateurs de Coe-Rexecode) suit une zone géographique déterminée et une thématique particulière :

- **Les économistes du pôle conjoncture** suivent les politiques de retraites et d'épargne, la compétitivité et la politique économique, la politique industrielle et industrie de façon transversale ainsi que les politiques de financement des collectivités locales.
- **Les économistes du pôle études** suivent le développement des entreprises de services, les politiques du logement, l'industrie pharmaceutique et les politiques de santé, l'économie numérique ainsi que la fiscalité verte et les politiques environnementales.
- **Les pôles documentation et statistique** interviennent de manière transversale en fonctions supports pour les économistes en ajustant leur offre de services suivant les besoins ponctuels ou les orientations sur le long terme de l'organisation.

### 8.2.1 Des besoins d'informations différenciés

Coe-Rexecode est une structure de 25 personnes avec des profils métiers assez différenciés. Un organigramme de la structure est proposé en **annexe 1**. Des complémentarités existent entre les 3 pôles d'analyse économique (études, conjonctures et statistiques), mais les besoins et les sources d'information sollicitées sont très différents :

- Le pôle perspectives produit avec une grande régularité des documents de travail, des analyses et des publications récurrentes pour l'extranet client et est très demandeur d'informations conjoncturelles.
- Le pôle modélisation produit des études plus ciblées et techniques. Le besoin d'information est pointu et les économistes du pôle font très souvent appel à la documentation pour des revues de littérature.
- Le pôle statistiques irrigue les pôles perspectives et modélisation et met à jour des séries à partir de sources externes, soit manuellement, soit de manière procédurale

## 8.3 Le déploiement de Sharepoint 2010

### 8.3.1 Une réponse à des besoins métiers spécifiques

En décembre 2012, le pôle documentaire en charge des systèmes d'information chez Coe-Rexecode a procédé à l'extinction de l'ancien Intranet documentaire (conformément au calendrier prédéfini) en basculant vers Sharepoint Server 2010. Le passage vers un outil collaboratif de gestion de contenu avec des capacités de recherche étendues répond à plusieurs objectifs :

- **Accès à l'information :** fédérer les ressources informationnelles, faciliter la recherche d'information dans ces ressources et améliorer leur visibilité au sein d'une plateforme unique
- **Travail collaboratif :** faciliter les processus de partage d'informations et de coédition de documents et ainsi favoriser une certaine transversalité entre les pôles (notamment entre les analystes et les statisticiens)
- **Stratégie de marketing documentaire :** répondre à une nouvelle stratégie d'offre de services définie par le pôle documentation qui soit en meilleure adéquation avec les besoins des utilisateurs (suivre les pratiques documentaires des utilisateurs et, le cas échéant, ajuster les services avec une plus grande réactivité)

## 8.4 Un bilan en demi teinte

Pourtant, 6 mois après le déploiement de Sharepoint, le bilan est mitigé et les remontées successives des utilisateurs ont permis de faire les constats suivants :

- il existe chez certains utilisateurs des réactions d'incompréhension voire de rejet vis-à-vis du nouvel outil car jugé trop complexe
- une partie des utilisateurs exprime son insatisfaction concernant les modalités d'accès à l'information (arborescence des sites et sous-sites de Sharepoint, fonctionnalités de recherche, difficultés à retrouver l'information pertinente)
- l'absence de période de transition entre l'ancien intranet et la mise en service de Sharepoint a été vécue par certains utilisateurs comme une « cassure » (malgré le travail de communication mené en amont par le pôle documentation)
- une partie des collaborateurs de Coe-Rexecode n'utilise jamais ou presque jamais le nouvel outil
- l'usage de certaines fonctionnalités et services proposés par Sharepoint est inexistant ou limité car jugés trop compliqués ou inefficaces : fonctionnalités de recherche, partage et coédition de documents

Ces observations ont permis de mettre en évidence un certain nombre de dysfonctionnements réels et supposés. Il a été décidé de façon conjointe avec la responsable des services d'informations et avec la direction de Coe-Rexecode de réaliser, dans un premier temps, un audit des besoins afin de définir dans un deuxième temps des axes d'amélioration notamment autour des fonctionnalités de recherche dans Sharepoint.

## **8.5 Une mission menée en mode projet**

Très tôt dans la définition des objectifs et du périmètre de la mission, la nécessité de mettre en place une méthodologie projet est apparue évidente. Cette méthodologie a permis de :

- s'assurer que les objectifs sont clairs et bien compris de tous les intervenants
- organiser la communication entre les acteurs du projet
- planifier les actions à entreprendre pour atteindre les objectifs
- chiffrer les ressources humaines nécessaires à l'accomplissement des tâches du planning
- analyser les éléments qui pourraient freiner l'atteinte des objectifs
- suivre le déroulement du projet et réajuster éventuellement les objectifs à atteindre
- faire un bilan pour tirer des enseignements du projet

### **8.5.1 Des outils de suivi et de capitalisation**

Afin de mener à bien cette mission et pleinement satisfaire aux exigences du pôle documentation, la mise en place d'une série d'outils de suivi et de reporting afin de capitaliser à toutes les étapes du projet s'est avérée nécessaire. Ces outils sont à la fois quantitatifs et qualitatifs [05, Marchat] :

**Une note de cadrage** validée par la responsable des systèmes d'information en début de projet pour définir le périmètre, les objectifs et les attendus de la mission ainsi que les différents livrables (voir section [8.5.2](#))

**Un plan d'action** afin de suivre les actions menées en fonction de leur priorité (haute, moyenne ou basse), de leur état d'avancement (en cours, réalisées ou reportées) ou de leur nature (livrable, suivi de projet, communication, paramétrage)

**La mise en place d'outils de planification :**

- Un macro planning pour estimer grossièrement la charge projet (temps et ressources sollicitées) (voir **annexe 2**)
- Un planning détaillé réévalué en cours de projet pour gérer les actions, les différentes phases du projet et les ressources sollicitées

### La mise en place d'un tableau de bord pour :

- la gestion des risques projet avec une pondération en fonction de leur impact, de leur probabilité et de leur criticité :
  - ⇒ Risques internes : liés à la méthodologie projet ou à des problématiques techniques
  - ⇒ Risques externes : liés au contexte organisationnel
- mesurer l'avancement du projet à l'aide d'une check list
- gérer les documents édités pendant la mission (comptes rendus de réunion, livrables, documents intermédiaires ainsi que leur versioning)

### 8.5.2 Les livrables attendus et leurs objectifs

Le tableau ci-après récapitule l'ensemble des livrables et leurs objectifs respectifs :

Intitulé du livrable	objectifs
<b>Livrable n°1</b> - Audit de l'existant	<ul style="list-style-type: none"><li>• Analyser des besoins informationnels des utilisateurs et évaluer leurs attentes, leur perception vis-à-vis de l'outil Sharepoint ainsi que son adéquation en fonction de leur segmentation métier (accès, partage, recherche d'information)</li><li>• Proposer des axes d'amélioration</li></ul>
<b>Livrable n°2</b> - Elaboration d'un référentiel de mots clés pour définir les « meilleurs résultats » de recherche (voir section <a href="#">11.2</a> )	<ul style="list-style-type: none"><li>• Outil de capitalisation : possibilité de s'appuyer sur un référentiel commun de mots clés afin de définir les meilleurs résultats de recherches et favoriser la remontée optimale des pages et bibliothèques de documents pertinents</li><li>• Outil d'amélioration continue du moteur de recherche : ajuster la pertinence des mots clés et synonymes en fonction des remontées utilisateurs et des indicateurs quali/quantitatifs de Sharepoint (analyse des logs de recherche, entretiens individuels)</li></ul>
<b>Livrable n°3</b> - Benchmark des solutions de moteur de recherche d'entreprise et solutions sémantiques	<ul style="list-style-type: none"><li>• comparatif des différentes versions de Sharepoint à travers leurs fonctionnalités de recherche respectives</li><li>• gains et limites des solutions retenues</li><li>• repères méthodologiques concernant les moteurs de recherche d'entreprise</li></ul>
<b>Livrable n°4</b> - bilan de fin de mission <sup>80</sup>	<ul style="list-style-type: none"><li>• recenser les actions menées durant la mission ainsi que les livrables produits et lister les préconisations pour l'optimisation future de la recherche</li></ul>

Tableau 4 : Listing des livrables attendus

<sup>80</sup> Le livrable n°4 n'a pas été listé dans la première version de la note de cadrage. En revanche, il l'a été dans la version finale de cette dernière.

## 9 Recueil des besoins utilisateurs

---

L'objectif de cette étude des besoins utilisateurs étant de recueillir des éléments nouveaux, le choix a été fait de l'étendre à l'ensemble des collaborateurs de Coe-Rexecode (tous pôles confondus) et de mobiliser uniquement un intervenant extérieur à la structure afin de limiter « l'auto-censure » et de bénéficier d'une meilleure prise de recul.

### 9.1 Méthodologie de l'analyse des besoins

Afin de préparer cette analyse de besoin, des entretiens exploratoires ont été réalisés auprès de 3 collaborateurs de Coe-Rexecode : un analyste du pôle conjoncture, un analyste du pôle études et un statisticien. Cette phase exploratoire a été nécessaire pour avoir une connaissance fine du rôle de l'information documentaire pour le public cible, identifier les segments de public en fonction des contraintes métier et enfin évaluer les profils d'activité des utilisateurs [19, Anobis].

Une approche qualitative par entretiens semi directifs individuels a ensuite été privilégiée. Un guide d'entretien a été élaboré à partir des éléments recueillis pendant la phase exploratoire. Cette trame d'entretiens s'appuie sur 4 grandes entrées thématiques et leurs sous thématiques associées (voir **annexe 3** pour le détail):

- APPRECIATION GENERALE DE SHAREPOINT
  - ⇒ Intégration dans le processus métier
  - ⇒ La fonction recherche dans Sharepoint
  - ⇒ Utilisation / fréquentation globale de l'outil
- TRAVAIL COLLABORATIF
  - ⇒ Intégration dans le processus métier
  - ⇒ Partage des documents dans Sharepoint
  - ⇒ Partage et stockage des documents en dehors de Sharepoint
- ACCES A L'INFORMATION
  - ⇒ La structuration de l'information dans Sharepoint
  - ⇒ La navigation dans les sites Sharepoint
  - ⇒ La visibilité / remontée des informations primordiales
- CHANGEMENTS / ACTIONS CORRECTIVES SUGGEREES
  - ⇒ Dans les processus collaboratifs / partage / accès de l'information
  - ⇒ Dans le fonctionnement / paramétrage actuel de Sharepoint

Chaque personne sollicitée a fait l'objet d'une action de communication personnalisée pour expliciter la démarche et l'objectif final de l'entretien. Cette action a consisté à prévenir par oral la personne et à envoyer un mail accompagné d'une fiche individuelle reprenant les principaux thèmes susceptibles d'être abordés en entretien ainsi que l'estimation du temps nécessaire à la réalisation de ce dernier. Cette action spécifique a permis d'optimiser l'entretien, de le rendre plus efficace et de mettre la personne interviewée plus à l'aise car elle connaissait à l'avance la teneur de l'entretien [06, Marchat].

L'analyse des résultats a consisté à établir un tableau de synthèse en regroupant le contenu des différents entretiens selon les thématiques abordées et en fonction des différents pôles : études, conjoncture, statistique, administratif et documentation.

## **9.2 Résultats des entretiens utilisateurs**

Cette étape de la mission correspond au livrable n°1 (voir section [8.5.2](#)). L'analyse des données collectées à partir des entretiens utilisateurs et des questionnaires a permis de mettre en lumière des tendances très marquées reprenant globalement la segmentation des différents pôles de Coe-Rexecode :

La majorité des utilisateurs du pôle statistique perçoit l'outil comme peu adapté voire même totalement inadapté à leur pratique métier. Sharepoint n'entre pas dans le cadre de la gestion des bases de données statistiques et est donc par conséquent peu utilisé.

Les utilisateurs du pôle administratif considèrent que les informations chargées sur Sharepoint sont globalement « calibrées » pour les économistes et peu adaptées à leurs usages. L'utilisation de l'outil est donc sporadique, à la marge, y compris sur les fonctions collaboratives les concernant directement, comme la gestion et le partage des agendas.

Pour les pôles études et conjoncture, la situation est très contrastée : certains analystes perçoivent clairement le passage à Sharepoint comme un vrai plus, tant d'un point de vue informationnel que qualitatif. D'autres analystes, en revanche, perçoivent Sharepoint comme en « décalage » et sont perturbés par la complexité de l'outil, l'absence de visibilité de certaines informations et l'inefficacité (réelle ou supposée) des fonctionnalités de recherche.

Quant au pôle documentation, la mise en place de l'outil a eu un fort impact sur la pratique journalière. Il est venu se substituer complètement à l'ancien intranet. Sharepoint est perçu comme beaucoup plus chronophage notamment pour les modifications de pages, le chargement et la préparation des documents en vue de leur intégration dans les différentes bibliothèques. Sharepoint est clairement positionné comme un outil de travail collaboratif et de mise à disposition / partage de l'information mais il ne révolutionne pas les pratiques de travail et de partage de l'information en vigueur chez Coe-Rexecode : consultation variable

selon les économistes des produits documentaires, consultation et pratique très variable selon les utilisateurs (du nouvel outil comme de l'ancien intranet).

### **9.2.1 Des tendances fortes et marquées**

**Perception générale de l'outil :** Sharepoint est massivement utilisé (et perçu) comme un outil d'intranet documentaire (capitalisation et stockage de ressources internes ou externes déjà publiées et dans leur version finale) et pas comme un outil très poussé de fonctions collaboratives.

**Les fonctionnalités collaboratives ne sont clairement pas perçues comme une priorité :** C'est en revanche un sujet d'intérêt pour les utilisateurs mais il existe encore peu d'attentes étant donné les circuits de validation et d'approbation tels qu'ils sont structurés actuellement.

**Une sous-utilisation globale de l'outil :** Sharepoint n'occupe pas une place centrale pour les utilisateurs et n'est pas considéré comme un point d'accès unique à l'information produite en interne ou capitalisée à l'externe.

**Une courbe d'apprentissage loin d'être optimale :** L'utilisation de l'outil se fait parfois de façon très sporadique. Certains utilisateurs passent plusieurs jours sans se servir de Sharepoint. Les automatismes de manipulation de l'outil ne sont pas acquis (ex. Où trouver telle ou telle information ? Comment retrouver une fonctionnalité spécifique ?)

**Un intérêt manifeste pour Sharepoint :** Il n'y a pas de rejet franc et massif de l'outil en lui-même mais plutôt des attentes très spécifiques qui ne sont pour l'instant que partiellement comblées ainsi qu'une segmentation forte par pôles.

**Un outil perçu comme peu intuitif et parfois complexe :** Les utilisateurs s'attendent à une ergonomie web avec des réflexes de navigation typiquement web mais Sharepoint ne respecte pas intégralement ces codes. Les problèmes ergonomiques identifiés sont :

- Un logo non cliquable pour revenir à la page d'accueil du portail,
- Un fil d'ariane absent,
- Des liens URL confusants,
- Une difficulté à sortir d'un sous-site pour revenir à la page portail

**Une solution surdimensionnée :** L'outil est très souvent considéré comme trop « générique », parfois trop complet et pas assez dimensionné aux besoins spécifiques des utilisateurs. Le contenu de certaines pages (notamment la page d'accueil du portail) est jugé trop chargé et/ou trop généraliste.

**Des tableaux de bord personnalisés :** Les utilisateurs expriment dans leur grande majorité une très forte demande autour de la personnalisation des contenus via la constitution de pages personnelles « sur mesure » suivant 2 axes :

- En fonction du périmètre opérationnel : segmentation par pôles et segmentation individuelle
- En fonction de critères de pertinence, zones d'intérêt, et sujets suivis (zones géographiques, organismes / institutions, secteurs)

**Une trouvabilité et une accessibilité immédiate à l'information :** Les utilisateurs expriment une véritable frustration lorsqu'ils sont confrontés à la recherche d'informations précises dans Sharepoint. De nombreuses demandes concernent l'amélioration des fonctionnalités de recherche et de la structuration de l'information. Les systèmes simples comme les alertes sont plébiscités ainsi que tout contenu envoyé en push de façon régulière sans avoir à aller chercher l'information.

**Sur-utilisation des mails et du serveur commun pour le partage d'informations :** Certains utilisateurs n'ont pas le réflexe de partager les documents et informations directement dans Sharepoint. L'information transite encore énormément par mail et par le répertoire « COMMUN » (avec des risques de doublons). Cette réalité est encore renforcée, notamment à la statistique, qui considère l'outil comme inadapté à leur pratique métier.

### **9.2.2 La perception des fonctionnalités de recherche**

Le moteur de recherche inclus à Sharepoint 2010 est sous-exploité et délaissé par les utilisateurs. Dans leur grande majorité, ces derniers éprouvent une grande insatisfaction face aux résultats de leurs requêtes. Ils estiment que les résultats, dans la plupart des cas, sont peu pertinents : trop d'informations parasites, des réponses très éloignées de leur requête d'origine ou qui ne correspondent pas à leur demande initiale. Les concepts d'étendues de recherche (voir section [11.1.1](#)), de recherche avancée et de filtres à facette ne sont globalement pas compris.

Le tableau suivant présente de manière synthétique les résultats des entretiens relatifs aux fonctionnalités de recherche et classés par pôles :

<b>Pôle administration</b>	L'usage des fonctionnalités de recherche est très limité. Les utilisateurs éprouvent des difficultés pour appréhender les fonctionnalités proposées par le moteur de recherche. L'outil est perçu comme peu utile et peu adapté aux usages du pôle administration. Ils ne perçoivent pas Sharepoint 2010 et son moteur de recherche comme point d'accès unique à l'information.
<b>Pôle études</b>	<p>Il existe deux visions utilisateurs radicalement différentes et contrastées :</p> <ul style="list-style-type: none"> <li>• Certains économistes estiment que les réponses proposées par le moteur sont pertinentes et en phase avec leur besoin d'information. Le concept d'étendue de recherche et la fonctionnalité de recherche avancée sont connus et utilisés (recherche avec le terme exact et spécification de la zone de recherche).</li> <li>• En revanche, pour certains autres économistes, les quelques requêtes lancées via le moteur ont donné peu de satisfaction. Par conséquent, ils ne sollicitent désormais plus le moteur de recherche et préfèrent passer directement par les documentalistes pour toute recherche d'information.</li> </ul>
<b>Pôle conjoncture</b>	<p>Le moteur de recherche est peu utilisé pour l'instant. Les économistes du pôle conjoncture estiment que les résultats fournis peuvent être pertinents dans d'autres contextes métier. Les réponses aux requêtes sont jugées trop génériques, pas assez ciblées et personnalisées en fonction de leurs besoins métier spécifiques ou de la zone géographique suivie.</p> <p>L'aspect recherche dans Sharepoint est systématiquement contourné. La plupart des utilisateurs n'ont pas le réflexe d'affiner à posteriori leur recherche à l'aide des filtres à facette ou de reformuler leur requête. Même si le document est présent dans la base Sharepoint, ils préfèrent, par souci de temps et d'efficacité, passer par les sites des organismes publics ou privés et capitaliser les informations ou les documents sur leurs postes, en local.</p>
<b>Pôle Statistique</b>	Les utilisateurs du pôle statistique considèrent les fonctionnalités de recherche (tout comme les autres fonctionnalités de l'outil) inadaptées à leur pratique métier, à leurs attentes et à leurs besoins d'informations.
<b>Pôle documentation</b>	Les fonctionnalités de recherche sont perçues comme peu optimisées. Les réponses sont peu pertinentes et ramènent de manière générale beaucoup de bruit. Pour les utilisateurs non spécialistes de l'information, le concept d'étendues de recherche est difficile à appréhender. Ce fractionnement de la recherche est d'autant plus confusant qu'il n'est pas évident de bien cibler la recherche si l'utilisateur n'a pas une connaissance préalable des contenus informationnels capitalisés dans « intranet » ou bien « Intradoc » (voir section <a href="#">11.1.1</a> ).

Tableau 5 : Perception des fonctionnalités de recherche par pôles

## 9.3 Analyse des risques et opportunités

L'analyse des entretiens utilisateurs à travers ses différentes tendances a mené à l'élaboration d'une matrice de type SWOT<sup>81</sup> afin de répondre à plusieurs objectifs opérationnels qui s'inscrivent dans une volonté d'anticipation et de planification :

- extraire les tendances issues de l'analyse des entretiens utilisateurs et les traduire en forces, faiblesses, opportunités et menaces
- identifier les facteurs de risques associés à chaque tendance et les classer en fonction de leur criticité ou de leur importance
- déterminer les axes d'amélioration possibles, notamment concernant les fonctionnalités de recherche et l'appropriation globale de l'outil
- arbitrer et anticiper sur la stratégie à adopter pour le reste de la mission et pour la mise en œuvre réaliste et pragmatique de ces améliorations

### 9.3.1 Utilisation de la matrice SWOT

Différents types de risques peuvent être identifiés : humains (résistance au changement), informationnels (difficultés réelles ou ressenties d'accéder à l'information pertinente), ou encore technologiques (difficultés réelles ou ressenties de manipulation de Sharepoint). Cette matrice SWOT peut se lire ainsi :

- les forces sont les facteurs positifs intrinsèques au projet de déploiement de l'outil Sharepoint
- les faiblesses sont les aspects négatifs internes à Coe-Rexecode (aspects fonctionnels, organisationnels ou humains)
- les opportunités sont les éléments positifs issus des résultats de l'analyse des entretiens utilisateurs ou liées aux possibilités qu'offre Sharepoint
- les menaces sont les obstacles ou problèmes externes à l'outil qui peuvent entraver le développement de la solution à moyen terme

---

<sup>81</sup> En anglais, acronyme de « Strengths, Weaknesses, Opportunities, Threats »

<b>FORCES</b>	<b>FAIBLESSES</b>
<ul style="list-style-type: none"> <li>• L'outil est globalement bien perçu par les utilisateurs : pas de rejet franc et massif.</li> <li>• Solution choisie puissante et offrant un large panel de fonctionnalités personnalisables.</li> <li>• Fonctionnalités sociales et innovantes (outils orienté web et pratiques 2.0).</li> <li>• Equipe porteuse du projet spécialisée dans la gestion documentaire.</li> <li>• Soutien actif et appui opérationnel de la direction (actions de communication et de promotion régulières de l'outil).</li> <li>• Population interne d'utilisateurs avec une capacité d'adaptation à l'outil informatique.</li> <li>• Remontée régulière des besoins métiers et paramétrages spécifiques réalisés en interne (forte réactivité).</li> <li>• Ecoute active des besoins utilisateurs et lancement des actions correctives dans des délais assez courts.</li> </ul>	<ul style="list-style-type: none"> <li>• Outil complexe et peu intuitif : terminologie Sharepoint peu accessible, demande une période d'adaptation plus ou moins longue suivant les utilisateurs.</li> <li>• Navigation et ergonomie qui ne respectent pas entièrement les standards du web (logo non cliquable pour revenir à la page portail, fil d'ariane absent).</li> <li>• Réponses aux requêtes lancées via le moteur de recherche parfois peu pertinentes (trop de réponses pas assez ciblées, pas de critères par pertinence).</li> <li>• Segmentation entre les différents pôles : attentes diverses en termes de besoins d'informations et de fonctionnalités. Pratiques métiers et documentaires très différentes qu'il faut gérer avec un outil jugé comme trop « générique ».</li> </ul>
<b>OPPORTUNITES</b>	<b>MENACES / RISQUES</b>
<ul style="list-style-type: none"> <li>• Nombreux enjeux en termes organisationnels et fonctionnels : Traçabilité des documents, gain de temps, moins de perte de documents, sécurité, rationalisation des outils informatiques.</li> <li>• Transversalité et communication renforcée entre les pôles et amélioration continue des processus métiers (recherche, rédaction, publication online et offline).</li> <li>• Simplification des audits et mise en place possible d'indicateurs sur l'outil pour mesurer la participation (qualitative - quantitative) et établir plus facilement un plan d'actions correctives si besoin.</li> <li>• Mutualisation des compétences, gestion des connaissances favorisée par la mise en place de communautés d'intérêt ou de sites d'équipes dédiés aux projets spécifiques (à moyen terme).</li> </ul>	<ul style="list-style-type: none"> <li>• Outil globalement sous-utilisé : <ul style="list-style-type: none"> <li>⇒ Fonctions collaboratives sous-utilisées : faible appropriation des fonctionnalités avancées de partages et d'édition de documents collaboratifs</li> <li>⇒ Forme de résistance au changement : Démobilisation d'un certain nombre d'utilisateurs face à la complexité ressentie et/ou vécue de l'outil avec risque de "décrochage".</li> </ul> </li> <li>• Faible appropriation des fonctions de recherches (indifféremment simples ou avancées). Difficulté à retrouver les informations considérées comme essentielles par les utilisateurs.</li> <li>• Attentes fortes vis-à-vis de l'outil et parfois pas toujours comblées (risques de déception ou d'incompréhension).</li> </ul>

Tableau 6 : Analyse SWOT du projet de déploiement Sharepoint

### 9.3.2 Choix de la stratégie à mettre en œuvre

Grâce à l'analyse de cette matrice, il est possible d'identifier 4 stratégies<sup>82</sup> pouvant être mises en œuvre [21, Poulet] :

- **La stratégie « Forces – Opportunités »** ou stratégie offensive d'expansion. Il s'agit de mettre l'accent sur les points forts internes pour développer les opportunités.
- **La stratégie « Faiblesses – Opportunités »** ou stratégie défensive d'expansion. L'amélioration des faiblesses identifiées permet d'exploiter au mieux des opportunités.
- **La stratégie « Forces – Menaces »** ou stratégie dite défensive. Elle consiste à développer les forces internes afin de se prémunir au maximum des menaces de l'environnement.
- **La stratégie « Faiblesses – Menaces »** ou stratégie de diversification ou de repositionnement. Elle consiste à réduire au maximum les faiblesses afin d'être moins vulnérable aux menaces.

La stratégie reposant sur le diptyque « forces - menaces » a été privilégiée dans le cadre de ce projet. Il est rapidement apparu que, face à la faible appropriation de l'outil et des fonctionnalités de recherche proposées, il était nécessaire de proposer des axes d'amélioration tangibles répondant de manière pragmatique aux besoins informationnels des utilisateurs. Cette stratégie repose sur trois axes.

1. L'élaboration d'un livrable afin de comparer les différentes versions de Sharepoint 2010 à travers leurs fonctionnalités de recherche respectives : mieux appréhender le périmètre fonctionnel pris en charge par la version actuellement déployée chez Coe-Rexecode, évaluer la pertinence d'implanter une solution de recherche tiers venant se substituer au moteur actuel et en dégageant pour chacune des solutions retenues leurs principaux gains ainsi que leurs principales limites
2. L'évaluation des performances du moteur de recherche actuel à travers l'analyse des logs ainsi que l'analyse des métadonnées pour les principaux formats indexés (pdf, documents office) et leur prise charge par le moteur
3. L'amélioration de la recherche par le paramétrage en backoffice de « meilleurs résultats » (voir section [11.2](#)) afin de faire remonter à la recherche une page ou une bibliothèque spécifique en fonction de mots clés prédéfinis et capitalisés dans un référentiel faisant l'objet d'un livrable

---

<sup>82</sup> Classification inspirée de « Stratégique 9ème édition » JOHNSON Gerry, SCHOLLES Kevan, WHITTINGTON, FRERY Frédéric. Pearson Education, 2011

## 10 Benchmark de solutions alternatives

---

L'analyse des besoins utilisateurs développée dans la partie précédente a permis de mettre en lumière une situation extrêmement contrastée ainsi qu'une appropriation incertaine de Sharepoint 2010 et plus spécifiquement des fonctionnalités de recherche d'entreprise proposées par l'outil. Cette seconde phase de la mission, qui correspond au livrable n°2 (voir section [8.5.2](#)), répond aux objectifs suivants :

- fournir un panorama sélectif des solutions existantes proposées par les différents éditeurs
- dégager pour chacune des solutions leurs forces et leurs limites en fonction du contexte spécifique à Coe-Rexecode
- établir un comparatif des fonctionnalités de recherche des différentes versions de Sharepoint 2010 afin de juger de la pertinence d'ajouter l'extension FAST à la version de Sharepoint 2010 actuellement déployée

### 10.1 Descriptif des solutions retenues

L'analyse des solutions adopte une segmentation par domaine de spécialité et par typologie d'acteur du marché : le choix s'est porté sur deux acteurs majeurs du secteur, des « Global players » Microsoft, DS Exalead ainsi que sur des éditeurs de taille plus modeste et plus spécialisés, des « pure players » Antidot, Polyspot et enfin une solution de « niche » ou « niche player »<sup>83</sup> proposée par Temis. Des solutions exclusivement open source comme Solr ont été écartées pour les raisons qui ont été déjà évoquées dans la section [6.1.2](#), page 33. Le tableau suivant décrit les différentes solutions et leurs positionnements respectifs :

Typologie	Domaine de spécialité	Solutions
Global players	Moteur de recherche intégré	FAST Search for Sharepoint
	Moteur de recherche sémantique	Exalead Cloud View
Pure players	Solutions mixtes : enrichissement sémantique / moteur de recherche	Antidot - Linked Enterprise Data
		Polyspot Enterprise Search
Niche players	Enrichissement sémantique / text mining	Temis Luxid

Tableau 7 : Solutions par typologies et domaines de spécialité

---

<sup>83</sup> La classification « global player », « pure player » et « niche player » est inspirée de la classification élaborée par le cabinet Gartner.

[http://www.gartner.com/technology/research/methodologies/research\\_mq.jsp](http://www.gartner.com/technology/research/methodologies/research_mq.jsp)

### 10.1.1 Gains et limites des solutions retenues

L'analyse des différentes solutions à travers leurs fonctionnalités ainsi que leurs positionnement respectifs a permis de dégager pour chacune d'entre elles des forces et des limites comme le décrit le tableau suivant de manière synthétique.

Solutions	Gains	Limites
<b>FAST</b> <b>Search for</b> <b>Sharepoint</b>	<ul style="list-style-type: none"> <li>• Infrastructure logicielle et matérielle déjà présente</li> <li>• Ne nécessite pas de développements spécifiques lourds : FAST est une extension de Sharepoint 2010.</li> <li>• En interne : le travail de structuration de l'information en cours d'optimisation profitera à terme à la pertinence et à l'efficacité du moteur dans le cas d'un passage à FAST</li> <li>• Fonctionnalités susceptibles de faire adhérer les utilisateurs au système d'information dans son ensemble : prévisualisation des documents Word, recherche par similarité, point d'accès unique et fédéré à l'information, intégration avec les autres fonctionnalités de Sharepoints</li> </ul>	<ul style="list-style-type: none"> <li>• Ne règle pas le problème de l'alimentation systématique des métadonnées de base : titre du document et auteur. Une absence de métadonnées, des métadonnées incomplètes ou des métadonnées de mauvaise qualité ont un impact négatif sur la pertinence des réponses et rendent le moteur inefficace (voir section <a href="#">11.1.2</a>).</li> <li>• Capacités de traitements linguistiques limitées par rapport à d'autres solutions dédiées</li> <li>• Solution surdimensionnée par rapport aux besoins documentaires actuels (la capacité de 500 millions de documents ne sera jamais atteinte)</li> </ul>
<b>Exalead</b> <b>Cloudview</b>	<ul style="list-style-type: none"> <li>• Fédération de contenu et vision à 360° de l'information : solution orientée Search Base Application, création de tableaux de bords personnalisés</li> <li>• Fonctionnalités sémantiques avancées</li> <li>• Interrogation en langage naturel : synonymes, similarité, singulier/pluriel</li> <li>• Qualité du connecteur Sharepoint</li> <li>• Gestion et indexation multilingue</li> </ul>	<ul style="list-style-type: none"> <li>• Coût élevé de la solution : de 250k€ à 300k€ et 80k€ de maintenance par an</li> <li>• Solution surdimensionnée pour les besoins actuels</li> <li>• Fonctionnalités très poussées mais n'offrant que peu de valeur d'usage à l'information produite actuellement dans l'organisation</li> <li>• Solution trop générique : ne prend pas en compte les spécificités du vocabulaire macro-économique et demande en amont un travail de paramétrage conséquent</li> </ul>

<p><b>Antidot - Linked Enterprise Data</b></p>	<ul style="list-style-type: none"> <li>• Coût peu élevé de la solution (licence de 20k€ à 40k€ par an)</li> <li>• Solution souple, légère et évolutive</li> <li>• S'appuie sur des standards ouverts, non propriétaires</li> <li>• Fédération de contenu et vision à 360° de l'information : solution orientée Search Base Application</li> </ul>	<ul style="list-style-type: none"> <li>• Back office : Effet « Boîte noire », produit très technique, pas facile d'accès</li> <li>• Front office : Outil complexe à appréhender pour un utilisateur moyen</li> <li>• Qualité du connecteur et de l'intégration Sharepoint ?</li> </ul>
<p><b>Polyspot Enterprise Search</b></p>	<ul style="list-style-type: none"> <li>• Coût initial « peu élevé » dans la mise en œuvre de la solution (du même ordre de grandeur que Antidot)</li> <li>• S'appuie sur des standards ouverts, non propriétaires</li> <li>• Connecteurs pour Sharepoint 2010, MS Exchange, eZ Publish</li> <li>• Indexation et enrichissement des données en synchrone et asynchrone</li> <li>• Prise en charge de plans de classements automatiques</li> </ul>	<ul style="list-style-type: none"> <li>• Coûts élevés pour la maintenance des référentiels sémantiques et de la performance générale de l'outil (nécessite la mise en place d'un plan d'amélioration et de suivi continu)</li> <li>• N'aide pas à l'appropriation de Sharepoint par les utilisateurs : l'interface de Polyspot devient le point d'accès unique et prend le pas sur Sharepoint.</li> </ul>
<p><b>TemisLuxid For Sharepoint 2010</b></p>	<ul style="list-style-type: none"> <li>• Ne nécessite pas de déployer un nouvel outil de recherche. La solution s'accommode du moteur de recherche déjà existant.</li> <li>• Dégage l'utilisateur de la responsabilité de renseigner les métadonnées complexes</li> <li>• Pertinence accrue et profondeur des métadonnées extraites par un processus automatique</li> <li>• Cartographie ou « mapping » de contenus : relier un organisme à ses publications ; relier un auteur à ses publications avec l'extraction de concepts manipulables par le moteur de recherche</li> <li>• Qualité du connecteur et de l'intégration avec Sharepoint 2010</li> </ul>	<ul style="list-style-type: none"> <li>• Enorme travail de paramétrage : recensement du vocabulaire « métier », des spécificités et des concepts en macro-économie ou utilisés plus spécifiquement dans l'organisation (multi-sémantisme)</li> <li>• Les périodes de test utilisateurs et de recettage peuvent être longues et complexes, nécessite un accompagnement (appel à une ressource externe)</li> </ul>

Tableau 8 : Gains et limites des solutions retenues

## 10.2 Comparatif des versions de Sharepoint 2010

Le comparatif porte sur 3 versions de Sharepoint et plus spécifiquement sur leurs fonctionnalités de recherche respectives : **Sharepoint Foundation** (version de base gratuite), **Sharepoint Server 2010** (version actuellement déployée chez Coe-Rexecode incluant toutes fonctionnalités de Foundation auxquelles s'ajoutent des fonctionnalités comme la gestion de contenu d'entreprise) et enfin **FAST search server 2010 for Sharepoint** (extension de Sharepoint Server 2010 et incluant le moteur FAST racheté au Suédois *Fast Search and Development par Microsoft* en 2008)

	Sharepoint foundation 2010	Sharepoint server 2010	FAST search server 2010 for Sharepoint
<b>Fonctions majeures</b>			
<b>Enterprise search</b> : 100 millions de documents		x	
<b>extensibilité extrême</b> : 500 millions de documents			x
Recherche par des fonctionnalités basiques	x	x	x
<b>Meilleurs résultats « visuels »</b> : des mots clés et des synonymes sont définis par un administrateur pour améliorer les résultats de recherche. Dans le cas de FAST uniquement, une « section » d'informations supplémentaires et pertinentes est affichée en plus des résultats de recherches textuelles : par exemple une bannière d'image, un logo ou une source d'information annexe		limité	x
<b>Etendues de recherches:</b> Les utilisateurs peuvent filtrer la recherche en utilisant des étendues prédéfinies		x	x
<b>Améliorations de la recherche de base en fonction du contexte utilisateur ou métier</b> : optimisation de la requête en fonction de critères prédéfinis par l'administrateur, mise en avant d'un certain nombre de sources ou documents avec la définition de sous-groupes utilisateurs aux besoins spécifiques → Personnalisation de la recherche			x
<b>Propriétés et métadonnées personnalisées:</b> gérer finement les métadonnées indexées et leur traitement dans les résultats de recherche		x	x

<b>Extraction de propriétés et d'entités nommées<sup>84</sup></b> : extrait les informations clés (noms de personnes, lieux, noms de société) à partir d'un texte non structuré (document Word, PDF) en complément des propriétés gérées traditionnelles (actuellement limité au titre, à l'auteur et à la date)		limité	x
<b>Requête fédérée</b> : fédère les résultats provenant de sources de recherche multiples		x	x
<b>Suggestions de requête (fonctionnalité prédictive)</b> : fournit une aide à la formulation de requête basée sur ce que l'utilisateur tape ou a déjà tapé (FAST rajoute la possibilité d'activer l'auto-complétion)		limité	x
<b>Résultats par similarité</b> : génère automatiquement et en parallèle une nouvelle recherche en fonction du résultat de recherche sélectionné			x
<b>Trier les résultats en fonction des métadonnées gérées ou du "ranking"</b> : trier les résultats en fonction des propriétés gérées sélectionnés ou par l'algorithme de traitement FAST			x
<b>Aperçu du document et vignettes de prévisualisation</b> : affiche les vignettes des documents Word et PowerPoint. (Affichage en ligne avec possibilité de visualiser les pages intérieures du document)			x
<b>Windows 7 avec recherche fédérée</b> : La recherche de données d'entreprise est proposée à l'utilisateur dans le navigateur traditionnel Windows		x	x
<b>Fonctionnalités secondaires</b>			
<b>Recherche "sociale"</b> : exploitation de la recherche fédérée pour trouver des utilisateurs de l'organisation par leur domaine d'expertise (publications, domaines traités)		x	x
<b>Exploitation d'une taxonomie</b> : s'appuie sur les mots clés libres générés par les utilisateurs, influence les classements de recherche et améliore l'expérience utilisateur		x	x

Tableau 9 : Fonctionnalités de recherche des différentes versions de Sharepoint 2010

<sup>84</sup> Mots ou groupes de mots qui font référence à une entité du monde concret : les noms d'entités (organisations, personnes et lieux), les expressions temporelles (dates et autres désignations temporelles) et les expressions numériques (grandeurs mesurables, quantités et pourcentages) [28, Yonnet]

### **10.3 Bilan du comparatif**

L'analyse des différentes alternatives a permis de faire les constats suivants : des solutions comme Exalead et FAST sont, dans le contexte spécifique de Coe-Rexecoce surdimensionnées soit par la capacité d'indexation (500 millions de documents pour FAST), soit par leur coût élevé et difficilement supportable. Les solutions proposées par Antidot et Polyspot semblent redondantes avec les fonctionnalités de recherche d'entreprise actuellement déployées et n'apportent pas réellement de valeur d'usage supplémentaire. Temis Luxid, en revanche, présente l'avantage de proposer une extension au moteur de Sharepoint Server 2010 qui, par sa capacité à enrichir à l'aide de métadonnées significatives le contenu présent dans les bibliothèques de documents le rend tout-à-fait pertinent. Mais sa mise en place semble, dans le contexte actuel de Coe-Rexecode, un peu prématuré

En effet, l'appropriation de Sharepoint est loin d'être acquise. L'introduction d'un nouvel outil, quel qu'il soit, serait hasardeuse. Solliciter les utilisateurs une nouvelle fois si peu de temps après le déploiement de Sharepoint, alors que les fonctionnalités de recherche sont si peu utilisées (et comprises), semble peu pertinent.

# 11 Optimisation du moteur de recherche

---

Afin de mener à bien cette étape d'optimisation des fonctionnalités de recherche, une étude des spécificités fonctionnelles a été réalisée ainsi qu'une série de tests afin de déterminer quelles sont les métadonnées et propriétés des documents qui sont priorisées par le moteur.

## 11.1 Spécificités fonctionnelles

Les tests ont montré que la pertinence est principalement basée sur les métadonnées titre et auteur. La présence ou l'absence de ces propriétés impacte de manière significative la remontée optimale du ou des documents qui pourraient être considérés comme pertinents pour l'utilisateur. Si les deux métadonnées sont présentes, elles seront fortement priorisées. L'occurrence du ou des termes de la requête trouvés dans le document améliore également de façon significative le classement de ce dernier dans la liste de résultats.

Si la métadonnée titre est absente des propriétés du document, le paramétrage par défaut du moteur fait remonter dans les résultats de recherche son nom informatique. Cet aspect peut être problématique lorsque le nom informatique est peu révélateur du contenu du document et qu'aucune autre métadonnée n'est présente (voir section [11.1.2.1](#)).

Pourtant la présence de la métadonnée titre ne garantit en rien son exploitation optimale par le moteur. Les tests menés sur l'indexation des documents Office (Word, Excel et Powerpoint) ont montré que même si la propriété titre est présente, le moteur va systématiquement l'ignorer en lui substituant le nom informatique du document.. Une action corrective a été menée pour modifier ce paramétrage par défaut du moteur (voir section [11.2.2](#)).

Le moteur ignore également partiellement la présence éventuelle de mots clés et de descripteurs dans les propriétés du document ou via des métadonnées spécifiquement gérées par Sharepoint. Le moteur n'est pas conçu (ou pas paramétré) pour effectuer une recherche directement sur les mots clés d'un document. Le fait de rechercher un mot présent dans le texte et un éventuel mot clé présent dans une propriété du document n'améliore pas le classement de ce dernier. En revanche, l'alimentation de mots clés via un référentiel se justifie par leur exploitation dans les filtres à facette (voir section [11.1.1](#)).

Le résumé automatique du document proposé à l'utilisateur est produit à partir de l'extraction d'éléments textuels contenant le ou les termes de la requête initiale. L'éventuelle présence d'un résumé dans une métadonnée gérée par Sharepoint ou via la création d'un descriptif dans les propriétés du document ne sera pas prise en compte ou exploité par le moteur.

Métadonnées	Documents Office	Ressources PDF
<b>Métadonnées de base</b>		
Titre		x
Auteur	x	
Langue	x	x
<b>Métadonnées personnalisées</b>		
Contributeur	x	
Pays/zone	x	x
DATE_PUB	x	x
THEME_GEO	x	x

Tableau 10 : métadonnées exploitées par Sharepoint 2010 Server Search

Comme le montre le tableau ci-dessus, le moteur ne traite pas de la même manière les documents office et les ressources PDF qui constituent la majeure partie des ressources disponibles dans les bibliothèques Sharepoint. Dans le cas des ressources PDF, le moteur va ignorer la métadonnée auteur renseignée via le formulaire Sharepoint et va prioriser la métadonnée auteur embarquée nativement dans le PDF (quand elle est renseignée).

### 11.1.1 Modes d'interrogation de filtres et présentation des résultats

L'utilisateur a la possibilité d'interroger l'index de recherche de deux manières différentes : en recherche simple depuis le portail Sharepoint et ses sous-sites dans une zone de saisie placée dans la barre de navigation supérieure ainsi que depuis un espace dédié, le centre de recherche, accessible via le portail et ses sous-sites.

L'utilisateur a également la possibilité d'interroger toute ou partie de la base Sharepoint à l'aide des étendues de recherche. Cette option est accessible indifféremment en recherche simple à l'aide d'une liste déroulante ou en recherche avancée à l'aide de cases à cocher. Une étendue de recherche définit un sous-ensemble d'informations dans l'index de recherche afin de limiter une recherche à des emplacements ou un contenu qui englobe généralement des rubriques spécifiques et des sources de contenu qui sont importantes et communes aux utilisateurs de l'organisation<sup>85</sup>. Six étendues de recherche ont été définies :

- Tous les sites : permet de lancer une recherche sur tous les sites et sous-sites
- Sur ce site : concentre la recherche sur le site actif et les sous-sites
- Intranet : concentre la recherche sur les documents d'activité de l'espace Intranet

<sup>85</sup> <http://office.microsoft.com/fr-fr/sharepoint-server-help/definition-detendues-pour-les-recherches-HA010241119.aspx>

- Intradoc : concentre la recherche sur l'espace Intradoc (pages pays, pages organismes)
- Documents et archives Coe-Rexecode
- PressBook : documents externes et articles relatifs à Coe-Rexecode
- Personnes : concentre la recherche sur les auteurs - collaborateurs de Coe-Rexecode

Le sous-site « centre de recherche » est scindé en 2 onglets. Le premier onglet permet d'interroger le moteur à l'aide d'une zone de saisie simple « à la google ». Le deuxième onglet permet de basculer vers la recherche dite « avancée » avec la possibilité pour l'utilisateur de spécifier un certain nombre de critères et de contraintes :

- spécifier si les termes de la requêtes contiennent : tous ces mots, la phrase exacte, n'importe quel mot ou bien aucun de ces mots
- contraindre à une ou plusieurs étendues de recherche et à un type spécifique de document
- restreindre la recherche à une ou plusieurs propriétés des documents avec la possibilité d'interroger directement avec des opérateurs booléens de types ET, OU, SAUF

Une fois la liste de réponses affichée, l'utilisateur a la possibilité d'affiner sa recherche grâce à des filtres à facettes placés dans une navigation contextuelle. Les facettes de navigation sont de deux types :

- des facettes génériques : les valeurs sont extraites automatiquement des propriétés des fichiers (type de document, auteur, date et langue)
- des facettes Coe-Rexecode : les valeurs, prédéfinies par l'administrateur Sharepoint sont modifiables, extensibles via des listes hiérarchiques et s'appuient sur des métadonnées gérées par Sharepoint. Pour l'instant, seuls la zone géographique, le pays et la métadonnée « contributeur » (en complément de l'auteur) sont exploitées

Sharepoint Search Server propose également à l'utilisateur deux types de classement des résultats de recherche : un classement par pertinence et un classement par date chronologique ou anté-chronologique.

Les documents restitués par l'index de recherche sont accessibles dans le navigateur. Les niveaux d'accès en lecture et en écriture sont restreints par les droits des collaborateurs définis dans l'annuaire LDAP.

### **11.1.2 Limites et écueils du moteur**

Il existe une qualité très inégale ainsi qu'une disparité et une hétérogénéité des métadonnées, notamment des propriétés de base, titre et auteur. Certains documents en sont même complètement dépourvus. C'est le cas, entre autre, des documents (rapports, études, analyses ...) issus d'organismes externes (FMI, BCE, Banque Mondiale...).

Ces deux métadonnées de base sont pourtant essentielles. Comme évoqué en section [11.1](#), le modèle de pertinence est principalement basé sur le titre et l'auteur (qui peut être une personne ou un organisme). Ces deux attributs de contextualisation importants pour les collaborateurs de Coe-Rexecode permettent d'évaluer la pertinence des réponses fournies par l'index de recherche ainsi que leur adéquation avec la requête initiale.

Pour les documents produits en interne la situation est extrêmement contrastée. Si certains documents embarquent des métadonnées de bonne qualité (c'est notamment le cas des documents issus du pôle documentation), la qualité est plus variable pour les autres pôles. Cette hétérogénéité s'explique en partie par le fait qu'un nombre important de fichiers est basé sur des modèles de documents prédéfinis. Le titre du document n'étant pas systématiquement mis-à-jour par son rédacteur, pour un gain de temps à sa création, il en résulte qu'un grand nombre de fichiers sont indexés sous le même titre.

Pour les documents externes, la propriété auteur est fréquemment peu ou mal renseignée. Leur formats et leurs libellés sont extrêmement variables, provoquant des entrées multiples pour un même auteur dans les filtres à facette ou dans la liste de résultats (exemple : Ocde, oecd, ocde / oecd).

Le modèle de pertinence du moteur pose également question. Les tests ont permis d'établir de manière empirique que la pertinence reposait sur la présence ou l'absence de certaines métadonnées et sur le nombre d'occurrences des termes de la requête dans l'index de recherche, mais sur quel modèle exact repose cette pertinence ? La documentation en ligne de Microsoft ne permet pas en l'état de déterminer sur quels critères s'appuie le calcul de pertinence. Cette absence de documentation accentue l'effet « boîte noire » du moteur et limite le paramétrage de ce dernier.

## **11.2 Actions correctives menées**

### **11.2.1 Gestion des « meilleurs résultats »**

Les « meilleurs résultats » sont constitués à partir de mots clés prédéfinis et de synonymes (mots apparentés étroitement au mot clé) en vue d'améliorer les résultats de recherche portant sur un contenu spécifique (documents, sites, bibliothèques de documents). Les « meilleurs résultats » permettent de guider les utilisateurs vers des emplacements

recommandés, des pages ou des bibliothèques de documents faisant autorité et constituent un moyen rapide, non invasif d'améliorer la pertinence des réponses fournies à ces derniers.

Les « meilleurs résultats » relatifs à ces mots clés s'affichent en évidence dans la page des résultats de recherche, au-dessus des résultats généraux (voir section [11.2.2](#)). Chaque entrée de mot clé comprend une liste de synonymes, ainsi que les adresses URL pointant vers des ressources prédéfinies ou vers un contenu fortement mis en valeur.

En définissant les termes les plus fréquemment recherchés ou les plus pertinents en tant que mots clés d'interrogation, on peut établir un référentiel standard de noms d'organismes, de processus et de concepts constituant les « connaissances communes » partagées par les membres d'une organisation<sup>86</sup>. En outre, le professionnel ou l'équipe en charge du paramétrage et de l'optimisation du moteur peut progressivement aligner les « meilleurs résultats » ainsi que les contenus faisant autorité en fonction des points d'intérêts ponctuels ou sur le long terme d'un utilisateur ou d'un groupe d'utilisateurs.

Une liste exhaustive des mots clés prédéfinis est proposée en **annexe 4**, ainsi qu'un extrait du référentiel ayant servi à définir, puis injecter les mots clés, les synonymes et les liens URL dans le backoffice de Sharepoint en **annexe 5**. Ce référentiel a fait l'objet d'un livrable (voir section [8.5.2](#)).

#### **11.2.1.1 Planification des mots clés et des synonymes**

Cette planification a consisté à identifier et à définir un contenu jugé prioritaire pour les économistes des pôles études et conjoncture. L'optimisation par mots clés et meilleurs résultats associés a porté sur les bibliothèques de documents et les pages Sharepoint de la partie Intradoc (voir section [11.1.1](#)) :

- zones géographiques et pays suivis par les analystes de Coe-Rexecode (France, Allemagne, zone asie ...)
- organismes nationaux (INSEE, Cour des comptes, Banque de France...)
- organismes internationaux (OCDE, BCE, Banque mondiale...)

Dans le contexte spécifique à Coe-Rexecode, un économiste peut désormais effectuer une recherche en tapant indifféremment OCDE (forme française) ou OECD (forme anglaise) et se verra proposer dans les premiers résultats de recherche un ou plusieurs liens vers les ressources Sharepoint se rapportant à l'OCDE sans qu'il ait à connaître à l'avance le chemin d'accès des ressources dans l'arborescence des sites et sous-sites Sharepoint.

---

<sup>86</sup> <http://technet.microsoft.com/fr-fr/library/hh133422%28v=office.14%29.aspx>

De la même manière, en sollicitant le moteur de recherche avec le terme INSEE<sup>87</sup> ou avec l'un de ses synonymes associés (statistiques insee; stats insee; statistiques France; documents insee), l'utilisateur se voit proposer plusieurs liens vers les ressources et les bibliothèques de publications INSEE (statistiques, notes de conjonctures, analyses et rapports)

The screenshot shows a search results page for OCDE. The main content area displays a list of results, with the top two being highlighted with star icons. The sidebar on the left contains various filters such as 'Type de résultat', 'Site', 'Auteur', and 'Date de modification'.

Type de résultat	Résultats 1-10 sur environ 16 002
Toute catégorie T...	
Adobe PDF	
Page Web	
Site	
Toute catégorie S...	
sp-2010:15731	
Auteur	
Toute catégorie A...	
Constanza_L	
Fabienne Besson-L...	
Fbl	
Mace_S	
McIntosh_J	
Murielle Prévost	
Narbeburu_O	
Ocde	
Oecd	
Oecd / Ocde	
Oecd/Ocde	
Oiseau_S	
Ollivaud_P	
Sylvie Foutrier	
moins de résultats ▾	
Date de modificat...	
Toute catégorie D...	
Semaine dernière	
Mois dernier	
Six derniers mois	
Plus tôt	
Pays/Zone	
Toute catégorie Pay...	

★	<a href="#">Page Intradoc - OCDE</a> Point d'entrée pour les publications et les statistiques de l'OCDE
★	<a href="#">Bibliothèque de documents - OCDE</a>
	<a href="#">OCDE</a> Les perspectives économiques de l'OCDE n°93 (juin 2013) sont mises à jour (persocde.bnk ... Tableau de bord de l'OCDE de la science, de la technologie et de l'industrie ... Auteurs : Fabienne Besson-Lhoste Date : 10/07/2013 Taille : 240KB <a href="https://sp-2010:15731/GED/Pages/Pages_organismes/OCDE.aspx">https://sp-2010:15731/GED/Pages/Pages_organismes/OCDE.aspx</a>
	<a href="#">OCDE</a> L'emplacement de cette page est : ... Cette liste : OCDE ... OCDE-Ilibrary-aide ... OCDE-Ilibrary- themes ... Gérer les plans d'actualisation du cache ... Modifier la définition de la source de données ... Date : 19/07/2013 <a href="https://sp-2010:15731/ged/OCDE/Forms/AllItems.aspx">https://sp-2010:15731/ged/OCDE/Forms/AllItems.aspx</a>
	<a href="#">Réunion Perspectives</a> L'emplacement de cette page est : ... Espace dédié à la préparation des réunions perspectives ... Gérer les plans d'actualisation du cache ... Perspectives économiques de l'OCDE ... Date : 26/07/2013 Taille : 137KB <a href="https://sp-2010:15731/\$Conjoncture/reunion_perspectives">https://sp-2010:15731/\$Conjoncture/reunion_perspectives</a>
	<a href="#">Perspectives économiques de l'OCDE</a> 93 MAI 2013Les Perspectives économiques de l'OCDE paraissent sous la responsabilité du Secrétaire général de l'OCDE. Les appréciations portées sur l'évolution ... 41 1.11. Il n'est pas ... Auteurs : Fabienne Besson-Lhoste, Ocde Date : 12/07/2013 Taille : 6MB <a href="https://sp-2010:15731/GED/OCDE/Perspectives/PERSPECTIVES-OCDE.pdf">https://sp-2010:15731/GED/OCDE/Perspectives/PERSPECTIVES-OCDE.pdf</a>
	<a href="#">OCDE-PPA.pdf</a> européennes participe aux travaux de l'OCDE. Les Éditions OCDE assurent une large diffusion aux travaux de l ... Auteurs : Fabienne Besson-Lhoste Date : 19/03/2013 Taille : 3MB <a href="https://sp-2010:15731/GED/OCDE/statistiques/PPA/OCDE-PPA.pdf">https://sp-2010:15731/GED/OCDE/statistiques/PPA/OCDE-PPA.pdf</a>
	<a href="#">Perspectives économiques de l'OCDE 2007 11</a> OECD ECONOMIC OUTLOOK Preliminary Edition 82 DECEMBER 2007 OECD ECONOMIC OUTLOOK PRELIMINARY EDITION 3 TABLE OF CONTENTS EDITORIAL ... 206 3. CORPORATE SAVING AND INVESTMENT: RECENT TRENDS ... Auteurs : Fabienne Besson-Lhoste Date : 11/07/2013 Taille : 2MB <a href="https://sp-2010:15731/GED/OCDE/Perspectives/peo82.pdf">https://sp-2010:15731/GED/OCDE/Perspectives/peo82.pdf</a>
	<a href="#">PERSPECTIVES DE L'OCDE SUR LES PME</a> Auteurs : Fabienne Besson-Lhoste, Ocde Date : 19/03/2013 Taille : 3MB <a href="https://sp-2010:15731/GED/OCDE/2000/pme.pdf">https://sp-2010:15731/GED/OCDE/2000/pme.pdf</a>

Figure 8 : Page de résultats OCDE après optimisation des meilleurs résultats

Comme le montre l'illustration ci-dessus, les ressources OCDE sont désormais clairement positionnées dans les premiers résultats fournis par l'index de recherche. Deux liens sont proposés, l'un renvoyant vers la page OCDE et ressources associées et l'autre renvoyant plus spécifiquement à la bibliothèque de documents OCDE.

<sup>87</sup> L'utilisateur peut interroger le terme indifféremment en majuscule ou en minuscule, le moteur effectue un transcodage automatique. En revanche, dans la définition des synonymes, on doit prévoir une forme avec accentuation et une forme sans accentuation : par exemple « bibliothèque insee » et « bibliotheque insee »

## 11.2.2 Actions complémentaires

Deux actions correctives supplémentaires ont été menées :

- Le choix d'une bibliothèque de document « pilote » afin de réaffecter de manière procédurale les métadonnées titre et auteur pour améliorer leur visibilité dans la liste de réponse fournie à l'utilisateur et ainsi améliorer la pertinence. Cette réaffectation par lot de métadonnées est rendue possible par l'utilisation du mode « feuille de données » de Sharepoint<sup>88</sup>.
- La réindexation de la base Sharepoint<sup>89</sup> afin de corriger le paramétrage par défaut du moteur qui consiste à ignorer la métadonnée titre d'un document Office même lorsque celle ci est présente ou renseignée par le rédacteur du document

---

<sup>88</sup> Possibilité de réaffecter par lot un ensemble de métadonnées (par exemple titre ou auteur) sur un grand nombre de documents inclus dans une bibliothèque Sharepoint

<sup>89</sup> Procédure de réindexation : <http://romaincs.wordpress.com/2013/04/23/le-titre-du-doc-apparait-a-la-place-du-titre-du-document-dans-les-resultats-de-la-recherche/>

# Conclusion

Le moteur de recherche d'entreprise, par sa capacité à indexer les données de manière transversale, n'est pas qu'un simple outil susceptible de pallier au désordre et au cloisonnement des silos informationnels. Sa mise en œuvre, comme projet d'entreprise, déclenche inévitablement une réflexion profonde et interroge la capacité de l'organisation à se transformer, tant d'un point de vue organisationnel que technologique.

La mission menée au sein de Coe-rexencode montre que les dispositifs TIC, au premier plan desquels figurent les moteurs de recherche, s'inscrivent dans cette démarche de transformation. Ils soulèvent des questionnements et engendrent de nouveaux objectifs informationnels. Ils obligent l'organisation à formuler et à (re)penser de nouvelles procédures ou routines de travail, à initier de nouvelles pratiques collaboratives et managériales, et enfin à envisager une politique de gouvernance de l'information.

Ainsi, la mise en œuvre d'un moteur de recherche n'est pas une démarche qui repose uniquement sur une approche technologique. Lorsqu'un collaborateur recherche de l'information, il le fait dans un but précis afin de réaliser une tâche particulière : prendre une décision, acquérir de nouvelles connaissances, etc. Cette démarche de recherche d'informations nécessite de la part de l'entreprise d'avoir avant tout une compréhension précise des besoins de ses collaborateurs.

La mise en place d'un moteur intégré comme celui de Sharepoint Server 2010 est emblématique de ces problématiques : il entraîne des réflexions sur le périmètre des sources à indexer et à interroger, sur le niveau « d'opacité » ou de visibilité des données et des connaissances ainsi que sur la politique d'accès aux informations. Il provoque également une remise à plat du système d'information et de l'organisation documentaire à l'échelle de l'organisation. La question des métadonnées, auparavant sous-jacente et peu stratégique, devient désormais primordiale. Afin d'offrir son plein potentiel aux utilisateurs, un moteur de recherche doit s'appuyer sur des métadonnées de bonne qualité, qu'elles soient générées de manière automatique ou manuelle.

Au delà de l'aspect purement technologique, la performance d'un moteur de recherche repose sur un travail de structuration, de management de l'information et de paramétrage mené en amont. Les éditeurs sont souvent prompts à propager l'idée que « l'intelligence » du moteur serait automatique et uniquement orchestrée par des algorithmes standards et des ressources linguistiques intégrées. La limite de cette logique est que l'algorithme de pertinence du moteur porte un modèle spécifique et particulier de la connaissance. Il doit être paramétré pour les besoins des utilisateurs et en fonction des processus en vigueur dans l'organisation.

Au cœur de cette structuration et de ce paramétrage, les gestionnaires et spécialistes de l'information occupent une place centrale. Leur rôle va consister à analyser l'existant ainsi que les usages en termes de recherche, de consultation, de partage et de circulation de l'information, à évaluer la juste adéquation de l'offre de services, dont le moteur fait partie, avec les besoins utilisateurs pour déterminer le cas échéant des actions correctives à mettre en œuvre et enfin à superviser la constitution de référentiels linguistiques et de vocabulaires d'enrichissement nécessaires à la production de métadonnées de haute qualité.

Au-delà des enjeux technologiques et documentaires, la mise en place d'un moteur de recherche d'entreprise coordonné avec la mise en place d'un ECM, comme ce fut le cas chez Coe-Rexecode, est un projet d'envergure qui demande un travail de communication et de conduite du changement. Il doit être explicité et défendu au sein de l'organisation par l'équipe qui en a la charge afin de s'assurer de l'adhésion des utilisateurs. Un tel projet ne peut être légitimé que si les finalités à long terme sont comprises par tous les collaborateurs. Sans quoi, il sera assimilé à un simple effet de mode lié à l'introduction d'un nouvel outil.

Cette mission menée au sein de Coe-Rexecode m'a permis de mieux cerner les principes d'une gestion de projet efficace à travers trois dimensions :

- la dimension « Projet » : un engagement de la direction au-delà d'objectifs court-termistes, un positionnement projet en phase avec les objectifs de l'organisation, une équipe projet spécialiste de la gestion de l'information et capable de dialoguer avec toutes les parties prenantes (direction, informaticiens, collaborateurs)
- la dimension « Processus » : un partage clair des rôles et responsabilités au sein de processus décisionnels dont les enjeux sont compris et explicités, une évaluation continue du projet et la mise en œuvre rapide d'actions correctives.
- la dimension « Ressources » : une équipe projet réactive et à l'écoute, des relais au sein de la direction clairement identifiés, des utilisateurs consultés et impliqués dans les différentes phases du projet

# **Sigles**

**API** : Application Programming Interface *ou* interface de programmation

**BI** : Business Intelligence *ou* informatique décisionnelle

**CRM** : Customer Relationship Management ou gestion de la relation client

**ECM** : Enterprise Content Management

**ERP** : Enterprise Resource Planning *ou* progiciel de gestion intégré

**ESP** : Enterprise Search Platform

**GED** : Gestion Electronique de Document

**IR** : Information Retrieval *ou* recherche d'informations

**ISO** : International Standards Organisation

**KM** : Knowledge Management

**ROC** : Return On Collaboration *ou* retour sur collaboration

**ROI** : Return On Investment *ou* retour sur investissement

**SBA** : Search Base Application *ou* application orientée recherche

**TAL** : Traitement Automatique du Langage

**TIC** : Technologies de l'Information et de la Communication

# **Glossaire**

**Autocomplétion** : Mode d'assistance à la saisie dans les champs de formulaire. L'autocomplétion consiste à proposer des mots complets lorsqu'un ou plusieurs caractères ont été saisis par l'utilisateur dans un formulaire. Elle peut être proposée par le moteur de recherche en utilisant l'historique des saisies (Source ADBS).

**Clustering** : Techniques statistiques (méthode des mots associés avec matrices de cooccurrences) qui permettent d'extraire automatiquement des termes dans les pages de résultats trouvées par un moteur ou un méta-moteur, afin d'affiner ou de réorienter une recherche. Les clusters ainsi formés servent à mettre en évidence des thématiques présentes dans les documents trouvés par les outils de recherche. Cela peut présenter un intérêt en recherche documentaire classique (classement automatique de résultats selon des thématiques prédéfinies ou générées automatiquement par le moteur) ou en veille (Source ADBS).

**Entité nommée** : la reconnaissance d'entités nommées est une sous-tâche de l'activité d'extraction d'information dans des corpus documentaires. Elle consiste à rechercher des objets textuels (c'est-à-dire un mot, ou un groupe de mots) catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, quantités, distances, valeurs, dates, etc. (Source Wikipedia)

**Informations non structurées** : Tout document, fichier, image, rapport, formulaire,... qui n'a pas de structure définie standard qui permettrait de le classer facilement sous forme d'enregistrement complet ou par un mode de traitement similaire automatisé. Les informations non structurées ne peuvent pas être définies en termes de lignes et de colonnes comme c'est le cas de données structurées issues de bases de données.

**Référentiel** : Ensemble d'éléments documentaires, pris pour référence, mis sous gestion de configuration et ne pouvant être modifiés après acceptation initiale que par l'intermédiaire d'une procédure formalisée.

**Search Base Application** : Applications basées sur la recherche (SBA), sont des applications logicielles dans lesquelles une plate-forme de moteur de recherche est utilisée comme infrastructure de base pour l'accès à l'information et la communication. Les SBA utilisent les technologies sémantiques pour agréger, normaliser et classer les contenus non structurés (documents Word), semi-structurés (documents Excel) et / ou structurés (Bases de données statistiques) dans plusieurs référentiels, et emploient des technologies de langage naturel pour accéder à l'information agrégée.

**Taxonomie** : Décomposition hiérarchique de catégories de sujets partant du très général au très spécifique. La taxonomie offre un moyen de localiser l'information dans le flux

croissant des données, grâce à la catégorisation et à l'indexation dans une structure que la personne qui effectue la recherche comprend facilement.

**Text mining** : Fouille de texte en français. C'est un ensemble de traitements informatiques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité dans des textes produits. La fouille de texte peut consister en l'indexation d'un ensemble de textes par rapport aux mots qu'ils contiennent. On peut ensuite interroger l'index ainsi créé pour connaître les similarités entre une requête et la liste de textes. (Source Wikipedia)

**Thesaurus** : Un thesaurus ou thésaurus est une liste de termes normalisés et reliés entre eux par des relations sémantiques de type générique /spécifique. Le thésaurus est l'outil de base de l'indexation des documents.

**Travail collaboratif** : Le travail collaboratif est le processus par lequel plusieurs individus, groupes ou systèmes travaillent ensemble, mais à un niveau nettement supérieur à la coordination ou à la coopération ; la collaboration implique généralement la planification conjointe, les ressources partagées et la gestion des ressources conjointes. Elle met en jeu des processus de communication synchrones ou asynchrones.

# **Bibliographie**

La bibliographie analytique proposée a été arrêtée au 23 Août 2013.

Elle a pour objectif d'offrir au lecteur des références jugées pertinentes pour la rédaction de ce mémoire. Cette bibliographie vise également à permettre au lecteur de poursuivre sa réflexion et d'approfondir la question du moteur de recherche d'entreprise sous un angle organisationnel, conceptuel ou encore technologique.

La rédaction des références bibliographiques est conforme aux normes :

- Z44-005. décembre 1987. Documentation. Références bibliographiques : contenu, forme et structure et à la norme
- NF ISO 690-2 Février 1998 Information et documentation. Références bibliographiques Documents électroniques, documents complets et parties de documents

Les références bibliographiques sont précédées d'un chiffre entre crochet qui correspond à l'ordre de classement des références dans la bibliographie et non à l'ordre d'appel dans le corps du texte. Chaque référence est suivie d'un texte en italique expliquant pourquoi elle a été retenue.

Le classement thématique est le suivant :

- Capital immatériel
- Gestion de projet et analyse des besoins utilisateurs
- Gouvernance de l'information
- Indexation et métadonnées
- Management documentaire en entreprise
- Mémoires et travaux universitaires
- Moteur de recherche d'entreprise : aspects sémantiques
- Moteur de recherche d'entreprise : marché et tendances
- Moteur de recherche d'entreprise : spécificités et nouveaux positionnements
- Moteur de recherche d'entreprise : stratégie et mise en œuvre
- Recherche d'information et traitement des langues
- ROC, ROI et gestion de contenu
- Technologies de l'information et connaissances
- Technologies de l'information et travail

Dans le corps de ce mémoire, certains sites internet ou sources externes sont mentionnés en notes de bas de page. Les liens sont tous actifs à la date du 15 Octobre 2013.

## Capital immatériel

**[01]** ATOS ORIGIN. De la gestion de contenu à la gestion de l'information. Livre blanc [En ligne] Paris, 2008. [Consulté le 24 Juillet 2013]

[http://www.fr.atosconsulting.com/NR/rdonlyres/E67BAD4D-E53D-4787-8A29-B19FBD7DF0EF/0/ATOS\\_ORIGIN\\_LIVRE\\_BLANC\\_ECM.pdf](http://www.fr.atosconsulting.com/NR/rdonlyres/E67BAD4D-E53D-4787-8A29-B19FBD7DF0EF/0/ATOS_ORIGIN_LIVRE_BLANC_ECM.pdf)

*Ce livre blanc à destination des dirigeants d'entreprises et des spécialistes des technologies de l'information montre en quoi une bonne gestion des actifs immatériels améliore la performance collective en positionnant clairement les moteurs de recherche comme une brique essentielle d'un ECM. Les auteurs proposent des axes d'amélioration contribuant à une meilleure structuration de l'information ainsi qu'une optimisation des outils de recherche.*

**[02]** BOURRET Christian et al. Capital immatériel et information professionnelle. L'émergence d'un concept nouveau : l'information durable. Documentaliste-Sciences de l'Information, 2008/4, vol. 45, p. 4-12. DOI : 10.3917/docsi.454.0004

*Dans cette étude, les auteurs définissent la notion de capital immatériel et en précisent clairement son cadre d'application au sein des organisations. Ils pointent également la difficulté de quantifier précisément l'apport de l'information et de mesurer sa valeur ajoutée à l'aide d'indicateurs traditionnels. L'étude met également en lumière les risques d'une information pléthorique et non maîtrisée.*

**[03]** CIGREF. DSI et Capital immatériel : Maturité et mise en œuvre [pdf en ligne]. Livre blanc. Paris, CIGREF, 2006, 84 p. [Consulté le 12 Juin 2013]. <http://www.cigref.fr/dsi-et-capital-immateriel-maturite-et-mise-en-oeuvre>

*Ce rapport propose une approche double du capital immatériel : l'approche analytique et l'approche systémique. L'auteur développe également l'idée de création de valeur par le système d'information en mettant en parallèle les besoins utilisateurs et les services à mettre en place pour les satisfaire.*

**[04]** FUSTEC Alan, MAROIS Bernard. Valoriser le capital immatériel de l'entreprise. Editions d'Organisation, 2006. 169 p. ISBN 2708136658

*Cet ouvrage, ancré dans la pratique, développe dans un premier temps une réflexion sur l'économie de l'immatériel en définissant ses enjeux et présente dans un second temps une méthode d'évaluation de la valeur financière du capital immatériel à l'aide d'une grille de notation et d'un tableau de bord.*

## Gestion de projet et analyse des besoins

[05] MARCHAT Hugues. Le kit du chef de projet, 4eme édition. Eyrolles, 2010, 226 p.

*Cet ouvrage permet de traverser de manière pragmatique toutes les strates de la méthode 3P de gestion de projet proposée par son auteur. Ouvrage retenu par sa facilité d'accès, sa modularité appliquée à toutes les phases d'un projet et ses fiches projet pré-formatées.*

[06] MARCHAT Hugues. La gestion de projet par étapes : l'analyse des besoins, 2eme édition. Paris, Editions d'Organisation, 2011, 225 p. ISBN 2-7081-3485-X

*Cet ouvrage propose une approche à la fois qualitative et quantitative afin d'appréhender la phase délicate et stratégique de l'expression de besoins utilisateurs. Premier volume d'une série de quatre ouvrages consacrés à la gestion globale de projet, il est extrêmement utile pour initier de bonnes pratiques dès les phases d'initialisation.*

## Gouvernance de l'information

[07] SERDALAB. La gouvernance documentaire dans les organisations françaises [pdf en ligne]. Livre blanc. Paris, SerdaLAB, mars 2011. [Consulté le 24 Juillet 2013]

[http://www.serdalab.com/Medias/Livres%20blancs-%C3%A9tudes%20gratuites/Livre blanc serdaLAB gouvernance documentaire.pdf](http://www.serdalab.com/Medias/Livres%20blancs-%C3%A9tudes%20gratuites/Livre%20blanc%20serdaLAB%20gouvernance%20documentaire.pdf)

*Résultats de l'enquête publique menée auprès d'un certain nombre d'organisations confrontées aux difficultés d'une gestion documentaire globale. L'enquête est particulièrement riche en données chiffrées en lien avec les thématiques liées à la gouvernance documentaire (réduction du temps de recherche d'informations, valorisation du capital immatériel, etc...).*

## Indexation et métadonnées

[08] AMAR Muriel. Nouvelles pratiques d'indexation, nouveaux enjeux documentaires ? In Urfist Paris [site], Paris, Urfist de Paris, mise à jour le 27 avril 2009, [Consulté le 10 Juin 2013].

<http://urfist.enc.sorbonne.fr/sites/default/files/file/traitementdoc/Pratiques-d%27indexation-support.pdf>

*Ce support de cours propose une vision globale et transversale des différents modes d'indexation, notamment documentaire et/ou automatique et dégage pour chacun d'entre eux leurs points forts et leurs points faibles. L'auteur aborde également les problématiques liées à l'ingénierie linguistique ainsi que les enjeux du web de données.*

**[09]** BENNETT Mark, LEHMAN John. Building a taxonomy. In New Idea Engineering [site], Santa Clara, New Idea Engineering Inc., mise à jour en janvier 2009, [Consulté le 10 Juin 2013]. <http://www.ideaeng.com/building-a-taxonomy-0102>

*Ce court article de blog propose une approche méthodologique visant à l'élaboration d'une taxonomie en explicitant très clairement son processus de création ainsi que ses différentes strates de développement. L'auteur insiste fortement sur la structure hiérarchique d'une taxonomie.*

**[10]** CHAUDIRON Stéphane. Technologies linguistiques et modes de représentation de l'information textuelle. Documentaliste-Sciences de l'information, 2007/1, vol. 44, pp. 30-39. ISSN 0012-4508.

*Cet article démontre l'utilité des technologies de TAL dans les processus d'indexation, de classification et de catégorisation de l'information non structurée. L'auteur revient également sur la difficulté pour les utilisateurs des systèmes d'information d'accéder à une information pertinente liée au sentiment de « surcharge informationnelle ». Il décrit également les quatre phases représentatives de l'analyse textuelle du moteur de recherche : l'indexation automatique, le résumé automatique, la clusterisation et la catégorisation.*

**[11]** MICROSOFT CORPORATION. Adding structure to unstructured content for enhanced findability. [En ligne]. Redmond, Microsoft Corporation, 2010, [Consulté le 10 Juin 2013]. <http://www.microsoft.com/en-us/download/details.aspx?id=14688>

*Après avoir défini le concept de métadonnées et bien marqué la différence entre un moteur web et un moteur d'entreprise, les auteurs de ce livre blanc démontrent l'inefficacité d'une indexation manuelle pour générer des métadonnées de haute qualité. L'analyse qui est faite du rejet d'un moteur de recherche par ses utilisateurs, faute de métadonnées valables, est très pertinente.*

**[12]** AFNOR NF Z47-102 Octobre 1993 Information et documentation - Principes généraux pour l'indexation des documents (norme)  
<http://www.boutique.afnor.org/norme/nf-z47-102/information-et-documentation-principes-generaux-pour-l-indexation-des-documents/article/702600/fa026779>

**[13]** REAMY Tom. To metadata or not to metadata. Econtent Magazine [en ligne], Information Today Inc., Numéro Octobre 2004. [Consulté le 10 Juin 2013]. <http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=7118&PageNum=1>

*Cet article souligne le problème de l'efficacité et de la valeur d'usage des métadonnées en abordant la question de leur impact financier en termes de maintenance et de pertinence.*

*L'auteur développe l'idée d'une « infrastructure de métadonnées » et d'une équipe mixte documentaire / informatique dédiée à l'enrichissement et à la maintenance de celle-ci.*

## **Modèles d'organisation de l'information**

**[14]** ANTIDOT. Réinventer les référentiels. In Blog Antidot [blog], Paris, Antidot, mise en ligne le 25 mars 2010, [Consulté le 10 Juin 2013]. <http://blog.antidot.net/2010/03/25/reinventer-les-referentiels/>

*Ce billet de blog définit ce que sont des référentiels et décrit leur utilité ainsi que leur champ d'application dans le cadre d'une utilisation par un moteur de recherche. L'auteur donne également une liste non exhaustive de bonnes pratiques en vue de leur élaboration.*

**[15]** COTTE Dominique. Documents numériques au travail. In Broudoux Evelyne, Chartron Ghislaine. Traitements et pratiques documentaires – vers un changement de paradigme ? Actes de la deuxième conférence Document numérique et Société, 2008. ADBS Editions. Paris, ADBS, 2008. Deuxième partie Acteurs de l'offre et traitements documentaires, pp.209-222. ISBN 978-2-84365-116-8.

*Cet ouvrage propose une analyse de la documentation numérique et de ses modes d'exploitation dans les organisations (description, circulation, stockage, recherche). Son approche de l'espace numérique de travail et la logique d'indexation documentaire est particulièrement pertinente.*

**[16]** GARNIER Alain. L'information non structurée dans l'entreprise : usages et outils. Hermès Sciences. Paris, Lavoisier, 2007. 245 p. ISBN 978-2-7462-1605-1.

*L'auteur propose une série d'indicateurs et d'outils opérationnels pour analyser et optimiser l'usage de l'information non structurée. Il place le moteur de recherche comme vecteur et élément central de cette optimisation.*

**[17]** LE FOLL Laurent, COUILLAULT Alain. Valorisation de l'information non structurée. [En ligne]. Paris, Aproged, octobre 2007. [Consulté le 10 Juin 2013]. <http://www.aproged.org/index.php/Demarrer-telechargement/Publications/6-Livre-Blanc-Valorisation-de-l-information-non-structuree.html>

*Ce livre blanc propose une vision globale, stratégique et managériale de l'information non structurée en définissant son cadre d'application et sa valeur pour les organisations. En annexe, des user cases sont présentés et commentés (Par exemple, Temis et sa solution d'enrichissement automatique des métadonnées).*

[18] Jean-Max NOYER, Maryse CARMES. Les interfaces machiniques comme problème sémio-politique, in Broudoux E. et Chartron G. (dir), Enjeux politiques du document numérique, eds ADBS, 2010, p. 193-216.

*Cet article propose une lecture transversale des notions qui entourent la question de la recherche d'information et des points d'application de la politique des moteurs de recherche. Les auteurs décrivent également les modèles sociotechniques qui sous-tendent les processus de recherche chez les usagers des systèmes d'information.*

## **Mémoires et travaux universitaires**

### **Analyse des besoins utilisateurs**

[19] ANOBIS Amandine. Choix et mise en place d'une méthodologie d'enquête de besoin dans le cadre d'une démarche marketing : le cas d'un cabinet de conseil en stratégie et management. 2009, 125 p. Mémoire, Sciences de l'information, INTD - CNAM, 2009

*Dans ce mémoire, l'auteur décrit les mutations inévitables à prendre en compte et les changements organisationnels nécessaires lors de la mise en place de nouveaux services documentaires.*

### **Gouvernance de l'information**

[20] HUGONIE Elsa. Construire un plan de gouvernance de l'information dans un grand groupe industriel : le cas de Thales Services. 2011, 118 p. Mémoire, Sciences de l'information, INTD - CNAM, 2011

*L'auteur amorce des pistes de réflexion concernant l'instauration d'une gouvernance documentaire transversale, notamment dans le cadre de la gestion d'un flux important de documents et de données.*

### **Management de l'information - utilisation des TIC**

[21] POULET Céline. La gestion de projet d'un déploiement ECM au sein d'un groupe industriel managé par processus. 2012, 112 p. Mémoire, Sciences de l'information, INTD - CNAM, 2012

*Ce mémoire développe la thématique de l'ECM à travers ses différentes composantes : organisationnelles, techniques, technologiques et financières, L'auteur propose une analyse de la création de valeur par l'outil et par les indicateurs appropriés pour mesurer le retour sur investissement.*

## Moteur de recherche

[22] HELMLINGER Johanna. Optimisation d'un moteur de recherche : implications organisationnelles et techniques. 2003, 51 p. Mémoire, Sciences de l'information, INTD - CNAM, 2003

*L'auteur propose une réflexion sur l'environnement de travail du moteur, les spécificités des documents qu'il doit traiter et les fonctionnalités d'analyse (statistiques et linguistiques) qu'il doit intégrer pour répondre au mieux aux besoins utilisateurs.*

[23] HERIGAULT Myriam. Moteur de recherche d'entreprise : déploiement du moteur sémantique Exalead à la R&D de Diagnostica Stago. 2012, 90 p. Mémoire, Sciences de l'information, INTD - CNAM, 2012

*Ce mémoire décrit les modalités de mise en place d'un accès unique à l'information par le biais d'une application de recherche associée à des fonctionnalités sémantiques. L'approche globale et thématique (notamment dans la partie bibliographique) voulue par l'auteur est extrêmement pertinente.*

[24] IMBERT Xavier. Choisir un moteur de recherche en open source : le cas de la base de données de Nomination. 2012, 69 p. Mémoire, Sciences de l'information, INTD - CNAM, 2012

*L'auteur présente les conditions nécessaires à l'amélioration d'un moteur de recherche existant dans le cadre d'un projet de restructuration documentaire. L'approche purement techniciste de ce dernier est particulièrement intéressante.*

## Moteur de recherche d'entreprise : aspects sémantiques

[25] AUTONOMY. A Unique Combination of Technologies. In Autonomy.com [site], Cambridge, Autonomy Corp., mise en ligne en septembre 2009, [Consulté le 12 Juin 2013]. [http://publications.autonomy.com/pdfs/Power/White Papers/Autonomy Technology/](http://publications.autonomy.com/pdfs/Power/White%20Papers/Autonomy%20Technology/)

*Ce livre blanc explore la technologie et les fonctionnalités de recherches conceptuelles intégrées au moteur AUTONOMY IDOL. Ce document est particulièrement pertinent dans sa description des différents stades d'évolution des applications de recherche et dans sa vision du concept de « Pan-Enterprise Search ».*

[26] POIBEAU Thierry. Sur le statut référentiel des entités nommées. [En ligne]. Villetaneuse, Poibeau Thierry, 2005, [Consulté le 12 Juin 2013]. <http://hal.inria.fr/docs/00/03/74/98/PDF/taln05-poibeau.pdf>

*Après avoir défini ce que sont des entités nommées et rendu compte de leurs méthodes de d'extraction et de catégorisation, l'auteur précise la nature polysémique de ces entités ainsi que la difficulté d'automatiser leur reconnaissance de façon univoque.*

[27] REAMY Tom. Enterprise Content Categorization – The business strategy for a semantic infrastructure, White paper. [En ligne]. Kaps Group, 2010, [Consulté le 12 Juin 2013]. <http://www.kapsgroup.com/presentations.shtml>

*Ce livre blanc très complet aborde les limites des méthodes statistiques d'analyse linguistique et du traitement automatique intégral. L'auteur y souligne l'importance des vocabulaires d'indexation contrôlés. Il définit également les critères optimaux d'une catégorisation de contenus en adoptant une vision stratégique.*

[28] YONNET Philippe. Vers un moteur de recherche conceptuel grâce à la sémantique ? Lettre Recherche & Référencement, Abondance, janvier 2011, n°122. [consultable via abonnement à la lettre professionnelle], pp. 19-29.

*Dans cet article, l'auteur confronte les moteurs de recherches syntaxiques et les moteurs de recherche sémantiques ou conceptuels et décrit les raisons pour lesquelles les moteurs syntaxiques restent pour l'instant plus performants que les moteurs de recherche dits « conceptuels ».*

## **Moteur de recherche d'entreprise : marché et tendances**

[29] INTRANET FOCUS. Enterprise Search trends and developments – Research note. [En ligne]. Horsham, Intranet Focus Ltd, juin 2012, [Consulté le 12 Juin 2013]. <http://www.intranetfocus.com/resources/downloads>

*Cette étude de marché porte sur le regain d'intérêt récent pour les applications de recherche, la segmentation du marché et les tendances émergentes telles que le Big data ou la recherche sur application mobile. Un chapitre entier est consacré à Sharepoint, à ses capacités d'interrogation et de recherche.*

[30] LEFASSY Olivier, How Agile Enterprise Search Infrastructure Can Help Cios, PolySpot [site], [Consulté le 12 Juin 2013] <http://www.polyspot.com/en/resources/documentation/how-agile-enterprise-search-infrastructure-can-help-cios.html>

Ce support de communication destiné à un public de décideurs décrit les technologies qui orchestrent la solution proposée par l'éditeur Polyspot.

[31] MARTIN James A. How to evaluate enterprise search options. CIO [en ligne], CXO Media, 8 février 2012, [Consulté le 12 Juin 2013].

[http://www.cio.com/article/699793/How\\_to\\_Evaluate\\_Enterprise\\_Search\\_Options?page=2&taxonomyId=3002](http://www.cio.com/article/699793/How_to_Evaluate_Enterprise_Search_Options?page=2&taxonomyId=3002)

*L'auteur présente dans un premier temps les différentes typologies d'éditeurs présents sur le marché des moteurs de recherche d'entreprise et les répartit en 3 catégories distinctes. Il donne quelques conseils méthodologiques concernant l'implantation d'une application de recherche.*

[32] OWENS Leslie. Market overview: enterprise search - Google, Microsoft and Autonomy face credible competitors. [En ligne]. Forrester, 2 septembre 2011, [Consulté le 14 Août 2013]. <http://public.dhe.ibm.com/common/ssi/ecm/en/iml14354usen/IML14354USEN.PDF>

*Cette étude compare 12 éditeurs de solutions de recherche. L'auteur complète l'étude en listant un certain nombre de bonnes pratiques et de recommandations afin de s'assurer de l'alignement du moteur de recherche avec la stratégie globale de l'entreprise.*

[33] WHIT Andrews. MarketScope for Enterprise Search. In Gartner [site], Stamford, Gartner, mise en ligne le 22 novembre 2010, [Consulté le 12 Juin 2013].

[http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/www.google.com/fr/enterprise/search/files/Gartner\\_MarketScopeForEnterpriseSearch.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/fr/enterprise/search/files/Gartner_MarketScopeForEnterpriseSearch.pdf)

*Cette étude liste les principales caractéristiques ainsi que les forces et les faiblesses des éditeurs qui dominent ce marché et des éditeurs de taille plus modeste. L'évaluation de Microsoft et de sa stratégie future est particulièrement éclairante.*

[34] WHITE Martin. Enterprise Search, 1er édition. O'Reilly Media, 2012, 245 p. ISBN 978-1-449-33044-6.

*L'auteur dresse un panorama de l'offre qui existe actuellement en matière de solutions d'applications de recherche en entreprise. Il liste également 12 facteurs qui conditionnent la réussite d'un projet de moteur de recherche. Cet ouvrage est extrêmement pertinent dans son approche transversale, pragmatique et très « orientée utilisateur » (User Centric).*

## **Moteur de recherche d'entreprise : spécificités et nouveaux positionnements**

[35] BALMISSE Gilles, La recherche d'information n'est pas l'accès à l'information, Le journal du net [en ligne], Avril 2007. [Consulté le 10 Juin 2013]

<http://www.journaldunet.com/solutions/0704/070418-3q-gilles-balmisse.shtml>

*Cet article propose une vision stratégique et non techniciste de la recherche d'information. L'auteur préconise une série de bonnes pratiques à mettre en place, notamment l'optimisation et l'harmonisation des métadonnées à l'échelle de l'organisation.*

**[36]** BALMISSE Gilles. Moteur de recherche en entreprise, la pertinence en question. IT-expert [en ligne], Press & Communication France, mars-avril 2009, n°78, pp. 32-40 [Consulté le 14 Août 2013]. <http://www.it-expertise.com/it-expert-n-78-marsavril-2009/>

*En introduction, l'auteur détaille les principes de fonctionnement d'une application de recherche, notamment l'acquisition des données et le pipeline de traitements linguistiques, de la requête utilisateur à la présentation des résultats. Une explication très claire du principe de pertinence est donnée ainsi que ses modes de calcul.*

**[37]** LARDY Jean-Pierre. Méthodes de tri des résultats des moteurs de recherche [pdf en ligne], Université Claude Bernard - Lyon I [Consulté le 05 Septembre 2013]. [http://archivesic.ccsd.cnrs.fr/docs/00/06/20/56/PDF/sic\\_00000053.pdf](http://archivesic.ccsd.cnrs.fr/docs/00/06/20/56/PDF/sic_00000053.pdf)

*Ce court article revient sur les différentes notions de tris utilisées par un moteur de recherche sur le web ainsi que sur le mode de fonctionnement basique de toute application de recherche quelque soit sa sophistication.*

**[38]** BENNETT Mark. 20+ Differences Between Internet vs. Enterprise Search – And Why You Should Care. In New Idea Engineering [site], Santa Clara, New Idea Engineering Inc., mise en ligne en février 2008, [Consulté le 10 Juin 2013]. <http://www.ideaeng.com/inet-enterprise-search-p1-0502>

*Cet article souligne la nature mixte de l'information en entreprise faite de données structurées et non structurées. L'auteur positionne le moteur de recherche comme un moyen de briser les silos d'information notamment à l'aide de la recherche fédérée.*

**[39]** FREYERMUTH Jérôme, GUIZIOU Erwan. Le moteur de recherche : Clé de voûte du système d'information. [pdf en ligne]. Voirin Consultants, 2013. 38 p. [Consulté le 12 Août 2013] <http://www.voirin-consultants.com/fr/management/actualites/publications/livres-blancs/le-moteur-de-recherche>

*Ce livre blanc consacré au fonctionnement et aux enjeux de la recherche d'information en entreprise propose une vue d'ensemble des technologies ainsi qu'un panorama des acteurs et des solutions du marché actuel de la recherche. Les auteurs consacrent un chapitre entier aux critères de différenciation entre un moteur de recherche web et un moteur de recherche d'entreprise.*

**[40]** DEBONNE Eric. Moteur de Recherche Internet versus Entreprise. In solaci.com [blog], Paris, mise en ligne le 22 novembre 2007, [Consulté le 10 Juin 2013] <http://www.solaci.com/blog/2007/11/moteur-de-recherche-internet-versus.html>

*L'auteur définit 5 critères de comparaison afin d'évaluer les spécificités respectives des moteurs généralistes web et des moteurs de recherche d'entreprise. Il rappelle notamment que l'exhaustivité est primordiale dans un contexte d'information professionnelle.*

**[41]** FORMOSA Pierre. Les SBA ouvrent le décisionnel à la sémantique et le rendent (enfin) user-friendly ! In Decideo.fr [site], Arche Numérique Médias, mise en ligne le 17 octobre 2011. [Consulté le 10 Juin 2013]. <http://www.decideo.fr/Les-SBA-ouvrent-le-decisionnel-a-la-semantique-et-le-rendent-enfin-user-friendly-a4599.html>

*L'auteur introduit le concept de « Business Intelligence étendue » et envisage une convergence des applications de recherche et des applications de BI classiques. Cet article souligne le rôle central du moteur de recherche dans la mise en valeur des données non-structurées.*

**[42]** SEARCH TECHNOLOGIES. Federated Search: The Options. In Search Technologies [site], Herndon, Search Technologies Corporation, [Consulté le 10 Juin 2013]. <http://www.searchtechnologies.com/federated-search.html>

*Cet article technique porte sur les 3 variantes pouvant être mises en œuvre dans le cadre du déploiement d'un module de recherche fédérée. L'approche schématique est particulièrement parlante.*

**[43]** SEARCH TECHNOLOGIES. What Every CIO Needs to Know About Enterprise Search and Search Engines. [En ligne]. Herndon, Search Technologies Corporation, [Consulté le 12 Juin 2013]. <http://www.searchtechnologies.com/search-engine-white-papers.html>

*Ce court document à destination de décideurs propose des éléments de réflexion en énumérant 5 points clés qui conditionnent le succès d'une application de recherche. Leurs recommandations sur la qualité des métadonnées et la croissance informationnelle sont très éclairantes.*

**[44]** WHITE Martin. The Sorry State of Search Satisfaction. Econtent Magazine [en ligne], Information Today Inc., 18 mars 2010. [Consulté le 12 Juin 2013]. <http://www.econtentmag.com/articles/articlereader.aspx?articleid=61565>

*Cet article est consacré au degré de satisfaction des utilisateurs vis-à-vis de leur application de recherche. L'auteur divulgue certains chiffres clés de l'étude annuelle récurrente « Global Intranet Trends » menée par la consultante Jane McConnell.*

## Moteur de recherche d'entreprise : stratégie et mise en œuvre

[45] INTRANET FOCUS. Developing an enterprise search strategy – Research note. [En ligne]. Horsham, Intranet Focus Ltd, Janvier 2013, [Consulté le 15 Juillet 2013]. <http://www.intranetfocus.com/resources/downloads>

*Cette note de recherche présente un cadre méthodologique pour le développement d'une stratégie de recherche d'entreprise efficace. L'auteur préconise notamment la prise en compte des besoins utilisateurs, la bonne gestion du cycle de vie des informations et une approche technologique.*

[46] INTRANET FOCUS. Enterprise Search Team Management – Research note. [En ligne]. Horsham, Intranet Focus Ltd, janvier 2012, [Consulté le 10 Juin 2013]. <http://www.intranetfocus.com/wp-content/uploads/Enterprise-Search-Team-Management.pdf>

*L'auteur définit le cadre d'une bonne gestion de l'application de recherche par la constitution d'une équipe en charge de l'implémentation et dédiée à la maintenance des référentiels linguistiques, à l'amélioration continue de la satisfaction des utilisateurs.*

[47] JUNG Marie. Bien choisir un moteur de recherche d'entreprise. 01net Entreprises [en ligne], NextInteractiveMedia, 29 janvier 2009. [Consulté le 12 Juin 2013]. <http://pro.01net.com/editorial/402884/bien-choisir-un-moteur-de-recherche-dentreprise/>

*Cet article liste un certain nombre de critères fonctionnels et technologiques ainsi que des points de vigilance à prendre en compte dans le choix d'un moteur de recherche d'entreprise.*

[48] MINDMETRE. Mind the Enterprise Search Gap, A MindMetre research report analyzing the gap between enterprise search expectations and real-life experience. [En ligne]. MindMetre Research, 2011, [Consulté le 12 Juin 2013]. <http://www.smartlogic.com/home/knowledge-zone/white-papers/1600-mindmetre-research-report-sponsored-by-smartlogic>

*Résultats d'une enquête internationale menée auprès de 2000 responsables d'entreprises Nord Américaines et Européennes. L'étude révèle des chiffres très disparates entre les différents territoires suivis et spécifiquement sur les attentes vis-à-vis des outils de recherche et en termes de satisfaction utilisateurs.*

## Recherche d'information et traitement des langues

[49] RUSSELL-ROSE Tony. Taxonomy of Enterprise Search and Discovery. In Information Interaction [blog], Guildford, Russell-Rose Tony, mise en ligne le 2 novembre 2011, [Consulté le 23 Août 2013]. <http://isquared.wordpress.com/2011/11/02/a-taxonomy-of-enterprise-search-and-discovery>

*Cette étude décrit les modes de recherche et de découverte d'informations en entreprise. L'auteur étend la notion théorique de recherche d'information (IR : Information Retrieval) vers une notion plus opérationnelle et plus en phase avec les objectifs des utilisateurs en l'entreprise.*

[50] COUILLAULT Alain, DEBONNE Eric, GARNIER Alain. Le traitement automatique des langues dans les industries de l'information. GFII, 2005. 65 pages. [Consulté le 25 Juillet [www.technolangue.net/IMG/pdf/LB\\_GRIIIL.pdf](http://www.technolangue.net/IMG/pdf/LB_GRIIIL.pdf)

*Ce livre blanc présente les bénéfices et les principes techniques du traitement automatique des langues. La deuxième partie propose, en annexe, des user cases appliqués à différents secteurs de l'économie.*

[51] ERTZSCHEID Olivier, GALLEZOT Gabriel. Chercher faux et trouver juste, Serendipité et recherche d'information. [Consulté le 25 Juillet 2013] [http://archivesic.ccsd.cnrs.fr/sic\\_00000689/](http://archivesic.ccsd.cnrs.fr/sic_00000689/)

*Cet article s'intéresse au cycle de l'information et plus spécifiquement à la construction de connaissances dans les processus de transformation et de recherche d'information. Les auteurs proposent trois modèles de recherche d'informations centrés sur l'utilisation du moteur de recherche.*

## ROC, ROI et gestion de contenu

[52] AIIM. Capitalizing on Content: A Compelling ROI for Change [en ligne]. AIIM, 2011. 22 p. [consulté le 29 Juillet 2013] [http://www.aiim.org/pdffdocuments/MIWP\\_Capitalizing-on-Content\\_2011.pdf](http://www.aiim.org/pdffdocuments/MIWP_Capitalizing-on-Content_2011.pdf)

*Cette étude présente certains facteurs de retour sur investissement liés à l'efficacité opérationnelle de la mise en œuvre d'un ECM. Très riche en données chiffrées, cette enquête est extrêmement pragmatique dans son approche, tant sur le plan du ROI que du « chaos informationnel » constaté dans les organisations répondantes. Les considérations propres à un ECM développées dans cette étude sont transposables aux problématiques de moteur de recherche.*

[53] PONCIER Anthony. Entreprise 2.0 : Faut-il mesurer le ROI ou la satisfaction ? [blog], mise en ligne le 09 Décembre 2009. <http://poncier.org/blog/?p=997> [Consulté le 28 Août 2013]

*Dans ce cours article de blog, l'auteur pose la question de la pertinence des indicateurs classiques pour juger du ROI des applications 2.0 en entreprise. Il donne également quelques données chiffrées tirées de l'enquête Cisco 2008 sur l'entreprise collaborative*

[54] LESPERANCE Luc, ULMER Galadrielle. L'entreprise 2.0, ROI et analyse de la valeur. [pdf en ligne]. Voirin Consultants, 2011. 34 p.

<http://www.voirin-consultants.com/fr/management/actualites/publications/livres-blancs/l-entreprise-20-acte-iv-roi-et-analyse-de-la-valeur> [Consulté le 12 Août 2013]

*Les auteurs de cette étude démontrent la création de valeur sur le long terme par la mise en place d'outils collaboratifs ou de type « 2.0 » en entreprise. Cette étude est particulièrement intéressante dans son utilisation de schémas et de graphiques.*

## **Technologies de l'information et connaissances**

[55] BOURDONCLE François. L'intelligence collective d'usage, in NOYER, JM., JUANALS, B.(dir), Technologies de l'Information et Intelligences Collectives, éditions Hermès-Lavoisier, avril 2010, chap.4, p139-152

*Cet article aborde le problème de la structuration des données non structurées à travers l'usage particulier des savoirs collectifs et reposant sur une approche émergente (dite « bottom-up »), par opposition à l'approche du web sémantique qui repose sur une approche normative des savoirs et des connaissances (dite « top-down »).*

[56] BALMISSE Gilles. Gestion des connaissances - Outils et applications du knowledge management. Paris, Vuibert, 2002. 259 p. Collection Entreprendre informatique. ISBN 2-7117-8697-8

*Cet ouvrage définit la gestion des connaissances, ses origines et ses domaines d'application. L'auteur développe le concept de Knowledge Management sous 3 angles : individuel, organisationnel et technologique et envisage la recherche d'information à travers son exploitation, sa structuration et sa diffusion.*

[57] ZACKALD Manuel. Évaluation des systèmes d'organisation des connaissances. Les Cahiers du numérique, 2010/3 Vol. 6, p. 133-166.

*Cet article propose une analyse des SOC (Systèmes d'Organisation des Connaissances) adaptés à la gestion de l'information documentaire numérique. L'auteur différencie deux*

*types de SOC : manuels et automatiques. L'article est particulièrement éclairant sur l'utilisation des index de moteurs de recherche en tant que SOC automatiques.*

**[58]** LEVY Pierre. « Au-delà de Google » Les voies de « l'intelligence collective », *Multitudes*, 2009/1 n° 36, p. 45-52. DOI : 10.3917/mult.036.0045

*Cet article propose une analyse des freins et des limites du moteur de recherche Google, notamment sur l'absence de personnalisation des algorithmes de recherche et la pertinence des réponses associées aux requêtes des utilisateurs.*

## **Technologies de l'information et travail**

**[59]** GUYOT Brigitte. Sciences de l'information et activité professionnelle. [En ligne] CNRS Editions. 2004. Hermès, n° 38. pp. 38-45. [consulté le 29 Juillet 2013] <http://documents.irevues.inist.fr/handle/2042/9422?show=full>

*L'auteur de cet article s'interroge sur le positionnement des sciences de l'information au sein des organisations ainsi que sur une nouvelle médiation entre les usagers des SI et les gestionnaires de l'information. L'auteur propose un modèle centré sur « l'usage » des informations et leur adéquation avec les besoins utilisateurs.*

**[60]** ISAAC Henri, CAMPOY Eric, KALIKA Michel. Surcharge informationnelle, urgence et TIC. L'effet temporel des technologies de l'information. in revue *Management & Avenir*, vol. 12, 2007 [consulté le 29 Juillet 2013]

[http://halshs.archives-ouvertes.fr/docs/00/15/51/19/PDF/HIM\\_A.pdf](http://halshs.archives-ouvertes.fr/docs/00/15/51/19/PDF/HIM_A.pdf)

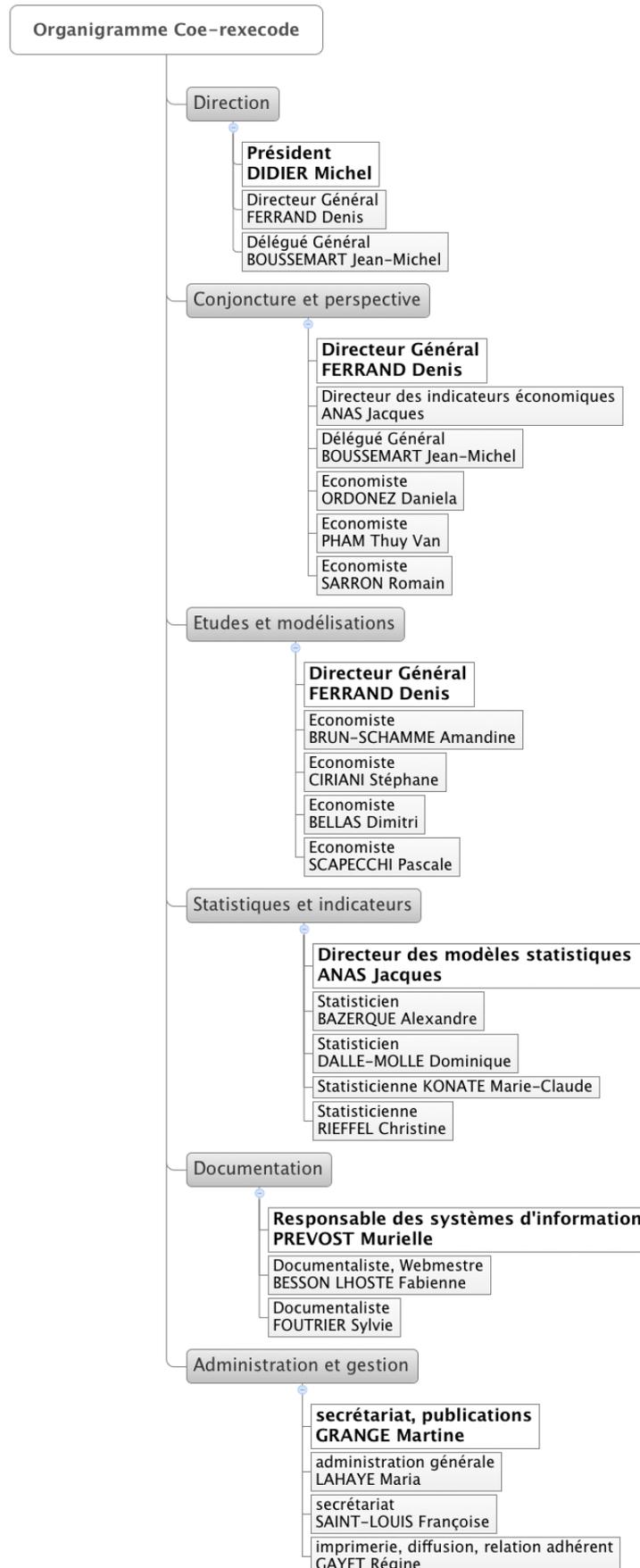
*Cet article présente les résultats d'une analyse de données collectées auprès de 12000 salariés sur une période de 5 ans. Les auteurs s'appuient sur la notion de surcharge informationnelle et montrent en quoi les TIC ont un rôle « aggravant » dans la perception de surcharge d'activité et d'urgence informationnelle.*

**[61]** VACHER Béatrice. Les enjeux de la manutention de l'information. *Système d'Information et Management*. 1998. pages 65-83. Ed. ESKA, vol 2. [consulté le 29 Juillet 2013] <http://www.revuesim.org/sim/article/view/45>

*Malgré une facilité de traitement accrue par l'utilisation des TIC, la gestion opérationnelle de l'information reste une activité chronophage constituée de différentes petites tâches et obligations pour les utilisateurs des systèmes d'information. L'auteur prend appui sur des exemples concrets pour illustrer son propos.*

# **Annexes**

# Annexe 1 : Organigramme de Coe-Rexecode



## Annexe 2 : Macro-planning et phasage

Macro planning du projet - Phase 1 (15 jours)			Macro planning de charge du projet
Intitulés des étapes	Date de début	Date de fin	Ressources internes
Cadrage et planification du projet	17/06/13	18/06/13	Murielle Prévost, Jean-Yves Vuillequez
Rédaction du projet de questionnaires	18/06/13	20/06/13	Jean-Yves Vuillequez, équipe du pôle doc
Validation du projet de questionnaires	20/06/13	21/06/13	Murielle Prévost, Jean-Yves Vuillequez
Lancement des entretiens individuels	24/06/13	28/06/13	Jean-Yves Vuillequez, autres ressources internes Coe-Rexecode
Analyse des entretiens et rédaction du livrable	28/06/13	03/07/13	Jean-Yves Vuillequez
<b>Livrable n°1 : audit de l'existant</b>	-	04/07/13	Murielle Prévost, Denis Ferrand, Jean-Yves Vuillequez
Présentation de l'audit	-	05/07/13	Murielle Prévost, Denis Ferrand, Jean-Yves Vuillequez

Macro planning du projet - Phase 2 (25 jours)			Macro planning de charge du projet
Intitulés des étapes	Date de début	Date de fin	Ressources internes
Optimisation du moteur de recherche	08/07/13	09/08/13	Murielle Prévost, Jean-Yves Vuillequez, équipe doc
Mise en œuvre actions correctives suite phase 1	05/07/13	09/08/13	Jean-Yves Vuillequez, équipe doc
<b>Livrable n°2 : référentiel de mots clés</b>	05/07/13	19/07/13	Murielle Prévost, Jean-Yves Vuillequez, équipe doc
Livrable n°3 intermédiaire (validation des axes)	02/07/13	22/07/13	Murielle Prévost, Jean-Yves Vuillequez
<b>Livrable n°3 : benchmark solutions moteurs</b>	08/07/13	02/08/13	Murielle Prévost, Denis Ferrand, Jean-Yves Vuillequez
<b>Livrable n°4 : bilan et préconisations</b>	05/08/13	09/08/13	Murielle Prévost, Jean-Yves Vuillequez, équipe du pôle doc

## Annexe 3 : Guide d'entretien

Thématiques	Questions	Remarques	Validation
<b>SHAREPOINT (APPRECIATION GENERALE)</b>			
<b>Intégration dans le processus métier</b>	<p>Pouvez-vous me décrire votre activité ?</p> <p>Comment s'insère Sharepoint dans votre activité journalière ?</p> <p>Quelles sont selon vous les fonctions utiles pour la conduite de votre activité ?</p> <p>Quelles étaient vos attentes initiales vis-à-vis de l'outil ?</p> <p>Quels sont les écarts constatés entre les attentes initiales et la réalité actuelle ?</p>	<p><b>Faire manipuler l'utilisateur</b> : utilisation de la solution, mode de navigation, création de sites, création d'alertes, modes de consultation</p>	-
<b>La fonction recherche dans Sharepoint</b>	<p>Satisfaction aux requêtes / recherches demandées ? fluidité ? pertinences des réponses ?</p> <p>Quels sont les motifs d'abandon ? : Trop de réponses ? pas assez de réponses ?</p> <p>Ergonomie : La barre de recherche est-elle correctement positionnée ? Est-elle facilement identifiable et visible ?</p>	<p>moteur de recherche (simple ou avancé), facilité de manipulation, évaluation des blocages</p>	-
<b>Utilisation / fréquentation globale de l'outil</b>	<p>Quel bilan tirez-vous 8 mois après le déploiement de l'outil ?</p> <p>L'outil est-il correctement positionné et répond-t-il aux attentes définies initialement ?</p> <p>Comment percevez-vous l'outil ? trop complexe ? accessible ? Ergonomique ? convivial ? peu intuitif ?</p> <p>Concernant le volet formation à l'outil :</p> <p>Estimez-vous avoir été assez préparé au basculement sur le nouvel outil ? Manque de formations spécifiques et ciblées ? Pas ou peu de temps à y consacrer ?</p>	<p>Croiser les résultats d'entretiens avec l'analyse des logs (évaluer l'adéquation de l'outil avec les pratiques métiers et les attentes)</p>	-
<b>TRAVAIL COLLABORATIF</b>			
<b>Intégration dans le processus métier</b>	<p>Quels sont les procédures qui nécessitent un travail collaboratif ? (échange de documents ? rédaction ? validation ? approbation)</p> <p>Quelle serait selon vous l'utilité de Sharepoint dans ces phases ?</p>	<p>Etablir une liste des activités collaboratives en amont des entretiens pour exhaustivité (réévaluation de la liste à postériori)</p>	-

<b>Partage des documents dans de Sharepoint</b>	<p>Avez-vous le réflexe de charger régulièrement des documents ou informations dans Sharepoint ?</p> <p>Si oui, quels sont les blocages (ou les facilités) rencontrés pour charger des documents dans Sharepoint ?</p> <p>Fonctionnalité perçue comme pas essentielle à la conduite du travail ? Procédures de chargement trop complexes ?</p> <p>Utilisation des sites dédiés aux équipes projets : Avez-vous le réflexe de créer des sites d'équipes projet et de les alimenter en contenu ?</p>	Identifier les points de blocages et les points d'améliorations éventuels ?	-
<b>Partage et stockage des documents en dehors de Sharepoint</b>	<p>Quel est le canal favorisé pour les échanges d'informations ? téléphone, mails, présentiel ?</p> <p>Utilisation du mail ? quelles modalités d'échange ? modalités de validation des documents (gestion des versions intermédiaires et définitives) ? Gestion des mails engageants ?</p>	proposer des solutions ultérieures pour une bascule vers Sharepoint → Passage du local au global (avec la gestion mixte du papier et du numérique)	-
<b>ACCES A L'INFORMATION</b>			
<b>La structuration de l'information dans Sharepoint</b>	Quels problèmes avez-vous identifié (liés l'architecture actuelle de SHAREPOINT) quand vous recherchez ou utilisez des informations ?	RAS	-
<b>La navigation dans les sites Sharepoint</b>	<p>La navigation dans les sites et sous-sites est-elle perçue comme fluide ?</p> <p>Les entêtes de menus sont-elles explicites ?</p> <p>Arrivez-vous à vous repérer dans l'arborescence actuelle ? Vous semble-t-elle claire ?</p>	Observations et manipulation utilisateurs	-
<b>La visibilité / remontées des informations primordiales</b>	<p>Systèmes d'alerte : régularité suffisante ? Nombre d'alertes trop important ? Alertes non pertinentes ?</p> <p>Trouvez-vous pertinente l'information mise à disposition sur les pages vous concernant plus spécifiquement ? (pages pays par exemple)</p>	Définir chaque profil d'utilisateurs (s'appuyer sur l'étude exploratoire)	-
<b>CHANGEMENTS / ACTIONS CORRECTIVES SUGGEREES (QUESTIONS OUVERTES)</b>			
<b>Dans les processus collaboratifs / partage / accès de l'information</b>	Selon, vous, quels sont les 3 axes majeurs à améliorer dans l'utilisation et le paramétrage actuel de Sharepoint	Voir listing des questions ouvertes	-

<p><b>Dans le fonctionnement / paramétrage actuel de Sharepoint</b></p>	<p>Sur la recherche ?          Sur la structuration de l'information ?          Sur le partage des documents ?          Sur la structuration des pages (ergonomie) ?          Sur les éléments qui ne sont pas présents et que vous souhaiteriez y trouver ?          Sur la disponibilité des informations et des documents ?</p>	<p>Sur la recherche : identifier et évaluer la pertinence des réponses (fonctionnement «à la google »? )          Sur la disponibilité : évaluer le ou les types de documents à forte valeur en fonction du profil utilisateur</p>	<p>-</p>
---	--	--	----------

## Annexe 4 : Listing des « meilleurs résultats » par mots clés

### Listing des organismes et zones géographiques<sup>90</sup>

Chine	Pages pays	Espagne
Asie	Ernst & Young	Etats-Unis
Consensus	JP Morgan	France
Russie	Fonds Monétaire International	Italie
Europe de l'Est	IHS-Global Insight	Japon
INSEE	Morgan Stanley	Royaume-Uni
infos rapides	morgan	Afrique et Moyen-Orient
insee première	Banque Asiatique de développement	Amérique latine
OCDE	Banque Centrale Européenne	Amérique
Agence Internationale de l'Energie	CEPII	Union Européenne
ONU	EUREN	Europe
Onu-Afrique	AIECE	Conseil d'Analyse Economique
Onu-ESCAP	EFN	
Onu-ESCWA	EFN EUROFRAME	
social commission	Cour des comptes	
Perspectives	Standard & Poor's	
Perspectives Allemagne	SESSI	
Perspectives France	OFCE	
Perspectives OFCE	PWC	
Perspectives OCDE	Oxford Economics	
Perspectives FMI	OMC	
Perspectives Banque mondiale	Moody's	
Perspectives BCE	KPMG	
Perspectives Commission Européenne	IRES	
Natixis	INED	
Eviews	IFF	
Banque Mondiale	HSBC	
Allemagne	CESE	

<sup>90</sup> Listing arrêté au 01/08/13

## Annexe 5 : Extrait du référentiel de mots clés pour les « meilleurs résultats »

Concept / Organisme lié	Mot clé	Synonymes	Titre / Meilleurs résultats
Asie	Chine	page chine;china	Page Intradoc - Chine
Asie	Asie	zone asie;inde;indonésie;indonesie;taiwan;thaïlande;philippines;malaisie;hong-kong;singapour	Page Intradoc - Asie
consensus	Consensus	Consensus Economics;Consensus forecasts;consensus economics;economics	Page Intradoc - Consensus Economics
consensus	Consensus	consensus economics;consensus forecasts;consensus forecast	Dossier Consensus Economics dans Perspectives
Europe de l'Est	Russie	federation de russie;federation russie	Page Intradoc - Russie
Europe de l'Est	Europe de l'Est	europe est;est europe ;est;ukraine;croatie	Page Intradoc - Europe de l'Est Page Intradoc - Russie
insee	INSEE	insee;statistiques insee;stats insee;statistiques;statistiques France	Page Intradoc - INSEE
insee	INSEE	documents insee;bibliothèque insee;bibliotheque insee	Bibliothèque de documents - INSEE
insee	infos rapides	info rapide	INSEE - infos rapides
insee	insee première	insee premiere;premiere insee;première;premiere	INSEE - première
ocde	OCDE	oecd	Page Intradoc - OCDE
ocde	OCDE	documents ocde;bibliothèque ocde;bibliotheque ocde;documents oecd;bibliothèque oecd;bibliotheque oecd	Bibliothèque de documents - OCDE
ocde	Agence Internationale de l'Energie	International Energy Agency;iea;aie;agence internationale énergie;agence internationale energie	Bibliothèque de documents - AIE
Agence Internationale de l'Energie	Agence Internationale de l'Energie		Données de l'Agence Internationale de l'Energie
onu	ONU	un;united nations	Bibliothèque de documents - ONU
onu	Onu-Afrique	onu afrique	Bibliothèque de documents - ONU-Afrique

onu	Onu-ESCAP	asia pacific;escap;escap forecasts;escap forecast;escap surveys;escap survey;economic and social commission for asia and pacific	Bibliothèque de documents - ONU-ECSAP
onu	Onu-ESCWA	western asia;escwa;escwa forecasts;escwa forecast;escwa surveys;escwa survey;economic and social commission for western asia	Bibliothèque de documents - ONU-ESCWA
onu	social commission		Bibliothèque de documents - ONU-ECSAP Bibliothèque de documents - ONU-ESCWA
perspectives - prévisions - outlook - forecast	Perspectives	perspective;prévisions;prevision;réunion perspectives;réunion perspective;forecast;outlook;forecasts;perspectives internationales	Page Intranet - Réunion perspectives
perspectives - prévisions - outlook - forecast	Perspectives Allemagne	perspective allemagne;prévisions allemagne;prevision allemagne;allemagne perspective;allemagne prévisions;allemagne prevision;outlook allemagne;allemagne outlook	Perspectives Allemagne
perspectives - prévisions - outlook - forecast	Perspectives France	perspective france;prévisions france;prevision france;france perspective;france prévisions;france prevision;perspectives économiques france;perspective économique france;perspectives finances publiques;perspective finances publiques	Perspectives France
perspectives - prévisions - outlook - forecast	Perspectives OFCE	perspectives de l'ofce;perspective ofce;ofce perspectives;ofce perspective;prévisions ofce;previsions ofce;prévision ofce;prevision ofce;ofce prévisions;ofce prévisions;ofce prevision;ofce prevision	Perspectives ofce
perspectives - prévisions - outlook - forecast	Perspectives OCDE	perspectives de l'ocde;perspectives économiques de l'ocde;perspectives économiques de l'ocde;perspectives économiques ocde;perspective ocde;prévisions ocde;previsions ocde;prévision ocde;prevision ocde;ocde prévisions;ocde prévisions;ocde prevision;ocde prevision	Perspectives OCDE
perspectives - prévisions - outlook - forecast	Perspectives FMI	world economic outlook;fmi world economic outlook;imf world economic outlook;world economic outlook fmi;world economic outlook imf	Perspectives FMI

perspectives - prévisions - outlook - forecast	Perspectives Banque mondiale	banque mondiale global economic prospects;global economic prospects;global economic prospect;banque mondiale perspectives;banque mondiale perpective	Perspectives Banque mondiale
perspectives - prévisions - outlook - forecast	Perspectives BCE	perspectives de la bce;perspective bce;bce perspectives;bce perspective;prévisions bce;previsions bce;prévision bce;prevision bce;bce prévisions;bce previsions;bce prévision;bce prevision	Perspectives BCE
perspectives - prévisions - outlook - forecast	Perspectives Commission Européenne	european economic forecast;commission européenne Forecast;commission européenne forecasts;commission europe Forecast;commission europe Forecasts; commission européenne perspectives;commission européenne prévisions;prévisions commission européenne;ue perspectives;ue perspective; perspectives ue;perspective ue	Perspectives Commission Européenne
	Natixis	artus;CDC;flash	Page Intradoc - Natixis
	Eviews		Page Wikitek - Eviews
	Banque Mondiale	la banque mondiale;the world bank;world bank	Page Intradoc - La Banque Mondiale
	Banque Mondiale	la banque mondiale;the world bank;world bank;documents la banque mondiale;bibliothèque the world bank;documents banque mondiale;bibliothèque world bank;	Bibliothèque de documents - La banque Mondiale
	Allemagne	page allemagne	Page Intradoc - Allemagne
	Allemagne	documents allemagne;bibliothèque Allemagne;bibliotheque allemagne	Bibliothèque de documents - Allemagne
	Espagne	page espagne	Page Intradoc - Espagne
	Espagne	documents espagne;bibliothèque Espagne;bibliotheque Espagne	Bibliothèque de documents - Espagne
	Etats-Unis	page etats-unis;page états-unis;usa;us;états- unis;états unis;etats unis;Amérique du nord;amerique du nord;Amérique nord;Amérique nord;page etats unis;page états unis;	Page Intradoc - Etats-Unis
	Etats-Unis	documents usa;bibliothèque usa;bibliotheque usa;documents us;bibliothèque us;bibliotheque us	Bibliothèque de documents - Etats-unis
	France	page france	Page Intradoc - France
	Italie	page italie	Page Intradoc - Italie
	Italie	documents italie;bibliothèque Italie;bibliotheque italie	Bibliothèque de documents - Italie

Asie	Japon	page japon	Page Intradoc - Japon
Asie	Japon	documents japon;bibliothèque japon;bibliotheque japon	Bibliothèque de documents - Japon
	Royaume-Uni	page royaume-uni;page uk;page angleterre;uk;angleterre;royaume uni	Page Intradoc - Royaume-Uni
	Royaume-Uni	documents uk;bibliothèque uk;bibliotheque uk;documents angleterre;bibliothèque angleterre;bibliotheque angleterre	Bibliothèque de documents - Royaume-Uni