



# Systèmes d'information et Intelligence économique : enjeux et perspectives

Sahbi Sidhom, Malek Ghenima, Philippe Lambert

## ► To cite this version:

Sahbi Sidhom, Malek Ghenima, Philippe Lambert. Systèmes d'information et Intelligence économique: enjeux et perspectives. NT2S Consulting Inc. 4ème Colloque International sur l'Intelligence Économique et le Knowledge Management - IEMA-4, May 2010, Alger, Algérie. NT2S Consulting Inc., 1, 2010, L'Intelligence Économique et le Knowledge Management. <inria-00549766>

**HAL Id: inria-00549766**

**<https://hal.inria.fr/inria-00549766>**

Submitted on 22 Dec 2010

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

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

# Systemes d'information et Intelligence économique : enjeux et perspectives

**Sahbi SIDHOM** (LORIA/KIWI & Nancy Université - France), **Malek GHENIMA** (ESCE & Université de la Manouba - Tunisie), **Philippe LAMBERT** (VinaLor Nancy - France).

e-Mails : sahabi.sidhom@loria.fr, Malek.Ghenima@escem.rnu.tn, philippe.lambert@vinalor.fr

**Résumé** — Poser « une stratégie économique » cela nécessite de disposer d'un maximum d'informations pertinentes, recueillies et produites dans un contexte légal. Le rôle de l'intelligence économique (IE) est donc de fournir aux décideurs des informations pertinentes synthétisées et orientées. Ces informations constituent une source indispensable pour prendre la bonne décision permettant d'engager la pérennité de l'organisation.

Par anticipation dans ce contexte, l'intelligence économique est à la fois un terrain de réflexions, une pratique avec des retours d'expérience, qui nécessite une approche inter- et pluri- disciplinaire pour être correctement appréhendée, avec un investissement de haut niveau par les acteurs scientifiques, économiques et socio-politiques. L'ensemble doit être structuré en processus communicants, avec les acteurs, les ressources et les outils, notamment les systèmes d'information (SI), pour permettre de convertir les données d'une organisation en informations décisionnelles.

Ainsi, les enjeux et les perspectives dans la conjoncture des systèmes d'information et de l'intelligence économique (SIIE) sont plus que jamais posés. C'est dans cet ordre de réflexion que nous abordons le sujet pour clarifier des hypothèses fondamentales (les concepts) pour valider nos approches expérimentales par la modélisation et de méthodologies appliquées.

Comme étude de cas et retour d'expérience dans le contexte de SIIE, nous présentons un système d'information pour l'aide à la décision dans la gestion des maladies chroniques sous l'initiative du Haut Conseil de la santé publique en France.

Le Haut Conseil en question a mis en place une série de groupes de travail thématiques pour formuler des recommandations pour la future politique de santé. Pour la gestion des maladies chroniques, un système d'information sur le sujet a été proposé. Le système était conçu pour permettre aux experts et groupes de travail de mieux comprendre comment la gestion des maladies chroniques dans les pays d'Europe de l'Ouest a été gérée. Sur le terrain d'études, l'institut d'information scientifique et technique (INIST), filiale du Centre National de Recherche Scientifique (CNRS France), a accepté d'accueillir ce projet et son étude avec ses ressources et banques de données en IST (information scientifique et technique). Ce travail couvre toutes les phases de l'exploration de données depuis l'analyse des besoins jusqu'à la réalisation des livrables, comme un système d'information pour l'aide à la décision.

Dans le volet recherche, c'était la situation de tester la plateforme linguistique NooJ, comme une boîte à outils pour le traitement automatique du langage naturel (TALN), pour l'extraction de connaissances afin d'assurer l'exploration par la visualisation de données.

Dans cette étude, d'abord, nous considérons l'approche formelle utilisée pour extraire des connaissances à partir d'un corpus de

références bibliographiques que nous avons formé pour le suivi scientifique des maladies chroniques. Dans cette approche, Sahbi Sidhom dans ses recherches, a construit une plate-forme d'analyse pour l'indexation automatique de contenus (IAC) et la recherche d'information (RI), à l'aide de diagrammes syntaxiques sur corpus.

Certains résultats sont présentés par le traitement de NooJ sur le corpus en utilisant les ressources linguistiques: dictionnaires électroniques et les règles morpho-syntaxiques de la langue. Plus précisément, nous analysons les unités de haut poids sémantiques extraites du corpus pour permettre la visualisation dans un graphe sémantique de concepts.

Dans la dernière partie, nous discutons des bases de résultats sur le traitement d'informations, sa représentation en connaissances et sa visualisation pour faciliter le processus de décision. Ce travail a couvert toutes les phases d'un travail de veille informationnelle, depuis l'analyse stratégique des besoins jusqu'à la réalisation du livrable, un système d'information d'aide à la décision.

**Mots-clés** — système d'information d'aide à la décision, intelligence économique, traitement automatique du langage (TAL), analyse stratégique des besoins, management des connaissances (KM), maladies chroniques, santé, veille informationnelle.

## INTRODUCTION GENERALE

L'intelligence économique, qui est déclinée dans la terminologie anglaise en « Business Intelligence » (BI), a été perçue dans la communauté scientifique comme étant l'activité de production des connaissances, servant les buts économiques et stratégiques d'une organisation.

Poser « une stratégie économique » cela nécessite de disposer d'un maximum d'informations pertinentes, recueillies et produites dans un contexte légal. Le rôle de l'intelligence économique est donc de fournir aux décideurs des informations pertinentes synthétisées et orientées. Ces informations constituent une source indispensable pour prendre la bonne décision permettant d'engager la pérennité de l'organisation.

Des outils informatiques puissants sont mis à la disposition des collecteurs (pour recueillir des informations électroniques), analystes et veilleurs pour identifier les informations nécessaires, utiles jusqu'à la prise de décision par les acteurs qualifiés (décideurs). Parce qu'en intelligence

économique (IE ou BI) la décision est capitale, le terme "informatique décisionnelle" est souvent associé au concept et aux processus inhérents au domaine. L'informatique décisionnelle est fondée sur l'exploitation de données diverses et transversales provenant de sources informationnelles différentes.

Par anticipation dans ce contexte, l'IE est à la fois un terrain de réflexions, une pratique, qui nécessite une approche inter- et pluri- disciplinaire pour être correctement appréhendée, avec un investissement de haut niveau par des acteurs scientifiques, économiques et socio-politiques. L'ensemble doit être structuré de processus, de personnes et d'outils, notamment les systèmes d'information (SI), va permettre de convertir les données d'une organisation en informations décisionnelles.

Partant de ce constat, les entreprises, les universités et même les états sont inmanquablement appelés à s'investir dans cette nouvelle discipline qui joue un rôle essentiel pour la performance, la compétitivité et la protection du patrimoine matériel et immatériel, qui sont devenues des enjeux majeurs. La question du rapport entre information - valeur ajoutée, profilage - acteurs, connaissance et décision est plus que jamais posée. C'est ce que nous allons présenter dans cet article par les aspects conceptuels sur les modélisations [25], [26] et les processus en systèmes d'information et intelligence économique, afin de clarifier les enjeux et les perspectives, dans la partie I.

Pour l'aspect pratique, en partie II. par une étude de cas, nous vous présentons ce que le Gouvernement Français a mis en place en 2007, par un ambitieux plan sur l'amélioration de la qualité de vie des personnes atteintes de maladies chroniques qui sera décliné jusqu'en 2011. En application de ce plan, le projet de loi de financement de la sécurité sociale 2007 dispose que les caisses nationales d'assurance maladie peuvent mettre en place des programmes d'accompagnement des patients atteints de maladies chroniques visant à leur apporter des conseils en termes d'orientation dans le système de soins et d'éducation à la santé [14].

Pour la plupart des acteurs du système de santé et en premier lieu pour les patients, la prise en charge des maladies chroniques se résume au dispositif médico-administratif des affections de longue durée (ALD). Or les ALD, simples ou exonérantes, sont presque toujours des maladies chroniques.

Dans ce contexte, Le Haut Conseil de la Santé Publique en France a mis en place une série de groupes de travail thématiques pour formuler des recommandations pour la future politique de santé. Pour ce besoin stratégique en matière de santé publique et spécifiquement pour la gestion des maladies chroniques, nous avons proposé un système d'information.

Le système était conçu pour permettre aux experts du groupe de travail de mieux comprendre comment la gestion des maladies chroniques dans les pays d'Europe de l'Ouest a été gérée. Sur le terrain d'études, l'Institut d'information scientifique et technique (INIST), filiale du Centre National de Recherche Scientifique (CNRS France), a accepté d'accueillir ce projet et son étude avec ses ressources et banques de données en IST. Ce travail couvre toutes les phases de

l'exploration de données depuis l'analyse des besoins jusqu'à la réalisation des livrables, comme un système d'information pour l'aide à la décision.

Dans cet article, d'abord, nous considérons l'approche formelle utilisée pour extraire des connaissances à partir d'un corpus de références bibliographiques que nous avons formé pour le suivi scientifique des maladies chroniques. Dans cette approche, Sahbi Sidhom dans la présentation de son approche a construit une plate-forme d'analyse pour l'indexation automatique et la recherche d'information (RI), à l'aide de diagrammes syntaxiques sur les corpus. Sur le volet recherche, c'était la situation de tester la Plateforme linguistique NooJ, comme une boîte à outils pour le traitement du langage naturel (NLP), pour l'extraction de connaissances afin d'assurer l'exploration par la visualisation de données.

Certains résultats sont présentés par le traitement de NooJ sur le corpus en utilisant les ressources linguistiques: dictionnaires électroniques et les règles morpho-syntaxiques de la langue. Plus précisément, nous analysons les unités de haut poids sémantiques extraites du corpus pour permettre la visualisation dans un graphe sémantique.

Dans la dernière partie, nous discutons des bases de résultats sur le traitement d'informations, sa représentation en connaissances et sa visualisation pour faciliter le processus de décision. Ce travail a couvert toutes les phases d'un travail de veille informationnelle, depuis l'analyse stratégique des besoins jusqu'à la réalisation du livrable, un système d'information d'aide à la décision.

## I. SYSTEMES D'INFORMATION : IMPLICATION DE L'IE ET DU K.M

Nous remarquons l'apparition de plus en plus d'outils de veille et de systèmes d'information natifs à la démarche de l'intelligence économique. Ces outils sont souvent conçus pour aider les entreprises à maîtriser leur organisation interne et à comprendre leur environnement externe par l'acquisition, le classement, l'analyse et l'interprétation d'informations selon des modèles d'exploitation automatiques, itératifs et intégratifs. Ces outils sont classés en deux catégories : (i) le premier concerne les systèmes de gestion de l'information, qui permettent une manipulation massive des données opérationnelles et l'extraction des connaissances ; (ii) le second consiste à automatiser la collecte et l'analyse des informations à partir d'un environnement concurrentiel pour l'aide à la prise de décision stratégique. On observe bien que le processus de "veille" est inhérent sans le désigner en tant que tel dans les deux catégories de systèmes d'information.

Cependant, la phase de l'analyse des besoins d'information n'a pas bénéficié d'autant de recherche et de réflexion de la part des acteurs en entreprises et des chercheurs dans les universités ou la branche R&D des organisations, qui s'intéressent aux synergies des deux domaines, les systèmes d'information (outils et architectures fonctionnels) et l'intelligence économique (processus et méthodologies sur les

pratiques de terrain). La définition du besoin d'informations constitue logiquement une étape fondamentale dans la démarche de l'intelligence économique [19], elle ne peut pas être comprise sans prendre en considération le contexte, l'environnement, l'objet sur les risques, les acteurs impliqués et l'enjeu encouru dans l'organisation si non actions adaptées.

Partant de ce constat, ce paragraphe propose de préciser l'apport déterminant du Management des Connaissances (ou KM. dans la terminologie anglaise du Knowledge Management) pour la compréhension et l'explicitation sur le développement d'une méthodologie efficace pour réussir cette phase de conjonction entre SI et IE (par le processus KM).

Tout d'abord, le courant classique qui s'intéresse principalement aux aspects techniques et fonctionnels des systèmes d'information présente des recherches actives avec le développement des moteurs de recherche et des langages d'indexation ; En deuxième, un autre courant qui se développe sur les aspects cognitifs de l'utilisateur en considérant le système d'information comme un système de communication entre un producteur d'informations et les usagers/acteurs potentiels ; En dernier, le courant qui cherche à définir la relation entre les usages dans un système et les usagers de l'information. Ce dernier montre l'importance accordée à l'utilisateur (ou acteur) avec la nécessité de calculer (par le système) l'utilisation de l'information dans un processus de recherche d'informations (RI). Dans les actuelles recherches sur le domaine, le paradigme de l'utilisateur nécessite: (i) de porter attention sur les besoins réels de l'utilisateur et son environnement avec les impacts dans le travail collaboratif ou coopératif et (ii) d'améliorer la qualité du service rendu par un système informatique à un utilisateur identifié ou non (si le profil est calculé de l'utilisateur avec ses traits et préférences par les ressources d'un système de recommandation).

Le point commun de ces différentes approches est de proposer une modélisation des usagers avec leurs profils, préférences et comportements. Le problème à résoudre peut s'exprimer par la détermination de la ou les ressources les plus pertinentes à présenter à l'utilisateur de manière à maximiser sa satisfaction à tout moment, dans tout contexte et en tout lieu afin d'initier un processus de décision en retour sur la base d'informations pertinentes (par un processus de veille) [10].

Dans cette dynamique, l'information comme une ressource, elle détermine la production de la connaissance pour l'utilisateur ou l'acteur économique et social. Le KM ne peut y être écarté dans une telle discipline en considérant que les connaissances « de l'entreprise » constituent un patrimoine et qu'à ce titre elles doivent être gérées. Intrinsèquement, la gestion des connaissances ne peut se présenter que par « un processus de création, de transfert et d'utilisation des connaissances dans le but d'améliorer le rendement de l'organisation (Hamilton, 1998) » [12], [13]. A ce titre, la gestion des connaissances est liée à deux types d'activités :

- les activités qui aident à s'informer et à s'approprier les connaissances individuelles, qui les transforment en capital et mémoire de l'organisation et servant à diffuser ce savoir explicite pour les acteurs de l'organisation.

- les activités qui facilitent les échanges entre les acteurs, dans le cadre desquelles ils partagent un savoir tacite, non codifié, pour devenir par évolution des pratiques opérationnelles dans un SI.

En synthèse de notre réflexion et des travaux de plusieurs auteurs dans le domaine, comme (Argyris, 1995), (Oubrich, 2003), (Bournois et Romani, 2000), (Prahalad et Hamel, 1990), (Prusak, 1997), (Spender et Grant, 1996) etc., l'IE et le KM, par la vision sur les processus, consistent en une gestion stratégique de l'information et de la connaissance qui visent à fournir « la bonne information, à la bonne personne, au bon moment » [18], [22]. Les deux domaines regroupent un ensemble d'activités reliées entre elles par des flux d'information, qui sont organisées dans le temps. Leur interférence permet d'obtenir des informations et des connaissances à valeur ajoutée, comme « un savoir à la fois valide et pouvant servir des actions stratégiques » par les acteurs dans la vie de l'organisation.

## 1. SYSTEMES D'INFORMATION : LES ENJEUX

L'usage professionnel de l'information est rythmé par le progrès technologiques des plateformes d'information et de communication. Grâce à l'ouverture des réseaux professionnels de communication, les techniques avancées et leurs capacités à rendre notre système d'information communicant avec la robustesse pour traiter des données volumineuses et hétérogènes. Ainsi, la progression quasi-généralisée des professionnels avec les systèmes d'accès aux ressources distribuées reste un bon indicateur de cette dynamique technologique et d'adaptation des acteurs sans l'inertie face aux mutations des outils et systèmes.

Dans la pratique, les possibilités offertes se focalisent à quelques composants logiciels simples mais efficaces, tels que la messagerie, la navigation sur le Web ou encore la communication instantanée sur les réseaux. Tous ces outils exploitent pleinement ces nouveaux moyens de communication et rendent plus accessibles les usages avec les possibilités aisées d'alimenter les contenus informationnels et de disposer des avis des acteurs sur la valeur informationnelle des ressources et contenus.

Dans le contexte professionnel des usages de l'information, les réflexes des acteurs deviennent plus complexes : eux-mêmes, ils dépendent fortement de l'environnement informatique installé, des enjeux évolutives de l'organisation et des besoins (ie. les problèmes décisionnels posés). Par contre, l'évolution des SI suit un rythme beaucoup plus lent, nécessitant bien souvent le recours à des entreprises extérieures pour innover et migrer leur SI existant [1]. A la perte possible de données s'ajoute celle du savoir intrinsèque créé autour de l'application existante. Ce savoir détenu par les acteurs (de l'entreprise) nécessite pour le préserver un accompagnement au changement [2]. Malheureusement, ce savoir implicite est peu connu, peu décrit et rarement intégré [Nourrissier et Sidhom, 2009] nécessite de le capitaliser.

Dans un autre contexte, il existe un nouveau réflexe dans les réseaux sociaux, blogs, forums ouverts, archives ouvertes, etc., où chaque usager a la possibilité de satisfaire ses propres besoins informationnels ou avec d'autres en collectif, s'il appartient à une communauté virtuelle.

Pour décrire l'échange à la fois propre à l'utilisateur et coopératif avec les autres, s'il veut partager son temps avec les autres, cela détermine une forme d'"altruisme" réciproque. Chaque acteur possède un intérêt à capitaliser et à partager son information qu'elle soit ou non décrite dans le contexte de sa création. L'objectif est surtout d'affiner ses connaissances [4] et de tester leurs fiabilités par recoupement [3]. Il reste à savoir ce que l'on souhaite donner au regard de ce que l'on obtient. La prudence est de rigueur, et ce partage s'effectue principalement par étape, afin d'éviter les pièges d'un déséquilibre trop important : le cas de désinformation, de lobbying, de réseaux d'influence ou de l'impact d'un leadership sur la communauté.

Cette problématique se rencontre quotidiennement en entreprise, même si les raisons diffèrent. Au sein d'une même organisation, les acteurs ont un intérêt commun à partager, sans forcément disposer de moyens techniques suffisants qui nécessite d'identifier le profil des acteurs dans une communauté, la hiérarchie et les degrés de confiance (acteur, information).

Dans de nombreuses petites et moyennes entreprises, les supports physiques de stockage résident dans un serveur de fichiers unique, accessible sur le réseau, et contenant l'ensemble des archives de la société (ou mémoire d'entreprise). Dès lors, on peut s'interroger sur le devenir de toutes ces informations collectées sur des supports de plus en plus mobiles, de la clé USB aux ordinateurs et téléphones portables avec mémoire. Chacune de ces unités possède les capacités à créer de l'information que quand elle est intégrée dans le SI fera un ajout au patrimoine informationnel. Nos usages nous poussent à simplifier notre gestion quotidienne de l'information au détriment de sa structuration. Cette tendance très forte impacte notre organisation de travail. Aujourd'hui, nous devons concevoir des systèmes d'information tenant compte de ces supports, usages et les capacités d'intégration de nouvelles données dans le système [29].

L'enjeu est donc d'aider les acteurs dans la gestion quotidienne de leurs informations dans l'environnement de l'organisation : nos réflexions tentent de définir les dimensions impliquées dans cette gestion complexe. Nous identifions les dimensions concernées par : l'information collectée (I : information), l'utilisateur/acteur en activités (U : user) et connaissance représentée (K : knowledge), afin d'améliorer la gestion individuelle et collective, en exploitant au mieux les nouvelles technologies et les projections de ses différentes dimensions pour la décision (D : decision) comme quatrième dimension dans la modélisation. L'objectif à moyen terme est d'établir des méthodes simples de coopération qui guideraient l'activité, les contenus pour permettre leur projection en connaissance. Leurs superpositions de les contextes d'usage déclinent les actions à entreprendre et les décisions adaptées (Fig.I.1).

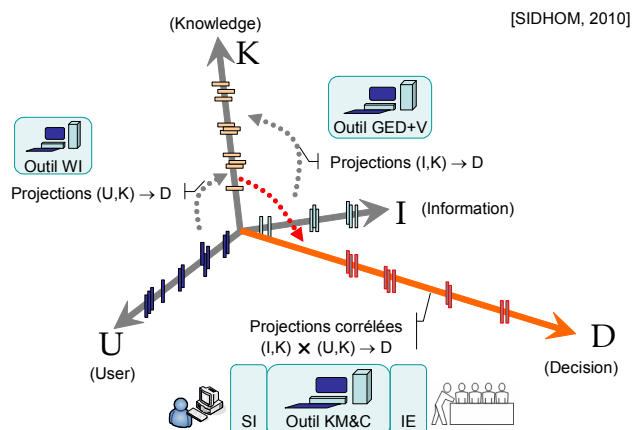


FIG.I.1 : LES ENJEUX EN SIIE : DIMENSIONS D'ETUDE A TRANSFORMER EN PROCESUS [26].

## 2. INTEGRATION PAR LES PROCESSUS : L'IE

Les technologies de l'information et de la communication, en particulier l'informatique et ses applications sur le Web (visible, invisible, social et intelligent), permettent d'intégrer des informations multi-sources et multi-formes. Par exemple, on peut s'intéresser aux informations provenant des bases de données d'entreprise [11], des ressources documentaires (mémoire d'une entreprise, archive, ressources électroniques, etc.) et des médias sociaux. S'agissant des ressources informationnelles : (i) primaires (les productions sources), (ii) secondaires (les notices bibliographiques avec/sans descriptions de contenu), (iii) tertiaires [27] (annotations, informations à valeur ajoutée, avis/votes sur les contenus, archives ouvertes augmentées, réindexation par les usagers), sont désormais disponibles par ces nouvelles technologies. Alors que les informations primaires sont les œuvres directes des producteurs (auteurs), les informations secondaires sont des transformations en « modèles réduits » des informations primaires, pour alimenter des bases documentaires, des banques de données pour GED et des archives électroniques (par des documentalistes, veilleurs et spécialistes de l'information des réseaux professionnels) et les informations tertiaires sont des avis de tout venant sur un contenu pour construire une information à valeur ajoutée [20], [21]. Les informations à valeur ajoutée sont les produits d'analyse et de synthèse de ces différents types d'information avec l'interaction des usagers (anonymes ou identifiés) sur le Web pour apporter des suggestions, pour voter sur la valeur/qualité informationnelle, pour suggérer des préférences (subjectives et/ou objectives) ou pour réindexer le contenu selon un contexte (social, professionnel, culturel, économique) donné. Ainsi, l'information (interne à l'organisation, exogène par les interactions) est de plus en plus utilisée comme un objet de référence à la décision. Sa notation variable (besoin, usage, contexte, pertinence) en fonction du besoin, la rend à valeur stratégique dans l'environnement de l'organisation.

En l'occurrence, l'usage de l'information dans le processus d'IE repose, en particulier, sur l'utilisation de systèmes d'informations stratégiques (SIS). Les systèmes d'informations (SI), en général, ont subi des évolutions profondes soit (i) dans l'objectif d'utilisation-exploitation, (ii) soit dans l'objectif de gestion et de management (par la nature de l'information à gérer dans un contexte donné) et (iii) soit par la combinaison des deux modes (objectif, gestion-management). C'est ainsi que se manifeste la déclinaison du terme SIS en la composition de :

- « système d'information » + « stratégique » : (SI+S) et
- « système » + d'« information stratégique » : (S+IS).

Faciliter le processus de recherche d'informations dans un SI nécessite des techniques de modélisation avec la prise en compte du modèle de l'utilisateur. C'est pour cette raison, des techniques ont été adaptées à la nature des besoins informationnels des utilisateurs. Certains besoins sont stables, qui nécessitent la proposition des techniques de filtrage (automatique, collaboratif, etc.) de l'information. D'autres sont dynamiques, qui demandent l'emploi des techniques de modélisation implicites ou explicites. Ces dernières peuvent être globales où il s'agit de stéréotypage, profilage ou caractéristique à l'individu [15].

La diversité des besoins informationnels de l'utilisateur fait appel à un nouveau type de SI multi-sources (SIMS). Il est constitué, de notre point de vue, d'un ensemble de sources de données hétérogènes [23], [24], d'un ensemble de tâches décrivant les besoins de l'utilisateur, sur ces sources et d'un ensemble de contraintes pour la formulation des besoins et des accès aux données. Il existe de nombreux exemples de SIMS : (i) les systèmes multibases répartis, (ii) les entrepôts de données (datawarehouse DW, ETL, etc.), (iii) les portails Web sur Internet ou Intranet (avec l'association de Web services), (iv) les systèmes d'information embarqués, (v) le Web Intelligent et outils associés.

Les caractéristiques communes à l'ensemble de ces systèmes sont :

- une forte hétérogénéité des sources de données traduisant soit une diversité des technologies disponibles et des plates-formes d'accès, soit une diversité concurrentielle sur les représentations sémantiques des données (DW) ;
- un besoin de vision globale de la structure d'un SI, des règles de correspondance avec les sources qui enrichissent le contenu et des contraintes d'accès aux sources pour des traitements spécifiques (outils KM, Web services, etc.),
- un souci de qualité sur les données disponibles, notamment en ce qui concerne leur cohérence, leur fiabilité, leur interopérabilité et leur complétude.

Une différence importante entre les systèmes d'information (SI) classiques et les systèmes d'information multi-sources (SIMS) réside dans leur approvisionnement-supervision en données. Dans les SI classiques, les mises à jour de la base de données sont effectuées par les utilisateurs via les programmes d'application. Dans les SIMS, les mises à jour de données sont faites par propagation des modifications survenues dans les

sources [28]. De nombreux problèmes caractérisent la conception des SIMS, nous en citons quelques-uns parmi les plus importants :

- l'intégration des données multi-sources, soit au niveau du schéma des traitements, soit au niveau des données, qui nécessite des opérations de transformation et des épurations,
- l'optimisation des requêtes basée sur la matérialisation de vues, qui est considérée comme une technique d'optimisation multi-requêtes,
- la détection des changements dans les sources et leur propagation jusqu'aux vues de l'utilisateur, qui entraîne des mises à jour par des processus d'incrémental et de rafraîchissement des données afin de satisfaire les contraintes de qualité.

Le développement d'un SIMS met en œuvre un grand nombre de composants logiciels pour le Filtrage, pour l'Extraction et pour la Transformation de données ; aussi bien d'outils de médiation entre SGBD, entrepôts de données, outils d'analyse et de veille [16], etc.

Toutes ces problématiques inhérentes aux modèles de SI et associées aux modèles de données laissent penser aux enjeux de l'IE. L'investissement technologique est nécessaire avec la mobilisation des acteurs dans le management de l'information. Ces deux aspects ne peuvent être qu'en cohérence dans la sphère de Management des Connaissances [17] et des Compétences qu'on décline par le KM&C.

### **3. CAPITALISATION DES CONNAISSANCES : PERSPECTIVES DU KM&C PAR LES ACTEURS**

D'un point de vue d'une information tertiaire, une annotation parvenant d'un acteur économique contient essentiellement des informations hétérogènes et elle est conçue dans l'objectif d'être lue, assimilée et partagée de manière aisée, avec tout autre acteur. Sa lecture est perçue par le canal visuel et/ou auditif en relation avec le monde du KM : cela suppose que la perception des concepts sera effective par l'annotation.

C'est ainsi que l'annotation associée à une ressource informationnelle (primaire ou secondaire) peut prendre des formes diverses, comme écrite (texte), orale (séquence audio), graphique (schéma, représentation, images, etc.), vidéo (séquence de vidéoconférence, etc.).

Indépendamment de la forme de présentation d'un contenu d'informations, une annotation prend habituellement une forme complémentaire ou différente de la source. Cette différence reste complémentaire et permet de produire des concepts inhérents à la culture de l'annotateur qui a voulu les introduire en terme de valeurs ajoutées. Les éléments à valeur ajoutée ne peuvent que faciliter la lecture des objets d'annotation, aider à structurer les concepts pour une disposition pertinente des informations sources et celles de l'acteur annotateur (culture, profil, préférences, etc.). C'est un nouveau processus interprétatif à l'usage d'un contenu afin de rendre l'interprétation accessible dans des objectifs

d'appropriation de la valeur informationnelle et de partage dans le travail collaboratif.

Ainsi, chaque objet informationnel peut-être associé à des annotations (ie. de nouveaux objets informationnels ou "informations interprétées") par des mots, des concepts, des ontologies, des éléments terminologiques, des phrases, des liens hypermédia, des mises en forme typographiques, des séquences multimédia. Soient des objets d'annotation qui peuvent avoir des traits homogènes ou hétérogènes avec la source d'information annotée.

Dans un contexte historique, l'idée de Vannevar Bush dans sa communication "As We May Think" (1945), était d'un intérêt fondé sur le travail collaboratif et qui reste d'actualité. L'annotation sera construite sur l'idée d'un travail collaboratif : les documents sont mis en concordance avec un public porteur de nouvelles idées (informations, connaissances, savoirs et savoir-faire) avec l'emploi d'un vocabulaire commun, pour s'exprimer sur des thèmes proches avec des concepts partagés.

Par ce rapprochement sur le mode « collaboratif », l'annotation pouvait s'adresser à l'auteur du document, au lecteur du document, au public intéressé à la fois par le document, ses annotations, dans le but de relever ou partager des informations pertinentes et s'en servir à bon escient comme pour la prise de décision. Ainsi, nous considérons que les annotations rapportées à un contenu informationnel dans un processus (RI, indexation, valorisation, réindexation, veille et décision) vont lever des ambiguïtés sur des éléments d'information, informer sur le véritable contenu sujet de qualité et faciliter l'analyse comme aide à la décision. Il faut également souligner que dans l'activité d'annotation, il y a l'aspect d'enrichissement de contenus par des informations implicites à la source et l'aspect d'accumulation des interprétations propres aux acteurs annotateurs par domaine de spécialité.

En conséquence, il est essentiel dans ce nouveau concept (d'annotation) d'introduire des objets sémantiques pour faciliter le filtrage d'informations pertinentes. On cite quelques éléments essentiels dans les objectifs d'annotation [27] et sa portée sémantique dans un processus de veille ou d'aide à la décision:

- construire une représentation externe au contenu,
- introduire des évaluations sur la valeur informationnelle d'un contenu : témoignage, apport, constat, démonstration, réfutation, etc.,
- permettre une prise de vue globale d'un contenu qui soit indépendante de la présentation de l'auteur,
- fournir une traçabilité sur l'exploitation d'un tel contenu dans tel objectif,
- favoriser le raisonnement et l'évaluation par le contenu,
- orienter l'information vers un groupe d'acteurs,
- filtrer l'information,
- insérer des marqueurs sémantiques sur un contenu et son public cible.

Sur l'aspect conceptuel d'un modèle d'annotation, la démarche consiste à faire le rapprochement entre le processus d'annotation à celui d'indexation pour s'associer au processus KM. Nous avons observé que l'annotation peut servir cette orientation sur trois points complémentaires :

1. Dans le cas d'une indexation automatique, il n'y a pas d'interprétation subjective car les concepts d'indexation sont issus des informations explicites du contenu. Un aspect sur l'insuffisance du processus d'indexation dans le contexte d'IE ou de veille, c'est qu'on ne dispose pas des éléments de l'interprétation sémantique sur le contenu et le point de vue de l'utilisateur lambda sur la valeur informationnelle ;
2. Dans le cas d'une indexation intellectuelle, souvent le processus d'indexation atteint une forme de stabilité interprétative car l'indexeur humain introduit une part de son interprétation sur le contenu et un aspect complémentaire sur la valeur informationnelle sous forme d'une valeur ajoutée par rapport à la source. Ceci conduit l'indexeur humain qui peut avoir la fonction de veilleur, à définir des grilles d'analyse ou des règles d'indexation « normalisées » par des contraintes. Ce qui conduira en résultat à des descriptions avec des informations pertinentes stables sur le contenu ;
3. Comme prévision au processus d'annotation, l'indexation automatique et/ou intellectuelle d'un contenu sont indispensables pour expliciter les relations sémantiques existantes, leurs compositions et structures, afin de traduire les éléments informationnels existants en éléments conceptuels pour les projeter en éléments de connaissance. Cela consiste à expliciter les attributs et les valeurs dans les deux logiques (d'indexation, d'interprétation et d'annotation) pour étendre les fonctionnalités de la connaissance représentée dans la logique KM.

C'est à ce propos que nous allons déterminer les conditions de transition d'un processus d'annotation, dans l'architecture de capitalisation des connaissances en IE, en faisant appel aux concepts KM, comme suit :

- la formalisation : les informations annotées sont amenées à être structurées complètement (dans un SI ou un processus d'IE) afin d'identifier les attributs et valeurs de son contenu. Ce qui permettra dans l'objectif de valorisation de déterminer : (i) la pertinence des informations associées aux annotations, (ii) les réponses d'un système sur un besoin utilisateur et son degré de satisfaction. En l'occurrence, la pertinence d'une requête et ses réponses font parties des facteurs clés de succès d'un système RI quand il s'agit d'un produit de veille ou d'IE ;
- l'explicitation : une annotation ne suffit pas à elle-même. Elle est souvent destinée à un ou plusieurs acteurs d'une organisation. Donc, elle nécessite des adaptations (la levée des ambiguïtés) pour son usage et

ses usagers : des précisions sur la langue source, la liste de mnémonique utilisée, des indicateurs, des conventions, une table associative (objet, structure, valeur, ...);

- la traduction : au niveau de la structuration d'une information ou une annotation, elle doit intégrer les propriétés de son rôle à jouer dans la communication. En l'occurrence, l'annotateur qui est le producteur d'une annotation, peut décliner la visée de son annotation aux prospecteurs potentiels (les acteurs à cibler par l'annotation). Ces derniers, dans le contexte d'IE, est un acteur de l'organisation (veilleur, analyste, infomédiaire ou décideur) qui s'outille dans sa démarche pour capitaliser l'information pertinente dans une logique "triptyque" reliant l'information, les annotations associées et le champ opératoire (ie. des connaissances induites par le raisonnement). Dans cette considération, un agent logiciel (ie. plate-forme d'outils informatiques, SI, outils GED ou veille) est nécessaire pour construire ce travail incrémental dans une architecture fonctionnelle de KM.

En synthèse, la complexité du sujet traité, sur les " *systèmes d'information* " et l' " *intelligence économique* ", nous a fourni un cadre d'analyse sur des problématiques liées sur le contenu informationnel, en évolution avec les modalités de la gestion de contenus : modélisation de l'acteur dans un processus, modélisation de l'information et sa valeur ajoutée dans la prise de décision (par le processus d'annotation) et les outils d'analyse (processus d'analyse, d'indexation et de capitalisation) pour construire une architecture fonctionnelle de KM.

Dans ce travail, les enjeux et les perspectives dans SIIE nous permettent d'évoluer vers l'expérimentation. Il est question de faire fonctionner les processus en architecture fonctionnelle par l'apport de chaque modèle dans un processus. L'accessibilité aux contenus pour retrouver l'information pertinente [GUARINO & WELTY, 2000a-b] et satisfaire une demande stratégique d'un décideur reste un véritable défi. Plusieurs aspects fondamentaux sont à prendre en considération comme les modélisations à mettre en œuvre pour l'analyse, l'extraction et la validation des résultats.

Ces questions nécessitent encore des efforts d'étude et la continuité dans les propositions pour apporter des solutions tangibles dont les retombées techniques et économiques sont considérables [14].

Dans ce qui suit, nous faisons le lien par une étude pour valider certaines de nos hypothèses de réflexion. C'est une avec un retour d'expérience dans le contexte de SIIE pour présenter un système d'information pour l'aide à la décision concernant la gestion des maladies chroniques sous l'initiative du Haut Conseil de la santé publique en France.

## II. CAS D'APPLICATION : EXTRACTION DES CONNAISSANCES ET VISUALISATION DANS LE PROJET « CHRONISANTE »

Franchise, parcours de soins, maladies de longue durée.... Devant le déficit de l'assurance maladie qui ne peut être juguler, les gouvernements tentent de trouver les meilleures solutions devant un dilemme : maîtriser l'augmentation des dépenses de santé tout en préservant la pérennité d'un système de soin efficace. Pour la France, le système de santé est fondé dans son caractère universel. Au fil du temps, de nouvelles questions qui sous-tendent cette problématique sont venues se greffer à une logique purement financière :

(i) Pourquoi les dépenses de santé augmentent-elles partout dans le monde ? Et, (ii) pourquoi plus vite dans certains pays ?

Les réformes proposées par les différents pays concernés par (i) et (ii) doivent cependant répondre à quatre objectifs :

- 1- assurer la viabilité financière des systèmes de santé,
- 2- permettre l'égalité de l'accès aux soins,
- 3- la qualité de ceux-ci (ie. soins), et enfin
- 4- la liberté et le confort des patients et des professionnels dans le système.

En France, plus de 15 millions de personnes sont atteintes de maladies chroniques qui, par leur caractère durable et évolutif, engendrent des incapacités, des difficultés personnelles, professionnelles et sociales importantes.

De nombreuses initiatives ont déjà permis d'améliorer la prévention, le traitement et la prise en charge de ces pathologies liées aux maladies chroniques.

En exemple, en matière thérapeutique, des avancées majeures ont été enregistrées dans le domaine au cours de la dernière décennie pour des pathologies telles que : le diabète, le cancer, le SIDA, etc.

En 2007, le Gouvernement français a mis en place un ambitieux plan sur l'amélioration de la qualité de vie des personnes atteintes de maladies chroniques qui sera décliné jusqu'en 2011. En application de ce plan, le projet de loi de financement de la sécurité sociale 2007 dispose que les caisses nationales d'assurance maladie peuvent mettre en place des programmes d'accompagnement des patients atteints de maladies chroniques visant à leur apporter des conseils en termes d'orientation dans le système de soins et d'éducation à la santé [1].

Pour la plupart des acteurs du système de santé et en premier lieu pour les patients, la prise en charge des maladies chroniques se résume au dispositif médico-administratif des affections de longue durée (ALD)<sup>1</sup>. Or les ALD, simples ou exonérantes, sont presque toujours des maladies chroniques.

C'est dans ce contexte qu'a été pensé et réalisé « *ChroniSanté* », un système d'information d'aide à la décision, dans le cadre d'un travail de Master en IST et Intelligence Economique, regroupant des facettes R&D dans le contexte professionnel de l'INIST en France (ie. Institut de

<sup>1</sup> Les ALD sont des affections comportant un traitement prolongé et une thérapeutique particulièrement coûteuse (Définition de la Haute Autorité de Santé).



l'Information Scientifique et Technique - Unité du CNRS pour la fourniture de l'information spécialisée).

Le système *ChroniSanté* avait pour objectif d'aider un groupe de travail pluridisciplinaire sur les maladies chroniques du Haut Conseil de la santé Publique (HCSP) à émettre une série de recommandations sur la réforme du système de soins en France. Ce projet a été élaboré en partenariat entre le HCSP et l'INIST.

Dans ses spécifications applicatives, le système a été doté de deux modules distincts : – une application de type Gestion de Contenu ou CMS pour « Content Management System » permettant la gestion de documents multi-formats et – une autre application de type veille bibliographique sur les maladies chroniques, leurs pathologies et leurs soins. Le projet *ChroniSanté* est consultable librement sur l'Internet (<http://hcsdp.demo.inist.fr>) avec la possibilité pour l'utilisateur de procéder à un étiquetage spécifique (ie. insertion de nouveaux tags) dans chaque ressource consultée [2].

Nous présenterons dans une première partie la méthodologie et les outils mis en oeuvre pour ce travail. De part la commande initiale du HCSP, le projet *ChroniSanté* s'inscrit dans une logique d'élaboration de recommandations stratégiques. La méthodologie d'intelligence économique a été privilégiée débouchant sur une pratique de veille informationnelle. Cette démarche a nécessité l'élaboration d'un outil spécifique à notre thématique.

La deuxième partie fera état des résultats obtenus après l'analyse des notices bibliographiques par un ensemble de transducteurs à état fini puis de leur rendu visuel pour l'émergence de nouvelles connaissances.

Enfin, la dernière partie sera consacrée à la discussion de ces résultats.

## 1. METHODOLOGIE ET OUTILS

### A. Méthodologie appliquée : processus d'intelligence économique (IE)

La question émise par le HCSP consistait à définir comment les systèmes de santé européens gèrent-ils le problème des maladies chroniques.

L'approche employée pour cette problématique s'appuie sur le modèle WISP (ie. Watcher Information and Search Problem) développé par Philippe Kislin (2007) [3]. Ce modèle est l'extension d'une approche IE pour décrire les besoins informationnels et aider l'utilisateur à les formuler.

Dans le contexte de ce travail, la formulation des besoins a été orientée au regard des références bibliographiques obtenues après consultation des bases de données métiers (cf. B.4). Dans l'approche WISP, la notion de point de vue de l'utilisateur permet d'intégrer quatre dimensions du problème :

1°/ la première dimension est analytique, elle correspond à la compréhension d'un triplet « Demande, Enjeu et Contexte » à la définition d' « Indicateurs » avec leurs propriétés et leurs évolutions ;

2°/ la deuxième dimension est méthodologique, elle est constituée en, un premier niveau, par la traduction du problème décisionnel en problème(s) de recherche d'informations et en, un second niveau, par les moyens

d'identification de l'information et d'acquisition des connaissances sur le problème posé ;

3°/ la troisième dimension est opérationnelle, elle correspond à la sélection des plans d'actions et à la mise en place des différentes étapes pour la résolution du problème décisionnel.

Les trois dimensions permettent une caractérisation de l'objectif (ie. la résolution d'un problème d'IE) en adéquation avec l'expression du besoin (ie. la demande formulée).

L'enjeu stratégique de ce travail, du côté veilleur, consiste à le formaliser par l'équation :

<<Enjeu>> = si on n'agit pas sur l'<OBJET> sachant l'état du <SIGNAL>, alors le risque est l'<HYPOTHESE> attendue. Où –l'enjeu est défini par un OBJET de l'environnement sur lequel il est possible d'agir, –un SIGNAL qui incite le décideur à déclencher le problème et –une HYPOTHESE qui correspond au risque encouru, comme conséquences attendues, si on ne réagit pas.

L'application de cette approche, dans l'optique de cibler au mieux les besoins informationnels du commanditaire du projet, le HCSP, a permis de traduire l'enjeu stratégique en une série de dimensions liée au problème avec un ensemble d'indicateurs de recherche d'information (RI) :

Le couple (Enjeu, RI) a également donné sa structure au système d'information dans lequel les dimensions ont pris le rôle de catégories thématiques (TAB.II.1).

Dimensions	Indicateurs de Recherche d'information
"Prise en charge des patients atteints de maladie chronique"	"Maladies chroniques"
<ul style="list-style-type: none"> <li> → Dimension sociale</li> <li> → intégration sociale</li> <li> → perception sociale du malade</li> <li> → inégalités sociales</li> <li> → dimension médico-psychologique</li> <li> → qualité des soins</li> <li> → accompagnement du malade</li> <li> → relations soignant /soigné</li> <li>    → Thérapeutique</li> <li> → éducation thérapeutique</li> <li> → dimension politique</li> <li> → la législation</li> <li> → Protection sociale</li> <li> → Financement assurance maladie</li> <li> → offre du panier de soins</li> <li> → harmonisation européenne</li> <li> → dimension innovation</li> <li>    → innovations technologiques</li> <li>     → Automesures</li> <li>     → dossier médical personnel</li> <li>     → Télémédecine</li> <li>     → innovations biomédicales</li> <li> → dimension prospective</li> <li>     → Modèle de prise en charge des MC</li> </ul>	<ul style="list-style-type: none"> <li> → définitions</li> <li> → critères actuels</li> <li> → synonymes en langues européennes</li> <li><b>Système de santé</b></li> <li> → organisation interne du pays cible</li> <li> → législation en place</li> <li><b>Pays européens de l'étude</b></li> <li> → Allemagne</li> <li> → Belgique</li> <li> → Danemark</li> <li> → Espagne</li> <li> → Italie</li> <li> → Pays-Bas</li> <li> → Royaume-Uni</li> <li>(...)</li> </ul>
(...)	

Tab.II.1 : INDICATEURS METHODOLOGIQUES.

## B. Processus d'extraction d'informations

### B.1. NooJ : environnement linguistique pour le traitement du langage naturel

NooJ est un environnement linguistique développé par Max Silberztein (2005) [4] de l'université de Franche-Comté (France). NooJ est fondé sur la technologie .NET et reconnaît un grand nombre de formats de documents. Outre cet avantage, l'utilisation de l'outil est facile avec une prise en main relativement rapide. Dans l'optique d'un repérage terminologique pour notre corpus bibliographique, NooJ offre la possibilité de créer des grammaires locales (ie. automates finis) complètement paramétrables pour l'extraction d'informations. Les ressources de NooJ sont principalement constituées de dictionnaires (de la langue) et de graphes syntaxiques de type transducteurs à état fini, permettant le repérage d'expressions complexes, l'extraction de lemmes et l'annotation automatique de ressources textuelles.

### B.2. Stratégie d'analyse morpho-syntaxique du corpus

Dans ses travaux de recherche, Sahbi Sidhom (2002) a élaboré une plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'informations [5]. Elle est composée d'un noyau d'indexation (ie. processus d'indexation) qui utilise le modèle des syntagmes nominaux comme descripteurs (ie. concept d'indexation) de l'information textuelle.

Pour reprendre la définition d'un syntagme nominal (SN) d'après Michel Le Guern (1989) [6], placer le mot du lexique dans un univers de discours qui, de facto, le place dans une logique extensionnelle, donne au SN un statut référentiel, segment de la réalité qui lui est associée.

Dans notre contexte, le SN se révèle être porteur d'une charge sémantique qui fait de lui un élément central et pertinent pour l'analyse des informations bibliographiques. C'est vers cette sémantique cherchée qu'on oriente nos analyses de corpus.

Ainsi, la grammaire de reconnaissance du SN s'articule autour de trois niveaux logiques présentés par le schéma (Fig.II.1).

Ces trois niveaux logiques sont :

1°/ le *niveau intensionnel* (ie. propriétés de la langue) est représenté par le niveau **N** : les unités considérées sont des prédicats libres simples (ie. les propriétés du nom) ou complexes (ie. les propriétés du nom modifiées par d'autres éléments : adjectivaux A', des expansions prépositionnelles EP, etc.) ;

2°/ le niveau intermédiaire ou niveau **N'** (ie. la prise en compte de l'univers du discours considéré) est la transition de l'intensionnel vers l'extensionnel ;

3°/ le niveau extensionnel ou niveau **N''** (ie. le syntagme nominal et sa complexité) est l'opération de fermeture au moyen d'un quantificateur qui sélectionne un élément précis dans la classe N des nominaux. Ce sont les objets du monde existants, référés ou construits par la pensée.

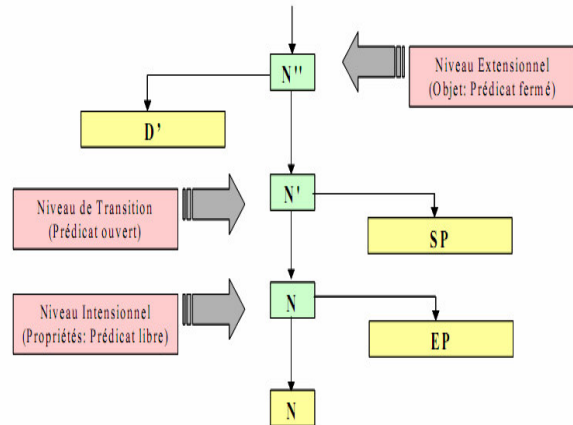


FIG.II.1 : NIVEAUX LOGIQUES DU SN.

Dans ce travail, cette grammaire morpho-syntaxique du SN a été réécrite pour NooJ en deux étapes : –dans un premier temps, le travail a consisté à reformater les ressources linguistiques (dictionnaires et grammaires) en notre possession puis, –dans un second temps, nous avons élaboré le transducteur à état fini du syntagme nominal. Les étiquettes existantes des dictionnaires ont été harmonisées pour correspondre au graphe syntaxique du SN. Celui-ci se compose d'un ensemble de cinq graphes (FIG. II.2) reprenant la structure présentée en FIG. II.1

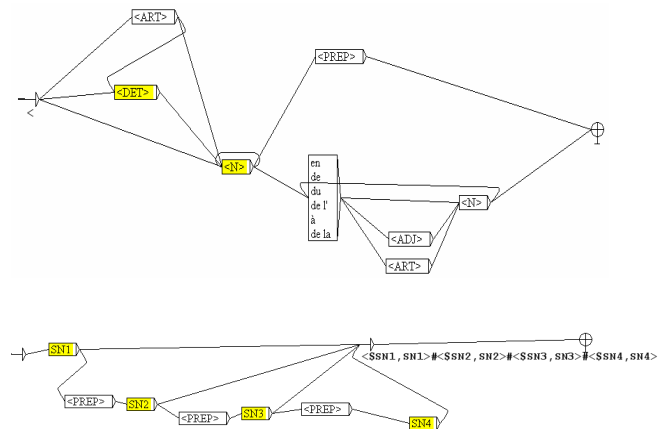


FIG.II.2 : EXEMPLE DE GRAPHES SYNTAXIQUES DANS NOOJ.

### B.3. NODEXL : outil de cartographie de l'information

Un des objectifs de notre travail est de pouvoir proposer des cartes sémantiques permettant de visualiser des espaces informationnels venant en appui à notre démarche d'intelligence économique et à la conception du système d'information *ChroniSanté*.

Sur l'aspect de la visualisation, la méthodologie présentée dans l'ouvrage de S. K. Card, J. D. Mackinlay et B. Shneiderman (1999), intitulé "Readings in information visualization : using vision to think" [7] offre un fondement de réflexions dans une logique de rendu visuel des informations traitées à destination des utilisateurs.

Dans ce recueil d'articles, les auteurs [7] définissent la visualisation de l'information comme « *l'utilisation de représentations visuelles interactives et informatisées de données pour amplifier la cognition* ». Le modèle de référence proposé, qui a rapidement fait autorité, présente les principes de transition des données aux formes visuelles dans l'espace de représentation (TAB. II.2).

Données		Formes visuelles
(Données brutes) → (Tables de données)	⇒	(Structures visuelles) → (Vues)

TAB. II.2 : TRANSITION DONNEES - STRUCTURE VISUELLE.

En particulier, dans l'activité de veille informationnelle, ce procédé constitue un vecteur principal d'émergence des liens significatifs après les phases de collecte, de traitement et d'analyse dans une masse importante de données et d'informations.

L'outil que nous avons utilisé est un modèle de document Excel baptisé « NodeXL » développé par Microsoft (<http://nodexl.codeplex.com>). Ce choix a été fait principalement pour la gratuité de l'outil, sa performance graphique et son intégration logicielle.

#### B.4. Application

##### 1° Collecte des données

L'INIST met à disposition des chercheurs du CNRS un portail baptisé « *Bibliovie* » qui regroupe divers services et des ressources documentaires libres ou payantes dédiés à l'information scientifique et technique (IST).

Dans le cadre de la construction de notre corpus, nous avons principalement effectué des recherches bibliographiques multi-bases via l'application « *Webspir* », un outil qui a été remplacé début 2009 par la plateforme « *OvidSP* » (<http://www.ovid.com>) avec des fonctionnalités quasi-équivalentes mais plus robustes pour les utilisateurs.

Trois bases de données ont été choisies pour la constitution de notre corpus de notices bibliographiques : « *Pascal*<sup>2</sup> », « *PsycInfo*<sup>3</sup> » et « *Medline*<sup>4</sup> ».

<sup>2</sup> *Pascal* : Produite par l'INIST-CNRS, PASCAL® est une base de données internationale et multidisciplinaire qui recense la littérature en sciences, technologies et médecine. Les publications françaises et européennes y sont particulièrement bien représentées (45% du fonds) ce qui en fait un précieux complément des bases de données d'origine américaine

<sup>3</sup> *PsycInfo* : Base de données de l'American Psychological Association (APA) donnant accès à des articles de périodiques (plusieurs étant en texte intégral), des chapitres de livres et des livres, des rapports de recherche et des thèses et mémoires en psychologie et domaines connexes (médecine, soins infirmiers, sociologie, etc.) du 19e siècle à nos jours. Elle intègre une collection en texte intégral d'environ 550 périodiques avec comité de lecture. Ses références bibliographiques proviennent de plus de 2 400 périodiques.

<sup>4</sup> *Medline* : Medline est une base de données bibliographiques produite par la National Library of Medicine (NLM-USA). Elle couvre tous les domaines biomédicaux : biochimie, biologie, médecine clinique, économie, éthique, odontologie, pharmacologie, psychiatrie, santé publique, toxicologie, médecine vétérinaire. Medline indexe les références bibliographiques ainsi que les résumés de 4 800 journaux biomédicaux publiés aux Etats-Unis et dans 70 autres pays. La plupart des références sont en anglais ou possèdent un résumé en anglais. La mise à jour est quotidienne. Medline ne couvre pas la

Le choix d'utiliser ces trois sources d'information se justifie par notre souhait de couvrir le plus complètement possible la thématique sur la prise en charge des maladies chroniques. La base *Pascal* a l'avantage de présenter des notices européennes et comporte des notices issues de la Banque de données en santé publique (BDSP). La base *Medline* est centrée autour des publications américaines, tout comme *PsycInfo* avec cependant des thématiques élargies au domaine des sciences sociales.

Contrairement à une logique de synthèse qui demande une exhaustivité thématique, nos équations de recherche ont été élaborées pour donner des résultats les plus larges possibles afin, d'une part, de couvrir l'ensemble des sous-thématiques des « *chronic disease* » et, d'autre part, de pouvoir identifier des sous-thématiques nouvelles auxquelles nous n'avions pas pensé initialement.

Le butinage sur les trois bases de données a rapporté 2097 notices pour *Pascal*, 6110 pour *Medline* et 2177 notices pour *PsycInfo*. Nous avons ensuite raffiné nos résultats pour ne sélectionner que celles produites entre 2001 et 2009 et de langue française. L'échantillon obtenu se compose de 397 notices et qui passe à 303 une fois dédoublonné. Ce corpus de synthèse constituera enfin la base de notre travail. Ces premiers résultats au niveau de la RI montre que les « *chronic disease* » est un concept nouveau en France du fait de la singularité du modèle du système de santé français.

##### 2° Analyse du corpus

Une première phase de préparation des données bibliographiques a consisté à vérifier l'orthographe et à apporter les modifications nécessaires aux traitements par NooJ. En effet, la quasi-totalité des notices comportaient soit des voyelles non accentuées, soit des erreurs orthographiques, qui ont nécessité leurs corrections.

Pour le moment, ce traitement est effectué à la main en attendant l'intégration d'un outil adapté (ie. un parseur morphologique à implémenter sous NooJ).

Une deuxième phase de préparation des sources bibliographiques a consisté à vérifier la granularité des notices dans le corpus et à apporter les modifications nécessaires. NooJ autorise l'import de données textuelles multi-formats mais impose la délimitation des unités textuelles (ie. Text Units) en indiquant un caractère frontière. Dans le corpus des notices ce caractère frontière est absent. Ainsi, nous avons été amenés à créer des fichiers séparés de notices. Chaque notice bénéficiant ainsi d'un code unique pour coupler le lemme à sa source.

## 2. RESULTATS

Lors de l'analyse automatique sur NooJ, le graphe

---

littérature médicale du monde entier, et les monographies et résumés de congrès ne sont pas indexés.

complexe du syntagme nominal a rapporté 637 formes incluant la forme la plus petite extraite (ie. le lemme N) jusqu'au syntagme nominal complexe (ie. N+N'+N'').

L'intérêt de cette approche consiste à retrouver des concepts auxquels le veilleur ne pense pas forcément lors de sa recherche d'indicateurs, c'est-à-dire : *la phase de traduction d'un problème décisionnel en problème(s) de recherche d'informations*, dans le contexte d'IE.

En l'occurrence, si l'on considère le concept « patient » qui est central pour notre thématique. Dans la pratique, nous avons tendance à établir notre recherche d'indicateurs dans une **acceptation passive** avec les concepts : « suivi du patient », « prise en charge du patient », « éducation du patient », etc. Mais, non dans une **acceptation active**, comme : « implication du patient », « participation active du patient », etc.

Egalement, la recherche de syntagmes nominaux dans les titres des notices a mis en exergue des notions qui, apparemment n'ont pas de rapport étroit avec notre thématique, mais qui, pourtant, reviennent plusieurs fois dans différentes notices. En l'occurrence, sur la thématique de la « consommation de cannabis », celle-ci met un lien pour la déclaration des maladies de longue durée.

Devant ces constats, nous avons pris parti de sélectionner les concordances les plus longues dans les SN. Ceci correspond à l'emboîtement des concepts : du concept emboîtant, le plus long et le plus informatif, au concept emboîté, le plus court et le moins précis. Cette richesse sémantique, qui en ressort, permet l'identification d'une collection informationnelle de concepts, pertinents, complexes et hiérarchiques. Ces caractéristiques risquent de passer inaperçu dans l'analyse linéaire d'un corpus.

Ainsi, le processus de traitement du corpus de notices bibliographiques sur la plateforme NooJ a démontré des résultats satisfaisants en se fondant sur les SN et leurs relations sémantiques.

Concernant la visualisation des informations, nous avons testé l'application NodeXL sur notre base de notices bibliographique puis sur les résultats d'extraction par NooJ (FIG. II.3 et FIG. II.4).

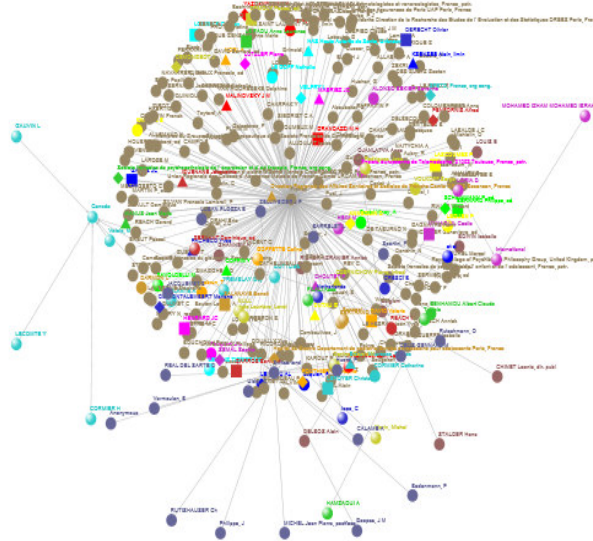


FIG. II.3 : RESEAU SEMANTIQUE DE LA BASE DE NOTICES BIBLIOGRAPHIQUES

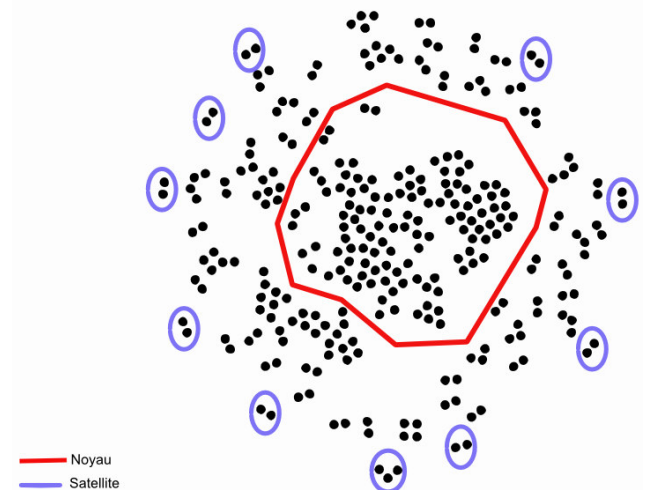


FIG. II.4 : DECOUVERTE DE NOUVEAUX CONCEPTS : LES CONCEPTS DU NOYAU ET CEUX EN SATELLITE.

Sur l'analyse du réseau sémantique SN, on observe que le terme de « prise en charge » se démarque largement de l'ensemble des concepts du corpus. Cette prépondérance du terme est à mettre en corrélation avec les recherches effectuées sur les bases bibliographiques. Une des fonctions de NodeXL permet de se rapprocher des différents noeuds du graphe afin d'affiner la visualisation.

L'utilité de ce type de document ou « structure visuelle », pour un veilleur, un expert ou un décideur, n'est pas à démontrer. Il permet au veilleur de présenter un document interactif susceptible de faire émerger de nouvelles connaissances. *In extenso*, une telle logique sémantique permet à mieux cerner la complexité des dimensions à prendre en compte dans un processus de veille ou d'IE.

### 3. DISCUSSION

Rappelons-le cependant, notre objectif n'était pas d'effectuer un traitement statistique de nos données bibliographiques mais de cartographier les unités sémantiques les plus représentatives pour notre projet.

Cette expérience montre que les techniques que nous avons utilisées demandent une automatisation pour arriver à un état de performance, de robustesse et d'efficacité acceptable.

Pour le processus de veille, la visualisation des informations extraites à partir des syntagmes nominaux s'avère utile aux acteurs du projet *ChroniSanté* sur plusieurs aspects :

Premièrement, la visualisation permet de faciliter l'indexation de documents pour un système d'information d'aide à la décision. En exemple, pour un article à traiter, l'extraction des syntagmes nominaux dans son contenu ou dans son titre peut être transformé en étiquettes (tags). Cette solution permet de présenter à l'utilisateur les concepts essentiels des ressources dans la base d'information. Cela peut également encourager une logique de réindexation par les usagers : aux étiquettes automatiquement créées, l'utilisateur y ajoutera des étiquettes subjectives créatrices, à leur tour, de valeur ajoutée [8].

Deuxièmement, la visualisation permet la production de nouvelles connaissances. Les noeuds d'un réseau sémantique visualisé peuvent en effet faire l'objet d'une analyse dans un groupe de travail pour l'identification de nouvelles thématiques en relation avec la veille stratégique selon les couvertures, les convergences et les divergences des concepts représentés. C'est, pour reprendre la logique de Humbert Lesca, un processus heuristique permettant une création collective de sens [9].

En dernier, si l'on considère les fonctionnalités de NooJ et les ressources associées (ie. dictionnaires, grammaires et méta-grammaires) via une communauté de professionnels et/ou de chercheurs, on peut envisager de greffer, dans cette logique d'analyse et de visualisation, une nouvelle étape de traduction de concepts (ex. des syntagmes nominaux extraits) dans un contexte multilingue. Cette phase correspondrait à une nouvelle approche de veille multilingue qui se développe de plus en plus dans les cellules de veille stratégique des grands groupes nationaux et internationaux.

Ce dernier aspect concerne l'évolution d'un processus d'IE dans un environnement économique en mutation et impliquant des acteurs issus de différents milieux et cultures.

#### CONCLUSION GENERALE

Poser « une stratégie économique » cela nécessite de disposer d'un maximum d'informations pertinentes, recueillies et produites dans un contexte légal. Le rôle de l'intelligence économique (IE) est donc de fournir aux décideurs des informations pertinentes synthétisées et orientées. Ces informations constituent une source indispensable pour prendre la bonne décision permettant d'engager la pérennité de l'organisation.

Par anticipation dans ce contexte, l'intelligence économique est à la fois un terrain de réflexions, une pratique avec des retours d'expérience, qui nécessite une approche inter- et pluri- disciplinaire pour être correctement appréhendée, avec un investissement de haut niveau par les acteurs scientifiques, économiques et socio-politiques. L'ensemble doit être structuré en processus communicants, avec les acteurs, les ressources et les outils, notamment les systèmes d'information (SI), pour permettre de convertir les données d'une organisation en informations décisionnelles.

En synthèse, la complexité du sujet traité, sur les « systèmes d'information » et l' « intelligence économique », nous a fourni un cadre d'analyse sur des problématiques liées sur le contenu informationnel, en évolution avec les modalités de la gestion de contenus : modélisation de l'acteur dans un processus, modélisation de l'information et sa valeur ajoutée dans la prise de décision (par le processus d'annotation) et les outils d'analyse (processus d'analyse, d'indexation et de capitalisation) pour construire une architecture fonctionnelle de KM.

Comme étude de cas et retour d'expérience dans le contexte de SIIE, nous avons présenté un système d'information pour l'aide à la décision dans la gestion des maladies chroniques sous l'initiative du Haut Conseil de la santé publique en France.

Cette expérience montre que les techniques que nous avons utilisées demandent une automatisation pour arriver à un état de performance, de robustesse et d'efficacité acceptable.

Pour le processus de veille, la visualisation des informations extraites à partir des syntagmes nominaux s'avère utile aux acteurs du projet *ChroniSanté*.

#### BIBLIOGRAPHIE

- [1] Ministère de la Santé et des Solidarités (2007). Plan pour l'amélioration de la qualité de vie des personnes atteintes de maladies chroniques 2007 - 2011. Projet Min. 2007.
- [2] P. Lambert (2009). *ChroniSanté: Un système d'information d'aide à la décision Pour la prise en charge des patients atteints de maladie chronique*. Rapport d'activités INIST-CNRS France. juillet 2009.
- [3] P. Kislin (2007). *Modélisation du problème informationnel du veilleur dans la démarche d'Intelligence Économique*. Thèse de doctorat, Université de Nancy2. 2007.
- [4] M. Silberstein (2005). *NooJ: a linguistic annotation system for corpus processing*. Proceedings of HLT/EMNLP on Interactive Demonstrations. 2005, p. 11.
- [5] S. Sidhom (2002). *Plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information: de l'écrit vers la gestion des connaissances*. Thèse de doctorat, Université Claude Bernard-Lyon1, 2002.
- [6] M. Le Guern (1989). *Sur les relations entre terminologie*



- et lexicque. in *Meta*, vol. 34, 1989.
- [7] S. Card, J. Mackinlay, et B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*. 1999, Morgan Kaufmann (Eds.).
- [8] A. Harbaoui, M. Ghenima, S. Sidhom (2009). Enrichissement des contenus par la réindexation des usagers : un état de l'art sur la problématique. in *Proceedings SIIE'2009 international conference*, Feb. 2009, Hammamet Tunisia.
- [9] H. Lesca, S. Kriaa-Medhaffer, A. Casagrande (2009). Veille stratégique: Un Facteur d'échec paradoxal largement avéré: la surinformation causée par l'Internet. Cas concrets, retours d'expérience et piste de solutions. in *Proceedings SIIE'2009 international conference*, Feb. 2009, Hammamet Tunisia.
- [10] BAUMARD P. (1991). *Stratégie et surveillance des environnements concurrentiels*, Eds. Masson.
- [11] BLOCH A. (1999). *Intelligence Economique*, *Economica*, Collection Poche Gestion, 1997; deuxième édition.
- [12] Bounfour A. (2000). "Gestion de la connaissance et systèmes d'incitation : entre théorie du "hau" et théorie du "ba"", *Systèmes d'Information et Management*, 5 (2), pp.7-40.
- [13] Bournois F. et Romani P-J. (2000). *l'intelligence économique et stratégique dans les entreprises Françaises*", Edition *Economica*, Paris.
- [14] Lambert P., S. Sidhom, (2010). Knowledge extraction and visualization: case study on ChroniSanté project in France. in *Proceedings SIIE'2010 international conference*, Feb. 2010, (Sousse) Tunisia.
- [15] CIGREF. (2005). *Intelligence économique et stratégique : les systèmes d'information au coeur de la démarche*.
- [16] DOU H. (1995). *Veille technologique et compétitivité – l'intelligence économique au service du développement industriel*, Dunod, Paris.
- [17] Grundstein, M. (2005). *Pilotage d'un Projet Global de Knowledge Management (PGKM). Rapport de Recherche #12*. Retrieved January 25, 2005 from <http://www.mgconseil.fr>
- [18] Lesca H. (1994). "Veille stratégique pour le management stratégique, état de la question et axes de recherche", *Economies et Société*, Série SG, N°20.
- [19] Lévét J-L. (2001). *Intelligence économique : mode de pensée, mode d'action*, *Economica*, Paris.
- [20] LEVET, J-L. et PATUREL, R. (1996). L'intégration de la démarche d'intelligence économique dans le management stratégique , *Vème Conférences internationale de Management stratégique*, 13-15-15 Mai, Lille.
- [21] Martre H. (1994). *Intelligence économique et stratégie des entreprises*, *Oeuvre Collective du Commissariat au Plan*, Paris, la documentation Française.
- [22] OUBRICH, M. (2003). *Processus d'intelligence économique : Transformer l'information en connaissance*, Colloque AIM.
- [23] Pentland B.T. (1999). "Building process theory with narrative: From description to explanation", *Academy of Management Review*, 24 (4), pp.711-724.
- [24] QUEYRAS, J. (2005). *L'intelligence économique territoriale dans un centre d'information du service public*, thèse de doctorat en Sciences de l'information et de la communication de l'université du sud Toulon-Var.
- [25] SALLES, M. (2000). (avec la collaboration de Zid, T. et Hussein, S.). *Projet MEDESIIIE (tâche WP11) : Rapport de synthèse sur les outils conceptuels et modèles à retenir*. Université de Toulouse I.
- [26] S. Sidhom (2010). *Projet de recherche et valorisations. Rapport interne de synthèse sur la "Réindexation des contenus par les usages" : outils conceptuels et modèles*. LORIA -Equipe KIWI.
- [27] S. Sidhom, A. David. (2006). Automatic indexing of multimedia documents as a starting point to annotation process. In *ISKO : 9th International Conference 2/ISSN 0938-5495* (2006).
- [28] A. David, S. Sidhom. (2005). *Intégration de la démarche d'Intelligence Économique dans l'architecture fonctionnelle d'un système d'information* In *Conférence Nationale sur Le Système National d'Information Economique : Etat et perspectives*. CERIST-Alger (2005) [Conférenciers invités].
- [29] P. Nourrissier, S. Sidhom, A. (2009). *Le travail coopératif comme vecteur d'évolution de nos systèmes d'information*. in *Proceedings SIIE'2009 international conference*, Feb. 2009, Hammamet Tunisia.

#### Auteurs —

Dr. **Sahbi SIDHOM** est Maître de Conférences à Nancy Université (Nancy2) en France et Chercheur permanent au laboratoire LORIA en informatique et ses applications (équipe de recherche KIWI: Knowledge, Information and Web Intelligence). Ces principales recherches sont dans le domaine du traitement automatique des langues naturelles, la représentation des connaissances (indexation de contenus, systèmes de recherche d'informations, réindexation dans le Web social), la gestion et le management des informations des connaissances et des compétences dans un système d'intelligence économique.

Dr. **Malek GHENIMA** est Directeur de l'Ecole supérieure de commerce électronique ([www.escem.rnu.tn](http://www.escem.rnu.tn)) à l'Université de la Manouba à Tunis (Tunisie). Ces principales recherches sont dans le domaine du traitement automatique des langues naturelles (la langue arabe) et le management des informations, des connaissances dans les systèmes e-Commerce.

Sahbi SIDHOM et Malek GHENIMA sont fondateurs du projet de la conférence internationale sur les systèmes d'information et l'intelligence économique (SIIE) depuis 2007 ([www.siiie.fr](http://www.siiie.fr)).

Dr. **Philippe LAMBERT** est professionnel dans le pôle technologique des nano-technologies à l'Institut Jean Lamour (Nancy-Université et CNRS France). Ces principales recherches sont dans le domaine de la veille technologique, l'IST et l'intelligence économique. Il est fondateur de Vinalor Nancy ([www.vinalor.fr](http://www.vinalor.fr)).