

Université
de Toulouse

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Informatique

Présentée et soutenue par :

Anass EL HADDADI

le : vendredi 9 décembre 2011

Titre :

Fouille multidimensionnelle sur les données textuelles visant à extraire les réseaux sociaux et sémantiques pour leur exploitation via la téléphonie mobile

Ecole doctorale :

Mathématiques Informatique Télécommunications (MITT)

Unité de recherche :

IRIT - SIG/EVI - UMR 5505

Directeur(s) de Thèse :

Ilham BERRADA, Professeur, Université Mohamed V - Souissi, Rabat - Maroc

Bernard DOUSSET, Professeur, Université Toulouse III, Toulouse - France

Rapporteurs :

Abdelhadi FENNAN, Professeur, Université Abdelmalek Essaadi - FST, Tanger - Maroc

Michel LAMURE, Professeur, Université Lyon 1, Lyon - France

Membre(s) du jury :

Khalid EL HIMDI, Professeur, Université Mohamed V - Agdal, Rabat - Maroc --- Examineur

Brigitte GAY, Professeur, ESC de Toulouse, Toulouse - France --- Invitée

Christian LONGEVIALLE, Professeur, Université de Marne la Vallée, Paris - France --- Invité

Klaus Solberg Søylen, MC, Halmstad University - SBE, Halmstad - Suède --- Examineur

Anass EL HADDADI

Fouille multidimensionnelle sur les données textuelles visant à extraire les réseaux sociaux et sémantiques pour leur exploitation via la téléphonie mobile

Directeurs de thèse :

Directrice Marocaine : Ilham BERRADA, Professeur à l'Université Mohamed V – ENSIAS – Rabat, Maroc

Directeur Français : Bernard DOUSSET, Professeur à l'Université Toulouse III – IRIT – Toulouse, France

Résumé

La concurrence est un concept fondamental de la tradition libérale et des sciences économiques qui oblige les entreprises à pratiquer l'Intelligence Economique (IE) pour bien se positionner sur le marché ou tout simplement pour survivre. Mais souvent, ce n'est pas le plus fort qui survit, ni le plus intelligent, mais celui qui est le plus sensible au changement, facteur dominant dans la société actuelle. Les changements proviennent de l'extérieur ou naissent au sein même de l'entreprise et peuvent l'affecter plus ou moins durablement.

Dès lors, les entreprises sont appelées à rester constamment en veille pour guetter le moindre changement en vue d'y apporter la solution adéquate en temps réel. Cependant, pour une veille réussie, on ne doit pas se contenter uniquement de surveiller les opportunités, mais avant tout, d'anticiper les menaces. Malheureusement, cette veille se déroule en oubliant l'indispensable volet sécurité de l'IE à savoir la sécurité des données manipulées et celle des procédés suivis pour atteindre les objectifs de l'IE.

Nos travaux de recherche consistent à proposer un Système d'Intelligence Economique (SIE) Généraliste et Mobile. Ce SIE intègre une approche de modélisation du contrôle d'accès aux données et aux traitements pour sécuriser toutes les informations et les flux d'interaction durant son cycle de vie. Le besoin en matière de sécurité dans un SIE provient du fait que les informations manipulées sont d'ordre stratégique ayant une valeur assez importante. Une telle sécurité ne doit pas être considérée comme une option supplémentaire qu'offre un SIE pour se distinguer d'un autre. D'autant plus que la fuite de ces informations n'est pas le fait de faiblesses inhérentes aux systèmes informatiques des entreprises, mais c'est avant tout une question organisationnelle.

La mobilité présente un choix stratégique pour notre SIE "XPlor EveryWhere", dont le but est de permettre aux utilisateurs de nos solutions de veille de continuer à rechercher, surveiller, valider et rediffuser des informations stratégiques au cours de leurs déplacements. Ils n'ont ainsi plus besoin d'être assis face à leur ordinateur pour accéder à des données utiles dans l'instant (préparation d'une réunion, nouvel ordre du jour, information sur un interlocuteur, une technologie, un marché, demande urgente d'une analyse ou d'un focus spécifique). Il leur suffit simplement d'utiliser discrètement, et en toute sécurité, leur appareil mobile. Grâce à XPlor EveryWhere, ils sont en permanence au cœur de l'information pertinente.

Anass EL HADDADI

Multidimensional Text Mining to extract social and semantics networks for their operation via mobile phone

Supervisors:

Moroccan Supervisor: Ilham BERRADA, Professor at the University of Mohamed V – ENSIAS – Rabat, Morocco

French Supervisor: Bernard DOUSSET, Professor at the University of Toulouse III – IRIT - Toulouse, France

Abstract

Competition is a fundamental concept of the liberal economy tradition that requires companies to resort to Competitive Intelligence (CI) in order to be advantageously positioned on the market, or simply to survive. Nevertheless, it is well known that it is not the strongest of the organizations that survives, nor the most intelligent, but rather, the one most adaptable to change, the dominant factor in society today.

Therefore, companies are required to remain constantly on a wakeful state to watch for any change in order to make appropriate solutions in real time. However, for a successful vigil, we should not be satisfied merely to monitor the opportunities, but before all, to anticipate risks. The external risk factors have never been so many: extremely dynamic and unpredictable markets, new entrants, mergers and acquisitions, sharp price reduction, rapid changes in consumption patterns and values, fragility of brands and their reputation.

To face all these challenges, our research consists in proposing a Competitive Intelligence System (CIS) designed to provide online services. Through descriptive and statistics exploratory methods of data, Xplor EveryWhere display, in a very short time, new strategic knowledge such as: the profile of the actors, their reputation, their relationships, their sites of action, their mobility, emerging issues and concepts, terminology, promising fields etc.

The need for security in Xplor EveryWhere arises out of the strategic nature of information conveyed with quite a substantial value. Such security should not be considered as an additional option that a CIS can provide just in order to be distinguished from one another. Especially as the leak of this information is not the result of inherent weaknesses in corporate computer systems, but above all it is an organizational issue.

With Xplor EveryWhere we completed the reporting service, especially the aspect of mobility. Lastly with this system, it's possible to: View updated information as we have access to our strategic database server in real-time, itself fed daily by watchmen. They can enter information at trade shows, customer visits or after meetings.

Anass EL HADDADI

Fouille multidimensionnelle sur les données textuelles visant à extraire les réseaux sociaux et sémantiques pour leur exploitation via la téléphonie mobile

Directeurs de thèse :

Directrice Marocaine : Ilham BERRADA, Professeur à l'Université Mohamed V – ENSIAS – Rabat, Maroc

Directeur Français : Bernard DOUSSET, Professeur à l'Université Toulouse III – IRIT - Toulouse, France

Mots-clés

- Analyse de données, Fouille de texte (Text Mining), Fouille de Web (Web Mining), Fouille de données (Data Mining), entrepôt de données stratégique.
- Infométrie, Scientométrie, Bibliométrie.
- Système d'information, Système d'information stratégique, système d'intelligence économique.
- Intelligence économique, Veille stratégique, Veille Scientifique, Veille Technologique (Brevets).
- XPlor, Xplor Everywhere

Anass EL HADDADI

Multidimensional Text Mining to extract social and semantics networks for their operation via mobile phone

Supervisors:

Moroccan Supervisor: Ilham BERRADA, Professor at the University of Mohamed V – ENSIAS – Rabat, Morocco

French Supervisor: Bernard DOUSSET, Professor at the University of Toulouse III – IRIT - Toulouse, France

Keywords

- Data Analysis, Text Mining, Web Mining, Data Mining, Strategic Data Warehouse.
- Scientometrics, bibliometrics.
- Information Systems, Strategic Information Systems, Competitive Intelligence Systems.
- Competitive Intelligence, Strategic Watch, Scientific Watch, Technologic Watch.
- XPlor, Xplor EveryWhere

Je dédie cette thèse

à l'âme de ma grand-mère,

à mon grand-père,

à ma mère,

à mon père,

à Yasmine,

à Ahlam,

à Tarik,

à Amine

et

à Oaima

Remerciements

C'est avec un immense plaisir que j'écris aujourd'hui ces lignes et que je remercie toutes les personnes qui ont fait que ce travail de thèse arrive à son terme dans les meilleures conditions.

Je souhaiterais en premier lieu, exprimer ma gratitude à Messieurs le Professeur Claude CHRISMENT, responsable de l'équipe SIG, le Professeur Radouane MRABET, président de l'université Mohamed V – Souissi, ancien directeur de l'ENSIAS, pour m'avoir accueilli au sein de leurs équipes de recherche. Je les remercie pour leur rôle qu'ils ont joué dans la mise au point de ce travail et pour les conseils qu'ils m'ont prodigués.

Je tiens à exprimer ma profonde gratitude à Monsieur le professeur Bernard DOUSSET et Madame la professeur Ilham BERRADA qui ont été, pour moi, à la fois un(e) encadrant(e), un(e) guide et un(e) confident(e). De nombreuses idées sont nées de nos remue-ménages et de nos discussions aussi scientifiques que philosophiques. Je les remercie tout particulièrement de la confiance qu'ils m'ont témoignée, de leur clairvoyance et de leur grande disponibilité.

Un très grand merci à mes deux rapporteurs, dont la lecture approfondie de ce mémoire a permis d'en améliorer la qualité : Monsieur Michel LAMURE, Professeur de l'Université Lyon 1, et Monsieur Abdelhadi FENNAN, Professeur à l'Université Abdelmalik Essaidi –FST de Tanger.

Je tiens également à remercier Messieurs Klaus Solberg Soilen, Professeur à Halmstad University School of Business and Engineering, Khalid EL HIMDI, professeur à l'Université Mohamed V – Agdal de Rabat, pour l'intérêt qu'ils ont porté à mes travaux en examinant ce mémoire et pour l'honneur qu'ils me font en participant à ce jury.

Je tiens également à remercier Monsieur Chritian LONGEVIALLE, Professeur de l'Université de Marne la Vallée, pour l'honneur qu'il me fait en participant à ce jury.

Je tiens également à remercier tous les membres de l'équipe SIG avec qui j'ai pu travailler ou encore simplement partager de bons moments. Merci à

Messieurs Mohand BOUGHANEM, Gilles HUBERT, Taoufiq DKAKI, Max CHEVALIER, Olivier TESTE, Guillaume CABANAC ..., ainsi que Mesdames Maryse SALLES, Florence SEDES, Lynda TAMINE-LECHANI, Wahiba BAHSOUN... Un merci particulier à Josiane MOTHE, pour l'attention qu'elle a portée à l'avancement de mes travaux de recherche.

Mes pensées vont aussi pour tous les membres de l'équipe AL-BIRONI à l'ENSIAS. Je remercie très particulièrement Monsieur le doctorant Hatim HAMID, pour sa collaboration, son encouragement et son soutien tout au long de ces années de réflexion.

Bien entendu, ce travail ne serait pas ce qu'il est devenu sans les rencontres et les discussions avec les chercheurs qui ont croisé mon chemin au cours des conférences, sans leurs écrits qui ont été pour moi autant de chemins de lecture et qui ont apporté le ciment à cet édifice. Je tiens à remercier tout particulièrement Madame Brigitte GAY, professeur à l'ESC pour ses conseils, son professionnalisme et ses compétences et Monsieur Khalid EL HIMDI, professeur à l'Université Mohamed V – Rabat, pour son encouragement et son soutien.

Toute mon amitié va vers les doctorants, actuels ou anciens, de l'équipe ou de l'IRIT et en particulier Reda EL JOURANI, Ilham GHALLAMALAH, Eloïse LOUBIER, Dieudonné TCHUENTE, Grégory CLAUDE, Ana-Maria MANZAT, Hamdi CHAKER, Faten ATIGUI, Houssem JERBI, Lamjed Ben JABEUR, Damien DODOGNON, Léa LAPORTE, ...

Je tiens à remercier aussi tous les étudiants de SID, PINT, InfoCom, IDE auprès desquels j'ai effectué mon enseignement et en particulier ceux qui ont accepté d'évaluer le prototype résultant de mes travaux de recherche.

Je remercie l'équipe de choc : Fadwa, Youssef et Hamid pour avoir toujours été là, même dans les mauvais moments.

Enfin, je remercie très tendrement mes parents, mes sœurs, mes frères – qui m'ont supporté patiemment, et ont partagé mes préoccupations –, tous les membres de ma famille et tous mes amis qui m'ont encouragé et supporté pendant cette période.

Il y a un peu de chacun de vous dans ce travail. Merci à vous d'avoir donné à ce "nous", la force de cette première personne du pluriel...

Table des matières

<i>Résumé</i>	1
<i>Abstract</i>	3
<i>Mots-clés</i>	5
<i>Keywords</i>	7
<i>Remerciements</i>	11
<i>Table des matières</i>	15
<i>Introduction générale</i>	19
Contexte de travail & Problématique	19
Contributions	20
Organisation du mémoire	23
<i>L'Intelligence Economique ? Un processus de management stratégique de l'information</i>	25
1.1. Introduction	27
1.2. L'intelligence économique	28
1.2.1. Historique de l'intelligence économique	28
1.2.2. Elargissement géographique de l'intelligence économique	30
➤ Aux États-Unis	30
➤ En France	31
➤ Au Maroc	31
➤ En Asie	32
1.2.3. Définition de l'Intelligence Economique	34
1.3. L'analyse stratégique :	38
1.3.1. L'analyse de l'environnement	38
1.3.2. L'analyse de la concurrence	40
1.3.3. La synthèse SWOT	41
1.4. La veille :	42
1.4.1. La veille peu formalisée et passive :	43
1.4.2. La veille formalisée et proactive	43
1.4.3. La typologie de la veille	43
1.4.4. Processus de la veille	45
➤ Analyse des besoins	46
➤ Recherche et collecte d'informations	47
➤ La collecte des informations	48
➤ Traitement et analyse des informations	48
➤ Mémorisation et diffusion de l'information utile	49
1.5. La protection de l'information	49
1.6. Synthèse	50
<i>Chapitre 2.</i>	53
<i>Le système d'intelligence économique</i>	53
2.1. Introduction	55
2.2. Les systèmes d'information et la stratégie	56
2.2.1. L'émergence des systèmes d'information stratégique	56
2.2.2. La dimension interne des S-IS	57
2.2.3. La dimension externe des S-IS ; veille et intelligence économique	58
2.2.4. L'émergence de systèmes d'intelligence économique	59
2.3. Les méthodes de conception des systèmes d'intelligence économique	60
2.3.1. La méthode d'analyse MEDESIIE	60
2.3.2. Les modèles SITE	61

2.3.3.	Le modèle XPlor	62
➤	Analyse des besoins	62
➤	Recherche et collecte des informations	65
➤	Homogénéisation et structuration des données	66
➤	Traitement des structures de données	66
➤	Définitions des objets dans les méta-données	67
➤	Descripteurs de formats spécifiques	68
➤	Exemple de descripteur générique	68
2.4.	Les systèmes d'intelligence économique	70
2.5.	Le système d'intelligence économique "XPlor EveryWhere"	74
2.5.1.	Architecture service de XPlor EveryWhere	75
➤	Le service « Sources et traitements »	75
➤	Entrepôt de données stratégique	76
➤	Service Analyse et restitution	76
➤	Service Sécurité et Administration	77
2.5.2.	La structure statique de XPlor EveryWhere	77
2.5.3.	Architecture technique de XPlor EveryWhere	79
2.5.4.	Architecture logicielle du système	81
2.6.	Synthèse	83
Chapitre 3.	Entrepôt de données stratégique	85
3.1.	Introduction	87
3.2.	Le modèle de représentation multidimensionnelle des documents	87
3.2.1.	Définition de la présentation multidimensionnelle	90
3.2.2.	L'entrepôt de données stratégique	91
3.2.3.	L'entrepôt sous forme de matrice carée	92
3.2.4.	L'entrepôt sous forme d'une matrice rectangulaire	93
3.3.	Le modèle de calcul des indicateurs	94
3.3.1.	Définition d'un indicateur	94
3.3.2.	Les indicateurs dans l'entreprise	94
3.3.3.	Les indicateurs de l'activité scientifique et technologique	95
➤	Les indicateurs qualitatifs	95
➤	Les indicateurs quantitatifs	96
➤	Les indicateurs relationnels	97
3.3.4.	Le modèle d'indicateur	98
➤	La fonction de filtrage	98
➤	La fonction combinatoire	99
3.4.	Approche de sécurité d'un système d'intelligence économique	100
3.4.1.	Risques de sécurité dans un système d'intelligence économique	100
➤	Risques liés à l'importation des données pour le traitement et l'analyse	102
➤	Risques liés à l'utilisation de logiciels de collecte	102
➤	Risques liés aux sources d'information	102
➤	Risques liés aux aspects juridiques	102
➤	Risques liés aux procédures de traitement et d'analyse des données	102
➤	Risques liés à la livraison de l'information	102
3.4.2.	Approche proposée pour le contrôle d'accès à un SIE	103
➤	Politique de sécurité dans un SIE	103
➤	Au niveau de la collecte et de l'importation des données	103
➤	Au niveau du traitement et de l'analyse des données	104
➤	Au niveau de la restitution de l'information aux décideurs	104
3.4.3.	Mise en œuvre dans un cas pratique : Sécurisation de XPlor EveryWhere	104
➤	Contrôle d'accès aux données et aux traitements dans XPlor EveryWhere	105
➤	Modélisation des processus de la plateforme XPlor EveryWhere	105
3.4.4.	Application du modèle de contrôle d'accès RBAC.	105
➤	Etablissement des différents rôles et les tâches associées	106
➤	Les permissions aux objets à protéger	107

➤ L'infrastructure de sécurité	108
3.4.5. Application des ontologies de sécurité	108
3.5. Synthèse	109
Chapitre 4. Expérimentations et évaluation de XPlor EveryWhere	113
4.1. Introduction	115
4.2. Les limites de la plateforme Tétralogie	115
4.3. Les limites des premières versions du prototype Xplor	118
4.4. XPlor EveryWhere	119
4.4.1. Accès sécurisé aux analyses stratégiques	119
➤ Gestion de la base de données	121
➤ Alimentation de la base de données	123
➤ Schéma de la base de données	123
➤ Gestion des droit d'accès	124
4.4.2. Analyse de données	125
➤ Dépôt et suivie d'une étude	125
➤ Analyse descriptive	126
➤ Menu Filtrage	128
➤ Carte géostratégique	130
4.5. Expérimentations : analyse stratégique sur les scaffolds	131
4.5.1. Besoin informationnel	133
4.5.2. Source d'information	135
4.5.3. Structuration et homogénéisation des données	136
4.5.4. Présentation multidimensionnelle	137
4.5.5. Analyse et interprétation	138
➤ Analyse de réseaux des inventeurs	138
➤ Analyse de collaborations internationales	142
➤ Analyse de réseau sémantique	145
➤ Analyse de déposants	146
➤ Synthèse de l'analyse des scaffolds	147
4.6. Evaluation de XPlor EveryWhere	148
4.6.1. Méthodologie	148
4.6.2. Population étudiée	149
4.6.3. Résultats de l'enquête	151
➤ Niveau de difficulté des fonctionnalités XPlor EveryWhere	151
➤ Analyse de l'utilité de XPlor EveryWhere	158
4.7. Synthèse	160
Conclusion et perspectives	161
5.1. Résumé des contributions	161
5.2. Perspectives de recherche	166
Liste des figures	167
Liste des tableaux	172
Liste des articles	175
Bibliographie	178
Glossaire	193
Annexes	209
6.1. Questionnaire d'évaluation de la plate forme Xplor EveryWhere – Système d'Information d'Intelligence Economique	210
6.2. Le prototype Xplor EveryWhere pour iPhone (iXEW)	215

Introduction générale

« Le futur est pareil à la beauté, tout est dans l'œil de celui qui regarde. » Pablo Picasso

Contexte de travail & Problématique	19
Contributions	20
Organisation du mémoire	23

Contexte de travail & Problématique

Jusqu'à une période récente, la stratégie de l'entreprise pouvait sembler un art facile (Ducieux et al., 2010) : les évolutions étaient lentes, les niveaux de rentabilité souvent stables, les marchés bien définis et protégés à l'intérieur de frontières géographiques, les actionnaires peu exigeants, les technologies en progression régulière mais sans rupture majeure. Puis la tourmente est arrivée. Au niveau mondial, les transformations de l'environnement économique au cours de ces dernières années ont eu des conséquences majeures sur l'activité des entreprises. Dans un contexte de plus en plus dynamique, instable et exigeant, les entreprises doivent concevoir de nouvelles stratégies (Roy, 2010) (Magakian et Payaud, 2010). Un rapide regard en arrière souligne à quel point l'environnement a opéré de nombreuses mutations. Depuis quelques années, les entreprises sont ainsi confrontées à une instabilité croissante des marchés (ruptures technologiques, déréglementation, crise financière 2007/2008 – 2011, crise sociale, printemps arabe, crise japonaise, etc.), un accroissement de l'intensité concurrentielle (apparition de nouvelles formes de concurrence : entreprise 2.0, entreprises en ligne, pays émergents), une course perpétuelle sous toutes ses formes, une internationalisation des marchés, etc. Réunis, ces facteurs ont indiscutablement eu une influence sur les habitudes managériales des entreprises : l'entreprise classique – avec une hiérarchie bien structurée, une centralisation de la prise de décision et une planification formelle – était convenable pour le fonctionnement dans un environnement stable et prévisible. Pourtant dans le contexte actuel, la prise de décision tend vers la décentralisation dans un esprit transversal de « community manager », c'est-à-dire les réseaux des décideurs (dirigeants d'entreprise, responsables politiques, etc.). A titre d'exemple le CAC 40¹ et le sommet de Davos².

¹ Le CAC 40, qui prend son nom du système de "Cotation Assistée en Continu", est le principal indice boursier sur la place de Paris. Créé par la Compagnie des Agents de Change, il est déterminé à partir des cours de 40 actions cotées en continu sur le Premier Marché parmi les 100 sociétés dont les échanges sont les plus abondants sur Euronext Paris. Source Interactive Trader, <http://www.interactive-trader.fr>, consulté le 22/07/2011

² Le Forum économique mondial (World Economic Forum) est une fondation à but non lucratif dont le siège est à Genève. Le Forum est connu pour sa réunion annuelle à Davos, en Suisse, qui réunit des dirigeants d'entreprise et des responsables politiques du monde entier ainsi que des intellectuels et des journalistes, afin de débattre des problèmes les plus urgents de la planète, y compris dans les domaines de la santé et de l'environnement. Source wikipedia http://fr.wikipedia.org/wiki/Forum_%C3%A9conomique_mondial, consulté le 22/07/2011.

Ces décideurs – d'un grand groupe international, d'une startup ou d'une collectivité locale – peuvent être des décideurs qui prennent de nouvelles fonctions, qui recherchent nécessairement de l'information pour agir et construire le changement ou des décideurs qui bougent pour être connus et reconnus, pour accéder à des réseaux d'influence, etc. Un point commun les rassemble : ils veulent être en mesure d'analyser l'environnement avant d'agir. Ils ont besoin d'être informés en temps réel de tout ce qui peut avoir un impact sur leur activité.

Le premier type de décideur demande souvent la cartographie de l'environnement global ou proche de leur entreprise, des synthèses, des détails d'actualités et des analyses stratégiques, afin d'avoir une vision à court terme. Le deuxième type de décideur doit avoir des informations sur les différents acteurs qu'ils vont rencontrer lors de leurs déplacements, pour tenir compte du marché, de l'influence d'un tel type de concurrent, de partenaire, de pays, etc. Ainsi, deux scénarios sont envisagés :

- Le premier consiste à approfondir l'analyse sur des sujets clés, c'est-à-dire de faire des zooms plus importants sur les alliances d'une entreprise, le réseau d'un interlocuteur, le réseau sémantique d'un nouveau terme de recherche, etc. Autrement dit, ces décideurs doivent être armés d'un « couteau suisse » de l'information stratégique qui répond à de multiples besoins informationnels pour avoir plus de détails sur un sujet.
- Le deuxième scénario présente la possibilité de rencontrer des interlocuteurs inconnus, des entreprises émergentes, des nouvelles technologies inexistantes dans le pays d'origine, etc. D'où la panique, puisque les décideurs sont loin de leurs bureaux, sans accès aux outils d'aide à la décision, sans possibilité de demander une analyse stratégique en temps réel pour avoir des informations stratégiques en urgence sur l'acteur en question.

Ma thèse s'inscrit parfaitement dans ce contexte caractérisé par un environnement incertain et imprévisible. Notre objectif consiste à donner des précisions et des réponses à la question : comment armer les entreprises et les décideurs en déplacement pour maîtriser le prévisible et éventuellement être prêt à affronter l'imprévisible ?

Contributions

Face aux défis présentés dans le contexte et la problématique de nos travaux de recherche, rencontrés sur le terrain, l'intelligence économique s'impose, plus que jamais dans la démarche stratégique des entreprises et des collectivités. "Par ses actions de veille et d'analyse, de sécurité économique et d'influence, elle permet aux organisations d'anticiper et de prévoir les menaces et les opportunités afin de dessiner leur stratégie de connaissance de cause" (Blanc, 2010). La construction de cette connaissance est conditionnée par la capacité de l'entreprise à collecter, interpréter, protéger, diffuser et utiliser l'information à des fins stratégiques. C'est cette union entre l'information et la stratégie qui est au centre de la notion d'Intelligence Economique (IE). Ainsi que le décrivent Christian Marcon et Nicolas Moinet (Marcon, 2006) : "Qu'il s'agisse de surveiller son champ concurrentiel, de protéger ses informations stratégiques, de capitaliser ses connaissances ou d'influencer son environnement, la question est bien celle de l'utilisation stratégique de l'information". L'IE intègre ainsi la notion d'intelligence collective parce qu'elle fait travailler en réseau les divers acteurs concernés par chacune des étapes, de la collecte à l'utilisation de l'information, faisant converger leurs activités dans une même vision stratégique. En ce sens, le vocable système d'intelligence économique désigne l'ensemble coordonné des pratiques et des stratégies d'utilisation de l'information utile déployées par une organisation. Cette information est devenue un actif extrêmement riche et complexe sur lequel se fondent les décisions et la stratégie. Si cet actif n'est pas soumis à des règles visant à le maîtriser, alors

l'opacité des prises de risque, quel que soit le secteur d'activité, ne fera que croître et les crises se succéderont.

La mise en place d'un Système d'Intelligence Economique (SIE) représente, pour l'organisation, la réponse indispensable aux défis d'un environnement mondialisé, de plus en plus complexe et imprévisible. Les SIE vont permettre à l'entreprise d'être non seulement réactive mais aussi proactive. Réactive, celle-ci visera à s'adapter aux changements de son environnement avec davantage de célérité que ses concurrentes. Proactive, elle cherchera à adopter une attitude anticipatrice, à mieux appréhender le sens de la dynamique de l'environnement. Elle identifiera les menaces actuelles ou potentielles pour s'en prémunir ainsi que les opportunités à saisir. On trouve plusieurs SIE sur le marché, mais ils présentent des outils partiels puisque ils ne remplissent que quelques fonctions d'IE. A titre d'exemple, nous pouvons citer les outils d'aide au sourcing (AcqURL, Check&Get, Compass et PowerMarks), les outils de collecte d'information (les agents de recherche, les agents de surveillance et les aspirateurs web), les outils d'analyse et mise en forme (les logiciels de classification, les outils de résumé automatique, les logiciels de cartographie) et les progiciels et les solutions spécialisées (Digimind, Iscope, Cybion et Datops).

Les solutions spécialisées se focalisent sur des problématiques particulières telles que la veille brevets. L'offre est assez riche et de nombreux acteurs proposent des solutions : CentreDoc, Delphion, INPI, Qwam, Matheossoftware.

Ma contribution consiste à restituer les études stratégiques, quel que soit le domaine (économique, politique, recherche, social, etc.), via un SIE Mobile Généraliste, qui facilite la navigation dans l'environnement relationnel d'un élément connu (entreprise, centre de recherche, chercheur, inventeur, mot-clé, etc.). Il devient ainsi possible de retrouver, à partir de cet élément, toute ou partie de l'information qui lui est connexe (alliances, concurrents, équipes, collaborations, concepts, émergences, mot associés, etc.), en toute sécurité grâce à la mobilité, aux modèles d'analyse multidimensionnelle, de calcul des indicateurs et de sécurisation.

En effet, la mobilité présente un choix stratégique pour notre SIE "XPlor EveryWhere", dont le but est de permettre aux utilisateurs de nos solutions de veille de continuer à rechercher, surveiller, valider et rediffuser des informations stratégiques au cours de leurs déplacements. Ils n'ont ainsi plus besoin d'être assis face à leur ordinateur pour accéder à des données utiles dans l'instant (préparation d'une réunion, nouvel ordre du jour, information sur un interlocuteur, une technologie, un marché, demande urgente d'une analyse ou d'un focus spécifique). Il leur suffit simplement d'utiliser discrètement, en toute sécurité, leur appareil mobile de type Android, iPhone, iPad, Blackberry ou autre. Grâce à XPlor EveryWhere, ils sont en permanence au cœur de l'information pertinente.

Le modèle d'analyse multidimensionnelle couvre les différentes étapes du cycle de l'analyse de données textuelles structurées, semi-structurées ou non structurées : analyse du besoin, recherche d'information, prétraitements, traitements statistiques, analyse et visualisation des données. Les techniques d'analyse multidimensionnelle sont actuellement bien maîtrisées pour toutes les données quantitatives disponibles sur les systèmes d'information des entreprises ou des administrations à condition que le SGBD s'y prête, que le schéma de la base soit adapté et que les données soient de qualité (homogènes, à jour, complètes, à la bonne granularité, compatibles,...). Il est toujours possible d'extraire les données utiles vers une base construite sur mesure pour la fouille multidimensionnelle, en réalisant, au passage, toutes les corrections nécessaires à leur exploitation. Par contre, pour les données textuelles issues de l'ensemble des sources électroniques (articles scientifiques, brevets, flux RSS, presse, etc.) ce type d'analyse est difficile à mettre en place : les données sources n'ont pas les mêmes formats ou sont même non structurées, elles sont réparties, hétérogènes et la multitude des cas rencontrés ne permet pas de trouver

une structure universelle ou les regrouper. Afin de pouvoir standardiser la fouille multidimensionnelle sur les données textuelles de toute provenance, des premiers travaux de recherche sur cette thématique ont lieu au sein de l'équipe Exploration et Visualisation d'Information (EVI) de l'Institut de Recherche en Informatique de Toulouse (IRIT) (Sosson et al., 2001)(Dousset, 2003)(Ghalamallah, 2009), qui proposent une structure unifiée permettant de stocker toutes les relations entre les items rencontrés dans les documents analysés, en se basant sur l'outil Tétralogie d'IE, développé au sein de l'équipe. Cette technique permet de construire des cubes croisant deux variables quelconques et le temps. Un premier outil XPlor a ainsi été proposé en 2001 pour mettre en ligne ce type de structure en mode client-serveur, afin de réaliser une fouille personnalisée via divers outils de restitution graphique des résultats. Toutes les données textuelles sont alors dans une même structure et bénéficient donc d'outils communs d'investigation interactive. Il est possible avec un tel modèle de pousser plus loin l'analyse en traduisant de façon dynamique l'évolution des relations dans les réseaux des entreprises, les réseaux des inventeurs, les réseaux des chercheurs, les réseaux sémantiques, les alliances, les fusions, les acquisitions, les concurrences, ... par le biais de l'analogie espace/temps présent dans XPlor Everywhere.

Le modèle de calcul d'indicateurs consiste à définir les indicateurs à calculer et à évaluer pour comprendre le besoin informationnel d'un décideur, synthétiser et interpréter l'environnement informationnel du sujet d'analyse et d'évaluer les résultats obtenus par domaine d'analyse, d'un point de vue autant quantitatif que qualitatif.

Le modèle de sécurisation dans notre SIE provient du fait que les informations manipulées sont d'ordre stratégique ayant une valeur assez importante. Une telle sécurité ne doit pas être considérée comme une option supplémentaire qu'offre un SIE pour se distinguer d'un autre. D'autant plus que la fuite de ces informations n'est pas le fait de faiblesses inhérentes aux systèmes informatiques, mais avant tout, c'est une question organisationnelle. Il s'agit donc de sécuriser à la fois des données (IE étant un produit) et des procédés (IE en tant que méthode de traitement). De là est partie notre contribution de proposer une approche de contrôle d'accès aux données et aux traitements dans un SIE. Il s'agit d'un modèle générique pour intégrer le contrôle d'accès dans le cycle de vie d'un tel type de système afin que toutes les tâches dans un processus d'IE soient exécutées uniquement par les utilisateurs autorisés et habilités.

Le SIE proposé fournit à l'utilisateur un espace interactif et collaboratif, facilitant la compréhension et découverte de l'information pertinente. Ce système répond aux différents points suivants :

- Avoir une interprétation (reporting) des connaissances claire et compréhensible.
- Consulter des informations à jour, en accédant au serveur de base de données stratégique en temps réel, lui-même alimenté quotidiennement par des veilleurs.
- Faire remonter des informations « terrain » lors des salons, de visites en clientèle ou à l'issue de réunions.
- Demander des renseignements spécifiques en urgence qui seront mis en ligne par des veilleurs.
- Permettre aux utilisateurs de proposer des explications ou de prendre des décisions,
- Permettre la surveillance et l'anticipation des évolutions.

Grâce à l'évolution de la technologie et de la communication, un tel système permet de gagner en efficacité et en réactivité. A tout moment, il est possible d'accéder à toutes les informations stratégiques utiles au décideur nomade, lui-même pouvant faire remonter très vite l'information « terrain » qui peut éventuellement déclencher d'autres analyses stratégiques. Cependant, l'un des problèmes majeurs est la protection des données sensibles ainsi que les traitements effectués sur le SIE. D'où différents besoins et domaines de recherche « **Les Systèmes d'intelligence économique** », « **La fouille de texte** », « **La mobilité** », « **la lisibilité** » et « **la sécurité** ».

Organisation du mémoire

Ce mémoire est composé de quatre chapitres. L'objectif du premier chapitre est de présenter le contexte général de nos travaux. Au cours de ce chapitre nous présenterons l'IE comme étant à la fois un processus et un produit. En tant que processus, l'IE est l'ensemble des méthodes légales et éthiques qu'une entreprise utilise pour mobiliser l'information qui l'aide à réussir dans un environnement mondialisé. En tant que produit, l'IE est l'ensemble des synthèses sur les activités des concurrents à partir de sources publiques et privées, et son champ d'application est le comportement présent et futur des concurrents, des fournisseurs, des clients, de la technologie, des acquisitions et fusions, des marchés, des produits et services, et de l'environnement des affaires en général (Bichard, 2005).

Dans ce contexte, comment modéliser un SIE, pour le management stratégique de l'information utile à l'entreprise ? La réponse à cette question fait l'objet du deuxième chapitre de ce manuscrit, dans lequel nous allons présenter la genèse des systèmes d'information, les systèmes d'information stratégique, l'émergence des SIE et leurs modèles d'analyse et de conception. Au cours de ce chapitre, notre premier axe de contribution est développé, concernant la proposition d'un système d'information adapté à la démarche d'IE. Il permet de réaliser une fouille personnalisée via divers outils de restitution graphique des résultats servant comme outil d'aide à la décision en se basant sur un modèle d'analyse multidimensionnelle et le modèle de calcul des indicateurs qui feront l'objet du troisième chapitre. Celui-ci sera complété par une présentation de notre second axe de contribution concernant la sécurisation des systèmes d'intelligence économique.

Dans le quatrième chapitre, nous expérimentons les contributions présentées, en évaluant les méthodes selon les critères précédemment retenus. Puis, une enquête est réalisée auprès d'une population ciblée, constituée d'individus informaticiens, statisticiens, documentalistes, de l'IE et de personnes dont la spécificité n'est pas liée directement aux domaines étudiés dans ce mémoire. Les résultats sont comparés à ceux de notre équipe et un bilan est effectué pour cibler les points forts et les points faibles du SIE proposé.

La conclusion générale présente nos apports méthodologiques, scientifiques, ainsi que nos implications managériales. Nous sommes conscients que notre travail comporte des limites ; elles sont également explicitées. Enfin, nous ferons part au lecteur de quelques unes des recherches futures que nous souhaiterions entreprendre.

Chapitre 1.

L'Intelligence Economique ? Un processus de management stratégique de l'information

« Une bataille entre concurrents peut être complètement menée par l'intermédiaire d'annonces, avant qu'un seul dollar de ressources ait été dépensé. » Michael Porter

<i>L'Intelligence Economique ? Un processus de management stratégique de l'information</i>	25
1.1. Introduction	27
1.2. L'intelligence économique	28
1.2.1. Historique d'intelligence économique	28
1.2.2. Elargissement géographique d'intelligence économique	30
➤ Aux États-Unis	30
➤ En France	31
➤ Au Maroc	31
➤ En Asie	32
1.2.3. Définition de l'Intelligence Economique	34
1.3. L'analyse stratégique :	38
1.3.1. L'analyse de l'environnement	38
1.3.2. L'analyse de la concurrence	40
1.3.3. La synthèse SWOT	41
1.4. La veille :	42
1.4.1. La veille peu formalisée et passive :	43
1.4.2. La veille formalisée et proactive	43
1.4.3. La typologie de la veille	43
1.4.4. Processus de la veille	45
➤ Analyse des besoins	46
➤ Recherche et collecte d'informations	47
➤ La collecte des informations	48
➤ Traitement et analyse des informations	48
➤ Mémorisation et diffusion de l'information utile	49
1.5. La protection de l'information	49
1.6. Synthèse	50

1.1. Introduction

L'observation du monde au cours de ces dernières années montre à quel point les entreprises sont confrontées à des changements intervenant dans leur environnement micro-économique. L'émergence du web 2.0, l'augmentation des prix des matières premières, le printemps arabe, la crise japonaise ont contribué à transformer le fonctionnement des marchés. Face à ce nouveau contexte économique, les entreprises doivent anticiper et prévoir les menaces et les opportunités afin de dessiner leur stratégie.

Cette stratégie est un art très ancien (Ducreux et al., 2010), sans doute pratiqué depuis le commencement des temps. « *Le grand théoricien militaire chinois Sun Tzu a décrit les principes de la guerre dès 500 ans av. J.-C... Quelques siècles plus tard, la bataille de Cana est un superbe exemple d'une stratégie militaire bien conçue et bien exécutée. En effet, en 216 av. J.-C., l'Empire romain, qui s'appuie sur des nombreuses légions est en difficulté ; la seconde guerre punique est en cours : Hannibal, le commandant de Carthage est en guerre avec l'Empire romain, et après avoir traversé les Alpes, descend du Nord vers le Sud et traverse les plaines de l'Italie. A Cana, il se heurte aux légions romaines qui sont en nombre beaucoup plus important que ses propres troupes. Il conçoit alors la stratégie qui consiste à prendre en tenaille les légions romaines en détachant sa cavalerie de chaque côté des légions de les attaquer à revers. Cette stratégie couronnée de succès montre qu'un avantage durable ou définitif peut être obtenu face à un adversaire ou à un ennemi sans nécessairement bénéficier de ressources supérieurs ; en effet la stratégie est l'art d'allouer ses ressources, par définition disponibles en quantité limitée, de manière à se créer un avantage durable sur le champ de bataille. Cependant, le mot « stratégie » a été peu utilisé dans la littérature jusqu'au milieu de XIXe siècle. En fait on retrouve le mot « stratégie » dans le dictionnaire d'Oxford en 1810. Puis chez le général prussien Karl Von Clausewitz, qui publie son Traité de la Guerre en 1832 dans lequel il consacre une partie de son ouvrage aux succès militaires de Napoléon* » extrait du grand livre de la stratégie, page 13.

Le mot « stratégie » aurait pu être appliqué aux affaires économiques dès cette période en s'appuyant sur les principes de stratégie militaire (Ducreux et al., 2010). Pourtant, il n'en a rien été. C'est seulement au milieu des années 1960 qu'il a été utilisé dans un contexte économique ; d'abord par Alfred Chandler, professeur au MIT, dans son livre *Strategy and structure* publié en 1963, puis de manière quasi simultanée en 1965 par Igor Ansoff, dans son livre *Corporate Strategy* et par Bruce Henderson, fondateur de Boston Consulting Group. Il a été le premier à articuler les grands principes de la stratégie dans le domaine de l'entreprise ainsi que l'apparition du terme Intelligence Economique (Competitive Intelligence) qui était défini comme « *l'activité de production de connaissance servant le but économique et stratégique d'une organisation, recueilli et produit dans un contexte légal et à partir de sources ouvertes* » en 1967 par Harold Wilensky, dans son ouvrage *Organisational intelligence : knowledge and policy in government and industry*. C'est cette union entre l'information et la stratégie qui est au centre de la notion d'IE qu'on peut définir comme étant le processus de management stratégique de l'information.

La littérature actuelle nous montre que le terme stratégie est utilisé de plus en plus dans la gestion de l'information (Blanc, 2010), ce qui explique le passage de l'ère industrielle à l'ère de savoir et de savoir-faire.

Dans ce premier chapitre, nous souhaitons montrer que l'IE est un processus de management stratégique de l'information pour une entreprise, ou plus largement une organisation dans un but d'adaptation à l'environnement. Et nous espérons le faire de façon à ce que les praticiens et les théoriciens puissent, si possible, facilement accéder à notre vision.

Ce chapitre est organisé en quatre sections. Dans la première section (1.2) nous définissons la démarche d'IE, après une présentation de l'historique dans les différentes parties du monde. Dans la seconde section (1.3), nous traitons les concepts liés à notre démarche d'IE à savoir l'analyse stratégique. Nous

allons ainsi définir l'analyse d'environnement d'une entreprise, son analyse concurrentielle avant de modéliser ces deux analyses sous forme d'une synthèse SWOT. La troisième section (1.4) de ce chapitre a pour objectif de définir la veille, sa typologie, et de détailler son cycle. La quatrième section (1.5) fera l'objet d'une discussion sur l'aspect protection de l'information dans une démarche d'IE. Ces quatre premières sections vont donc exprimer notre vision de la démarche d'IE dans le cadre de nos travaux de recherche. Enfin, nous concluons par une synthèse sur le contexte général de nos travaux.

1.2. L'intelligence économique

1.2.1. Historique de l'intelligence économique

L'origine fondatrice de l'IE connue et pratiquée aujourd'hui est l'intelligence militaire. Le commandement militaire a en effet compris que pour vaincre, il est nécessaire de connaître toutes les composantes du champ de bataille : la mission, le terrain, la météorologie, les plans B de l'ennemi. Ces efforts d'information préalable portent aux USA le nom d' « Intelligence Preparation of the Battlefield (IPB) ». C'est à partir de l'IPB qu'en 1950 est née l'« Intelligence Preparation of the Market Place (IPMP) » (Bonnecarrere et al., 2011).

Le premier élément marquant de cette avancée historique est la création de la première « war room³ » d'analyse stratégique et compétitive, fin 1950, chez General Electric aux USA, par J. Howard Westing de l'université du Wisconsin. Il s'agissait de l'un des tout premiers transferts de l'intelligence militaire et de ses applications (employées lors de la Seconde Guerre mondiale puis la guerre froide). Cette création d'une « war room » civile ouvrait aussi une première intrusion de l'intelligence économique dans le domaine de l'analyse marketing de la concurrence, domaine dont l'importance allait croître de manière extrêmement rapide entre 1950 et 1980.

De cette expérience est née l'une des toutes premières définitions de la discipline de Pinkerton (Prescott, 1999) selon laquelle l'intelligence économique était :

« le savoir dérivé de l'analyse et de l'interprétation de certaines informations économiques nécessaires à la prise de décisions d'organisation, stratégiques ou tactiques utiles à toutes les fonctions de l'entreprise. Elle impliquait que cette nouvelle fonction soit dotée d'un système organisé de collecte, de production et de diffusion de « l'intelligence » aux unités opérationnelles de l'entreprise ».

En 1968, le Dr William T. Kelley de la Wharton School publie le premier livre utilisant les termes « Marketing Intelligence » comme titre et les deux premiers articles sur le sujet publiés dans le Journal of Marketing étaient également signés par T. Kelley.

Le premier mémoire (Ph. D) sur le sujet revenait à G. Albaum : « A new approach to the information function in Marketing » (1962).

En 1980, à Harvard, l'un des pères de l'analyse de la concurrence, à savoir Michael Porter apporte sa contribution avec son ouvrage « Competitive Strategy » traduit en français sous le titre : Choix stratégique et concurrence, technique d'analyse des secteurs et de la concurrence dans l'industrie en

³ Un lieu où des personnes, bénéficiant d'un soutien matériel et logiciel, se basent sur une méthodologie commune afin de définir puis de mettre en oeuvre une stratégie. Source <http://www.oodoc.com/37795-war-room-salle-operation-guerre.php> , consulté le 25 juillet 2011

1990. Dans cette ouvrage, Porter soutient notamment « la nécessité d'un système de renseignement sur la concurrence » et propose une grille d'analyse des « signaux du marché » où il indique notamment un des principes de l'utilisation de l'information pour agir sur la concurrence : « une bataille entre concurrents peut être complètement menée par l'intermédiaire d'annonces, avant qu'un seul dollar de ressources ait été dépensé ».

Ces premiers pas furent suivis jusqu'à la fin des années 1980 par une série impressionnante d'avancées théoriques ou de bonnes pratiques, due à Leonard Fuld (1995) sur les sources et les bases de données, Ben Gilad (1988) sur la structure d'un système d'intelligence économique et bien sûr John E. Prescott (université de Pittsburg), numéro un de la recherche en intelligence économique dans le monde académique, avec Michael Porter (Harvard). Cela a permis l'apparition de nouveaux concepts et méthodologies aux Etats-Unis tels que TBIS (Transaction Based Information System) de R. Glazer en 1991, EIS (Executive Information Systems) de F. Crockett (1992), les matrices SWOT et MIMS d'analyse de la concurrence développés par le Boston Consulting Group, l'EEI (Element of Essential Information) en amont du cycle classique de production de « l'intelligence » avec ses deux éléments clés : « need to know » opposé au « nice to know », etc.

L'IE à l'américaine ou « competitive intelligence », s'est vraiment étendue en tant que discipline sous l'impulsion du SCIP (Society of Competitive Intelligence Professionals) créé en 1986 à Washington DC. Cette association réunissait alors des universitaires, des praticiens et des consultants en « competitive intelligence ». Trois hommes issus du SCIP sont à l'origine d'un référentiel d'enseignement en intelligence économique : Leonard Fuld, Ben Gilad et Jan Herring. Ensemble, ils fondent en 1996 le premier établissement de formation entièrement dédié à l'IE : « le Fuld-Gilad-Herring Academy of «competitive intelligence» ». Cet établissement a organisé sa formation autour de dix points :

1. Améliorer la prise de décision et la gestion du temps (Improve planning and decision-making) : le site de l'établissement précise que la CI permet de prendre des décisions rationnelles sous la pression (making reasoned, rational decisions while under pressure) et que leurs étudiants sont placés en situation réelle (real-life situations) où ils doivent appliquer des outils et des techniques et rendre des décisions précises au bon moment.
2. Prévenir et anticiper les menaces concurrentielles (Enhance early warning and forecasting competitive threats)
3. Gérer et réduire les risques (Manage and reduce risk)
4. Connaître les limites légales et éthiques (Know ethical and legal limits)
5. Rendre la connaissance rentable (Make knowledge profitable) : le site précise ici qu'il ne faut pas confondre la «competitive intelligence» (CI) avec le knowledge management (KM). Ainsi, alors que la CI a toujours un but précis, celui d'accroître le profit et les bénéfices, le KM cherche à exploiter toutes les informations que contient une entreprise sans que son but ne soit pas établi clairement.
6. Éviter la surcharge ou le surplus d'information (Avoid information overload)
7. Éviter les angles morts ou les œillères (Avoid blindspots) : il s'agit d'être capable de repérer les changements dans l'environnement stratégique afin de changer sa manière de voir et de s'adapter de manière proactive.

8. Développer des compétences de management du renseignement critique (Develop critical intelligence management skills) : le professionnel en CI doit être en mesure de conduire les cinq étapes du cycle du renseignement :
 - Planifier et définir les objectifs de l'organisation, en particulier en définissant des besoins en renseignements clés (Key Intelligence Topics) ;
 - Apprendre à utiliser les sources secondaires pour ne pas crouler sous le volume d'information ;
 - Rassembler les informations principales en respectant les directives légales et éthiques ;
 - Appliquer les modèles d'analyse appropriés à la situation ;
 - Apprendre à fournir l'information à la direction de manière efficace.
9. Empêcher les « fuites » et conduire le contre espionnage (Prevent “leaks” and conduct counter-intelligence) : c'est-à-dire protéger les informations d'une organisation.
10. Avoir une vision stratégique (Gain strategic perspective).

1.2.2. Elargissement géographique de l'intelligence économique

Essentiellement appliquée aux Etats-Unis jusqu'à la fin des années 1980, l'IE s'internationalise, sous des vocables divers, et la créativité produit ses effets dans des nombreux pays développés. Comme c'est souvent le cas, « les expériences avancées » sont fréquemment provoquées par les circonstances et par des spécificités propres à chaque continent, région ou pays.

➤ Aux États-Unis

Aux États-Unis, l'intelligence économique a très vite été perçue comme un instrument essentiel pour se faire une place dans la compétition économique mondiale (Chauna et al., 2011b). Dès la fin de la guerre froide, la stratégie des États-Unis a pris en compte l'importance de la puissance économique. C'est ainsi qu'en 1993, le président Clinton oriente le travail de la CIA vers le renseignement économique. Dans les faits, cela se traduit par la création du « National Economic Council » (NEC) cette même année. Cet organisme s'est donné quatre objectifs principaux : coordonner les politiques économiques intérieures avec les grands défis de l'économie mondiale, coordonner les conseils économiques du président, s'assurer que les programmes et les décisions politiques qui concernent l'économie concordent avec les buts affichés par le président et mettre en œuvre son programme de politique économique. Le directeur de ce conseil est nommé par le président américain et est composé de nombreux départements et agences dont les prises de décision ont un impact direct sur l'économie du pays. Chaque secteur d'activité a ses spécialistes et est représenté (l'agriculture, l'énergie, les marchés financiers, la santé, la sécurité sociale et le travail). Concrètement, la structure vient en aide aux entreprises américaines qui sont positionnées sur les marchés prometteurs ou sensibles. A l'heure actuelle, il s'agit surtout de domaine des nouvelles technologies, de l'énergie, de l'armement ou de la défense.

Ces quelques points donnent une idée de ce que l'on entend par IE ou « competitive intelligence » outre-Atlantique.

➤ En France

En France, l'IE apparaît plus tardivement qu'aux Etats-Unis. En premier lieu, c'est la surveillance de l'environnement et la veille qui sont privilégiées. Aujourd'hui chacun s'accorde à considérer la surveillance et la veille comme des processus de la démarche d'IE. Après avoir analysé les grandes articulations dans l'évolution du concept, on peut résumer les grandes dates de l'IE en France comme suit :

- 1994 : rapport du commissariat au Plan sur l'IE et la stratégie des entreprises, dit « rapport Henri Martre » ;
- 1995 : suite au rapport Martre, création d'un comité pour la compétitivité et la sécurité économique ;
- Juin 2003 : rapport de la mission parlementaire conduite par Bernard Carayon ;
- Décembre 2003 : Alain Juillet est nommé haut responsable à l'IE par Jacques Chirac, président de la République ;
- Juin 2008 : le livre blanc de la défense et de la sécurité nationale propose la création de la Direction Centrale des Renseignements Intérieurs (DCRI, fusion de la DST et des Renseignements Généraux) qui a pour principales missions, la protection du patrimoine matériel et immatériel du secteur économique français (entreprises, laboratoires, centres d'études et de recherche), les actions de contre-ingérence et le maintien d'un environnement de sécurité pour les entreprises.
- 2009 : lancement opérationnel de la DCRI et mise en place des DDRI (Directions Départementales des Renseignements Intérieurs) ;
- 17 septembre 2009 : parution du décret d'institution d'un délégué interministériel à l'IE placé auprès du secrétaire général du ministère chargé de l'Economie. Il a pour mission d'élaborer et de proposer la politique publique d'IE, d'en assurer la mise en œuvre et d'en évaluer l'efficacité. Il est rattaché à un comité directeur de l'IE, placé auprès de la présidence de la République.

➤ Au Maroc

C'est dans l'optique de compétitivité nationale et internationale que l'intelligence économique a fait son apparition dans les années 90 au Maroc. En effet les grandes entreprises telles que l'ONA (Omnium Nord Africain), l'ONE (Office Nationale d'Electricité) ou encore l'OCP (Office Chérifien de Phosphate), se sont dotées, à cette époque, de cellules de veille pour protéger leur développement et leurs investissements.

En 2003 le Roi Mohamed VI, dans son discours du Trône du 30 juillet, annonce la création de l'Institut Royal des Etudes Stratégiques (IRES) :

"... Notre pays vit une transition globale qui nécessite le renforcement de ses capacités d'analyse, d'adaptation et d'anticipation. Aussi, avons-nous décidé de créer un institut royal des études stratégiques pour remplir cette mission essentielle, afin d'être en interaction permanente avec les changements et de maîtriser et d'agir sur les mutations profondes qui s'opèrent aux niveaux interne et externe ...".

L'IRES a été créée en novembre 2007. Elle « assure également une fonction de veille, notamment en recueillant et en traitant les informations et les données rendues publiques dans les domaines politiques, économiques et sociaux sur le plan national, mais aussi au niveau international. » (IRES, 2010).

Parmi les attributions de cet institut figurent également la mise en place et l'actualisation d'un système d'information prospective (Mouline, 2010) ayant pour objet de recueillir et de traiter les données en relation avec les questions stratégiques que ce soit au niveau national ou international.

D'autres initiatives plus générales ont été menées par des organismes ou des associations. Citons l'exemple de l'association R&D Maroc qui organise régulièrement des séminaires sur l'innovation, la veille et la compétitivité et l'Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes (ENSIAS) en collaboration avec l'Institut de Recherche en Informatique de Toulouse (IRIT) qui organise des journées scientifiques sur l'IE. Ces différentes initiatives ont pu sensibiliser les différents acteurs, privés et publics, aux enjeux liés à la mise en place d'une structure de veille, d'adopter une démarche d'IE pour la gouvernance stratégique de l'entreprise.

En outre la plupart des observatoires nationaux, tourisme, emploi, sécheresse, observatoire des TIC de l'ANRT... ont pour principale mission la veille stratégique.

➤ En Asie

A l'autre bout du monde, le Japon a été souvent pris comme modèle d'intelligence économique. La réussite économique du Japon est fondamentalement associée à son comportement de traqueur d'information (Chauna et al., 2011a). L'attention accordée par les japonais à l'information est une caractéristique culturelle, un état d'esprit qui remonte à des temps historiques puisqu'en 1868 figurait dans la constitution japonaise la phrase suivante : « *Nous irons chercher la connaissance dans le monde entier afin de renforcer les fondements du pouvoir impérial* » (Rostaing, 1993).

Les Chinois n'utilisent pas le terme d'IE mais ils en font (Chounac, 2010). Il n'y a pas non plus de dispositif centralisé d'IE mais une multitude d'acteurs qui participent au système de renseignement économique chinois : Le Ministère des Sciences et Technologies, en charge de la veille technologique ; Le Ministère du Commerce, chargé d'accompagner les entreprises chinoises dans leurs négociations commerciales ; le Bureau de recherche des affaires du Conseil d'Etat, chargé d'acquérir des renseignements stratégiques au profit du Premier Ministre Chinois ; le Ministère de la sécurité de l'Etat qui regroupe l'ensemble des services secrets chinois chargé de l'espionnage économique ; le Ministère des Affaires étrangères dont les diplomates recueillent des informations dans le domaine technologique ; on peut citer également la Commission des Sciences et Technologies de l'Industrie de défense.

Il faut savoir que toutes ces institutions se sont mises au renseignement économique offensif à partir véritablement d'une directive de 1995 du Comité central du Parti communiste chinois : « Décision sur l'accélération du développement scientifique et technique ». Leurs cibles sont évidemment occidentales, japonaises ou encore sud-coréennes, notamment celles qui développent des produits de haute technologie.

Mais la Chine ne s'attaque pas seulement à ces entreprises. Elle sait très bien que d'autres pays émergents sont ses principaux compétiteurs. Elle s'intéresse donc de près à ces entreprises, notamment à celles de l'Asie du Sud dans les pays qu'on nomme Tigre et Dragons (Malaisie, Thaïlande, Vietnam...).

La carte géostratégique suivante nous montre l'évolution de la recherche en IE ces deux dernières années. On remarque très bien la puissance de la Chine en domaine de recherche sur l'IE, ainsi qu'une absence de la Russie et de la majorité des pays africains et arabes.

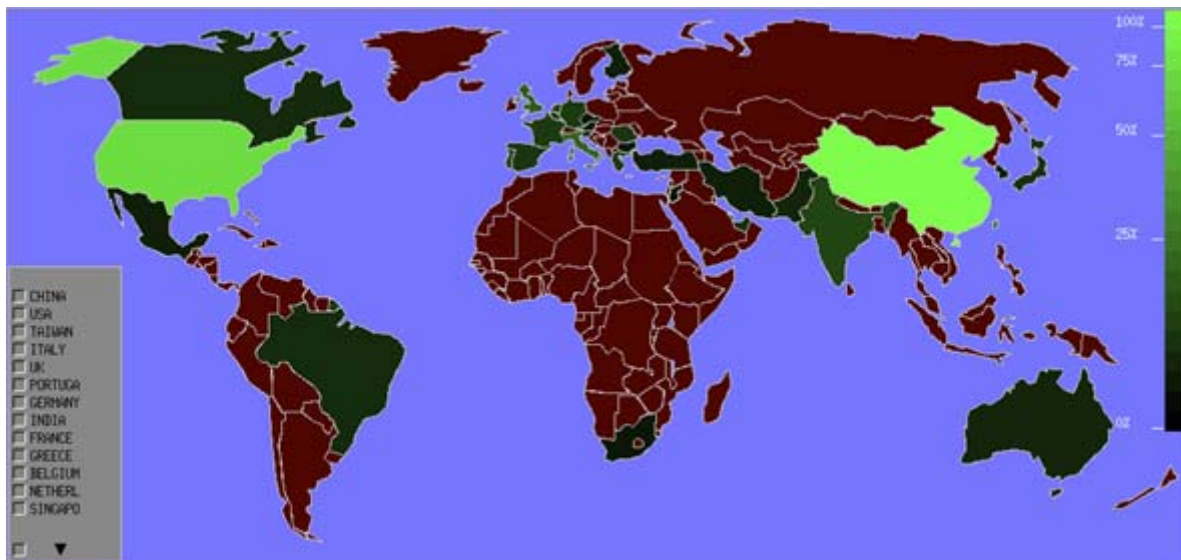


Figure 1. La carte géostratégique de la recherche en intelligence économique

En termes de collaborations, on remarque qu'il y a deux clusters :

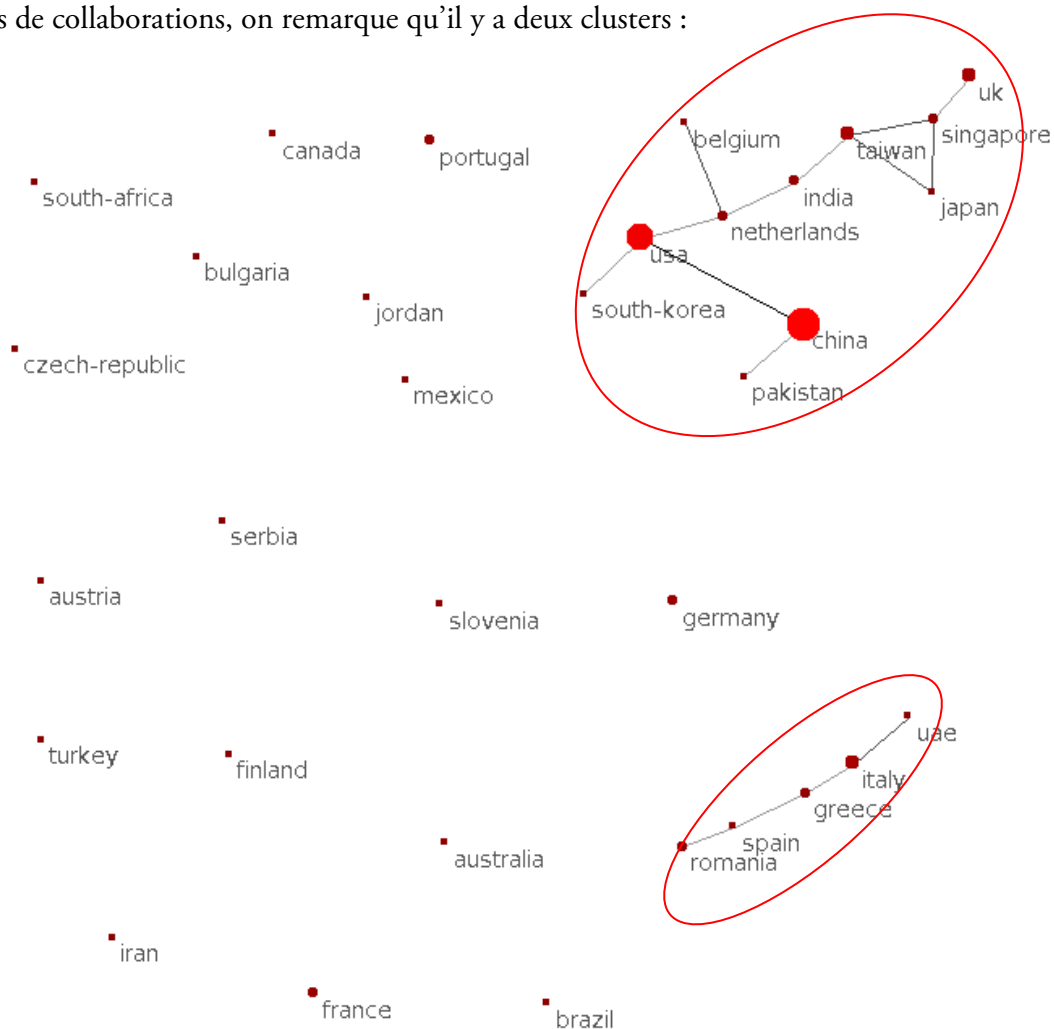


Figure 2. Les collaborations internationales de la recherche en intelligence économique

- Cluster 1 : Etats-Unis, Chine, Pakistan, Corée du Sud, Pays-Bas, Belgique, Inde, Japon, Singapour, Taiwan et Royaume-Uni.
- Cluster 2 : Espagne, Roumanie, Grèce, Italie et Emirats Arabes Unis

Les autres pays comme la France, Canada et l'Allemagne ne collaborent pas en termes de recherche sur l'intelligence économique.

Après ce petit tour mondial, qui nous a permis de voir un peu l'historique d'IE, nous proposons une approche d'étude de l'IE selon sa définition, ses fonctions, son processus et ses produits.

1.2.3. Définition de l'Intelligence Economique

L'émergence du concept d'IE et sa compréhension font appel à de multiples champs scientifiques (figure 3). Il se rattache aux sciences de la sociologie des organisations, aux sciences politiques, aux sciences de l'information, aux technologies de l'information et à la communication et aux sciences de gestion de l'information et des connaissances. En fait on peut considérer que sa genèse vient de différentes disciplines de recherche extrêmement vastes et prometteuses.

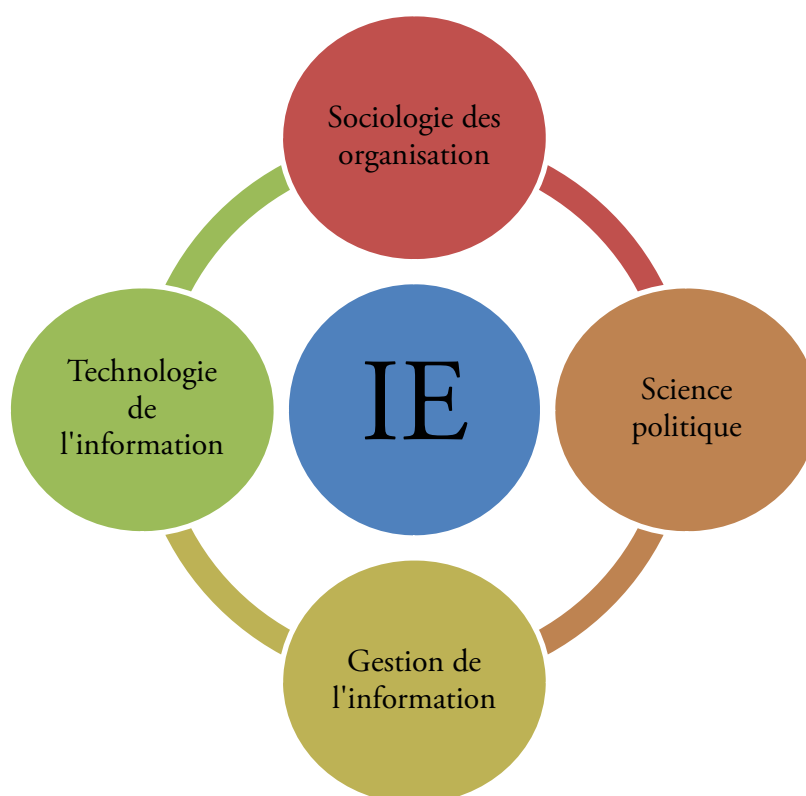


Figure 3. Emergence de l'IE

L'univers de l'IE étant très vaste. Cela rend la notion d'IE très riche en concept et en domaine d'application. Nous retenons pour nos travaux de recherche quatre définitions de l'IE :

La première définition de l'IE moderne date de 1967 par Harold Wilensky, dans son ouvrage intitulé : « *Organisational intelligence : knowledge and policy in government and industry* » (Wilensky, 1967). Il définit l'IE comme étant :

l'activité de production de connaissance servant le but économique et stratégique d'une organisation, recueilli et produit dans un contexte légal et à partir de sources ouvertes.

La deuxième définition de l'association professionnelle d'IE (SCIP)⁴:

*Competitive Intelligence is a systematic and ethical program for gathering, analyzing, and managing external information that can affect your company's plans, decisions, and operations. Put another way, CI is the process of enhancing marketplace competitiveness through a greater -- yet unequivocally ethical -- understanding of a firm's competitors and the competitive environment. Specifically, it is the legal collection and analysis of information regarding the capabilities, vulnerabilities, and intentions of business competitors, conducted by using information databases and other "open sources" and through ethical inquiry. Effective CI is a continuous process involving the legal and ethical collection of information, analysis that doesn't avoid unwelcome conclusions, and controlled dissemination of actionable intelligence to decision makers.*⁵

La troisième définition du rapport du Commissariat général du plan « *Intelligence économique et stratégie des entreprises* », document de référence dans la matière. Il a été conçu comme un ensemble de repères sur lequel repose le socle théorique de l'IE en France. Selon le rapport (Martre, 1994):

L'intelligence économique peut être définie comme l'ensemble des actions coordonnées de recherche, de traitement et de distribution, en vue de son exploitation, de l'information utile aux acteurs économiques. Ces diverses actions sont menées légalement avec toutes les garanties de protection nécessaires à la préservation du patrimoine de l'entreprise, dans les meilleures conditions de délais et de coûts. L'information utile est celle dont ont besoin les différents niveaux de décision de l'entreprise ou de la collectivité, pour élaborer et mettre en œuvre de façon cohérente la stratégie et les tactiques nécessaires à l'atteinte des objectifs définis par l'entreprise dans le but d'améliorer sa position dans son environnement concurrentiel. Ces actions, au sein de l'entreprise, s'ordonnent autour d'un cycle ininterrompu, générateur d'une vision partagée des objectifs de l'entreprise.

Les définitions de SCIP et du rapport Martre ne sont pas donc très éloignée de celle de Wilensky. Elles y ajoutent la notion de coordination (Colletis, 2007) ou ce qu'on peut nommer **l'intelligence collective**. Sans oublier que l'IE est à la fois un processus et un produit (Bichard, 2005).

En tant que processus, l'IE est l'ensemble des méthodes légales et éthiques qu'une entreprise utilise pour mobiliser l'information qui l'aide à réussir dans un environnement mondialisé. En tant que produit, l'IE est l'ensemble des informations sur les activités des concurrents à partir de sources publiques et privées, et son champ d'application et le comportement présent et futur des concurrents, des fournisseurs, des clients, de la technologie, des acquisitions et fusions, des marchés, des produits et services, et de l'environnement des affaires en général.

Avec l'avancement des travaux de recherche, une nouvelle génération de définitions à émergé. De définitions centrées sur la description des processus et des techniques d'IE, elles sont passées à des

⁴ SCIP est une association professionnelle à but non lucratif dans le domaine d'IE Créée en 1986, SCIP est basée aux Etats-Unis. Elle comprend près de 6000 adhérents, rassemblés par « chapitres » dans 50 pays. SCIP France a été créée en 1992 par Robert Guillaumot, Yves-Michel Marti, Bruno Martinet et Jean-Pierre Bernat (Elf). Elle est le premier réseau français des professionnels de l'IE et de Gestion des connaissances.

⁵ <http://competitive-intelligence.mirum.net/business-intelligence/definition-competitive-intelligence.html>, consulté le 18 janv. 11

définitions incluant des objectifs stratégiques puis, depuis peu, à des définitions incluant les fonctions de gestion de connaissance, d'apprentissage collectif, de coopération (Salles et al., 2000), de management stratégique de l'information et des système d'information (Xiaobo et lian, 2010) (Olusoji et Amos, 2010) (Frion et Moinet, 2010) (Guoliang et al., 2009) (Weidong, 2008) (Sauter, 2008) (Ginsburg et al., 2009) (Li et Dong, 2010) (Li et al., 2007) (Chen et al., 2008) (Chen et al., 2009) (Radum, 2006) (liu et Xia, 2009) (Tan et al., 2004) (Li et al., 2007). Cette vision valorise l'intelligence collective (Zara, 2008) en passant d'une entreprise hiérarchisée, isolée face à son environnement, à l'entreprise 2.0 (Yvan, 2010). Nous retrouvons à partir de ces différentes définitions, un ensemble de fonctions qui sont associées à la notion d'IE, nous le résumons dans le tableau suivant :

Fonctions	Auteurs
Gestion des connaissances	(Salles et al., 2000) (Guilhon, 2003) (Levet, 2001) ; (Jackobiak, 2004) (Moinet, 2008) (Moinet et al., 2009a)
Intelligence Collective	(Salles et al., 2000) (Colletis, 2007) (Zara, 2008), (Moinet et Jdey, 2009) (Casselman et Samson, 2011)
Management stratégique de l'information	(Xiaobo et lian, 2010) (Olusoji et Amos, 2010) (Frion et Moinet, 2009) (Guoliang et al. 2009) (Weidong, 2008) (Ginsburg et al., 2009) (Li et Dong, 2010) (Li et al., 2007) (Chen et al., 2008) (Chen et al., 2009)(Radum, 2006) (liu et Xia, 2009) (Tan et al., 2009) (Li et al., 2007)
Gouvernance des systèmes d'information	(Chauna et al., 2011a)
Entreprise 2.0	(Yvan, 2010)

Tableau 1. Les Fonctions associées à la notion d'IE

Finalement, suite à ce bref aperçu de principales définitions, nous définissons l'IE de la manière suivante :

l'IE est un processus, un produit et un service. En tant que processus l'IE est le workflow du management stratégique de l'information, pour une prise de décision collaborative. En tant que produit, l'IE est un système d'information stratégique et organisationnel. En tant que service, l'IE est une cartographie de l'environnement de l'entreprise.

Selon cette définition, le processus d'IE est alors une coordination des processus d'analyse stratégique et de la veille basée sur les systèmes d'information. On retrouve dans cette présentation schématique, les quatre étapes essentielles du cycle du renseignement à savoir :

- La compréhension du besoin,
- La recherche et la collecte d'information,
- Le traitement de l'information,
- La diffusion de l'information.

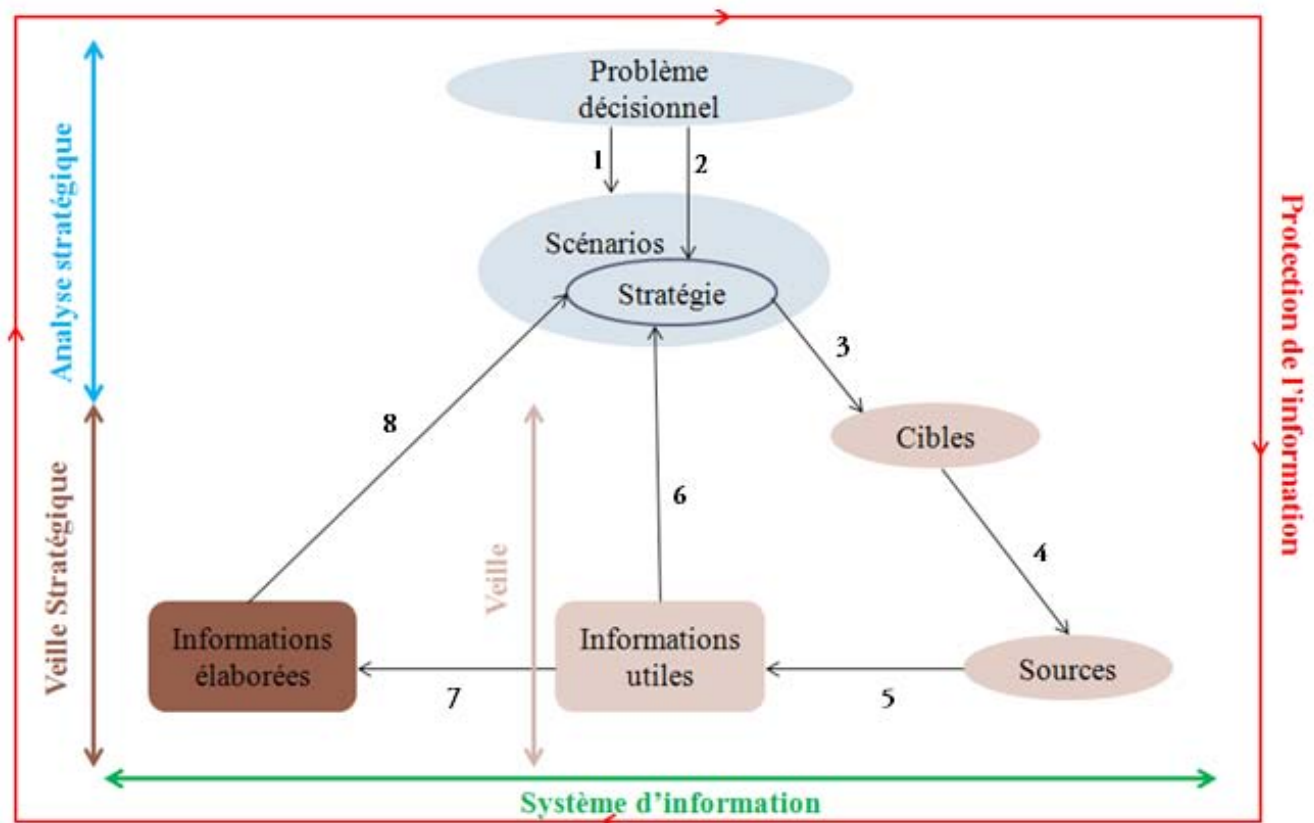
Le passage de ces étapes au niveau du processus d'IE leur a fait subir des transformations au niveau des termes qui les décrivent ainsi qu'une adaptation aux diverses situations :

- La compréhension du besoin est nommée par plusieurs expressions synonymes selon les auteurs à savoir : Ciblage (Lesca, 1994), Expression du besoin, Orientation générale (De

vasconcelos, 1999), Identification du problème décisionnel et sa traduction en besoin informationnel (Bouaka, 2004).

- Pour la recherche et la collecte d'information, nous retrouvons : traquer (Lesca, 2004), trouver (Fuld, 1995), acquérir (Dupré, 1997), recueillir et rechercher (Jacobiak, 2001).
- L'exploitation devient traitement puis analyse (Martinet et Marti, 1995).
- Seule l'étape de diffusion semble faire l'unanimité bien qu'elle soit parfois ajoutée au milieu du processus.

Dans la littérature relative à l'IE, tous les auteurs s'accordent sur l'existence d'une étape primordiale consacrée à la compréhension du besoin. Le but de cette première phase consiste à définir un problème décisionnel par un ensemble de scénarios possibles et de proposer une stratégie d'entreprise. Elle permet de traduire les orientations stratégiques en axes de développement. Ces axes sont traduits en cibles prioritaires en déterminant les questions opérationnelles sur lesquelles il serait utile de s'informer. Ces cibles peuvent être les signaux forts, les nouveaux centres d'intérêts, les nouveaux produits, les concurrents, les nouveaux acteurs ou les nouveaux procédés.



1. Elaborer 2. Choisir 3. Déterminer & Hiérarchiser 4. Identifier & Sélectionner 5. Collecter & Evaluer 6. Valider & Diffuser 7. Analyser & Interpréter 8. Valider & Diffuser

Figure 4. Processus de l'Intelligence Economique

Le besoin informationnel déterminé, suit la phase de la veille qui permet d'analyser et de surveiller l'environnement informationnel, les cibles identifiées par la stratégie retenue.

En fonction des cibles définies dans la phase précédente, il s'agit de rechercher les informations les plus pertinentes, à l'intérieur et à l'extérieur de l'entreprise. La première étape consiste à identifier les sources d'information qui sont susceptibles de contenir ces informations. A partir de ces sources les informations utiles sont extraites et évaluées. Il existe deux types d'informations qui sont nécessaires à

l'élaboration de la stratégie : l'information utile qui peut être directement utilisée et l'information élaborée qui résulte de l'analyse et du traitement des informations utiles. L'étape de traitement et d'analyse est soutenue par les méthodes et les outils de fouille de données qui permettent de révéler les informations cachées, souvent statistiques, à partir des informations utiles.

La dernière phase du cycle consiste à restituer les résultats jugés pertinents aux utilisateurs potentiels qui vont établir et définir les actions stratégiques en se basant sur le degré de fiabilité des résultats obtenus. Cette phase peut engendrer une redéfinition ou un réajustement de la cible, le cycle devient alors itératif.

La protection de l'information est un processus qui s'instaure en parallèle du cycle de l'IE et s'applique à toutes ces étapes.

L'ensemble de ces étapes peuvent être soutenues par l'utilisation des systèmes d'information.

Après avoir défini l'IE et son processus, nous allons détailler les différents axes liés à la démarche d'IE : l'analyse stratégique, la veille, les sources d'information et la protection d'information.

1.3. L'analyse stratégique :

Selon le grand livre de la stratégie (Ducieux et al., 2010), l'analyse stratégique consiste à évaluer les choix ouverts à l'entreprise. Ceux-ci doivent prendre en compte tous les aspects de l'environnement où elle évolue, le potentiel de son portefeuille d'activités ainsi que l'analyse fine de ses bases d'expérience (la segmentation).

La compréhension de l'environnement et plus particulièrement de l'environnement concurrentiel, ainsi qu'une bonne évaluation des propres forces et faiblesses sont un préalable à toute décision stratégique, permettant de s'assurer de l'adéquation entre l'environnement externe et les ressources internes :

1.3.1. L'analyse de l'environnement

Comprendre l'environnement passe par plusieurs étapes : connaître l'environnement et les pays où opère l'entreprise, comprendre l'évolution du secteur et de ses caractéristiques, et notamment sa croissance ; construire une représentation des forces concurrentielles et anticiper la dynamique, faire une synthèse de l'ensemble de ces informations.

Certaines entreprises qui réalisent l'importance de cette fonction mettent en place des systèmes de veille permanents couvrant l'ensemble des éléments qui composent le « macro-environnement » de l'entreprise : démographie, technologie, législation, et le micro-environnement : demande, offre, publics concernés... Le schéma de la figure 5 visualise ces différents éléments :

L'analyse environnementale PEST consiste à décrire et analyser les quatre composantes majeures de l'environnement (politique, économique, socioculturel et technologique), et à évaluer leur dynamique et leur influence sur l'entreprise et sa stratégie.

L'environnement politique : Il s'agit de comprendre le cadre légal, juridique, fiscal et social dans lequel évolue l'entreprise : droit du travail, fiscalité, réglementation du commerce extérieur, protection et environnement. Deux difficultés majeures surviennent dans l'appréciation de l'environnement politique :

- La première est liée à l'instabilité, qu'elle soit purement politique comme c'est le cas de nombreux pays en voie de développement, comme les pays arabes, ou qu'elle soit réglementaire

comme de nombreux pays développés qui modifient fréquemment la réglementation et son interprétation ;

- La seconde difficulté résulte de la complexité croissante des cadres juridiques et fiscaux, liée à la superposition de réglementations et à l'interdépendance croissante avec des réglementations supranationales.

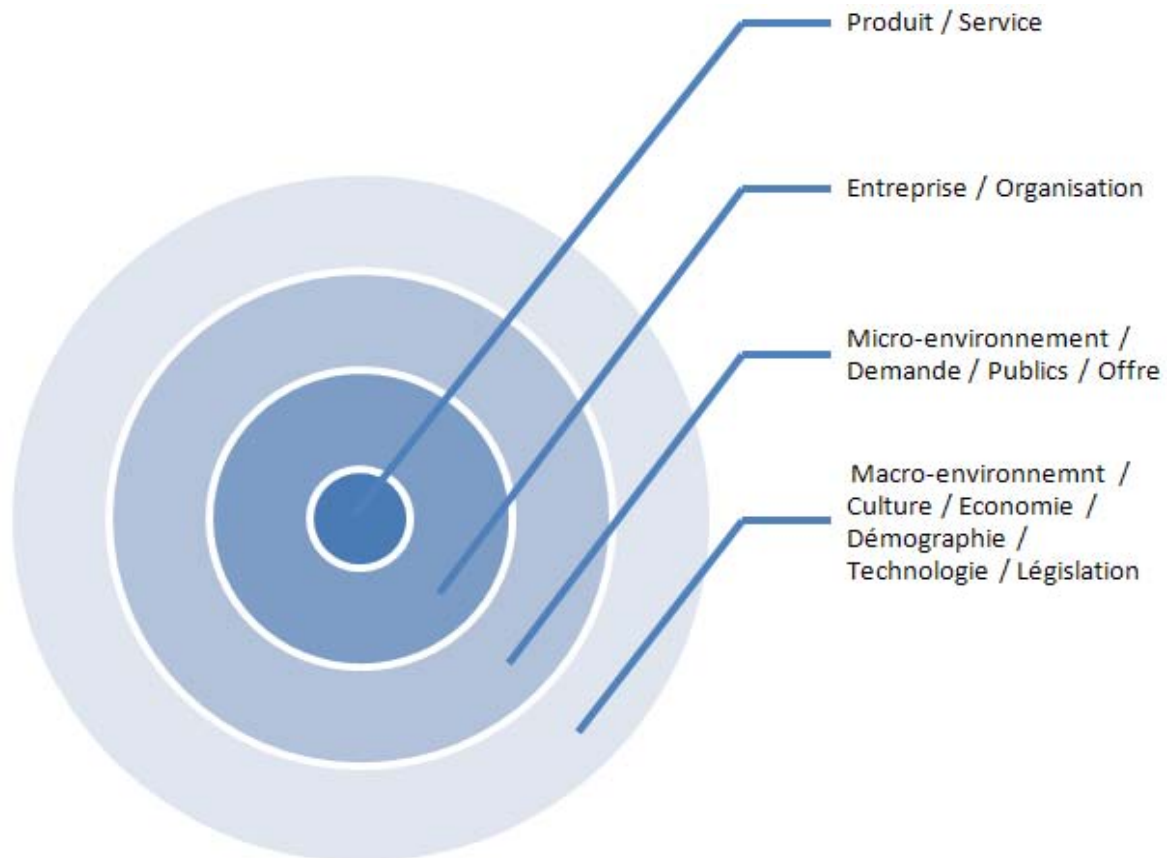


Figure 5. Les environnements de l'entreprise (Magakian et Payaud, 2010)

L'environnement économique : L'objectif est d'identifier et de comprendre les principaux paramètres intervenant au niveau macroéconomique : croissance, cycle économique, taux d'intérêt, main-d'œuvre, infrastructures. Tous ces éléments peuvent être quantifiés et leur impact pris en compte pour les choix stratégiques.

L'environnement socioculturel : Cet aspect concerne les éléments liés à la démographie, à l'éducation, au mode de vie, à la distribution des revenus, à la disponibilité et la mobilité des compétences.

L'environnement technologique : Il prend en compte le niveau de la qualité de la recherche et développement, les nouvelles découvertes, les nouveaux développements, les transferts technologiques et le rythme d'obsolescence, le niveau d'investissement en technologie (public et privé), la protection de la propriété intellectuelle... Dans la plupart des pays se sont créées des poches technologiques dans lesquelles se concentrent les activités liées à une technologie donnée : Silicon Valley aux Etats-Unis pour l'informatique, Toulouse en France pour l'aviation, Chartres/Orléans pour la cosmétologie, ... Ces poches technologiques ont un effet d'attraction très important car elles constituent un environnement favorable.

1.3.2. L'analyse de la concurrence

Une des approches les plus complètes de l'analyse concurrentielle est celle proposée par Michael Porter. elle consiste à considérer que toute entreprise cherche à obtenir un avantage concurrentiel et que, pour y parvenir, elle doit prendre en compte et maîtriser cinq forces :

- La rivalité existante entre les concurrents qui, elle-même soumise à d'autres forces, définit le niveau de l'intensité concurrentielle ;
- La menace d'entrants potentiels : quand l'entrée potentielle dans un secteur est facile, les concurrents existants sont peu protégés des nouveaux entrants ;
- Le rapport de force avec les clients : quand le nombre de clients est faible, et donc leur pouvoir individuel fort, l'attractivité du secteur est réduite ;
- Le rapport de force avec les fournisseurs : si le nombre de fournisseurs est réduit, leur pouvoir sur les entreprises est fort, diminuant ainsi l'attrait du secteur ;
- La menace des produits de substitution : lorsque les menaces de substitution sont fortes, le secteur est peu attractif.

L'attrait intrinsèque d'un secteur d'activité résulte du jeu de ces forces (figure 6).

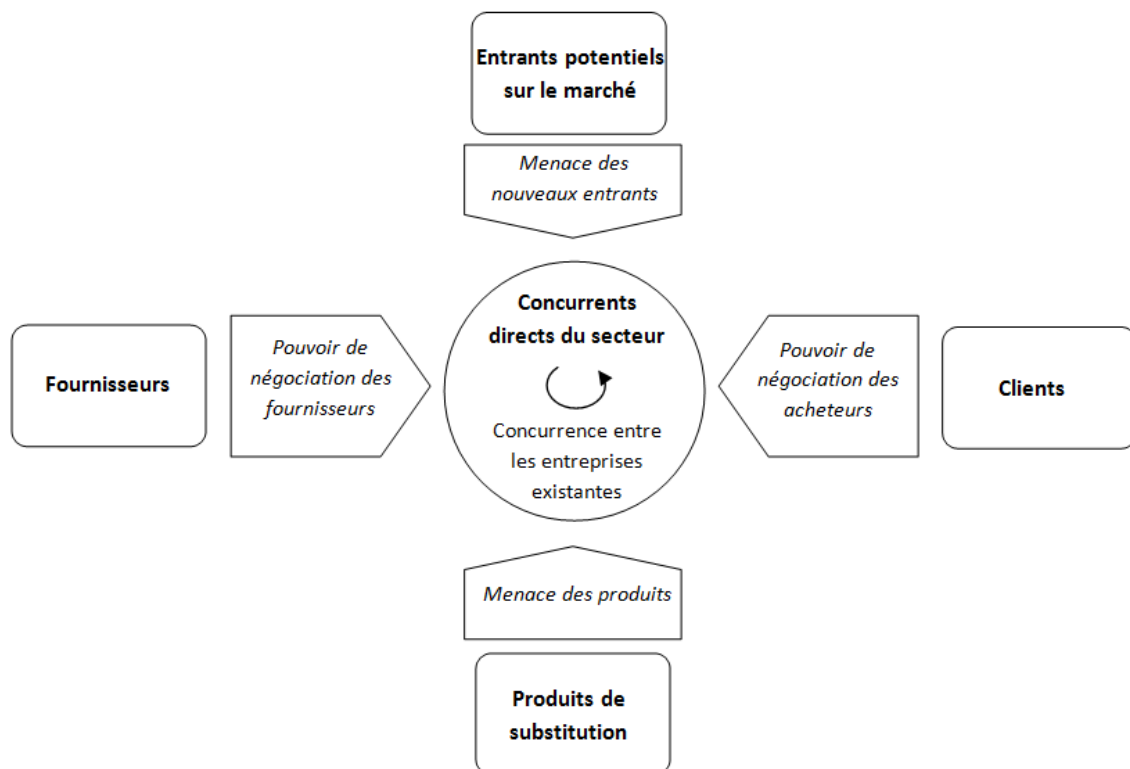


Figure 6. Schéma de Porter - l'analyse de l'environnement (Placet et Rarivoson, 2011)

1. Ces forces se combinent pour déterminer l'intensité de la lutte concurrentielle du secteur. Qui sont les concurrents ? Qui propose des biens et services ou des produits proches ? Quelle est leur capacité financière ? Quelle est la nature de leur avantage ? Quelle est la pérennité de leur stratégie ? L'intensité concurrentielle s'accroît d'autant plus que les quatre autres forces décrites ci-dessus sont défavorables.
2. Des nouveaux entrants peuvent être attirés par un nouveau marché et provoquer des bouleversements importants dans la structure de celui-ci. La menace que représentent

aujourd'hui les nouveaux entrants est plus importante qu'elle ne l'était dans le passé pour deux raisons principales. D'une part, la fluidité croissante des compétences entre entreprises par le biais des progiciels, des technologies et des mutations de salariés, rend les barrières traditionnelles moins opérantes : c'est ainsi que l'on peut voir de nouveaux entrants dans des secteurs très traditionnels en cours de consolidation tels que l'automobile avec l'arrivée de Smart. D'autre part, la volonté des entreprises d'utiliser aux mieux leurs actifs existants les amène à entrer dans des secteurs nouveaux qui partagent certains de ces actifs.

3. En aval de l'entreprise se trouvent les distributeurs et les clients finaux. Selon la structure de la distribution, selon le nombre de clients, le rapport de force entre l'entreprise et ses distributeurs ou ses clients n'est pas le même.
4. En amont de l'entreprise se trouvent les fournisseurs, là aussi s'établit un rapport de force entre l'entreprise et ses partenaires. Les rapports de force avec les fournisseurs sont régis par leurs tailles respectives.
5. La cinquième force prise en compte par le modèle de Michael Porter est celle des produits de substitution qui, par la concurrence qu'ils exercent sur les produits existants, peuvent menacer le succès stratégique d'une entreprise. Cette force est extrêmement vive puisqu'elle représente le cœur même du système : c'est elle qui permet de remplacer des produits existants par des produits nouveaux présentant un avantage de coûts ou des fonctionnalités nouvelles.

1.3.3. La synthèse SWOT

Enfin, le stratège va faire une synthèse des principaux points mise en évidence dans les analyses précédentes en répertoriant, d'une part forces et faiblesses, et d'autre part, opportunités et menaces liées à l'environnement. Développé à Harvard au milieu des années 1960, le modèle SWOT (Strength, Weakness, Opportunity, Threat) ou FFOM en français (Forces, Faiblesses, Opportunités, Menaces), également appelé modèle LCAG (car proposé par quatre professeurs de la Harvard Business School, Learned, Christensen, Andrews et Guth), permet de synthétiser l'analyse interne et l'analyse externe.

Les opportunités et menaces

Les opportunités et menaces peuvent être évaluées en analysant les caractéristiques de l'offre et de la demande du marché concerné :

- Caractéristique du secteur : taille, croissance, nature de la demande, barrière à l'entrée, poids relatif produit/service, phénomènes de substitution, etc. ;
- Capacité de production : structure de coûts, c'est-à-dire main d'œuvre et investissements, ticket d'entrée, effet d'échelle, surcapacité ou sous-capacité ;
- Technologie : rythme d'innovation, brevets, rapidité des renouvellements technologiques, rôle de l'information, vitesse de propagation de l'information ;
- Distribution : spécificité des réseaux, stabilité, phénomènes de substitution, nouveaux modèles économiques...
- Concurrence : concentration, intégration, forces et faiblesses de la concurrence, vitalité des nouveaux modèles économiques, nouveaux entrants, concurrents directs et indirects ;
- Impact des méga-tendances : capacité de l'entreprise à en tirer parti ou à déjouer les menaces, etc. ;

Les forces et faiblesses

Les forces et faiblesses de l'entreprise peuvent être analysées en fonction de critères tels que :

- Principales bases d'expérience et position concurrentielle relative ;
- Client : connaissance des clients, veille marketing et concurrentielle, part de marché, existence de marques et valorisation, fidélité des clients à l'entreprise, qualité des produits, niveaux de prix, taille des clients, force du réseau commercial, savoir-faire marketing etc. ;
- Production : délai de réaction, flexibilité, capacité de réaction, fiabilité, niveau de coûts, compétitivité, etc. ;
- R&D et propriété intellectuelle : bit ratio, lead time (temps de gestation), brevets, potentiel et flux (pipe), protection juridique, etc. ;
- Compétences : gestion des ressources humaines, risques de perte de compétences, systèmes de promotion, gestion des hauts potentiels, taux de départ (attrition rate), etc. ;
- Capacité de l'entreprise à tirer parti des méga-tendances les plus importantes, ouverture, diversité culturelle, dispersion des âges, etc. ;
- Puissance et valeur de la vision existante.

Le tableau SWOT présenté ci-dessous, permet en une seule vision d'avoir un aperçu d'ensemble de la problématique de l'entreprise.

Externes	Opportunités	<ul style="list-style-type: none"> - Dématérialisation des produits et multiplication des opportunités de différenciation - Développement des modes de distribution du fait de facteurs exogènes (Internet) et du fait de l'évolution de l'encadrement réglementaire - Multiplication des accès à l'information (descendante et montante) - Opportunité de réussir un marketing individualisé
	Menaces	<ul style="list-style-type: none"> - Concurrence vive : guerre des prix, nouveaux entrants, nouveaux modèles économiques - Banalisation des produits et des services et multiplication de l'offre - Communication de grande envergure menée par la concurrence pour toucher le particulier - Saturation des marchés - Délocalisation de la production des offres
Internes	Forces	<ul style="list-style-type: none"> - Produit concurrentiel de par son rôle central - Anticipation du développement via Internet - Contrats auprès de grands groupes
	Faiblesses	<ul style="list-style-type: none"> - Manque d'expertise marketing - Pas d'outils informatiques adaptés pour une analyse approfondie de la clientèle - Stratégie commerciale fondée sur le produit et non sur le consommateur

Tableau 2. Analyse SWOT (Ducreux et al., 2010)

1.4. La veille :

Il existe dans la littérature deux définitions non concordantes de la veille, une peu formalisée et passive et l'autre formalisée et proactive. Cette discordance est due à la présence de « l'incertitude » qui caractérise l'environnement (Kislin, 2007) (El-haddadi, 2008).

1.4.1. La veille peu formalisée et passive :

Elle correspond à une activité “en tâche de fond” de surveillance selon certains axes prédéfinis avec des besoins informationnels pas nécessairement exprimés.

- Selon Lesca (Lesca, 1994) la veille est « le **processus informationnel** par lequel l'entreprise se met à **l'écoute anticipative** des signaux faibles de son environnement dans le but créatif de **découvrir des opportunités** et de réduire son **incertitude** ».
- Selon Martinet (Martinet et Marti, 1995) elle est « une attitude plus ou moins organisée d'écoute des signaux provenant de l'environnement de l'entreprise et susceptible de mettre en cause ces options stratégiques ».
- Selon Atamer et Calori (Atamer et Calori, 1993) elle est comme « un système permettant d'alerter le plus tôt possible les décideurs sur les points sensibles de l'entreprise (...) Elle permet d'engager rapidement des **actions correctives** ou de remettre en question les choix stratégiques en temps réel ».
- Selon Rouach (Rouach, 1999) la veille est « **l'art de repérer, collecter et traiter, stocker** des informations et des signaux pertinents (forts, faibles) qui vont irriguer l'entreprise à tous les niveaux de rentabilité, permettre d'orienter le futur (technologique, commercial...) et également de **protéger le présent et l'avenir** face aux attaques de la concurrence».

1.4.2. La veille formalisée et proactive

Elle nécessite une méthodologie et des objectifs clairement identifiés.

- Selon Jakobiak (Jakobiak, 2004a) la veille est comme « l'observation et l'analyse de l'environnement, suivies de la **diffusion bien ciblée** des informations sélectionnées et traitées, utiles à la prise de décision stratégique »
- Selon Tarondeau (Tarondeau, 1998), il s'agit plutôt d'un **équilibre à acquérir** : « Dans un environnement turbulent et fluctuant, l'entreprise trouve, comme le funambule, son équilibre dans le mouvement. ». L'auteur entend par là opposer les stratégies ‘statiques’ dites de positionnement aux stratégies ‘dynamiques’ de changement permanent.
- S. Lardera et B. Quinio (Laedera et Quinio, 1996) distinguent quant à eux, deux types de veille : la **veille interne** et la **veille externe**. La veille interne a pour but de faciliter la créativité des individus et leur réceptivité aux nouveautés, qu'elles proviennent de l'intérieur ou de l'extérieur. La veille interne, à l'image de l'organisation apprenante, s'appuie sur les outils de gestion des compétences, de diffusion des savoirs et de travail en groupe. L'objectif de la veille externe est de capter les signaux en provenance de l'environnement pour les inclure dans la réflexion stratégique. Elle correspond à l'appréciation des possibilités extérieures en tant que préambule à toute décision stratégique.

1.4.3. La typologie de la veille

Selon le modèle de Porter (figure 6) la veille peut se décliner en plusieurs types de veilles (Martinet, 1989). Tout d'abord, quatre types de veilles, qui se résument en veille environnementale, veille technologique, veille commerciale et veille concurrentielle (figure 7).

La veille concurrentielle

- L'objectif est de suivre les mouvements de ses concurrents sur les sujets critiques ou potentiellement critiques.
- Les informations recherchées : stratégies, politiques tarifaires, nouveaux produits ou services, résultats financiers, recrutement, clients perdus ou gagnés, partenariats, rachats, nouveaux entrants...
- Les bénéfices attendus : une comparaison de son entreprise avec ses concurrents.

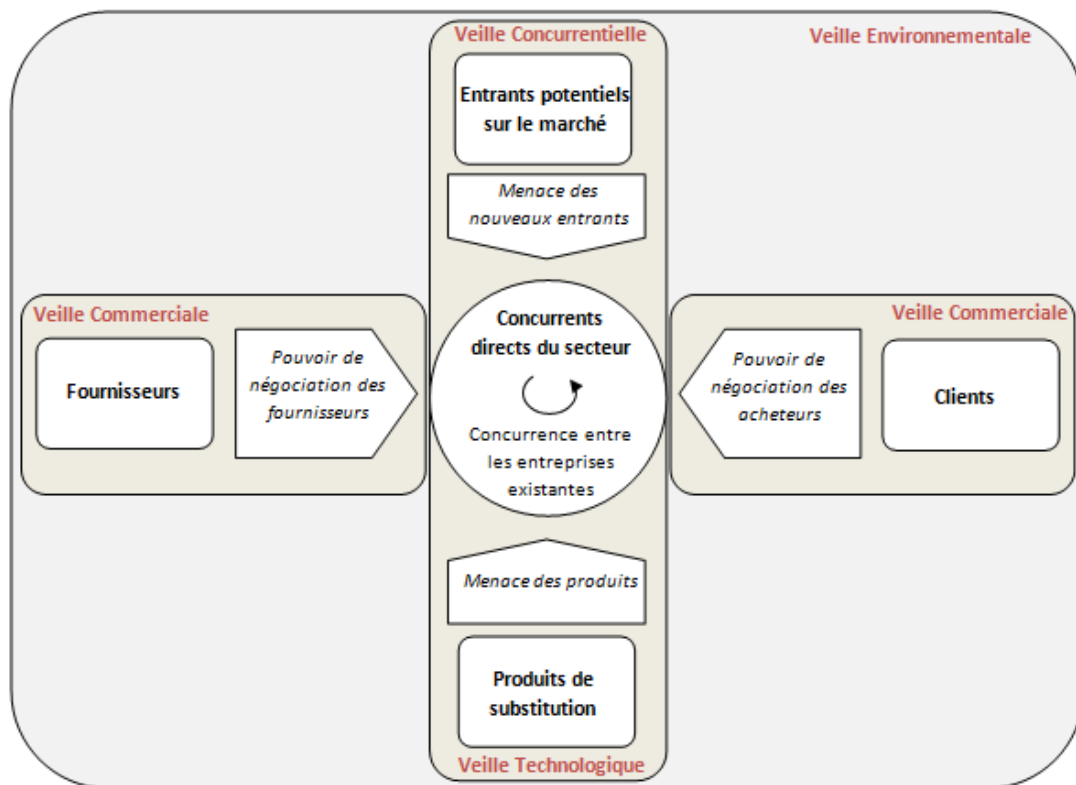


Figure 7. Typologie de la veille

La veille sectorielle

- L'objectif est de surveiller son secteur d'activité, composé des acteurs au sens large que sont les concurrents, les partenaires, les fournisseurs, les clients, la législation et l'environnement.
- Les informations recherchées sont l'actualité sur ses fournisseurs et clients, les innovations, les projets de loi, les tendances comportementales et sociétales.
- Les bénéfices attendus : anticiper les évolutions de ses marchés.

La veille technologique

- L'objectif est de surveiller les brevets, l'évolution des normes, les projets de normes.
- Il est atteint en recherchant des informations dans les bases de données de brevets.
- Les bénéfices attendus : l'anticipation des ruptures et les évolutions technologiques.

La veille scientifique

- L'objectif est de surveiller la recherche scientifique, les nouveaux domaines émergents

- Il est atteint en recherchant des informations sur les publications et travaux de recherche.
- Les bénéfices attendus : évaluation de la recherche, orientation de la R&D.

La veille réglementaire

- L'objectif est de relever les projets de loi, les textes et normes réglementaires.
- Les informations recherchées sont : les sujets légaux concernant le secteur de l'entreprise, les lois, décrets et jurisprudence, les débats parlementaires et propositions de loi, les propositions patronales et syndicales.
- Les bénéfices attendus : s'assurer d'être en amont des nouvelles règles légales.

La veille financière

- L'objectif est de détecter les mouvements sur les marchés financiers affectant potentiellement l'entreprise.
- Les informations recherchées sont les rumeurs de fusion-acquisition, cession, les écarts significatifs de cours de bourse, de matières premières.
- Les bénéfices attendus : apprécier l'environnement financier avec anticipation.

La veille territoriale

- L'objectif est de suivre les aspects socio-économiques, politiques et socioculturels d'un territoire.
- Les informations recherchées : activité des acteurs clés, initiatives civiles, économiques, culturelles, initiatives réussies dans d'autres régions, grands tendances sociétales, économiques, technologiques pouvant impacter le territoire.
- Les bénéfices attendus : apprécier les attentes et résultats des porteurs de projets pour affiner sa politique.

On peut ajouter à ces différentes veilles, la veille acteurs (suivi des cadres, recrutement des entreprises, réseaux sociaux des individus), les veilles sanitaires (relatives aux problèmes de santé publique, à la maîtrise des épidémies), les veilles éducatives (courants pédagogiques, évaluations des compétences scolaires) et les "écoveilles" mises en place dans le contexte du développement durable.

1.4.4. Processus de la veille

Le processus de veille peut se décrire comme un système, composé lui-même, selon Dutton et coll. (Dutton et Coll, 1983) de sous-systèmes, traversé par des flux informationnels provenant de sources de toute nature (humaine, documentaire, ...) et de tout type (formel, informel, ...). Ce système est influencé par le mode de management, les ressources, la culture et la structure de l'organisation dans lequel il opère et bien sûr de la considération, de la motivation et de l'implication des acteurs qui l'implantent et l'utilisent. Diverses représentations de processus existent :

- le cycle de renseignement militaire qui s'appuie sur un cycle itératif et continu de quatre phases distinctes : l'acquisition de l'information suite à une expression des besoins, l'interprétation, l'appréciation et enfin la communication de cette information.
- Le processus de la veille pour le management stratégique proposé par H. Lesca (Lesca, 1994) (Lesca et Rouibah, 1997) (Lesca, 2004) qui a comme principaux objectifs : de définir l'orientation de la veille ; d'identifier les acteurs et types d'informations à surveiller ; d'aider

les utilisateurs à ne sélectionner que l'information cruciale ; de définir le portrait d'un bon traqueur de signaux faibles (signaux d'alerte précoce) ; de déterminer les critères d'un bon outil pour la communication des informations de type signaux faibles ; de transformer les signaux faibles en information utile et significative ; d'intégrer les informations ainsi recueillies dans un processus décisionnel ; et de faire un diagnostic de la veille pour progresser.

- Le processus de surveillance sectorisée de Jakobiak (Jakobiak, 2004a) (Ibekwe-Senjuan, 1997) est également construite autour de huit étapes s'inspirant fortement du cycle du renseignement, Jakobiak reconnaît au moins trois étapes fondamentales dans le processus : la collecte, le traitement et la diffusion de l'information et y ajoute l'expression des besoins, et la mémorisation de l'information dans un corpus de veille.

Pour résumer, la veille s'articule selon un mouvement itératif permettant d'optimiser et de réorienter chacune des quatre actions principales qui le composent :

- Analyse des besoins ;
- Recherche et collecte des informations ;
- Traitement et analyse des informations ;
- Mémorisation et diffusion de l'information utile.

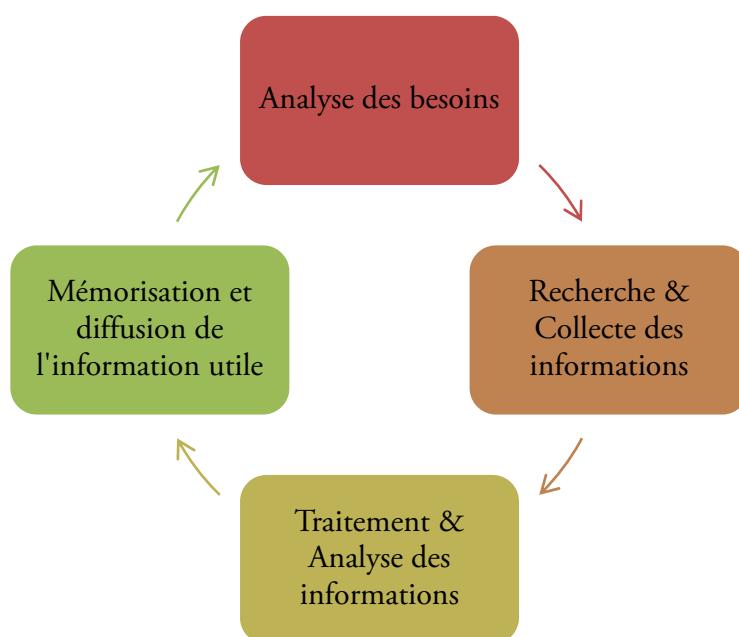


Figure 8. Cycle de la veille

➤ Analyse des besoins

L'analyse des besoins, première phase de veille stratégique, conditionne sa réussite dans la mesure où elle définit les besoins réels de dirigeants (ou destinataires des résultats de la veille). Phase de communication et d'échange, c'est l'étape dans laquelle les enjeux sont définis et traduits en axes de veille. Nécessitant rigueur et méthode, c'est l'une des phases les plus difficiles de la démarche de veille.

La phase **analyser** constitue le socle de la détermination de l'univers à explorer. Elle donne toutes les bases des usages courants des dirigeants. Elle repose sur une étude exhaustive des sources d'information

et services existants ainsi que leurs usages. La réussite de cette phase repose sur la capacité de ceux qui vont analyser le besoin à l'extraire et le traduire en axes de veille.

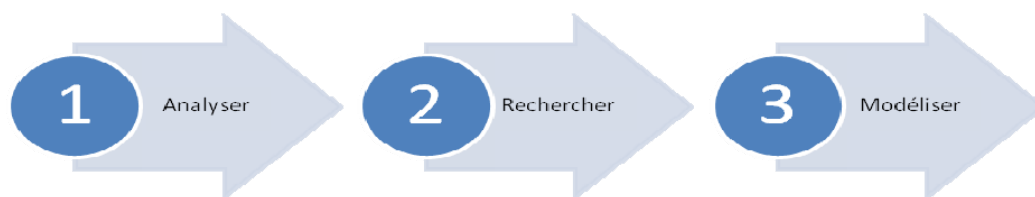


Figure 9. la phase analyse des besoins de la veille

La phase **rechercher** est normalement la plus créative. Elle permet à la fois de faire émerger toutes les idées possibles mais aussi de replacer ces idées dans le champ de contraintes (temporelles et financières).

La phase **modéliser** est l'aboutissement de l'étude. Elle consiste à rédiger les scénarii de collecte, tel que chaque axe de veille pourra être exploré selon un ou plusieurs scénarii, recouvrant l'univers défini. Ils permettent de concentrer la veille sur des signaux d'alerte potentiels en fonction des événements prévisibles imaginés.

➤ Recherche et collecte d'informations

Une fois que l'on sait ce que l'on cherche, il faut s'avoir où trouver l'information. Les sources sont nombreuses, diverses, et plus ou moins facilement accessibles. Cependant, il existe plusieurs types de sources à surveiller : les bases de données scientifiques, les bases de brevets, les médias, la presse, les flux RSS, l'Internet, l'Intranet, les forums, les blogs, les wikis, les sites de micro-blogging, les bookmarking sociaux, les mashup, les crowdsourcing, les réseaux sociaux, etc. Nous définissons deux types de sources d'information :

Les sources formelles

L'information est dite formelle dès qu'elle est publiée sur support papier, informatique, microfilm, etc. Elle peut être structurée ou non, mais il s'agit dans tous les cas d'une information directement accessible et exploitable. Ce type de sources correspond à l'information blanche.

Les sources formelles sont composées principalement de la presse, la télévision, la radio, les livres, les banques de données et CD-ROM, les brevets, les informations légales, les études réalisées par des prestataires publics ou privés, Internet. Ces sources ont l'avantage d'être sûres, assez exhaustives et faciles d'accès.

Dans un contexte de veille stratégique, les bases de données les plus consultées sont à dominance scientifique, technologique, réglementaire et se trouvent sur des bases bibliographiques. Parmi les bases de données les plus intéressantes, nous pouvons citer par exemple Factiva dans le domaine économique, Inspec dans le domaine physique et Kompass Europe orientée entreprise, Pascal qui représente une source multidisciplinaire et PubMed spécialisée dans le domaine médical.



Figure 10. Les sources formelle pour l'intelligence économique

Les sources informelles

L'information informelle est constituée de toutes les informations non formalisées et non disponibles directement. Il est donc nécessaire d'entreprendre des démarches directes auprès des détenteurs supposés de cette information. Ce type de sources correspond à l'information grise. Ces sources peuvent être les expositions et les salons, les fournisseurs, les colloques, les congrès, les clubs. On y échange et communique des informations. L'information qui circule est souvent d'une grande valeur stratégique telles que celles des concurrents à l'occasion de portes ouvertes, ses communications commerciales et financières, ses publications dans des journaux internes, etc. Sans oublier les sources internes de l'entreprise qui représentent plus de 80% des informations de ce qu'un décideur peut chercher comme information alors qu'elle se trouve dans son entreprise, des sites personnels, des études de recherche menées par un groupe d'étudiants ou de thésard, etc. ainsi que dans les réseaux personnels dans la limite de la légalité et de la déontologie.

➤ La collecte des informations

La collecte s'effectue à partir des sources sélectionnées. Elle peut être :

- Semi-automatique, consistant à suivre l'actualité à l'aide de portail agrégateurs (type netvibes) alimentés par une sélection de flux RSS issus des sites majeurs. Mais se limiter à cela, c'est accepter le risque d'avoir une vision déformée de la réalité, aussi bien en terme de quantité que de qualité. Les agrégateurs réduisent le champ de vision et faussent la veille : la réalité est alors déformée.
- Automatique avec l'aide d'outils informatiques.

➤ Traitement et analyse des informations

C'est l'étape déterminante du cycle, celle qui conditionne le succès de la veille. En effet, c'est à ce stade que la réflexion entre en jeu pour interpréter les diverses informations obtenues et ainsi apporter une réelle valeur ajoutée au processus.

L'information étant collectée, évaluée et traitée, l'analyse peut s'effectuer selon une méthodologie, des outils et des tableaux de bord permettant :

- d'évaluer l'évolution du domaine étudié dans le temps ;
- de détecter l'apparition d'un nouveau concept ou sujet ;

- de mettre en évidence, par des outils de cartographie, des réseaux de spécialistes, des réseaux sémantiques, des collaborations, des alliances, etc.
- de mettre en évidence des signaux faibles.

L'analyse repose sur les techniques de la fouille de texte (ou text mining), de découverte de connaissance et les statistiques décisionnelles. Cela permet d'extraire des connaissances sur les documents concernés sur une période définie, voire une zone géographique choisie. C'est possible grâce au système d'intelligence économique que nous proposons.

➤ **Mémorisation et diffusion de l'information utile**

Le but de cette étape et de faire un bon reporting, en se basant sur les techniques de visualisation, donc cette étape doit permettre :

- Une représentation graphique des informations.
- Un archivage spécifique des informations utiles aux décideurs en fonction de leur durée de vie (validité).
- De générer des rapports à la demande pour les décideurs.
- De gérer des groupes de discussion avec droits d'accès, échange et partage de l'information ; chacun profite des informations récoltées par les autres.
- Un pilotage plus pointu, une gestion des feed-back.
- Un ciblage par rapport aux axes de développement de l'entreprise, un suivi des acteurs, des thèmes les plus actifs et des sources surveillées.
- Un pilotage en temps réel de l'activité de veille, grâce à des tableaux de bords permettant de contrôler et réorienter l'activité en fonction des objectifs.

1.5. La protection de l'information

Le besoin en matière de sécurité dans un SIE provient du fait que les informations manipulées sont d'ordre stratégique ayant une valeur assez importante. Une telle sécurité ne doit pas être considérée comme une option supplémentaire qu'offre un tel système pour se distinguer d'un autre, d'autant plus que la fuite de ces informations n'est pas le fait de faiblesses inhérentes aux systèmes informatiques des entreprises, mais c'est avant tout une question organisationnelle.

Dès lors, les mêmes procédures organisationnelles, relatives à l'utilisation des ressources informatiques et visant à assurer la fonction de veille stratégique et concurrentielle dans un SIE, devraient aussi s'appliquer au domaine du contrôle de l'information. L'établissement d'une politique de sécurité, la sensibilisation des salariés, la mise en garde quant au contenu diffusé et l'adhésion à des bonnes pratiques sont des outils potentiels pour réduire le risque des fuites d'information.

Nous rappelons que l'IE, est à la fois un processus et un produit. En tant que processus, l'IE est l'ensemble des méthodes légales et éthiques qu'une entreprise utilise pour mobiliser l'information qui l'aide à réussir dans un environnement mondialisé. En tant que produit, l'IE est l'ensemble des informations sur les activités des concurrents à partir de sources publiques et privées. Son champ d'application est le comportement présent et futur des concurrents, des fournisseurs, des clients, de la technologie, des acquisitions et fusions, des marchés, des produits et services et de l'environnement des affaires en général.

Il s'agit donc de sécuriser à la fois des données (IE étant un produit) et des procédés (IE en tant que méthode de traitement). De là est partie notre contribution de proposer une approche de contrôle d'accès aux données et aux traitements dans un SIE. Il s'agit d'un modèle générique pour intégrer le contrôle d'accès dans le cycle de vie des SIE afin que toutes les tâches dans un processus d'IE soient exécutées uniquement par les utilisateurs autorisés et habilités.

La notion de protection est plus large qu'il n'y paraît. Protéger en enfermant ou protéger en diffusant ? Il n'existe pas de recette miracle mais des choix stratégiques à opérer. Nous traitons cet aspect dans le troisième chapitre de ce manuscrit.

1.6. Synthèse

L'intelligence économique est le développement d'une véritable activité de collecte et de traitement des informations qu'il est possible de se constituer sur l'environnement, afin de gérer les enjeux stratégiques d'une entreprise. Ces enjeux ne s'identifient pas seuls, de façon explicite. La pratique de la stratégie nécessite une forme de veille concurrentielle pour laquelle les informations constituent la principale source d'efficacité. Quelle que soit la stratégie poursuivie, la collecte et le traitement des informations sont incontournables. Pour illustrer les différentes applications de la veille stratégique, Bruno Martinet et Jean-Michel Ribault (Martinet et Ribault, 1988) (Delbecque, 2006) (Coutenceau et al., 2010) proposent de s'appuyer sur le cadre des cinq forces de Porte :

- la veille concurrentielle comprend l'observation et la collecte de données sur les concurrents et les nouveaux entrants potentiels ;
- la veille technologique s'intéresse aux pratiques des concurrents et aux entreprises susceptibles de fournir des produits de substitution ;
- la veille commerciale se concentre sur les fournisseurs et les clients ;
- enfin, la veille environnementale reprend les principes de l'analyse stratégique sectorielle.

Cependant, la valeur apportée par les informations dépend du cadre de collecte, de traitement et de diffusion. B. Martinet et Y.M. Marti proposent la représentation suivante :

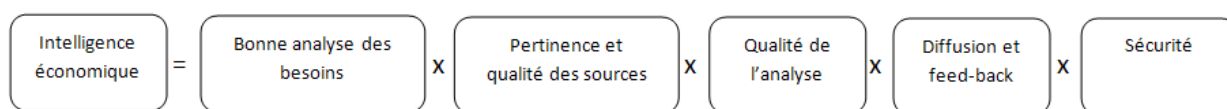


Figure 11. La valeur d'intelligence économique (Coutenceau et al., 2010)

Grâce à l'application du processus l'intelligence économique, c'est-à-dire la mise en place d'une démarche forte du traitement des informations, quatre implications stratégiques sont identifiées (Ducieux et al., 2010):

- La vision, souvent résultat collectif porté par le leadership du dirigeant ;
- L'innovation, aboutissement d'une problématique spécifique ;
- La création, ouverture vers un possible jusqu'alors inconnu ;
- Et l'anticipation, laquelle combine problématisation, issue des données de l'environnement et intentionnalité.

Pour adresser ce problème, nous proposons un système d'intelligence économique, qui va plus loin que les systèmes d'information mais comprend ces derniers comme une étape pour la diffusion, la disponibilité d'interprétation et pratique au stockage et à la sécurité de l'ensemble. Selon les contextes et les données à collecter ou à traiter, chaque membre de l'organisation peut agir sous la triple forme de décideur, collecteur ou spécialiste.

Pour résumer, le concept d'IE induit une culture collective de l'information accompagnée d'actions offensives, avec le souci constant de protéger le capital informationnel de l'organisation. Concept à part entière, il est un assemblage de diverses notions, comme le montre le réseau sémantique suivant.



Figure 12. Le réseau sémantique de l'intelligence économique

Nos travaux vont dès lors consister à définir un modèle de traitement et de gestion des données dans le contexte du processus d'IE. Ainsi, notre modèle d'analyse multidimensionnelle se base sur les cinq principales étapes de l'IE, à savoir « La formulation du besoin, La collecte et le traitement des données, L'analyse, La restitution et l'interprétation des résultats et La sécurisation des informations et des traitements ». Ce modèle d'analyse multidimensionnelle que nous proposons a pour objectif principal la création de nouvelles connaissances souvent implicites et à visée stratégique, qui se déduisent le plus souvent de l'analyse de l'évolution des réseaux d'acteurs (auteurs, inventeurs, sociétés, villes, régions, pays, journaux, etc.), des réseaux sémantiques (mots-clés, multi-termes, ontologies, etc.) et des interactions (acteur – sémantique).

Notre démarche consiste alors à définir un système d'intelligence économique dont l'objectif est:

- De supporter les différentes étapes de la démarche en offrant toutes les fonctionnalités présentées dans les sections 1.2, 1.3 et 1.4,

- D'extraire et traiter les données textuelles issues des sources identifiées dans la section 1.4.4,
- De répondre aux besoins d'analyse de la démarche de l'IE en offrant des connaissances synthétiques de l'environnement informationnel de l'entreprise, sur un domaine donné,
- De Restituer les résultats sous forme d'un portail Web / Mobile.

Afin de répondre à ces besoins d'analyse, nous proposons le prototype d'un système d'information d'intelligence économique qui couple les besoins des étapes du processus d'IE avec les principes et techniques issus du domaine de la découverte des connaissances à partir du texte. Dans ce contexte, nous présentons, dans le deuxième chapitre de ce manuscrit, les systèmes d'information d'intelligence économique.

Chapitre 2.

Le système d'intelligence économique

« N'entrons pas dans l'avenir à reculons, c'est-à-dire avec retard et avec regret » Paul Valéry

2.1.	Introduction	55
2.2.	Les systèmes d'information et la stratégie	56
2.2.1.	L'émergence des systèmes d'information stratégique	56
2.2.2.	La dimension interne des S-IS	57
2.2.3.	La dimension externe des S-IS ; veille et intelligence économique	58
2.2.4.	L'émergence de système d'intelligence économique	59
2.3.	Les méthodes de conception des systèmes d'intelligence économique	60
2.3.1.	La méthode d'analyse MEDESIIE	60
2.3.2.	Les modèles SITE	61
2.3.3.	Le modèle XPlor	62
➤	Analyse des besoins	62
➤	Recherche et collecte des informations	65
➤	Homogénéisation et structuration des données	66
➤	Traitement des structures de données	66
➤	Définitions des objets dans les méta-données	67
➤	Descripteurs de formats spécifiques	68
➤	Exemple de descripteur générique	68
2.4.	Les systèmes d'intelligence économique	70
2.5.	Le système d'intelligence économique "XPlor EveryWhere"	74
2.5.1.	Architecture service de XPlor EveryWhere	75
➤	Le service « Sources et traitement »	75
➤	Entrepôt de données stratégique	76
➤	Service Analyse et restitution	76
➤	Service Sécurité et Administration	77
2.5.2.	La structure statique de XPlor	77
2.5.3.	Architecture Technique de XPlor EveryWhere	79
2.5.4.	Architecture logicielle du système	81
2.6.	Synthèse	83

2.1. Introduction

Plus qu'un demi-siècle de progrès phénoménal des technologies de l'information tel que 1954 présente l'année zéro des systèmes d'information avec l'Electronic Data Processing (EDP). Puis les Management Information Systems (MIS) ont émergé vers le milieu des années 1960. L'objectif de ces SI est de fournir de l'information aux décideurs, à la différence du traitement automatique de données (EDP), qui se concentrait sur l'automatisation de procédures (administratives) existantes, comme le traitement de fiche de paie (Vidal et Petit, 2009).

A la fin des années 1960 et au début des années 1970, le domaine des systèmes d'information est encore embryonnaire et certains concepts ne sont pas encore bien définis. Bien souvent, on assimile le système d'information à un système de traitement automatique de données, outil essentiel d'une « gestion automatisée » avec l'idée que ces SI doivent être capables de fournir la bonne information, à la bonne personne, au bon moment (la règle des R3 : the Right information at the Right time to the Right person). Quelque année après, les principaux paradigmes du système d'information organisationnel (SIO) sont présents dans l'ouvrage « Les systèmes d'information dans les organisations » de J.-L. Le Moigne. Dans les années 1980, le domaine connaît une maturité avec la création des premiers centres de recherche sur les SI, les universités continuent à créer des départements indépendants de recherche sur les MIS et mettent en place des programmes spécifiques d'enseignement sur les systèmes d'information. De plus, la crise économique des années 1980 a modifié les pratiques sur le terrain, ce qui a donné naissance aux systèmes d'information stratégique. C'est C. Wiseman qui, le premier, constate l'association, sur le terrain, entre informatique et stratégie (Vidal et Petit, 2009). Il suggère de changer de lunettes conceptuelles, c'est-à-dire d'abandonner l'approche classique des systèmes d'information, et de se placer dans une perspective stratégique. Cette dimension stratégique des systèmes d'information fera dès lors l'objet de nombreux travaux : H. Tardieu et B. Guthmann (1991) ont affiné la définition des systèmes d'information stratégique et posé les bases du triangle stratégique « qui met en œuvre de façon coordonnée la stratégie de l'entreprise, sa structure et les technologies de l'information ».

Actuellement, on parle de plus en plus des systèmes d'intelligence économique. Grâce à l'application du processus l'intelligence économique, mentionnée dans la synthèse du premier chapitre, un système d'intelligence économique, va plus loin que les systèmes d'information classiques mais comprend ces derniers comme une étape pour la diffusion, la disponibilité d'interprétation et à la sécurité de l'ensemble. Selon les contextes et les données à collecter ou à traiter, chaque membre de l'organisation peut agir sous la triple forme de décideur, collecteur ou spécialiste.

Mais que présente vraiment la notion de « système d'intelligence économique » ? S'agit-il simplement d'un nouveau terme à la mode ? Ou s'agit-il d'une association entre des systèmes d'information stratégique & collaboratifs ? Ou bien est-ce l'art du pilotage des systèmes d'information ? D'abord, nous ferons le point sur les systèmes d'information stratégique dans la première section de ce chapitre (2.2). Ensuite, dans la seconde section (2.3), nous présenterons les systèmes d'intelligence économique, ainsi que les différents modèles proposés dans la littérature pour la conception de ce type de systèmes d'information. Enfin, la troisième section (2.4) fait l'objet de la proposition de notre modèle de système d'intelligence économique.

2.2. Les systèmes d'information et la stratégie

2.2.1. L'émergence des systèmes d'information stratégique

Les premiers travaux dédiés à la stratégie d'entreprise mentionnent rarement les systèmes d'information. L'émergence des systèmes d'information stratégique est attribuée à C. Wiseman à la fin des années 1980. Il définit la stratégie comme un nouveau rôle des systèmes d'information comme le montre le tableau suivant, qui introduit le lien entre la stratégie et SI :

Rôle fonction	Automatiser les procédures de base	Satisfaire les besoins en information	Supporter ou former la stratégie concurrentielle
Gestion des transactions	Système d'information opérationnels	Systèmes d'information stratégiques	
Recherche et analyse	Aide à la décision		

Tableau 3. Les rôles des systèmes d'information (Vidal et Petit, 2009)

Au début des années 1990, deux modélisations très proches présentent une synthèse des approches antérieures. La première est proposée par une équipe du MIT, menée par M. Scott Morton (1995), et réalisée à la demande de grands entreprises (American Express, Kodak, General Motors, etc.). Son objectif est de décrire l'impact des technologies de l'information sur les organisations, et en quoi cet impact peut modifier l'organisation des entreprises, et plus largement la donne concurrentielle.

Cette étude aboutit à la conception d'un cadre général d'étude, dit « MIT90 », qui propose une nouvelle représentation de l'entreprise. Cette représentation nous montre que le SI peut impacter le fonctionnement de l'entreprise. Ainsi, elle insiste sur la technologie, facteur primordial de l'émergence du concept de SI dans le cadre de la réflexion stratégique (Vidal et Petit, 2009).

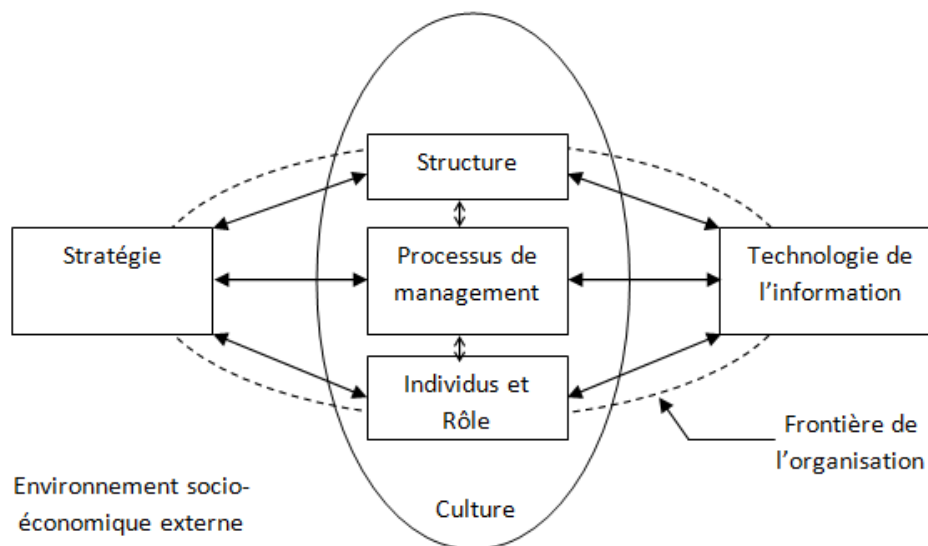


Figure 13. Le cadre général de l'étude MIT 90 (Source : Pascal. Vidal, 2009, p. 25)

Le schéma de M. Scott Morton est repris par H. Tardieu (1991), qui propose une modélisation du concept recentrée autour de la notion de triangle stratégique.

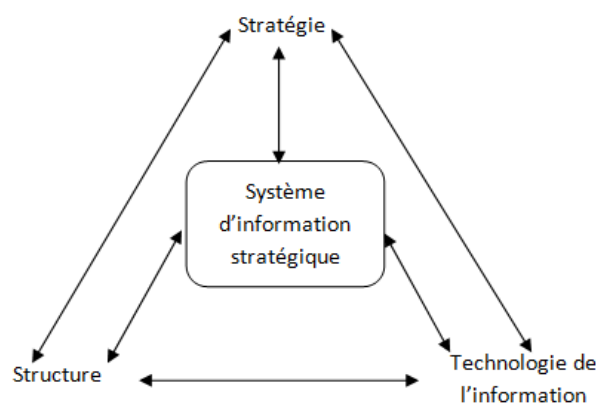


Figure 14. La dimension stratégique des systèmes d'information (Source : Pascal. Vidal, 2009, p. 26)

Si cette modélisation paraît moins développée que le cadre « MIT90 », elle introduit cependant plusieurs conceptions intéressantes :

- Elle explique le concept de SI stratégique, mais qui demeure implicite dans le cadre MIT90.
- Le SI stratégique devient un équilibre entre la stratégie et les SI.

H. Tardieu et B. Guthman distinguent deux types de systèmes d'information stratégiques :

- Le **SI-S**, ou **système d'information – stratégique**, utilise le système d'information pour apporter un avantage concurrentiel à l'entreprise. Celui-ci devient donc le support de la stratégie.
- Le **S-IS**, ou **système – d'informations stratégiques**, apporte aux dirigeants l'ensemble des informations et/ou modélisation nécessaires pour enrichir le processus de décision stratégique. Le SI est la base de la décision.

Quels que soient les modèles mobilisés, que l'on parle de diagnostic interne ou externe, la problématique reste la même : la prise de décision stratégique, et le management stratégique de l'information parce qu'elle est naturellement plus difficile que les autres, nécessite de recueillir et d'organiser une grande variété d'informations qui concernent l'entreprise et son environnement au sens large (concurrent, environnement légal, etc.).

Le S-IS a donc une double dimension :

- Une dimension interne : la capacité de l'entreprise à collecter, traiter et organiser les différentes données opérationnelles afin de construire la représentation la plus précise possible de son propre fonctionnement ;
- Une dimension externe : collecter dans l'environnement les informations qui sont susceptibles d'impacter à court, moyen et long terme la stratégie de l'entreprise.

2.2.2. La dimension interne des S-IS

L'analyse stratégique est une décision collective. Prise au plus haut niveau hiérarchique de l'organisation, elle doit être transversale et doit concerner l'organisation dans son ensemble. Le système d'information qui va collecter les informations internes nécessaires à la prise de la décision stratégique doit donc présenter plusieurs caractéristiques :

- Capacité à agréger les informations « verticalement », puisque la décision stratégique est prise au plus haut niveau de l'organisation. Le S-IS interne doit proposer une représentation agrégée de toutes les informations qui concernent une fonction particulière, depuis les informations opérationnelles de base.
- Capacité à agréger les informations « horizontalement ». Le S-IS interne doit rassembler des informations provenant de fonction habituellement cloisonnées et disposant souvent de systèmes d'information indépendants.

Pour toutes ces raisons, dès que la technologie l'a permis, les concepteurs de systèmes d'information ont envisagé la mise en place de systèmes aptes à enregistrer l'ensemble des informations issues de l'entreprise dans des programmes uniques, pour fournir aux managers des données cohérentes et disponibles rapidement. Ainsi sont nés les ERP (Entreprise Resource Planning), systèmes d'information « globaux » aptes à collecter, organiser et rendre disponibles des informations issues de toute l'organisation.

2.2.3. La dimension externe des S-IS ; veille et intelligence économique

La stratégie de l'entreprise a toujours une double dimension « intégrative » (centrée sur l'interne et le développement des compétences clés) et « adaptative » (tournée vers l'externe et la recherche d'opportunités stratégiques). Le système d'informations stratégiques centré sur la recherche des informations issues de l'entreprise ne recouvre donc qu'une partie du problème. La conception de la stratégie nécessite également de disposer d'informations externes, par nature beaucoup moins disponibles que les informations internes.

Une fois encore, c'est la technologie qui a permis le développement de systèmes d'information centrés sur cet objectif, par la conjonction de deux phénomènes connexes :

- La capacité technologique à concevoir des systèmes aptes à parcourir, sélectionner, trier des volumes d'information très importants. Ceux-ci doivent disposer de performances de stockage considérables et de traitement très rapides (afin de disposer d'informations actualisées).
- La disponibilité des informations. Il y a quinze ans, même avec de telles capacités, trouver l'information recherchée, la matière première, se révélait difficile. Certes, les salons et la presse, entre autres, pouvaient déjà renseigner l'organisation. Mais c'est bien entendu l'explosion d'Internet qui fait progresser de manière exponentielle la quantité d'informations disponibles... et donc le besoin d'outils pour la traiter.

La veille stratégique et l'intelligence économique dégagent une nouvelle problématique. Mais elles fondent l'existence de systèmes d'information dédiés, en créant une fonction supplémentaire, un nouveau champ dans lequel s'exprime le besoin de SI. En même temps, elles assurent aussi leur propre légitimité sur de tels systèmes, car sans le secours des SI, serait-il possible d'envisager des procédures de veille ou d'intelligence économique, et géographiquement dispersées ?

Un S-IS externe ne repose pas sur la même logique qu'un système intégré comme un ERP. Il doit traiter des informations moins cohérentes, donc moins fiables, et proposer des rapprochements entre des données provenant de sources différenciées. Il doit également pouvoir en évaluer le poids relatif, ou du moins les représenter sous une forme décodable par le décideur.

Les S-IS externes sont aujourd'hui en phase d'émergence et beaucoup moins répandus que les outils fondés sur la collecte de l'information interne. La complexité de leur développement et la fiabilité de l'information traitée ralentissent certainement leur développement.

2.2.4. L'émergence de systèmes d'intelligence économique

Toute présentation stratégique dans les systèmes d'information s'appuie aujourd'hui sur une dimension informationnelle. Cette dimension utilise des systèmes de nature différente :

- Les S-IS, orientés sur la recherche d'information afin d'enrichir le processus de décision stratégique.
- Les SI-S, conçus pour donner à l'organisation un avantage sur la concurrence.

Mais la distinction entre S-IS et SI-S est artificielle (Vidal et Petit, 2009) : l'exercice d'un SI-S est naturellement générateur d'informations pour un S-IS et les informations collectées et organisées dans le cadre d'un S-IS pourront constituer une base intéressante pour le développement des SI-S. C'est ce que montrent H. Tradieu et B. Guthmann, en évoquant les boucles récursives qui lient les systèmes.

L'alliance S-IS et SI-S donne naissance au système d'intelligence économique, qu'on peut définir comme étant : « un système d'information stratégique externe, qui a comme mission la collecte dans l'environnement des informations qui sont susceptibles d'impacter à court, moyen et long termes la stratégie de l'entreprise ainsi que la gouvernance interne de l'entreprise ». Comment peut-on modéliser un tel système d'information ? Quelle est l'architecture adéquate pour développer ce type de système d'information ? La prochaine section nous permettra de répondre à ces deux questions.

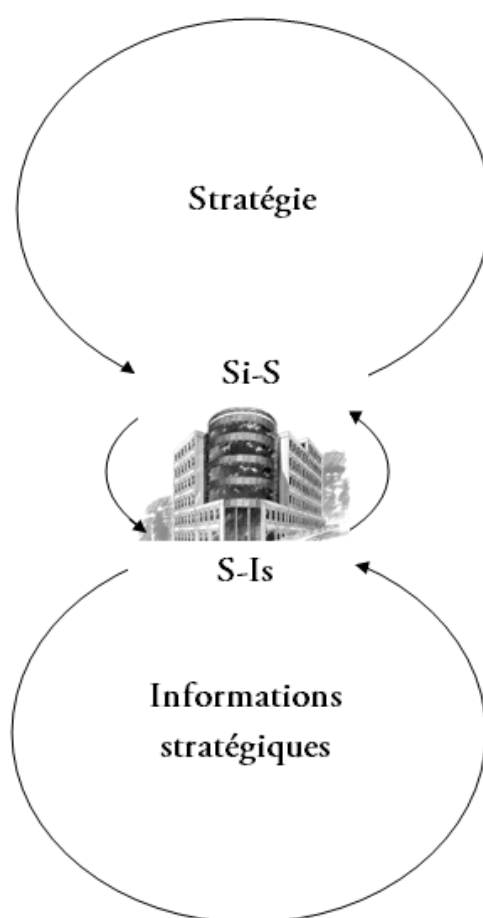


Figure 15. Stratégie et SI, boucles récursives (Extrapoler de la Source : Pascal. Vidal, 2009, p. 141)

2.3. Les méthodes de conception des systèmes d'intelligence économique

Depuis deux décennies de travaux de recherche en IE diverses difficultés se posent toujours lors de la conception des systèmes d'intelligence économique, surtout, qu'on trouve dans plusieurs travaux (Afolabi, 2007) (Boukrara, 2010) (Besson et al., 2010) (Othenin et al., 2011) une confusion entre les systèmes d'intelligence économique et les systèmes de business intelligence. Ceci peut s'expliquer par le manque de méthodes, la mauvaise traduction du terme intelligence économique. Or, la plupart des sources en domaine sont des constats des experts et rapport des hommes d'états, qui ignorent la dimension théorique du concept d'IE. Ces travaux de recherche montrent que ces modèles d'analyse d'intelligence économique se focalisent sur quatre dimensions clés à savoir (Ghallamalah, 2009) :

- La dimension environnementale d'une entreprise qui regroupe les éléments pouvant influencer d'une manière directe ou indirecte l'évolution stratégique d'une entreprise. Elle est caractérisée par les partenaires, les concurrents, les marchés, les clients, etc.
- La dimension humaine qui englobe les acteurs impliqués dans une démarche d'IE, qu'ils soient internes ou externes à l'entreprise. Elle est caractérisée par les réseaux de collaborations, d'interactions et de communications entre les différents acteurs impliqués dans cette démarche.
- La dimension stratégique qui correspond aux différents modèles d'analyse permettant l'élaboration des stratégies d'une entreprise allant de l'identification des objectifs à la décision et à la définition d'actions.
- La dimension technologique qui rassemble toutes les méthodes, outils et techniques utilisés pour répondre aux besoins informationnels dans une démarche d'IE, par les procédures de recherche, collecte, traitements et diffusion de l'information.

La prise en compte de l'une ou plusieurs de ces dimensions permet de concevoir plusieurs modèles servant d'analyse (conceptuelle ou pratique) de l'IE. Nous avons retenu deux modèles universitaires bâtis sur ces dimensions : le modèle MEDESIIE, le modèle SITE.

2.3.1. La méthode d'analyse MEDESIIE

La démarche d'IE proposée par Maryse Salles dans le cadre du projet MEDESIIE⁶ est consacrée entièrement à l'analyse du besoin en IE des PME. MEDESIIE considère le système d'IE comme le système de représentation des connaissances de l'entreprise, concernant les relations, réelles ou potentielles, qu'elle entretient avec son environnement. Un tel système d'IE a pour vocation première d'apporter une aide à la décision, pour toutes les décisions qui affectent les relations de l'entreprise avec son environnement, de quelque manière que ce soit et avec quelque intensité que ce soit⁷.

L'architecture conceptuelle de ce système s'inspire de la définition proposée par Seligmann pour la conception des systèmes d'information. Seligmann décrit une méthode comme constituée obligatoirement de quatre composants, qu'il désigne comme des "manières de" (way of) (Salles, 2003): the way of thinking (le paradigme, le point de vue), the way of modelling (les modèles à construire), the way of organizing (la démarche à suivre), the way of supporting (un support rassemblant les outils).

⁶ Projet MEDESIIE (Méthode de définition de système d'information pour l'intelligence économique), consacré entièrement à l'analyse du besoin en intelligence économique (i.e.) des PME, et financé par la région Midi-Pyrénées.

⁷ Source : <http://ieut1.irit.fr/introduction/intro2.htm> (Projet MEDESIIE)

MEDESIIE (Salles, 2005), propose des modèles pour décrire l'entreprise, sa stratégie, son environnement, son besoin relevant de l'IE et ses produits et services.

- Un modèle d'entreprise, est décrit selon ses différentes fonctions (productive, économique (lien avec le marché), financière et d'innovation (système d'information)). Chaque fonction est elle-même composée d'un ensemble de sous-fonctions de management.
- Un modèle de stratégie, est représenté par un ensemble de choix stratégiques structurels et d'axes de développement (recherche d'indépendance, croissance de ses activités, augmentation de ses profits, etc.) ;
- Un modèle d'environnement, est décrit dans un premier temps par les fonctions de l'entreprise et des relations qu'elle développe avec son environnement. Dans un second temps, il est décrit par l'environnement exogène à l'entreprise, selon huit grands axes (la géométrie spatiale des marchés, la demande, la technologie, la concurrence, le régime financier, les conditions d'offre, le cadre réglementaire, l'environnement politique et géopolitique).
- Un modèle du besoin, sert de cadre à la collecte du besoin, sa formalisation, son analyse et sa validation. Le besoin exprimé est représenté par un ensemble d'unités de besoin. Le modèle d'unité de besoin est décrit en fonction de trois dimensions : le niveau de pilotage de la décision pour lequel l'unité est exprimée (sa valeur : opérationnelle, tactique ou stratégique), la phase du processus de décision IDC associé à l'unité et son contenu informationnel (identification de sa valeur et de sa fonction).
- Un modèle des produits et services, toute fourniture d'aide à la décision portant sur l'environnement, de l'entreprise qui consiste en un outil de prototypage réalisé en fonction de la satisfaction des besoins recueillis afin d'en définir les coûts, la portée et de pouvoir en évaluer, a priori, les effets.

2.3.2. Les modèles SITE

Le modèle SITE regroupe les différents modèles d'IE proposé par l'équipe de recherche SITE qui est dirigée par A. David et rattaché au laboratoire LORIA. La problématique de l'équipe est d'étudier la modélisation et le développement de systèmes d'informations stratégiques dans le cadre de l'IE.

Ces modèles se basent sur la mise en relation de trois espaces : l'espace des problèmes décisionnels, l'espace des problèmes informationnels et l'espace des médiations qui les relient par l'intermédiaire des échanges entre les deux types d'acteurs (décideur et veilleur).

Les travaux de l'équipe ont en commun la prise en compte de l'utilisateur dans les systèmes d'information. Ils proposent des modèles permettant de définir les différents acteurs, leurs interactions et leurs positionnements dans le processus d'IE.

Nous retenons trois modèles à savoir : EquAte (Explore Query Analyse Annote), MEPD (Modèle pour l'explication d'un problème décisionnel), WISP (Model for Information Retrieval query Annotations Based on Expression Levels).

Le modèle EquAte (David et Thierry, 2002), représente une situation de recherche d'information qui implique les phases cognitives suivantes :

- L'exploration du monde d'information
- L'interrogation de la base d'information

- L'analyse de la base d'information
- L'annotation fondée sur différentes préférences

Le modèle MEPD (Bouaka, 2004), consiste à définir les différentes facettes d'un problème décisionnel. Il se base sur :

- La modélisation du décideur, par son identité, les traits de sa personnalité, son style cognitif et son expérience,
- La modélisation de l'environnement, par l'environnement immédiat (les clients, les fournisseurs, les concurrents) et global (social, économique, politique, etc.),
- La modélisation de l'organisation, par l'environnement, son signal, les hypothèses que le décideur peut déduire de la détection des signaux recueillis.

Le modèle WISP (Kislin, 2007), est associé au modèle MEPD dans lequel il s'intègre en lui empruntant quelques paramètres. Le modèle WISP est un modèle tridimensionnel, multi-facettes, qui intègre la notion de point de vue :

- Une dimension analytique qui englobe la compréhension demande-enjeu-contexte (DEC), la définition des indicateurs informationnels et toutes les opérations d'analyse et de création de connaissances qui peuvent être réalisées par l'étude des éléments mémorisés,
- Une dimension méthodologique qui est constituée, à un premier niveau, par les compétences de traduction du problème décisionnel en problème informationnel et à un second niveau par les stratégies de recherche par lesquelles l'information est identifiée et les connaissances sont acquises.
- Une dimension opérationnelle qui correspond à la sélection des plans d'action et à la mise en place des différentes étapes de résolution de la méthodologie associée au modèle WISP.

2.3.3. Le modèle XPlor

En se basant sur l'aspect modélisation du rôle / utilisateurs dans un système d'intelligence économique proposé dans les modèles Equate, MEDP et WISP et le modèle du besoin présenté dans MEDESIIE nous proposons notre modèle d'analyse et de conception d'un système d'intelligence économique. Le modèle proposé se base sur trois modèles :

- Un modèle d'analyse exploratoire multidimensionnelle, que nous nommons Xplor, proposé par les membres de notre équipe (Sosson et al., 2001)(Ghallamalah, 2009) qui repose sur l'extraction de connaissances à partir des données textuelles par l'analyse des données relationnelles et de leur évolution.
- Un modèle d'indicateur qui a pour objectif la définition d'un ensemble d'indicateurs de calcul, selon les besoins d'analyse, à partir de la représentation multidimensionnelle des documents.
- Un modèle de sécurisation qui repose sur le modèle Rôle / Utilisateurs, pour sécuriser les informations et les traitements effectués sur notre système d'intelligence économique.

➤ Analyse des besoins

La première étape de ce modèle est l'analyse des besoins qui consiste à traduire le problème décisionnel en un problème informationnel. Elle est établie à partir du problème informationnel exposé par les

décideurs. L'objectif de cette activité est de décrire la démarche de pilotage du processus d'analyse. Nous définissons cette activité selon le principe de 5W-1H (Jakobiak, 2004b) (Ghallamalah, 2009) (El-Haddadi et al., 2011): What, Why, Who, When, Where, How (Quoi, Quand, Où, Qui, Comment, Pourquoi).

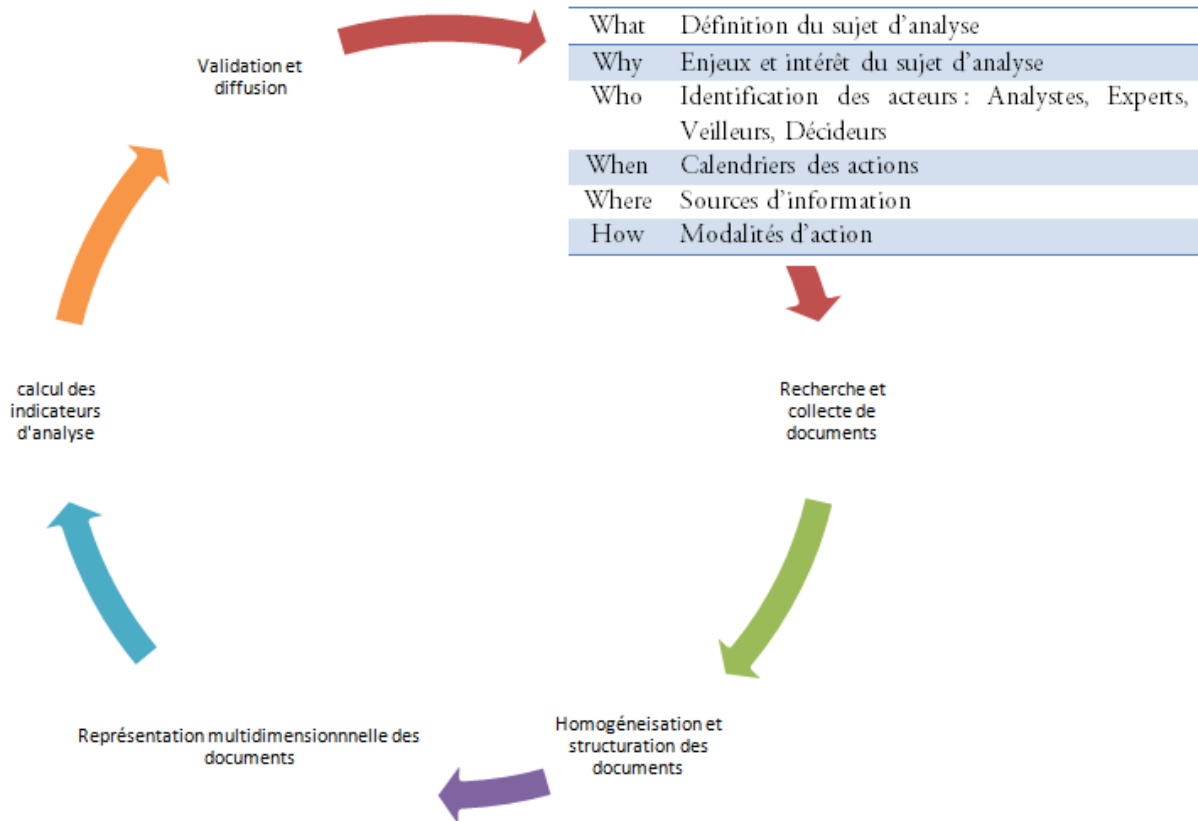


Figure 16. Le processus du modèle Xplor (Ghallamalah, 2009)

Ce principe nous permet de décrire le besoin informationnel posé et d'orienter l'analyse. Toutes les sources d'information identifiées, qu'elles soient formelles ou informelles, sont susceptibles de contenir des informations utiles au sujet de l'analyse. Calculer les indicateurs pour synthétiser et interpréter l'environnement informationnel du sujet de l'analyse.

A ce niveau, il faut introduire les différents indicateurs associés au thème. Chaque indicateur est analysé de manière à identifier ses attributs ciblés, sa granularité, ses valeurs et ses relations. L'objectif de cette décomposition, est d'éclaircir les objets manipulés lors du traitement.

Définition. Un besoin informationnel défini par $\langle S^A, Th^A, Ind^A, Att^A, Val^A \rangle$

S^A : représente le contexte général du besoin pour l'analyse A.

$Th^A = \langle th^A_1, th^A_2, \dots, th^A_m \rangle$, représente les thèmes fixés pour le sujet S_A .

$Ind^A = \{ \langle th^A_i, \langle ind^A_{i1}, ind^A_{i2}, \dots, ind^A_{in} \rangle \rangle \}$, représente les indicateurs associés à chaque thème.

$Att^A = \{ \langle ind^A_{ij}, \langle att^A_{ij1}, att^A_{ij2}, \dots, att^A_{ijp} \rangle \rangle \}$, représente les attributs identifiés pour les indicateurs.

$Val^A = \{ \langle att^A_{ijk}, \langle val^A_{ijk1}, val^A_{ijk2}, \dots, val^A_{ijkq} \rangle \rangle \}$, représente les valeurs des

attributs spécifiés pour chaque type d'acteurs.

Dans la figure suivante, nous définissons une hiérarchie de concepts associés à la décomposition des indicateurs.

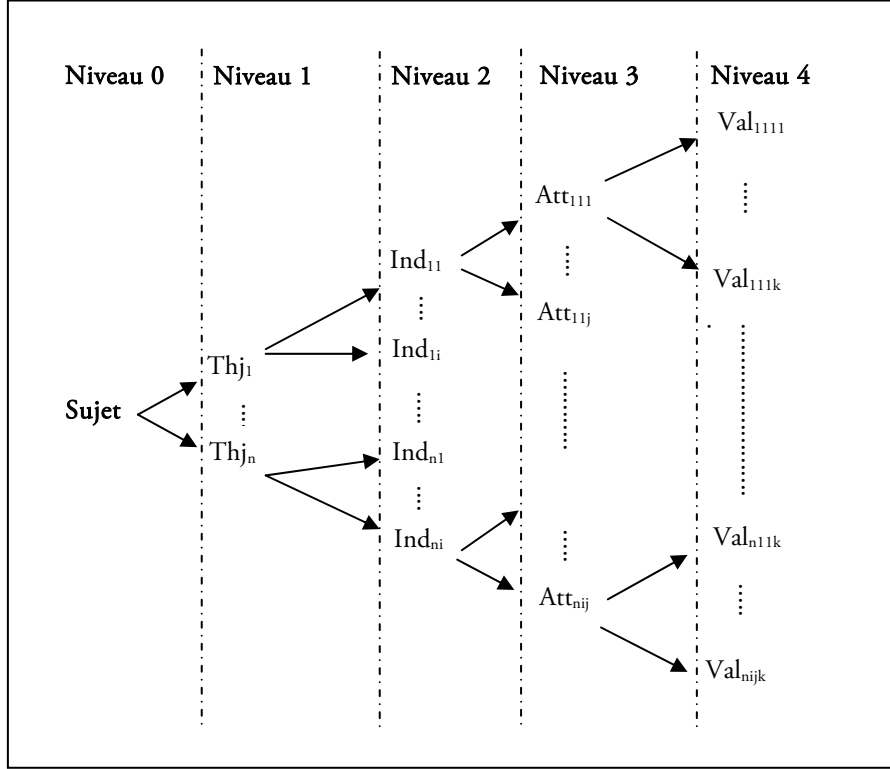


Figure 17. Hiérarchie de concept associés à la spécification des besoins (Ghallamalah, 2009)

Exemple : Le besoin est défini comme suit :

$$B_s = \langle S_A^s, Th_A^s, Ind_A^s, Att_A^s, Val_A^s \rangle$$

$S_A^s = \langle \text{évaluation de la recherche scientifique marocaine} \rangle$

$Obj_A^s = \langle \text{'Les collaborations'}, \text{'La production scientifique'}, \text{'Les thématiques de recherches'} \rangle$

$Ind_A^s = \langle \langle \text{'Les collaborations'}, \langle \text{'Les collaborations entre auteurs'}, \text{'Les collaborations entre organismes'}, \text{'Les collaborations entre pays'} \rangle \rangle, \langle \text{'La production scientifique'}, \langle \text{'Le nombre de publications'}, \text{'L'évolution du nombre de publications'} \rangle \rangle, \langle \text{'Les thématiques de recherche'}, \langle \text{'L'évolution des thématiques de recherche des auteurs'} \rangle \rangle \rangle$

$Att_A^s = \langle \langle \text{'Les collaborations entre auteurs'}, \langle \text{Auteur} \rangle \rangle, \langle \text{'Les collaborations entre organismes'}, \langle \text{Auteur}, \text{Organismes} \rangle \rangle, \langle \text{'Les collaborations entre pays'}, \langle \text{Auteur}, \text{Pays} \rangle \rangle, \langle \text{'Le nombre de publications'}, \langle \text{Auteur} \rangle \rangle, \langle \text{'L'évolution du nombre de publications'}, \langle \text{Auteur}, \text{Date} \rangle \rangle, \langle \text{'L'évolution des thématiques de recherche des auteurs'}, \langle \text{Auteur}, \text{Descripteurs}, \text{Date} \rangle \rangle \rangle$

$Val_A^s = \langle \langle \text{Auteur}, \text{Descripteurs}, \text{Date} \rangle, \langle 2008, 2009, 2010, 2011 \rangle \rangle$

La hiérarchie de concepts associés est comme suit :

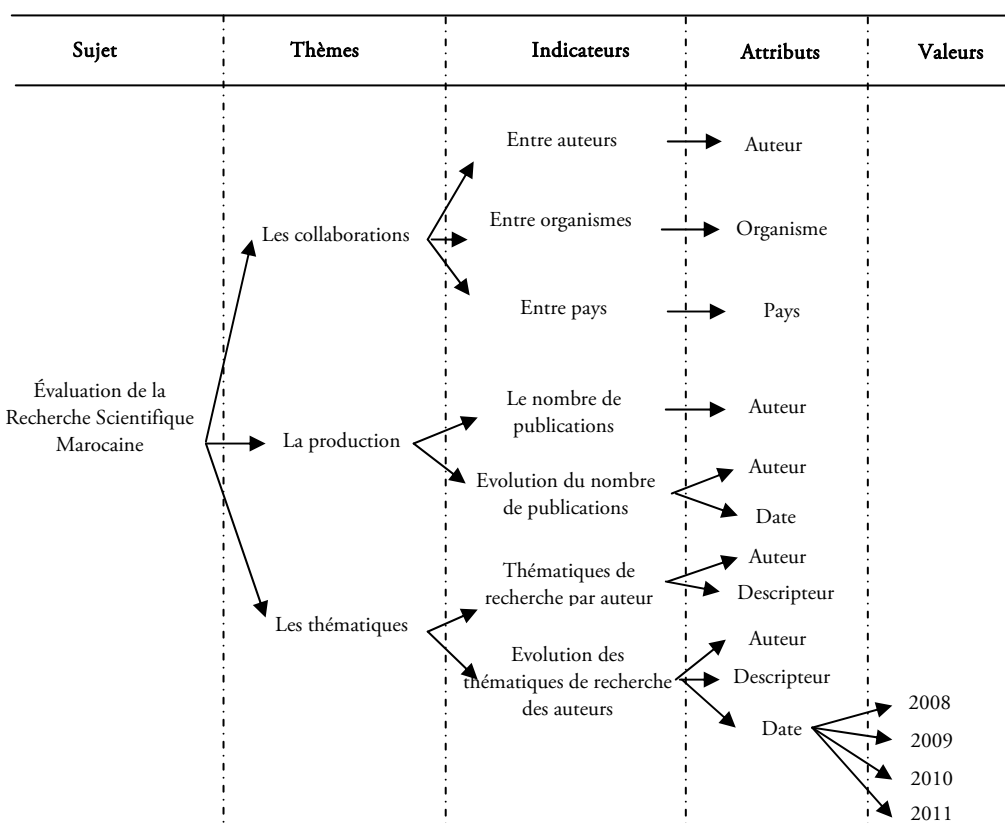


Figure 18. Exemple de la hiérarchie de concepts associés au sujet de la recherche scientifique marocaine

Une fois que l'analyse des besoins, la première étape de notre modèle, est validée, la deuxième étape qui consiste à rechercher et collecter des informations peut débuter.

➤ Recherche et collecte des informations

Cette étape permet de se focaliser, suivant des critères prédéfinis, sur des données supposées à la fois « interprétables » et à fort potentiel informatif. De plus, la préparation des données consiste, dans un premier temps, à les sélectionner en accord avec les objectifs que l'on s'impose, en ayant recours aux techniques de recherche d'information, (Büttcher et al., 2010) (Croft et al., 2011). Ce processus (Saltan et McGill, 1984) cherche à mettre en correspondance une collection de documents et le besoin de l'utilisateur (Maniez et Grolier, 1991), traduit sous la forme d'une requête (Kleinberg, 1999) à travers un système d'information. Ce dernier est composé d'un module d'indexation automatique ou semi-automatique, d'un module d'appariement document-requête et éventuellement d'un module de reformulation de la requête.

Différents modèles sont utilisés dans les moteurs de recherche pour l'appariement entre la requête et le document, tels que les modèles probabilistes (Sparck-Jones, 2000), booléens (Salton, 1971), (Roberston, 1977), flous (Paice, 1984), connexionnistes (Mothe, 1994), (Boughanem et al. 2000), flexibles (Sauvagnat, 2005), etc.

Les sources n'étant pas équivalentes, il s'avère ainsi intéressant de les combiner pour avoir une information plus exhaustive et moins biaisée. Deux bases qui indexent le même journal ne choisissent pas les mêmes articles. L'une va proposer le résumé et de bons mots-clés, l'autre choisira les citations et toutes les adresses. Faut-il les dédoubler ou doit-on conserver toutes les informations complémentaires ? La question reste posée. Le choix est toujours laissé aux utilisateurs, mais nous préconisons de garder

toutes les versions disponibles puisqu'une information redondante est moins gênante qu'une information partielle ou manquante.

D'un autre point de vue, compte tenu du fait que les sources sont complémentaires, elles permettent d'une part d'aboutir à un corpus multi-formats représentatif du domaine étudié, et d'autre part, d'avoir une vision stratégique, d'un sujet (produit ou procédé), à la fois scientifique (les articles scientifiques) et technologique (les brevets). Cette hétérogénéité des sources permet de :

- Suivre les évolutions de la recherche au service de la technique ;
- mettre en évidence les signaux faibles (par exemple les nouveaux centres d'intérêt) ;
- mettre en évidence les nouveaux acteurs ;
- mettre en évidence les meilleures pratiques.

➤ **Homogénéisation et structuration des données**

Un corpus représentatifs du domaine étudié se compose des sources souvent hétérogènes mais complémentaires. . Chaque source est décrit par des méta-données de premier niveau (Md1) qui analysent finement sa structure, décomposent son contenu en unités sémantiques (Dkaki et al., 1997) (Dkaki et al., 2000) (Chrisment et al., 2006) (Chrisment et al., 2008) et définissent les modes d'extraction les mieux appropriés à chaque cas. Des méta-données de second niveau (Md2) permettent de mettre en correspondance les contenus de même type (mais pas de même format) issus de chaque source et servent d'interface unique avec les techniques d'identification, d'extraction, de dénombrement et de croisement des informations utiles.

Ensuite les informations sont croisées (Mothe, 2000) (Dousset, 2003), soit au sein d'un même type (associations) soit entre deux types différents afin de réaliser une première étude statique de leurs influences mutuelles.

➤ **Traitement des structures de données**

Une première contribution de notre travail (Dousset, 2003) (El Haddadi et al., 2010) a consisté à traiter les données dans leurs formes natives à la différence de la majorité de logiciels qui imposent aux utilisateurs un format de données propriétaire. Cette restriction engendre une perte de temps considérable de restructuration de données et nécessite parfois une expertise élevée en informatique. Le format natif offre plusieurs avantages à savoir une meilleure réactivité, une mise à jour du corpus facilitée et une préservation de l'ensemble des informations. Néanmoins, afin de s'adapter à quasiment toutes les structures, il est nécessaire d'utiliser les méta-données qui sont des outils de description des formats dont le principe est de :

- Trouver une technique permettant de différencier les documents les uns des autres (ou les unités textuelles) ;
- Déterminer les balises des champs sémantiques présentes dans la base et leur donner un nom et un sigle standard ;
- Définir leur utilité et leur priorité ;
- Déterminer des techniques judicieuses de découpage permettant d'extraire chaque type d'information.

Notons par ailleurs que plus de 90% des cas rencontrés peuvent être traités sans aucun reformatage. En effet, il est possible de travailler simultanément sur plusieurs formats et sur plusieurs sources en élaborant des règles de correspondance entre les champs utiles à l'aide de méta-données de second niveau. Celles-ci permettent à la fois d'orchestrer la synchronisation de tous les formats et de les interfacer de façon unique avec les outils d'extraction sémantique. Chaque source a un format spécifique qui lui-même a un descripteur spécifique (méta-données de premier niveau). Une collection de formats est gérée par un descripteur générique (méta-données de second niveau : le chef d'orchestre).

➤ Définitions des objets dans les méta-données

Il existe différents types d'objets dans les méta-données que nous décrivons ci-dessous.

Bannière de synchronisation. Elle sert à détecter le changement de document à l'intérieur d'un fichier séquentiel contenant un ensemble de documents au même format. A titre d'exemple, si une ligne vide sépare deux documents qui ne comportent pas eux mêmes de saut de ligne, cette séparation peut éventuellement servir de bannière de synchronisation « Vide ».

Nom standard d'un champ. C'est le nom donné, dans la langue de l'utilisateur, à un type d'information balisée. Il doit être le même dans tous les descripteurs compatibles avec les analyses multi-bases (exemple : Titre, Résumé, Auteurs, Dates, Pays, etc.).

Sigle standard d'un champ. Ce sigle permet de nommer de façon standard (si possible avec deux lettres) les répertoires, les dictionnaires et matrices (PA.ind, PA.Filtre, PA.Syn, MC-PA, AU-AU-DT,...). Il s'avère très utile pour normaliser la structure des analyses et leur appliquer des traitements automatiques (diffusion au format html, compilation dans une base de données pour alimenter un portail, etc.).

Bannière du champ. C'est le nom pris par un champ de la base au moment de sa collecte (long ou court) ou à l'issue d'un éventuel reformatage (html) : TI:, Title:, TI-, Titre :, etc.

Drapeau d'utilisation. Il sert à masquer certains champs inutiles à l'analyse ou temporairement peu utiles. Les interfaces utilisateur des fonctions de la plate-forme « Tétralogie » (Dousset, 2003) n'affichent ainsi que les champs requis.

Liste des séparateurs. Elle définit l'ensemble des chaînes de caractères qui servent à séparer les unités sémantiques à extraire de chaque champ. Certains jokers sont prévus : le saut de ligne « \n », l'espace « b », ... Un opérateur permet aussi de ne conserver, à l'extraction, que l'élément de rang i : « ORDi », « ORD0 » pour le dernier.

➤ Descripteurs de formats spécifiques

Pour chaque base structurée ou semi-structurée, il convient de définir son descripteur de format spécifique qui permet de l'interfacer définitivement avec notre plate-forme de traitement de l'information « Tétralogie ». Voici un exemple partiel de descripteurs de formats :

Article Numéro 9

Notice 9

TI: Implementation of transformable business structures with factory

AU: Dworski, E.A.1; Aldinger E.

AF: lInt. Perform

SO: Industrie Man

Publisher: GITO-V

LN: German

AB: This article

of decision makin

elements of class

Hereby, it provid

production perfor

of the production

KWIn: competitive

small-to-medium e

KWPh: transformable factory cockpit; decision making; business intel

factories; digital factories; small-and-medium sized companies; produ

CA: C7480 Production engineering computing

Bannière de synchronisation

Notice

descripteurs des champs de la base INSPEC-2010

# nom	abrev	champ	visible	Separateurs
AU	AU	AU:	True	"
Re	Re	Re:	True	"
AF	AF	AF:	True	"
SO	SO	SO:	True	"
Country	PA	AF:	True	",";
Date	DP	SO:	True	b",";
CounPub	PAPub	SO:	True	b"
LN	LN	LN:	False	"
Resume	AB	AB:	True	s,"s."s-"s;"s?"sb"s)"s]"'","."-";
KWInspec	MC	KWIn:	True	;"
KWPh	MC1	KWPh:	True	;"
CA	CA	CA:	True	b"
TCA	TCA	TCA:	True	."

Non standardisé du champ

Figure 19. Le descripteur de la base Pubmed format Medline

➤ Exemple de descripteur générique

Tel qu'il a été décrit dans les sections précédentes, pour réaliser une étude simultanée sur différentes bases, il est nécessaire de définir un descripteur générique en donnant pour chaque champ utile un nom standard, un sigle standard et un drapeau d'utilisation (Dousset, 2003) (El Haddadi et al., 2010). Les autres éléments (bannière de synchronisation, nom effectif du champ dans chaque base et liste des séparateurs) sont tour à tour empruntés aux méta-données de chaque base. Voici un descripteur générique compatible simultanément avec trois bases (Pub-Med, Pascal et Science Direct) :

TY ← Non utilisé

descripteurs des champs des meta donnees

# nom	abrev	champ	visible	Separateurs
Titre	TI	T1:	True	"
Resume	AB	AB:	True	"
Auteurs	AU	AU:	True	"
Organisme	OR	AD:	True	"
Ville	VI	AD:	True	"
Pays	PA	AD:	True	"
Date	DP	PY:	True	"
Classifs	CC	CC:	True	"
Mots-cles	MC	KW:	True	"
Index-E	IE	CE:	True	"
Journal	JN	JO:	True	"
E-mail	EM	EM:	True	"
FIN	FIN	FIN	FIN	"

Non utilisés

Nom standardisé du champ

Figure 20. Descripteur générique de trois bases : Pub-Med, Pascal et Science Direct

- La bannière de synchronisation est dans cet exemple inutile puisqu'elle est spécifique à chaque base. Elle est toutefois conservée par soucis d'unicité du format des descripteurs.
- La première colonne permet de désigner explicitement chaque champ utile. Son nom standard permet de faire coïncider les champs de même nature dans les différentes bases simultanément analysées.
- La seconde colonne permet de nommer les répertoires et matrices de croisement de façon standard (habituellement deux lettres en majuscule).
- La troisième colonne n'est pas utilisée puisqu'elle est remplacée par celle qui est spécifique à chaque base (liste des noms exacts des bannières des champs de cette base).
- La quatrième colonne sert à rendre actif les champs utiles pour toutes les bases.
- La cinquième colonne est remplacée par celle provenant du descripteur de chaque base car il s'agit de définir les séparateurs spécifiques à chaque champ.

L'extraction de la terminologie venant des trois bases est alors pilotée par le descripteur générique. En revanche, les spécificités de chaque base (séparation des documents, noms des champs et techniques de découpage) sont tour à tour prises en charge par le descripteur associé à chaque base. Nous avons ainsi pu fusionner, sans aucun reformatage, les contenus de tous les champs utiles de trois bases bibliographiques issues du Web invisible.



Figure 21. Principe des méta-données de premier et second niveau

Nous donnons, ci-dessous, une illustration générale de la séquence des opérations de traitement des structures de données.

Cette approche propose de traiter les informations dans leurs formats natifs dans le but d'avoir une meilleure réactivité et une facilité de mise à jour du corpus. Une telle démarche permet d'optimiser le processus d'intelligence économique et de surveiller tout type d'information qui peut être traité sans aucun reformatage préalable dans plus de 90% des cas étudiés.

L'objectif final est d'obtenir une vue globale et unifiée des sources collectées qui sera utilisé tout au long du processus d'analyse. Cette vue doit répondre aux objectifs suivants :

- Une vue homogène, partagée par les différentes données quelles que soient leurs sources.
- Une vue globale et réduite de l'information pour faciliter et accélérer le traitement sans passer par des reformatages.
- Une vue qui facilite l'analyse de tout type d'information et restitue l'information dans des délais très courts tels que l'exige l'intelligence économique.

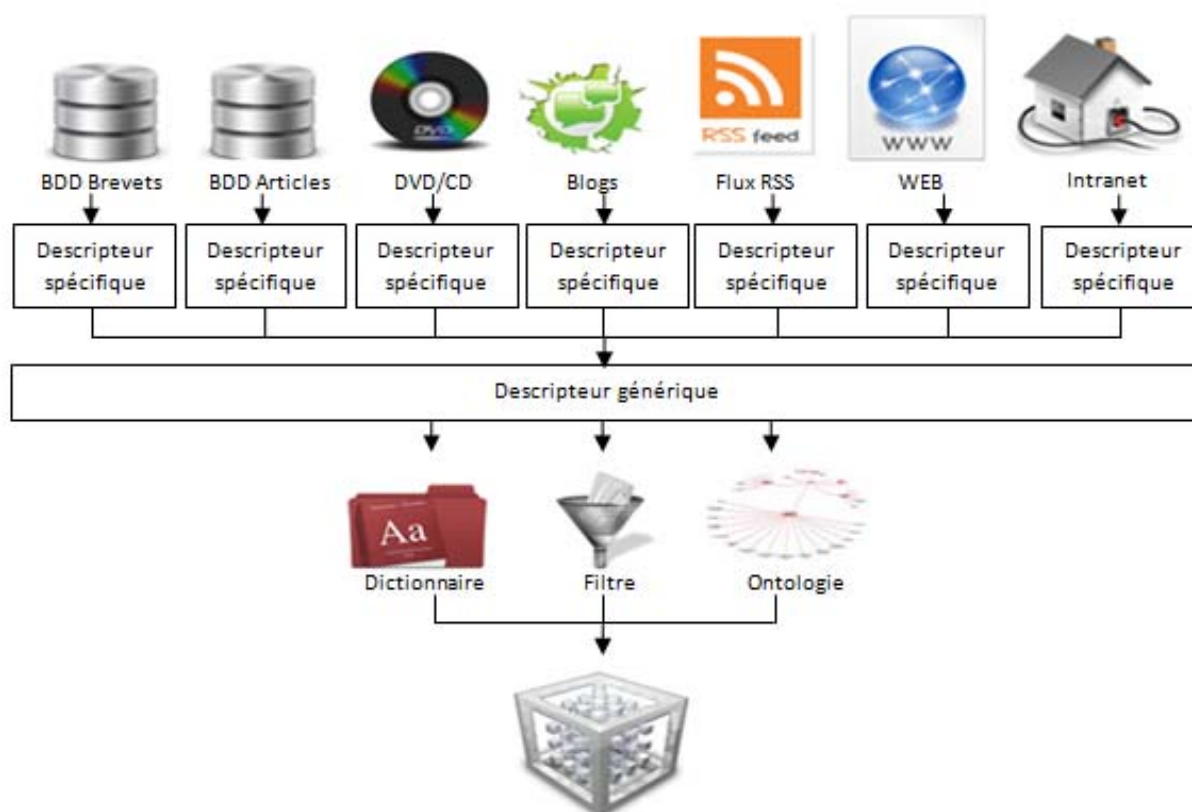


Figure 22. Homogénéisation des sources d'information

Cette vue unifiée associée au corpus ciblé correspond à une représentation logique, structurée, prédéfinie de l'ensemble de ses collections sous forme d'un entrepôt de donnée stratégique qui fera l'objet du troisième chapitre de ce manuscrit.

Maintenant que notre modèle d'intelligence économique est défini, nous allons présenter les systèmes d'informations adaptés à la démarche d'intelligence économique. Et nous présenterons notre SIE Xplor EveryWhere.

2.4. Les systèmes d'intelligence économique

Un système d'intelligence économique, se présente comme un compromis de S-IS et SI-S. C'est un système d'information qui va plus loin que les systèmes d'informations cités précédemment mais qui comprend ces derniers comme une étape pour :

- la diffusion,
- la disponibilité d'interprétation,

- le stockage,
- la sécurité de l'ensemble,
- l'aspect collaboratif dans la prise de décision.

Le système d'intelligence économique induit donc une culture collective de l'information accompagnée d'actions offensives, avec le souci constant de protéger le capital informationnel de l'organisation.

On trouve dans la littérature différentes propositions de conception des SIE. Pour Patrick Romagni et Valérie Wild, la définition d'un système d'information adapté à la démarche d'IE est la suivante (Romagni et Wild, 1998):

Ensemble organisé de procédures permettant, à tout moment, de donner aux décideurs une représentation de la place de l'entreprise dans son environnement et sur son marché. Il produit de l'information pour assister les individus dans les fonctions d'exécution, de gestion et de prise de décision.

Il doit :

- Faciliter les décisions, en automatisant un certain nombre d'actions ou en mettant à disposition des décideurs des éléments nécessaires à la prise de décision,
- Coordonner les actions pour le traitement de l'information,
- Stocker de manière durable et stable les informations,
- Traiter les données, c'est la fonction la plus importante du système d'information car il entraîne la création d'informations directement utilisables par les décideurs.

La plupart des autres propositions (cf. (Xiaobo et lian, 2010) (Olusoji et Amos, 2010) (Moinet et Frion, 2009) (Guoliang et al. 2009) (Weidong, 2008) (Ginsburg et al., 2009) (Li et Dong, 2010) (Chen et al., 2008) (Chen et al., 2009) (Radum, 2006) (Liu et Xia, 2009) (Tan et al., 2004) (Li et al, 2007)) sont d'accord au niveau de la conception et des composants, avec Patrick Romagni et Valérie Wild. Cette conception permet de concevoir des systèmes d'information permettant aux organisations une meilleure gestion de l'information et d'offrir une base de coordination des actions entre les différents acteurs. Cette coordination transversale est soutenue par les motivations suivantes :

- Les objectifs d'une démarche d'IE sont inter-reliés. Ils ne peuvent être traités séparément.
- Le besoin de partage d'informations entre les différents acteurs,
- Le partage des connaissances acquises au cours d'une démarche,
- L'organisation de l'entreprise de manière fonctionnelle est transversale.

Cette conception doit évoluer d'une architecture verticale vers une architecture transversale qui permet une gestion globale de l'information.

L'architecture transversale est basée sur une architecture modulaire et évolutive structurée autour des projets de l'entreprise, l'infrastructure globale d'information va permettre selon (Cigref, 2004) :

- Une réduction des nombreuses coordinations verticales en diminuant les échelons hiérarchiques,
- Une meilleure surveillance de l'environnement des unités décentralisées,
- Un décloisonnement par une communication transversale,

- Des relations axées sur la complémentarité des métiers,
- Une meilleure adaptation à la dynamique du marché.

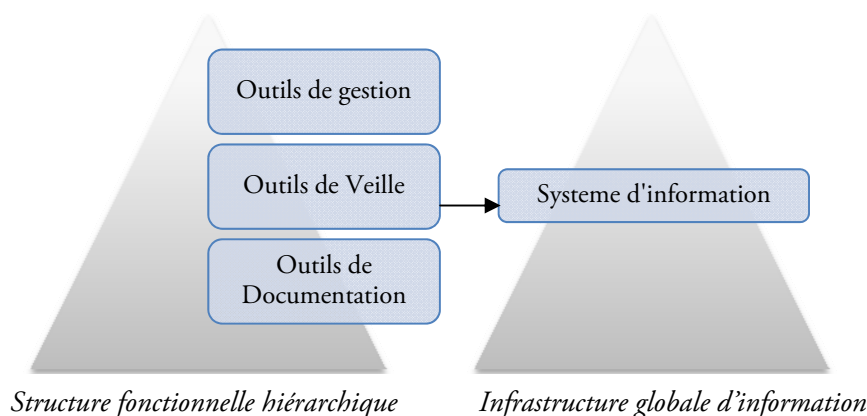


Figure 23. Infrastructure globale d'information (Cigref, 2004)

Cette analyse confirme l'émergence des systèmes d'information stratégique. La démarche d'IE s'inscrit alors dans le schéma global du système d'information stratégique de l'entreprise.

D'après le rapport du Cigref « *Intelligence économique et stratégique* », le portail de gestion de l'information est actuellement l'outil le mieux adapté pour implanter le concept et la culture de l'IE à travers un réseau. Ce portail est bâti autour d'une solution logicielle dite d'IE. L'avantage de ce type de système d'information s'appuie sur les points suivants :

Gestion du projet de veille

- Favoriser la prise de conscience collective, chacun comprenant les enjeux de sa contribution au dispositif.
- Porter à la connaissance de tous, les thèmes non confidentiels de veille et leur actualisation.
- Donner une identification des acteurs et de leur rôle en matière de projets et de gestion de l'information.

Partage de l'information

- Améliorer la qualité de l'information et créer de la valeur ajoutée (échange de points de vue et consolidations).
- Améliorer la circulation en offrant la possibilité d'assurer un feed-back.
- Créer des flux plutôt que des stocks d'informations vers des destinataires susceptibles d'en tirer profit.
- Favoriser la participation du plus grand nombre, avec possibilité de communication descendante, montante et transversale.
- Retenir des formats de diffusion ou de mémorisation adaptés à chaque type d'information et de destinataire.
- Avoir une traçabilité des demandes d'information et des réponses.

Les fonctionnalités du portail doivent offrir aussi :

Interface personnalisée

- Personnalisation de l'interface (information du jour, par thème, moteur de recherche, etc.) accès rapide à l'information.

Collecte plus précise et ciblée

- Moteur de recherche, sur internet, web invisible, etc.
- Remontée rapide des sources informelles (clients, fournisseurs, commerciaux, réseaux personnels, etc) via certains outils (assistants numériques, courrier électronique, téléphone mobile, internet, etc.).
- Création d'annuaires de liens Internet pour l'ensemble des membres d'un réseau.

Traitement, analyse, stockage

- Représentation graphique des informations.
- Archivage spécifique des informations utiles aux décideurs en fonction de leur durée de vie (validité).

Diffusion

- Une représentation graphique des informations.
- Un archivage spécifique des informations utiles aux décideurs en fonction de leur durée de vie (validité).
- De générer des rapports à la demande pour les décideurs.
- De gérer des groupes de discussion avec droits d'accès, échange et partage de l'information ; chacun profite des informations récoltées par les autres.
- Un pilotage plus pointu, une gestion des feed-back.
- Un ciblage par rapport aux axes de développement de l'entreprise, un suivi des acteurs, des thèmes les plus actifs et des sources surveillées.
- Un pilotage en temps réel de l'activité de veille, grâce à des tableaux de bords permettant de contrôler et réorienter l'activité en fonction des objectifs.

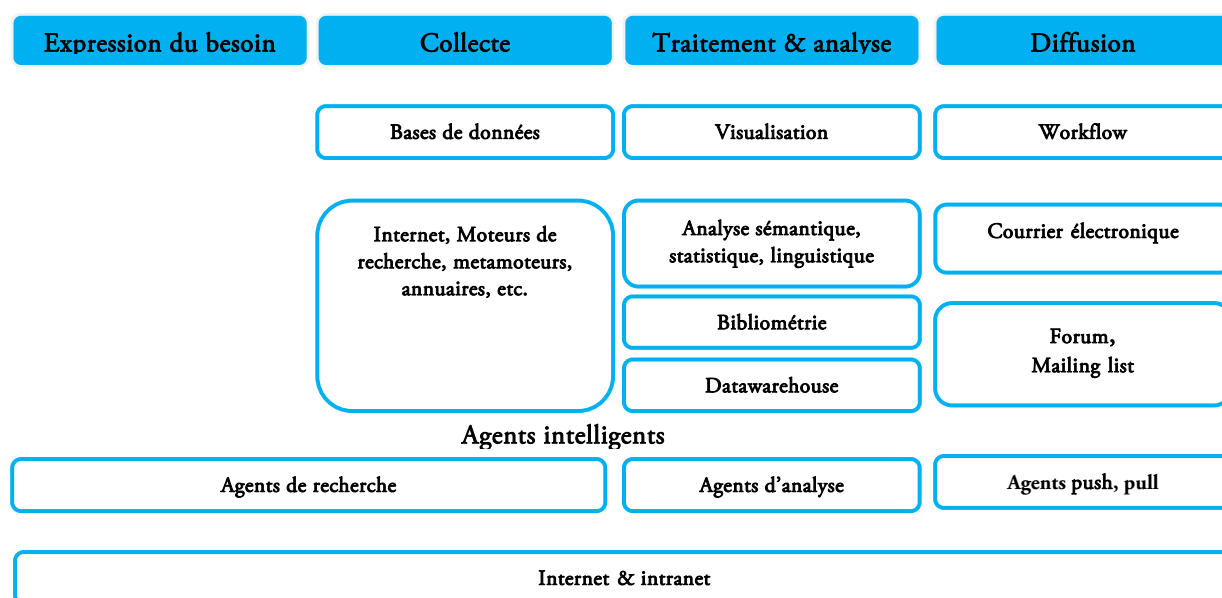


Figure 24. Le couplage IE, Intelligence Collective et Mining (Ghallamalah, 2009)

Le couplage des besoins identifiés dans une démarche d'IE avec les différentes techniques existantes de travail collaboratif ou de Business Intelligence tels que : workflow, groupware, datawarehouse, data mining, text mining, gestion électronique des documents, visualisation, etc. permet d'optimiser chaque étape du cycle de l'IE. La figure suivante résume l'ensemble de ces techniques pour chaque étape de la démarche d'IE. La phase de collecte se fait généralement par l'utilisation des bases de données, Internet, agents de recherche, moteurs de recherche, l'étape de traitement et d'analyse peut être supportée par les outils de visualisation, la bibliométrie, l'analyse statistique, les entrepôts de données, etc. l'étape de diffusion peut s'appuyer sur les agents push-pull, les courriers électroniques, etc.

2.5. Le système d'intelligence économique “XPlor EveryWhere”

Tétralogie est un outil bien adapté aux analyses macroscopiques, il permet de dégager les signaux forts, les signaux faibles et les tendances à partir d'un corpus de documents collectés sur un sujet précis. L'information élaborée qui en est issue représente une synthèse obtenue par diverses méthodes d'analyse de données et diffusée via des visualisations graphiques (carte géostratégique, graphe évolutif, ACP, AFC, CAH, etc.). Mais à l'issue des nombreuses analyses stratégiques que nous avons déjà réalisées avec ce logiciel, il est apparu que les utilisateurs finaux des analyses produites veulent, en complément de l'aspect global (vue macroscopique), des analyses microscopiques sur des éléments déjà identifiés (concurrence, marchés, nouveaux produits ou procédés, partenaires potentiels, ...) ou afin d'en découvrir d'autres. A posteriori, de nombreux experts ou décideurs sont demandeurs de plus de détails sur les éléments traditionnels de leur environnement. Notamment, pour tout ce qui concerne leur vocabulaire spécifique, les acteurs qu'ils côtoient, les marchés qu'ils convoitent, les alliances qu'ils projettent.

Nous proposons donc de continuer à valider le modèle proposé et de compléter les analyses macroscopiques par un système de navigation en ligne au cœur de l'information relationnelle obtenue par des recoupements statistiques, des classifications ou des analyses multidimensionnelles. Le but est de privilégier l'extraction d'information en fonction du contexte général et non exclusivement par décryptage du contenu de quelques documents pris séparément. Il devient ainsi possible de retrouver, à partir d'un élément connu (acteur, mot clé), toute ou partie de l'information qui lui est connexe (équipes, collaborations, concepts, émergences, mots associés, ...) et ce par l'utilisation de nombreux opérateurs d'association ou de filtrage et de fonctions de reporting pertinentes.

Le prototype développé permet d'effectuer des analyses stratégiques sur des corpus d'information textuelle issus des sources les plus diverses comme les bases en ligne (publications scientifiques, brevets, portails, annuaires), les Cd, le Web visible et invisible, les news, la presse, les traces de connexions aux sites, les bases internes, les flux RSS, les réseaux sociaux ... et donne la possibilité aux décideurs de mener par eux même leurs investigations sans la participation d'un analyste confirmé ou d'un expert.

Ses applications sont très diverses :

- identification des thématiques et des acteurs du domaine,
- mise en évidence des stratégies de développement et de coopération, ...
- proposition de scénarii d'évolution technologique (innovation),
- extraction des signaux faibles.
- Consulter des informations à jour, car nous accédons à notre serveur de base de donnée stratégique en temps réel, lui-même alimenté quotidiennement par des veilleurs.

- Faire remonter des informations « terrain » lors de salons, de visites en clientèle ou à l'issue de réunions.
- Demander des renseignements spécifiques en urgence qui seront mis en ligne par les veilleurs.

2.5.1. Architecture service de XPlor EveryWhere

Notre approche pour la conception du système XPlor Every Where repose sur une architecture service à quatre niveaux (figure 25):

➤ Le service « Sources et traitements »

Ce service permet d'alimenter l'entrepôt de donnée stratégique, à partir des corpus de données textuelles. Il permet le passage de la représentation des documents textuels (données qualitatives) sous forme de données quantitatives. Il concerne le traitement de l'hétérogénéité des informations, d'un point de vue :

- contenu sémantique : scientifique, technique, etc.,
- structurel : fortement structuré (brevet) à non structuré (e-mails),
- linguistique (multilinguisme) : chinois, arabe,
- format du support : Word, html, pdf, etc.,
- taille : définition de l'unité d'information à analyser (granularité de l'information).

Pour cette partie du traitement, les techniques employées s'appuient sur les fonctionnalités du système Tétralogie développé antérieurement par les membres de l'équipe SIG-EVI (section 2.3.3 de ce chapitre).

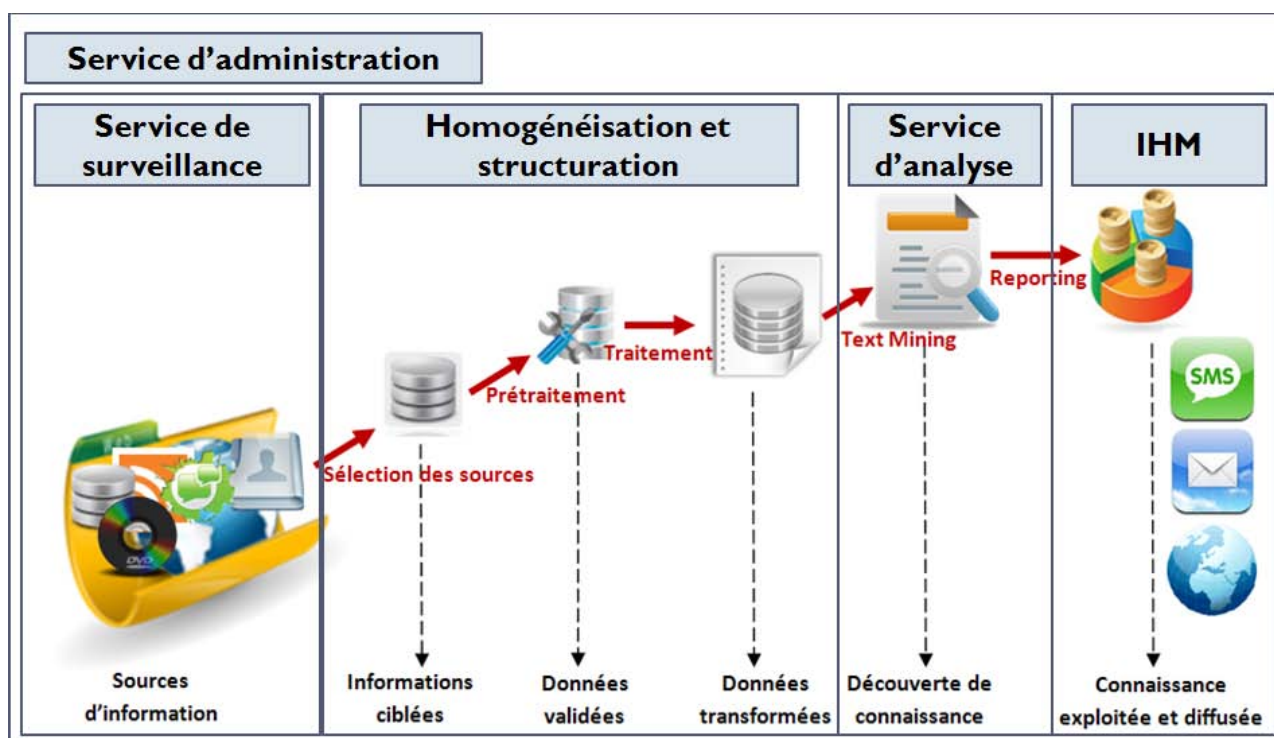


Figure 25. L'architecture service de XPlor

➤ Entrepôt de données stratégique

Ce service, qui fait l'objet du prochain chapitre, est un espace de stockage qui à un premier niveau permet d'offrir une vue unifiée du corpus cible, et l'extraction et le stockage des données sources structurées sous forme d'une représentation multidimensionnelle. Le second niveau porte sur les traitements de création de l'entrepôt de données. Il repose sur la création des matrices spécifiques au type de variable étudiée : le tableau suivant synthétise les différents cas :

Type de variable	Uni-modale	Multi-modale à modalités exclusives	Multi-modales à modalités non exclusives
Uni-modale	- Contingence - Présence/ absence	- Co-occurrences simples - Présence/ absence	- Co-occurrences simples - Co-occurrences multiples - Présence/ absence
Multi-modales à modalités exclusives	- Co-occurrences simples - Présence/ absence	- Co-occurrences simples - Présence/ absence	- Co-occurrences simples - Co-occurrences multiples - Présence/ absence
Multi-modales à modalités non exclusives	- Co-occurrences simples - Co-occurrences multiples - Présence/ absence	- Co-occurrences simples - Co-occurrences multiples - Présence/ absence	- Co-occurrences simples - Co-occurrences multiples - Présence/ absence - Pondération

Tableau 4. Type de matrices possibles selon le type de variables croisées

Notre approche consiste à utiliser les composantes du système Tétralogie pour la création des matrices à deux dimensions et réajuster la structure des résultats obtenus pour faciliter le transfert des données dans l'entrepôt de données du prototype. Cet entrepôt ne contient que les données sous forme relationnelle tel que : N°doc-Auteur, Auteur-Auteur-Date, etc. Cette représentation permet de synthétiser toutes les relations existantes entre les différents attributs constituant les documents à analyser. La base de données est interfacée grâce au système XPlor EveryWhere sur Intranet ou Internet, afin que l'utilisateur puisse lui même mener ses propres investigations.

➤ Service Analyse et restitution

Ce service permet d'effectuer des analyses multidimensionnelles en ligne sur les données issues de l'entrepôt et restituer les résultats à l'utilisateur par des fonctions de la visualisation d'information (reporting). Il comporte les fonctionnalités de navigation et d'analyse en ligne. Ces fonctions sont : administration des données, exploration des données relationnelles et visualisation des résultats. Chaque attribut peut être filtré au moyen de fonctions relationnelles prédéfinies en se servant des liens complexes qu'il possède avec lui-même et les autres attributs de la base. Des statistiques descriptives interactives sont alors disponibles pour chaque extrait (fréquences, équivalences, etc.) ainsi que sur l'évolution de leurs relations. Dans ce service, il s'agit de représenter dans un espace physique sous forme de graphiques une information souvent abstraite (Hinnum et al., 2005) (Loubier, 2009) et ainsi de fournir à l'utilisateur final la possibilité de proposer des explications ou de prendre des décisions (Meliker et al., 2005) (Loubier, 2009).

➤ Service Sécurité et Administration

Ce service s'intéresse d'une part à l'administration du système et d'autre part à la protection des données sensibles qui proviennent des différentes analyses stratégiques et des différents traitements dans un système d'intelligence économique. L'approche sécuritaire proposée intègre le contrôle d'accès dans le cycle de vie de système XPlor Everywhere afin que toutes les opérations soient exécutées uniquement par les utilisateurs autorisés et habilités. Les différents utilisateurs qui interagissent avec notre système, sont représentés dans le diagramme de contexte suivant :

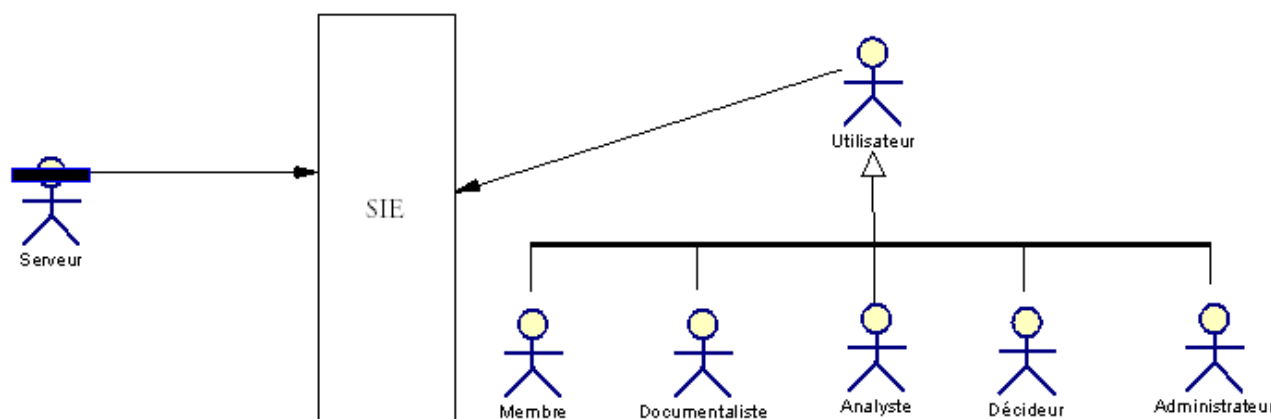


Figure 26. Les différents utilisateurs de XPlor EvryWhere

Pour distinguer visuellement les utilisateurs « humains » des acteurs « non-humains » dans le diagramme de contexte, les diagrammes de cas d'utilisation, et les diagrammes d'interactions, nous présenterons le stéréotype d'acteur « non-humaine » avec le graphique suivant :



Figure 27. Le stéréotype d'acteur « non-humain »

2.5.2. La structure statique de XPlor EvryWhere

Dans le but d'appréhender, d'un point de vue logique, la structure statique du XPlor EvryWhere. Nous définissons la structure des objets composant le système et les liens structurels entre ces objets. Nous réalisons le diagramme de Classes de notre SIE.

Les premières classes candidates identifiées dans cette phase doivent être des concepts connus des utilisateurs du système, ce qu'on appelle couramment des objets métiers. Pendant notre étude de cas, nous avons pu identifier les objets suivants : "Sujet d'analyse", "Catégorie d'analyse", "Liste des analyses", "Corpus", "Membre", "Rôle de membre", "décideur", "Analyste", "Documentaliste".

Exemple « demande d'une étude » :

Scénario : Envoyer une demande

- L'abonné s'authentifie
- Il demande une étude (Brevets / Publications scientifiques)
- Il remplit le formulaire associé à cette étude

- Une vérification de saisie se fait automatiquement par le système
- L'abonné valide la saisie et demande l'enregistrement
- Le système enregistre les informations saisies dans la BDD
- Le système crée une tâche à l'état « en étude » au niveau de la BDD
- Le système associe cette tâche au profil de l'abonné et de l'analyste.

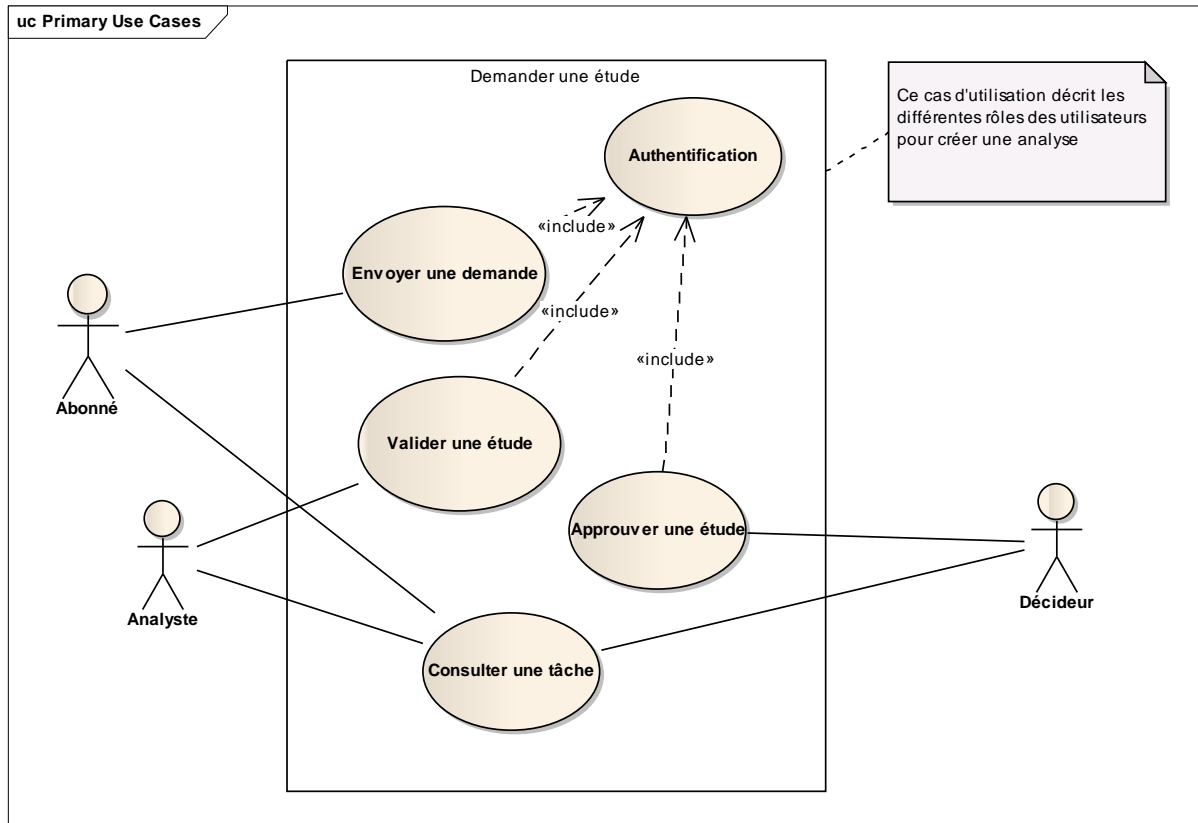


Figure 28. Diagramme de cas d'utilisation « demande d'une étude »

Scénario : Valider une étude

- L'analyste s'authentifie
- Il consulte sa file de travail
- Il choisit l'étude qui a une tâche de l'état « en étude »
- Il consulte les informations associées à cette tâche et valide l'étude
- Il modifie l'état de la tâche à « en analyse »
- Le système enregistre la modification au niveau de la BDD
- Le système associe la tâche au profil du décideur

Scénario : Approuver une étude

- Le décideur s'authentifie
- Il consulte sa file de travail
- Il choisit l'étude qui a une tâche de l'état « en analyse »

- Il consulte les informations associées à cette tâche et approuve l'étude
- Il modifie l'état de la tâche à « terminé »
- Le système enregistre la modification au niveau de la BDD
- Le système notifie l'abonné par e-mail ou SMS

Scénario : Consulter une tâche

- L'utilisateur s'authentifie
- Il consulte sa file de travail
- Il choisit la tâche à consulter
- Il consulte les informations associées à cette tâche.

Diagramme de classes :

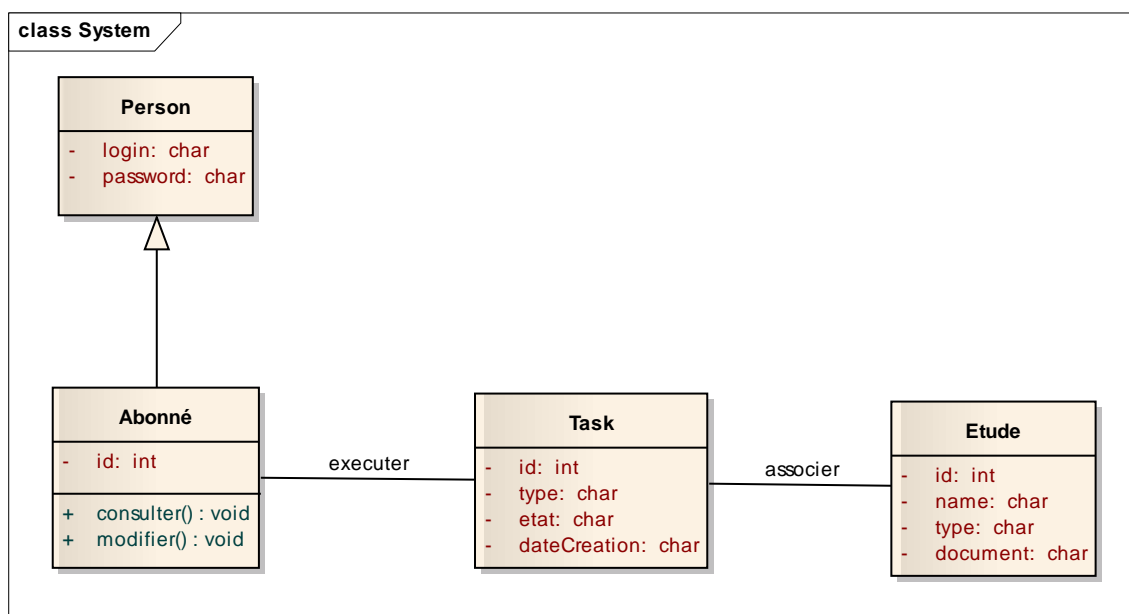


Figure 29. Diagramme de classes « demande d'une étude »

2.5.3. Architecture technique de XPlor EveryWhere

L'étude fonctionnelle consistait à dégager les exigences purement fonctionnelles tout en ne se préoccupant pas du niveau technologique. La présente partie est dédiée à dégager les exigences d'un système d'intelligence économique à la lumière des besoins fonctionnels. Le but est d'éviter les risques techniques et les problèmes que le prototype pourrait rencontrer comme les problèmes d'intégration, d'interopérabilité, d'évolution, ...

Les diagrammes de déploiement montrent la disposition physique des matériels qui composent le système et la répartition des composants sur ces matériels. Les ressources matérielles sont représentées sous forme de nœuds connectés entre eux, à l'aide d'un support de communication dont la nature et les caractéristiques peuvent être précisées. La configuration matérielle de notre SIE est schématisée par le diagramme de déploiement UML suivant :

L'architecture technique de XPlor EveryWhere (figure 30) (figure 31) est constituée de trois composants principaux:

Serveurs : un serveur web et le serveur BDD.

Clients : les navigateurs web et les mobiles 3G

Services et connexions :

- 3G connexion : Pour l'échange des messages entre le client mobile et le serveur Web.
- TCP connexion : il assure le transfert des requêtes et réponses entre le navigateur et le serveur Web.
- JDBC Connector : le driver du SGBD utilisé.

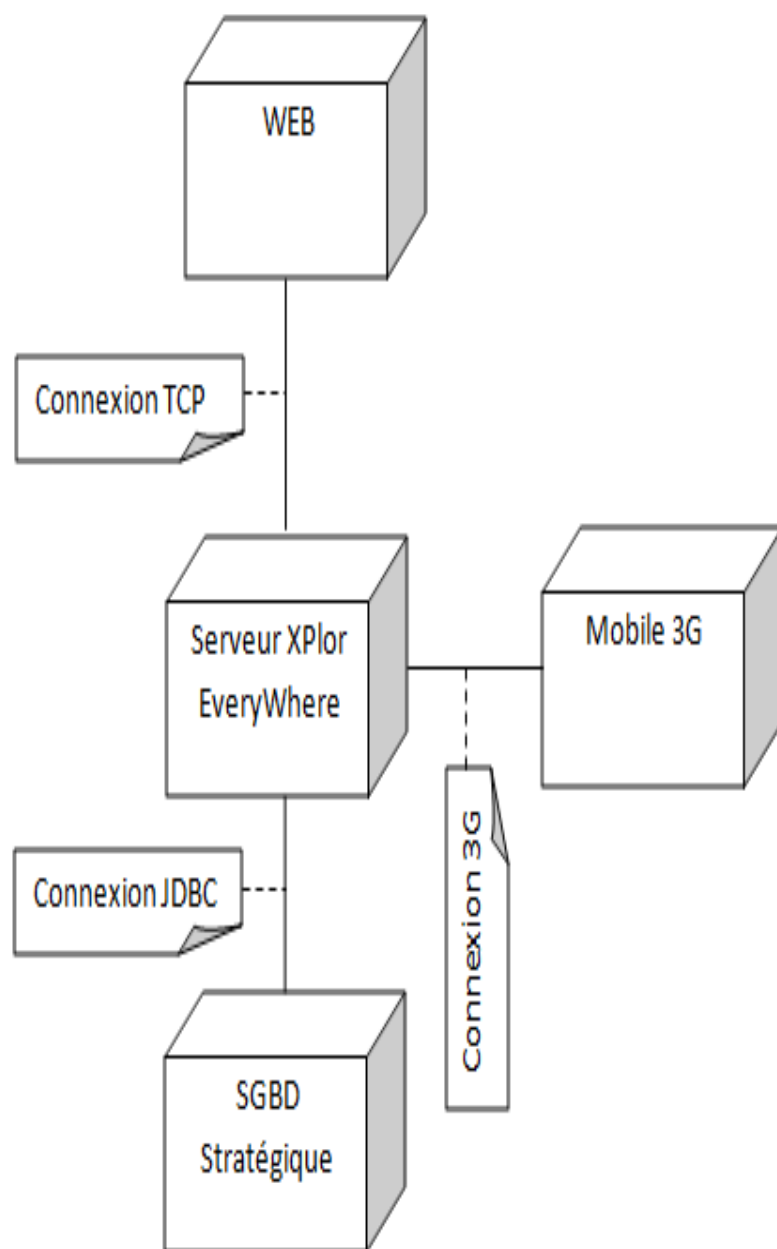


Figure 30. Diagramme de déploiement de XPlor EveryWhere

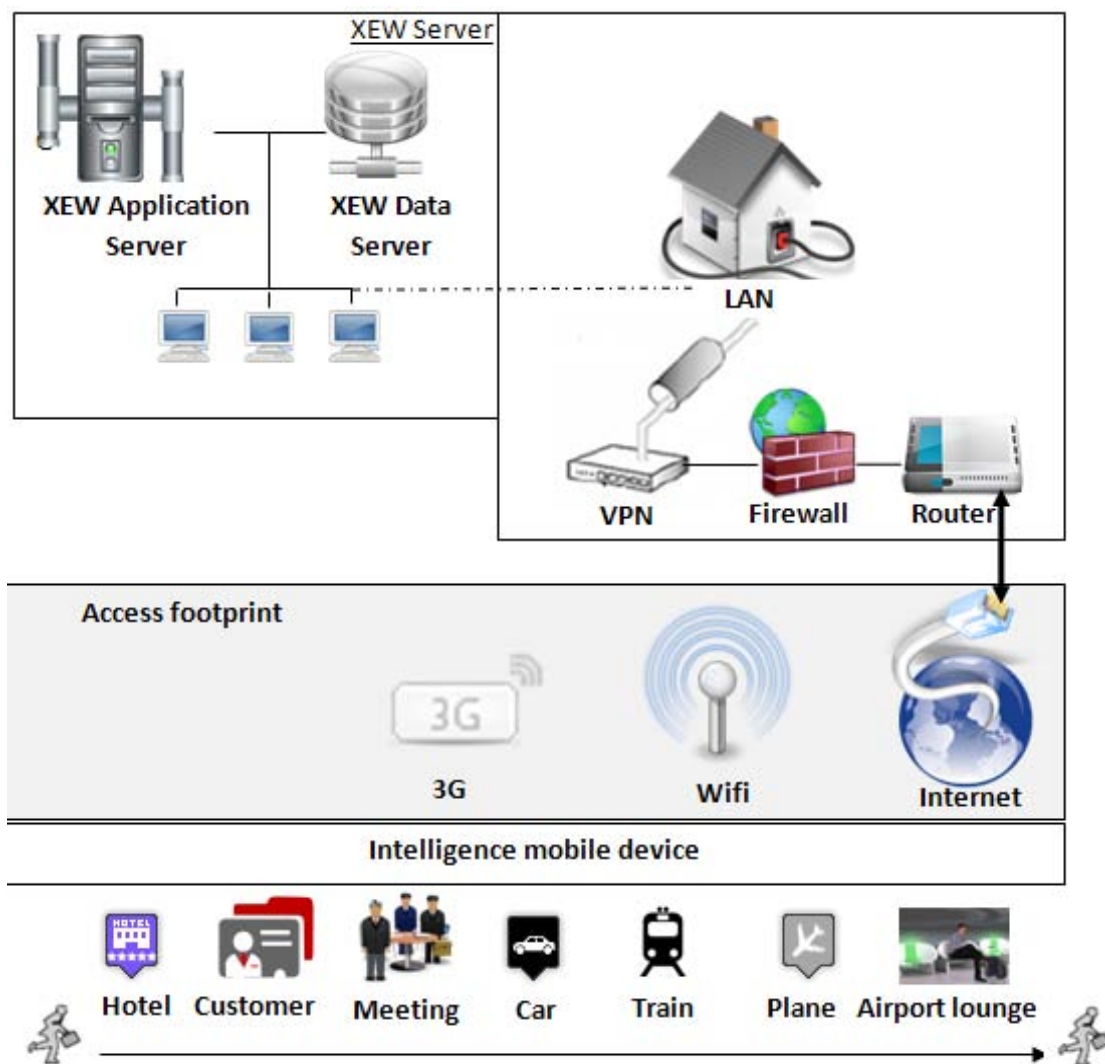


Figure 31. Architecture technique de Xplor EveryWhere

2.5.4. Architecture logicielle du système

Parmi les différentes façons de structurer une architecture, la mieux adaptée et maîtrisée en informatique est l'approche par couches. Une couche est une division logique et horizontale d'un système qui fournit une abstraction particulière du système aux couches supérieures. Chaque couche possède des responsabilités spécifiques. Dans une structuration par couches, les couches inférieures prennent en charge des responsabilités qui offrent un socle de base pour les couches supérieures, permettant d'abstraire celles-ci de problématiques qui ne les concernent pas.

Ainsi, nous avons adopté un découpage en quatre couches (figure 32). Une telle architecture permet également d'obtenir un bon niveau de réutilisabilité, à travers la ré-application des solutions aux problèmes rencontrés pour chaque couche, dans d'autres applications utilisant la même architecture.

- **Couche Présentation** : Cette couche est présente principalement pour gérer le domaine visuel des applications et pour gérer les interactions avec les utilisateurs. Elle est chargée de dessiner les fenêtres et autres composants graphiques.
- **Couche Métier** : Elle se compose de deux sous-couches :

- **Couche Service** : Son but principal est de fournir des services spécifiques au domaine (métier). Ces services correspondent a priori aux processus métiers définis lors de la phase d'analyse. Cette couche offre également des interfaces vers les autres couches.
- **Couche Modèle** : Cette couche est l'espace où réside le modèle du système. Sa charge principale consiste à modéliser toutes les entités métier du système. Cette couche est basée sur le modèle objet.

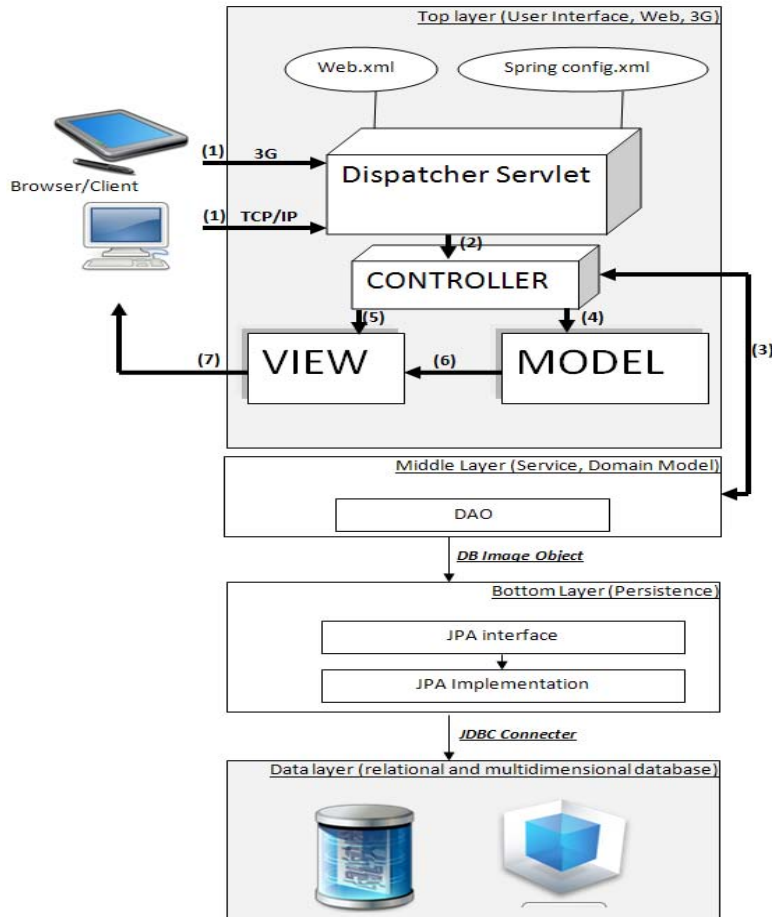


Figure 32. Architecture multi-couche de XPlor

- **Couche Mapping et persistance** : Cette couche, est certainement l'une des plus importantes. C'est ici que l'on trouve les fonctionnalités de base qui permettent de créer, rechercher et supprimer des entités métiers dans le respect des propriétés transactionnelles. C'est également dans cette couche que les mécanismes de conversion objet/relationnel peuvent en partie prendre place.
- **Couche Données** : Cette couche, est responsable du stockage physique de données. Elle assure le support transactionnel. Dans notre cas, cette couche se base sur le modèle relationnel.

Ainsi, l'architecture choisie est basée sur les quatre couches. Chaque couche fournit aux autres couches des services et utilise les services des autres.

2.6. Synthèse

L'objectif du deuxième chapitre de ce manuscrit est de présenter les systèmes d'intelligence économique. Dans l'introduction, nous avons présenté la genèse des systèmes d'information, jusqu'à l'émergence des systèmes d'informations stratégiques. Ces nouveaux systèmes ont fait l'objet de la première section de ce chapitre (section 2.2), ce qui nous a permis d'éclaircir et de différencier entre les systèmes d'intelligence économique et les systèmes de business intelligence, puisque le premier se présente comme la dimension externe des systèmes d'information stratégique, et le deuxième comme étant la dimension interne de ces derniers. Dans la section 2.3, nous avons traité les modèles d'analyse et de conception en intelligence économique à savoir le modèle MEDESSIE, Equate, MEDP et WISP. En nous basant sur l'aspect modélisation du rôle / utilisateur dans un SIE proposé dans les modèles Equate, MEDP et WISP et le modèle du besoin présenté dans MEDESSIE, nous avons proposé un modèle d'analyse multidimensionnelle de données.

Pour résumer, un système d'intelligence économique, se présente comme un compromis de Système – Information Stratégique et Système Information - Stratégique. Il a comme mission de :

- Faciliter les décisions, en permettant d'automatiser un certain nombre d'actions ou en mettant à disposition des décideurs des éléments nécessaires à la prise de décision,
- Coordonner les actions par le traitement de l'information,
- Stocker de manière durable et stable les informations,
- Traiter les données, c'est la fonction la plus importante du système d'information car il entraîne la création d'informations directement utilisables par les décideurs.

Dans le but de s'adapter à ces missions, nous proposons donc un système d'intelligence économique « XPlor EveryWhere » dans la section 2.4, considérant l'objectif d'interactivité avec les utilisateurs nomade afin de faciliter la navigation dans la connaissance, et d'offrir les bonnes conditions de la veille scientifique et technologique. Ce système d'intelligence économique repose sur :

- un modèle générique d'analyse multidimensionnelle, qui couvre les différentes étapes de cycle d'analyse de données textuelles structurées, semi-structurées ou non structurées : analyse de besoins, recherche, prétraitement, traitement, analyse et visualisation des données.
- un modèle de calcul des indicateurs
- un modèle de sécurisation.

La présentation des trois modèles fera l'objet du prochain chapitre.

Chapitre 3.

Entrepôt de données stratégique

« Comment trouver un diamant dans un tas de charbon sans se salir les mains ? » (SAS Enterprise Miner)

3.1.	Introduction	87
3.2.	Le modèle de représentation multidimensionnelle des documents	87
3.2.1.	Définition de la présentation multidimensionnelle	90
3.2.2.	L'entrepôt de données stratégique	91
3.2.3.	L'entrepôt sous forme de matrice symétrique	92
3.2.4.	L'entrepôt sous forme d'une matrice asymétrique	93
3.3.	Le modèle de calcul des indicateurs	94
3.3.1.	Définition d'un indicateur	94
3.3.2.	Les indicateurs dans l'entreprises	94
3.3.3.	Les indicateurs de l'activité scientifique et technologique	95
➤	Les indicateurs qualitatifs	95
➤	Les indicateurs quantitatifs	96
➤	Les indicateurs relationnels	97
3.3.4.	Le modèle d'indicateurs	98
➤	La fonction de filtrage	98
➤	La fonction combinatoire	99
3.4.	Approche de sécurité de système d'intelligence économique	100
3.4.1.	Risques de sécurité dans un système d'intelligence économique	100
➤	Risques liés à l'importation des données pour le traitement et l'analyse	102
➤	Risques liés à l'utilisation des logiciels de collecte	102
➤	Risques liés aux sources d'information	102
➤	Risques liés aux aspects juridiques	102
➤	Risques liés aux procédures de traitements et d'analyses des données	102
➤	Risques liés à la livraison de l'information	102
3.4.2.	Approche proposée pour le contrôle d'accès à un SIE	103
➤	Politique de sécurité dans un SIE	103
➤	Au niveau de la collecte et de l'importation des données	103
➤	Au niveau du traitement et de l'analyse des données	104
➤	Au niveau de la restitution de l'information aux décideurs	104
3.4.3.	Mise en œuvre dans un cas pratique : Sécurisation de Xplor EveryWhere	104
➤	Contrôle d'accès aux données et aux traitements dans Xplor EveryWhere	105
➤	Modélisation des processus de la plateforme Xplor	105
3.4.4.	Application du modèle de contrôle d'accès RBAC.	105
➤	Etablissement des différents rôles et les tâches associées	106
➤	Les permissions aux objets à protéger	107
➤	L'infrastructure de sécurité	108
3.4.5.	Application des ontologies de sécurité	108
3.5.	Synthèse	109

3.1. Introduction

La quantité d'informations, que les entreprises doivent gérer et surveiller, ne cesse pas d'augmenter, environs :

- 170 Terabytes des pages web,
- 35.000.000.000/ jour (400.000 Terabytes / an) des e-mails,
- 17,3 Exabytes / an de SMS,
- Plus que 5000 banques de données professionnelles (Brevets, articles scientifiques et techniques, données économique, etc.),
- l'émergence du web 2.0 .

Le traitement, l'analyse et l'exploitation de ces sources d'information dépassent la capacité de la majorité des entreprises. Ils se trouvent de plus en plus surinformé.

Face à cette situation, les techniques d'extraction de connaissance à partir de texte et le modèle de management stratégique de l'information visent à assurer le traitement, afin de pouvoir surveiller l'environnement de l'entreprise, d'une façon continue. Dans ce contexte la plateforme Tétralogie nous permet d'effectuer des analyses stratégiques sur des corpus d'information textuelle (brevets, articles scientifiques, flux RSS, etc.). L'information élaborée qui en est issue représente une synthèse de l'ensemble des documents collectés telle que : l'identification des acteurs et de leurs relations, les sous-sujets cohérents, les signaux forts et faibles, les tendances, les composantes stratégiques. Cependant, l'utilisateur final aimerait pouvoir lui-même zoomer facilement sur son propre environnement, pour connaître, par exemple, le positionnement de ses principaux concurrents, les procédés alternatifs connexes à son activité, les marchés potentiels, etc.

Le modèle d'analyse multidimensionnelle proposé répond à ces besoins par la prise en compte de l'aspect temporel au sein de l'analyse. Il repose sur l'extraction des connaissances à partir des données textuelles par l'analyse des données relationnelles évolutives. Cette spécification temporelle du modèle permet de situer les événements, les stratégies et les actions aussi bien dans le passé par reconstruction de la chronologie, que dans le présent par orientation temporelle ou le futur par anticipation, pour tout ce qui concerne les organisations successives d'un réseau, telles que les collaborations, alliances, fusions, acquisitions, co-citations, co-signatures, co-occurrences de tout ordre.

Ce modèle repose sur le modèle de représentation multidimensionnelle des documents, ce qui nous permet de générer l'entrepôt de données stratégique. La présentation du modèle fera l'objet de la première section de ce chapitre. Le modèle de calcul des indicateurs sera traité dans la section 3.3. Puis, nous détaillerons l'approche générique de sécurisation des données stratégique dans la section 3.4.

3.2. Le modèle de représentation multidimensionnelle des documents

Les techniques d'analyse multidimensionnelle sont actuellement bien maîtrisées pour toutes les données quantitatives disponibles sur les systèmes d'information des entreprises ou des administrations à condition que le SGBD s'y prête, que le schéma de la base soit adapté et que les données soient de qualité (homogènes, à jour, complètes, granularisées, compatibles,...). Il est toujours possible d'extraire les données utiles vers une base construite sur mesure pour la fouille multidimensionnelle, en réalisant, au passage, toutes les corrections nécessaires à leur exploitation. Par contre, pour les données textuelles

issues de l'ensemble des sources électroniques, ce type d'analyse est difficile à mettre en place (Sosson et al., 2001) (Tournier, 2007) (Ravat, 2007) (Teste, 2009) (El Haddadi et Dousset, 2009) (El Haddadi et al., 2010): les données sources n'ont pas les mêmes formats ou sont même non structurées, elles sont réparties, hétérogènes et la multitude des cas rencontrés ne permet pas de trouver une structure universelle ou de les regrouper. Les premiers travaux dans le domaine consistent à traiter le format XML, puisque ce type de document fournit un environnement permettant non seulement de considérer la structure mais aussi le contenu du document (Fuhr et al., 2001). La littérature scientifique nous présente différents approches répondant à la problématique de la modélisation multidimensionnelle intégrant les documents.

Une approche qui se base sur le schéma en étoile (figure 33) classique se limitant à des indicateurs numériques et ne permettant donc pas l'analyse du contenu des documents a été proposée (McCabe et al., 2000) (Jensen et al., 2001) (Mothe et al., 2003) (Kheith et al., 2005) (Tseng et Chou, 2006).

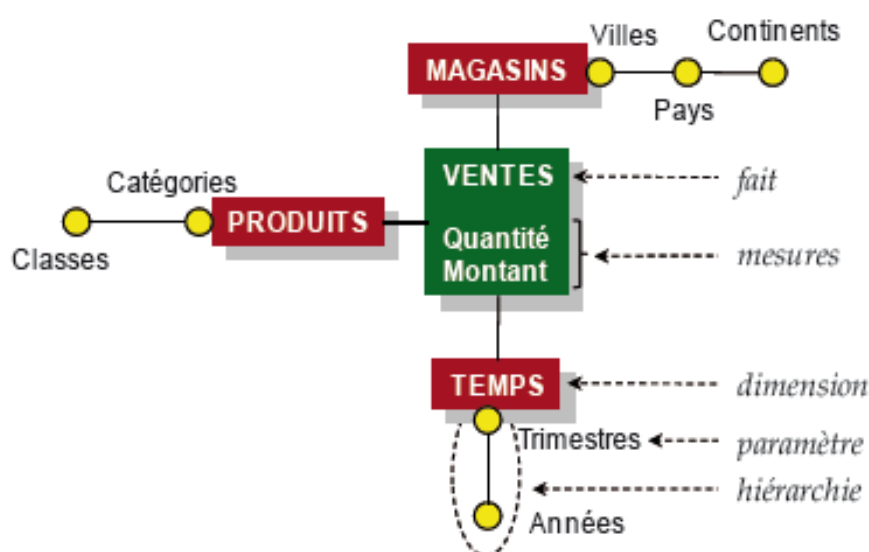


Figure 33. Exemple de schéma en étoile (Teste, 2009)

Un autre type d'approche propose d'intégrer les documents XML dans des entrepôts de documents spécifiquement dédiés à leur stockage (Khrouf et al., 2004) (Wang et al., 2003, 2005) (Wiwatwattana et al., 2007). Cependant, ces travaux restent limités à des analyses numériques, à des comptages d'instances, d'où la problématique de l'analyse de contenu textuel.

Pour répondre à cette problématique, les auteurs de l'article de (Park, et al., 2005) intègrent les techniques de fouille de données dans la modélisation multidimensionnelle. Nous pouvons aussi citer l'approche de (Pérez et al., 2008) qui propose de combiner la modélisation multidimensionnelle et les techniques de recherche d'information pour fournir les documents pertinents relevant de l'analyse en cours. Ces travaux proposent de lier des informations contenues dans des documents à des données multidimensionnelles afin d'expliquer les faits. Cependant, les travaux de l'équipe SIG-ED (Teste, 2009) consistent à analyser directement les informations contenues dans les documents. Les membres de l'équipe ont proposé deux approches. Une première approche qui consiste à étendre les modèles de représentation multidimensionnelle de données numériques aux documents (Teste, 2009) (Pujolle et al., 2008) (Ravat et al., 2007a, 2008). Cette approche concerne la définition de dimensions modélisant la structure des documents (figure 34). Une dimension « structure » est constituée à partir des structures extraites des documents via la structure arborescente des documents XML (DTD ou Schéma). Chaque

paramètre de ces dimensions modélise les différents niveaux de granularité d'un même document (section, sous-section, paragraphe...).

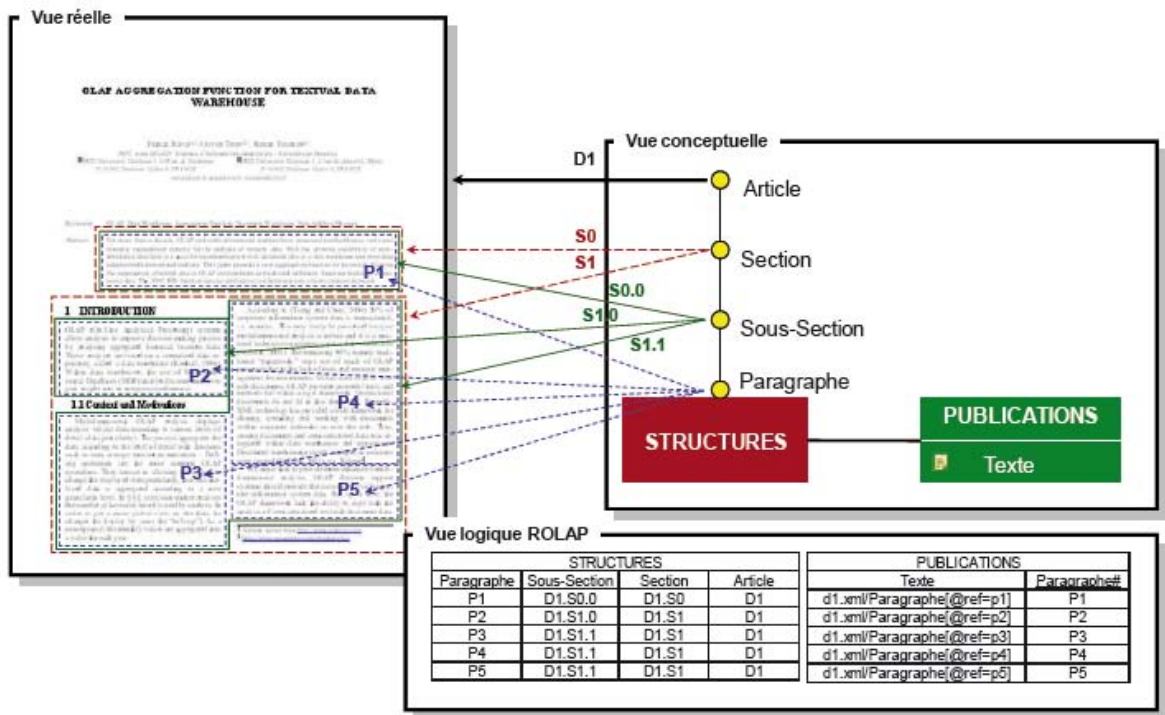


Figure 34. Exemple d'une dimension « structure » (Teste, 2009)

Une deuxième approche propose le modèle « Galaxie », pour pallier les difficultés des modèles par extension (Tournier, 2007) (Ravat et al., 2007b) (Teste, 2009). Cette modélisation en galaxie (figure 35) repose sur l'idée d'utiliser un concept unique pour représenter les données qui peuvent être employées de manière symétrique en tant que sujet ou axe d'analyse.

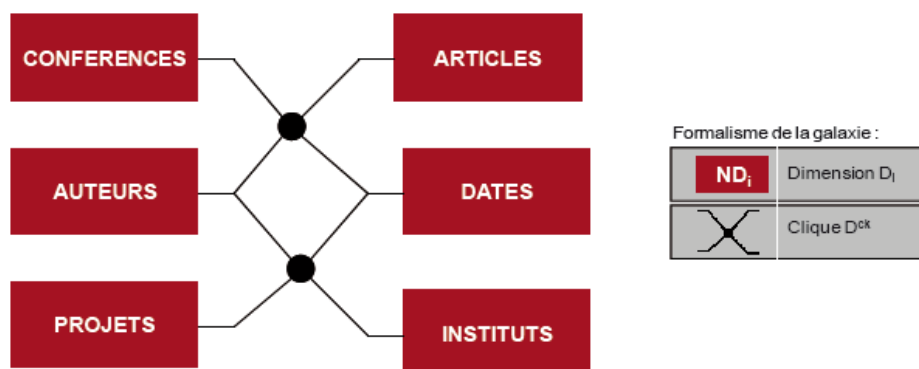


Figure 35 : Exemple d'un galaxie (Teste, 2009)

Ces travaux proposent des modélisations multidimensionnelles des documents XML. Notre objectif va plus loin puisque nous souhaitons pouvoir analyser et modéliser tous les types de document, quel que soit le format en combinant les techniques de la fouille de texte et la modélisation multidimensionnelle. Cette approche consiste, comme expliqué dans la section 2.3.3, à traiter les données dans leurs formes natives.

Le format natif offre plusieurs avantages : une meilleure réactivité, une mise à jour du corpus facilitée et une préservation de l'ensemble des informations. Néanmoins, afin de s'adapter à la plupart des structures, il est nécessaire d'utiliser les méta-données qui sont des outils de description des

formats. Notons par ailleurs que plus de 90% des cas rencontrés peuvent être traités sans aucun reformatage.

Donc, pour standardiser la fouille multidimensionnelle sur les données textuelles de toutes provenances, nous avons proposé une structure unifiée permettant de stocker toutes les relations inter-items rencontrées dans les documents analysés. Cette technique permet de construire des cubes croisant deux variables quelconques et le temps.

Cette modélisation permet l'extraction des relations de dépendances existantes entre les différents attributs de corpus d'information traités. Notre but est de présenter sous forme d'une modélisation multidimensionnelle les relations de dépendances entre les variables présentes dans des grandes collections de documents. La mise en évidence de ces relations et leur analyse permettent d'établir des scénarii tendant à expliquer les mécanismes complexes qui gèrent le fonctionnement de l'environnement d'un domaine ou d'un acteur. Le but est de réduire l'espace informationnel afin de mieux le maîtriser, en éliminant les éléments indépendants, pour ne garder que les relations les plus significatives en termes de stratégie. De nombreuses mesures de dépendance sont utilisables : covariances, corrélations, coïncidences, contingences, co-occurrences, proximités. Elles donnent des visions différentes mais complémentaires d'une même réalité.

3.2.1. Définition de la présentation multidimensionnelle

Définition. Une présentation multidimensionnelle PM est définie par $(Ch^{PM}, Att^{PM}, T^{PM})$ où

- $Ch^{PM} = \{ch^{PM}_1, \dots, ch^{PM}_n\}$ est l'ensemble des champs du corpus,
- $Att^{PM} = \{att^{PM}_1, \dots, att^{PM}_{2n}\}$ est l'ensemble des attributs ou les valeurs d'un champ,
- $T^{PM} = \{t^{PM}_1, \dots, t^{PM}_n\}$ est l'ensemble des variables temporelles.

La présentation multidimensionnelle est alors une modélisation à trois dimensions. Elle nous permet de définir les différentes relations de dépendance entre les valeurs des champs avec la prise en compte de la structure temporelle. Ces relations peuvent être des variables de différents type (Dousset, 2003) (Ghalamallah, 2009) :

- Qualitatives ordinales : année de publication, de dépôt, heures de connexion, jours de la semaine ou du mois, ...
- Qualitatives hiérarchiques : thésaurus hiérarchiques, zones géographiques, inclusions sémantiques, chemins d'accès aux fichiers, ...
- Qualitatives nominales : auteurs, revues, pays, dictionnaires de mots-clés, ...

De plus les variables qualitatives peuvent être :

- Uni-modales : présence ou absence d'une caractéristique.
- Multi-modales à modalités exclusives : année, revue, langue, type de document, source, ... (une seule modalité de cette variable est alors requise obligatoirement pour chaque document).
- Multi-modales à modalités éventuellement multiples : auteurs, mots-clés, classifications, citations complètes, ... (plusieurs modalités différentes de cette variable peuvent apparaître une seule fois chacune dans le même document).

- Multi-modales à modalités multiples éventuellement redondantes: mots du texte libre, affiliations, pays et villes des auteurs dans le cas de plusieurs adresses, auteurs cités, revues citées, ... (une même modalité peut alors apparaître plusieurs fois dans le même document).

Pour un corpus de notices dont la structure d'extraction est comme suit :

Définition. La structure d'extraction $SE = \langle NU, DP, AU, JN, PA, MC, OR \rangle$

où

- NU est le numéro de la notice bibliographique,
- DP est la date de publication, regranularisée,
- AU est le champ auteurs, filtré,
- JN est le journal, harmonisé,
- PA est le champ pays (des auteurs), extrait du champ adresse,
- MC est le champ mots-clés s'il y a une seule base ou le champ multi-terms s'il y en a plusieurs,
- OR est le champ organismes (des auteurs), extrait du champ adresse.

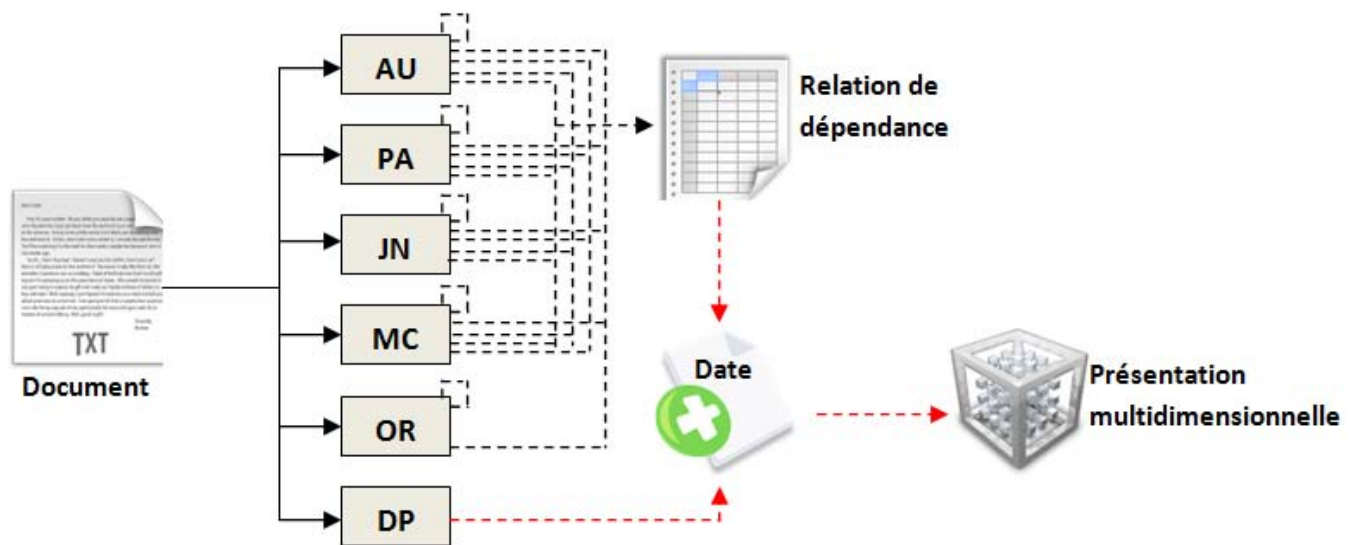


Figure 36. Dépendances intra document à trois dimensions et l'élément temporel.

Nous proposons de construire des matrices à trois dimensions qui permettent de définir les relations de dépendances existantes entre les variables du corpus en y intégrant systématiquement la variable temporelle, c'est-à-dire l'élément « Date ».

3.2.2. L'entrepôt de données stratégique

La présentation multidimensionnelle vise à identifier toutes les relations de dépendances existantes entre les différents variables de sujet d'analyse. Ces relations sont définies par des matrices de co-occurrences, qui indiquent la présence simultanée des modalités de deux variables qualitatives dans un

document. Nous adoptons ces matrices en y rajoutant une troisième variable temporelle (Année, Mois, Jours, Heures), ce qui consiste à indiquer la présence d'une telle relation, dans un tel moment.

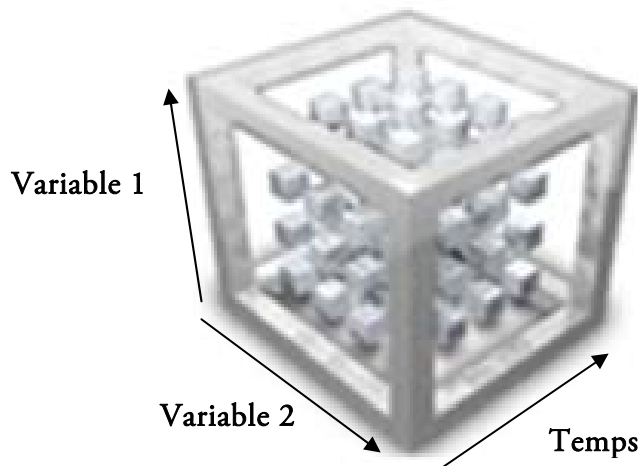


Figure 37. Cube de données

L'entrepôt permet de regrouper les relations existantes dans un corpus en périodes. Nous identifions deux types de forme de cube :

- Cube sous forme de matrice carrée (symétrique) : dans le cas où nous considérons la coprésence des modalités d'une même variable et la variable temporelle dans un document.
- Cube sous forme de matrice rectangulaire (asymétrique) : dans le cas où nous considérons la présence des modalités de deux variables distinctes et la variable temporelle dans un document.

3.2.3. L'entrepôt sous forme de matrice carée

Définition. Soient X et Y deux variables qualitatives distinctes à p et q modalités décrivant un ensemble de n individus. Et T une variable qualitative temporelle à r modalités.

Soit χ l'ensemble des modalités $\{x_1, \dots, x_p\}$ de la variable X .

Soit γ l'ensemble des modalités $\{y_1, \dots, y_q\}$ de la variable Y .

Soit τ l'ensemble des modalités $\{t_1, \dots, t_r\}$ de la variable T .

L'entrepôt sous forme de matrice carée est un cube à p lignes et p colonnes qui a pour élément générique le nombre n_{ijk} d'individus tel que $x_i \in \chi$ et $x_j \in \gamma$ et $t_k \in \tau$

$X \times X \times T$	X	x_1			x_j			x_p		
	T	t_1	...	t_r	t_1	...	t_r	t_1	...	t_r
X												
x_1		n_{111}	...	n_{11r}	n_{1j1}	...	n_{1jr}	n_{1p1}	...	n_{1pr}
x_2		n_{211}	...	n_{21r}	n_{2j1}	...	n_{2jr}	n_{2p1}	...	n_{2pr}

\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
X_i	n_{i11}	...	n_{i1r}	n_{ij1}	...	n_{ijr}	n_{ip1}	...	n_{ipr}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
X_p	n_{p11}	...	n_{p1r}	n_{pj1}	...	n_{pjr}	n_{pp1}	...	n_{ppr}

Figure 38. Cube sous forme d'une matrice carée

3.2.4. L'entrepôt sous forme d'une matrice rectangulaire

Définition. L'entrepôt sous forme d'une matrice rectangulaire est un cube à p lignes et q colonnes qui a pour élément générique le nombre n_{ijk} d'individus tel que $x_i \in \chi$ et $y_j \in \gamma$ et $t_k \in \tau$

$X \times Y \times T$	Y	y_1			y_j			y_q		
	T	t_1	...	t_r	t_1	...	t_r	t_1	...	t_r
X												
X_1		n_{111}	...	n_{11r}	n_{1j1}	...	n_{1jr}	n_{1q1}	...	n_{1qr}
X_2		n_{211}	...	n_{21r}	n_{2j1}	...	n_{2jr}	n_{2q1}	...	n_{2qr}
\vdots		\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
X_i		n_{i11}	...	n_{i1r}	n_{ij1}	...	n_{ijr}	n_{iq1}	...	n_{iqr}
\vdots		\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
X_p		n_{p11}	...	n_{p1r}	n_{pj1}	...	n_{pjr}	n_{pq1}	...	n_{pqr}

Figure 39. Cube sous forme d'une matrice rectangulaire

Grâce à la structure du cube, nous construisons le corpus multidimensionnel. L'alimentation du corpus se base sur la prise en compte des relations de dépendances existantes dans la structure du cube par suppression des éléments indépendants. Afin de construire le corpus multidimensionnel, nous ne garderons que les cases dont les valeurs sont supérieures ou égales à un.

Nous définissons le corpus multidimensionnel associé au cube comme suit :

Soit la structure du corpus multidimensionnel (3 dimensions) SC^{M3D} définie comme suit :

$$SC^{M3D} = \{D_i, D_j, D_t\}$$

Le corpus multidimensionnel C^{M3D} est l'ensemble non vide \emptyset de l'union des trois attributs :

$$(Att_i \cup Att_j \cup Att_t)$$

Avec

- $Att_i \in D_i$ est l'ensemble des attributs $\{att_1^i, \dots, att_p^i\}$ de la dimension « i »,
- $Att_j \in D_j$ est l'ensemble des attributs $\{Att_1^j, \dots, Att_q^j\}$ de la Dimension « j »,
- $Att_t \in D_t$ est l'ensemble des attributs $\{Att_1^t, \dots, Att_r^t\}$ de la Dimension « t ».

Exemple. La figure suivante présente une présentation multidimensionnelle formé des collaborations entre les auteurs en cellule et de trois arêtes graduées respectivement par les thématiques de recherche, les organismes et des dates de publication. Cette présentation ne se limite pas à trois axes mais se généralise en hyper-cube où le nombre d'axes est quelconque pouvant aller jusqu'à plusieurs dizaines.

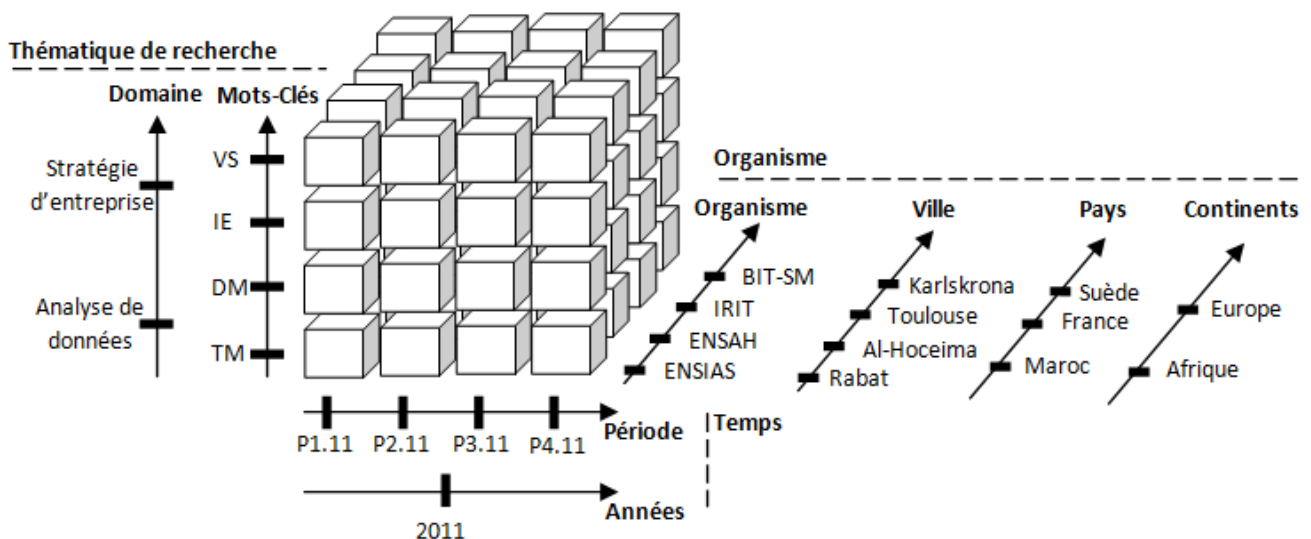


Figure 40. Cube d'évaluation de recherche par auteur

3.3. Le modèle de calcul des indicateurs

3.3.1. Définition d'un indicateur

Un indicateur est un outil d'évaluation et d'aide à la décision, qui sert à fournir des indications, des renseignements sur la valeur d'une grandeur. Il a le rôle de résumer l'information complexe, d'évaluer la performance d'une organisation, de politiques, de progrès, de tendances, d'identifier des mécanismes sectoriels.

Cet indicateur doit refléter réellement les variations de ce qu'il est censé de synthétiser ou mesurer. Il doit mettre en évidence les liens entre les différents composants du système ou de l'écosystème.

3.3.2. Les indicateurs dans l'entreprise

Au sein de l'entreprise, les managers utilisent des indicateurs pour mesurer la performance des différentes organisations et piloter l'activité selon les objectifs principaux de métier. On peut citer quelques indicateurs selon le domaine :

- Economique : le Produit Intérieur Brut (PIB), dont on surveille le taux de croissance afin de mesurer la croissance économique, et le Produit National Brut (PNB) qui permet de comparer les puissances économiques des différentes nations. Industriel : le taux de rendement synthétique ou le taux de rendement global.
- Finance : le ROI (retour sur investissement), la marge nette.
- Logistique : le taux de service, la rotation de stock, les coûts.
- Marketing : la part de marché, le taux de marge, le taux de pénétration du marché.
- Achats : le Pareto des fournisseurs.
- Etc.

3.3.3. Les indicateurs de l'activité scientifique et technologique

La bibliométrie est l'exploitation statistique des publications scientifiques (articles, brevets, ...). Cette analyse permet de rendre compte de l'activité des producteurs (chercheur, laboratoire, institut, université, etc.) et des diffuseurs (périodique, éditeur, etc.) de l'information scientifique, tant d'un point de vue quantitatif que qualitatif (Académie des sciences, 2011) (Sahel, 2011) (Man et al., 2009) (Filliatreau, 2001). Dans les deux cas, ces techniques sont utilisées pour rassembler les informations, dénombrer les publications et l'exploitation statistique. Elles aboutissent cependant à des traitements de natures différentes et des résultats qui n'ont ni la même finalité, ni les mêmes usages :

- Des indicateurs uni-variés permettant des comparaisons quantitatives entre des ensembles de publications caractérisées.
- Des indicateurs relationnels destinés à cartographier, de façon figée ou évolutive, les domaines couverts par des ensembles de publications.

On peut ainsi cumuler des points de vue complémentaires, afin d'avoir une idée plus juste et multiforme de l'interprétation.

On peut présenter les différents indicateurs bibliométriques qu'on trouve dans la littérature scientifique (Kermarrec et al., 2007) (Académie des sciences, 2011) (Sahel, 2011) (Filliatreau, 2001), en trois parties : la première est consacrée aux principaux indicateurs qualitatifs, la deuxième présente les indicateurs quantitatifs, tandis que la troisième traite des indicateurs relationnels, c'est-à-dire les indicateurs mesurant les liens entre les différents attributs du corpus informationnel : les réseaux sociaux, sémantiques, internationaux, les alliances, etc.

➤ Les indicateurs qualitatifs

Le facteur d'impact : c'est l'application la plus connue des techniques bibliométriques. Le terme et la fonction ont été inventés par Eugène Garfield, fondateur de l'Institute for Scientific Information (ISI). Il était destiné à l'origine aux professionnels de l'édition. Son rôle a été détourné de sa fonction initiale par les chercheurs. Il est défini par le nombre moyen de citations attribuées à chaque revue, pendant une période donnée, des articles publiés dans cette revue. Le FI est une première approximation en bonne corrélation avec la qualité de la revue, malgré le fait que la période prise en compte pour son calcul est probablement trop courte (deux ou cinq ans dans la base JCR de ISI) ce qui ne donne pas de place aux articles donnant lieu à des citations pendant de très nombreuses années (souvent les articles qui ont eu le plus d'impact sur le progrès scientifique). Il convient, en outre, de remarquer que les facteurs d'impact peuvent être l'objet de manipulations par certaines grandes revues.

L'indice H (ou indice de Hirsch) : c'est un indice essayant de quantifier la productivité scientifique et l'impact d'un scientifique en fonction du niveau de citation de ses publications. Il peut aussi s'appliquer à un groupe de scientifiques, tel qu'un département, une université ou un pays. Le facteur H est défini de la façon suivante : il est calculé en classant les publications par ordre décroissant en fonction du nombre de citations auxquelles elles ont donné lieu. Le rang de la publication pour lequel le nombre de citations est égal à ce rang est le facteur H. Ainsi un chercheur ayant un facteur H de 47 a publié 47 articles ayant chacun été cités au moins 47 fois. Le facteur H dont le niveau moyen dépend beaucoup des disciplines, on y reviendra, est intéressant mais souffre de plusieurs faiblesses. Il avantage les chercheurs seniors ayant une activité de recherche reconnue sur le long terme (le facteur H augmente régulièrement avec l'âge), et les normalisations - par exemple la division par le nombre d'années de carrière - sont artificielles et peu utilisables.

L'indice G : cet indice a été introduit par L. Egghe pour pallier certains défauts de l'indice H et distinguer l'excellence dans la productivité. Un chercheur aura par exemple un facteur G de 83 si ses 83 articles les plus cités totalisent au moins 6889 citations, c'est à dire G au carré. Ce facteur présente l'avantage de valoriser les articles très cités qui ont un temps de vie élevé. Ces articles contribuent à faire croître G dans le temps alors qu'ils ne modifient pas H. On peut aussi envisager toutes les dérivés possibles de G afin de mieux cerner la forme de la distribution du nombre des citations en fonction des articles les plus cités. L'indice G est aujourd'hui moins connu et moins utilisé que l'indice H, peut-être en partie par le fait que sa signification est moins évidente en première lecture et que son accès est moins répandu.

Le classement de Shanghai : Le classement académique des universités mondiales ou classement de Shanghai (appellation commune du Academic Ranking of World Universities en anglais) est un classement des principales universités mondiales, établi par des chercheurs de l'université Jiao-Tong de Shanghai en Chine. Ces institutions sont classées selon six critères comme le nombre de publications dans deux revues scientifiques et le nombre de prix Nobel attribués aux élèves et aux équipes pédagogiques. Les créateurs du classement soulignent eux-mêmes certaines de ses limites, notamment un biais en faveur des pays anglophones et des institutions de grande taille et les difficultés à définir des indicateurs adéquats pour classer les universités spécialisées dans les sciences sociales.

➤ Les indicateurs quantitatifs

Le nombre de publications : Cet indicateur illustre la production scientifique, mesurée par le comptage de publications scientifique (articles, revues, journaux, livres, comptes-rendus, rapports, etc.). Le comptage fournit une première mesure simplifiée et approximative de la quantité de travail par un producteur (chercheur, laboratoire, institut, université, pays, etc.). Ainsi, le dynamisme d'un producteur peut être suivi et son évolution dans le temps mesurée. Cependant, cet indicateur ne tient pas compte de la qualité des publications.

Le nombre de citations : c'est un indicateur de l'impact d'un article cité, de son intérêt. Cependant, un chercheur peut citer ces propres travaux, augmentant ainsi le nombre de citations qui lui seront créditées. C'est ce qu'on appelle le phénomène de « auto-citations » qui justifie les critiques sur la fiabilité de cette mesure.

Le nombre de co-signatures : c'est un indicateur qui vise à identifier la coopération nationale ou internationale. C'est un paramètre pour mesurer la croissance (ou le déclin) des recherches en coopération. L'analyse chronologique des co-signatures est une mesure de la pénétration de la coopération internationale dans la production de la science et de la technologie nationale.

Le nombre de brevets : Le comptage des brevets a comme objectif l'identification de la place d'une invention et le rôle de chaque inventeur dans la mise au point de nouvelles techniques. C'est donc une mesure de l'activité novatrice et de la capacité technologique à l'échelle des nations et des branches et des entreprises industrielles.

Quatre types de données de brevets sont concernés :

- le nombre de « demandes résidentes » qui émanent d'inventeurs résidant dans le pays concerné pendant une période donnée : cet indicateur donne une idée de la production de technologie ; on peut en tirer des informations supplémentaires comme, par exemple, le « coefficient d'inventivité » (= demandes résidentes/10 000 habitants) ;
- le nombre de « demandes non-résidentes » de brevets, déposées par les inventeurs ne résidant pas dans le pays : cet indicateur reflète la pénétration technologique ;
- le nombre de demandes nationales de brevets qui est la somme des demandes résidentes et non-résidentes : il indique, en quelque sorte, la taille du « marché technologique » que représente ce pays ;
- le nombre de demandes de brevets déposées à l'étranger par les inventeurs résidant dans le pays concerné : cet indicateur reflète la « diffusion technologique » de ce pays.

Le nombre de citations des brevets : Cet indicateur mesure l'impact de la technologie.

➤ Les indicateurs relationnels

Les co-publications : Une co-publication est le résultat d'une coopération entre les représentants de chaque entité et pays participant à un programme de recherche défini et exécuté en commun. Cet indicateur permet d'identifier les principaux partenaires des opérations de recherche et de décrire les réseaux scientifiques.

L'indice d'affinité : L'indicateur utilisé pour évaluer le taux relatif des échanges scientifiques d'un pays donné (A) avec un autre pays (B), pendant une période de temps donnée (et éventuellement dans un domaine scientifique précis), et ceci par rapport à l'ensemble de la coopération internationale de ces deux mêmes pays pendant la même période, est appelé « indice d'affinité ».

Cet indicateur mesure non seulement les liens entre pays mais aussi le niveau d'équilibre entre eux, autrement dit les « rapports de force » qui caractérisent ces échanges. Il montre aussi bien les domaines les plus forts que les domaines les plus faibles. En examinant l'évolution des indices d'affinité en fonction du temps, on obtient une indication des changements intervenus dans les rapports scientifiques bilatéraux.

Les liens scientifiques des citations : Cet indicateur mesure l'influence des rapports entre communautés scientifiques. A travers les citations, on peut dessiner les réseaux d'influence entre différentes communautés scientifiques.

Corrélations entre publications scientifiques et brevets : Cet indicateur illustre des liens (interactions) entre les sciences (mesurées par les publications) et les technologies (reflétées par les brevets).

Les co-citations : Les co-citations mesurent le nombre de fois que deux publications sont citées simultanément dans un même article. Cet indicateur illustre des réseaux thématiques et l'impact des auteurs. La méthode des co-citations représente, dans sa finalité, les réactions de la communauté scientifique aux résultats de la recherche.

La co-occurrence des mots : Cet indicateur examine le nombre de fois que deux mots ont été utilisés ensemble dans des publications ou des brevets, ce qui permet la visualisation de réseaux sémantique pour représenter les thèmes majeurs du domaine étudié et leurs relations.

Aucun de ces indicateurs, en fait, n'est complètement satisfaisant. Il paraît indispensable dans la pratique, de ne jamais utiliser l'un d'entre eux isolément. Ainsi, des efforts importants doivent être faits dans les analyses stratégiques, quel que soit le contexte d'analyse, pour cumuler des points de vue complémentaires, afin d'avoir une idée plus juste est multiforme de l'interprétation. Pour faire progresser cette réflexion, nous proposons le modèle d'indicateur, afin de faciliter la navigation dans l'information stratégique selon le besoin des décideurs d'une part, et de contrôler la qualité des traitements avec les indicateurs de contrôle d'autre part.

3.3.4. Le modèle d'indicateur

Les techniques de la bibliométrie sont adoptées initialement par les institutions académiques à des fins d'évaluation du rendement (science policy of research management). Les méthodes issues de la bibliométrie voient aujourd'hui leurs domaines d'application déborder vers le secteur privé et se répandre en tant qu'outils d'analyse dans :

- Les centres décisionnels des organisations commerciales,
- L'analyse de l'environnement scientifique et technique,
- La veille technologique et concurrentielle.

Notre objectif est l'application des techniques de la bibliométrie dans le domaine de l'intelligence économique, ce qui nous permet de définir un ensemble d'indicateurs grâce à l'analyse des différents corpus d'informations. Le modèle proposé ne se limite pas aux notices bibliographiques types, articles ou brevets et à l'application des indicateurs standards de la bibliométrie (impact factor, indice H, indice G, etc.), mais tend vers l'application de la bibliométrie et la définition d'autres indicateurs de calcul sur autre type d'information (les réseaux sociaux, les flux RSS, etc.).

Le modèle d'indicateur présente deux fonctions majeures : une fonction de filtrage qui consiste à visualiser l'environnement d'un attribut et une fonction combinatoire qui consiste à combiner les différents indicateurs de la bibliométrie, cités dans la section précédente, selon les besoins d'analyse, le contexte et les intérêts stratégiques.

➤ La fonction de filtrage

La fonction « filtrage » permet dans un premier temps de faire une analyse descriptive. C'est-à-dire il nous permet de sélectionner un attribut (ou plusieurs) dans le corpus informationnel, de fixer un seuil d'analyse, puis de visualiser l'environnement de l'attribut (ou plusieurs) o-occurrence. Dans un deuxième temps nous permet d'étudier l'évolution de l'environnement avec l'étude des attributs temporels. Donc la fonction « filtrage » est l'ensemble des fonctions « sélection », « seuil », « environnement » et « évolution ».

Définition. La fonction F_{filtrage} est définie comme suit :

$$F_{\text{filtrage}}(Att_i) = \text{Ens. } (F_{\text{sélection}}, F_{\text{seuil}}, F_{\text{environnement}}, F_{\text{évolution}})$$

La fonction sélection permet de sélectionner un attribut (ou plusieurs) dans le corpus informationnel.

Définition. La fonction sélection $F_{selection}$ est définie comme suit :

C^{M3D}

$$F_{selection}(Att_i) = \{Val_i^{M3D}\}$$

Avec

$\{Val_i^{M3D}\}$ est l'ensemble des valeurs de l'attribut sélectionné.

La fonction seuil a pour objectif de filtrer les relations en fixant un nombre minimum soit à la valeur d'un attribut, soit au nombre d'occurrences d'une relation entre deux attributs.

La fonction environnement permet de visualiser les différentes interactions de l'attribut choisi (ou plusieurs), ce qui facilite la navigation entre les différents liens relationnels, c'est-à-dire la navigation entre les réseaux sociaux, sémantiques, internationaux, etc.

Définition. La fonction environnement $F_{environnement}$ est définie comme suit :

$$F_{environnement}(Att_i) = \{Att_j^{M3D}, Val_j^{M3D}\}$$

Avec

- Att_j^{M3D} est l'attribut j associé à l'attribut i,
- Val_j^{M3D} est la valeur de l'attribut j,
- $\{Att_j^{M3D}, Val_j^{M3D}\}$ est l'ensemble des attributs qui présente l'environnement de l'attribut i.

La fonction évolution permet une analyse chronologique du relationnel, d'un attribut (ou plusieurs).

Définition. La fonction $F_{evolution}$ est définie comme suit :

$$F_{evolution}(Att_i) = \{Att_t^{M3D}, Val_t^{M3D}\}$$

Avec

- Att_t^{M3D} est l'attribut temporel associé à l'attribut i,
- Val_t^{M3D} est la valeur de l'attribut temporel,
- $\{Att_t^{M3D}, Val_t^{M3D}\}$ est l'ensemble des attributs temporels associés à l'attribut i.

➤ La fonction combinatoire

Chaque fonction (indicateur) a ses avantages et ses limites. Il ne faut pas les considérer comme des indices exacts ; ils sont complémentaires. Les différentes méthodes doivent être utilisées en parallèle. Il

est donc nécessaire de définir une fonction combinatoire qui nous permet de répondre aux besoins des analyses. Le tableau suivant présente quelques combinaisons de fonctions (indicateurs).

BESOINS INFORMATIONELS	Fonctions Combinatoires			
Evaluation d'un institut de recherche	$F_{filtrage}(Institut)$	<i>Le facteur d'impact</i>	<i>L'indice h</i>	<i>Cl. Shangai</i>
La qualité d'un auteur	$F_{filtrage}(Auteur)$	<i>Le facteur d'impact</i>	<i>La notoriété</i>	

Tableau 5. *La fonction combinatoire des fonctions (indicateurs)*

Au cours de cette section, nous avons présenté le modèle des indicateurs qui permet de synthétiser l'environnement informationnel d'une analyse stratégique demandée. Le calcul de ces indicateurs demande des traitements et de contrôle d'accès aux données stratégiques, ce qui déclenche un modèle sécuritaire qui fera l'objet de la prochaine section.

3.4. Approche de sécurité d'un système d'intelligence économique

Le besoin en matière de sécurité dans un système d'intelligence économique provient du fait que les informations manipulées sont d'ordre stratégique ayant une valeur assez importante. Une telle sécurité ne doit pas être considérée comme une option supplémentaire qu'offre un SIE pour se distinguer d'un autre, d'autant plus que la fuite de ces informations n'est pas le fait de faiblesses inhérentes aux systèmes informatiques des entreprises, mais c'est avant tout une question organisationnelle.

Dès lors, les mêmes procédures organisationnelles relatives à l'utilisation des ressources informatiques, visant à assurer la fonction de veille stratégique et concurrentielle dans un SIE devraient aussi s'appliquer au domaine du contrôle de l'information : établissement d'une politique de sécurité, sensibilisation des salariés, mise en garde quant au contenu diffusé, et l'adhésion à des bonnes pratiques sont des outils potentiels pour réduire le risque des fuites d'information.

Nous rappelons que l'intelligence économique est à la fois un processus et un produit. En tant que processus, l'IE est l'ensemble des méthodes légales et éthiques qu'une entreprise utilise pour mobiliser l'information qui l'aide à réussir dans un environnement mondialisé. En tant que produit, l'IE est l'ensemble des informations sur les activités des concurrents à partir de sources publiques et privées, et son champ d'application est le comportement présent et futur des concurrents, des fournisseurs, des clients, de la technologie, des acquisitions et fusions, des marchés, des produits et services, et de l'environnement des affaires en général.

Il s'agit donc de sécuriser à la fois des données (IE en tant que produit) et des procédés (IE en tant que méthode de traitement). De là est partie notre contribution de proposer une approche de contrôle d'accès aux données et aux traitements dans une plateforme d'IE. Il s'agit d'un modèle générique pour intégrer le contrôle d'accès dans le cycle de vie d'un SIE afin que toutes les tâches dans un processus d'IE soient exécutées uniquement par les utilisateurs autorisés et habilités.

3.4.1. Risques de sécurité dans un système d'intelligence économique

L'enquête menée par le CLUSIF en 2008 (Clusif, 2008) montre que 73% des entreprises de plus de 200 salariés estiment lourde de conséquences une impossibilité de moins de 24 heures de leurs outils informatiques. Cette dépendance à l'informatique est d'autant plus aiguë dans le domaine de l'IE où toute l'activité de veille est complètement automatisée.

Une activité de veille dans un SIE, qu'elle soit externalisée via un fournisseur de service ou menée en interne, se déroule en plusieurs étapes, chacune exigeant une forme de sécurisation spécifique. La première étape, la collecte des données, exige un service de sécurité en matière de disponibilité, d'intégrité et de confidentialité. Toute altération à ce niveau implique des prédictions erronées et c'est tout le sens de l'IE qui est remis en cause. La deuxième étape est celle du traitement des données collectées. Elle exige un service de contrôle d'accès et met en œuvre des mécanismes d'authentification/autorisation afin de limiter l'accès aux personnes mandatées pour la réalisation des différentes tâches. Cette étape est nécessaire par exemple pour la préservation du secret du métier. Enfin, l'ultime étape est celle de la restitution de l'information jugée 'stratégique' aux clients. Elle exige une authentification mutuelle (plateforme/décideur) et une intégrité des données.

Les risques de sécurité sont encore plus importants pour une plateforme collaborative où les veilleurs, organisés en réseaux, peuvent travailler sous forme de forum. La gestion des droits d'accès des utilisateurs distants et accédant à des fonctionnalités réparties est assez laborieuse, car dans la plupart des cas, l'architecture ne reflète pas la politique de sécurité (Talbot et al., 2010). La solution est de proposer un système de sécurité "transparent", car s'il est compréhensible pour l'attaquant (comprendre un système de sécurité ne signifie pas forcément déceler une faille), il est tout d'abord compréhensible à l'administrateur de sécurité, une condition indispensable pour qu'il puisse monter la garde. La figure 41 montre quelques problèmes de sécurité pouvant compromettre l'activité de veille dans une plateforme d'IE. Ces problèmes sont à la fois externes à la plateforme (attaque de la part des concurrents par exemple), mais aussi internes à la plateforme (séparation des tâches entre les différents rôles impliqués, minimum de privilèges accordés à chaque utilisateur, éviter les conflits d'intérêt, etc...).

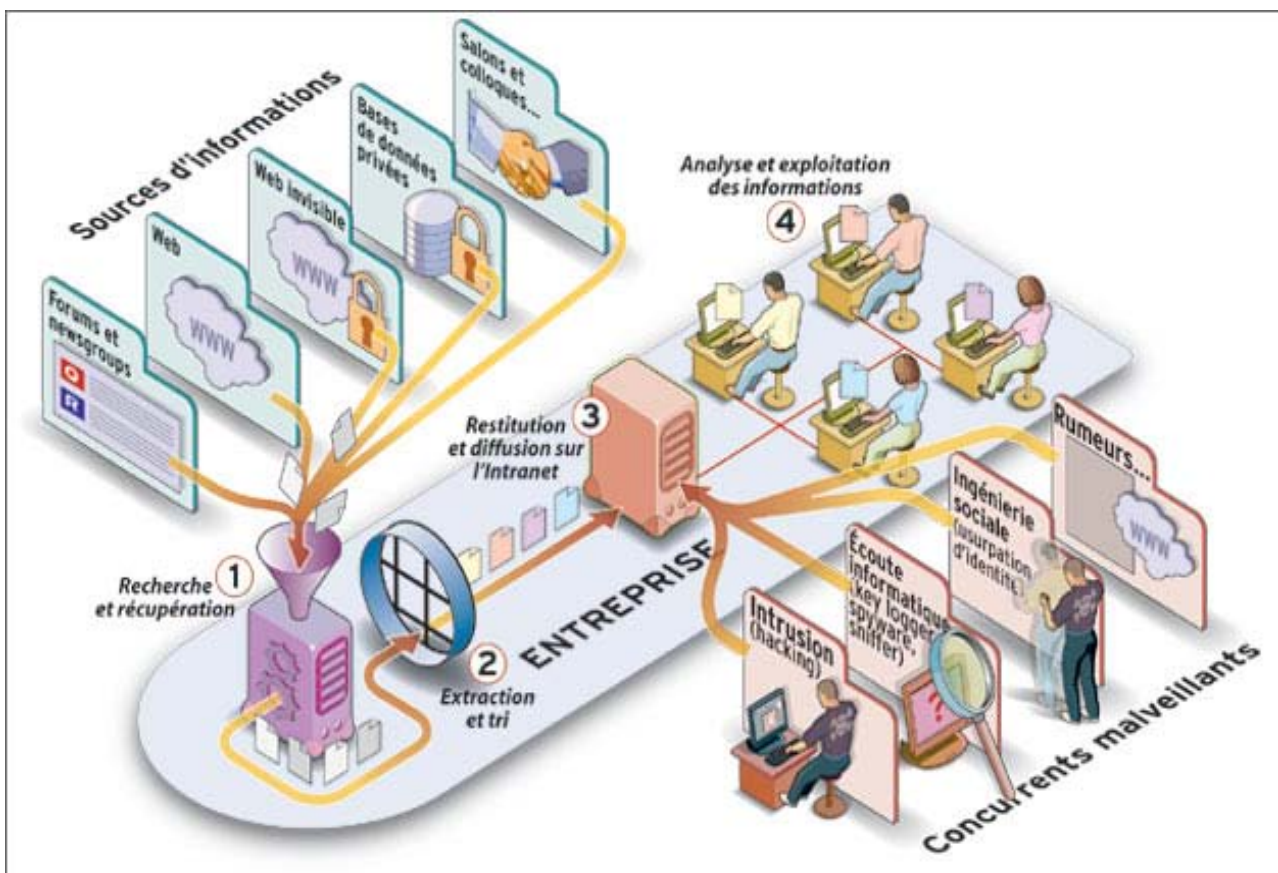


Figure 41. Interactions dans une plateforme d'IE (Bichard, 2005)

➤ **Risques liés à l'importation des données pour le traitement et l'analyse**

Les données sont importées via des outils qu'on achète ou qu'on développe chez soi tels que les méta-moteurs de recherche, les web crawlers, etc... pour surveiller tout type d'information sur le Web visible (forums de discussion, newsgroups, blogs, newsletters, Intranets, RSS, Blog, ...) et invisible (bases de données en ligne, sites nécessitant une authentification, moteurs de recherche, réseaux sociaux, etc.).

➤ **Risques liés à l'utilisation de logiciels de collecte**

Il y a des précautions à prendre lorsque ces outils sont des open sources qu'on doit intégrer à la plateforme. On doit s'assurer en recompilant soi-même le code avant son intégration que ces outils ne sont pas en fait des "malwares" destinés à renvoyer en retour le résultat des analyses des données qu'elles importent. "On ne peut avoir totalement confiance dans un code qu'on n'a pas tout à fait développé soi-même" (Talbot et al., 2010).

➤ **Risques liés aux sources d'information**

Tous les sites web ne sont pas authentiques : certains sont des miroirs qui redirigent le visiteur vers un site dont les services sont différents de ceux qu'ils affichent. Pour collecter des informations uniquement à partir de sources fiables, il faut se baser sur les certificats de la part d'organisation d'approbation de sites web qui vérifient entre autres, qu'un site Web possède bien une déclaration de confidentialité.

➤ **Risques liés aux aspects juridiques**

L'IE est complètement distincte de l'espionnage, l'activité de veille doit se faire d'une manière légale et une attention particulière doit être accordée à la préservation de la vie privée. Aux USA, les activités de veille stratégique et concurrentielle sont assujetties à des lois fédérales qui expliquent entre autre les modalités sous lesquelles une veille peut être menée (Cook et Pfleeger, 2010). La politique de sécurité au niveau de la plateforme doit donc spécifier à l'avance le contenu 'légalement' accessible pour une activité de veille. Par exemple, le système doit être en mesure de refuser les permissions à un veilleur s'il tente d'importer des données protégées car dans ce cas, les conséquences peuvent aller de la dégradation de la réputation à un procès intenté contre la plateforme. Le mieux est d'inviter les utilisateurs à adhérer à un référentiel de bonnes pratiques pour encadrer le comportement de chaque usager dans les différentes situations dans l'entreprise.

➤ **Risques liés aux procédures de traitement et d'analyse des données**

La perte ou la destruction d'informations sensibles est dans 80 % des cas le fait de maladresses internes (SCIE, 2009). Les dommages peuvent être accidentels ou intentionnels. A la différence des attaques provenant de l'extérieur et opérées par des utilisateurs non autorisés à accéder au système, ces vulnérabilités sont liées à l'utilisation abusive des privilèges par des personnes autorisées et correctement identifiées. Les travaux du projet MITRE (Caputo et al., 2009) concernant l'utilisation abusive des permissions légalement acquises ont conduit à proposer des techniques pour palier à ce problème.

➤ **Risques liés à la livraison de l'information**

Le résultat des traitements – la fameuse information stratégique – doit être livré au client ayant fait appel à une activité de veille stratégique. Deux méthodes permettent d'arriver à cette fin :

- La stratégie PULL : le client va chercher les informations sur le serveur. Dans ce cas, un portail internet de la plateforme doit fournir les résultats moyennant une authentification de la part du client. L'essor du commerce électronique ces dernières années a fourni des avancées considérables en matière de sécurité qui peuvent être réutilisées dans les portails internet des plateformes d'IE.
- La stratégie PUSH : le serveur envoie l'information sous forme d'alerte au client sous forme de message dans un réseau informatique (ordinateur relié à internet) ou télécom (smartphone).

Notre Système d'intelligence économique XPlor EveryWhere envoie les résultats au client dans son téléphone portable. Ces équipements (pour la plupart des smartphones), sont à considérer comme des ordinateurs et doivent donc être équipés de systèmes de protection puissants contre les attaques car désormais, ils ne sont pas seulement des périphériques d'affichage, mais contiennent des informations stratégiques. En effet, c'est grâce à leur faculté d'équipements réseau ouverts, programmables et pouvant fournir divers services de type PC, comme la messagerie, les services d'alertes et la navigation sur le Web qu'ils sont retenus pour la livraison de l'information. (Shabtal, 2010) décrit les risques encourus par l'utilisation de ces équipements en expliquant le mode d'exécution des processus et présente un Framework qui peut être utilisé pour la sécurité du terminal (téléphone) du client.

3.4.2. Approche proposée pour le contrôle d'accès à un SIE

Face aux risques de sécurité énumérés dans la section 3.4.1, rares sont les SIE qui sont dotées d'un service de sécurité qui s'intègre dans un projet digne de ce nom et qui respecte une politique de sécurité préétablie. Les réponses à une attaque de sécurité sont plutôt traitées au grès de leur apparition dans le dysfonctionnement du système (Hole et Netland, 2010).

Pour faire face à cette lacune, nous proposons une approche de contrôle d'accès d'une plateforme d'IE qui s'étale sur tout le cycle de vie de l'IE. Pour respecter l'aspect "générique" de l'approche (afin qu'elle puisse être intégrée dans toute plateforme d'IE), nous proposons que la sécurité du traitement des données se base sur le standard de contrôle d'accès RBAC (Role-Based Access Control) (ANSI, 2004).

➤ Politique de sécurité dans un SIE

La politique de sécurité décrit d'une manière exhaustive l'ensemble des dispositifs et des procédures techniques, administratives, juridiques qui doivent être mis en place pour travailler en toute confiance aussi bien au sein de la plateforme qu'avec les collaborateurs. Le Service de Coordination à l'Intelligence Economique (SCIE) (Caputo et al., 2009), rattaché aux ministères français de l'Economie et du Budget a élaboré un guide de bonnes pratiques en matière d'IE sur lequel nous nous basons, - en y introduisant les principes de contrôle d'accès les plus indispensables - pour formuler une politique de sécurité applicable à une plateforme d'IE. Nous l'exprimons d'une manière formelle dans ce mémoire (des travaux antérieurs que nous avons menés donnent une idée sur la formalisation d'une telle politique (El Bakkali et Hatim, 2009). Un administrateur de sécurité est chargé de sa mise en œuvre et d'observer les modalités selon lesquelles chaque étape du cycle de l'IE est réalisée conformément à ses dispositions.

➤ Au niveau de la collecte et de l'importation des données

- Application du principe SOD (Separation of Duty ou Séparation des tâches) (Botha et Eloff, 2001) pour éviter les conflits d'intérêt.
- Application de la cardinalité : pour les sujets particulièrement sensibles à diffusion restreinte, limiter le nombre de collaborateurs impliqués dans la collecte.

- Pour la collecte d'informations qui ne doivent en aucun cas être divulguées (même pour obtenir une information en retour), il faut les scinder en plusieurs sujets et diversifier les contacts.
- Après la collecte, supprimer régulièrement les cookies de l'ordinateur et déconnecter la messagerie.

➤ **Au niveau du traitement et de l'analyse des données**

- Identification des personnes qui doivent avoir accès à l'information
- Identification des rôles joués par chaque utilisateur.
- Identification des permissions associées à chaque rôle.
- Identification de chaque tâche associée à chaque rôle
- Identification des conflits entre les entités (rôles, utilisateurs, tâches)
- Identification des ressources
- Protection du corpus
- Respect du principe LP (Least Privilege ou minimum de privilèges) (Buyens et al., 2009) pour accorder à un utilisateur le plus petit ensemble de privilèges (ou permissions) qui lui sont nécessaires dans l'accomplissement de la tâche qu'il s'apprête à exécuter et non tous les privilèges liés à son rôle dans l'entreprise.
- Etablissement d'un référentiel d'historique des accès de tous les rôles, utilisateurs et tâches.

➤ **Au niveau de la restitution de l'information aux décideurs**

- Classification de l'information en fonction de son degré de sensibilité (générale, restreinte, confidentielle, etc...). Les informations confidentielles sont celles dont la divulgation procurerait un avantage à la concurrence ou aux partenaires ou réduirait l'avantage dont dispose l'entreprise (R&D, travaux d'innovation, savoir-faire technologique, contenu d'offres commerciales, structure des comptes, fichiers clients, projets de développement, fonctionnement de l'entreprise...)
- Définir les supports de diffusion de l'information (réunions, compte-rendus, messagerie électronique, etc...)
- Identification des clients qui doivent avoir accès à l'information.

3.4.3. Mise en œuvre dans un cas pratique : Sécurisation de XPlor EveryWhere

Le prototype XPlor EveryWhere est alimenté par des données relationnelles issues de la plateforme Tétralogie. Cette dernière permet d'effectuer des analyses stratégiques globales sur des données textuelles ou factuelles provenant de bases bibliographiques en ligne, sur CD-ROM, d'Internet ou de toute autre source informatisée, presse etc. Par l'intermédiaire de méthodes statistiques descriptives et exploratoires des données, XPlor EveryWhere fait apparaître, dans des temps très courts, de nouvelles connaissances stratégiques comme l'identité des acteurs, leur notoriété, leurs relations, leurs lieux d'action, leur mobilité, l'émergence et l'évolution des sujets et des concepts, la terminologie, les domaines porteurs.

L'approche adoptée, pour le développement du prototype XPlor EveryWhere, permet de combiner les techniques d'extraction de connaissance à partir des données textuelles et les techniques de stockage,

d'analyse et de visualisation des données relationnelles. Chacune de ces techniques est vue comme un composant aux fonctionnalités précises et délimitées. Plus simples à développer, plus robustes et testés dans des contextes différents, ces composants peuvent s'assembler de plusieurs manières pour créer ainsi des applications variées et adaptées aux besoins des utilisateurs.

➤ Contrôle d'accès aux données et aux traitements dans Xplor EveryWhere

Nous adaptons le standard RBAC (Role-Based Access Control) pour mettre en œuvre la politique de contrôle d'accès. Ceci est nécessaire pour une éventuelle extension en cas de collaboration avec d'autres plateformes d'IE qui respectent ce standard. Le modèle de contrôle d'accès n'a pas besoin d'être remis de fond en comble. Un framework construit autour des concepts de ce standard assure la gestion de tous les accès à la plateforme.

➤ Modélisation des processus de la plateforme Xplor EveryWhere

Pour une visibilité complète et une meilleure maîtrise des flux d'information pour l'administrateur de sécurité, nous représentons dans la figure 42 tous les processus métiers de Xplor EveryWhere par un workflow (un ensemble de tâches reliées par une relation de dépendances et de succession).

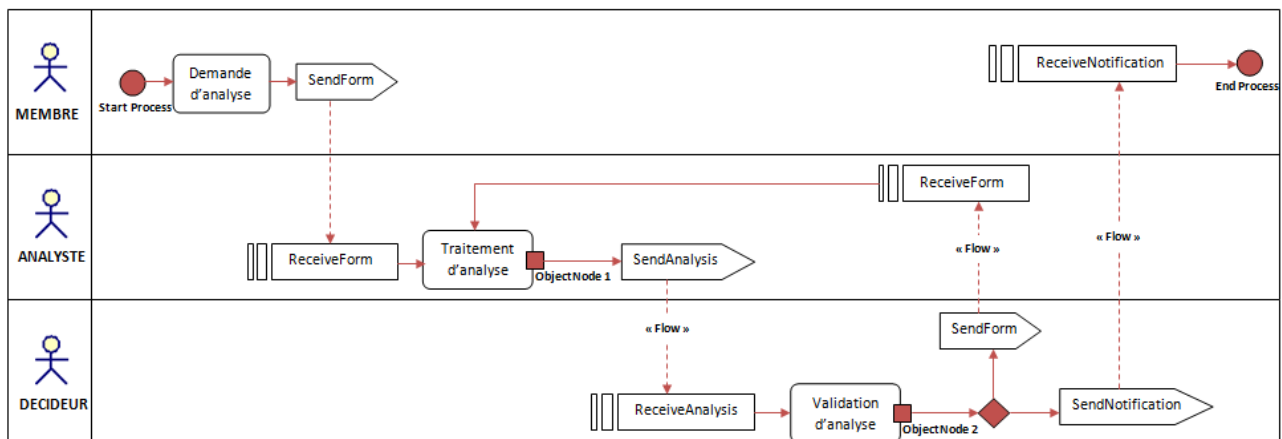


Figure 42. Représentation du workflow Xplor

3.4.4. Application du modèle de contrôle d'accès RBAC.

Adopté comme une norme ANSI / INCITS en 2004, RBAC (Role-Based Access Control) est le modèle de contrôle d'accès le plus répandu dans les systèmes informatiques (ANSI, 2004). Dans ce modèle, les permissions sont accordées aux rôles joués par les utilisateurs plutôt qu'aux utilisateurs eux-mêmes. Puisque dans une entreprise on dénombre quelques rôles alors qu'on a plusieurs milliers d'employés, l'administration de la sécurité se trouve largement facilitée.

Définition. Les principales entités dans le RBAC sont les utilisateurs, les rôles, les autorisations et les sessions. Nous définissons :

- U = Ensemble des utilisateurs, $\{u_1, u_2, \dots, u_k\}$. Un utilisateur peut être un humain ou un processus automatique.
- R = Ensemble des rôles, $\{r_1, r_2, \dots, r_l\}$.

- $P = L$ ensemble des autorisations, $\{p_1, p_2, \dots, p_m\}$.
- $WS = \{ws_1, ws_2, \dots, ws_n\}$, est l'ensemble des sessions.

L'administrateur de sécurité associe ces différentes entités, pour gérer les tâches des utilisateurs et les permissions. Nous proposons deux fonctions :

- $UA \subseteq U \times R$: la relation utilisateur – rôle (a many-to-many user-to-role assignment relation).
- $PA \subseteq P \times R$: la relation autorisation – rôle (a many-to-many permission-to-role assignment relation).

➤ Etablissement des différents rôles et les tâches associées

Concept central du standard RBAC, les rôles décrivent les positions administratives et opérationnelles que remplissent les acteurs impliqués dans le fonctionnement d'une plateforme d'IE. Ces rôles auxquels on associe les privilèges d'accès aux ressources sont établis en fonction des besoins du métier de l'IE et les utilisateurs les utilisent pour accéder à la plateforme. Il faut noter que l'accès au système n'est possible que sur la base du rôle joué pour exécuter une tâche. Le tableau 1 répertorie les différents rôles dans XPlor ainsi que les tâches correspondantes.

Rôle	Tâche
Administrateur sécurité	Protéger le système informatique et les données sensibles
Collecteur-Veilleur	Collecter les données au moyen de : <ul style="list-style-type: none"> - Flux Rss : S'informer de l'actualité sur les sites sélectionnés sans avoir à se connecter en utilisant des lecteurs de flux Rss - Les agents d'alerte : Détection de changements sur une page web - Les requêtes : Recherche à partir de mots-clés dans les moteurs de recherche - Les newsletters
Analyste-Veilleur	<ul style="list-style-type: none"> - Trier les informations et ne retenir que celles qui sont pertinentes - Appliquer les techniques de data-mining pour extraire des informations utiles - Organiser, structurer, hiérarchiser et rapprocher les informations - Interpréter et synthétiser les principaux résultats de l'analyse
Administrateur	<ul style="list-style-type: none"> - Valider la crédibilité de la source d'information de la part du collecteur - Valider les informations retenues : évaluer leur exactitude et leur fiabilité
Auditeur	<ul style="list-style-type: none"> - S'interroger sur la conformité des procédures suivies dans les différentes activités avec la politique de sécurité - Tester les résultats
Stagiaire	Remplir l'une des tâches du rôle auprès duquel il est admis

Tableau 6. Les différents rôles et tâches associés à l'utilisateur de XPlor

Une autre méthode proposée se base sur les diagrammes des cas d'utilisation. En fait, selon l'approche proposée par S. Supakkul et L. Chung (Supakkul, 2005), le méta-modèle de cas d'utilisation contient quatre méta-éléments qui sont : l'acteur, le cas d'utilisation, l'association entre acteurs et cas d'utilisation et, enfin, le système. Pour ces éléments, nous devons définir les exigences de sécurité associées. Par exemple, dans la Figure 43, la validation d'une analyse stratégique a besoin de trois services de sécurité : la confidentialité, l'intégrité et la non-répudiation. À l'aide de cette approche, nous définissons les services de sécurité nécessaires à chaque cas d'utilisation.

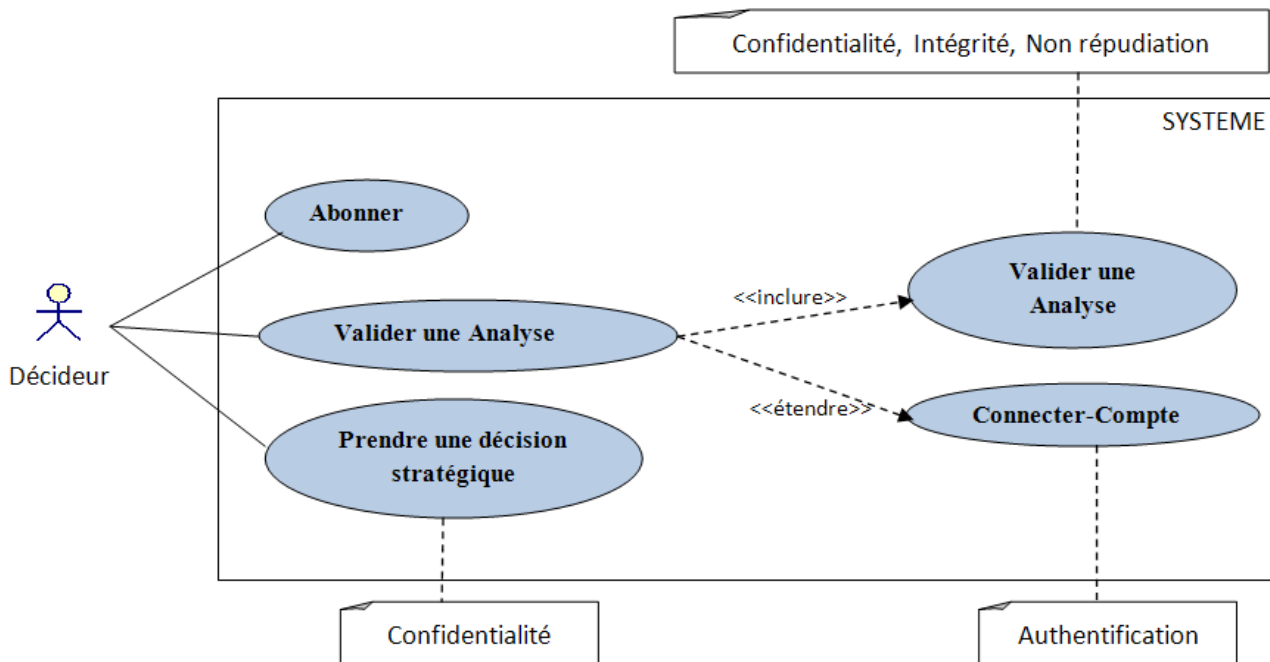


Figure 43. Service de sécurité pour un cas d'utilisation

➤ Les permissions aux objets à protéger

D'une manière logique, un rôle n'est rien d'autre qu'un ensemble de permissions. Les permissions accordées à un rôle sont des opérations portant sur des objets. Concernant la plateforme XPlor Everywhere, les objets à protéger sont les suivants

- Le portail web de la plateforme
- Les documents internes : tableaux de bord, documents de gestion, guides de procédure interne, compte-rendus de réunion, de visites, les bilans ...
- Les documents externes : courriers et courriels reçus, CV, etc...
- Les logiciels et applications : méta-moteur de recherche, explorateur de web, outil de data-mining, etc ...
- Le corpus

Le corpus est l'ensemble des données collectées pour l'activité de veille au sein de la plateforme. Les données peuvent provenir de la presse en ligne, de sites internet de clients et de fournisseurs, de forums et de blogs, de bases de données gratuites ou payantes, de réseaux sociaux sur Internet, etc.

➤ L'infrastructure de sécurité

- Outils cryptographiques : Pour chiffrer les données sensibles lors des échanges, notamment lors de la livraison des informations aux décideurs
- Référentiel des historiques des accès : Pour la traçabilité de toutes les activités menées. A tout moment, pour un utilisateur donné, ce référentiel renseigne sur les rôles activés, les tâches exécutées, ainsi que les durées nécessaires.
- Gestionnaire de worklist : Pour une session donnée (chaque fois qu'un utilisateur active un rôle pour exercer une tâche donnée), une worklist correspondante est générée conformément à la politique de sécurité et sur la base du référentiel des historiques. Elle comporte la succession de toutes les actions possibles pour l'utilisateur en question et permet ainsi à l'administrateur d'octroyer ou de refuser l'accès à chaque étape.
- Référentiels.

3.4.5. Application des ontologies de sécurité

En raison de la diversité des acteurs, des rôles, des contraintes, des temps d'exécution, des tâches complexes de la sécurité, un modèle de base de connaissances doit être utilisé pour exprimer ces conditions. Ces solutions sont adaptées à un usage spécifique. Ce que nous ciblons c'est le partage et la mise à disposition des concepts qui seront utilisés de façon constante à travers la plateforme d'IE.

Un patron de connaissance formelle des concepts de la sécurité mobile fournit une base solide pour proposer une approche applicable aux plateformes Web-Mobile. C'est pour cette raison que nous avons adopté le formalisme des ontologies pour exprimer les connaissances requises dans le domaine de la sécurité mobile. Cette ontologie a un double objectif : tout d'abord la mise en place d'une connaissance formelle sur la sécurité dans le cadre des applications mobiles, puis l'identification des contre-attaques possibles qui devraient être appliquées pour la sécurisation de la plateforme d'IE. Les différents concepts sont reliés sémantiquement.

Les ontologies de sécurité sont nombreuses dans la littérature (Ahmed et al., 2007)(Herzog et al., 2007) (Kim et al., 2005)(Raskin et al., 2001), chacune étant définie pour un type spécifique d'utilisation. Les importations et la réutilisation sont les bases du partage des connaissances pour les ontologies, en fait, nous avons partiellement fondé notre ontologie sur les travaux de Herzog, Shahmehri et Duma (Herzog et al., 2007), qui traitent de la sécurité des systèmes d'information, ainsi que sur les travaux de Beji et El Kadhi (Beji et El kadhi, 2009), qui proposent une ontologie de sécurité pour les applications mobiles. Ces auteurs ont décrit les concepts et les relations les plus significatifs dans le domaine : de la vulnérabilité, des menaces, des biens, des contraintes, des acteurs, des mécanismes, des ressources et des services. Leur proposition est sémantiquement composée de trois sous-ontologies :

- Active-Vulnerability-Threat Ontology (AVTO).
- Mobile Profile Ontology (MPO)
- Defense Mechanism Ontology (DMO)

La Figure 44 donne une vue d'ensemble de l'ontologie principale avec certaines relations sémantiques entre les sous-ontologies proposées par (Beji et El kadhi, 2009). Ces auteurs ont fait cette distinction afin de permettre la réutilisation et le partage.

La sous-ontologie MPO définit les fonctionnalités d'utilisation possibles avec les ressources associées. Chaque mobile utilisant certaines ressources, la construction de MPO se base sur une classification des

ressources disponibles pour les différents cas d'utilisations. La gestion Vulnérabilité-Menace est largement adaptée à une sous-ontologie (AVTO), où les classes supérieures et les relations principales sont les mêmes et seulement les sous-classes des vulnérabilités diffèrent d'un contexte à un autre. Enfin, la sous-ontologie DMO gère les services de sécurité et les mécanismes de sécurité associés. Notre objectif est l'enrichissement de cette ontologie avec des vulnérabilités particulières, pour la protection des données stratégiques et la création des relations sémantiques entre l'ontologie des acteurs ainsi que les différentes tâches possibles dont l'exploitation de la plateforme d'IE.

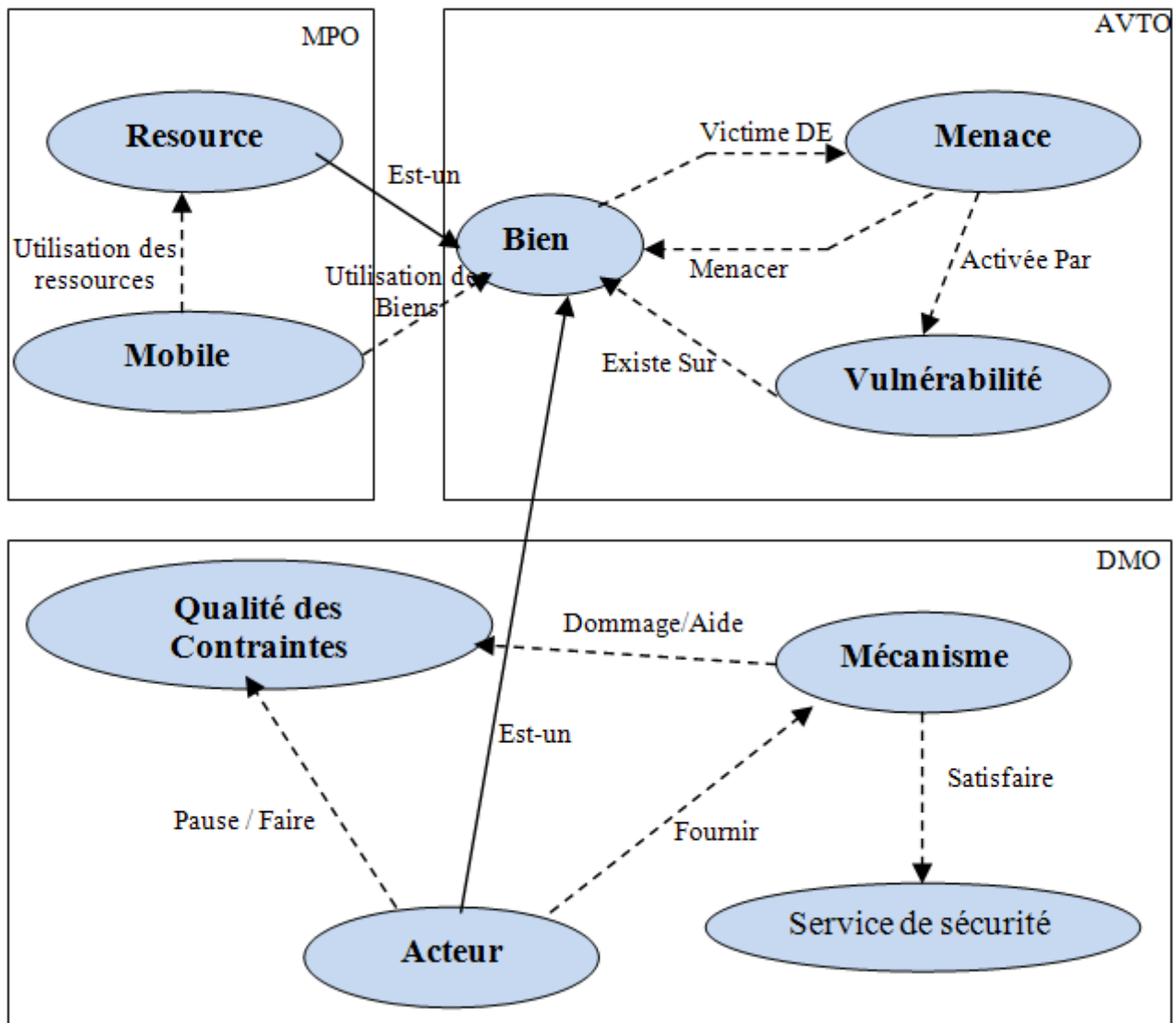


Figure 44. L'ontologie de sécurité (Beji et El kadhi, 2009)

3.5. Synthèse

Le système d'intelligence économique XPlor EveryWhere offre l'interactivité aux utilisateurs nomades, en facilitant la navigation dans la connaissance grâce au modèle d'analyse multidimensionnelle « XPlor », qui couvre les différentes étapes de la fouille de texte : analyse de besoins, recherche, prétraitement, traitement, analyse et visualisation des données.

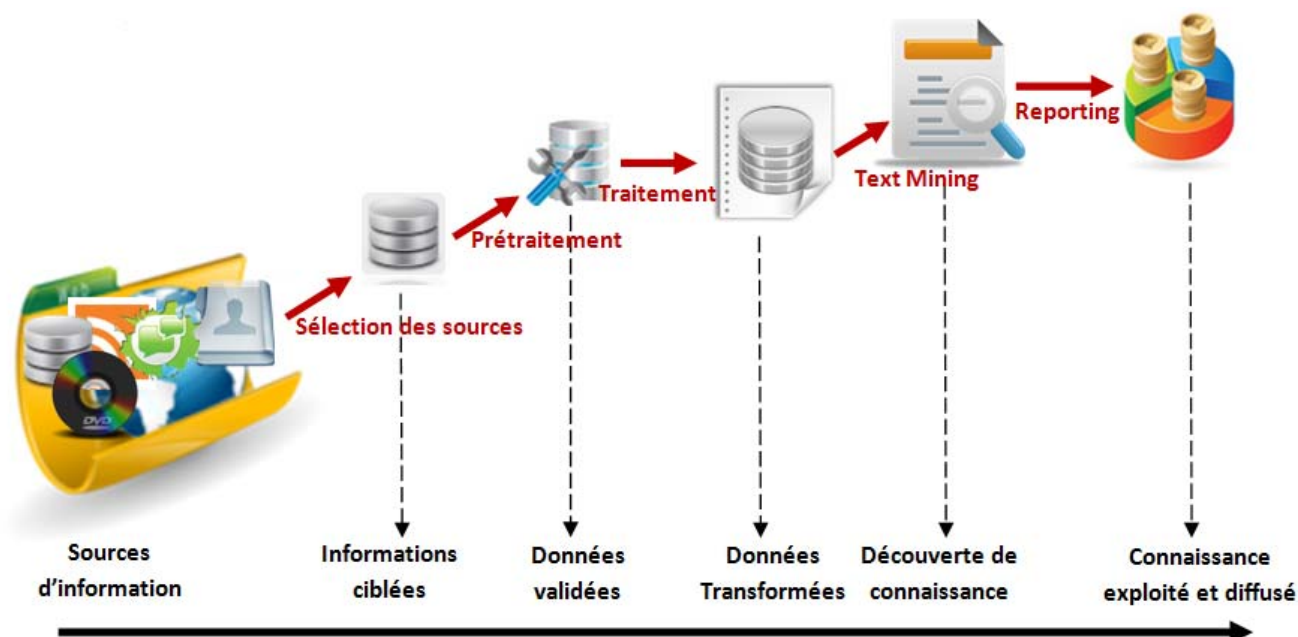


Figure 45. Les étapes du modèle d'analyse multidimensionnelle XPlor

Le modèle « XPlor » s'appuie sur trois principaux modèles : Le **modèle de présentation multidimensionnelle** des corpus textuels, l'objet de la section 3.3, qui nous permet de décrire les relations de dépendance basées sur des mesures de co-occurrences, dans un esprit chronologique, c'est-à-dire la prise en compte des variables temporelles (Date, Année, Mois, Jours). Ce type de présentation synthétise l'ensemble des liens statiques et dynamiques du domaine analysé.

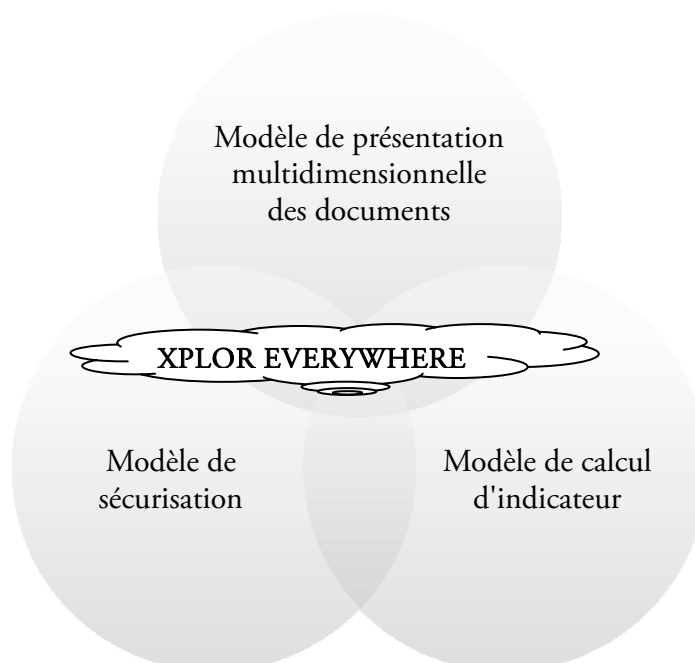


Figure 46. Les sous-modèles du modèle d'analyse multidimensionnelle XPlor Everywhere

Dans le but d'interpréter les différentes relations de dépendances, nous avons définis le **modèle des indicateurs**, traité dans la section 3.4. Le modèle proposé se base sur l'application des techniques de la bibliométrie dans le domaine d'intelligence économique. Il ne se limite pas aux notices bibliographiques types, articles ou brevets et l'application des indicateurs standards de la bibliométrie (impact factor,

indice H, indice G, etc.), mais tend vers l'application de la bibliométrie et la définition d'autres indicateurs de calcul sur d'autres types d'information (les réseaux sociaux, les flux RSS, etc.). Au cours de la section, nous avons relevé quelques indicateurs présentés dans la littérature scientifique, puis nous avons présenté notre indicateur de filtrage. Cependant, il ne faut pas considérer ces indicateurs comme des indices parfaits, ils doivent être vus comme complémentaires.. Grâce à cette complémentarité, le modèle d'indicateurs permet de synthétiser l'environnement informationnel d'une analyse stratégique.

Le calcul de ces indicateurs demande des traitements et des contrôles d'accès aux données stratégiques, ce qui déclenche **le modèle de sécurisation**. Ce modèle consiste à sécuriser un système d'intelligence économique, qui est un problème indécidable, puisqu'on ne peut conclure d'une manière définitive que le système est déjà sécurisé. Dans ce modèle, nous avons fait ressortir les risques de sécurité dans les SIE et le besoin d'avoir en leur sein un service de contrôle d'accès. Sur la base de ces vulnérabilités, les bases d'une approche générique de sécurité sont proposées et mises en pratique dans le SIE XPlor EveryWhere. De plus, nous avons proposé une autre solution de sécurité via les bases de connaissances grâce à la conception d'une ontologie de sécurité en se basant, sur trois ontologies MPO, DMO et AVTO, puis nous introduisons les services de sécurité nécessaires à chaque cas d'utilisation.

Après une présentation de nos contributions, de l'approche théorique de nos travaux de recherche, nous expérimentons les contributions présentées, en évaluant les méthodes selon les critères précédemment retenus, dans le prochain chapitre de ce manuscrit. Une enquête est réalisée auprès d'une population ciblée, constituée d'individus informaticiens, statisticiens, documentaliste, de l'IE et de personnes dont la spécificité n'est pas liée directement aux domaines étudiés dans ce mémoire. Les résultats sont comparés à ceux de notre équipe et un bilan est effectué pour cibler les points forts et les points faibles du système d'intelligence économique proposé.

Chapitre 4.

Expérimentations et évaluation de XPlor EveryWhere

« L'intelligence est le croisement de l'information et de la stratégie. Le prisme est large. Il va du cycle du renseignement - dont la définition « officielle » de l'intelligence économique s'est inspirée - à la manipulation de la connaissance en passant par la désinformation. Dans tous les cas, l'information est au service d'une stratégie : en amont pour définir et comprendre son environnement pertinent, prévenir les risques, détecter les opportunités...; en aval pour décider, leurrer l'adversaire, le paralyser, ... » (Moinet)

4.1.	Introduction	115
4.2.	Les limites de la plateforme Tétralogie	115
4.3.	Les limites des premières versions du prototype Xplor	118
4.4.	XPlor EveryWhere	119
4.4.1.	Accès sécurisé aux analyses stratégiques	119
➤	Gestion de la base de données	121
➤	Alimentation de la base de données	123
➤	Schéma de la base de données	123
➤	Gestion des droit d'accès	124
4.4.2.	Analyse de données	125
➤	Dépôt et suivie d'une étude	125
➤	Analyse descriptive	126
➤	Menu Filtrage	128
➤	Carte géostratégique	130
4.5.	Expérimentations : analyse stratégique sur les scaffolds	131
4.5.1.	Besoin informationnel	133
4.5.2.	Source d'information	135
4.5.3.	Structuration et homogénéisation des données	136
4.5.4.	Représentation multidimensionnelle	137
4.5.5.	Analyse et interprétation	138
➤	Analyse de réseaux des inventeurs	138
➤	Analyse de collaborations internationales	142
➤	Analyse de réseau sémantique	145
➤	Analyse de déposants	146
➤	Synthèse de l'analyse scaffolds	147
4.6.	Evaluation de XPlor EveryWhere	148
4.6.1.	Méthodologie	148
4.6.2.	Population étudiée	149
4.6.3.	Résultats de l'enquête	151
➤	Niveau de difficulté des fonctionnalités XPlor EveryWhere	151
➤	Analyse de l'utilité de XPlor EveryWhere	158
4.7.	Synthèse	160

4.1. Introduction

Ce chapitre présente la démarche de validation scientifique mise en place pour expérimenter le prototype Xplor Everywhere. Nous présentons, en section 4.2, les limites de la plateforme Tétralogie et les limites de la version Xplor proposée par Ilhem Ghalamallah en 2009 dans la section 4.3. Une présentation des fonctionnalités de notre prototype fera l'objet de la section 4.4. Dans la section 4.5, nous présenterons l'expérimentation effectuée pour valider chaque étape de développement du SIE Xplor Everywhere. Cette expérimentation a été effectuée dans le cadre des bureaux d'études du M1-SID (Statistique Informatique Décisionnelle), M2-PINT (Propriété Intellectuelle et Nouvelle Technologie) et DUT info-com GIDO (Gestion de l'Information et du Document), formations pour lesquelles j'ai dispensé des cours à l'université Paul Sabatier – Toulouse III. Les résultats de ces évaluations, section 4.6, nous ont servi dans un premier temps à améliorer l'utilisation de Xplor Everywhere, puis d'envisager de nouvelles perspectives pour nos travaux à venir.

4.2. Les limites de la plateforme Tétralogie

La plateforme Tétralogie (Dousset et Benjamaa, 1988) (Dkaki, 1993) développée au sein de notre équipe SIG-IRIT est dédiée à l'extraction des connaissances à partir de textes. Dans le contexte de la veille stratégique (VS), Tétralogie est un outil particulièrement bien adapté aux analyses macroscopiques, il permet en effet de dégager les signaux forts, les signaux faibles et les tendances à partir d'un corpus de documents collectés sur un sujet précis. L'information élaborée, qui en est issue, représente une synthèse obtenue par diverses méthodes d'analyse de données et diffusée via des visualisations graphiques.

L'approche suivie par Tétralogie est essentiellement basée sur la découverte de connaissance à partir de l'information textuelle de type bibliographique ou de brevets issue de bases bibliographiques en ligne, sur CD-ROM, d'Internet ou de toute autre source informatisée. Par l'intermédiaire de méthodes statistiques, d'analyse exploratoire des données et de cartographie interactive, il fait apparaître, dans des temps très courts, de nouvelles connaissances stratégiques jusque là inexploitées comme l'identité des acteurs, leur notoriété, leurs relations, leurs lieux d'action, leur mobilité, l'émergence et l'évolution des sujets et des concepts, la terminologie, les domaines porteurs, que lire et où publier, avec qui collaborer.

La plateforme est composée de plusieurs modules complémentaires :

- Module de manipulation de corpus permettant d'interfacer les sources d'information (Dousset et Benjamaa, 1988) (Dkaki et al., 1996) (Dkaki et al., 1998) (Mothe et al., 2002) (Dousset, 2009).
- Module de représentation réduite de données, qui permet de structurer selon les objectifs visés les relations (présence- absence, contingence, co-occurrence, co-occurrence multiple, etc.) existantes dans le corpus analysé (Dousset, 1987) (Dousset et Benjamaa, 1988) (Dousset et al., 1989) (Dousset et Karouach, 2005),
- Module d'analyse de données, qui recouvre toutes les techniques de classification, d'association et de séquences (ACP, AFC, CAH, CPP, etc.) (Dousset, 1987) (Dousset et al., 1989) (Dousset et al., 1993) (Dousset, 2006) (Dousset et Karouach, 2007).

- Module de visualisation, qui offre des résultats graphiques sous forme de carte factorielle en 2D, 3D et 4D (Dousset et al., 1991) (Banammar, 1999) (Dousset et Karouach, 2002) (Dousset, 2003).
- Module d'analyse de graphe, qui permet d'explorer l'évolution des matrices par des graphes évolutifs (Karouach, 2003) (Dousset et Karouach, 2005) (Loubier et al., 2007) (Loubier et al., 2008) (Loubier, 2009).

L'architecture modulaire, qui est adoptée, permet une adaptation fine à des besoins spécifiques qui sont essentiellement liés à la taille des entreprises ou des organisations, à leur secteur d'activité, à leur type de clientèle, aux impératifs en termes de coût, de qualification, de réactivité et de vision prospective. Plusieurs niveaux de compétences et de services sont proposés :

- les scénarios d'analyse préfabriqués couvrant la majorité des besoins élémentaires en information élaborée,
- des stratégies d'analyse plus complexes mises au point par des spécialistes pour chaque source spécifique d'information électronique,
- un logiciel interactif pour la réalisation d'analyses spécifiques très détaillées pouvant être conduites de façon collégiale via Internet.

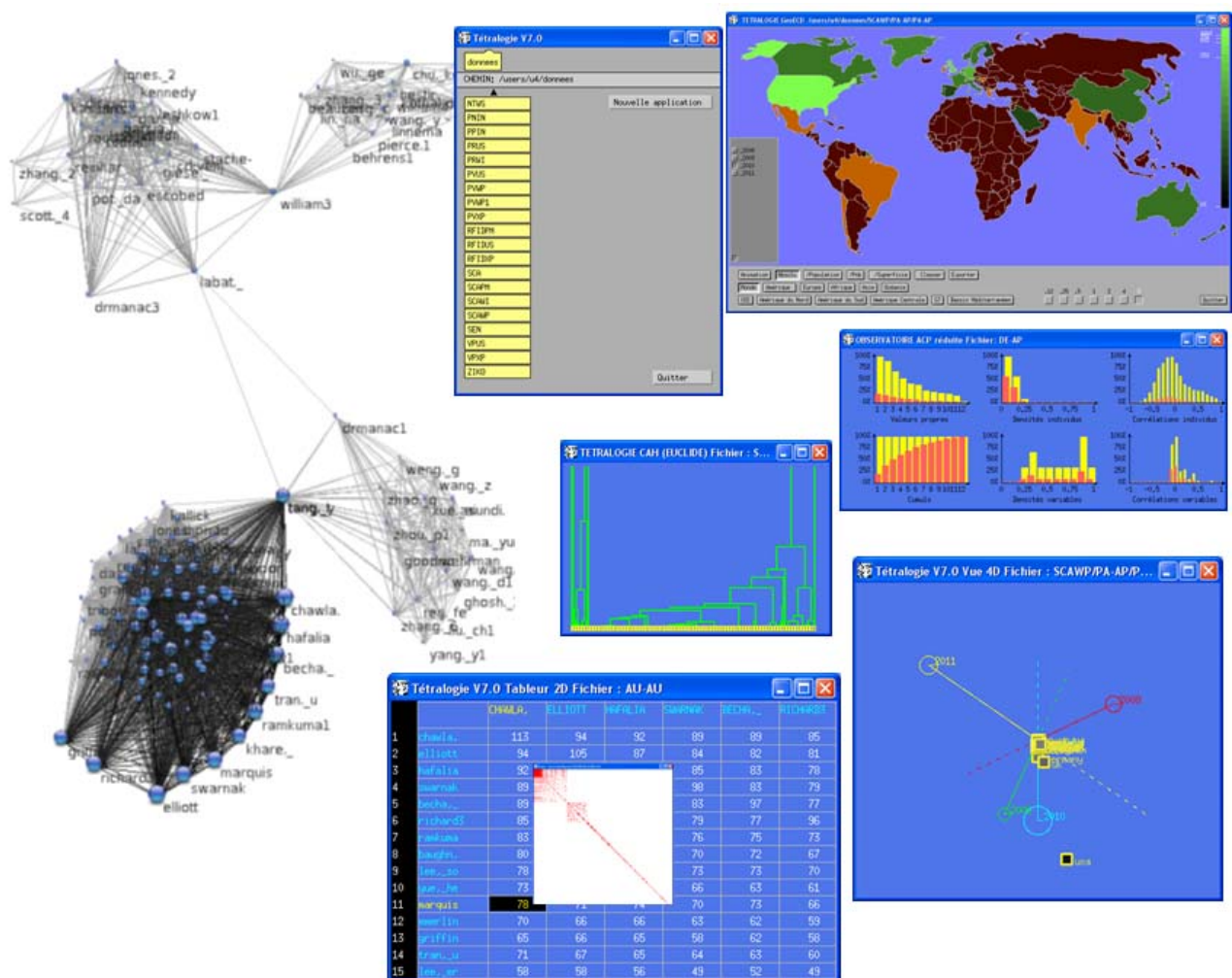


Figure 47. Les différentes sorties de la plateforme Tétralogie

A l'issue des analyses stratégiques réalisées avec Tétralogie, il est apparu que les utilisateurs finaux cherchent, en complément de l'aspect global et stratégique (connaissances générales), des zooms plus précis sur certains points et ce afin de satisfaire leurs besoins spécifiques (connaissances ciblées) d'information sur des éléments qu'ils ont déjà identifiés (concurrence, marchés, nouveaux produits ou procédés, partenaires potentiels, ...) ou afin d'en découvrir d'autres. A posteriori, de nombreux experts ou décideurs sont demandeurs de plus de détails sur les éléments traditionnels de leur environnement, notamment, pour tout ce qui concerne leur vocabulaire spécifique, les acteurs qu'ils côtoient, les marchés qu'ils convoitent, les alliances qu'ils projettent. Le prototype visé doit donc offrir des possibilités d'exploration très fine et des représentations synthétiques de l'environnement et de son évolution, des informations recueillies et des nouvelles connaissances déduites. Il doit permettre une automatisation générique des indicateurs permettant aux décideurs d'appréhender les données massives qu'ils ont à traiter. En amont, il doit assurer le recoupement des informations collectées, la recherche, la sélection et le filtrage de l'information élaborée disponible dans des bases spécialisées internes, externes et sur Internet, ainsi que la prise en compte des données informelles collectées, sur le terrain, par les différents acteurs de la veille. En aval, pour la communication et la restitution des résultats, il doit privilégier l'ergonomie dans les fonctions de présentation, d'exploration, de navigation et de synthèse.

Tétralogie gère des fichiers contenant l'information brute pour en extraire, des matrices de croisement, ces matrices de grande taille contiennent plus de zéro que de valeurs non nulles (matrices creuses), ce qui limite un peu la puissance du système, soit en volume d'espace en mémoire et sur disque soit au niveau de la vitesse de lecture de ces matrices. Pour remédier à ces problèmes, au lieu de générer des matrices de co-occurrences entre tous les items, nous proposons de générer un fichier où ne sont enregistrés que les croisements effectifs entre items, ce qui permet de diminuer l'espace utilisé et d'augmenter la vitesse de lecture et de traitement de ce fichier.

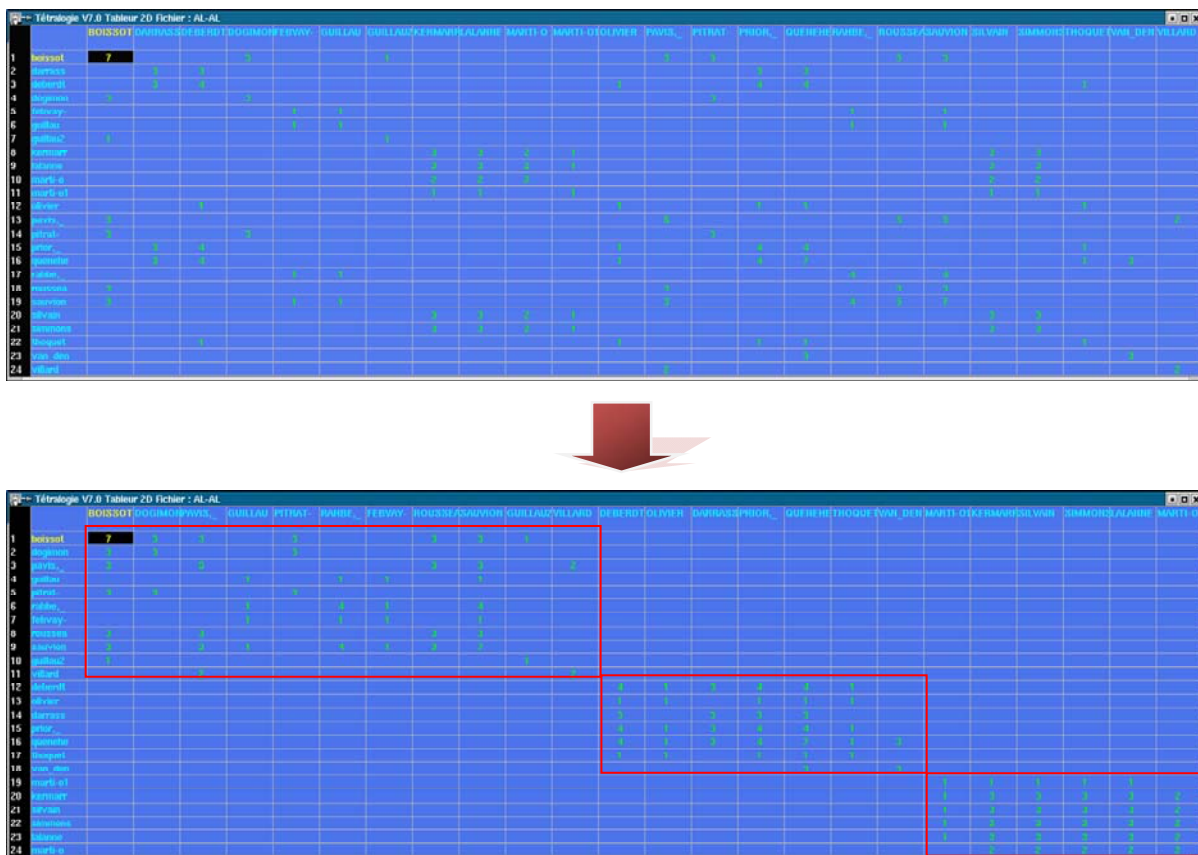


Figure 48. Exemple de matrice de co-occurrence Auteur-Auteur

4.3. Les limites des premières versions du prototype Xplor

Depuis 2001, La première version de XPlor a vu le jour, avec des multitudes d'opérations dont l'objectif est de représenter les relations des unités de même nature ou entre deux unités différentes. Ces deux représentations permettent une analyse visuelle et exploratoire statique des relations explicites, dans une optique de découverte de nouvelles relations implicites. Cependant, la conception de la base de données pose des problèmes majeurs au niveau de la mise à jour, la synthèse oblige à parcourir plusieurs pages - faute de l'architecture conceptuelle-, et on peut noter un manque relatif à l'aspect temporel.

En 2009, une deuxième version est proposée avec une simplification du schéma de la base de données, l'apport d'une troisième dimension temporelle. Les analyses ne sont ainsi plus simplement statiques mais deviennent entièrement dynamiques. La page d'accueil (Figure 49) du Xplor V2 s'affiche au niveau du navigateur de l'utilisateur sous forme d'un ensemble de liens hypertexte. Le menu de la page d'accueil du prototype s'articule autour de quatre composantes principales :

- Informations et connaissances portant sur la Veille et l'IE.
- Classification des liens des principaux acteurs de l'IE.
- Analyse Générale : administration et diffusion des analyses stratégiques du système Tétralogie.
- Analyse Détaillée.



Figure 49. Page d'accueil du prototype Xplor V2

Le passage à l'échelle pose des problèmes à cause de l'architecture de la base de données, puisqu'on est obligé de créer une table par type d'analyse, ce qui implique un changement de la conception de la base à chaque demande d'analyse. La fonction de filtrage proposée ne permet pas de visualiser l'environnement détaillé de l'attribut choisi et on ne peut pas naviguer, ni afficher les différents types de réseaux : le profil de l'attribut, le réseau social, le réseau sémantique, etc.

L'absence d'une approche de sécurisation, malgré les données sensibles traitées dans ce type d'analyse, pose un problème majeur de déploiement.

Il n'existe pas d'aspect relatif à la mobilité pour ce prototype, puisque il est incompatible avec les appareils téléphoniques (Androïde, Apple Store, etc), ce qui présente un problème aux décideurs nomades dans leurs déplacements.

4.4. XPlor EveryWhere

XPlor EveryWhere restitue les études stratégiques macroscopiques de l'outil Tétralogie, sous forme d'études microscopiques. Il facilite la navigation dans l'environnement relationnel d'un élément connu (entreprise, centre de recherche, chercheur, inventeur, mot clé, etc.). Il devient ainsi possible de retrouver, à partir de cet élément, toute ou partie de l'information qui lui est connexe (alliances, concurrents, équipes, collaborations, concepts, émergences, mot associés, etc.), en toute sécurité.

Ce SIE offre à l'utilisateur un espace interactif et collaboratif, facilitant la compréhension et la découverte de l'information pertinente. Ce système répond aux différents points suivants :

- Avoir une interprétation (reporting) des connaissances claire et compréhensible.
- Consulter des informations à jour, en accédant au serveur de base de données stratégique en temps réel, lui-même alimenté quotidiennement par des veilleurs.
- Faire remonter des informations « terrain » lors des salons, de visites en clientèle ou à l'issue de réunions.
- Demander des renseignements spécifiques en urgence qui seront mis en ligne par des veilleurs.
- Permettre aux utilisateurs de proposer des explications ou de prendre des décisions,
- Permettre la surveillance et l'anticipation des évolutions.

Grâce à l'évolution de la technologie et de la communication, un tel système permet de gagner en efficacité et en réactivité. A tout moment, il est possible d'accéder à toutes les informations stratégiques utiles au décideur nomade, lui-même pouvant faire remonter très vite l'information « terrain » qui peut éventuellement déclencher d'autres analyses stratégiques, en toute sécurité grâce à l'approche de sécurisation implémenté dans le système.

4.4.1. Accès sécurisé aux analyses stratégiques

Le déploiement d'un SIE nécessite une sécurisation bien maîtrisée, afin de protéger les données sensibles ainsi que les traitements effectués à chaque instant. Nous adaptons l'approche de sécurisation présentée dans la section 3.4 pour mettre en œuvre la politique de contrôle d'accès, comme nous la présente la figure 50. Les utilisateurs doivent saisir le login et le password, ainsi que l'approche attribut les analyses stratégique propre à chaque utilisateur.



Figure 50. Page d'accueil du prototype Xplor EveryWhere

La sécurisation du système est gérée par le service administration (figure 51), qui propose deux fonctions : gestion de la base de données et gestion des comptes utilisateurs.

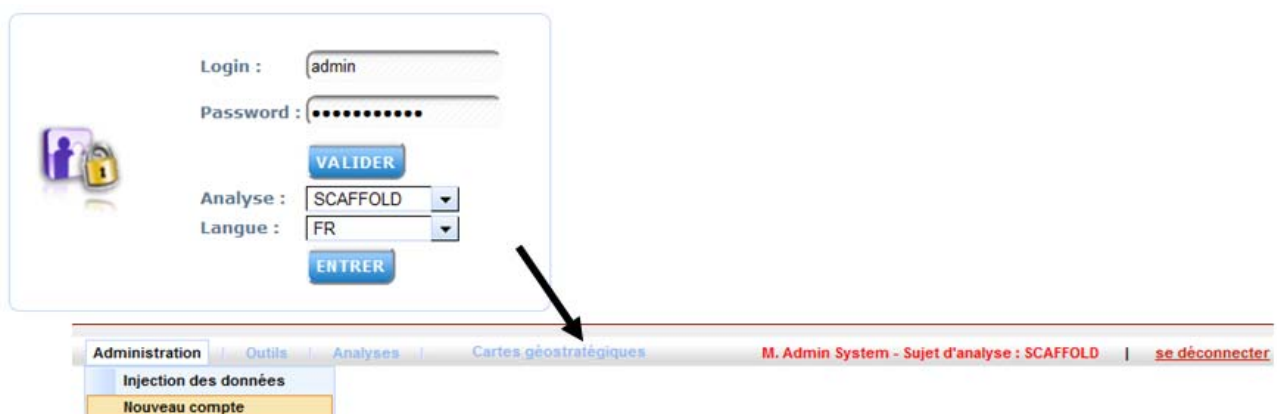


Figure 51. Service Administration du prototype Xplor EveryWhere

➤ Gestion de la base de données

Le prototype XPlor EveryWhere est alimenté par des données relationnelles issues de la plateforme Tétralogie. Cette méthode présente de nombreux avantages :

- Compatibilité totale avec l'analyse par Tétralogie,
- Ne nécessite pas de système d'extraction complémentaire,
- Permet de compléter des analyses déjà prêtes,
- Renforce par la navigation la pertinence du rapport d'analyse.

Elle est essentiellement destinée à compléter les analyses Tétralogie en permettant à l'utilisateur final de naviguer à sa guise, afin de préciser certains passages de l'analyse et de les ramener dans le contexte et l'environnement requis. Pour certaines analyses généralistes disponibles en ligne, cette approche permet à chacun de compléter son interprétation par des conclusions toujours un peu stéréotypées de ce type de démarche globale. Pour des analyses plus pointues sur des sujets très précis, la taille plus réduite des dictionnaires utilisés permet de conserver toute l'information utile, notamment au niveau des champs sémantiques. Dans ce cas, cette approche nous semble la mieux indiquée.

L'utilisation de la plateforme Tétralogie, offre la possibilité de créer des matrices à trois dimensions de type co-occurrences. Pour traiter et exploiter les résultats de ces matrices dans le prototype Xplor EveryWhere, nous avons redéfini les fonctions de créations des matrices de co-occurrences de Tétralogie. Cette réadaptation intervient principalement sur le format des résultats.

Au préalable, le résultat de la fonction « Co-occurrence-3D », était sous forme d'un fichier texte décrivant un tableau à deux dimensions. Pour chaque valeur prise par l'attribut « Temps », un fichier était généré.

Par exemple, pour la création des matrices de co-occurrence de type (figure 52):

- Auteur-Auteur-Temps
- Auteur-KeyWord-Temps

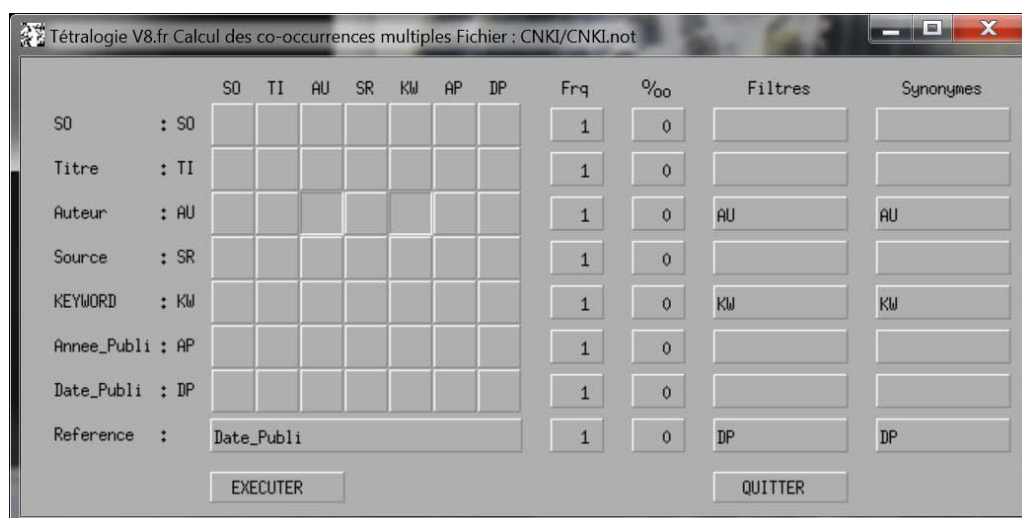


Figure 52. Co-occurrences multiples de Tétralogie

Les résultats obtenus par tétralogie sont décrits dans la figure 53.

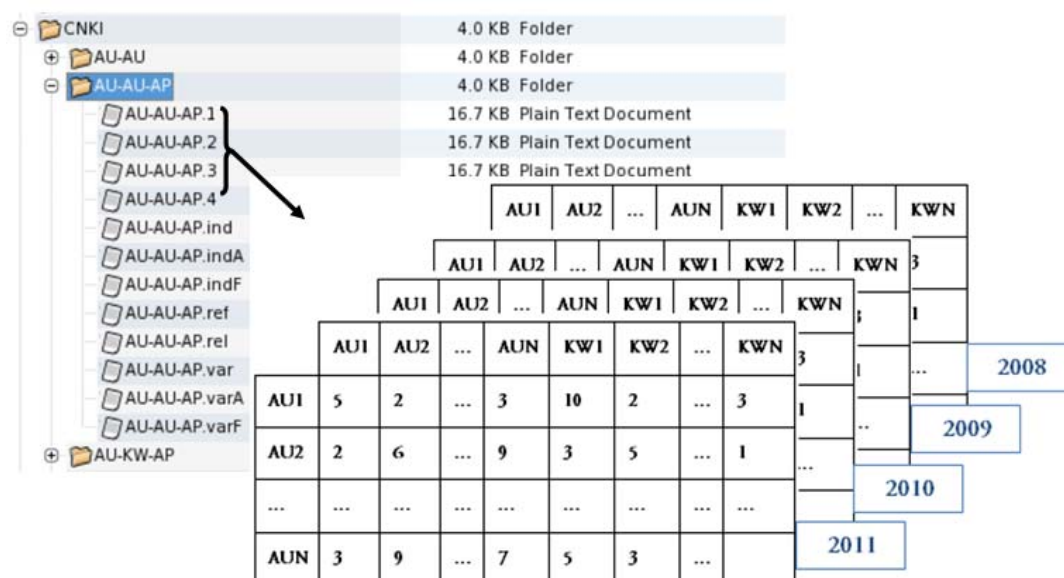


Figure 53. Les résultats Tétralogie de la fonction Co-occurrences Multiples

Ces résultats ne sont pas adaptés à la structure de la base de données du prototype Xplor EveryWhere. Pour y remédier, nous proposons une fonction « TetraToXplor » pour générer trois fichiers tels que :

- Le premier fichier « Champs », décrit les attributs ciblés. Chaque attribut est identifié par un numéro, un code et un type.
- Le second fichier « Items », représente les valeurs de chaque attribut ciblé. Chaque valeur est identifiée par son libellé, son numéro et le numéro de son attribut.
- Le troisième fichier « Co-occurrence », concerne toutes les relations de co-occurrences identifiées dans le corpus ciblé. Chaque co-occurrence est identifiée par le numéro de la valeur du premier attribut, le numéro de valeur du second attribut, la valeur de la co-occurrence et le libellé de l'attribut « Temps » (Attribut1-Attribut2-Co-occurrence-Temps).

Ces représentations nous offrent une nouvelle structure qui permet de décrire de façon synthétique les résultats de la compilation des matrices. Cette structure a été proposée de façon à respecter le modèle de données associé à la base de données du prototype Xplor EveryWhere. Ainsi, les données obtenues sont facilement exportées vers le prototype.

La figure 54 est un exemple de la nouvelle structure des résultats de la fonction « Matrice de co-occurrences à trois dimensions ».

numero_champ	champ	type
1	NO	Numero
2	AU	Auteurs
3	JN	Journaux
4	DP	Date
5	MT	Termes
6	PA	Pays
7	AG	Agents
8	DE	Deposants

Fichier Champs

idCooc	numero_1	numero_2	co_ocurrence	date
1	1	1	6	8
2	2	2	16	8

Fichier Cooccurrence

idItem	nom_long	numero_item	idInnovChamp
1	HOOVER, KENLEY, K.	1	7
2	FOLEY & LARDNER LLP	2	7

Fichier Items

Figure 54. Les résultats de la fonction TetraToXplor

➤ Aliménation de la base de données

Le service administration permet d'alimenter automatiquement les résultats de la fonction TetraToXPlor (figure 55).

Figure 55. Injection des données

➤ Schéma de la base de données

- Module « gestion des analyses »

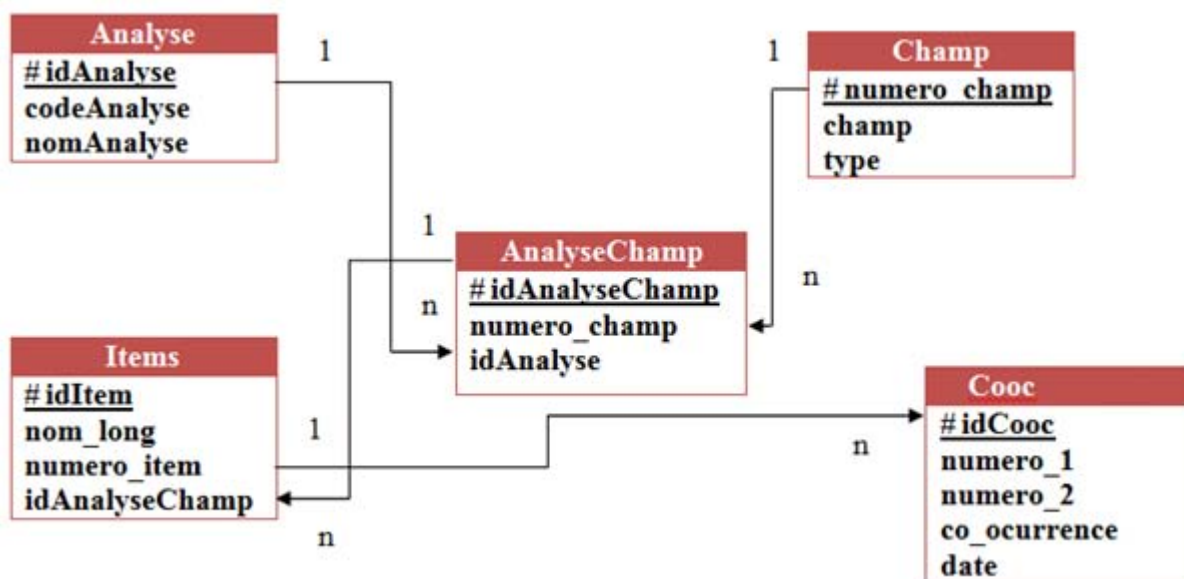


Figure 56. Module gestion des analyses

- Module « Gestion des droits d'accès »

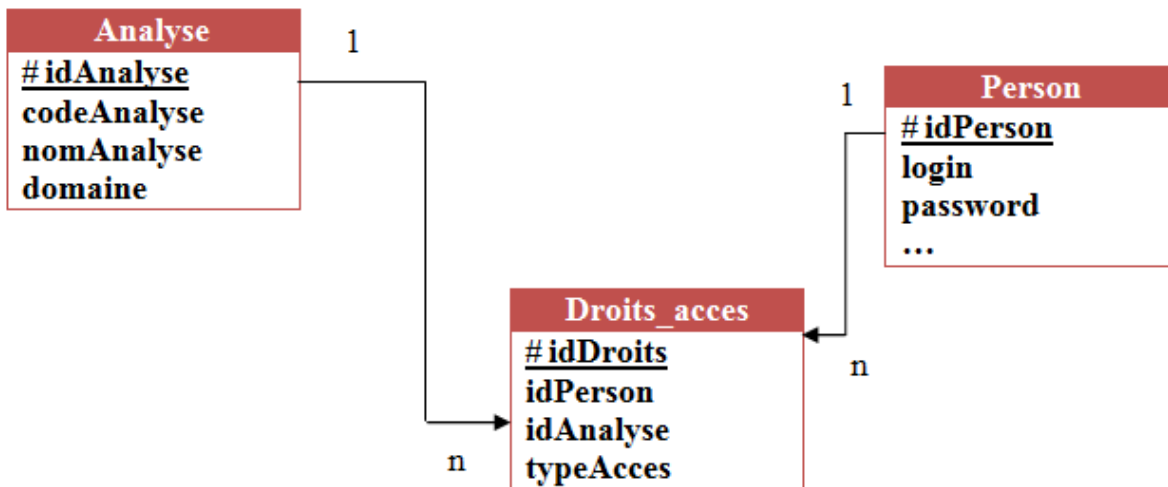


Figure 57. Module gestion des droits d'accès

- Module « Gestion de la file de travail »

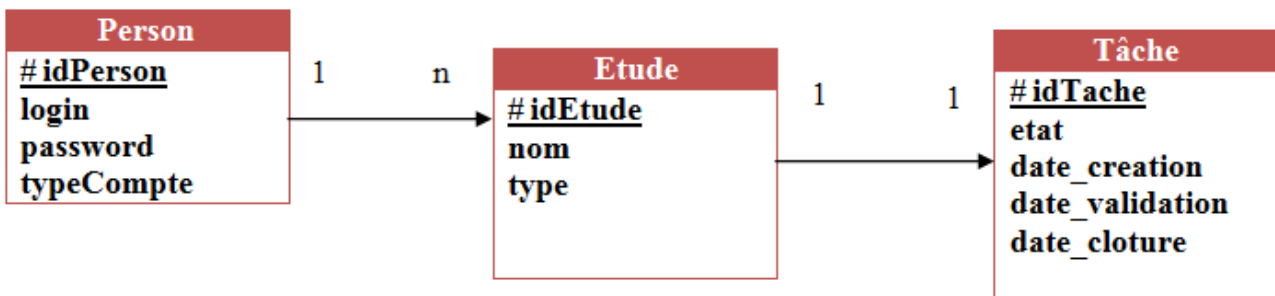


Figure 58. Module gestion de la file de travail

➤ Gestion des droit d'accès

L'approche adoptée, pour le développement du prototype Xplor Everywhere, permet de combiner les techniques d'extraction de connaissance à partir des données textuelles et les techniques de stockage, d'analyse et de visualisation des données relationnelles. Chacune de ces techniques est vue comme un composant aux fonctionnalités précises et délimitées. Plus simples à développer, plus robustes et testés dans des contextes différents, ces composants peuvent s'assembler de plusieurs manières pour créer ainsi des



Figure 59. Gestion des droits d'accès

applications variées et adaptées aux besoins des utilisateurs. Donc pour la sécurisation il faut attribuer des tâches à des rôles bien précis, d'où l'approche de contrôle d'accès, qu'on a présenté dans la section 3.4.

Pour bien assurer la demande des utilisateurs en termes d'analyse, nous proposons trois niveaux de type d'accès aux analyses:

- Niveau 0 : qui correspond à l'analyse statique.
- Niveau 1 : qui correspond à l'analyse dynamique.
- Niveau 3 : qui correspond au zoom avancée.

Gestion des comptes

Attribution des droits

AJOUTER NOUVEAU

MODIFIER

SUPPRIMER

ENREGISTRER

ANNULER

Prénom

Administrateur

Nom

System

Email

tetra.everywhere@gmail

Profession

Administrateur

Téléphone

0033 5 61 55 67

Ville

Toulouse

Pays

France

Adresse

IRIT, 118 Route de Narbonne - 31062

Rechercher

Liste des comptes						
Login	Prénom	Nom	Email	Profil	Ville	Pays
admin	Administrateur	System	tetra.everywhere@c	admin	Toulouse	France
anass	Anass	EL HADDADI	haddadi@irit.fr	Analyste	Toulouse	France
ilham	Ilham	BERRADA	iberrada@ensias.m	Expert IE	Rabat	Maroc
bernard	Bernard	DOUSSET	dousset@irit.fr	Expert IE	Toulouse	France

Figure 60. Gestion des compte utilisateur Xplor EveryWhere

4.4.2. Analyse de données

L'espace analyse de données (figure 61) permet aux utilisateurs d'explorer les analyses mises en ligne tout en gardant l'aspect collaboratif entre l'utilisateur, le veilleur et l'analyste.

Administration

Outils

Analyses

Cartes géostratégiques

M. ANASS EL HADDADI - Sujet d'analyse : SCAFFOLD

se déconnecter

Déposer une étude

Mon fichier de travail

Changer de mot de passe

Figure 61. Espace analyse de données

➤ Dépôt et suivie d'une étude

Cet espace, permet de déposer une étude c'est-à-dire le déclenchement d'une étude stratégique à la demande avec une liste des règles à respecter si l'étude est de type veille scientifique ou brevet.

Demander une analyse

Nom d'analyse :

Télécharger le document associé

[+ Add...](#)

[ENVOYER](#)

Règles à respecter

Dépôt d'Etudes brevet ou publications scientifiques

Le fichier doit être :

- zipper (en .rar ou autre)
- contenir :
 - les dossiers pour chaque croisement
 - les fichiers de synonymie (type .syn .Filtre .indF ...)

Nomenclature des dossiers :

Obligatoirement en majuscule

Correspondance Abréviation/Signification :

--- Acteurs ---

AU Auteurs

JN Journaux

PA Pays

VI Villes

OR Organismes

EM E-mails

[Close](#) | [Next »](#)

Figure 62. Déposer une étude

L'espace collaboratif offre le suivi d'état d'avancement des analyses.

Liste des tâches						
Nom d'analyse	Etat de la tâche	Date de création	Date de validation	Date de clôture	Nombre de champs	Période
Test 2	EN ETUDE	21/02/2010	09/10/2010	29/10/2010		
Test 3	EN ETUDE	21/02/2010	12/10/2010	27/10/2010		
Test 5	EN ETUDE	04/10/2010	05/10/2010	29/10/2010		

Figure 63. Suivi des analyses

➤ Analyse descriptive

Le menu analyse (figure 64), permet d'une part d'effectuer une analyse détaillée statique et dynamique et d'autre part d'appliquer des filtres afin de naviguer dans l'environnement d'un acteur donné.



Figure 64. Menu analyse

Pour extraire l'information stratégique de la base de données nous avons interrogé la base en utilisant les fonctions de requêtage issues du langage HQL⁸, en formulant des requêtes selon les besoins de l'utilisateur. Chaque requête représente un indicateur générique.

Parmi ces requêtes nous pouvons citer :

- Sélection d'un ou plusieurs champs d'analyse,
- Indicateur uni-varié
 - o Nombre d'occurrences d'un item,

⁸ Hibernate Query Language : C'est le langage de requêtes du framework Hibernate. Il permet d'effectuer des requêtes sur une base de données de manière orientée objet. Ces requêtes seront ensuite transformées par Hibernate en requêtes SQL natives pour être exécutées sur la base de données.

- Fréquence d'un item,
- Sélection des X meilleurs items,
- Indicateur relationnel
 - Sélection de l'évolution des X meilleurs items,
 - Sélection des X meilleurs pour une période précise,
 - Sélection de plusieurs items, et comparaison en fonction du temps pour étudier leur évolution.

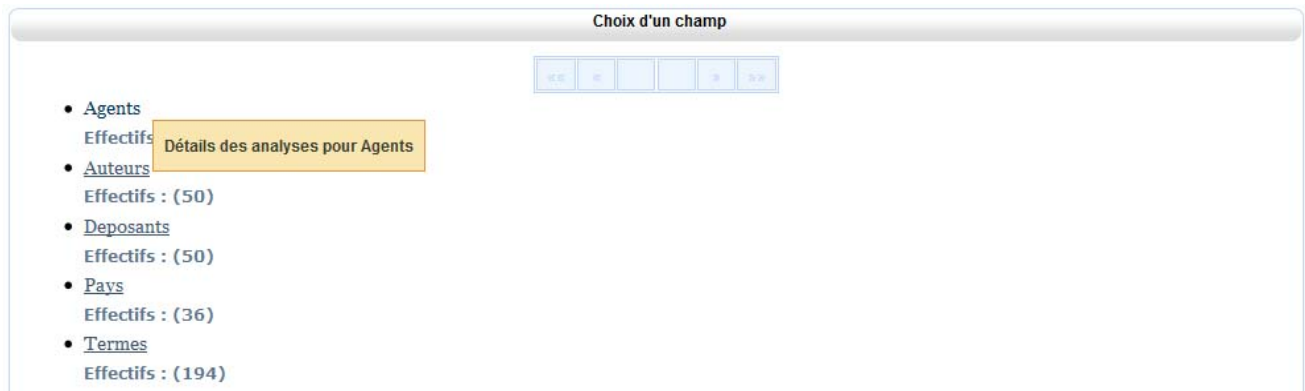


Figure 65. Choix du champ d'analyse (Top10, 20 ou 50)

Une fois le champ d'analyse choisi, un tableau de synthèse affiche la liste du top 10, 20 ou 50, avec la possibilité d'exportation sous format PDF ou Excel (figure 66).

Nous représentons ce tableau sous forme d'histogramme, de camembert, de barre pour une interprétation statique. Puis, une représentation de l'évolution des relations de dépendances (figure 67).

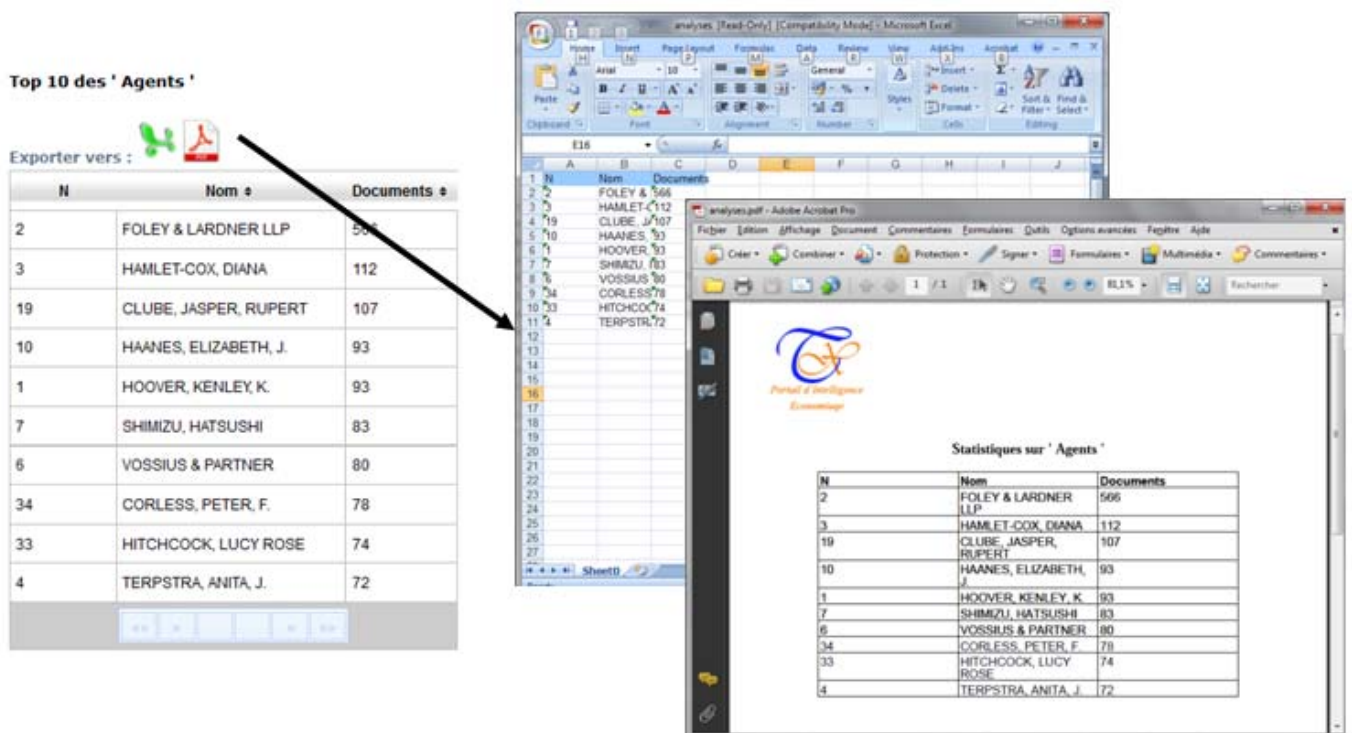


Figure 66. Le tableau de synthèse



Figure 67. Reporting Statistiques Descriptives

➤ Menu Filtrage

Dans le but de visualiser l'environnement relationnel d'un attribut, le menu filtrage (figure 68) permet à l'utilisateur final de naviguer à sa guise, afin de préciser certains passages de l'analyse et de les ramener dans le contexte et l'environnement requis.

Une fois le filtre choisi, le prototype permet de visualiser le type, le nombre de croisements de l'attribut (figure 69), ainsi que son environnement, son réseau social, son réseau sémantique, ses collaborations internationales sous forme de graphes statiques et dynamiques:

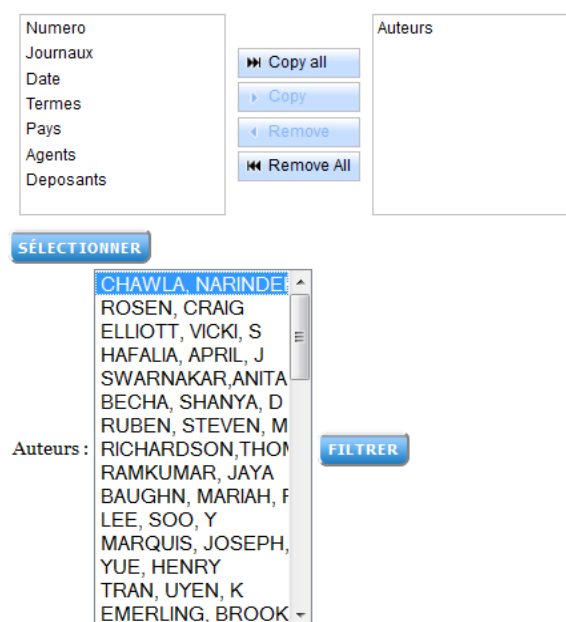


Figure 68. Choix de filtre

Résultat du filtrage

Statistiques			
Type	Nom	Nombre de croisements	Environnement
Auteurs	CHAWLA, NARINDER, K	969	Visualiser

[Environnement](#)
[Profils](#)
[Réseau social](#)
[Réseau sémantique](#)
[Réseau international](#)

Figure 69. Résultats de filtre

Graphes statiques

- Prise en compte de la valuation des arêtes.

Les algorithmes de type Force Directed Placement (FDP) (Loubier, 2009) sont conçus pour travailler sur des arêtes binaires (0/1). Lorsque nous avons voulu (dans XPlor) tenir compte des valuations des arêtes pour produire un graphe proche de la réalité évaluée, le dessin n'aboutissait plus (résultat décevant surtout si les valuations sont très différentes). Nous avons dû paramétrer l'ensemble des forces afin d'adapter l'algorithme à chaque cas rencontré avec une méthode de réglage automatique des forces en fonction de la topologie du graphe (nombre de sommets et d'arêtes, diamètres du graphe, nombre maximum d'adjacences, nombre de clusters, plage de valuation, ...).

- Positionnement de graphe.

Le Markov clustering (Mcl) (Loubier, 2009) permet de partitionner un graphe en tenant compte de sa topologie (liens entre sommets). Il est alors possible de travailler sur des graphes de grande taille sans être confronté à une grande complexité. En effet, le graphe de cluster (graphe réduit) permet une étude macroscopique de la topologie. Chaque cluster peut ensuite être détaillé et dessiné à part avec ses connecteurs externes. Il est même possible de revenir au dessin global en fixant le représentant le plus significatif de chaque classe. Ainsi, il est possible d'appréhender à la fois la structure macroscopique du graphe et les structures locales de chaque classe.

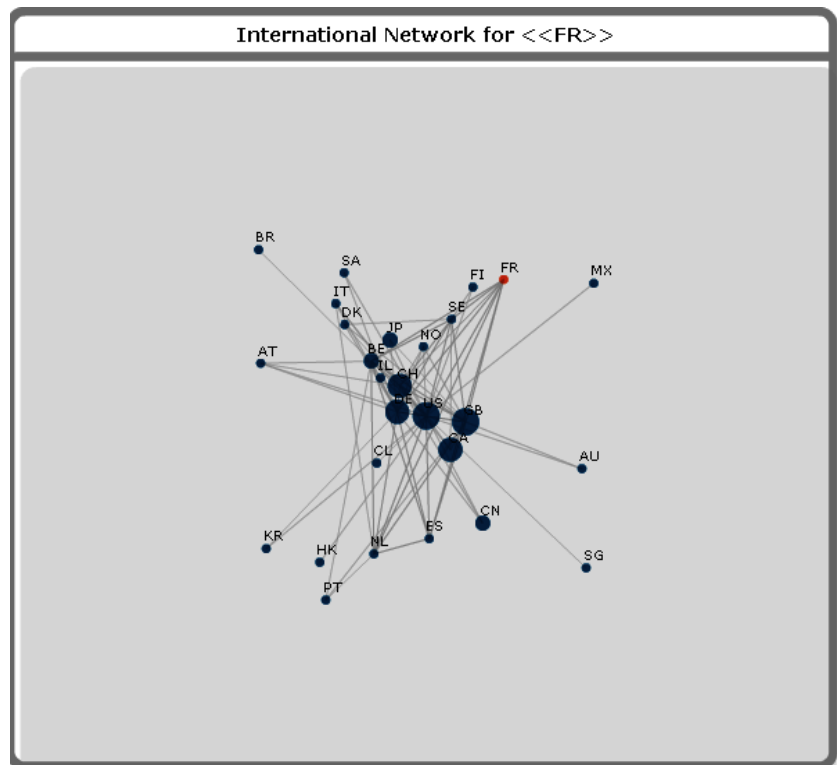


Figure 70. Graphe Statique

Graphes dynamiques

Il s'agit en fait d'une séquence de plusieurs graphes ayant les mêmes sommets et dont la somme des arêtes permet de retrouver le graphe global. Ceci permet de simuler l'évolution d'un graphe sur plusieurs instances.

Trois paramètres sont pris en compte :

- La décomposition du temps en instances.
 - La présence d'un sommet dans chaque instance.
 - La liaison temporelle entre deux sommets.
- Visualisation séparée des instances.

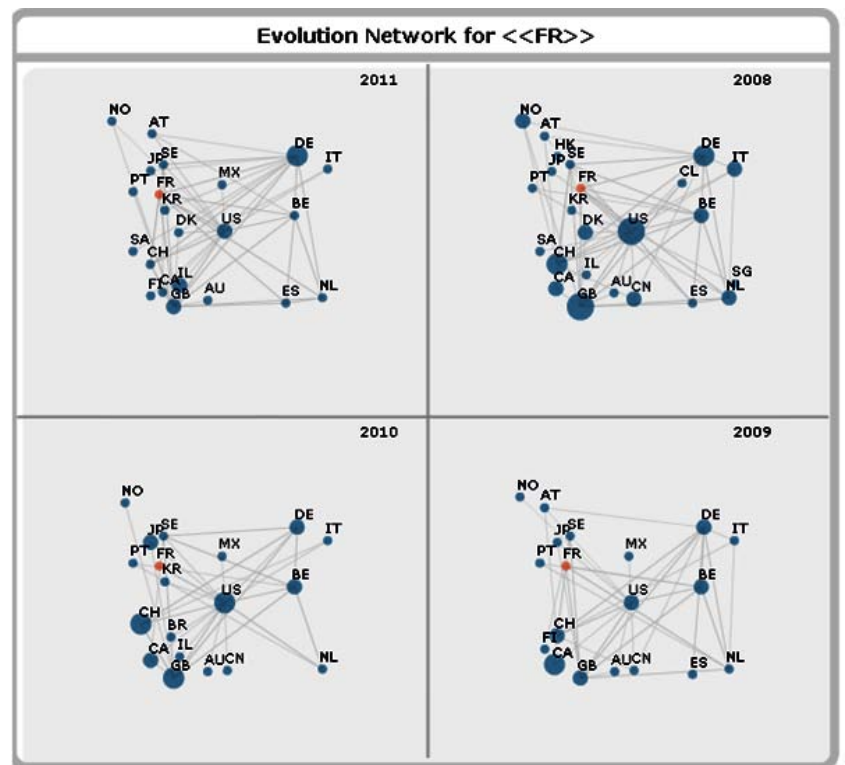


Figure 71. Graphe dynamique

Chaque instance peut être visualisée comme un graphe statique, les sommets restent immobiles (conservation de la carte mentale), ils gardent leur position définie par le dessin initial du graphe global. (Retour possible au graphe temporel à tout moment).

- Analogie avec une horloge

Pour matérialiser l'enchaînement des instances, nous disposons des repères temporels sur un cercle ou une ellipse. Ces repères sont ajoutés au graphe comme des sommets particuliers. Le dessin de graphe est alors déclenché afin de trouver un compromis entre sa topologie globale et ses topologies instanciées.

➤ Carte géostratégique

Pour la visualisation de collaborations internationales, Xplor EveryWhere propose les cartes géostratégiques avec la prise en compte de l'aspect temporel.

Les mêmes fonctionnalités sont développées pour les Smartphone (figure 73).

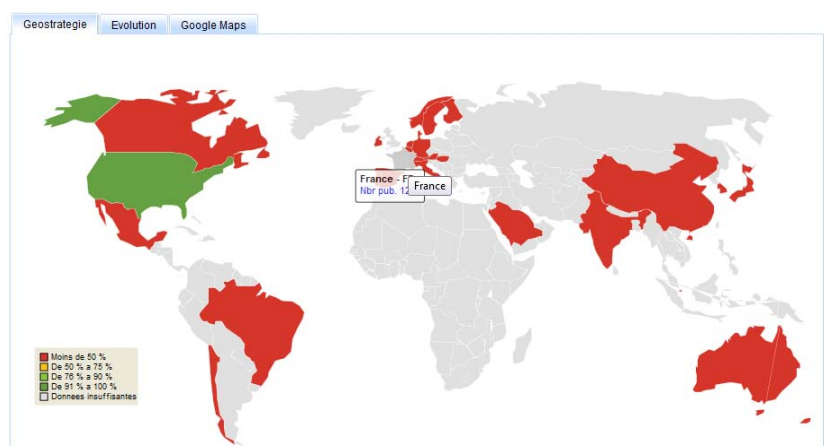


Figure 72. Carte Géostratégique



Figure 73. Le prototype XPlor EveryWhere pour iPhone

4.5. Expérimentations : analyse stratégique sur les scaffolds

Dans cette section, nous présentons l'expérimentation effectuée pour valider chaque étape de développement du SIE XPlor EveryWhere.

Le bureau d'études consiste à effectuer une analyse stratégique (veille scientifique et technologique) sur un sujet d'actualité. Il a été proposé au cours de l'année universitaire 2010-2011 de traiter les sujets suivants :

- M-Paiement
- Technologie RFID (Radio frequency identification)
- Séquenceurs d'ADN (notamment le GS20 de Roche)
- Génomique

La génomique est une discipline de la biologie moderne. Elle étudie le fonctionnement d'un organisme, d'un organe, d'un cancer, etc. à l'échelle du génome, et non plus limitée à celle d'un seul gène.

La génomique se divise en deux branches :

La génomique structurale, qui se charge du séquençage du génome entier ;

La génomique fonctionnelle, qui vise à déterminer la fonction et l'expression des gènes séquencés en caractérisant le transcriptome et le protéome.

- Puces à protéines
- Radars anticollision

- Véhicules hybrides
- Véhicules propres (électriques)
- Energie renouvelable
- Domotique
- Aquaculture
- Panneaux photovoltaïques
- Green computing

Green computing (en français informatique verte) ou encore green information technology (abréviation green IT) est un concept consistant à tenir compte des contraintes et des coûts en énergie (alimentation électrique et climatisation) des matériels informatiques. Les objectifs de l'informatique verte sont de mesurer et d'améliorer la performance énergétique des outils de production informatique. Pour être véritablement pertinente et crédible, une démarche de type informatique verte doit reposer sur des éléments concrets. Disposer d'informations quantifiées permet alors de prendre des décisions à la fois économiques et écologiques lorsqu'il s'agit d'opter pour tel ou tel équipement ou de mesurer l'impact financier d'un plan d'évolution.

Le Journal officiel de la République française du 12 juillet 2009 donne éco-TIC comme équivalent de green information technology et green IT ou encore green computing. Selon la définition qu'il en donne, les écotechniques de l'information et de la communication sont des techniques de l'information et de la communication dont la conception ou l'emploi permettent de réduire les effets négatifs de l'activité humaine sur l'environnement.

Le Journal officiel précise que « la réduction des effets négatifs de l'activité humaine sur l'environnement tient à la diminution de la consommation d'énergie et des émissions de gaz à effet de serre qui résulte du recours aux écotechniques ou à la conception même de ces techniques, qui s'attache à diminuer les agressions qu'elles pourraient faire subir à l'environnement au cours de leur cycle de vie ».

- **Cloud computing**

Le Cloud computing est un concept de déportation sur des serveurs distants des traitements informatiques traditionnellement localisés sur le poste utilisateur. Bien que l'anglicisme Cloud computing soit largement utilisé en français, on rencontre également les francisations « informatique dans le nuage », « informatique en nuage », « informatique dématérialisée », ou encore « infonuagique ».

Ce concept est vu comme une évolution majeure par certains analystes, ou comme un piège par d'autres comme notamment Richard Stallman.

- **Le graphène (matériau bidimensionnel)**

Le graphène est un cristal bidimensionnel (monoplan) de carbone dont l'empilement constitue le graphite. Il fut isolé en 2004 par Andre Geim, du département de physique de l'université de Manchester, qui a reçu pour cette découverte le prix Nobel de physique en 2010 avec Konstantin Novoselov. Il peut être produit de deux manières : par extraction mécanique du graphite (graphène exfolié) dont la technique a été mise au

point en 2004, ou par chauffage d'un cristal de carbure de silicium, qui permet la libération des atomes de silicium (graphène epitaxié).

- **L'informatique quantique**
- **Les peintures à nano particules (applicables sur n'importe quelle surface)**

Les peintures à nano particules, applicables sur n'importe quelle surface, permettant de remplacer les panneaux photovoltaïques avec un rendement cinq fois supérieur d'ici 5 à 7 ans.

- **Anticorps (antibodies)**

Les scaffolds – APTAMER – SMIP – ANTICALIN – PHYLOMER

XmAb – Affibody – EviBody(ies).....

Dans cette expérimentation, on prend comme exemple l'analyse sur les scaffolds effectué avec les étudiants Delbecq Axelle, Guasch Lucie, de Penfentenyo Marie, Truchot Cindy et Mounes Anthony de M2 PINT (figure 74).

Figure 74. L'analyse Scaffolds

4.5.1. Besoin informationnel

Cette étude a été menée de façon conjointe avec les étudiants du Master Innovation, Technologie et Entreprenariat de l'École Supérieure de Commerce de Toulouse qui veulent analyser deux entreprises pharmaceutiques **Molecular Partners** et **Novasep**:

- **Molecular Partners** est une société biopharmaceutique privée, située à Zurich, en Suisse. Créée en 2004, elle compte aujourd'hui 38 employés. Molecular Partners est précurseur dans le développement de bibliothèques de DARPins (Designed Ankyrin Repeat Proteins) à visée thérapeutique.
- **Novasep**, fondée en 1995, est spécialisée dans le domaine de la production de principes actifs (APIs) et des technologies de purification pour les industries des sciences de la vie. Elle compte aujourd'hui 1300 employés, et a généré un chiffre d'affaire annuel d'environ 300 millions d'euros en 2009. Historiquement basée à Nancy, en France, où se trouve son siège social, elle compte aujourd'hui plusieurs annexes à travers le monde (six usines de manufacture en France, en Allemagne, et aux Bahamas ; deux sites de production biopharmaceutique en Belgique ; et enfin plusieurs sites de R&D et de manufacture aux États-Unis et en Chine, sans oublier des bureaux au Japon).

Les deux entreprises aimeraient avoir un état des lieux de la recherche concernant les molécules similaires aux anticorps et notamment les équipes, les stratégies, les alliances, les points forts et faibles des autres entreprises... Nous nous intéresserons donc plus particulièrement à 12 molécules regroupées sous le nom de **Scaffold** et nous mènerons une étude stratégique visant à établir un rapport de recherche pertinent permettant à ces deux entreprises de s'adapter et de réagir rapidement de façon innovante face au marché.

Les immunoglobulines et leurs fragments représentent des outils thérapeutiques et de diagnostic très importants. En revanche, grâce aux progrès de l'ingénierie des protéines, de nouvelles molécules d'affinité, appelées scaffolds, ont été développées comme alternatives aux anticorps. Ces nouveaux scaffolds présentent certains avantages par rapport aux anticorps comme une petite taille, qui leur permet une meilleure pénétration dans les tissus, une forte stabilité, et la possibilité d'être produits dans des systèmes alternatifs et moins coûteux que les cellules de mammifères. Ces molécules sont donc caractérisées par le fait de pouvoir se lier à une autre. Ainsi, elles miment l'action des anticorps avec ce réel avantage d'être synthétisables chimiquement. C'est pourquoi les scaffolds constituent aujourd'hui une cible essentielle et privilégiée dans les activités de recherche, activités toujours plus croissantes depuis ces dernières années qui aboutissent par conséquent à de nombreux brevets d'invention. Les 12 molécules scaffolds que nous avons choisi d'intégrer dans notre étude sont les suivantes :

- **Aptamer** : Oligonucléotide synthétique (souvent ARN) capable de fixer un ligand spécifique. C'est un outil biochimique (synthétisés in vitro) utilisé dans des applications biotechnologiques, thérapeutiques... en fonction de la cible contre laquelle ils sont dirigés. Leur sélectivité et leurs propriétés de fixation sont similaires à celles des anticorps.
- **Spiegelmer Aptamer** : Énantiomères d'aptamères (synthèse chimique « en miroir ») avec les mêmes propriétés que les aptamères sauf qu'ils sont plus stables car ils ne sont pas dégradés par les enzymes et ont donc une durée de vie plus longue dans le plasma sanguin.
- **Affilin** : protéines utilisées en génie génétique qui ont la capacité de lier sélectivement des antigènes. Leur structure est dérivée d'une des deux protéines : gamma-B cristallin ou ubiquitine. Elles sont isolées par des techniques telles que le phage display. Elles ressemblent dans leurs affinités aux antigènes mais pas dans leur structure : ainsi ces protéines miment la fonction des Anticorps. Elles sont en train d'être développées en tant que drogues biopharmaceutiques potentielles.
- **Anticalin** : protéines artificielles qui sont capables de lier des antigènes, des protéines ou des petites molécules, mimant ainsi la fonction des anticorps sans en avoir une structure semblable. Elles sont dérivées de lipocalins humains qui sont une famille de protéines liantes naturelles. Les anticalins sont utilisés à la place d'anticorps monoclonaux, mais sont huit fois plus petits avec une taille d'environ 180 acides aminés et une masse de 20 kDa.
- **Evibodies** : Fragment de protéine de la famille des CTLA-4 (récepteur sur les globules blancs), plus petit que les anticorps et très spécifique.
- **Phylomer** : Peptides naturels (15 à 50 acides aminés), codés par le génome bactérien et sélectionnés pour leur affinité avec une cible spécifique.
- **Transferrin scaffold** : Protéine de liaison du fer modifiée et liée à une autre portion de protéine.
- **Adnectins** : dérivés de la fibronectine humaine, une abondante protéine extracellulaire permettant l'adhésion des cellules à la matrice extracellulaire. Il s'agit de protéines génétiquement modifiées qui sont capables de se lier à des antigènes, on les appelle « monocorps

» et malgré leurs noms, ces protéines n'ont pas une structure apparentée à celles des anticorps. Les adnectins ont été développées par *Adnexus*, une société de biotechnologies Américaine, dont le but a été de développer de nouveaux étiquetages thérapeutiques afin de lutter contre certaines maladies notamment des domaines de l'oncologie et de l'immunologie.

- **Affibody** : petites molécules de haute affinité qui ont été développées par une société Suisse de biotechnologies, *Affibody AB*. Ces molécules peuvent être génétiquement modifiées pour lier spécifiquement un nombre important de protéines d'étiquetage, imitant ainsi la fonction d'anticorps monoclonaux (famille des imitateurs d'anticorps). Les molécules Affibody ont été utilisées dans la recherche biochimique et ont été développées en tant que nouveaux médicaments.
- **Avimer** : abréviation de l'avidité des multimères, ce sont des protéines artificielles capables de se lier spécifiquement à certains antigènes grâce à de multiples sites de liaison (classées comme des mimétiques d'anticorps)
- **DARPin** : protéines à répétition de motifs ankyrines ; les DARPins sont dérivés de protéines naturelles ankyrine (classe de protéines qui permettent des interactions de haute affinité protéine-protéine dans la nature) et elles présentent ce motif répété plusieurs fois. Ces protéines sont utilisées comme outils d'investigation, de diagnostic et de thérapeutique.
- **Atrimers** : comprennent 3 domaines de liaison pour la fixation sur la cible d'intérêt. L'avantage de leur conception vient de l'utilisation d'une protéine du plasma humain de structure trivalente la tétranectine ; chacun des 3 domaines de liaisons de la tétranectine comprend 5 boucles d'acides aminés distincts qui peuvent être programmés pour se lier à n'importe quelle cible d'intérêt. L'intérêt de ces Atrimers dit "candidats-médicaments" réside dans leur structure unique pour la liaison avec la cible d'intérêt.

L'ensemble de ces molécules scaffolds constituera notre équation de recherche, point de départ de notre étude stratégique.

4.5.2. Source d'information

Parmi la masse d'informations disponibles, il est primordial d'identifier celles utiles pour notre étude. L'extraction de données pertinentes va reposer avant tout sur une équation de recherche la plus précise et adéquate en rapport avec notre sujet de recherche. Dès lors, après plusieurs combinaisons de mots clés, nous avons retenu l'équation de recherche suivante :

(Scaffold) AND ((Adnectin) OR (Affibody) OR (Anticalin) OR (Aptamer) OR (Spiegelmer) OR (Atrimers) OR (Darpins) OR (Evibody) OR (Phylomer) OR (Transferrin)).

Grâce à cette équation de recherche, nous avons interrogé plusieurs bases de données qui nous ont permis d'obtenir un certain nombre de brevets plus ou moins importants. Pour des raisons de complexité liée à la collecte d'informations, nous avons choisi de ne travailler qu'avec la base de données WIPO (1246 brevets).

En effet, la base de données USPTO ne met à disposition que des brevets américains et de ce fait, ne prend en compte que le point de vue américain. Or il est important de pouvoir comparer les informations provenant de différents pays. De plus, les brevets les plus récents ne sont pas répertoriés dans USPTO car cette base de données a un an de retard dans l'indexation de ses brevets.

Quant à Esp@cenet, la publication des brevets a l'inconvénient de présenter un risque en termes de perte d'informations pouvant avoir un intérêt majeur dans la collecte de nos données. En effet, la page de réponse à la requête ne comporte au maximum que 30 brevets et nous n'avons accès qu'au 500 premiers brevets. De plus, le moteur de recherche de la base ne permet de rechercher les mots clés que dans le titre ou le résumé. En faisant la recherche avec la même équation de recherche, nous n'obtenons que 3 brevets.

(WO/2011/113001) SCAFFOLD SYSTEM TO REPAIR CARDIOVASCULAR CONDITIONS

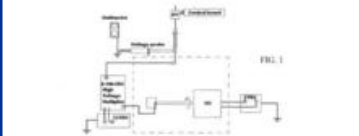
Biblio. Data	Description	Claims	National Phase	Notices	Documents
Latest bibliographic data on file with the International Bureau					
Pub. No.: WO/2011/113001 Publication Date: 15.09.2011		International Application No.: PCT/US2011/028201 International Filing Date: 11.03.2011			
IPC: A61F 2/06 (2006.01), A61F 2/82 (2006.01), A61L 27/14 (2006.01), A61L 27/22 (2006.01), A61M 29/02 (2006.01)					
Applicants: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West 7th Street, 7th Floor Austin, Texas 78701 (US) (All Except US). MASSEY, Jordan [US/US]; (US) (US Only). AGRAWAL, Mauli [US/US]; (US) (US Only). BAILEY, Stephen [US/US]; (US) (US Only).					
Inventors: MASSEY, Jordan; (US). AGRAWAL, Mauli; (US). BAILEY, Stephen; (US).					
Agent: MEYERTONS, HOOD, KIVLIN, KOWERT & GOETZEL, P.C.; Meyertons, Eric B. P.O. Box 398 Austin, Texas 78767-0398 (US) .					
Priority Data: 61/312,847 11.03.2010 US					
Title: SCAFFOLD SYSTEM TO REPAIR CARDIOVASCULAR CONDITIONS					
Abstract: A device for treating a cardiovascular condition includes an expandable scaffold positionable in a portion of a vasculature of a mammal; and maintained via stent technology, wherein the scaffold is comprised of electrospun fibers composed of a biodegradable compound. The biodegradable compound serves as a temporary template that allows the cardiovascular tissue to be rebuilt about the scaffold.					
					
Designated States: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW. African Regional Intellectual Property Org. (ARIPO) (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW) Eurasian Patent Organization (EAPO) (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM) European Patent Office (EPO) (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR) African Intellectual Property Organization (OAPI) (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).					
Publication Language: English (EN)					
Filing Language: English (EN)					

Figure 75. Exemple d'un brevets issus de la base WIPO

4.5.3. Structuration et homogénéisation des données

Avant de pouvoir travailler sur les 1246 brevets, il a fallu enlever le bruit, les redondances, les images et autres balises. Pour cela, nous avons défini un programme permettant de rechercher l'information utile à l'analyse dans les fichiers textes.

Le but de ce programme est multiple :

- éliminer la mise en forme du texte (html) qui n'apporte rien au contenu, mais qui représente 90% du poids du fichier,
- rebaliser le texte par des balises du type : TI:, AU:, OR:, etc.,
- rajouter des balises inexistantes dans le texte en les créant à partir de la mise en forme html : DP:, NO:,etc.
- conserver les informations qui sont codées en caractères latins ou en chiffres arabes comme les dates, les numéros ou certains termes comme les auteurs occidentaux, les formules ou les éléments techniques,
- éviter les doublons dus à la présence simultanée de deux versions du même terme (ex : noms des auteurs cliquables sur la page web).

Un corpus d'informations non structurées a alors été obtenu après ce premier prétraitement. Par la suite, il a donc été primordial de structurer l'ensemble de nos données. Pour cela, nous avons utilisé le descripteur Tétralogie (figure 76) dans le but de créer différents fichiers qui présentent la liste des attributs étudiés en se basant sur des dictionnaires de synonymies.

Les diverses commandes du descripteur Tétralogie ont donc permis d'ordonner plus précisément ces informations notamment en termes de fréquence ou de synonymie. Nous obtenons ainsi un fichier contenant la totalité des informations (.ind), un second dans lequel ces données sont classées par ordre alphabétique (.indA) et enfin, un dernier où celles-ci sont rangées par fréquence d'apparition (indF).

RR:				
Titre	TI	TI:	False	"
Num_Pub	NP	NP:	True	/"b"
Num_Appl	NA	NA:	True	/"b"
Abstract	MC	AB:	True	,".;" :"/""(")"b"
Inventeur	AU	AU:	True	;"."
Pays_Inventeur	PI	AU:	True	(")"
Deposant	DE	CO:	True	["
Pays_Deposant	PA	CO:	True	(")"
Num_Appl	NA	NA:	False	"
Date_Publi	DP	DP:	True	/"b"
Annee_Publi	AP	DP:	True	/"b"."
Date_Depot	DD	DD:	True	/"b"
Donnees_Prio	Do	Do:	True	b"."
Date_Prio	Dp	Dp:	True	b"."
Pays	Pa	PA:	True	,". "(")"
Class_WO	IC	IC:	True	," "("
Agent	AG	AG:	True	;"
Description	DE	DE:	False	"
Claims	CL	CL:	False	"
Langue_Publi	LP	LP:	False	"
Langue_Depot	LD	LD:	False	"
FIN	FIN	FIN	FIN	FIN

Figure 76. Descripteur de la base WIPO

4.5.4. Présentation multidimensionnelle

Grâce à la fonction « TetraToXplor », présenté dans la section 4.4.1, nous définissons la présentation multidimensionnelle du corpus d'information, ce qui nous permet de définir et identifier les différentes corrélations existantes entre les attributs d'un même document et leurs corrélations avec les autres documents du corpus ciblé. Le résultat de cette étape est la génération des fichiers : Champs, Items et Co-occurrences.

Une fois que la base de données est générée vient l'étape de traduction du besoin informationnel sous forme d'indicateurs d'analyse en se basant sur le modèle 5W-1H. Selon les différents besoins informationnels (lieux de la recherche concernant les molécules similaires aux anticorps et notamment les équipes, les stratégies, les alliances, les points forts et les points faibles des autres entreprises, etc.), nous présenterons une synthèse des indicateurs à calculer via XPlor EveryWhere :

- les tops 10, 20 ou 50 des principaux acteurs du domaine, leurs interactions et leurs évolutions.
- leur réseau sémantique et leurs évolutions.
- la spécialité des acteurs.
- les principaux organismes, leurs thématiques, leurs politiques et l'évolution de leurs environnements.
- La période d'analyse 2008-2011

4.5.5. Analyse et interprétation

Cette étape consiste à interpréter les croisements d'information réalisés et de mettre ainsi en évidence différents réseaux :

- les réseaux professionnels (inventeurs, mandataires, organismes déposants...)
- les réseaux sémantiques (mots clés)

La représentation sous forme de graphes ou de cartes géostratégiques nous permet ainsi d'expliquer les croisements effectués pour en formuler des hypothèses.

➤ Analyse de réseaux des inventeurs

A partir du zoom (figure 77) de la matrice Inventeurs X Inventeurs générée par Tétralogie, nous avons pu identifier plusieurs clusters mettant ainsi en évidence la collaboration de plusieurs équipes. Ce croisement nous permet d'obtenir le réseau professionnel des différents inventeurs ayant publié des brevets relatifs aux scaffolds.

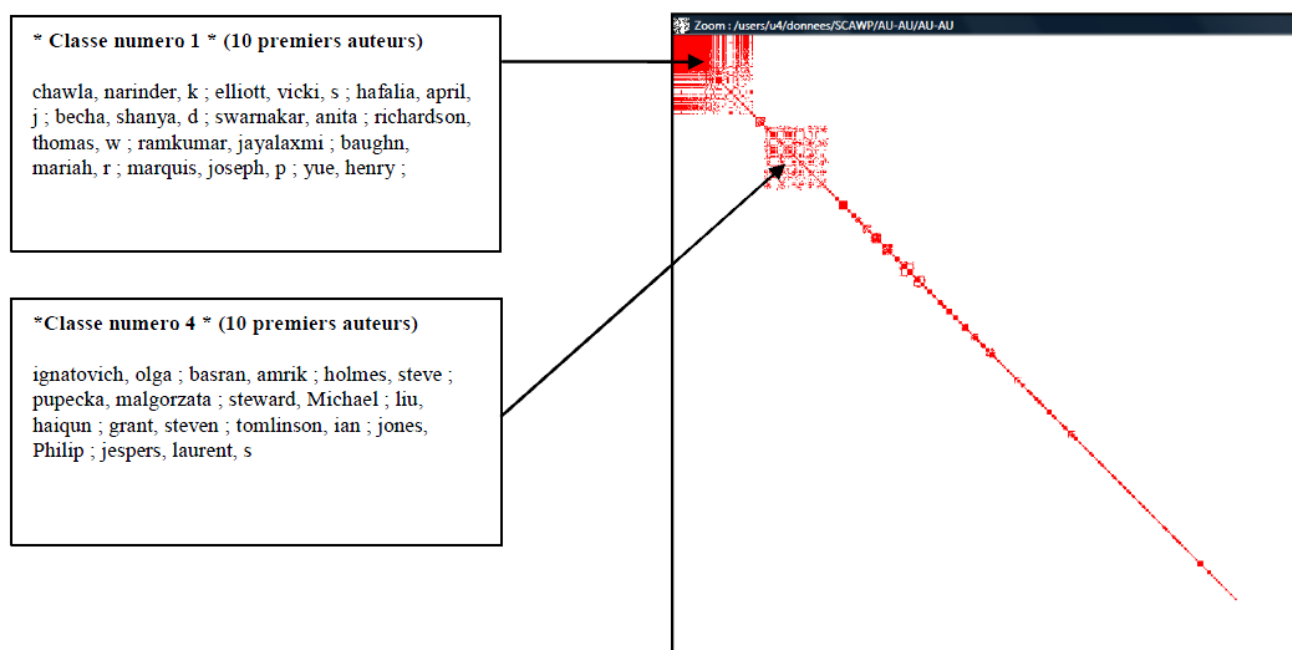


Figure 77. Zoom de la matrice cooccurrence Inventeurs X Inventeurs

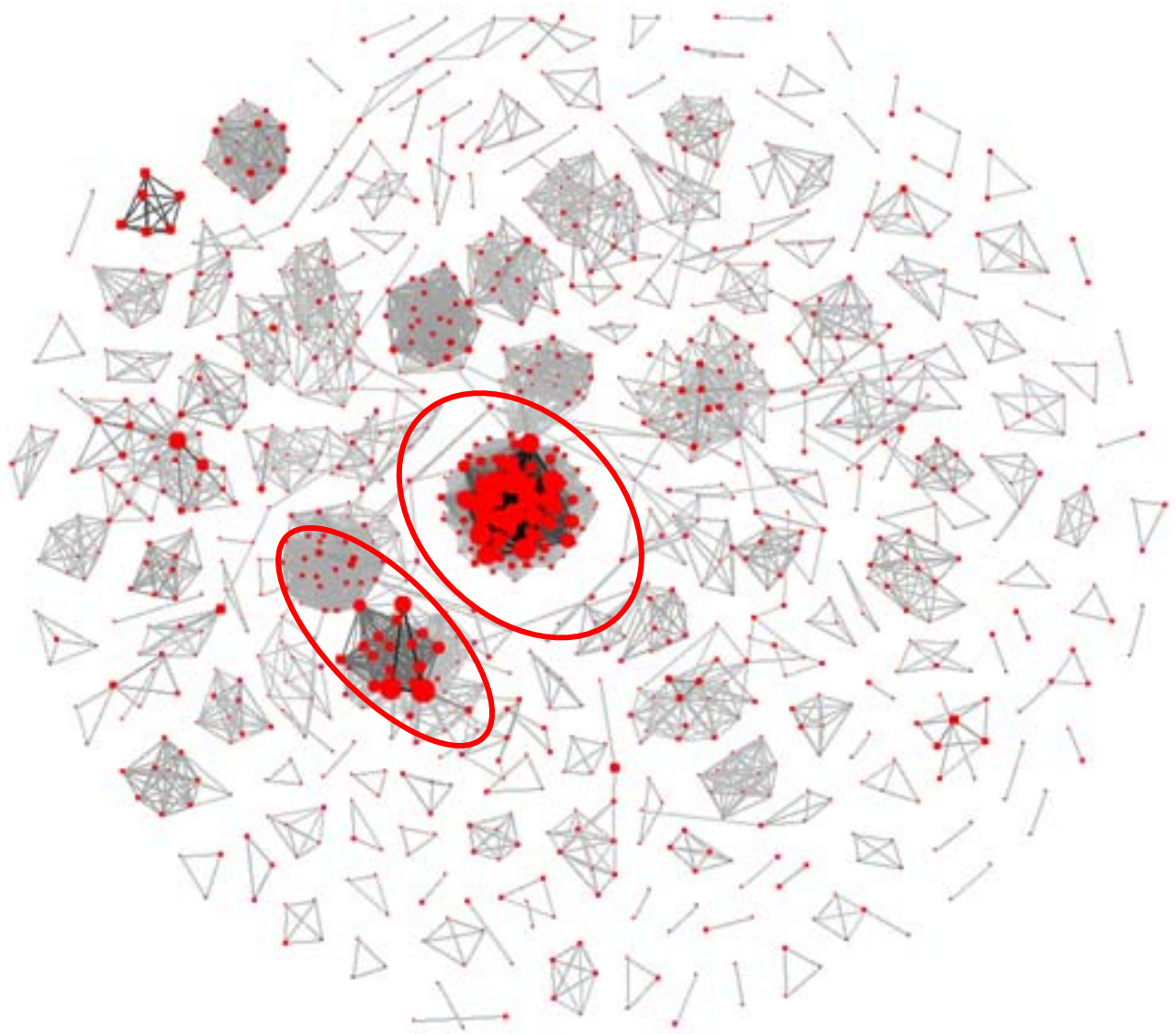


Figure 78. Graphe des clusters de réseau des inventeurs

Nous nous sommes plus particulièrement intéressés aux deux principaux clusters désignés sur la figure ci-dessus.

Interprétations :

- Cluster 1 : Quatre équipes de recherche en lien les unes avec les autres sont présentes, dont une rassemble un nombre important d'auteurs. Deux de ces équipes sont reliées par le biais d'un unique inventeur (Williams, Lewis), tandis que les deux autres équipes sont associées via plusieurs inventeurs.

Il est également à noter que Tang Tom, Drmnac (Radoje et Snezana) et Labet Ivan sont les inventeurs faisant le lien entre trois des quatre équipes.

- Cluster 2 : La principale information que nous pouvons retirer de cette représentation est la liaison de deux équipes unies par un seul inventeur (Baker Kevin).

De plus, nous pouvons identifier que les inventeurs du cluster (Ruben Steven, Barash Steven et Rosen Craig) qui publient le plus travaillent ensemble (forte intensité des liens).

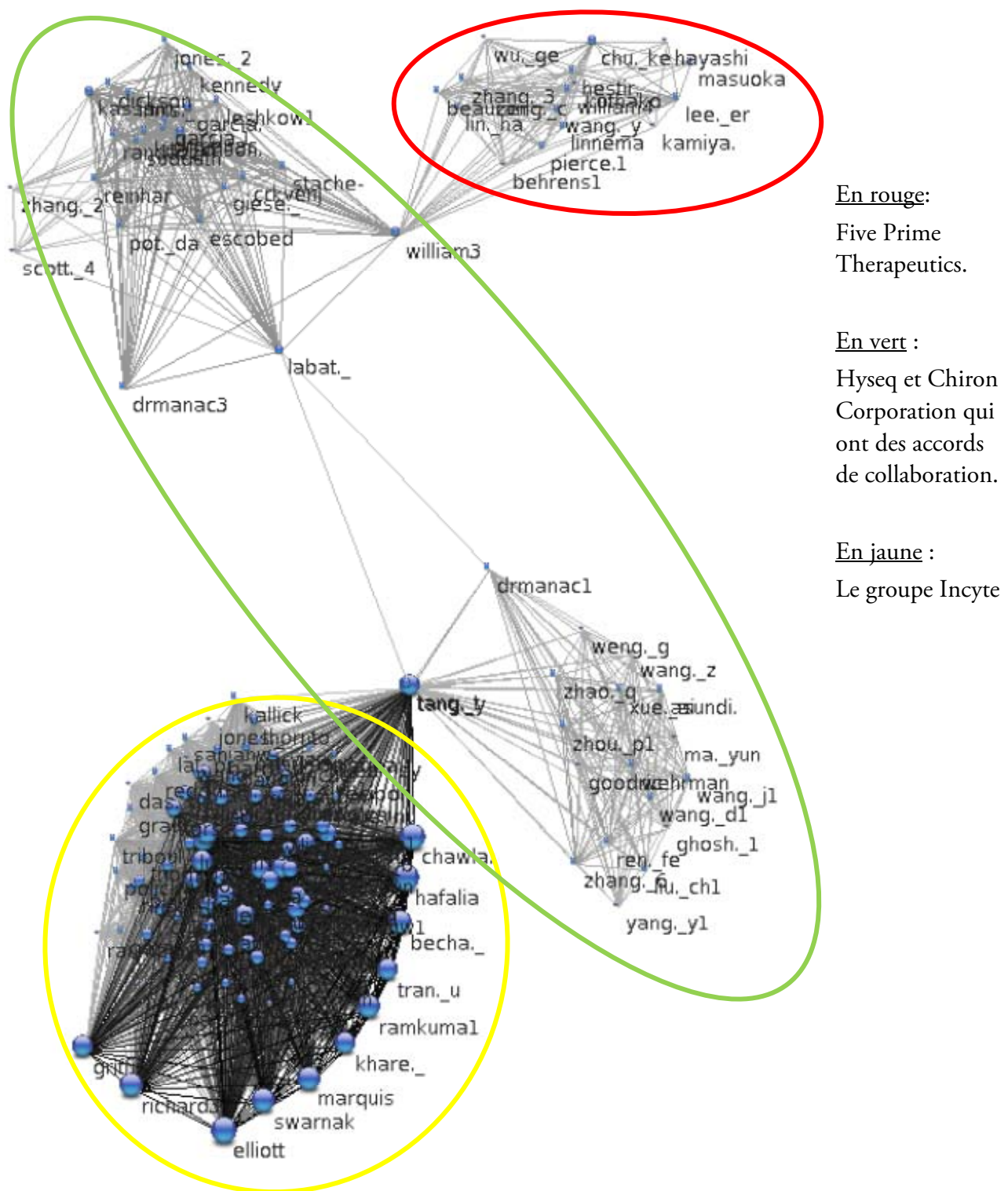


Figure 79. Visualisation du cluster 1

Les inventeurs entourés de rouges (figure 80) travaillent pour Human Genome Sciences et ceux entourés de jaune, pour Genentech. Baker fait le lien entre ces deux équipes car il a travaillé dans les deux. Cela ne veut pas dire qu'il y ait collaboration entre les deux équipes, le lien est à situer dans le temps.

Il semblerait qu'il y ait deux équipes au sein du groupe Genentech avec Ellen Filvaroff qui fait le lien entre les deux. Après quelques recherches dans le document initial, l'équipe de gauche n'a publié qu'en

2005 et 2006 et celle de droite avant, à partir de 2000. Ellen Filvaroff est restée toutes ces années et fait donc le lien entre l'ancienne équipe et la nouvelle.

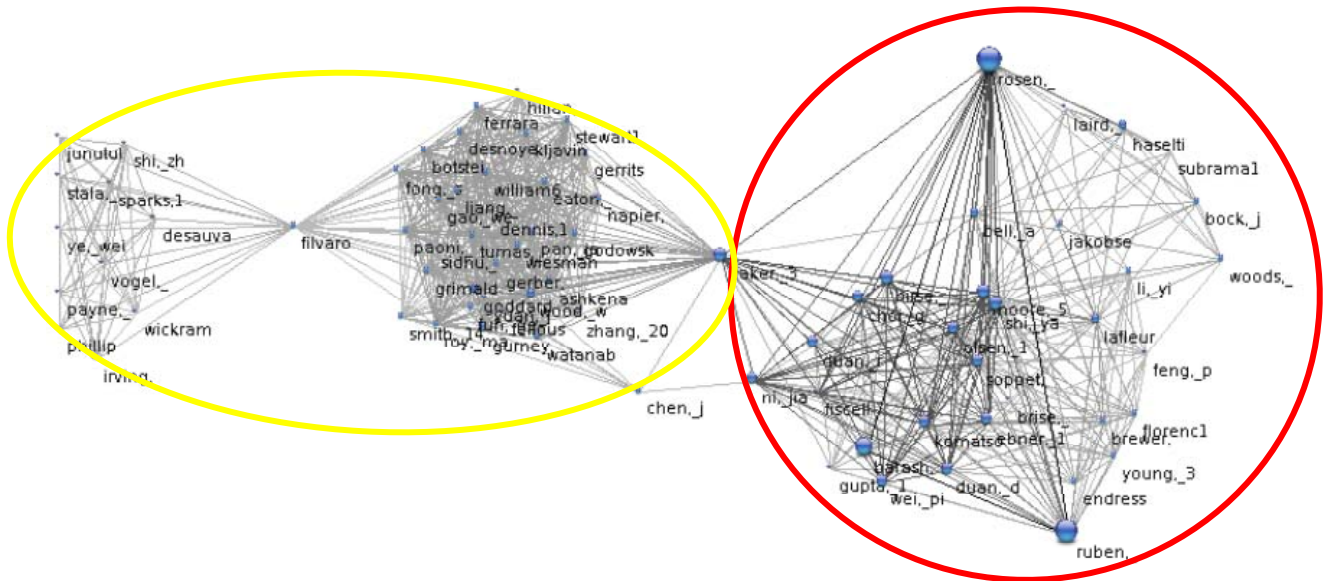


Figure 80. Visualisation du cluster 2

Cette étude macroscopique des réseaux d'inventeurs nous permet d'avoir une vision générale des équipes de recherche importantes. Dans le but d'étudier les profils des inventeurs, leurs spécialités, leurs évolutions, nous utilisons le SIE XPlor EveryWhere pour effectuer des études microscopiques.

La figure 81, nous montre l'évolution du top 10 des inventeurs. Les 10 premiers présentent le groupe Incyte avec l'inventeur Chawla en tête entre 2008 et 2011.

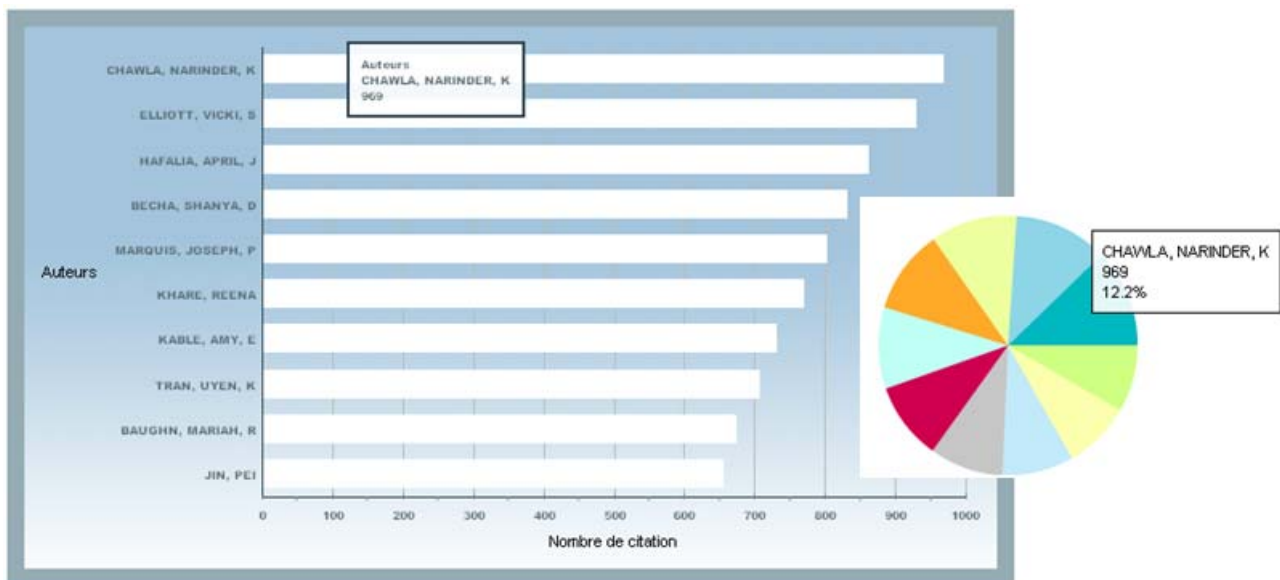


Figure 81. Le top 10 des inventeurs

La fonction de filtrage de XPlor EveryWhere nous permet naviguer dans l'environnement de leader, la figure 82 présente tout les cooccurrences de Chawla. Son profil, son réseau social, sa spécialité sont visualisés avec un simple clic (figure 83).

Statistiques			
Type	Name	Cooccurrences	Date
Agents	FOLEY & LARDNER LLP	13	2008
Agents	HAMLET-COX, DIANA	2	2008
Agents	FOLEY & LARDNER LLP	14	2009
Agents	HAMLET-COX, DIANA	3	2009
Auteurs	CHAWLA, NARINDER, K	16	2008
Auteurs	ELLIOTT, VICKI, S	13	2008
Auteurs	HAFALIA, APRIL, J	13	2008
Auteurs	BAUGHN, MARIAH, R	9	2008
Auteurs	BECHA, SHANYA, D	13	2008
Auteurs	MARQUIS, JOSEPH, P	12	2008

Figure 82. L'environnement de Chawla



Figure 83. Le profil, le réseau social et le réseau sémantique de Chawla

➤ Analyse de collaborations internationales

Dans ce réseau (figure 84), nous remarquons tout d'abord que la Hongrie est le seul pays non intégré au réseau de collaborations formé par les autres pays. On peut interpréter ceci comme le fait que la Hongrie travaille de façon indépendante sur les Scaffolds, soit par soucis de confidentialité soit parce qu'au contraire elle ne considère pas ce sujet comme étant prometteur et n'envisage pas de collaborer

pour des raisons stratégiques. A l'inverse, tous les autres pays sont en collaboration plus ou moins étroite formant un système au centre duquel apparaissent les Etats-Unis.

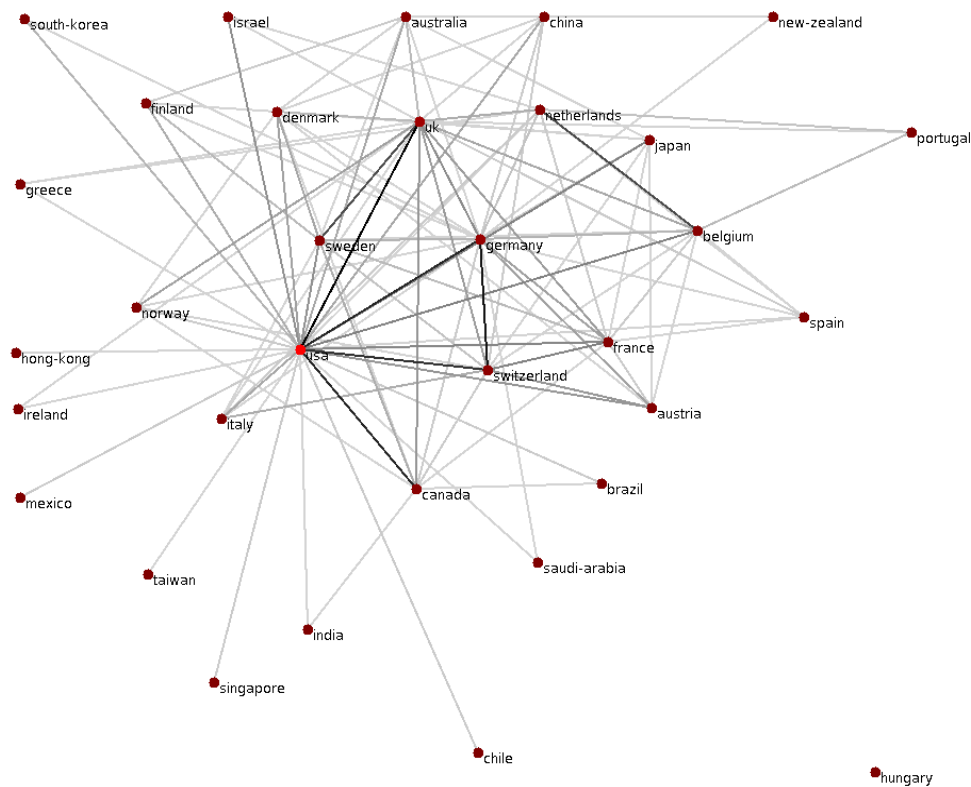


Figure 84. Réseau des collaborations internationales

La représentation sous forme de cartes géostratégiques illustre plus facilement l'évolution des pays déposants les brevets au cours des années.

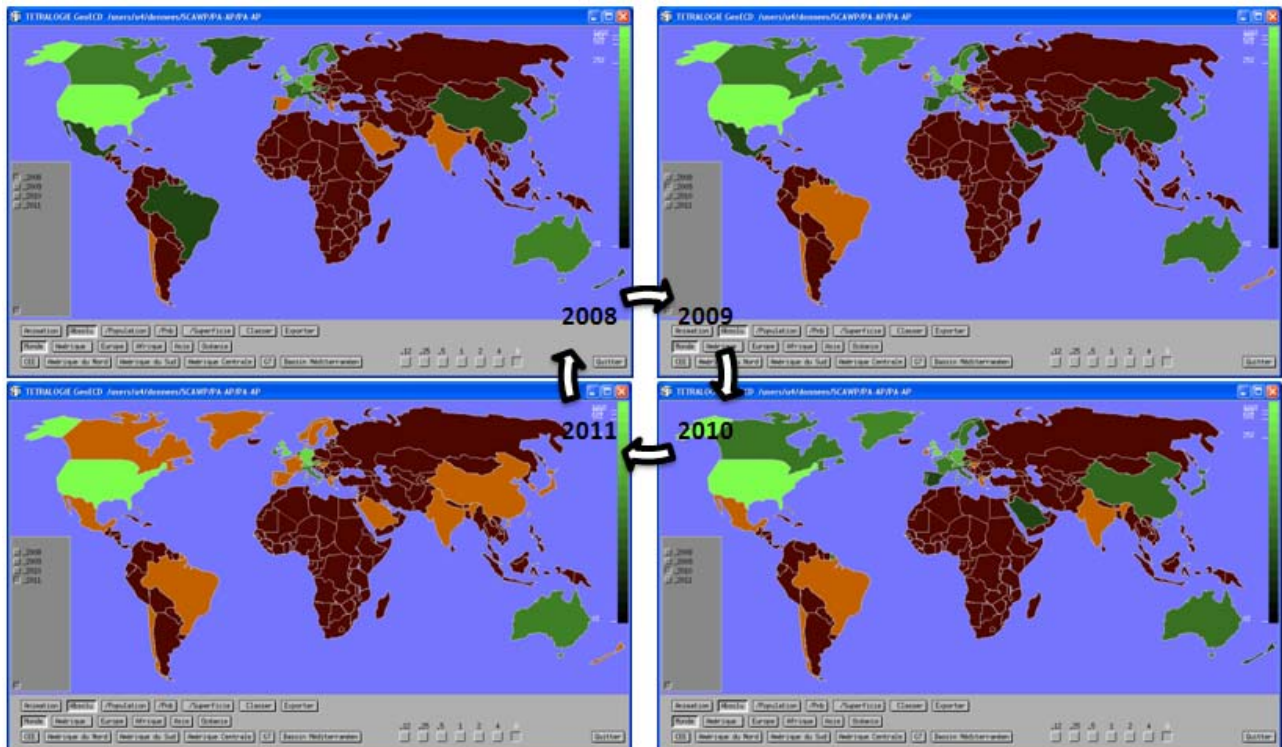


Figure 85. La carte géostratégique des pays déposants de brevets sur les Scaffold

Elle confirme notamment la position dominante des Etats-Unis et l'émergence de certains pays comme l'Allemagne, le Portugal ou encore l'Angleterre. De plus, le nombre important de pays en 2011 désignés comme des pays non déposants (pays en orange) s'explique une fois de plus par une année tout juste entamée. Cependant, cette représentation manque de précisions car interpréter une évolution en fonction d'un dégradé de couleurs n'est pas toujours évident et ne donne qu'une vue d'ensemble. Le logiciel XPlor EveryWhere nous permet d'apporter davantage de précisions pour notre interprétation.



Figure 86. La carte géostratégique XPlor EveryWhere

Cette figure confirme nos observations précédentes dans le sens où nous retrouvons les mêmes pays qui publient de façon régulière tels que les Etats-Unis, l'Angleterre, l'Allemagne ou encore la Suisse. On note que les Etats-Unis diminuent leur activité en 2010 contrairement à certains pays comme la Finlande, le Japon, l'Angleterre et la Suède qui eux se montrent plus présents. Les États-Unis perdent de l'importance en tant que pays déposant au cours des années car on a de nouveaux pays qui entrent sur le marché et investissent de plus en plus sur ce sujet (figure 87).

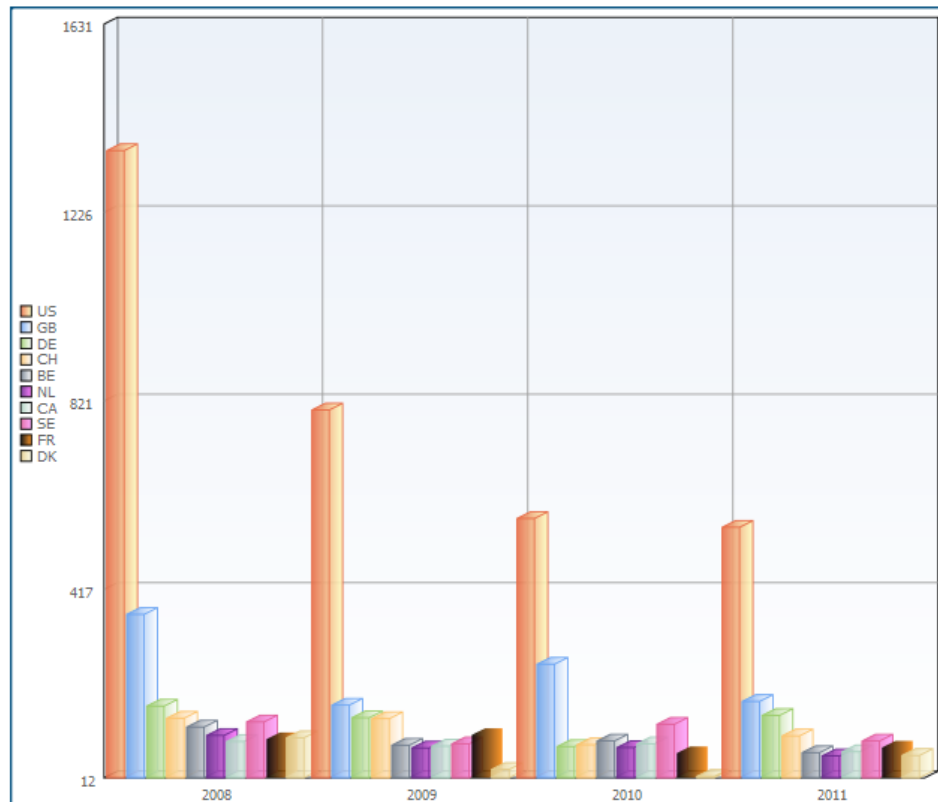


Figure 87. Evolution du tops 10 des pays

➤ Analyse de réseau sémantique

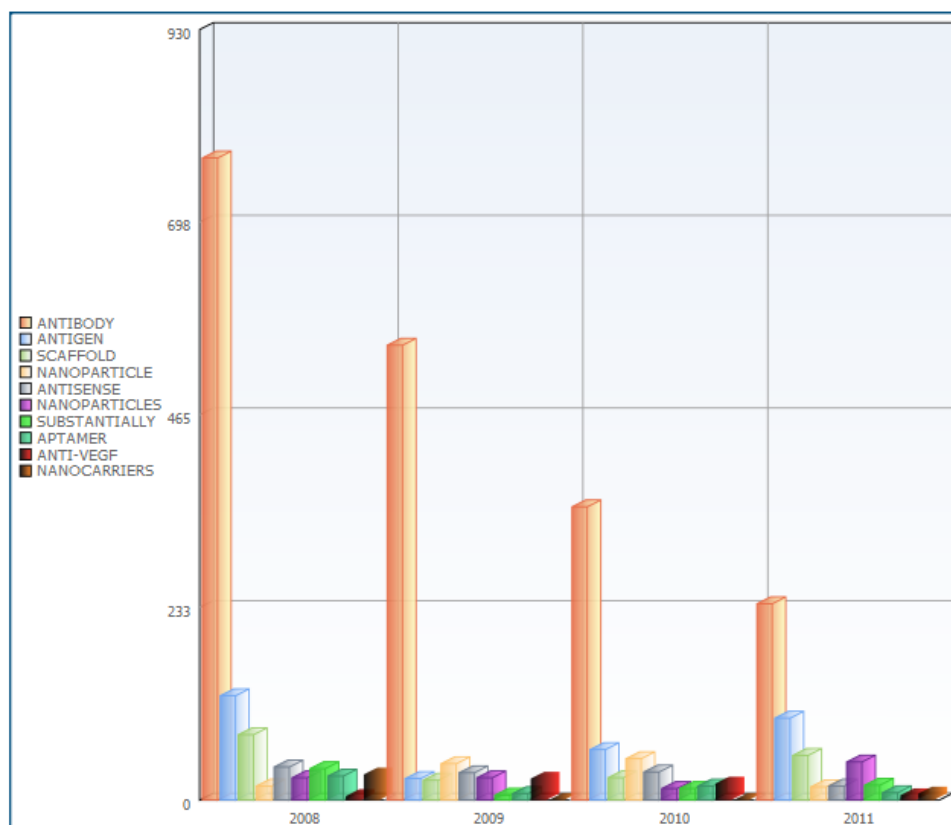


Figure 88. Évolution des mots-clés employés dans l'ensemble des brevets en fonction des années

Un ensemble de mots-clés s'agence autour de scaffolds avec des combinaisons de mots-clés majoritairement présents :

- Antibody-protein-polynucleotide
- Antibody-g-protein
- Antibody-antigen

Ce réseau sémantique facilite par exemple la recherche d'informations relatives aux scaffolds. Grâce à XPlor EveryWhere, on peut trouver assez facilement la spécialité de la France.

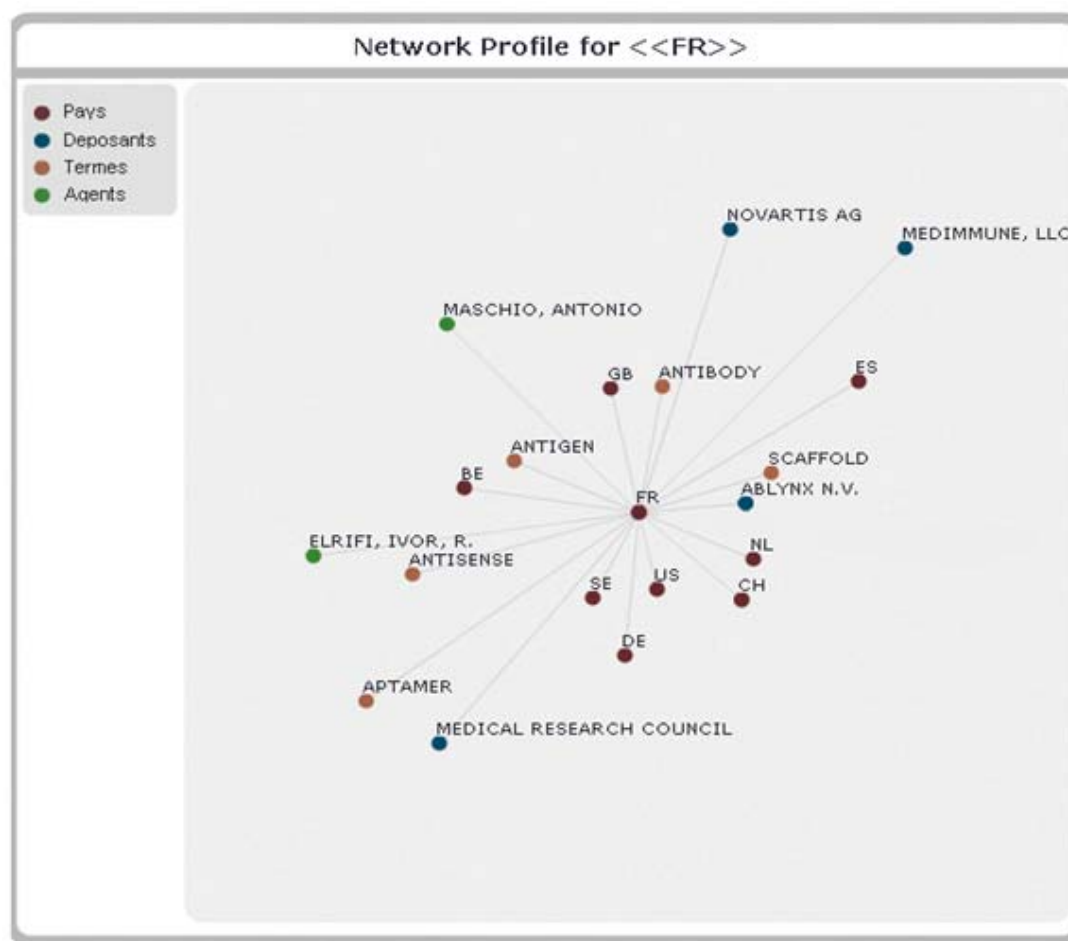


Figure 89. Spécialité de la France

Le mot-clé « antibody » a majoritairement été utilisé jusqu'en 2010 puis les mots clés « antigen » et « scaffold » ont pris de l'ampleur. Cette évolution peut s'expliquer par une prise de conscience de l'utilisation possible des Scaffolds pour leurs propriétés antigéniques avantageuses.

➤ Analyse de déposants

Nous remarquons tout d'abord que l'entreprise **Incyte** principalement présente en 2008 et 2009 ne l'est plus par la suite. On peut penser que cette entreprise n'a pas tiré suffisamment profit de ses recherches sur les scaffolds ou qu'elle a préféré se concentrer sur un autre sujet plus porteur. A l'inverse, elle a peut-être su suffisamment prendre d'avance sur le sujet et tire ainsi profit de son monopôle. Le cas de l'entreprise Glaxo est similaire à celui d'Incyte. On peut penser que l'absence de publication de ces deux entreprises dans le domaine des scaffolds a laissé la place à d'autres organismes comme Novartis.

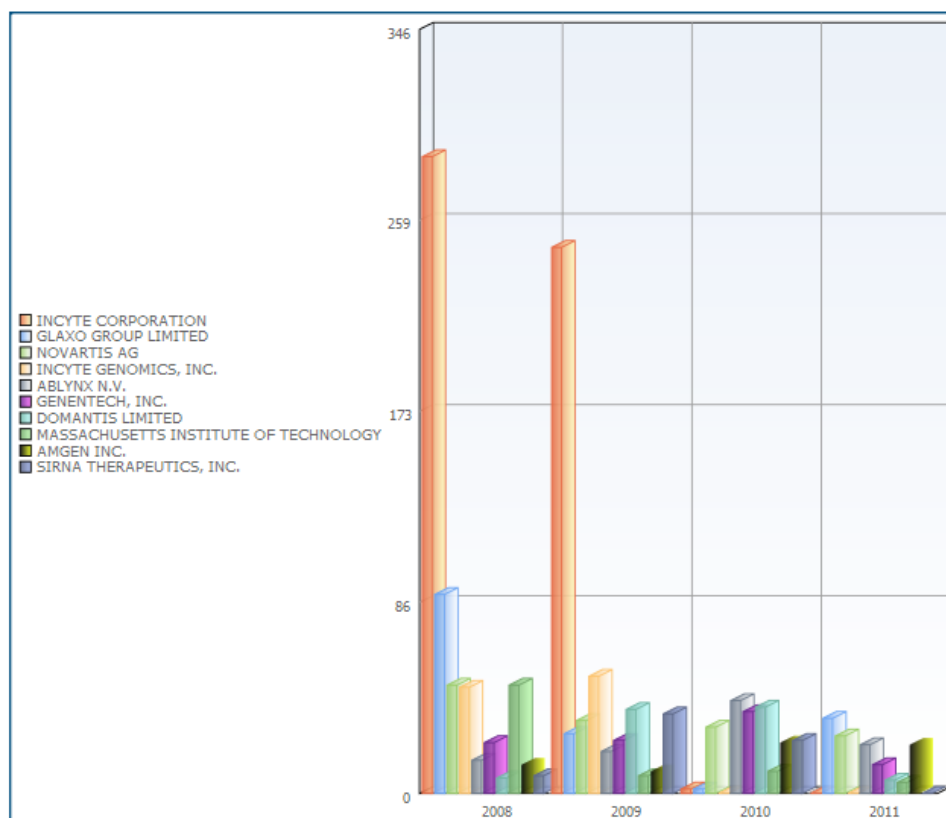


Figure 90. Évolution des déposants mondiaux en fonction des années

➤ Synthèse de l'analyse des scaffolds

Placée en amont des travaux des étudiants de l'ESC, cette étude avait pour but de leur apporter une vision globale de la recherche dans le domaine des scaffolds. De ce fait, nous avons pu établir différents réseaux, faire ressortir les auteurs et pays émergents et les domaines et entreprises à la pointe.

Tout d'abord, nous pouvons dire que la recherche sur les scaffolds a commencé à être bénéfique pour les entreprises en 2001, date à laquelle les premiers brevets ont été déposés. L'année 2008 a été la plus prolifique en termes de dépôts de brevets. Pour 2011, l'année est tout juste commencée et le pourcentage de brevets est déjà très haut. On peut penser que l'attribution d'une date de dépôt cesse avant la fin de l'année civile et qu'elle reprend au début de la suivante en rattrapant le retard d'un seul coup, ce qui donne l'impression d'un dépôt important en ce début d'année.

Grâce au réseau sémantique, nous avons pu constater qu'initialement la recherche se focalisait sur « antibody ». Au fil des années, de nouveaux sujets ont émergé, élargissant la recherche autour des termes « antigen » et « scaffolds » et révélant ainsi un intérêt grandissant pour de nouvelles applications des peptides et des nucléotides liants.

Outre les Etats-Unis, qui occupent une position dominante au sein de la recherche, d'autres pays comme la Suisse et l'Allemagne se détachent et investissent dans le domaine des scaffolds.

Des pays comme Israël et le Portugal, que nous ne pensions pas retrouver dans ce domaine, se sont finalement révélés très impliqués et très prometteurs. Les trois principaux inventeurs du secteur sont américains et travaillent dans des entreprises américaines.

- Chawla Narinder et Eliott Vicki travaillent dans le groupe Incyte et n'ont publié de brevets que pour cette entreprise.

- Rosen Craig a commencé sa carrière chez Human Genome Sciences et en a été le premier conseiller scientifique. Il a fondé Co-Genosys, spécialisée dans la pharmacologie et rachetée par une entreprise israélienne Teva Biopharmaceutical, permettant un bénéfice pour le groupe HGS qui avait gardé 13% de son capital.

La tendance générale montre que les inventeurs ne publient que pour une seule entreprise et préfèrent ainsi se spécialiser. D'un autre côté, nous avons pu mettre en évidence de grandes équipes de collaboration. En recherchant les lieux où les inventeurs travaillent, nous remarquons que les équipes se forment au sein d'une même entreprise et certains inventeurs ayant changé de compagnie, font le lien entre deux équipes. Par exemple, Baker a travaillé pour Human Genome Sciences et pour Genentech, il fait le lien entre deux équipes de recherche. L'analyse des entreprises déposantes met en évidence deux grands groupes dominants : Human Genome Sciences et Incyte. De plus, nous avons remarqué qu'aucune de ces entreprises ne collaborait, ce qui montre que ce domaine est très concurrentiel. Grâce à cette étude, nous avons pu appréhender la veille technologique et son intérêt. Les outils mis à notre disposition nous ont permis d'analyser et d'extraire des informations stratégiques à partir d'un corpus de données et ceci sans avoir à lire tous les brevets. Le logiciel Xplor EveryWhere permet de mieux visualiser les données. Son avantage est d'être accessible par tous grâce à sa facilité de maniement et partout grâce aux applications sur les smartphones. Ainsi, les décideurs ont accès à l'information structurée lors de leurs déplacements.

4.6. Evaluation de Xplor EveryWhere

4.6.1. Méthodologie

Pour l'évaluation de Xplor EveryWhere, un questionnaire a été réalisé afin de tester cet outil dans sa globalité dans des bureaux d'études en analyses de données (annexe p. 209). Les objectifs de cette évaluation sont :

- Cibler la population, à savoir si l'outil est destiné à un public plus ou moins restreint ;
- Tester toutes les fonctionnalités proposées ;
- Recenser les lacunes de l'outil,
- Etre à l'écoute des futures utilisateurs, quelle que soit leur origine scientifique, leur domaine d'expertise, afin de prendre en compte leurs besoins ;

Pour cette évaluation, un sondage est réalisé. Il s'agit d'une enquête ponctuelle qui consiste à construire un échantillon à partir d'une population ciblée. Les personnes faisant partie de l'échantillon sont interrogées à l'aide d'un questionnaire et les réponses obtenues sont ensuite extrapolées à la population de la base.

Chaque groupe de bureau d'étude doit effectuer une analyse sous Xplor EveryWhere. Chaque expérimentation est chronométrée afin d'évaluer la difficulté d'utilisation, selon le profil de l'utilisateur. Un utilisateur doit remplir une fiche de renseignement dans laquelle il indique son niveau d'enseignement supérieur, son secteur d'activité et son niveau de connaissance dans le domaine de l'analyse de données.

Les différentes fonctionnalités sont testées et l'utilisateur doit évaluer le niveau de difficulté pour les appliquer, en cochant, pour chacune d'entre elles, la case correspondante, à savoir :

- Très facile
- Facile
- Moyenne

- Difficile
- Très difficile

Les questions posées mènent la personne interrogée à effectuer une analyse, portant dans un premier temps sur l'analyse statique puis l'étude évolutive. Enfin, l'utilisateur interrogé évalue librement notre contribution en qualifiant l'intérêt de Xplor EveryWhere, ses points forts, ses points faibles et les recommandations pour faire évoluer cet outil.

L'enquête est réalisée en suivant les règles suivantes :

- Choisir des mots adaptés à la population interrogée.
- Utiliser un langage spécifique à une population ayant des bases scientifiques.
- Les questions sont indépendantes et le formulaire est divisé en plusieurs parties autonomes.
- Limiter l'appel à la mémoire.
- Ordonner les questions par thème. Pour chaque partie du questionnaire, les questions sont ordonnées selon un ordre chronologique d'utilisation de l'outil.
- Tester le questionnaire auprès d'un échantillon réel et réduit. Deux experts du domaine ont testé dans un premier temps le questionnaire afin de pouvoir apprécier le niveau de difficulté de l'enquête, le temps d'application, la cohérence des points étudiés.
- S'assurer que toutes les questions sont intelligibles.
- Corriger le questionnaire et refaire le test si nécessaire.

Plusieurs corrections ont été effectuées suite aux premiers tests du questionnaire sur l'échantillon réduit, d'un point de vue de l'adéquation entre les questions et les résultats graphiques, mais aussi au niveau de la précision de certaines formulations.

La synthèse des questions de l'enquête est présentée dans le tableau suivant :

	Profil utilisateur	Espace statistique descriptive	Espace carte géostratégique	Espace Filtrage	Espace administration	Synthèse	Total
Réponse à choix	6	4	6	8	2	12	38
Réponse libre	0	0	0	0	0	2	2
Nombre de questions	6	4	6	8	2	14	40

Tableau 7. Nombre de questions du questionnaire d'évaluation

4.6.2. Population étudiée

Les critères de qualité des sondages sont relativement clairs sur le plan théorique. La base du sondage doit idéalement comprendre tous les membres de la population à représenter.

Le choix de la population interrogée consiste à fixer des quotas basés sur les orientations scientifiques. La sélection des personnes interrogées repose sur les individus qui répondent aux critères permettant de remplir les quotas fixés. Ainsi, nous fixons plusieurs catégories de personnes interrogées selon leur domaine de compétence et de niveau supérieur ou égal à bac+2.

Elles constituent ainsi des sous-ensembles d'une population donnée, considérées comme un véritable modèle réduit de la population étudiée :

- Les informaticiens ;
- Les statisticiens et informaticiens décisionnels ;
- Les spécialistes du décisionnel ;
- Les veilleurs ;
- Les spécialistes de l'intelligence économique ;
- Les spécialistes de la propriété intellectuelle et des nouvelles technologies.

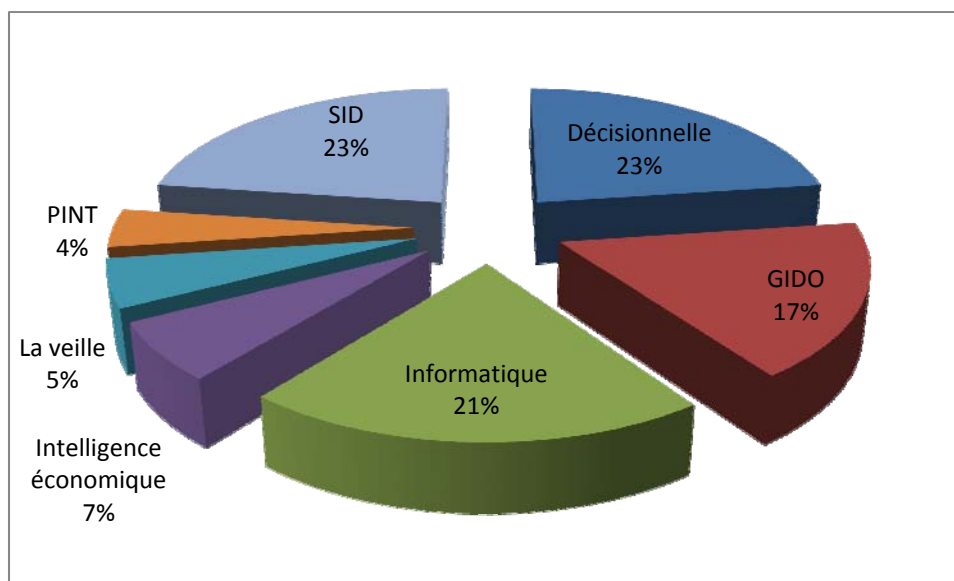


Figure 91. La population de l'enquête

Le tableau 7 indique, selon les disciplines, le nombre de personnes interrogées, la proportion ayant déjà manipulé Tétralogie ou d'autres outils de décisionnel.

Disciplines	Nombre de personne interrogées	% ayant manipulé Tétralogie sur le total par catégorie	% ayant manipulé Tétralogie sur le total global	% ayant manipulé d'autres outils de décisionnel sur le total par catégorie	% ayant manipulé d'autres outils de décisionnel et Tétralogie sur le total global	% n'ayant jamais utilisé un outil de décisionnel sur le total global
Décisionnel	21	0	0	100	23	0
GIDO	16	75	13	0	0	4
Informatique	19	0	0	100	21	0
Intelligence économique	6	0	0	0	0	6
La veille	5	0	0	100	5	5
PINT	4	100	4	0	0	0
SID	21	76	23	67	15	6
Total	92		40		64	23

Tableau 8. *Caractéristiques de la population étudiée par catégorie*

On remarque que 40,22% des personnes interrogées ont utilisé Tétralogie, 64,13% ont utilisé Tétralogie et d'autres outils de décisionnel (SPSS, SAS, R, etc.). En revanche, 22,83% n'ont jamais utilisé un outil décisionnel.

4.6.3. Résultats de l'enquête

Le premier point étudié est la durée nécessaire à chaque profil de personnes interrogées pour effectuer l'étude, afin de pouvoir situer le niveau de difficulté, selon l'utilisateur. Les résultats sont restitués dans le tableau 9, dont la dernière colonne correspond aux résultats des 10 personnes expertes du domaine qui ont participé dans un premier temps notre sondage afin de l'améliorer et qui servent de comparatif.

	Décisionnel	GIDO	Informatique	IE	La veille	PINT	SID	Experts
Durée minimale	36	46	25	50	52	43	41	20
Durée maximale	45	60	31	66	72	68	52	26
Moyenne	39,76	52,00	27,74	59,83	65,40	51,25	47,09	22,80
Ecart-type	3,05	4,75	2,16	5,67	8,26	11,59	2,95	2,20

Tableau 9. *Résultats pour la durée de l'expérimentation*

Pour cet aspect temporel, la difficulté d'utilisation reste raisonnable, par comparaison entre le temps d'expérimentation pour les personnes habituées à manipuler les outils de décisionnel, surtout Tétralogie et celles qui ne le sont pas.

Afin de restituer les résultats de notre enquête, les réponses aux questions sont croisées avec le profil des personnes interrogées. Les repères utilisés correspondent au niveau de difficulté des tâches, à savoir :

- Très facile (TF)
- Facile (F)
- Moyen (M)
- Difficile (D)
- Très difficile (TD)

➤ Niveau de difficulté des fonctionnalités XPlor EveryWhere

Les 26 questions sont classées en quatre types :

- Fonctionnalité menu statistique descriptive (tableau 10).
- Fonctionnalité menu carte géostratégique (tableau 11).
- Fonctionnalité menu filtrage (tableau 12).
- Fonctionnalité menu administration (tableau 13).

Questions	Codage
Lancer le menu statistique descriptif	Q1
Lancer les statistiques sur un champ	Q2
Faire le retour vers un autre champ	Q3

L'interprétation des résultats	Q4
--------------------------------	----

Tableau 10. Questions du menu statistique descriptif

Questions	Codage
Lancer le menu carte géostratégique	Q5
L'interprétation des résultats de la carte standard	Q6
Lancer la carte évolutive	Q7
L'interprétation des résultats de la carte évolutive	Q8
Lancer la carte google maps	Q9
L'interprétation des résultats de la carte google maps	Q10

Tableau 11. Questions du menu carte géostratégique

Questions	Codage
Lancer le menu filtre	Q11
Filtrer sur les auteurs	Q12
Visualiser l'environnement d'un auteur	Q13
Visualiser le profil	Q14
Visualiser le réseau social	Q15
Visualiser le réseau sémantique	Q16
Visualiser le réseau international	Q17
Visualiser le réseau international évolutif	Q18

Tableau 12. Questions du menu filtrage

Questions	Codage
Déposer une étude	Q19
Suivi des tâches	Q20

Tableau 13. Questions du menu administration

La décomposition matricielle s'effectue par critère d'évaluation. Chaque matrice correspond à une appréciation spécifique de difficulté et chacune des valeurs des tableaux recense le nombre de réponses pour un type de difficulté, de la catégorie en colonne, pour le critère en ligne.

Q1								
Difficulté	Décisionnel	GIDO	IE	Informatique	La veille	PINT	SID	Total
F		9	6		5	4	8	32
M		3						3
TF	21	3		19			15	58
Total	21	16	6	19	5	4	23	93

Tableau 14. Matrice de réponse sur Q1, croisant la difficulté et les catégories d'individus

Q2								
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	SID	Grand Total
F		10	6		5	4	10	35

M		2					2
TF	21	3		19		13	56
Grand Total	21	16	6	19	5	4	93

Tableau 15. Matrice de réponse sur Q2, croisant la difficulté et les catégories d'individus

Q3							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
D		2					2
F		7				4	19
M		5	6		5		17
TF	21	2		19			56
Grand Total	21	16	6	19	5	4	94

Tableau 16. Matrice de réponse sur Q3, croisant la difficulté et les catégories d'individus

Q4							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
D		1					1
F		11	6		5	4	41
M		3					4
TF	21	1		19			48
Grand Total	21	16	6	19	5	4	94

Tableau 17. Matrice de réponse sur Q4, croisant la difficulté et les catégories d'individus

Q5							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
F		9	6		5	4	32
M		3					4
TF	21	4		19			58
Grand Total	21	16	6	19	5	4	94

Tableau 18. Matrice de réponse sur Q5, croisant la difficulté et les catégories d'individus

Q6							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
D							1
F		11	6		5	4	38
M		3					6
TF	21	1		19			47
Grand Total	21	16	6	19	5	4	92

Tableau 19. Matrice de réponse sur Q6, croisant la difficulté et les catégories d'individus

Q7							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand Total
F		10	6		5	4	35
M		2					3
TF	21	4		19			56
Grand Total	21	16	6	19	5	4	94

Tableau 20. Matrice de réponse sur Q7, croisant la difficulté et les catégories d'individus

Q8							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand Total
D		2					3
F		11	6		5	4	38
M		2					5
TF	21	1		19			48
Grand Total	21	16	6	19	5	4	94

Tableau 21. Matrice de réponse sur Q8, croisant la difficulté et les catégories d'individus

Q9							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand Total
D							1
F		6				4	16
M		6	6		5		17
TF	21	2		19			57
Grand Total	21	14	6	19	5	4	91

Tableau 22. Matrice de réponse sur Q9, croisant la difficulté et les catégories d'individus

Q10							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand Total
D							2
F		8	6		5	4	33
M		2					7
TF	21			19			46
Grand Total	21	10	6	19	5	4	88

Tableau 23. Matrice de réponse sur Q10, croisant la difficulté et les catégories d'individus

Q11							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
F		8	5	1	5	4	12 35
M		7					1 8
TF	21	1	1	18			10 51
Grand Total	21	16	6	19	5	4	23 94

Tableau 24. Matrice de réponse sur Q11, croisant la difficulté et les catégories d'individus

Q12							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
D		1					1
F		10	5	1	5	4	12 37
M		4					3 7
TF	21	1	1	18			7 48
Grand Total	21	16	6	19	5	4	22 93

Tableau 25. Matrice de réponse sur Q12, croisant la difficulté et les catégories d'individus

Q13							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
F		11	5	1	5	4	13 39
M		2					3 5
TF	21	3	1	18			6 49
Grand Total	21	16	6	19	5	4	22 93

Tableau 26. Matrice de réponse sur Q13, croisant la difficulté et les catégories d'individus

Q14							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
F		10	5	1	5	4	15 40
M		3					2 5
TF	21	3	1	18			6 49
Grand Total	21	16	6	19	5	4	23 94

Tableau 27. Matrice de réponse sur Q14, croisant la difficulté et les catégories d'individus

Q15							
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	Grand SID Total
F		7		1		4	15 27
M		6	5		5		2 18

TF	21	2	1	18			6	48
Grand Total	21	15	6	19	5	4	23	93

Tableau 28. Matrice de réponse sur Q15, croisant la difficulté et les catégories d'individus

Q16								
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	SID	Grand Total
D							1	1
F		11	5	1	5	4	14	40
M		2					1	3
TF	21	2	1	18			7	49
Grand Total	21	15	6	19	5	4	23	93

Tableau 29. Matrice de réponse sur Q16, croisant la difficulté et les catégories d'individus

Q17								
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	SID	Grand Total
F		11	5	1	5	4	13	39
M		2					2	4
TF	21	2	1	18			6	48
Grand Total	21	15	6	19	5	4	21	91

Tableau 30. Matrice de réponse sur Q17, croisant la difficulté et les catégories d'individus

Q18								
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	SID	Grand Total
D							1	1
F		8	5	1	5	4	12	35
M		1					2	3
TF	21		1	18			3	43
Grand Total	21	9	6	19	5	4	18	82

Tableau 31. Matrice de réponse sur Q18, croisant la difficulté et les catégories d'individus

Q19								
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	SID	Grand Total
D		8	6		5		2	21
F	21	2		19		4	11	57
M		6					8	14
TF							2	2
Grand Total	21	16	6	19	5	4	23	94

Tableau 32. Matrice de réponse sur Q19, croisant la difficulté et les catégories d'individus

Q20								
Difficulté	Décisionnelle	GIDO	IE	Informatique	La veille	PINT	SID	Grand Total
D		1						1
F	21	5		19		4	15	64
M		9	6		5		6	26
TF							2	2
Grand Total	21	15	6	19	5	4	23	93

Tableau 33. Matrice de réponse sur Q20, croisant la difficulté et les catégories d'individus

Ces résultats nous montrent que 49% des appréciations sont notées très faciles, 39% faciles, 8% moyennes et 2% difficiles. Cela montre que le prototype XPlor EveryWhere propose des fonctionnalités simples à utiliser et à exploiter.

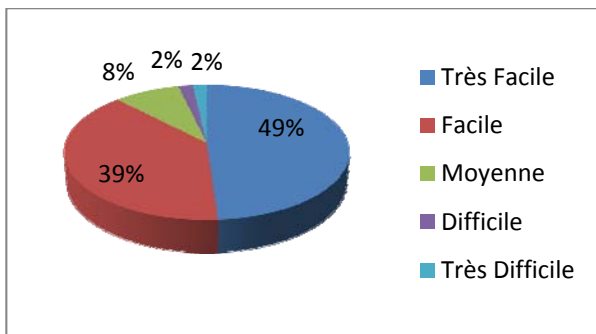


Figure 92. Difficulté des fonctionnalités de XPlor EveryWhere

Pour les documentalistes, plus de la moitié des fonctionnalités de XPlor EveryWhere sont faciles, 10 % sont très faciles, 23% moyennes et 5% sont difficiles. Rappelons qu'ils ont manipulé Tétralogie, avec un niveau débutant en informatique et en décisionnel.

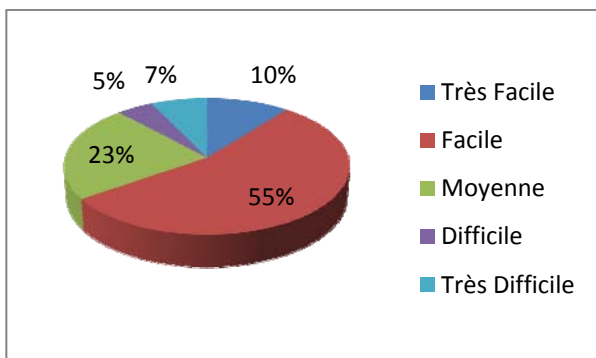


Figure 93. Difficulté des fonctionnalités de XPlor EveryWhere pour les documentalistes

Pour les statisticiens et les informaticiens en décisionnel, la moitié des fonctionnalités de XPlor EveryWhere sont faciles, 36 % sont très faciles, 10% moyennes et 2% sont difficiles. Rappelons qu'ils ont manipulé Tétralogie et d'autres outils de décisionnel comme SAS, SPSS, R, etc.

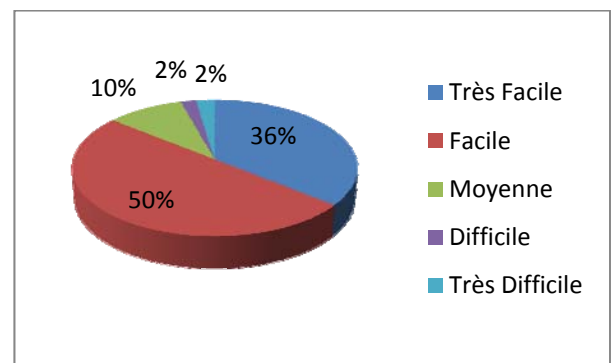


Figure 94. Difficulté des fonctionnalités de XPlor EveryWhere pour les statisticiens et les informaticiens en décisionnel

Pour les spécialistes de l'intelligence économique, 69% ont trouvé que les fonctionnalités de XPlor Everywhere sont faciles, 7 % sont très faciles, 19% moyennes et 5% sont difficiles.

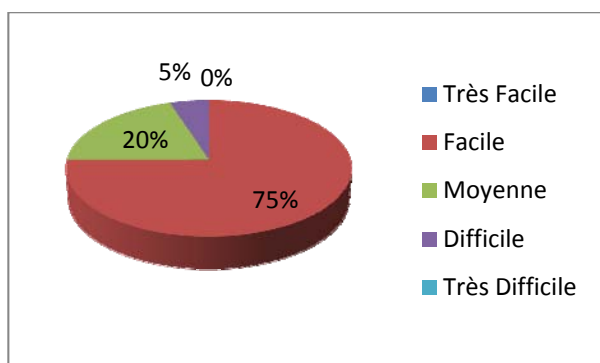


Figure 95. Difficulté des fonctionnalités de Xplor EveryWhere pour les spécialistes de l'intelligence économique

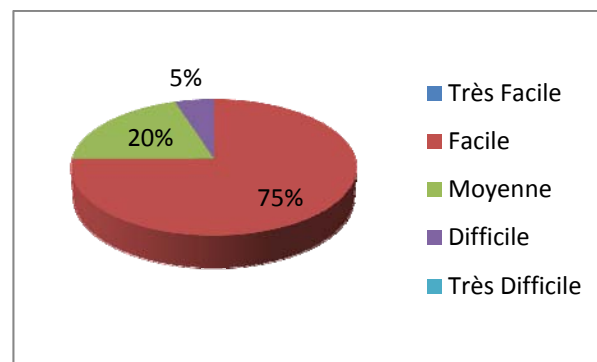


Figure 96. Difficulté des fonctionnalités de Xplor EveryWhere pour les spécialistes de la veille

Pour les spécialistes de la veille, 75% ont trouvé que les fonctionnalités de Xplor EveryWhere sont faciles, 20% moyennes et 5% sont difficiles.

De manière générale, la manipulation des fonctionnalités de Xplor EveryWhere n'implique pas d'obstacle pour les utilisateurs, ne requiert pas un bon niveau en informatique, ni en analyse de données. Cela est notamment dû à l'ergonomie et à l'interactivité proposée. La figure suivante nous montre les difficultés des fonctionnalités pour les différents individus interrogés.

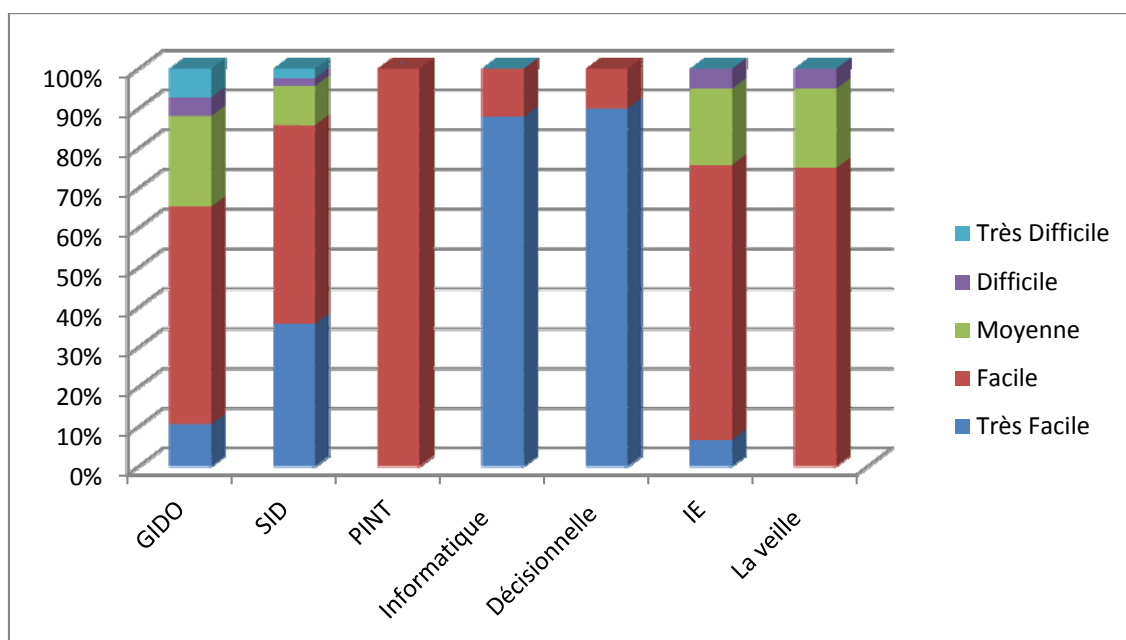


Figure 97. Difficulté des fonctionnalités de Xplor EveryWhere pour les différents individus

➤ Analyse de l'utilité de Xplor EveryWhere

La dernière partie de l'enquête porte sur le ressenti des testeurs vis-à-vis de Xplor EveryWhere. Le questionnaire propose 11 questions d'ordre général.

Questions	Codage
-----------	--------

Comment qualifiez-vous cet outil ?	QG1
La représentation des données	QG2
L'outil facilite-t-il l'exploration des données	QG3
La détection des différentes tendances, des signaux forts est-elle facilitée ?	QG4
Qualifieriez-vous l'outil d'interactif ?	QG5
Qualifieriez-vous l'outil d'ergonomique ?	QG6
L'accès aux différentes fonctionnalités est-il difficile ?	QG7
Avez-vous déjà effectué une analyse de données avec la plateforme « Tétralogie » ?	QG8
Si oui, que pensez-vous de l'ajout de ce module à cette plateforme ?	QG9
Que pensez-vous de la mobilité de Xplor (la version pour téléphone portable) ?	QG10
Qualifiez-vous Xplor de plateforme d'Intelligence Economique ?	QG11
Si votre domaine d'activité n'est pas directement lié à la visualisation de données, pensez-vous que cet outil pourrait vous être utile professionnellement ?	QG12

Tableau 34. Questions de l'utilité de Xplor EveryWhere

	Indispensable (%)	Très Utile (%)	Assez Utile (%)	Utile / Oui (%)	Sans grand Intérêt (%)	Inutile / Non (%)	Adaptée (%)	Sans avis (%)
QG1		35,90	61,54					2,56
QG2							89,74	10,26
QG3				94,87				5,13
QG4				61,54		12,82		25,64
QG5				84,62		10,26		5,13
QG6				84,62				15,38
QG7				7,69		87,18		5,13
QG8				70,00		30,00		
QG9	12,82	46,15		38,46				2,56
QG10	2,56	23,08		28,21	17,95			28,21
QG11				64,10		2,56		33,33
QG12				58,21		20,51		21,28

Tableau 35. Synthèse sur l'évaluation de Xplor EveryWhere

Les résultats du tableau de synthèse montrent une importante satisfaction de l'expérimentation de Xplor EveryWhere. 35.90% des utilisateurs trouvent que le prototype proposé est très utile, 61.54% assez utile. La représentation des données est adaptée avec 89,74% de satisfaction. 94,87% des différents utilisateurs (informaticiens, statisticiens, spécialistes de l'IE, de la veille et documentalistes) ont trouvé que le prototype facilite l'exploitation des données. 84,62% trouvent l'outil interactif et ergonomique.

La majorité des utilisateurs de Tétralogie trouve que Xplor EveryWhere est indispensable avec 12.82%, très utile avec 46.15% et utile 38.46%. La mobilité semble indispensable pour 2.56% des utilisateurs, très utile pour 23.08%, utile pour 28.21% contre sans grand intérêt pour 17.95%. 64,10% des utilisateurs qualifient le prototype de SIE. 58,21% réutiliseront Xplor EveryWhere dans leurs domaines professionnels contre 20.51% qui ne le recommandent pas pour le moment et 21.28% sans avis.

Intéressons-nous maintenant aux deux questions libres, qui traitent de l'aspect collaboratif du prototype et des conseils des utilisateurs pour l'améliorer. La majorité des utilisateurs trouve que l'aspect collaboratif est indispensable, important si l'accès est bien sécurisé. L'espace collaboratif est clair, lisible et simple ce qui met l'interprétation des résultats à la portée de tout le monde même s'il n'est pas spécialiste en décisionnel. La vision de l'évolution des tâches d'analyse est très importante. Cependant elle peut être encore améliorée, sous forme d'un bureau virtuel à distance, pour faciliter les réunions des décideurs en déplacement.

Les utilisateurs demandent l'ajout d'un moteur interne de recherche, d'ajuster les graphes, d'ajouter des légendes aux cartes géostratégiques statiques et évolutives et d'ajouter une aide sur le prototype afin de faciliter l'exploitation de toutes les fonctionnalités. Ces conseils et recommandations nous permettent d'établir des plans de prévisions sur les axes à améliorer, tels que l'ergonomie, l'accès aux fonctionnalités et davantage d'interactivité.

4.7. Synthèse

L'expérimentation du système d'intelligence économique XPlor EveryWhere montre une satisfaction des utilisateurs quel que soit leur domaine d'activité : informaticiens, statisticiens, analystes, veilleurs ou documentalistes, etc. La majorité qualifie le prototype d'interactif, d'ergonomique, de collaboratif, d'IE, et d'aide à la décision.

Le temps mis par chaque catégorie de la population étudiée a été analysé afin d'étudier les éventuelles difficultés, selon le domaine professionnel. Nous nous sommes rendus compte que les résultats ne variaient pas beaucoup.

L'aspect mobilité présente pour les utilisateurs une excellente valeur ajoutée pour les SIE. Ils peuvent maintenant bénéficier de l'avantage d'un système portable sécurisé pour rester à l'écoute de l'environnement. Cette expérimentation nous a permis de valider le prototype proposé et de prendre en compte des mesures pour l'amélioration de XPlor EveryWhere. Nous détaillons ces mesures dans nos perspectives de recherche, dans la dernière partie de ce manuscrit.

Conclusion et perspectives

«Ce n'est pas la fin. Ce n'est même pas le commencement de la fin. Mais, c'est peut-être la fin du commencement» (Winston Churchill, 1953)

5.1.	Résumé des contributions	161
5.2.	Perspectives de recherches	166

Les entreprises font face aujourd'hui à une concurrence accrue sur des marchés extrêmement dynamiques et imprévisibles : nouveaux entrants, fusions et acquisitions, baisses tarifaires brutales, évolution rapide des modes de consommation et des valeurs, fragilité des marques et de leur réputation, ... Les facteurs de changement et de risques externes n'ont jamais été aussi nombreux, accrus par le contexte de la crise financière actuelle. Pour mieux anticiper ces risques et guetter les opportunités, l'IE semble être la seule à l'avant-garde.

Le rapport Henri Martre définit l'IE comme étant l'ensemble des actions coordonnées de recherche, de traitement et de distribution de l'information utile aux acteurs pour permettre l'action et la prise de décision. Ces diverses actions sont menées légalement avec toutes les garanties de protection nécessaires à la préservation du patrimoine de l'entreprise, dans les meilleures conditions de délais et de coûts. L'information utile est celle dont ont besoin les différents niveaux de décision de l'entreprise ou de la collectivité, pour élaborer et mettre en œuvre de façon cohérente la stratégie et les tactiques nécessaires à l'atteinte des objectifs définis par l'entreprise dans le but d'améliorer sa position dans son environnement concurrentiel. Ces actions, au sein de l'entreprise, s'ordonnent autour d'un cycle ininterrompu, générateur d'une vision partagée des objectifs de l'entreprise.

5.1. Résumé des contributions

Pour arriver à ces fins, nos travaux consistent à proposer un système d'intelligence économique pour mettre en ligne ce type de structure, comme nous l'avons vu dans le premier chapitre. Ce qui permet aux décideurs nomades de consulter les données stratégiques lors de leurs déplacements en toute sécurité et mobilité.

Ce système repose sur l'étendu des étapes d'extraction et de visualisation des données. Comme nous l'avons vu dans l'expérimentation du quatrième chapitre, suite à une collecte des textes semi-structurés, sur un sujet défini, les informations sont extraites et prétraitées via la plateforme de veille stratégique Tétralogie, développée par notre équipe. Notre démarche d'analyse s'appuie en premier lieu sur les caractéristiques des données manipulées, en l'occurrence les données temporelles. Les entités sont croisées en matrices de co-occurrences, dont la valeur indique le nombre de fois ou la présence simultanée de deux informations est observée. Suivant la granularité temporelle choisie, la matrice peut être décomposée en plusieurs périodes. Une fois les entités organisées, notre contribution repose sur

l'étude des différentes relations caractérisant l'environnement informationnel. L'intérêt de notre contribution repose sur trois axes principaux :

- L'entrepôt de données stratégique, qui présente une vue fédératrice, homogène et partagée par les différentes données quelles que soient leurs sources. Il permet d'une part l'extraction et le stockage des données sources structurées sous forme d'une représentation multidimensionnelle les traitements de création de l'entrepôt de données d'autre part. Il repose sur la création de matrices à deux et trois dimensions (contingences, co-occurrences, présence absence) à partir de la vue unifiée. Notre approche consiste à utiliser les composantes du système Tétralogie pour la création des matrices à deux dimensions et réajuster la structure des résultats obtenus pour faciliter le transfert des données dans l'entrepôt de données du prototype. Cet entrepôt ne contient que les données sous forme relationnelle tel que : N°doc-Auteur, Auteur-Auteur-Date, etc. Cette représentation synthétise toutes les relations existantes entre les différents attributs constituant les documents à analyser. La base de données est interfacée grâce au système XPlor EveryWhere sur Intranet ou Internet, afin que l'utilisateur puisse lui-même mener ses propres investigations.
- La sécurisation des données (IE en tant que produit) et des procédés (IE en tant que méthode de traitement). Nous proposons une approche de contrôle d'accès aux données et aux traitements dans un SIE. Il s'agit d'un modèle générique pour intégrer le contrôle d'accès dans le cycle de vie des SIE afin que toutes les tâches dans un processus de veille soient exécutées uniquement par les utilisateurs autorisés et habilités.
- La mobilité qui présente un choix stratégique pour notre SIE "XPlor EveryWhere", dont le but est de permettre aux utilisateurs de nos solutions de veille de continuer à rechercher, surveiller, valider et rediffuser des informations stratégiques au cours de leurs déplacements. Ils n'ont ainsi plus besoin d'être assis face à leur ordinateur pour accéder à des données utiles dans l'instant (préparation d'une réunion, nouvel ordre du jour, information sur un interlocuteur, une technologie, un marché, demande urgente d'une analyse ou d'un focus spécifique). Il leur suffit simplement d'utiliser discrètement, en toute sécurité, leur appareil mobile.

Le SIE XPlor EveryWhere vient compléter l'offre d'outils dédiés à l'IE, par la possibilité de gérer et d'exploiter **n'importe où** l'environnement informationnel, et ce de manière **interactive** et en toute **sécurité**.

Cette expérimentation a été effectuée dans le cadre des bureaux d'études du M1-SID (Statistique Informatique Décisionnelle), M2-PINT (Propriété Intellectuelle et Nouvelle Technologie) et DUT info-com GIDO (Gestion de l'Information et du Document), formations pour lesquelles j'ai dispensé des cours à l'université Paul Sabatier – Toulouse III

Notre contribution est expérimentée à travers un sondage réalisé auprès d'une population ciblée. Cette enquête repose sur des analyses effectuées au sein des bureaux d'études du M1-SID (Statistique Informatique Décisionnelle), M2-PINT (Propriété Intellectuelle et Nouvelle Technologie) et DUT info-com GIDO (Gestion de l'Information et du Document), formations pour lesquelles j'ai dispensé des cours à l'université Paul Sabatier – Toulouse III. Ces analyses stratégiques via notre SIE ont permis d'évaluer la performance de l'outil sous différents aspects. Les résultats prouvent l'intérêt de XPlor EveryWhere dans un contexte d'Intelligence Economique ou plus précisément de veille stratégique.

En plus de cette évaluation par expérimentation, nous effectuons une seconde évaluation en nous basant sur les critères proposés par la société FULD & COMPANY. L'un des leaders mondiaux de l'IE particulièrement dans l'analyse des marchés et des concurrents ainsi que la veille stratégique et concurrentielle. Depuis 1998, cette compagnie a toujours évalué les outils technologiques dédiés à l'IE. Le dernier rapport d'évaluation de la compagnie « Intelligence Software Report 2008-2009 », porte sur

les outils de traitements avancés de l'information utilisés dans le cadre d'une démarche d'IE par les entreprises. Il se base sur l'auto-évaluation de 480 entreprises dans le monde.

L'évaluation des outils se base sur plus de 120 critères (représentées sous forme de fonctionnalités) qui couvrent les cinq étapes du cycle de traitement de l'information dans le cadre de l'IE, à savoir : Expression du besoin, Collecte d'Information publiée, Collecte d'Information primaire, Traitement et Analyse, Diffusion.

Expression du besoin : Cette partie consiste à définir : la stratégie et les objectifs de l'étude ainsi que son besoin informationnel, cela constitue le moteur des étapes de collecte et d'analyse.

Critères : l'expression du besoin est un processus purement humain, aucun logiciel ne peut définir dynamiquement les besoins d'une entreprise. Ils peuvent seulement offrir un cadre pour la gestion du projet. Dans ce contexte, les fonctionnalités que doit satisfaire l'outil sont :

- Une interface de gestion des droits d'accès des utilisateurs,
- Une interface pour introduire et gérer les orientations thématiques des axes de recherches,
- Une interface qui permet aux intervenants d'introduire et gérer les différents besoins informationnels liés à une étude,
- Une interface d'échange et de partage de données dans le cadre d'une étude pour permettre une collaboration entre les différents intervenants.

Collecte d'information publiée : Cette étape concerne l'identification des sources de données et la recherche d'information.

Critères : la collecte d'information se basent sur les types de source : Les données internes de l'entreprise : bases de données, documents de Microsoft Office, rapports, emails, graphiques et intranet,

- Les sources externes issues du web : pages web, sites internet, bases de données : Interroger simultanément plusieurs moteurs de recherches (requêtes inter-langues),
- Définir un profil utilisateur pour la recherche d'information (requêtes, thèmes, dates, etc.),
- Surveiller les flux RSS,
- Une interface de requêtes prédéfinies,
- Stockage des données,
- Filtrage automatique des données collectées selon les critères de l'utilisateur : concurrents, sources, dates, acteurs clés...etc.,
- Evaluation et validation des données collectées,
- Commentaires et annotation des données collectées,
- Classification et regroupement automatique des données collectées selon les critères de l'utilisateur.

Collecte d'information primaire : Concerne la collecte d'information informelle, auprès des acteurs internes à l'entreprise.

Critères : l'évaluation des savoir-faire de l'entreprise se base sur les points suivants :

- Recherche et collecte des données à partir de groupes, de nouvelles et de forums,
- Une interface pour saisir les informations informelles : entrevues, rapports, expertises, etc.

Traitement et analyse : Consiste à produire des informations utiles à partir des données collectées.

Critères : l'évaluation des techniques d'analyse des données qualitatives concerne les points suivants :

- Trier les données selon des règles définies par l'utilisateur,
- Interface pour le tri et la visualisation des données collectées,
- Visualisation des modèles : matrice SWOT, les cinq forces de Porter, etc.
- Affichage des données dans un ordre chronologique,
- Extraction des relations existantes entre les différents acteurs, dates, lieux, événements et autres corrélations importantes,
- Les méthodes de fouille de données textuelles pour extraire et analyser les variables sélectionnées par l'utilisateur,
- La capacité de transformer les données qualitatives aux quantitatives.
- Visualisation des résultats d'analyse sous différentes formes graphiques.

Diffusion : Porte sur les techniques utilisées pour exporter et diffuser les résultats obtenus.

Critères :

- Prise en compte d'une charte graphique ou de normes de présentation des rapports,
- Exporter les résultats sous Microsoft Office ou équivalent,
- Diffusion des rapports par mail, imprimé, etc.

Treize outils ont été sélectionnés parmi les 480 recensés. Ce sont des outils qui recouvrent une grande partie de la totalité des étapes du processus de l'IE. Chacun des outils est donc évalué selon les critères associés à ces cinq étapes. Les outils évalués ne couvrent pas de façon optimale toutes les étapes du cycle. Le pourcentage de couverture du cycle est en moyenne entre 50% et 80%. Ces outils se basent essentiellement sur les techniques de recherche d'information sur le web en utilisant les agents intelligents, seulement quatre d'entre eux (QL2 Software, Temis, Cymfony, Traction software) s'appuient sur les techniques de Text Mining et deux outils (Brimstone, Cymfony et Netro-City) sur les techniques de visualisation et de Reporting.

SIE 2008-2009	Expression des besoins	Collecte d'Information publiée	Collecte d'Information primaire	Traitement et Analyse	Diffusion	Moyenne de cycle
Brimstone	4	3	3	4	3	3,4
Cipher	5	3	4	3	4	3,8
Coemergence	5	3	5	2	4	3,8
Comintell	3	3	3	2	2	2,6
Global Intelligence	3	2	3	2	2	2,4
KB Intelligence	1	2	1	1	1	1,2
Netro City	1	3	4	3	3	2,8
QL2 Software	3	3	3	2	1	2,4
Stratégie Software	5	3	5	1	4	3,6
Temis	3	2	2	2	2	2,2
TNS Media / cymfony	3	3	3	2	3	2,8

Traction Software	4	5	4	2	4	3,8
Wincite	5	3	2	2	4	3,2

Tableau 36. Evaluation des SIE 2008-2009 par la société FULD & COMPANY⁹

Critères d'évaluation	XPlor EveryWhere	Moyenne de l'étape
Expression du besoin		
Une interface de gestion des droits d'accès des utilisateurs	5	3,25
Une interface pour introduire et gérer les orientations thématiques des axes de recherches	2	
Une interface de gestion des besoins informationnels liés à une étude	2	
Une interface d'échange et de partage de données	4	
Collecte d'information publiée		
Les données internes de l'entreprise : bases de données, documents Microsoft Office, rapports, emails, etc.	4	2,8
Les sources externes issues du web : pages web, sites internet, bases de données.	4	
Définir un profil utilisateur pour la recherche d'information (requêtes, thèmes, dates, etc.)	1	
Surveiller les flux RSS	4	
Une interface de requêtes prédéfinies	1	
Stockage des données	4	
Filtrage automatique des données collectées selon les critères de l'utilisateur : concurrents, sources, dates, acteurs	4	
Evaluation et validation des données collectées	3	
Commentaires et annotations des données collectées	0	
Classification et regroupement automatique des données collectées selon les critères de l'utilisateur	3	
Collecte d'information primaire		
Recherche et collecte des données à partir de groupes, de nouvelles et des forums	4	3,5
Une interface pour saisir les informations informelles : entrevues, rapports, expertises, etc.	3	
Traitement et analyse		
Trier les données selon des règles définies par l'utilisateur	4	4,25
Interface pour la visualisation de données collectées	5	
Visualisation des modèles : matrice SWOT, les cinq forces de Porter, etc.	0	
Affichage des données dans un ordre chronologique	5	
Extraction des relations existantes entre les différents acteurs, dates, lieux et autres corrélations importantes	5	
Les méthodes de fouille de données textuelles pour extraire et analyser les variables sélectionnées par l'utilisateur	5	
La capacité de transformer les données qualitatives aux quantitatives	5	
Visualisation des résultats d'analyse sous différentes formes graphiques	5	
Diffusion		
Prise en compte d'une charte graphique ou de normes de présentation des rapports	5	5

⁹ Source : le site <http://traction.tractionsoftware.com/traction/permalink/Press319>

Exporter les résultats sous Microsoft Office ou équivalent	5	
Diffusion des rapports par mail, imprimé, etc.	5	
Moyenne de cycle		3,76

Tableau 37. *Evaluation de Xplor Everywhere selon les critères de la société FULD & COMPANY*

Selon les critères de Fuld & Company, Xplor Everywhere serait parmi les quatre premiers SIE avec une moyenne de 3,76.

5.2. Perspectives de recherche

Comme nous l'avons souligné à plusieurs reprises dans ce manuscrit, l'analyse stratégique des différentes relations caractérisant l'environnement informationnel nous permet de comprendre les mécanismes de formation des réseaux sociaux, des réseaux sémantiques, des réseaux internationaux, etc., de connaître ou d'évaluer à un instant précis l'environnement d'une entreprise, d'un centre de recherche ou d'un chercheur, etc.

La visualisation des ces réseaux sur un grand écran ne présente aucun problème d'interprétation, de compréhension, ni de navigation. Cependant, on ne peut pas les visualiser sur les smartphones, ce qui présente un obstacle d'interprétation.

Pour cela, nous envisageons, le développement d'un algorithme d'optimisation des graphes. Cette optimisation se base sur les principes de permutation et d'hybridation des nœuds. En nous référant aux algorithmes génétiques de mutation et de reproduction.

Le deuxième axe de nos futurs travaux de recherche est de faire de Xplor Everywhere un référentiel des SIE basé sur l'ingénierie dirigée par les modèles ou IDM (en l'anglais MDE – Model Driven Engineering). Cette modélisation permet d'élaborer un modèle indépendant de l'information (Computation Independent Model, CIM), la transformation de celui-ci en modèle indépendant de la plate-forme (Platform Independent Model, PIM) et enfin la transformation de ce dernier en modèle spécifique à la plate-forme cible (Platform Specific Model, PSM).

Liste des figures

<i>Figure 1. La carte géostratégique de la recherche en intelligence économique</i>	33
<i>Figure 2. Les collaborations internationales de la recherche en intelligence économique</i>	33
<i>Figure 3. Emergence de l'IE</i>	34
<i>Figure 4. Processus de l'Intelligence Economique</i>	37
<i>Figure 5. Les environnements de l'entreprise (Magakian et Payaud, 2010)</i>	39
<i>Figure 6. Schéma de Porter - l'analyse de l'environnement (Placet et Rarivoson, 2011)</i>	40
<i>Figure 7. Typologie de la veille</i>	44
<i>Figure 8. Cycle de la veille</i>	46
<i>Figure 9. la phase analyse des besoins de la veille</i>	47
<i>Figure 10. Les sources formelle pour l'intelligence économique</i>	48
<i>Figure 11. La valeur d'intelligence économique (Coutenceau et al., 2010)</i>	50
<i>Figure 12. Le réseau sémantique de l'intelligence économique</i>	51
<i>Figure 13. Le cadre général de l'étude MIT 90 (Source : Pascal. Vidal, 2009, p. 25)</i>	56
<i>Figure 14. La dimension stratégique des systèmes d'information (Source : Pascal. Vidal, 2009, p. 26)</i>	57
<i>Figure 15. Stratégie et SI, boucles récursives (Extrapoler de la Source : Pascal. Vidal, 2009, p. 141)</i>	59
<i>Figure 16. Le processus du modèle XPlor (Ghallamalah, 2009)</i>	63
<i>Figure 17. Hiérarchie de concept associés à la spécification des besoins (Ghallamalah, 2009)</i>	64
<i>Figure 18. Exemple de la hiérarchie de concepts associés au sujet de la recherche scientifique marocaine</i>	65
<i>Figure 19. Le descripteur de la base Pubmed format Medline</i>	68
<i>Figure 20. Descripteur générique de trois bases : Pub-Med, Pascal et Science Direct</i>	68
<i>Figure 21. Principe des méta-données de premier et second niveau</i>	69
<i>Figure 22. Homogénéisation des sources d'information</i>	70
<i>Figure 23. Infrastructure globale d'information (Cigref, 2004)</i>	72
<i>Figure 24. Le couplage IE, Intelligence Collective et Mining (Ghallamalah, 2009)</i>	73
<i>Figure 25. L'architecture service de XPlor</i>	75
<i>Figure 26. Les différents utilisateurs de XPlor EvryWhere</i>	77
<i>Figure 27. Le stéréotype d'acteur « non-humain »</i>	77
<i>Figure 28. Diagramme de cas d'utilisation « demande d'une étude »</i>	78
<i>Figure 29. Diagramme de classes « demande d'une étude »</i>	79
<i>Figure 31. Architecture technique de XPlor EveryWhere</i>	81
<i>Figure 32. Architecture multi-couche de XPlor</i>	82
<i>Figure 33. Exemple de schéma en étoile (Teste, 2009)</i>	88
<i>Figure 34. Exemple d'une dimension « structure » (Teste, 2009)</i>	89
<i>Figure 35 : Exemple d'un galaxie (Teste, 2009)</i>	89
<i>Figure 36. Dépendances intra document à trois dimensions et l'élément temporel.</i>	91
<i>Figure 37. Cube de données</i>	92

<i>Figure 38. Cube sous forme d'une matrice carée</i>	93
<i>Figure 39. Cube sous forme d'une matrice rectangulaire</i>	93
<i>Figure 40. Cube d'évaluation de recherche par auteur</i>	94
<i>Figure 41. Interactions dans une plateforme d'IE (Bichard, 2005)</i>	101
<i>Figure 42. Représentation du workflow Xplor</i>	105
<i>Figure 43. Service de sécurité pour un cas d'utilisation</i>	107
<i>Figure 44. L'ontologie de sécurité (Beji et El kadhi, 2009)</i>	109
<i>Figure 45. Les étapes du modèle d'analyse multidimensionnelle XPlor</i>	110
<i>Figure 46. Les sous-modèles du modèle d'analyse multidimensionnelle XPlor EveryWhere</i>	110
<i>Figure 47. Les différentes sorties de la plateforme Tétralogie</i>	116
<i>Figure 48. Exemple de matrice de co-occurrence Auteur-Auteur</i>	117
<i>Figure 49. Page d'accueil du prototype XPlor V2</i>	118
<i>Figure 50. Page d'accueil du prototype XPlor EveryWhere</i>	120
<i>Figure 51. Service Administration du prototype XPlor EveryWhere</i>	120
<i>Figure 52. Co-occurrences multiples de Tétralogie</i>	121
<i>Figure 53. Les résultats Tétralogie de la fonction Co-occurrences Multiples</i>	122
<i>Figure 54. Les résultats de la fonction TetraToXPlor</i>	122
<i>Figure 55. Injection des données</i>	123
<i>Figure 56. Module gestion des analyses</i>	123
<i>Figure 57. Module gestion des droits d'accès</i>	124
<i>Figure 58. Module gestion de la file de travail</i>	124
<i>Figure 59. Gestion des droits d'accès</i>	124
<i>Figure 60. Gestion des compte utilisateur XPlor EveryWhere</i>	125
<i>Figure 61. Espace analyse de données</i>	125
<i>Figure 62. Déposer une étude</i>	126
<i>Figure 63. Suivi des analyses</i>	126
<i>Figure 64. Menu analyse</i>	126
<i>Figure 65. Choix du champ d'analyse (Top10, 20 ou 50)</i>	127
<i>Figure 66. Le tableau de synthèse</i>	127
<i>Figure 68. Choix de filtre</i>	128
<i>Figure 69. Résultats de filtre</i>	129
<i>Figure 70. Graphe Statique</i>	129
<i>Figure 71. Graphe dynamique</i>	130
<i>Figure 72. Carte Géostratégique</i>	130
<i>Figure 73. Le prototype XPlor EveryWhere pour iPhone</i>	131
<i>Figure 74. L'analyse Scaffolds</i>	133

<i>Figure 75. Exemple d'un brevets issus de la base WIPO</i>	136
<i>Figure 76. Descripteur de la base WIPO</i>	137
<i>Figure 77. Zoom de la matrice cooccurrence Inventeurs X Inventeurs</i>	138
<i>Figure 78. Graphe des clusters de réseau des inventeurs</i>	139
<i>Figure 79. Visualisation du cluster 1</i>	140
<i>Figure 80. Visualisation du cluster 2</i>	141
<i>Figure 81. Le top 10 des inventeurs</i>	141
<i>Figure 82. L'environnement de Chawla</i>	142
<i>Figure 83. Le profil, le réseau social et le réseau sémantique de Chawla</i>	142
<i>Figure 84. Réseau des collaborations internationales</i>	143
<i>Figure 85. La carte géostratégique des pays déposants de brevets sur les Scaffold</i>	143
<i>Figure 86. La carte géostratégique XPlor EveryWhere</i>	144
<i>Figure 87. Evolution du tops 10 des pays</i>	145
<i>Figure 88. Évolution des mots-clés employés dans l'ensemble des brevets en fonction des années</i>	145
<i>Figure 89. Spécialité de la France</i>	146
<i>Figure 90. Évolution des déposants mondiaux en fonction des années</i>	147
<i>Figure 91. La population de l'enquête</i>	150
<i>Figure 92. Difficulté des fonctionnalités de XPlor EveryWhere</i>	157
<i>Figure 93. Difficulté des fonctionnalités de XPlor EveryWhere pour les documentalistes</i>	157
<i>Figure 94. Difficulté des fonctionnalités de XPlor EveryWhere pour les statisticiens et les informaticiens en décisionnel</i>	157
<i>Figure 95. Difficulté des fonctionnalités de XPlor EveryWhere pour les spécialistes de l'intelligence économique</i>	158
<i>Figure 96. Difficulté des fonctionnalités de XPlor EveryWhere pour les spécialistes de la veille</i>	158
<i>Figure 97. Difficulté des fonctionnalités de XPlor EveryWhere pour les différents individus</i>	158
<i>Figure 98. Page d'accueil</i>	215
<i>Figure 99. Page de connexion</i>	215
<i>Figure 100. TableView des différents acteurs</i>	215
<i>Figure 101. Différents messages d'alerte possible</i>	216
<i>Figure 102. Barre de navigation de iXEW</i>	217
<i>Figure 103. Indicateur d'activité</i>	217
<i>Figure 104. Barre de recherche</i>	218
<i>Figure 105. Liste des agents</i>	219
<i>Figure 106. Histogramme de l'évolution des Etats-Unis</i>	220
<i>Figure 107. Courbes de l'évolution des Etats-Unis</i>	220
<i>Figure 108. Diagrammes de secteurs des Etats-Unis</i>	220
<i>Figure 109. La carte géostratégique de « SCAFFOLD »</i>	221
<i>Figure 110. Outils Leaks (1)</i>	222

<i>Figure 111. Outils Leaks (2)</i>	222
-------------------------------------	-----

Liste des tableaux

Tableau 1. Les Fonctions associées à la notion d'IE	36
Tableau 2. Analyse SWOT (Ducieux et al., 2010)	42
Tableau 3. Les rôles des systèmes d'information (Vidal et Petit, 2009)	56
Tableau 4. Type de matrices possibles selon le type de variables croisées	76
Tableau 5. La fonction combinatoire des fonctions (indicateurs)	100
Tableau 6. Les différents rôles et tâches associés à l'utilisateur de XPlor	106
Tableau 7. Nombre de questions du questionnaire d'évaluation	149
Tableau 8. Caractéristiques de la population étudiée par catégorie	151
Tableau 9. Résultats pour la durée de l'expérimentation	151
Tableau 10. Questions du menu statistique descriptif	152
Tableau 11. Questions du menu carte géostratégique	152
Tableau 12. Questions du menu filtrage	152
Tableau 13. Questions du menu administration	152
Tableau 14. Matrice de réponse sur Q1, croisant la difficulté et les catégories d'individus	152
Tableau 15. Matrice de réponse sur Q2, croisant la difficulté et les catégories d'individus	153
Tableau 16. Matrice de réponse sur Q3, croisant la difficulté et les catégories d'individus	153
Tableau 17. Matrice de réponse sur Q4, croisant la difficulté et les catégories d'individus	153
Tableau 18. Matrice de réponse sur Q5, croisant la difficulté et les catégories d'individus	153
Tableau 19. Matrice de réponse sur Q6, croisant la difficulté et les catégories d'individus	154
Tableau 20. Matrice de réponse sur Q7, croisant la difficulté et les catégories d'individus	154
Tableau 21. Matrice de réponse sur Q8, croisant la difficulté et les catégories d'individus	154
Tableau 22. Matrice de réponse sur Q9, croisant la difficulté et les catégories d'individus	154
Tableau 23. Matrice de réponse sur Q10, croisant la difficulté et les catégories d'individus	154
Tableau 24. Matrice de réponse sur Q11, croisant la difficulté et les catégories d'individus	155
Tableau 25. Matrice de réponse sur Q12, croisant la difficulté et les catégories d'individus	155
Tableau 26. Matrice de réponse sur Q13, croisant la difficulté et les catégories d'individus	155
Tableau 27. Matrice de réponse sur Q14, croisant la difficulté et les catégories d'individus	155
Tableau 28. Matrice de réponse sur Q15, croisant la difficulté et les catégories d'individus	156
Tableau 29. Matrice de réponse sur Q16, croisant la difficulté et les catégories d'individus	156
Tableau 30. Matrice de réponse sur Q17, croisant la difficulté et les catégories d'individus	156
Tableau 31. Matrice de réponse sur Q18, croisant la difficulté et les catégories d'individus	156
Tableau 32. Matrice de réponse sur Q19, croisant la difficulté et les catégories d'individus	156
Tableau 33. Matrice de réponse sur Q20, croisant la difficulté et les catégories d'individus	157
Tableau 34. Questions de l'utilité de XPlor Everywhere	159
Tableau 35. Synthèse sur l'évaluation de XPlor Everywhere	159

Tableau 36. *Evaluation des SIE 2008-2009 par la société FULD & COMPANY* _____ 165

Tableau 37. *Evaluation de XPlor Everywhere selon les critères de la société FULD & COMPANY* _____ 166

Liste des articles

Cette thèse a donnée lieu à la publication de 15 articles et deux prix :

Chapitre de livre international

- EL HADDADI A., DOUSSET B., BERRADA I., Discovering patterns in order to detect weak signals and define new strategies. Book: "Pattern Discovery Using Sequence Data Mining : Applications and studies" , IGI Global (à paraître en 2011).

Article de revue internationale

- EL HADDADI A., HATIM H., DOUSSET B., BERRADA I. and EL BAKKALI H. Reduce threats in competitive intelligence system: a generic information fusion access control model. Dans : International Journal of Database Management Systems, Academy & Industry Research Collaboration Center (AIRCC), Vol. 3 N. 1, p. 19-33, février 2011.

Article de revue nationale

- EL HADDADI A., DOUSSET B., BERRADA I., Sécurisation d'une plate-forme dédiée à l'intelligence économique. Revue Génie Logiciel, GL & SI, Vol. 94 , p. 22-25, Septembre 2010.

Articles de conférences et workshops internationaux

- EL HADDADI A., DOUSSET B., BERRADA I., Xplor EveryWhere – The Competitive Intelligence System for Mobile (regular paper). Dans : International Conference on Multimedia Computing and Systems (ICMCS 2011), Ouarzazate, Morocco, 07/04/2011-09/04/2011, IEEEExplore digital library, (à paraître en Avril 2011).
- EL HADDADI A., DOUSSET B., BERRADA I., Xplor : The competitive intelligence system based on a multidimensional analysis model (regular paper). SIIE 2011 (à paraître en février 2011).
- DOUSSET B., EL HADDADI A., MOTHE J., Content accessibility and Semantic networks processed on foreign natural language analysis (regular paper). SIIE 2011 (à paraître en février 2011).
- D. Tchunte, M. Canut, N. Jessel, A.Péninou, A.El Haddadi , Visualizing the evolution of user profiles from online social networks, 9-11 August, 2010, Odense , Denmark, IEEE Computer Society.

Articles de conférences et workshops nationaux

- EL HADDADI A., DOUSSET B., BERRADA I., Xplor EveryWhere a tool for competitive intelligence on the web and Mobile. VSST, Toulouse - France, 25 au 29 Octobre 2010
- HATIM H., EL HADDADI A., EL BAKKALI H. DOUSSET B. and BERRADA I. and Approche générique de contrôle d'accès aux données et aux traitements dans une plateforme d'intelligence économique. VSST, Toulouse - France, 25 au 29 Octobre 2010
- EL HADDADI A., DOUSSET B., BERRADA I., Sécurisation d'une plate-forme dédiée à l'intelligence économique. SEY- SY / INFORSID 2010, Marseille- France, 25 au 28 mai 2010
- EL HADDADI A., DOUSSET B., BERRADA I., LOUBIER I., Les multi-sources dans un contexte d'intelligence économique. EGC 2010, Hammamat – Tunisie, 26-29 Janvier 2010
- EL HADDADI A., BERRADA I., DOUSSET B., Portail Web de Veille Stratégique pou mobile, Les 1ères Journées Doctorales en Technologies et d'Information et de la Communication : JDTIC'09, Rabat– Maroc, 16-17 et 18 Juillet 2009.
- EL HADDADI A., DOUSSET B., BERRADA I., KASSOU I., Construction d'une ontologie de domaine fondée sut le Text Mining, 2ème séminaire de Veille Stratégique, Scientifique et Technologique : VSST'2009, Nancy – France, 30-31 mars 2009.
- EL HADDADI A., Portail Web de Veille Stratégique pou mobile, Forum Jeune Chercheur d'INFORSID, Toulouse – France, page 459-460, 26-27-28 et 29 Mai 2009.

Article de conférence sans actes publiés

- EL HADDADI A., DOUSSET B., Mise en ligne des analyse stratégique sur téléphone 3G. Journées Nationales du GDR en Intelligence Economique – Beaulieu sur Mer - France, 25 et 26 Novembre, 2009.

Prix de la meilleure présentation Poster. « Portail Web de Veille Stratégique pou mobile, Les 1ères Journées Doctorales en Technologies et d'Information et de la Communication ». Thème : Informatique dans les 1ères Journées Doctorales en Technologies de l'Information et de la Communication. Académie Hassan II des Sciences et Techniques à Rabat, le 18 Juillet 2009.

Best paper award SIIE'2011 N° 1, “Content accessibility and Semantic networks processed on foreign natural language analysis”, February 2011, Marrakech – Morocco.

Bibliographie

A

- (Acad. des sciences, 2011) **INSTITUT DE FRANCE – ACADEMIE DES SCIENCES.,** *Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs.* Rapport de l'Académie des sciences, 17 Janvier 2011.
- (Afolabi, 2007) **AFOLABI B. S.,** *La conception et l'adaptation de la structure d'un système d'intelligence économique par l'observation des comportements de l'utilisateur.* Thèse de doctorat, Université Nancy 2, Nancy, 2007.
- (Ahmed et al, 2007) **AHMED M., ANJOMSHOAA A., NGUYEN T. M., MIN T. A.,** *Towards an Ontology-based Organizational Risk Assessment in Collaborative Environments Using the SemanticLIFE.* International Conference on Availability, Reliability and Security, ARES 2007.
- (ANSI, 2004) **AMERICAN NATIONAL STANDARD FOR INFORMATION TECHNOLOGY.,** *Role based access control.* ANSI INCITS 359-2004, February 2004.
- (Atamer et Calori, 1993) **ATAMER T., CALORI R.,** *Diagnostic et décisions stratégiques.* Paris : Dunod Entreprises, 1993.

B

- (Banammar, 1999) **BEN AMMAR A., DOUSSET B.,** *Les métriques et l'analyse relationnelle : Visualisation en quatre dimensions.* 7^{ème} Conférence sur les systèmes d'information élaborée : Bibliométrie - Informatique stratégique - Veille technologique. Ile Rousse, 1999.
- (Beji et El kadhi, 2009) **BEJI S., EL KADHI N.,** *Security ontology proposal for mobile applications.* International Conference on Mobile Data Management: systems, Services and Middleware, pages: 580-587, 2009.
- (Besson et al., 2010) **BESSON B., DESCHAMPS C., ANDRIAMIAMINA H., ARCHANBAUD B., BARET C., BESNARD M., BLANCHET A., BOUAKAZ A., BAPTISTE B., DE LA RUE CAN B.,** *Méthodes d'analyses appliquées à l'intelligence économique,* ICOMTEC Poitiers, 2010.
- (Bichard, 2005) **BICHARD J. P.,** *De la veille stratégique à la sécurité de l'information,* Décision Informatique, N° 625, Mars 2005
- (Blanc, 2010) **BLANC A.,** *Management Stratégique de l'Information – de la veille stratégique à l'intelligence économique,* éditions WEKA, Paris, 2010.
- (Bonnecarrere et al., 2011) **BONNECARRERE T., DUBOIS N., ROLLET F., SOULA O.,** *Intelligence économique et stratégique: le cas des Etats-unis,* Dossier Master 2 IECS, Mars 2011.
- (Botha et Eloff, 2001) **BOTHA R. A., ELOFF J. H. P.,** *Separation of duties for access control enforcement in workflow environments,* IBM Systems Journal, vol 40, n° 3, 2001 (pp 666-682).

- (Bouaka, 2004) **BOUAKA N.**, *Développement d'un modèle pour l'explication d'un problème décisionnel: un outil d'aide à la décision dans un contexte d'Intelligence Economique*. Thèse de doctorat de l'université Nancy 2, 2004.
- (Boughanem et al. 2000) **BOUGHANEM M., CHRISMENT C., MOTHE J., SOULE-DUPUY C., TAMINE L.**, *Connectionist and genetic approaches to perform IR*. F. Crestani and G. Pasi editors. Soft computing in information retrieval: techniques and applications. Physica Verlag, pages 173-198, Heidelberg, 2000.
- (Boukrara, 2010) **BOUKRARA A., CHALAL R.**, *Une approche d'ingénierie des besoins pour la conception des systèmes d'information d'intelligence économique*. Veille stratégique, scientifique et technologique : VSST 2010, Toulouse 25-29 Octobre 2010.
- (Büttcher et al., 2010) **BÜTTCHER S., CLARKE C. L. A., CORMACK G. V.**, *Information Retrieval: Implementing and Evaluating Search Engines*, MIT Press, 2010.
- (Buyens et al., 2009) **BUYENS K., WIN B. D., JOOSEN W.**, *Resolving least privilege violations in software architectures*, In Proceedings of the ICSE Workshop on Software Engineering for Secure Systems (ICSE/SESS'09), May 19, 2009 (pp 9-16).

C

- (Caputo et al., 2009) **CAPUTO D. D., STEPHENS G. D., MALOOF M. A.**, *Detecting Insider Theft of Trade Secrets*, IEEE Security & Privacy, Vol 7, N° 6, November/December 2009
- (Casselmann et Samson, 2011) **CASSELMAN R. M., SAMSON D.**, Internal, *collaborative and competitive knowledge capability*. Proceedings of the 44th Hawaii International Conference on System Sciences – 2011.
- (Chauna et al., 2011a) **CHAUNAC A., CREMIERS F., GAIDOT A., VALLAS H.**, *Le système d'intelligence économique japonais*, Dossier ANAJ-IHEDN, Comité Défense Economique, Section Intelligence Economique, N°1 – Japan – Mars 2011.
- (Chauna et al., 2011b) **CHAUNAC A., DIAKITE Z., NAQUIN P., WISSNER M. H. R.**, *Le système d'intelligence économique américain*, Dossier ANAJ-IHEDN, Comité Défense Economique Section Intelligence Economique, N°2 – Avril 2011.
- (Chen et al., 2008) **CHEN T., HOU K., YU K.**, *Study on model about competitive intelligence system of enterprise based on data mining under electronic commerce environment*. International Seminar on Business and Information Management, IEEE 2008.
- (Chen et al., 2009) **CHEN T., NIU X., YANG W.**, *The application of web data mining*

technique in competitive intelligence system of enterprise based on XML. Third International Symposium on Intelligent Information Technology Application, IEEE 2009.

- (Chounac, 2010) **DE CHAUNAC A.**, *L'ingérence économique chinoise est multiforme*. Site Web Chinoiseries. 2010. Consulté le 12-02-2011, <http://chinoiseries.over-blog.fr/article-aymar-de-chaunac-l-ingerece-economique-chinoise-est-multiforme-60823028.html>
- (Chrisment et al., 2006) **CHRISMENT C., HERNANDEZ N., GENOVA F., MOTHE J.**, *D'un thesaurus vers une ontologie de domaine pour l'exploitation d'un corpus*. AMETIST, INIST, Vol.0, pages 59-92, 2006.
- (Chrisment et al., 2008) **CHRISMENT C., HAEMMERLE O., HERNANDEZ N., MOTHE J.**, *Méthodologie de transformation d'un thesaurus en une ontologie de domaine*. Revue d'Intelligence Artificielle (RIA) 22(1) :7-37, 2008.
- (Cigref, 2004) **CLUB INFORMATIQUE DES GRANDES ENTREPRISES FRANÇAISES.**, *Intelligence économique et stratégique*. Rapport Cigref, 2004.
- (Clusif, 2008) **CLUB DE LA SECURITE DE L'INFORMATION FRANÇAIS.**, *Menaces informatiques et pratiques de sécurité en France*, édition 2008.
- (Colletis, 2007) **COLLETIS G.**, *Intelligence économique : vers un nouveau concept en analyse économique ?* Cahier du GRES, N.3, France, 2007.
- (Cook et Pfleeger, 2010) **COOK I. P., PFLEEGER S. L.**, *Security Decision Support Challenges in Data Collection and Use*, IEEE Security & Privacy, Vol 8, N° 3, May/Juin 2010
- (Coutenceau et al, 2010) **COUTENCEAU C., BARBARA F., EVERETT W., GILLIERON A., JACQUIN X., POUILLAIN M., VALLE C., D'ARVIEA E. V.**, *Guide pratique de l'intelligence économique*, Eyrolles – éditions d'organisation, 2010.
- (Croft et al., 2011) **CROFT, B., METZLER, D, STROHMAN, T.**, *Search Engines: Information Retrieval in Practice*, Addison-Wesley, ISBN-13: 9780136072249, 2011.

D

- (David et Thiery, 2002) **DAVID A., THIERY O.**, *Application of "EQuA2te" Architecture*. Economic Intelligence, 2002.
- (De vasconcelos, 1999) **DE VASCONCELOS C.**, *L'intelligence économique et la stratégie de développement de la PME*, Thèse pour le doctorat en Sciences de. Gestion, Grenoble, 1999.
- (Delbecque, 2006) **DELBEQUE E.**, *L'intelligence économique*. Institut d'étude et de

recherche pour la sécurité des entreprises, 2006.

- (Dkaki et al., 1996) **DKAKI T., MOTHE J., DOUSSET B., CHRISMENT C.,** *Extraction et synthèse de connaissances à partir de bases de données hétérogènes.* INFORSID'96, pp. 287-308, Bordeaux, France, juin 1996.
- (Dkaki et al., 1997) **DKAKI T., DOUSSET B., MOTHE J.,** *Recherche de l'information stratégique dans les bases de données : veille scientifique et technique.* 15ème congrès INFORSID, INFORSID'97, pp 673-690, 1997.
- (Dkaki et al., 1998) **DKAKI T., DOUSSET B., MOTHE J.,** *Analyse d'informations issues du Web avec Tétralogie.* Veille stratégique, scientifique et technologique : VSST'98, pp 159-170, Toulouse, France, 1998.
- (Dkaki et al., 2000) **DKAKI T., DOUSSET B., EGRET D., MOTHE J.,** *Information discovery from semi-structured sources – Application to astronomical literature.* Computer Physics Communication, 2000.
- (Dkaki, 1993) **DKAKI T.,** *Outils informatiques et méthodes automatiques pour la veille technologique.* Thèse de doctorat de l'université Paul Sabatier, Toulouse III, 1993.
- (Dousset et al., 1989) **DOUSSET B., CAMBUS J.P., BIGAUT PH.,** *Réalisation d'une chaîne entièrement automatisée d'étude de la fibrinofomation et de la fibrinolyse.* Journées d'informatique médicale de Toulouse, pp. 104-113, Toulouse, France, 1989.
- (Dousset et al., 1991) **DOUSSET B., KOUSSOUBE S., GENESTAL M.,** *Un système expert d'aide à la conduite et à la surveillance en réanimation.* 11èmes journées internationales sur les systèmes experts et leurs applications, vol 3, pp 151-160, Avignon, France, 1991.
- (Dousset et al., 1993) **DOUSSET B., DKAKI T., LONGEVIALLE C.,** *Qualité de l'information et analyse des données.* 5èmes journées d'études sur les systèmes d'informations élaborées : Bibliométrie - Informatique stratégique - Veille technologique, Ile Rousse Corse France, 1993.
- (Dousset et Benjamaa, 1988) **DOUSSET B., BENJAMAA T.,** *Trilogie logiciel d'analyse de données.* Journées d'études sur les systèmes d'informations élaborées : Bibliométrie - Informatique stratégique Veille technologique. Ile Rousse, 1988.
- (Dousset et Karouach, 2002) **DOUSSET B., KAROUACH S.,** *Collaboration interactive entre classifications et cartes thématiques ou géographiques.* 9èmes rencontres de la société francophone de classification, 2002.
- (Dousset et Karouach, 2005) **DOUSSET B., KAROUACH S.,** *Manipulation de graphes de grande taille pour l'étude des réseaux d'acteurs et des réseaux sémantiques.* 10èmes journées d'études sur les systèmes d'information élaborée : Bibliométrie - Informatique stratégique - Veille technologique, (Ile Rousse Corse France),

2005.

- (Dousset et Karouach, 2007) **DOUSSET B., KAROUACH S.,** *Apports de la classification dans l'analyse des graphes de grande taille.* Veille stratégique, scientifique et technologique : VSST 2007, Marrakech – Maroc, 2007.
- (Dousset, 1987) **DOUSSET B.,** *Utilisation de systèmes experts dans la gestion et la consultation bibliographique.* Publication dans les actes du colloque IA/Santé de Toulouse, pp 167-172, 1987.
- (Dousset, 2003) **DOUSSET B.,** *Intégration de méthodes interactives de découverte de connaissances pour la veille stratégique.* Habilitation à diriger des recherches, Université Paul Sabatier, France, 2003.
- (Dousset, 2006) **DOUSSET B.,** *TETRALOGIE: a platform for scientific and technological survey.* International Workshop on Webometrics, Infometrics and Scientometrics & Seventh COLLNET Meeting, Nancy, (conférencier invité), LORIA, 2006.
- (Dousset, 2009) **DOUSSET B.,** *Extraction de l'information implicite par analyse textuelle de sites Internet en UNICODE,* Veille stratégique, scientifique et technologique : VSST 2009, 2009.
- (Ducieux et al., 2010) **DUCREUX J-M., ABATE R., KACHANER N.,** *Le grand livre de la stratégie,* The Boston Consulting Group, 2010.
- (Dupré, 1997) **DUPRE A., DUHARD N.,** *Les armes secrètes de la décision, la gestion de l'information au service de la performance économique.* Gualino éditeur, Paris, 1997.
- (Dutton et Coll, 1983) **DUTTON J.E., FABEY L., NARAYANAN V.K.,** *Toward understanding Strategic Issue Diagnosis, In Strategic Management Journal, 4, 1, 307-323, 1983.*

E

- (El Bakkali et Hatim, 2009) **EL BAKKALI H., HATIM H.,** *RB-WAC: New approach for access control in workflows,* The 7th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA'09), (pp 637-640), May 10-13, 2009.
- (El Haddadi et al., 2010) **EL HADDADI A., DOUSSET B., BERRADA I., LOUBIER I.,** *Les multi-sources dans un contexte d'intelligence économique.* EGC 2010, Hammamat – Tunisie, 26-29 Janvier 2010
- (El Haddadi et Dousset, 2009) **EL HADDADI A., DOUSSET B.,** *Mise en ligne des analyses stratégiques sur téléphone 3G.* Journées Nationales du GDR en Intelligence Economique – Beaulieu sur Mer - France, 25 et 26 Novembre, 2009.

- (El Haddadi, 2008) **EL HADDADI A.**, *Les apports du text mining et web mining à la veille stratégique*. Mémoire Master de recherche, Université Mohamed V – Souissi – ENSIAS, Rabat, 2008.

F

- (Filliatreau, 2001) **FILLIATREAU G.**, *Les indicateurs bibliométriques en recherche*. Education & Formation – n°59 – avril-juin 2001.
- (Fuhr, et al., 2001) **FUHR N., GROJOHANN K.**, *XIRQL: a query language for information retrieval in XML documents*, SIGIR, W.B. Croft, D.J. Harper, D.H. Kraft, J. Zobel (Eds.), ACM Press, New York, 2001, pp. 172–180.
- (Fuld, 1995) **FULD L.M.**, *The new Competitor Intelligence*. John Wiley, Chichester, 1995.

G

- (Gaidot et al., 2011) **GAIDOT A., DIAKITE Z., CAZABAT M.**, *De la guerre culturelle ou un soft power pas si soft*, Newsletter ANAJ-IHEDN, Comité Défense Economique, Section Intelligence Economique, N°4 – été 2011.
- (Ghallamalah, 2009) **GHALLAMALAH I.**, *Proposition d'un modèle d'analyse exploratoire multidimensionnelle dans un contexte d'intelligence économique*. Thèse de doctorat de l'université Paul Sabatier, Toulouse III, 18 décembre 2009.
- (Ginsburg et al., 2009) **GINSBURG M., KASS A., YAH P. Z.**, *Exploring two enterprise semantic integration systems*, Proceedings of the 42nd Hawaii International Conference on System Sciences – IEEE, 2009.
- (Guilhon, 2003) **GUILHON A.**, *Le processus d'IE et l'identité de la PME, l'IE dans la PME : Visions éparses, paradoxes et manifestations*, Editions Economica, 2003.
- (Guoliang et al. 2009) **GUOLING S., XIAORONG M., XU W.**, *Designing a network acquisition system of competitive intelligence*, Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009.

H

- (Herzog et al., 2007) **HERZOG A., SHAHMEHRI N., DUMA C.**, *An Ontology of Information Security*, International Journal of Information Security and Privacy, Volume 1, Issue 4, 2007.
- (Hinnum et al., 2005) **HINUM K., MIKSCH S., AIGNER W., OHMANN S., POPOW C., POHL M., RESTER M.**, *Gravi : Interactive Information Visualization to Explore Highly Structured Temporal Data*. J. UCS (JUCS), volume 11, pages 1792-1805, 2005.
- (Hole et Netland, **HOLE K. J., NETLAND L. H.**, *Toward Risk Assessment of Large-*

- 2010) *Impact and Rare Events*, IEEE Security & Privacy, Vol 8, N° 3, May/Juin 2010

I

- (Ibekwe-Senjuan, 1997) **IBEKWE-SANJUAN F.**, *Recherche des Tendances Thématiques dans les publications Scientifiques*. Définition d'une Méthodologie Fondée sur la Linguistique, thèse en Sciences : SIC, Université de Grenoble III, 1997.
- (IRES, 2010) **INSTITUT ROYAL DES ETUDES STRATEGIQUE.**, *Veille stratégique au sein de l'IRES*, 2010.

J

- (Jakobiak, 2004a) **JAKOBIAK F.**, *Intelligence économique en pratique : Comment bâtir sons propre système d'intelligence économique*, 2004.
- (Jakobiak, 2004b) **JAKOBIAK F.**, *Intelligence économique en pratique : la comprendre, l'implanter ; l'utiliser*, Les éditions d'Organisation, 2004.
- (Jensen, et al., 2001) **JENSEN M. R., MOLLER T. H., PEDERSEN T. B.**, *Specifying OLAP Cubes On XML Data*, 13th Int. Conf. on Scientific and Statistical Database Management (SSDBM), IEEE Computer Society, pp.101–112, 2001.

K

- (Karouach, 2003) **KAROUACH S.**, *Visualisations interactives pour la découverte de connaissances : concepts, méthodes et outils*. Thèse de Doctorat en informatique, Université Paul Sabatier, France, 2003.
- (Kermarrec et al., 2007) **KERMARREC A. M., FAOU E., MERLET J. P., ROBERT P. SEGOUN L.**, *Que mesurent les indicateurs bibliométrique ?*. Document d'analyse de la commission d'évaluation de l'INRIA, 12 Septembre 2007.
- (Kheith et al., 2005) **KEITH S., KASER O., LEMIRE D.**, *Analyzing Large Collections of Electronic Text Using OLAP*, APICS 29th Conf. in Mathematics, Statistics and Computer Science, Acadia University, pp. 17–26, 2005.
- (Khrouf et al., 2004) **K. KHROUF, C. SOULE-DUPUY**, *A Textual Warehouse Approach: A Web Data Repository*, M. Mohammadian (Ed.), Intelligent Agents for Data Mining and Information Retrieval, Idea Publishing Group, pp. 101–124, 2004.
- (Kim et al., 2005) **KIM A., LUO J., MYONG K.**, *Security Ontology for Annotating Resources*, LNCS 3761, pp. 1483 – 1499. Springer-Verlag Berlin Heidelberg 2005.
- (Kislin, 2007) **KISLIN PH.**, *Modélisation du problème informationnel du veilleur dans la démarche d'intelligence économique*. Thèse de doctorat de l'université Nancy

2, France, 2007.

- (Kleinberg, 1999) **KLEINBERG J. M.**, *Authoritative sources in a hyperlinked environment*. Journal of the ACM, pages 604-632, 1999.

L

- (Laedera et Quinio, 1996) **LARDERA S., QUINIO B.**, *Information et décision stratégique*, accordons les instruments, Paris : Masson, 1996.
- (Lesca et Rouibah., 1997) **LESCA H., ROUIBAH K.**, *Des outils au service de la veille stratégique*, In Systèmes d'information et management, 2, 2, 101-131, 1997.
- (Lesca, 1994) **LESCA H.**, *Veille stratégique pour le management stratégique : Etat de la question et axes de recherche*, Economies et Sociétés, Sciences de Gestion, 20, 5, 31-50, 1994.
- (Lesca, 2004) **LESCA H.**, *Veille stratégique : La méthode L.E.SCAning*, Paris : EMS, 2004.
- (Levet, 2001) **LEVET J. L.**, *IE : mode de pensée, mode d'action*. Economica, Paris, 2001.
- (Li et al, 2007) **LI K., YUCHEN F., XIAOKE Z., QUAN L., ZHIMING C.**, *Study on competitive intelligence system based on web*. Workshop Intelligent Information Technology Application, IEEE, 2007.
- (Li et Dong, 2010) **LI J., DONG Y.**, *Post-controlled vocabulary compiling in competitive intelligence system*. IEEE, 2010.
- (Liu et Xia, 2009) **LIU P., XIA X. H.**, *web mining technology in competitive intelligence system research*. International Symposium on Information Engineering and Electronic Commerce, IEEE, 2009.
- (Loubier et al., 2007) **LOUBIER E., BAHOUN W., DOUSSET B.**, *Visualization and analysis of large graphs*. ACM International Workshop for Ph.D. Students in Information and Knowledge Management (ACM PIKM 2007), ACM, 2007.
- (Loubier, 2008) **LOUBIER E., BAHOUN W., DOUSSET B.**, *VisuGraph : un outil pour la visualisation de données temporelles*. MANifestation des Jeunes Chercheurs STIC (MajecStic 2008), Aline Cauvin, Abbas Chamseddine, Nicolas Faessel, Sébastien Fournier (Eds.), Laboratoire des Sciences de l'Information et des Systèmes (LSIS), 2008.
- (Loubier, 2009) **LOUBIER E.**, *Analyse et visualisation de données relationnelles par morphing de graphe prenant en compte la dimension temporelle*. Thèse de doctorat de l'université Paul Sabatier, Toulouse III, 09 Octobre 2009.

M

- (Magakian et Payaud, 2010). **MAGAKIAN J. L., PAYAUD M. A.,** *100 fiches pour comprendre la stratégie d'entreprise*. Bréal, 2010.
- (Man et al., 2009) **MAN R., DAI Q., NI C., LI X.,** *An author co-citation analysis of information science in China with Chinese Google Scholar search engine, 2004-2009*. Scientometrics, Vol. 81, No. 1 pp. 33-46, 2009.
- (Maniez et Grolier, 1991) **MANIEZ, J., GROLIER, E.,** *A decade of research in classification*. 1991
- (Marcon, 2006) **MARCON C., MOINET N.,** *L'intelligence économique*. Editions Dunod, France, 2006.
- (Martinet et Marti, 1995) **MARTINET B., MARTI Y. M.,** *L'Intelligence Economique : les yeux et les oreilles de l'entreprise*. Organisation Editions, Entreprise-Gestion, 1995.
- (Martinet et Ribault, 1988) **MARTINET B., RIBAUT J.M.,** *La veille technologique, concurrentielle et commerciale*, Les Editions d'Organisation, Paris, 1988.
- (Martre, 1994) **MARTRE H.,** *IE et stratégie des entreprises*. Œuvre Collective du Commissariat au Plan, la documentation Française, France, 1994.
- (McCabe, et al., 2000) **MCCABE C., LEE J., CHOWDHURY A., GROSSMAN D. A., FRIEDER O.,** *On the design and evaluation of a multi-dimensional approach to information retrieval*, 23rd Int. ACM Conf. on research and development in Information Retrieval (SIGIR), ACM Press, pp. 363– 365, 2000.
- (Meliker et al., 2005) **MELIKER J., SLOTNICK M., AVRUSKIN G., KAUFMANN A., JACQUEZ J., NRIAGU J.,** *Improving exposure assessment in environmental epidemiology: Application of spatio-temporal visualization tools*. Journal of Geographical Systems (JGS), volume 7, pages 49-66, 2005).
- (Moinet et Frion, 2009) **MOINET N., FRION P.,** *Knowing is action: from noticing to sense-making*, 3d European Competitive Intelligence Symposium: Competing, Consuming and Collaborating in a Flat World, Stockholm - Suède., 2009.
- (Moinet et Jdey, 2009) **MOINET N., JDEY A.,** *La veille collaborative : le web 2.0 comme levier*, ECIG 2009 : eCommerce et gouvernance de l'internet, Sousse - Tunisie., 2009.
- (Moinet, 2008) **MOINET N.,** *La maîtrise de l'information stratégique dans les organisations au regard du couple agilité/paralysie : la notion de dispositif intelligent, L'information dans les organisations : dynamique et complexité*, in Christiane Volant, Tours, Presses Universitaires François Rabelais, 67-81., in C. Volant, 2008.
- (Mothe et al, 2003) **MOTHE J., CHRISMENT C., DOUSSET B., ALAUX J.,** *DocCube: Multi-Dimensional Visualization and Exploration of Large Document Sets*.

Journal of the American Society for Information Science and Technology, JASIST, Special topic section: web retrieval and mining, Vol. 7 N. 54, p. 650-659, 2003.

- (Mothe et al., 2002) MOTHE J., CHRISMENT C., ALAUX J., *Visualisation globale de collections de documents sous forme d'hyper cube - Le système DocCube*. Journées francophones d'Extraction et de Gestion des Connaissances, EGC2002, Hermès, pp. 131-142, Montpellier, 2002.
- (Mothe, 1994) MOTHE J., *Modèle connexionniste pour la recherche d'informations - Expansion dirigée de requêtes et apprentissage*. Thèse de doctorat de l'université Paul Sabatier, Toulouse III, 1994.
- (Mothe, 2000) MOTHE J., *Recherche et exploration d'informations -Découverte de connaissances pour l'accès à l'information*. Habilitation à diriger des recherches, Université Paul Sabatier, Toulouse, 2000.
- (Mouline, 2010) MOULINE T., *l'intelligence économique au service de la veille stratégique*. Journée IE, Rabat _ Maroc, 2010.

O

- (Olusoji et Amos, 2010) OLUSOJI B. O., AMOS D., *AMTEA : Tool for creating and exploiting annotations in the context of economic intelligence (competitive intelligence)*, IEEE, 2010.
- (Othenin et al., 2011) OTHENIN G. C., CARON C., GUILLEMETTE M. C., *When competitive intelligence meets geospatial intelligence*. Proceedings of the 44th Hawaii International Conference on System Sciences – 2011.

P

- (Paice, 1984) PAICE C., *Soft evaluation of Boolean search queries in information retrieval systems*. Information Technology: Research and Development, pages 33-42, 1984.
- (Park, et al., 2005) PARK B. K., HAN H., SONG I.Y., *XML-OLAP: A Multidimensional Analysis Framework for XML Warehouses*, 7th Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK), LNCS 3589, Springer, pp.32–42, 2005.
- (Pascal. Vidal, 2009) VIDAL P., PETIT V., *Systèmes d'information organisationnels*, Pearson Education France, 2009.
- (Pérez, et al., 2008) PEREZ J. M., LLAVORI R. B., ARAMBURU M. J., PEDERSEN T. B., *Contextualizing data warehouses with documents*, Decision Support Systems, Elsevier, n°45, pp.77–94, 2008.
- (Placet et Rarivoson, 2011) PLACTE J. L., RARIVOSON P., *L'homme au cœur de la stratégie – des ressources humaines aux actifs humains*, Collection Stratégies et

Management, Dunod, 2011.

- (Prescott, 1999) **PRESCOTT J. E.**, *The evolution of competitive intelligence*, Proposal Management, Spring 1999.
- (Pujolle et al., 2008) **PUJOLLE G., RAVAT F., TESTE O., TOURNIER R.**, *Fonctions d'agrégation pour l'analyse en ligne (OLAP) de données textuelles*, Revue des Sciences et Technologies de l'Information, ISI (Ingénierie des Systèmes d'Information), Hermes, Vol.13, N°6, p.61-84, 2008.

R

- (Radum, 2006) **RADUN V.**, *The internal and external communication of intelligence and the competitive intelligence process: the case of the higher educational sector in serbia*, PICMET Proceedings, 9-13 July 2006.
- (Raskin et al., 2001) **RASKIN V., HEMPELMANN C. F., TRIEZENBERG K. E., NIRENBURG S.**, *Ontology in Information Security: A Useful Theoretical Foundation and Methodological Tool*, Proceedings of the workshop on New security paradigms, Cloudcroft, New Mexico, Pages: 53-59, 2001.
- (Ravat, 2007) **RAVAT F.**, *Modèles et outils pour la conception et la manipulation de systèmes d'aide à la décision*. Mémoire d'habilitation à diriger des recherches, université Paul Sabatier, Toulouse III, 2007.
- (Ravat et al., 2008) **RAVAT F., TESTE O., TOURNIER R.**, *Multidimensional Analysis of XML Document Contents With OLAP Dimensions*, Chapitre VI de l'ouvrage "Progressive Methods in Data Warehousing and Business Intelligence: Concepts and Competitive Analytics" de la série "Advances in Data Warehousing and Mining", IGI Publishing, David Tanar, 2008.
- (Ravat et al., 2007a) **RAVAT F., TESTE O., TOURNIER R., ZURFLUH G.**, *Querying Multidimensional Databases*, 11th East-European Conference on Advances in Databases and Information Systems (ADBIS'07), Springer-Verlag, LNCS 4690, Y.E. Ioannidis, B. Novikov, B. Rachev, p.298-313, Varna (Bulgarie), septembre 2007.
- (Ravat et al., 2007b) **RAVAT F., TESTE O., TOURNIER R., ZURFLUH G.**, *A Conceptual Model for Multidimensional Analysis of Documents*, 26th International Conference on Conceptual Modeling (ER'07), Springer-Verlag, LNCS 4801, C. Parent, K.-D. Schewe, V. C. Storey, B. Thalheim, p.550-565, Auckland (New Zealand), novembre 2007.
- (Roberston, 1977) **ROBERSTON S. E.**, *The probability ranking principle in IR. Journal of Documentation*. Pages 294-304, 1977.
- (Romagni et Wild, 1998) **ROMAGNI, P., WILD V.**, *L'intelligence économique au service de l'entreprise, ou l'information comme outil de gestion*. Les Presses du Management, 1998.

- (Rostaing, 1993) ROSTAING H., *Veille Technologique et Bibliométrie : Concepts, Outils, Applications*, Thèse en Sciences : SIC, Université de Marseille III, 58-59, 1993.
- (Rouach, 1999) ROUACH D., *Management du transfert de technologie : L'art de coopérer, innover, veiller*, Gestion, Presses Universitaires de France, Paris, 185 p, 1999.
- (Roy, 2010) ROY P., *Les nouvelles stratégies concurrentielles*. Editions La Découverte, Paris, 2010.

S

- (Sahel, 2011) SAHEL J. A., *Quality Versus quantity: assenssing individual research performance*. Bibliometrics, Vol3, 25 May 2011.
- (Salles et al., 2000) SALLES M., CLERMONT PH., DOUSSET B., *Une méthode de conception de système d'IE*. Colloque IDMME'2000, Canada, 2000.
- (Salles, 2003) SALLES M., *Modélisation des situations de décision dans une méthode d'ingénierie du besoin en I.E*. Conférence IERA, Intelligence Economique : Recherches et Applications, Nancy, France, 2003.
- (Salles, 2005) SALLES M., *De l'analyse du besoin des PME en IE à l'Intelligence Territoriale*. Colloque Européen d'Intelligence Economique, Poitiers Futuroscope, ESCEM Poitiers, p. 414-427, Poitiers, France, 2005.
- (Saltan et McGill, 1984) SALTON G., MCGILL M., *Introduction to modern information retrieval*. McGraw-Hill Int. Book Co, 1984.
- (Salton, 1971) SALTON G., *A comparison between manual and automatic indexing methods*. Journal of American Documentation. Pages 61–71, 1971.
- (Sauter, 2008) SAUTER V. L., *Competitive Intelligence Systems*, International Handbooks on Information Systems, Handbook on Decision Support Systems 2, VII, Pages 195-210, 2008.
- (Sauvagnat, 2005) SAUVAGNAT K., *Modèle flexible pour la recherche d'information dans des corpus de documents semi-structurés*. Thèse de doctorat, Université Paul Sabatier, Toulouse, 2005.
- (SCIE, 2009) SERVICE DE COORDINATION A L'INTELLIGENCE ECONOMIQUE, *Guide des bonnes pratiques en matière d'intelligence économique*, Base de Connaissance AEGE, France, Février 2009. <http://www.bdc.aege.fr>
- (Shabtal, 2010) SHABTAL A., FLEDEL Y., ELOVICI Y., *Securing Android-Powered Mobile Devices Using SELinux*, IEEE Security & Privacy, Vol 8, N° 3, May/Juin 2010

- (Sosson et al., 2001) **SOSSON D., VASSARD M., DOUSSET B.,** *Portail pour la navigation en ligne dans les analyses stratégiques, scientifiques et technologiques. Veille stratégique, scientifique et technologique : VSST'2001, Barcelone - Espagne, 15/10/2001-19/10/2001, Vol. 1, p. 347-358, 2001.*
- (Sparck-Jones, 2000) **SPARCK J. K., WALKER S., ROBERTSON, S. E.,** *A probabilistic model of information retrieval: Development and comparative experiments.* Information Processing & Management, 36(6), 779-840, 2000.
- (Supakkul, 2005) **SUPAKKUL S., CHUNG L.,** *A UML Profile for Goal- Oriented and Use Case-Driven Representation of NFRs and FRs* , Proceedings of the 2005 Third ACIS Int'l Conference on Software Engineering Research, Management and Applications (SERA'05), 2005, IEEE.

T

- (Talbot et al., 2010) **TALBOT E. B., FRINCKE D., BISHOP M.,** *Demythifying Cybersecurity*, IEEE Security & Privacy, Vol 8, N° 3, May/Juin 2010
- (Tan et al., 2004) **TAN A. H., ONG H. L., PAN H., NG J., LI Q. X.,** *Towards personalised web intelligence.* Knowledge and Information Systems, Springer, 2004.
- (Tarondeau, 1998) **TARONDEAU J.C.,** *Le management des Savoirs, Paris : PUF, 1998.*
- (Teste, 2009) **TESTE O.,** *Modélisation et manipulation des systèmes OLAP : de l'intégration des documents à l'utilisateur.* Mémoire d'habilitation à diriger des recherches, université Paul Sabatier, Toulouse III, 2009.
- (Tournier, 2007) **TOURNIER R.,** *Analyse en ligne (OLAP) des documents.* Thèse de doctorat de l'université Paul Sabatier, Toulouse III, Décembre 2007.
- (Tseng et Chou, 2006) **TSENG F. S. C., CHOU A. Y. H.,** *The concept of document warehousing for multi-dimensional modeling of textual-based business intelligence*, J. of Decision Support Systems (DSS), vol.42(2), Elsevier, pp. 727-744, 2006.

W

- (Wang et al., 2003) **WANG H., LI, J., HE Z., GAO H.,** *X aggregation: Flexible Aggregation of XML Data*, 4th Intl. Conf. on Advances in Web-Age Information Management (WAIM), LNCS 2762, Springer, pp. 104-115, 2003.
- (Wang et al., 2005) **WANG H., LI, J., HE Z., GAO H.,** *OLAP for XML Data*, 5th Intl. Conf. on Computer and Information Technology (CIT), IEEE Computer Society, pp. 233-237, 2005.

- (Weidong, 2008) **WEIDONG G.**, *Establishment and application of competitive intelligence system in china*, Second International Symposium on Intelligent Information Technology Application, IEEE, 2008.
- (Wilensky, 1967) **WILENSKY H.L.**, *Organizational intelligence: knowledge and policy in government and industry*. Editions Basic Books, USA, 1967.
- (Wiwatwattana et al., 2007) **WIWATWATTANA N., JAGADISH H.V., LAKSHMANAN L.V.S., SRIVASTAVA D.**, *X3: A Cube Operator for XML OLAP*, 23rd Intl. Conf. on Data Engineering (ICDE), IEEE Computer Society, pp. 916–925, 2007.

X

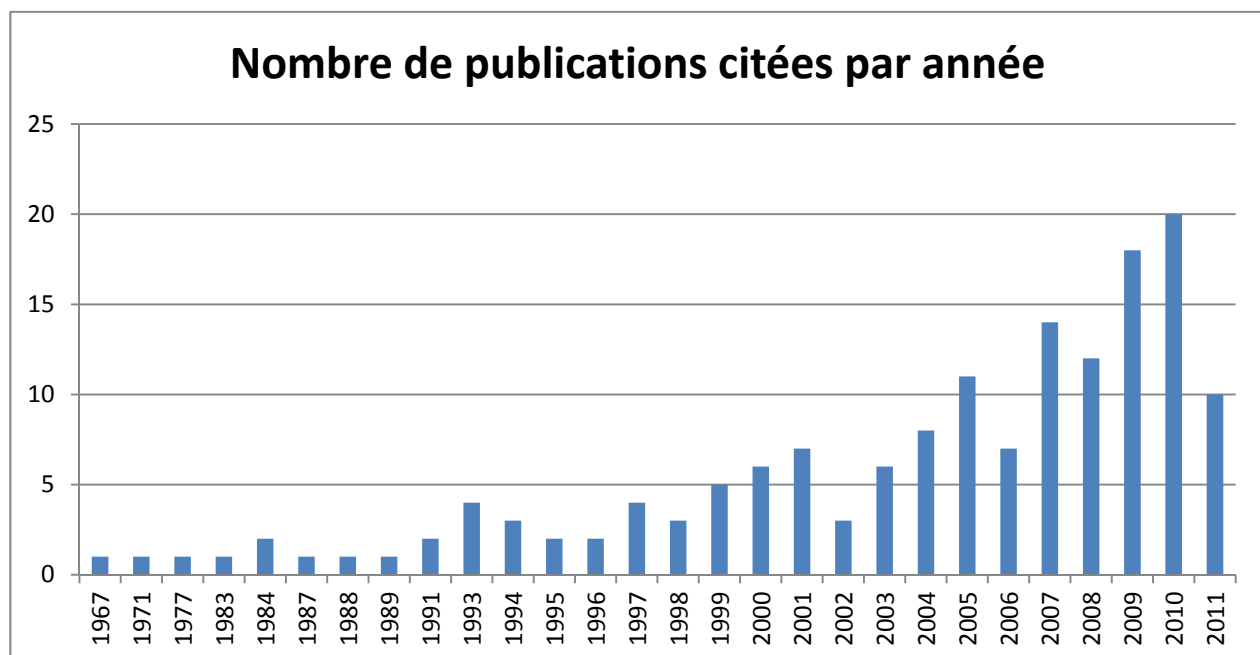
- (Xiaobo et lian, 2010) **XIAOBO T., LIAN L.**, *A integration method of competitive intelligence and knowledge management system for corporate decision-making*, IEEE, 2010.

Y

- (Yvan, 2010) **YVAN M.**, *L'entreprise 2.0 – Comment évaluer son niveau de maturité ?*, Afnor éditions, 2010.

Z

- (Zara, 2008) **ZARA O.**, *Le management de l'intelligence collective – Vers une nouvelle gouvernance*, M21 éditions, 2008.



Glossaire

Agent	Logiciel envoyé sur un réseau pour effectuer une tâche à la place de l'utilisateur et sans son intervention. Un agent est dit intelligent lorsqu'il utilise les techniques de l'intelligence artificielle.
Agent intelligent	Objet utilisant les techniques de l'intelligence artificielle : il adapte son comportement à son environnement et en mémorisant mémorise ses expériences., Il se comporte comme un sous-système capable d'apprentissage. : i Il enrichit le système qui l'utilise en ajoutant, au cours du temps, des fonctions automatiques de traitement, de contrôle, de mémorisation ou de transfert d'information.
Analyse concurrentielle	Elle a pour objectif d'identifier les forces et les faiblesses des différents concurrents sur un projet donné. Elle relève du diagnostic stratégique externe que doit réaliser toute entreprise. [norme afnor AFNOR xp x 50-053]
Analyse de liens	Détermine des relations entre les champs dans une base de données [the KKD process for extracting usefull knowledge from volumes. Of data, u. Fayyad, g. Piatetsky-shapiro, p. Smyth, cacm, vol. 39 num.11, pp 27-34, novembre 1996.]
Analyse des données	En statistique, méthodes développées dans les années 1930 et dont le succès actuel est du dû à l'essor des ordinateurs qui permettent d'automatiser les calculs. "les Les méthodes d'analyse des données permettent une étude globale des individus et des variables en utilisant généralement des représentations graphiques suggestives". [l'analyse des données de j-m. Bouroche et g. Saporta, puf, qsj numéro.1854, p3]
Analyse discriminante décisionnelle	Elle permet de prévoir les modalités du caractère à expliquer à partir des valeurs prises par les caractères explicatifs. [l'analyse des données de j-mbouroche et g. Saporta, puf, qsj numéro.1854, p109].
Analyse en composantes principales (ACP)	"Description des données contenues dans un tableau individus-caractères numériques (...) Ces derniers jouant tous le même rôle". "méthode de réduction du nombre de caractères permettant des représentations géométriques des individus et des caractères" dans un espace de dimension réduite. [l'analyse des données de j-m. Bouroche et g. Saporta, puf, qsj numéro.1854, pp 17 et 21]
Analyse factorielle des correspondances (AFC)	Proposée par JP Benzécri dans les années 60, elle permet l'analyse des données qualitatives (tableaux de contingence, de présence-absence, disjonctifs, d'enquêtes). C'est en fait une ACP sur les profils, dont les interprétations très riches sont particulièrement appréciées dans le traitement des données textuelles.
Analyse factorielle discriminante	Elle permet, à l'aide d'une visualisation dans un espace factoriel de dimension réduite, de décrire les liaisons entre les caractères à expliquer et les caractères explicatifs.

Analyse relationnelle	Elle permet d'analyser les matrices de variance-covariance ou de corrélation (tableaux de burt et condorcet) issues du croisement de deux informations non indépendantes et conduit le plus souvent à une classification qui les partitionne.
Analyse statistique	Elle permet, grâce aux méthodes et aux représentations graphiques de la statistique traditionnelle, de décrire les données textuelles (fréquences, corrélations, dispersion, co-occurrences, etc).
Aspirateur de site	Logiciel qui en transférant les données sur le disque dur de son utilisateur, permet de visualiser des pages html hors connexion. Le paramétrage permet de définir la profondeur de l'aspiration qui peut aller jusqu'au site entier.
Avantages concurrentiels	Sur chaque couple produit-marché, l'entreprise devra rechercher des avantages concurrentiels, en d'autres termes développer certaines caractéristiques de son offre qui lui permettront de rivaliser ou de surpasser ses concurrents. M. Porter (1982) met en avant trois formes d'avantages concurrentiels : la domination par les coûts, la différenciation et la focalisation.
Axe de surveillance	Description ou caractérisation de thèmes d'information. [norme afnor AFNOR xp x 50-053]
Base de données	Ensemble d'informations regroupées sous la forme d'enregistrements stockés sur un système de fichiers logiques (fichiers) ou physiques structurés et organisés de manière à pouvoir être facilement manipulés.
Blogs	Un blog (graphie alternative : blogue1) ou un cybercarnet2 est un type de site Web, ou une partie d'un site Web. Comme son étymologie l'indique (web log signifie journal de bord sur le web en anglais), un blog est censé contenir régulièrement de nouveaux billets, c'est-à-dire des notes ou des articles agglomérés au fil du temps sur un sujet donné.
CAH	<p>Les méthodes hiérarchiques produisent des partitions en classes imbriquées de plus en plus grandes, le nombre de classes n'est pas connu a priori, plusieurs partitions imbriquées peuvent être proposées. La classification ascendante hiérarchique (C.A.H.) consiste à regrouper les individus en classes en fonction de deux critères:</p> <ul style="list-style-type: none"> - les individus d'une même classe sont le plus semblables possible, - les classes sont les plus disjointes possibles. <p>Pour cela, nous avons besoin d'une mesure globale de la proximité des individus à l'intérieur de chaque classe et de la distance interclasse pour apprécier la qualité de la partition obtenue. Comme il n'est pas envisageable d'évaluer toutes les partitions pour ne garder que la meilleure au sens du critère choisi (problème trop fortement combinatoire), il est donc exclu de trouver cette meilleure partition</p>

Capital intellectuel	Ensemble des actifs, biens ou valeurs, incorporels, comme le savoir-faire technologique, la fidélité du client, l'originalité et la fiabilité des processus opérationnels, etc., sur lesquels reposent de plus en plus la survie et le développement d'une entreprise. [office québécois de la langue française – 2003]
Capteur	Élément humain ou technique d'un système de veille chargé de collecter et transmettre des signaux correspondant généralement à un axe de surveillance identifié. [norme AFNOR afnor xp x 50-053]
Cartographie (ou Mapping)	Représentation visuelle des résultats de recherche ou des rapports de veille. Ils se présentent généralement sous forme de cartes qui indiquent les relations directes et indirectes identifiées au cours du cycle de veille.
CdCD-ROM	Compact disc - read only memory
Champ de veille	Ensemble homogène des données faisant l'objet d'une veille ; ex : dans une veille concurrentielle : les sociétés concurrentes directes ou les produits de substitution.
Co-citation	En scientométrie, méthode qui permet d'identifier "les publications auxquelles se réfèrent le plus les scientifiques". "le Le calcul des co-citations va lisser les citations en éliminant celles qui sont isolées". [j-p. Courtial, introduction à la scientométrie, anthropos, parisParis, 1990, p 52].
Co-occurrence	Désigne la présence simultanée dans certains documents d'un corpus homogène d'éléments d'information que l'on veut mettre en relation.
Corpus	Ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique. En lexicométrie, il s'agit de l'ensemble de textes réunis à des fins de comparaison, servant de base à une étude quantitative.
Corrélation	Coefficient de corrélation linéaire : "mesure l'intensité de la liaison entre 2 caractères quantitatifs".
Co-signature	Rédaction et publication en commun d'un article ou d'un document de travail... Par plusieurs auteurs. L'analyse statique des co-signatures permet de déterminer les différentes collaboratoires collaborations d'un domaine, l'analyse dynamique conduit à l'étude de l'évolution des équipes et de leurs collaborations.
Crowdsourcing	Le crowdsourcing (en français, externalisation ouverte) est un des domaines émergents du management de la connaissance : c'est le fait d'utiliser la créativité, l'intelligence et le savoir-faire d'un grand nombre de personnes (des internautes en général), en sous-traitance, pour réaliser certaines tâches traditionnellement effectuées par un employé ou un entrepreneur. Ceci se fait par un appel ciblé (quand un niveau minimal d'expertise est nécessaire) ou par un appel ouvert à d'autres acteurs. Le travail est éventuellement rémunéré. Il peut s'agir de simplement

externaliser des tâches ne relevant pas du métier fondamental de l'entreprise, ou de démarches plus innovantes.

Data mining	Ensemble des technologies avancées susceptibles d'analyser l'information d'un datawarehouse pour en tirer des tendances, pour segmenter l'information ou pour trouver des corrélations dans les données. [j. M. Franco et eds-institut prométhéus – 1997]
Datawarehouse	Ensemble des données collectées parmi les données opérationnelles de l'entreprise, éventuellement complétées par des données externes, organisé organiséespour être mis à la disposition de ses décideurs. [association des informaticiens de langue française]
Descripteur	Mot ou locution contribuant à caractériser l'information contenue dans un document et à en faciliter la recherche. [commission ministérielle de terminologie informatique]
Diagramme de classes	Le diagramme de classes est un schéma utilisé en génie logiciel pour présenter les classes et les interfaces des systèmes ainsi que les différentes relations entre celles-ci. Ce diagramme fait partie de la partie statique d'UML car il fait abstraction des aspects temporels et dynamiques.
Diagramme de déploiement	un Un diagramme de déploiement est une vue statique qui sert à représenter l'utilisation de l'infrastructure physique par le système et la manière dont les composants du système sont répartis ainsi que leurs relations entre eux
diagramme Diagramme de contexte	C'est uUn diagramme définissant le plus haut niveau des interfaces d'un système. Ils définissent les composants de l'environnement d'un système tels qu'ils sont perçus du point de vue du système, en les identifiant comme une série de terminaisons rattachées aux interfaces du système. Il n'y a qu'un seul diagramme contextuel par système, bien qu'il puisse s'étendre sur plusieurs pages s'il est très complexe.
Diagrammes de cas d'utilisation	Les diagrammes de cas d'utilisation sont des diagrammes UML utilisés pour donner une vision globale du comportement fonctionnel d'un système logiciel. Un cas d'utilisation représente une unité discrète d'interaction entre un utilisateur (humain ou machine) et un système. Il est une unité significative de travail. Dans un diagramme de cas d'utilisation, les utilisateurs sont appelés acteurs (actors), ils interagissent avec les cas d'utilisation (use cases).
Diffusion sélective de l'information	fFourniture systématique et personnalisée d'informations à des usagers caractérisés par leur profil d'intérêt (profil documentaire). [AFNOR afnor - vocabulaire de la documentation]
Donnée	Représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement.

Entreprise 2.0	Entreprise 2.0 est un terme dérivé de du Web 2.0 pour décrire combien un logiciel peut être social, et est employé actuellement pour transformer l'intranet d'entreprise en une plateforme plus organisationnelle, collaborative et gérée par l'utilisateur. Le terme fut inventé par Andrew McAfee de la Harvard Business School dans le MIT Sloan Management Review du printemps 2006.
Environnement	L'environnement de l'entreprise est constitué par l'ensemble des éléments qui sont en relation avec elle sans pour autant être directement impliqués dans son fonctionnement interne. L'environnement de l'entreprise comprend aussi bien des acteurs que des structures économiques et sociales. Il comprend en particulier des acteurs économiques, des structures géographiques et démographiques, des structures sociales et économiques, un cadre juridique, un environnement technologique, des comportements et des cultures. Tous ces facteurs évoluent, changent d'orientation, parfois très très rapidement. L'environnement actuel des entreprises est complexe et incertain.
Flux RSS	<p>Un flux RSS est un fichier dont le contenu est produit automatiquement (sauf cas exceptionnels) en fonction des mises à jour d'un site Web. Les flux RSS sont souvent utilisés par les sites d'actualité ou les blogs pour présenter les titres des dernières informations consultables en ligne.</p> <p>Le terme RSS signifie que le contenu du fichier est informatiquement codé selon le standard RSS, qui s'appuie lui-même sur le langage informatique XML.</p>
Forum	Service permettant discussions et échanges sur un thème donné. cChaque utilisateur peut lire à tout moment les interventions de tous les autres et apporter sa propre contribution sous forme d'articles. [délégation générale à la langue française]
Fréquence (d'une unité textuelle)	Le nombre de ses occurrences dans le corpus.
Génie logiciel	Ensemble des méthodes, des techniques et des outils concourant à la production d'un logiciel, au-delà de la seule activité de programmation.
Gestion de l'information	Dans un système de traitement de l'information, ensemble des fonctions permettant de gérer l'acquisition, l'analyse, la mémorisation, la recherche et la diffusion de l'information. [nf z 61-001 - technologies de l'information]
Gestion des connaissances	Gestion, par des moyens informatiques, des informations significatives qui sont acquises par une entreprise et qui y circulent, ainsi que du savoir-faire développé par le personnel, de manière à créer un système interactif de formation maison continue qui débouche sur une meilleure qualité des produits et services, ainsi que sur une plus grande compétitivité de l'entreprise. [office de la langue française – 2001]

Gestion stratégique	Démarche globale qui inclut une anticipation de l'évolution des marchés et de la concurrence et qui, dans les activités de planification, d'organisation, d'impulsion et de contrôle menées par la direction générale, consiste à définir les orientations stratégiques de l'entreprise, à déterminer des objectifs explicites et à mettre en œuvre les actions et les moyens appropriés pour y faire face, grâce à l'amélioration et à l'utilisation des compétences de l'entreprise. [office de la langue française, 1998]
Gestion stratégique de l'information	Gestion de l'information dans laquelle l'information est considérée comme l'élément essentiel de la stratégie d'une entreprise qui désire optimiser son fonctionnement, assurer sa prospérité et occuper une place de choix dans son domaine, par rapport à ses concurrents. [office de la langue française, 1999]
Groupware	Un groupware (ou collecticiel) est un type de logiciel qui permet à un groupe de personnes de partager des documents à distance pour favoriser le travail collaboratif.
Html	Hypertext markup language est un langage de description de page avec lequel sont conçues les pages du web. Il sert à décrire la présentation de la page, et surtout les liens avec d'autres documents (textes, pages, etc.) Situés situés sur le réseau. [délégation générale à la langue française]
Indexation	Processus destiné à représenter par les éléments d'un langage documentaire ou naturel des données résultant de l'analyse du contenu d'un document ou d'une question. On désigne également ainsi le résultat de cette opération. [AFNOR afnor - vocabulaire de la documentation]
Indicateur	Variable ayant pour objet de mesurer ou apprécier un état, une évolution économique. [le petit robert – 1993]
Infométrie	Désigne l'ensemble des activités métriques relatives à l'information et au secteur de la documentation. [terme]Terme aussi utilisé pour indiquer l'extension du champ d'application de ces méthodes et techniques à un domaine beaucoup plus vaste de recherches, d'activités, à un domaine plus vaste de matériaux, d'indices. [j-m.noyer, les sciences de l'information, p10].
Information	Signification que l'homme attribue à des données à l'aide des conventions employées pour les représenter. [AFNOR afnor - vocabulaire de la documentation]
Information formelle	Donnée qui a été rédigée ou diffusée sur un support (papier, multi média, son, image...). [norme AFNOR afnor xp x 50-053]
Information informelle	Donnée recueillie auprès d'une source orale ou n'ayant pas été explicitement mise en forme pour publication. [norme AFNOR afnor xp x 50-053]

Information stratégique	Information contenant des éléments susceptibles de contribuer à la définition, l'infléchissement ou la remise en cause de la stratégie de l'organisation. [norme AFNOR afnor xp x 50-053]
Information utile	Information directement exploitable par son destinataire. [norme AFNOR afnor xp x 50-053]
Innovation	Terme désignant à la fois le processus de création par l'entreprise d'une offre considérée comme nouvelle et le résultat de ce processus : un nouveau produit, un nouveau service ou un nouveau procédé de fabrication.
Intelligence collective	L'intelligence collective désigne les capacités cognitives d'une communauté résultant des interactions multiples entre ses membres (ou agents)
Interactivité	L'interactivité est une activité nécessitant la coopération de plusieurs êtres ou systèmes, naturels ou artificiels qui agissent en ajustant leur comportement. L'interactivité est souvent associée aux technologies permettant des échanges homme-machine (voir interface homme-machine). Toutefois elle est présente dans toutes les formes de communication et d'échange où la conduite et le déroulement de la situation sont liées à des processus de rétroaction, de collaboration, de coopération entre les acteurs qui produisent ainsi un contenu, réalisent un objectif, ou plus simplement modifient et adaptent leur comportement.
Interface	Jonction entre deux matériels ou logiciels leur permettant d'échanger des informations par l'adoption de règles communes, physiques ou logiques.
Internet	Réseau mondial associant des ressources de télécommunication et des ordinateurs serveurs et clients, destiné à l'échange de messages électroniques, d'informations multimédias et de fichiers. Il fonctionne en utilisant un protocole commun qui permet l'acheminement de proche en proche de messages découpés en paquets indépendants. L'acheminement est fondé sur le protocole ip (internet protocol), spécifié par l'internet society (isoc). L'accès au réseau est ouvert à tout utilisateur ayant obtenu une adresse auprès d'un organisme accrédité. La gestion est décentralisée en réseaux interconnectés. [délégation générale à la langue française]
Intranet	Réseau de télécommunication et de téléinformatique destiné à l'usage exclusif d'un organisme et utilisant les mêmes protocoles et techniques que l'internet. [délégation générale à la langue française]
JDBC	JDBC (Java DataBase Connectivity) est une interface de programmation créée par Sun Microsystems -depuis racheté par Oracle Corporation-, pour les programmes utilisant la plateforme Java. Il permet aux applications Java d'accéder par le biais d'une interface commune à des sources de données pour lesquelles il existe des pilotes JDBC. Normalement, il s'agit d'une base de données relationnelle, et des pilotes

JDBC sont disponibles pour tous les systèmes connus de bases de données relationnelles.

Knowledge Management	Management La gestion des connaissances Système est mise en place dans l'entreprise pour identifier, collecter, traiter, stocker, diffuser les informations formelles détenues par les membres de l'entreprise. Elle sert ; également à transformer de l'information informelle en information formelle.
Liste de diffusion	Dans les groupes de discussion fonctionnant par courrier électronique, système permettant de diffuser un message, un article, une question, etc. A à un grand nombre de personnes préalablement inscrites sur une liste. [délégation générale à la langue française]
Lobbying	Terme anglais qui provient de lobby (groupe de pression) et qui fait appel à la pratique d'une pression exercée sur les parlementaires et par extension aujourd'hui sur les différents acteurs du système décisionnel pouvant agir sur l'environnement d'une organisation. Ces pressions peuvent prendre des formes multiples (information, argumentation, conseil, appel à l'opinion publique, diffusion de livre blanc...). [norme AFNOR afnor xp x 50-053]
Mashup	Un mashup en informatique est un site Web ou une application dont le contenu provient de la combinaison de plusieurs sources d'information. Le document composite produit est aussi appelé compendium.
Mémoire d'entreprise	Somme des informations, des connaissances, des idées, du savoir et du savoir-faire accumulés dans une entreprise au fil du temps, qui est recueillie par des moyens informatiques ou électroniques et qui est conservée dans des bases de données ou sur des supports électroniques destinés à cet effet. [office de la langue française]
Menace	Circonstance inopportune (irruption des concurrents sur le marché, évolution des goûts des consommateurs, nouvelle loi, etc.) A oà prendre en compte dans la définition d'une stratégie.
Meta-moteur	Un méta-moteur lance la recherche sur plusieurs moteurs et annuaires de sites et affiche les résultats fournis par les instruments de recherche qu'il a utilisés. L'interrogateur formule sa recherche à l'aide de mots sans savoir, dans le cas où des moteurs et des annuaires seraient simultanément sollicités, s'ils sont contenus dans les fichiers ou seulement dans leur description établie dans un annuaire. Le résultat consiste en listes de pages et/ou de sites. [association des bibliothécairesbibliothécaires français]
Méta-données	Une méta-donnée (mot composé du préfixe grec meta, indiquant l'auto-référence ; le mot signifie donc proprement « donnée de/à propos de donnée ») est une donnée servant à définir ou décrire une autre donnée quel que soit son support (papier ou électronique).

Les métadonnées sont à la base des techniques du Web sémantique. Elles

sont définies dans le cadre du modèle Resource Description Framework (RDF).

Micro-blogging	Le micro-blog (parfois appelé miniblogue) est un dérivé concis du blog, qui permet de publier un court article, plus court que dans les blogs classiques, les articles pouvant être de type texte court, mais peuvent également contenir une image ou même une vidéo embarquée, Les flux d'agrégation sont plus légers que dans les blogs traditionnels et peuvent contenir tout le message. La diffusion peut également être restreinte par l'éditeur à un cercle de personnes désirées.
Mot-clé	(keyword) Descripteur extrait du texte qu'il caractérise ou d'un thésaurus. [commission ministérielle de terminologie informatique]
Moteur de recherche	(Search Engine) Système d'exploitation de banque de données et, par extension, serveur spécialisé permettant d'accéder accéder sur la toile à des ressources (pages, sites, ...) à partir de mots mots-clés. [délégation générale à la langue française]
Notice bibliographique	Ensemble de données organisées sous forme de champs et qui décrivent un document (champ auteur / champ titre / champ année et lieu de publication / champ mot-clé...)
Objectif stratégique	Objectif dont la réalisation s'étend sur plusieurs années. [meta, journal des traducteurs – 1966]
Observatoire	Centre de regroupement et d'analyse de données qualitatives et quantitatives relatives à une problématiqueproblématique particulière pour en suivre les évolutions. [norme AFNOR afnor xp x 50-053]
Occurrence	Apparition d'un terme dans un corpus.
Plan de recherche	La synthèse de l'ensemble des tâches à réaliser pour recueillir l'information et l'expertise manquante. Bien sûr, le plan de recherche prévoit une répartition des rôles pour renforcer la collaboration entre des personnes ayant des préoccupations différentes (bureau d'étude, production, finance, commercial, etc.). [arist rhône-alpes]
Planification stratégique	Constitue la phase qualitative qui n'est pas emprisonnée dans un processus rigide, c'est la planification du changement. Cette phase n'a pas pour objet la rédaction d'un plan, son but est d'effectuer le choix des options stratégiques et de s'assurer de la cohérence des options retenues.
Position concurrentielle	Place détenue par une entreprise sur le marché en tenant compte de ses concurrents.
Programmation par objets	Mode de programmation dans lequel les données et les procédures qui les manipulent sont regroupées en entités appelées objets.
Propriété industrielle	Droits relatifs aux inventions, aux dessins ou aux modèles industriels, aux marques de fabrique ou de commerce, aux marques de service, aux noms

commerciaux, aux indications de provenance ou aux appellations d'origine, ainsi que certaines règles visant à réprimer la concurrence déloyale. [norme AFNOR afnor z 40-001 – 1979]

Prospective	Evaluation des futurs possibles, inscrits dans une vision globale, qualitative, volontariste et multiple (les scénarios). Elle est plus un art qu'une technique, et appelle plus à une attitude qu'un savoir universel.
Pull	Technologie fondée sur l'architecture client-serveur, qui est utilisée ordinairement dans le web lorsque l'internaute recherche de l'information et qui exige de lui une démarche active par l'intermédiaire de son navigateur qui lui présentera ensuite le résultat de cette recherche. [office de la langue française – 2000]
Push	Technologie basée sur la webdiffusion (webcasting) qui n'exige pas de démarche active de la part de l'internaute, puisqu'il lui suffit de s'abonner à une ou plusieurs chaînes thématiques pour recevoir l'information désirée. [office de la langue française – 2000]
Question fermée	Question dont les seules réponses possibles sont proposées explicitement à la personne interrogée.
Question ouverte	Question posée sans grille de réponse pré-établie, dont la réponse peut être numérique ou textuelle.
Renseignement	Désigne des connaissances de tous tout ordres sur un adversaire potentiel, utiles aux pouvoirs publics, au commandement militaire. (...). il s'agit d'une somme d'informations élaborées, vérifiées et synthétisées, destinées à un groupe restreint d'individus. Le renseignement a pour finalité de protéger les valeurs morales publiques, de canaliser les tendances spontanées de chacun à poursuivre des intérêts particuliers, de permettre au pouvoir politique d'exercer, de façon responsable, ses prérogatives.
Reporting	Un compte rendu (cf. J.O. du 26 octobre 2006) souvent appelé en anglais reporting est l'opération consistant, pour une entreprise, à faire rapport de son activité. C'est la présentation périodique de rapports et bilans analytiques sur les activités et résultats d'une organisation, d'une unité de travail ou du responsable d'une fonction, destinée à en informer ceux chargés de les superviser en interne ou en externe, ou tout simplement concernés par ces activités ou résultats.
Requête	Expression formalisée d'une demande.
Réseau sémantique	Un réseau sémantique est une structure de graphe dont la fonction est l'encodage des connaissances taxonomiques concernant des objets ainsi que leurs propriétés. Les cartes heuristiques ou les arbres heuristiques sont un des moyens de représenter des réseaux sémantiques
Réseaux	Ensemble fluctuant et multiforme d'entités individuelles ou collectives

impliquées dans un objectif commun et mises en relation les unes avec les autres. Le réseau associe des entités à la fois autonomes et liées, possède ses propres règles de fonctionnement et se transforme en fonction des objectifs poursuivis.

réseaux Réseau sociaux

Un réseau social est un ensemble d'identités sociales telles que des individus ou encore des organisations reliées entre elles par des liens créés lors des d'interactions sociales. Il se représente par une structure ou une forme dynamique d'un groupement social.

Robot de collecte d'information

Ils permettent, en tâche de fond, de récupérer toutes les pages web ayant répondu à un robot de recherche interactive, ils sont le plus souvent programmés dans des langages disponibles en free ware sur le réseau.

Robot de recherche

Sur internet, ils permettent de récupérer, en mode interactif, les pages web qui correspondent à une équation de recherche définie par l'utilisateur ou à une thématique donnée déjà indexée. Les plus connus sont: alta Alta vistaVista, yahooYahoo, magellanMagellan, lycosLycos, ...

Scientométrie

Discipline qui se rattache au courant économétrique de mesure de l'activité scientifique et de son évaluation [d'après j-p. Courtial, introduction à la scientométrie, de la bibliométrie à la veille technologique, anthropos, 1990, p 7].

Sécurité du système de traitement de données

Ensemble de mesures techniques et administratives appliquées à un système de traitement de données pour protéger le matériel, le logiciel et les données contre toute modification, destruction ou divulgation par accident ou malveillance. [iso – 1986]

Segment stratégique

Ensemble de produits ou services se caractérisant par une même combinaison de facteurs clés de succès.

Serveur

Système informatique destiné à fournir des services à des utilisateurs connectés et, par extension, organisme qui exploite un tel système. Un serveur peut par exemple permettre la consultation et l'exploitation directe de banques de données. [délégation générale à la langue française]

SGBD

Un système de gestion de base de données (abr. SGBD, en anglais DBMS) est un ensemble de logiciels informatiques qui sert à la manipulation des bases de données. Il sert à effectuer des opérations ordinaires telles que consulter, modifier, construire, organiser, transformer, copier, sauvegarder ou restaurer des bases de données. Il est souvent utilisé par d'autres logiciels ainsi que les administrateurs ou les développeurs.

Signal

Élément d'information indiquant aux veilleurs une évolution de tendances dans le champ de veille. Le signal peut être qualifié de faible (germe annonciateur) ou fort (confirmation d'une tendance repérée - synthèse des positions et menaces d'un acteur)

Signal critique	Information critique qui génère le déclenchement d'une analyse stratégique
Signal faible	Information qui annonce le changement d'une tendance et doit déclencher un travail de recherche d'informations complémentaires pour la confirmer et permettre le déclenchement d'une analyse stratégique. [norme AFNOR afnor xp x 50-053]
Signal fort	Information qui confirme une tendance actuelle et qui permet de bâtir une stratégie de réponse adéquate. [norme AFNOR afnor xp x 50-053]
Sociaux Social bookmarking	Le social bookmarking (en français « marque-page social », « navigation sociale » ou « partage de signets ») est une façon pour les internautes de stocker, de classer, de chercher et de partager leurs liens favoris.
Source	Emetteur de l'information diffusée qui doit être mentionnée lorsqu'elle est sortie de son contexte.
Sourcing	Phase du cycle de veille de recherche et de sélection des sources. Elles doivent être surveillées et mises à jour régulièrement.
Sql	Language standardisé développé par ibm et qui a pour fonction la définition, la manipulation et le contrôle des données d'une base de données relationnelle.
Start up	La Une startup est une entreprise en construction qui ne s'est pas encore lancée sur le marché commercial (ou seulement à titre expérimental). Elle est en phase plus ou moins longue de développement d'un produit, d'une idée, d'une étude de marché, etc., et de recherche de partenaires professionnels, techniques ou d'échec décuplés par rapport à des entreprises traditionnelles.
Stratégie	<p>Le mot stratégie est dérivé du grec stratos qui signifie « armée » et ageîn qui signifie « conduire », et par suite de l'Italien strategia. Ce terme recouvre plusieurs significations qui ne sont pas strictement équivalentes du fait de leur emploi dans des domaines distincts :</p> <p>Initialement, il s'agit de l'art de coordonner l'action de l'ensemble des forces de la Nation - politiques, militaires, économiques, financières, morales... - pour conduire une guerre, gérer une crise ou préserver la paix. Selon Charles de Gaulle : «La stratégie est de la compétence du gouvernement et de celle du haut-commandement des forces armées.»</p> <p>Par suite et par extension, c'est l'élaboration d'une politique, définie en fonction de ses forces et de ses faiblesses, compte tenu des menaces et des opportunités, dans d'autres domaines que celui de la défense, notamment dans les activités économiques (stratégie d'entreprise, stratégie commerciale, industrielle, financière, etc.) mais aussi dans des jeux complexes avec par exemple la stratégie échiquienne. Dans la didactique on parle aussi des stratégies d'apprentissage et de stratégies de</p>

communication.

Surveillance	<p>La surveillance est la fonction d'observer les activités de personnes ou de groupes d'individus.</p> <p>Dans le domaine purement technique, on parle aussi de supervision ou de monitoring.</p>
Système d'information	Ensemble des moyens (organisations, acteurs, procédures, systèmes informatiques) nécessaires au traitement et à l'exploitation des informations dans le cadre d'objectifs définis au niveau de la stratégie de l'établissement, des métiers, de la réglementation. [cnrs]
Système d'information stratégique	Système structuré qui permet de recueillir l'information sur l'environnement de l'entreprise susceptible d'impacter de façon favorable ou défavorable l'avenir et qui doit fournir l'information nécessaire pour bâtir la stratégie. [norme AFNOR afnor xp x 50-053]
Système de veille	Ensemble structuré réunissant les compétences répondant à des besoins de veille. [norme AFNOR afnor xp x 50-053]
TCP	Transmission Control Protocol (littéralement, « protocole de contrôle de transmissions ») abrégé TCP, est un protocole de transport fiable, en mode connecté, documenté dans la RFC 793 de l' IETF.
Tétralogie	Logiciel de VT développé à l'IRIT (Toulouse) et disponible sur le réseau. Il est basé sur l'analyse exploratoire des données et les méthodes de classification automatique. Il permet d'extraire de bases bibliographiques ou de données issues d'internet les informations endogènes qu'elles contiennent. Un tableur 3d 3D spécifique, des cartes factorielles en 4d 4D et des arbres de classification permettent de naviguer de façon interactive dans les analyses.
Text mining	<p>La fouille de textes ou l'extraction de connaissances dans les textes est une spécialisation de la fouille de données et fait partie du domaine de l'intelligence artificielle. Cette technique est souvent désignée sous l'anglicisme text mining.</p> <p>C'est un ensemble de traitements informatiques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité dans des textes produits par des humains pour des humains. Dans la pratique, cela revient à mettre en pratique un modèle simplifié des théories linguistiques dans des systèmes informatiques d'apprentissage et de statistiques.</p>
Traitement automatique des données	Ensemble des opérations réalisées par des moyens automatiques, relatifs à la collecte, l'enregistrement, l'élaboration, la modification, la conservation, la destruction, l'édition de données et, d'une façon générale, leur exploitation.
Traitement du langage	Discipline étudiant les techniques permettant à une machine d'analyser et

naturel	de comprendre le langage humain.
UML	UML (en anglais Unified Modeling Language ou « langage de modélisation unifié ») est un langage de modélisation graphique à base de pictogrammes. Il est apparu dans le monde du génie logiciel, dans le cadre de la « conception orientée objet ».
Veille	L'observation de l'environnement suivi de la diffusion bien ciblée des informations analysées, sélectionnées et traitées utiles à la prise de décision stratégique. [f. Jakobiak]
Veille active	Surveillance de l'environnement focalisé sur ce que l'on suppose a priori important et stratégique pour l'entreprise. [norme AFNOR afnor xp x 50-053]
Veille commerciale	Elle porte essentiellement sur les clients et les fournisseurs de l'entreprise, et également sur les sous-traitants et les partenaires de celle-ci. Elle permet d'identifier les capacités ou les difficultés des fournisseurs et sous-traitants, d'être à l'écoute des attentes et des besoins des clients, [cdies]
Veille concurrentielle	Elle cherche à identifier la concurrence actuelle ou à venir, d'une manière économique-financière (chiffre d'affaires, effectifs, investissements, ...). C'est par cette veille que l'entreprise peut s'organiser par rapport à ses forces et ses faiblesses. L'objectif est d'identifier clairement la stratégie des concurrents pour anticiper les actions de ceux-ci et agir de manière opérationnelle sur l'avenir de l'entreprise sur son marché. [cdies]
Veille environnementale	Cadre sociopolitique dont doit tenir compte l'entreprise (dont veille réglementaire, veille sociétale, veille des groupes de pression, ...)
Veille juridico-réglementaire	Elle doit permettre à l'entreprise de prendre connaissance des évolutions réglementaires au niveau national et européen, en ce qui concerne les normes, les lois (propositions, jurisprudences, décrets, ...) et réglementations (fiscalité, ...). Concrètement, la veille juridique permet à l'entreprise d'anticiper les changements liés à l'adoption d'un texte de loi et de concourir légalement sur un marché étranger. [cdies]
Veille passive	Ecoute de tous les signaux de l'environnement sans axe précis. [norme AFNOR afnor xp x 50-053]
Veille sectorielle	Offre, demande, conjoncture etc. sur un secteur d'activité
Veille sociale	Monde du travail : syndicats, législation sociale, état des relations humaines dans l'entreprise.
Veille sociétale	Cette veille est liée aux aspects socio-économiques, politiques, géopolitiques et socioculturels de la société. Elle permet de surveiller l'évolution des mentalités et des comportements des consommateurs, les risques liés à l'environnement, les mouvements sociaux, [cdies]
Veille stratégique	Regroupement de l'ensemble des activités de veille, consistant en un

processus anticipatif d'observation et d'analyse de l'environnement, suivi de la diffusion ciblée des informations utiles à la prise de décisions. La veille stratégique concerne les décisions qui engagent le devenir, l'évolution de l'entreprise ou de la collectivité en relation avec les changements de son environnement socio-économique.

Veille technologique

eElle a pour objectif de devancer les changements technologiques, d'appréhender les innovations techniques et de promouvoir la politique de recherche et de développement de l'entreprise. Elle cherche à identifier les produits, les technologies et les procédés de fabrication mis en œuvre par les clients, les fournisseurs et les concurrents de l'entreprise, afin de suivre les évolutions du marché, de se réorienter en fonction des changements technologiques, et de mettre en place des procédures de protection industrielle. [cdies]

Veille territoriale

eElle est une évolution de la culture locale, basée sur la collecte et le traitement des signaux destinés aux donneurs d'ordre pour lui fournir au moment opportun, l'information judicieuse. [cdies]

Web 2.0

Le Web 2.0 est une évolution du Web vers plus de simplicité (ne nécessitant pas de connaissances techniques) et d'interactivité (permettant à chacun de contribuer sous différentes formes). L'expression « Web 2.0 » désigne l'ensemble des techniques, des fonctionnalités et des usages du World Wide Web qui ont suivi la forme initiale du web en particulier les interfaces permettant aux internautes ayant peu de connaissances techniques de s'approprier les nouvelles fonctionnalités du web. Ainsi, les internautes contribuent à l'échange d'informations et peuvent interagir (partager, échanger, etc.) de façon simple, à la fois avec le contenu et la structure des pages, mais aussi entre eux, créant ainsi notamment le Web social. L'internaute devient, grâce aux outils mis à sa disposition, une personne active sur la toile.

Web invisible

Partie du web correspondant à l'ensemble des documents web qui ne sont pas indexés par les outils de recherche traditionnels. [office de la langue française – 2003]

Wiki

Un wiki est un site Web dont les pages sont modifiables par les visiteurs afin de permettre l'écriture et l'illustration collaboratives des documents numériques qu'il contient.

Workflow

Un workflow (anglicisme) est la représentation d'une suite de tâches ou opérations effectuées par une personne, un groupe de personnes, un organisme, etc. Le terme flow renvoie au passage du produit, du document, de l'information, etc., d'une étape à l'autre.

XML

eXtensible Markup Language est le langage universel pour les bases de données utilisées sur le web. Il définit les propriétés des données tandis que le protocole html explique comment les afficher. [microsoft – laboratoire des technologies]

Annexes

6.1. Questionnaire d'évaluation de la plate forme Xplor EveryWhere – Système d'Information d'Intelligence Economique

Les entreprises font face aujourd'hui à une concurrence accrue sur des marchés extrêmement dynamiques et imprévisibles : nouveaux entrants, fusions et acquisitions, baisses tarifaires brutales, évolutions rapides des modes de consommation et des valeurs, fragilité des marques et de leur réputation, ... Les facteurs de risque externes n'ont jamais été aussi nombreux, sans s'étendre sur la crise financière.

Dans ce nouvel espace, le système d'information d'intelligence économique (SIIE) est devenu la principale composante des évolutions stratégiques de l'entreprise. Les technologies de l'information ne sont maintenant plus considérées comme le support de l'organisation mais bien comme un actif stratégique fondateur, capable de conférer à l'entreprise un avantage concurrentiel durable.

Dans ce cadre, nous présentons notre système d'information **Xplor EveryWhere (XEW)** basé sur un modèle d'analyse exploratoire multidimensionnelle. L'objectif du système est d'offrir une référence méthodologique aux opérations de collecte, de traitement et d'analyse des informations. Il doit permettre d'observer et d'analyser l'environnement informationnel d'un problème décisionnel sous toutes ses dimensions. Ce modèle repose sur l'exploitation des informations par l'analyse de l'évolution de leurs interactions afin de comprendre et résumer le comportement de l'environnement par de nouvelles connaissances synthétiques. XEW est un Portail Web pour Mobile d'Intelligence Economique, qui permet aux utilisateurs de nos solutions de veille de continuer à rechercher, surveiller, valider et rediffuser des informations stratégiques au cours de leurs déplacements. Ils n'ont ainsi plus besoin d'être assis face à leur ordinateur pour accéder à des données utiles dans l'instant (préparation d'une réunion, nouvel ordre du jour, information sur un interlocuteur, une technologie, un marché, demande urgente d'une analyse ou d'un focus spécifique). Il leur suffit simplement d'utiliser discrètement leur appareil mobile de type iPhone ou Blackberry. Grâce à ce Portail Web Mobile, ils sont en permanence au cœur de l'information pertinente. Notre Portail Web pour Mobile se base sur les ontologies de domaines, sur les réseaux d'acteurs et une gestion de la sécurité elle-même contrôlée par une ontologie spécifique. Il est alors possible de :

Consulter des informations à jour, car nous accédons à notre serveur de base de donnée stratégique en temps réel, lui-même alimenté quotidiennement par des veilleurs.

Faire remonter des informations « terrain » lors de salons, de visites en clientèle ou à l'issue de réunions.

Demander des renseignements spécifiques en urgence qui seront mis en ligne par les veilleurs.

Grâce à l'évolution de la technologie, un tel portail pour mobile vous permettrait de gagner en efficacité et en réactivité car, à tout moment, il est possible d'accéder à toutes les informations stratégiques utiles au décideur nomade, lui-même pouvant faire remonter très vite l'information « terrain » qui peut éventuellement déclencher d'autres analyses stratégiques.

Ce questionnaire a pour objectif d'évaluer cet outil, en ciblant plusieurs axes :

- **L'interface : convivialité, facilité d'accès aux données, facilité de manipulation.**
- **L'opérabilité : exploitation des résultats, qualité des résultats.**
- **L'interactivité : possibilité pour l'utilisateur de maîtriser pleinement son application.**

Pour cela, nous vous proposons de tester notre outil sur plusieurs jeux de données. Nous utiliserons les notations suivantes :

- **TD=Très Difficile**
- **D = Difficile**

- M = Moyen
- F = Facile
- TF = Très Facile

Renseignements

- Niveau en informatique
Débutant ☐ Intermédiaire ☐ Expert ☐
- Expérience dans le décisionnel
Débutant ☐ Intermédiaire ☐ Expert ☐
- “ Avez-vous déjà utilisé des outils de visualisation de données ? ” Oui ☐ Non ☐

Si oui, citez les outils que vous avez déjà utilisés :

.....
.	.
.....
.	.
.....
.	.
.....
.	.
.....
.	.

- Niveau d'étude
Bac ☐ Bac+1 ☐ Bac+2 ☐ Bac+3 ☐ Bac+5 ☐ Bac+8 ☐
Autre
- Votre domaine d'activité ?

.....
.....

- Profession actuelle
Etudiant(e) ☐ Professeur ☐
Ingénieur ☐ Cadre et profession libérale ☐
Documentaliste ☐ Veilleur ☐
Autre

Menu Statistique Descriptive

La tâche	Niveau de difficulté				
	T	D	M	F	TF
• Lancer le menu statistique descriptif					
• Lancer les statistiques sur un champ					

Menu Carte Géostratégique

[illegible]

De manière générale...

- Comment qualifieriez-vous cet outil ?
Très utile ☐ Assez utile ☐ Peu utile ☐ Inutile ☐ Sans avis ☐
- La représentation des données est
Adaptée ☐ Non adaptée ☐ Sans avis ☐
- L'outil facilite-t-il l'exploration des données
Oui ☐ Non ☐ sans avis ☐
- La détection des différentes tendances, des signaux forts est-elle facilité ?
Oui ☐ Non ☐ sans avis ☐
- Qualifieriez-vous l'outil d'interactif ?
Oui ☐ Non ☐ sans avis ☐
- Qualifieriez-vous l'outil d'ergonomique ?
Oui ☐ Non ☐ sans avis ☐
- L'accès aux différentes fonctionnalités est-il difficile ?
Oui ☐ Non ☐ sans avis ☐
- Avez-vous déjà effectué une analyse de données avec la plate-forme « Tétralogie » ?
Oui ☐ Non ☐ sans avis ☐
- Si oui, que pensez-vous de l'ajout de ce module à cette plate-forme ?
Indispensable ☐ Très utile ☐ Utile ☐ Sans grand intérêt ☐ Inutile ☐ Sans avis ☐
- Que pensez-vous de la mobilité de Xplor (la version pour téléphone portable) ?
Indispensable ☐ Très utile ☐ Utile ☐ Sans grand intérêt ☐ Inutile ☐ Sans avis ☐
- Qualifieriez-vous Xplor de plate forme d'Intelligence Economique ?
Oui ☐ Non ☐ sans avis ☐
- Quels conseils donneriez-vous pour l'améliorer ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

- Si votre domaine d'activité n'est pas directement lié à la visualisation de données, pensez-vous que cet outil pourrait vous être utile professionnellement ?

Oui ☐ Non ☐ sans avis ☐

6.2. Le prototype Xplor EveryWhere pour iPhone (iXEW)

➤ Interface graphique :

A l'aide de l'Interface Builder, un bouton a été placé sur toute la fenêtre de la page d'accueil, il est cependant caché par l'image. Cela permet d'accéder à la page de connexion en cliquant n'importe où sur la page.



Figure 98. Page d'accueil

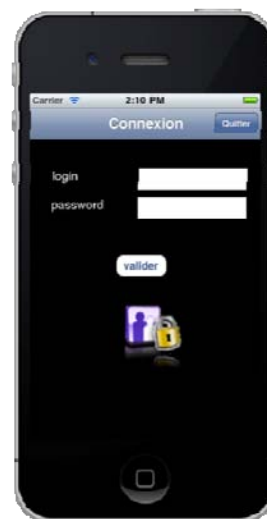


Figure 99. Page de connexion

➤ TableView :

Une tableView est une liste déroulante qui permet lorsque l'on clique sur une cellule d'effectuer une action. La cellule cliquée est enregistrée et permet d'afficher les informations correspondantes aux choix de l'utilisateur. Cette liste provient des données de la base. On peut ainsi récupérer grâce au parsing du fichier XML les différentes études stratégiques disponibles, la liste des catégories et la liste des personnes ou groupes associés à ces catégories. Toutes ces informations sont ensuite affichées dans la tableView.

<div>Analyses</div> <div>SCAFFOLD</div> <div>Quitter</div>	
Agents	>
Auteurs	>
Déposants	>
Journaux	>
Pays	>

Figure 100. TableView des différents acteurs

➤ Ergonomie Utilisateur :

Pour faciliter l'ergonomie du iXEW, plusieurs éléments ont été mis en place. En effet, il est essentiel qu'une application iPhone soit très facile à manipuler et doit respecter quelques règles de conception pour l'interface utilisateur décidée par Apple¹⁰. Par exemple, lorsque l'on décide de passer l'application en arrière plan, il faut pouvoir la retrouver là où on l'a quittée quand on décide de la reprendre.

- Message d'alerte :

Lors de la connexion, si l'utilisateur se trompe dans son login ou son mot de passe, un message apparaît à l'écran l'informant de son erreur (figure 101-1). Il doit alors recommencer sa connexion après avoir cliqué sur le bouton « ok » de cette nouvelle fenêtre.

Un autre message d'alerte peut apparaître si les choix de l'utilisateur ne donnent aucun résultat dans le top 10 et dans la liste complète. Ici on peut voir que l'étude « Scaffold » ne contient pas d'information pour la catégorie « Journaux » (figure 101 -2).

Lorsque l'on clique sur le bouton « quitter » un message apparaît (figure 101-3) demandant à l'utilisateur s'il est sûr de vouloir quitter l'application ou non. Si l'on clique sur le bouton « Continuer » on retourne alors sur la page d'accueil et il faut se connecter à nouveau pour accéder aux autres pages, « Annuler » permet de reprendre là où l'utilisateur a laissé l'application.

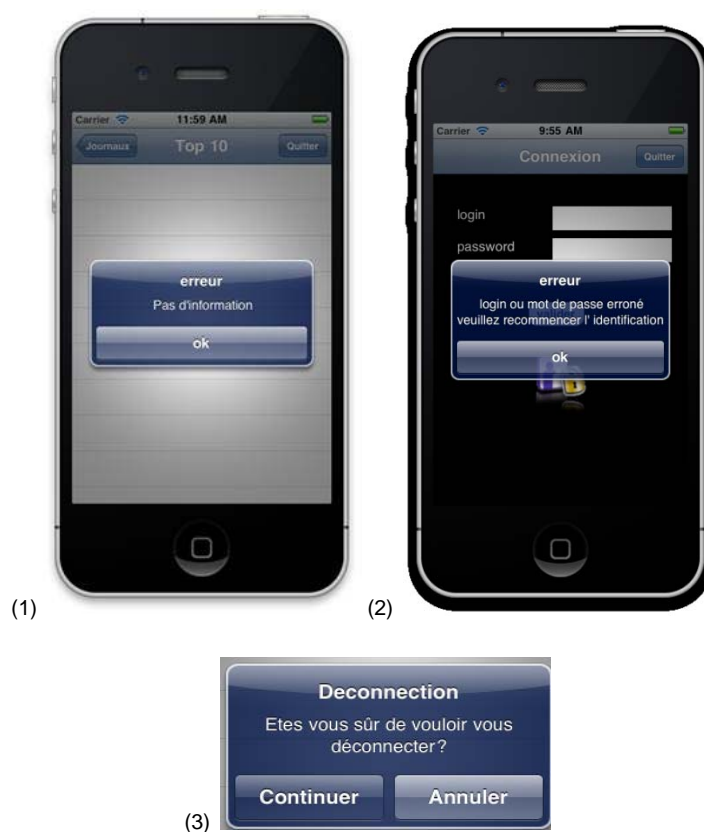


Figure 101. Différents messages d'alerte possible

- Barre de navigation:

¹⁰ Ces règles sont décrites dans le document suivant :

<http://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/MobileHIG.pdf>

- Sur chaque page de l'application apparaît une barre de navigation (figure 15-1) contenant un titre, mais aussi un bouton de retour à la page précédente (à gauche) et un bouton « quitter » (à droite) qui permet de revenir à la première page de l'application.



Figure 102. Barre de navigation de iXEW

Ici, nous avons choisi la vue « Année ». Nous pouvons retourner à la liste des choix possibles pour la catégorie 'Pays' (top10, liste complète, cartes géostratégiques), ou quitter l'application. Le bouton « Quitter » permet à l'utilisateur de se déconnecter, contrairement au bouton de gauche qui est un simple bouton de retour qui entraîne l'affichage de la vue précédente.

D'autres boutons peuvent apparaître comme « Analyse » (figure 102-2) qui permet de retourner directement sur la vue 'Analyse' sans devoir repasser par la vue 'Catégorie'. Cela permet un retour plus rapide à des vues précédentes sans devoir cliquer sur le bouton retour de chaque vue.

- Navigation entre les graphiques :

Pour naviguer entre les différents graphiques disponibles, il suffit de passer le doigt sur l'écran de droite à gauche pour afficher le graphique suivant et l'inverse pour le graphique précédent. Cela permet de changer très facilement de vue. Des petites flèches à gauche et à droite indiquent que l'on peut naviguer entre les graphiques.

Pour cela, il a été nécessaire de mettre en place un objet de type « UISwipeGestureRecognizer » qui détecte les mouvements d'un doigt sur l'écran. On appelle ensuite une méthode qui permet de changer de vue (même principe que pour le changement de vue du « navigation controller »).

- Indicateur d'activité :

Une petite roue tournante indique que le chargement des données est en train de s'effectuer. Elle permet ainsi de signifier qu'il n'y a pas de problème mais qu'il faut attendre pour voir s'afficher les informations. En effet le parsing du fichier XML peut prendre un peu de temps (quelques centièmes de secondes à plusieurs secondes) suivant la quantité d'information à traiter. Ici il s'agit d'une tableView sur le Top10 des auteurs, il faut donc calculer pour chaque auteur son nombre de co-occurrences et ensuite le classer par ordre décroissant puis de garder seulement les 10 premiers. Une fois le chargement des données terminé, l'indicateur d'activité est caché.



Figure 103. Indicateur d'activité

- Barre de recherche :

Une des possibilités de iXEW est d'obtenir la liste complète de tous les auteurs pour une étude donnée. Cette liste pouvant être longue, une barre de recherche peut permettre de trouver rapidement un auteur dans la liste en entrant les premières lettres. En cliquant ensuite sur le bouton « Search » nous obtenons la deuxième figure qui correspond au résultat de la recherche. Le bouton « Cancel » permet de revenir à la liste complète.

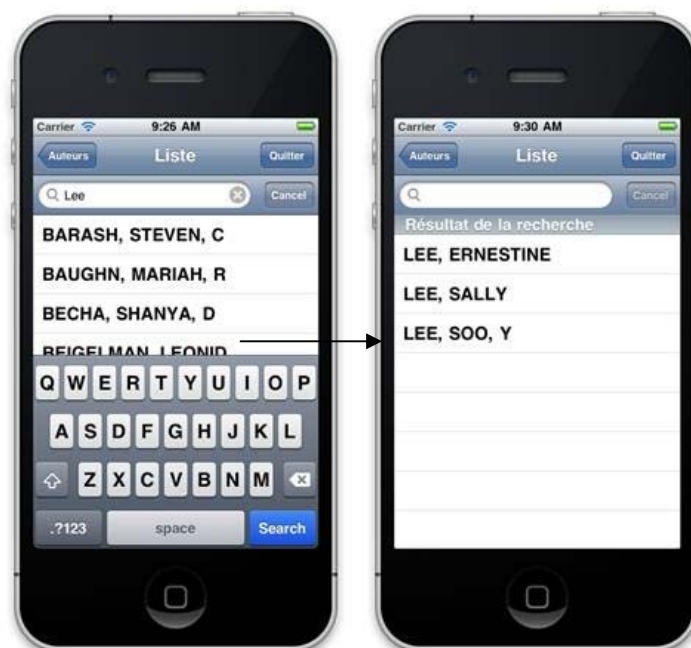


Figure 104. Barre de recherche

Nous pouvons ainsi voir qu'il y a 3 auteurs dans cette étude qui porte le nom de Lee. Ici, comme il s'agit d'une liste de noms, la recherche porte sur une correspondance avec le début du nom uniquement. On suppose que la personne connaît au moins la première lettre du nom qu'elle cherche. En changeant une expression du programme on peut chercher une correspondance dans tout le nom.

Il existe une méthode qui permet la rotation des vues selon la position de l'iPhone. Pour les tableView, il suffit simplement que cette fonction retourne « YES ». Pour les autres vues comme celle de la carte il faut indiquer dans le code ou avec l'Interface Builder (si l'objet a été créé sans code) le changement à effectuer pour chaque objet.



Figure 105. Liste des agents

➤ Statistiques descriptives:

Pour la création des graphiques il est nécessaire d'installer deux librairies que sont Coreplot et Quartzcore. Coreplot contient différentes classes permettant ensuite de créer des objets de type « barplot », « scatterplot » et « pieplot ». Les graphiques sont accessibles en cliquant sur une des cellules du TOP10. On obtient ainsi une évolution du relationnel depuis 2008.

Pour obtenir les données, on se base sur les clics de l'utilisateur et on parse les informations correspondantes. Par exemple, si l'utilisateur choisit l'étude « Scaffold » puis la catégorie « Pays » et le pays Etats-Unis, toutes les lignes où apparaissent les Etats-Unis vont être récupérées dans la table co-occurrences. Puis, l'année de publication sera considérée pour créer une variable de somme par année. Ces variables sont ensuite regroupées dans un tableau.

Les graphiques suivant représentent toutes les publications des Etats-Unis pour l'étude « Scaffold ».

- Histogrammes et Courbes :

Pour l'histogramme et la courbe, l'utilisateur peut utiliser le zoom et déplacer les graphiques. Un objet de type date a été créé ainsi qu'un objet intervalle de temps d'un an pour permettre d'afficher les années sur l'axe des abscisses.

L'axe des ordonnées est calculé en fonction de la valeur maximale du nombre de publications car parfois on peut avoir comme valeur maximale une dizaine de publications à plus de mille comme pour les Etats-Unis. On peut ainsi voir la tendance du nombre de publications. Ici on note qu'il y a une baisse par rapport à 2008 (figure 105 et 106).

L'histogramme représente l'évolution du relationnel. Au dessus de chaque barre, on a rajouté le total.

La courbe permet de voir les évolutions positives et négatives du nombre de publications par an. Lorsque l'utilisateur clique sur un point de la courbe, le nombre de publications pour l'année sélectionnée apparaît juste au dessus.

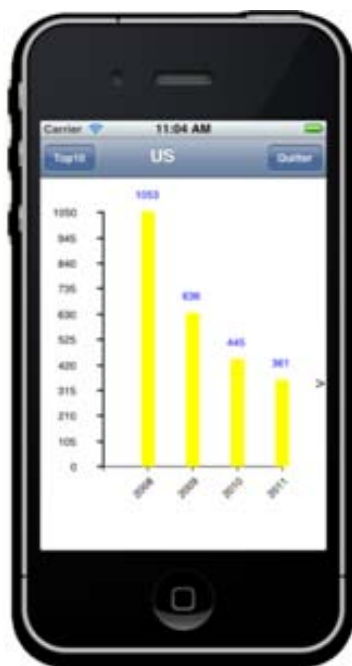


Figure 106. Histogramme de l'évolution des Etats-Unis



Figure 107. Courbes de l'évolution des Etats-Unis

- Diagrammes en secteurs

Le diagramme en secteurs permet de voir la répartition des publications depuis 2008. Les différents secteurs sont classés du plus grand au plus petit dans le sens des aiguilles d'une montre en partant du haut. Cela permet de repérer quelles sont les années où il y a eu le plus ou le moins de publications. Sur le graphique ci-dessous, on peut voir qu'un peu plus d'un tiers des publications des Etats-Unis datent de 2008.

Cela permet de noter si la répartition des publications est homogène entre les années. Ici on peut voir que 2008 représente une part très importante de leur publication totale.

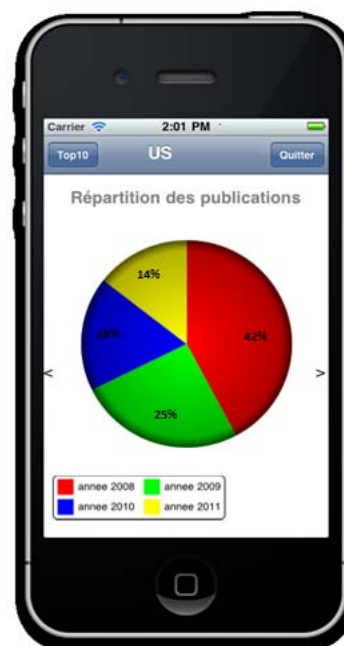


Figure 108. Diagrammes de secteurs des Etats-Unis

➤ Cartes

Pour la catégorie « Pays », l'utilisateur peut choisir d'afficher une carte géostratégique. L'installation de la librairie MapKit est nécessaire pour créer une carte Google Map. Cette carte a le nom « Google » inscrit en bas à gauche car il s'agit d'une condition obligatoire si on veut utiliser la carte dans iXEW. On peut ensuite ajouter des annotations à l'aide de pins placés avec une latitude et longitude données.

Lorsque l'on clique sur un pin, on peut avoir des informations qui nous intéressent. Les pays n'ayant pas de publications ne sont pas représentés par une annotation sur la carte.



Figure 109. La carte géostratégique de « SCAFFOLD »

Cette carte montre le nombre de publications par pays en 2011. On peut choisir l'année (ou le total) avec la page précédente qui est une tableView des années.

➤ Tests sur iPhone:

Dans un premier temps, iXEW est testée sur un simulateur d'iPhone permettant ainsi de voir si cela fonctionne. Cependant il peut y avoir parfois des problèmes quand on teste avec un iPhone. Ces problèmes peuvent être dus à une mauvaise gestion de la mémoire.

Pour pouvoir effectuer des tests sur iPhone ou iPad il faut avoir une licence « IOS Developer Program » qui est payante et valable un an. Elle permet de pouvoir tester son application sur 100 appareils enregistrés.

Une fois cette phase de tests terminée, il faudra soumettre iXEW à Apple. En effet, Apple doit vérifier que l'application fonctionne bien et correspond à la description donnée. Si l'application est rejetée, il faut alors corriger le ou les problèmes décelés avant de soumettre à nouveau l'application.

Une fois l'application validée, elle est ensuite mise sur l'appStore, et pourra ensuite être téléchargée par les utilisateurs. Apple transmet aux créateurs de l'application les bugs récurrents que peuvent rencontrer les utilisateurs afin de les corriger et de sortir une mise à jour de l'application.

➤ Gestion de la mémoire :

Lorsque l'on crée un objet (tableau, dictionnaire,...) on alloue un espace mémoire. Cependant la mémoire disponible sur l'iPhone pour une application est limitée. Pour éviter un plantage de l'application il faut donc gérer la mémoire.

Il faut libérer tout les objets créés avec le mot-clé alloc, new, copy ou retain comme propriété par un release ou autorelease.

- Outils Leaks :

L'outil Leaks permet de repérer les fuites de mémoire afin d'éviter les problèmes lors de l'utilisation de l'application. Pour cela, il suffit d'accéder à l'application via l'iPhone ou l'iPad, en parcourant les différentes vues on peut apercevoir les fuites de mémoire. Il suffit de lancer l'application en mode

‘Profile’ et de choisir l’instrument ‘Leaks’. On peut savoir quel est l’objet qui en est à l’origine est la corriger. Le simulateur permet aussi de détecter les fuites mais étant donné que les résultats sur le simulateur ne sont pas toujours identiques à ceux de l’iPhone ou iPad il ne permet pas de repérer toutes les fuites.

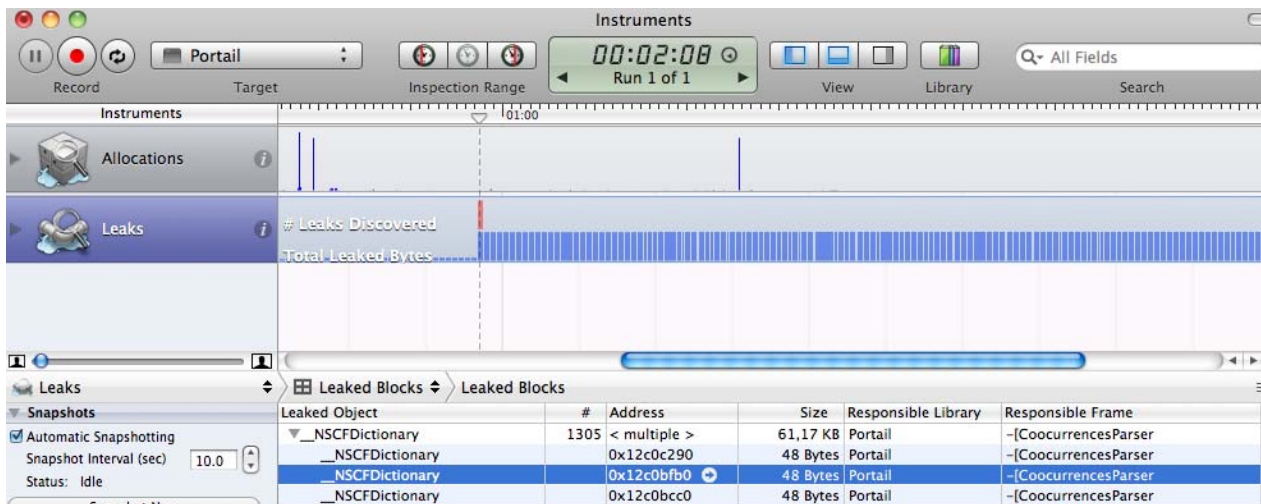


Figure 110. Outils Leaks (1)

Le pic orange (figure 110) signifie qu’il y a une fuite, on peut voir en dessous la liste des fuites, ici en surbrillance il s’agit d’un dictionnaire de la classe CooccurrencesParser.

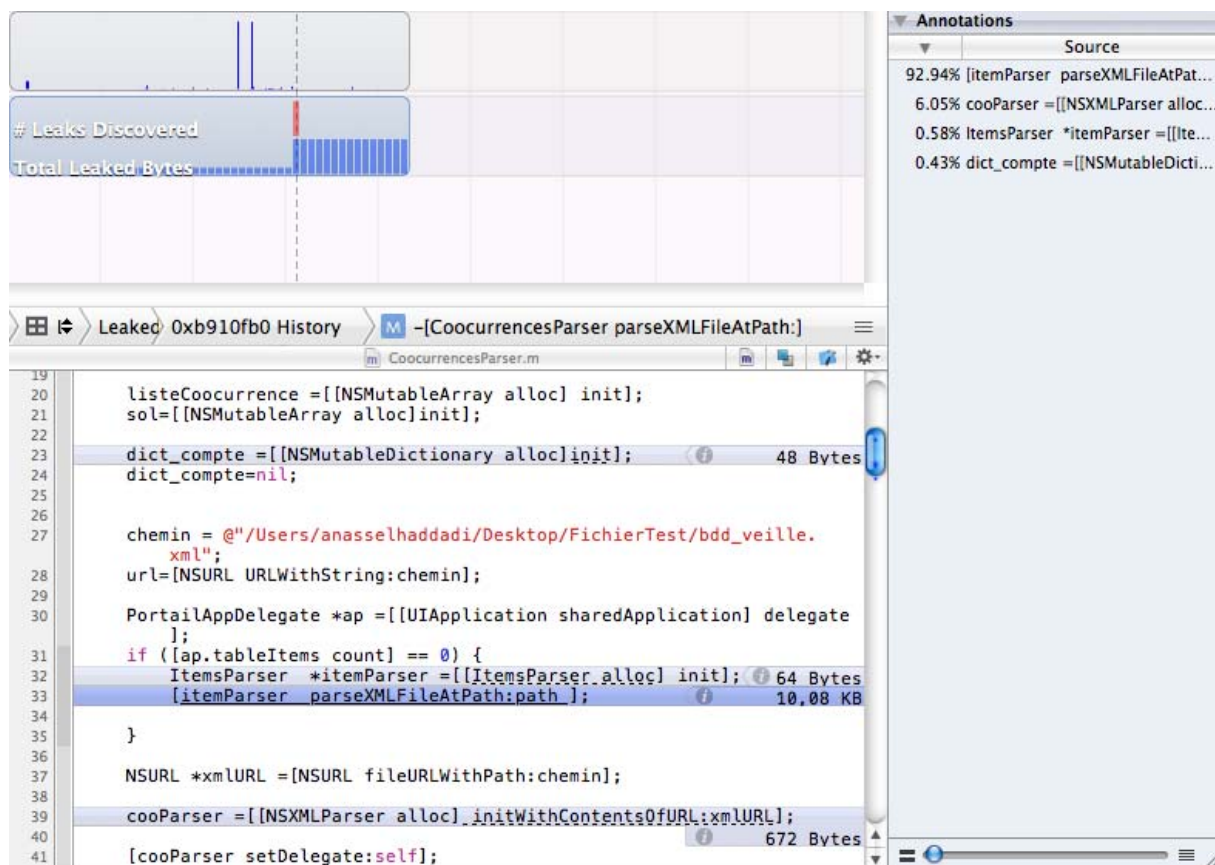


Figure 111. Outils Leaks (2)

En cliquant sur cette fuite, on arrive à trouver les lignes de codes qui sont à l'origine de ces fuites avec leur taille, et leur pourcentage par rapport au total. Ici une ligne correspond à plus de 90 % du total de la fuite des dictionnaires de la classe 'CooccurrencesParser'.