MODELE D'ANALYSE MULTIDIMENSIONNELLE DEDIE A L'INTELLIGENCE ECONOMIQUE

Ilhème Ghalamallah, **Bernard Dousset**
ghalamal@irit.fr, dousset@irit.fr

Institut de Recherche en Informatique de Toulouse, IRIT-SIG, Université Paul Sabatier,
118, route de Narbonne, 31062 Toulouse cedex 9, France

Mots clefs :
Analyse relationnelle, extraction de connaissances, évolution, intelligence économique.

Keywords:
Relational analysis, data mining, evolution, business intelligence.

Palabras clave :
Análisis emparentado, explotación minera de los datos, evolución, inteligencia de negocio.

Résumé

Le processus d'intelligence économique est une coordination des processus d'analyse stratégique et de veille. Par les modèles d'analyse stratégique une entreprise peut identifier et comprendre les opportunités et menaces issues de son environnement, et par le processus de veille elle satisfait son besoin informationnel.
Ainsi, nous considérons la démarche d'intelligence économique comme un processus qui englobe l'analyse stratégique et la veille. Dans ce cadre, nous présentons un modèle d'analyse multidimensionnel qui permet de satisfaire le besoin informationnel engendré par une démarche d'Intelligence économique (IE). 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. Dans un objectif de la création de connaissances.

Abstract

The process of competitive intelligence is a process of coordination of strategic analysis and scanning. By the models of strategic analysis a company can identify and understand the opportunities and threats from its environment and the scanning process satisfy her informational needs.
Thus, we consider the process of competitive intelligence as a process that includes strategic and scanning analysis. In this context, we present a multidimensional analysis model that can satisfy the information needs generated by a competitive intelligence. This model is based on the use of information by analyse the evolution of their interactions, to understand and summarize the behavior of the environment through new synthetic knowledge. With the aim of creating knowledge.

Introduction

L'IE est une réponse aux bouleversements de l'environnement global des entreprises. Dans une économie où tout se complexifie et bouge rapidement, l'avantage concurrentiel des entreprises s'inscrit de plus en plus dans la capacité de celles-ci à développer de nouvelle connaissances en vue de produire, de manière continue, des innovations [24]. Par ailleurs, les NTIC apportent des contraintes auxquelles les entreprises doivent s'adapter : un flot d'informations continu, une circulation beaucoup plus rapide de l'information, des techniques de plus en plus complexes (les logiciels sont de plus en plus difficiles à appréhender rapidement). Le risque est d'être submergé par cette information, de ne plus pouvoir distinguer l'essentiel du négligeable. En fait, avec l'avènement de la nouvelle économie dominée par le marché, la problématique industrielle de l'entreprise s'est complexifiée. Désormais pour être compétitive, l'entreprise doit savoir gérer son capital immatériel. L'IE est une démarche et un processus organisationnel qui permet à l'entreprise d'être plus compétitive, par la surveillance de l'environnement et des changements externes d'une part, et d'autre part par la surveillance des changements internes.

Dans le cadre de cet article nous proposons un modèle d'analyse multidimensionnelle dédié à l'intelligence économique. 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.

1 Intelligence Economique

L'IE reste encore, six ans après la définition canonique proposée par Martre [12], une notion aux frontières peu stables. Les dernières années ont vu les définitions de l'IE se multiplier en évoluant assez sensiblement, passant de définitions quasi exclusivement centrées sur la description des processus et techniques de l'IE, à des définitions incluant les objectifs stratégiques de l'IE, pour faire face depuis peu à des définitions incluant les notions de gestion des connaissances, d'apprentissage collectif ou de coopération [21].

Notre vision de l'IE est éminemment stratégique, c'est une démarche d'anticipation et de projection dans le futur, par la mise en évidence des liens unissant les acteurs dans un même secteur d'activités. L'IE repose sur une démarche d'anticipation individuelle et collective, une profonde connaissance de l'environnement et des réseaux existants afin de pouvoir agir et réagir en fonction de leur évolution. La coordination des actions dans le cas d'une stratégie commune requiert une forte capacité à saisir les variations et les réactions environnementales à chaque étape afin de repérer les facteurs de changement et d'en tenir compte par des corrections appropriées.

Dans le tableau de synthèses ci-dessous nous avons récapitulé l'historique des principaux axes de l'IE. Nous pouvons les décomposer en deux tendances : le cœur du domaine qui est stable depuis plus de 15 ans (Rassemblement, Traitement, Diffusion, Interprétation, Connaissances, Coordination, Prise de décision, Environnement) et les nouvelles préoccupations plus ponctuelles, mais récentes et qui font essentiellement intervenir le facteur temps (Immédiate, Ultérieure, Continue, Anticiper, Au bon moment). Notre proposition prend en compte ces deux dimensions : le cycle de l'IE avec sa composante relationnelle et sa synchronisation avec les besoins effectifs et instanciés des décideurs.

Dans le contexte de nos travaux, nous retenons la notion de l'IE telle qu'elle a été définit par Henri Martre c.à.d. en tant qu'ensemble des actions coordonnées de recherche, de traitement et distribution de l'information utile aux acteurs pour permettre l'action et la prise de décision. Ceci dépasse les actions partielles

désignées sous le nom de documentation, de veille (scientifique et technologique, concurrentielle, financière, juridique et réglementaire) et invite de surcroît à "passer d'un traitement individuel de l'information à la gestion de l'information et à un processus d'actions collectives" [21].

	Wilensky	Baumard	Martre et AL	Martinet et Marti	Levet et Paturel	Colletis	Revelli	Besson et Possin	De Vasconcelos	Levet	Paturel	Guilhon et Manni	Juillet
	1967	1991	1994	1995	1996	1997	1998	1998	1999	2001	2002	2003	2005
Rassemblement, Recherche, Collecte, Recueil	■		■		■		■	■	■		■	■	
Traitement, Tri, Mémorisation, Validation	■		■				■	■	■		■	■	
Savoir-faire, Acteurs		■	■			■		■			■		■
Diffusion, Distribution	■		■				■	■	■			■	
Interprétation, Analyse, Production	■				■			■		■	■		
Connaissances, Informations stratégiques								■	■	■		■	■
Coordinations, Collectives, Connexion, Combiner, Communication, Partage		■			■	■			■			■	
Prise de décision, Actions	■	■	■				■		■				■
Environnement				■						■		■	
Comprendre, Adapter				■						■			
Détecter, Surveillance active								■			■		
Immédiate					■								
Ultérieure					■								
Menaces, Opportunités								■					
Continue									■				
Anticiper										■			
Au bon moment											■		
Créativités, Compétences nouvelles												■	
Protection													■

Tableau 1 : Les principaux axes de l'IE

Dans la littérature nous retrouvons un ensemble de fonctions de l'IE, nous les résumons dans le tableau suivant :

Fonctions	Auteurs
Maîtrise du patrimoine scientifique et technique, et des savoirs faire	Clerc (1997) ; Levet et Paturel (1996)
Détection des menaces et des opportunités	Clerc (1997) ; Levet et Paturel (1996)
Influence	Clerc (1997) ; Levet et Paturel (1996)
Coordination des stratégies	Levet et Paturel (1996)
Coordinations des activités	Colletis (1997)
Renseignement	Besson et Possin (1996) ; Hassid et al (1997) ; Baud(1998) ; Lointier (2000) ; Larviet (2002)
Gestion de risque informationnel	Larivet (2002)
La création des connaissances	Levet et Paturel (1996) ; Besson et Possin (1998) ; De Vasconcelos (1999) ; Bournois et Romani (2000), AFDIE (2001) ; Guilhon et Manni (2003) ; Levet (2001) ; Jackobiak (2004)
L'aide à la décision	Bloch (1996) ; Revelli (1998) ; Bournois et Romani (2000) ;AFDIE (2001)
L'innovation	Martre (1994) ; Bloch (1996) ; Bournois et Romani (2000)

Tableau 2: Les fonctions de l'IE [25]

La fonction d'IE à laquelle nous nous intéressons, ici, est essentiellement la création de connaissance (Levet et Paturel (1996) ; Besson et Possin (1998) ; De Vasconcelos (1999) ; Bournois et Romani (2000), AFDIE (2001) ; Guilhon et Manni (2003) ; Levet (2001) ; Jackobiak (2004)).

2 Présentation du modèle d'analyse multidimensionnel

Le modèle d'analyse multidimensionnel que nous proposons, se base sur les quatre principales étapes du processus d'intelligence économique, à savoir « La formulation du besoin, La collecte et le traitement des donnée, L'analyse, La restitution et interprétation des résultats ». Dans l'objectif principal est la création de connaissances.

Dans le cadre de cet article nous allons détailler les étapes allant de la formulation du besoin à la représentation des données collectées, et présenter brièvement les étapes d'analyse et de restitution.

2.1 Formulation du besoin

Afin de faciliter la formulation du besoin Tournier [22] propose un pseudo langage qui permet une spécification plus précise d'analyse. Il définit une classification des questions classiques de recherche « Quoi ?, Qui ?, Où ?, Quand ?, Pour qui ?, Quelles données ? » selon trois clauses (clause Analyser, clause En Fonction, clause Pour). Nous enrichissant et adaptons cette classification en répondant à deux autres questions de recherches qui sont « Pourquoi ? et Comment ? », on définie une nouvelle clause « clause Dans l'Objectif ». Ainsi nous adoptons le pseudo langage comme suit :

Clause Dans l'Objectif : cette clause répond aux questions « Pourquoi ? ». Elle permet de définir le contexte général (Sujet) et le ou les objectifs de l'analyse décrit par les décideurs

Clause Analyser : cette clause réponds à la question « Quoi ? et Comment ? », Elle définit les axes d'analyses que les décideurs décrivent par un ou plusieurs indicateurs correspondant aux objectifs de la clause précédente.

Clause En Fonction : cette clause répond à la question « Qui ?, Où ?, Quand ? ». Elle indique les acteurs associés aux indicateurs de l'analyse.

Clause Pour : cette clause répond à la question « Pour qui ?, Quelles données ? ». Ceux sont les contraintes sur les données de l'analyse.

Nous définissons alors une hiérarchie des concepts de la spécification des besoins comme suit :

Un besoin B est défini comme suit :

$$B = < S_A, Obj_A, Ind_A, Act_A, Att_A >$$

- S_A : représente le contexte général du besoin pour l'analyse A,
- $Obj_A = < Obj_1, Obj_2, \ldots, Obj_m >$, représente les objectifs fixés pour le sujet S_A.
- $Ind_A = \{< Obj_i, < Ind_{i1}, Ind_{i2}, \ldots, Ind_{in} >>\}$, représente les indicateurs associés à chaque objectifs.
- $Act_A = \{< Ind_{ij}, < Act_{ij1}, Act_{ij2}, \ldots, Act_{ijp} >>\}$, représente les acteurs identifiés pour les indicateurs définies pour un objectif.
- $Att_A = \{< Act_{ijk}, < Att_{ijk1}, Att_{ijk2}, \ldots, Att_{ijkq} >>\}$, représente les attributs spécifiés pour chaque acteurs.

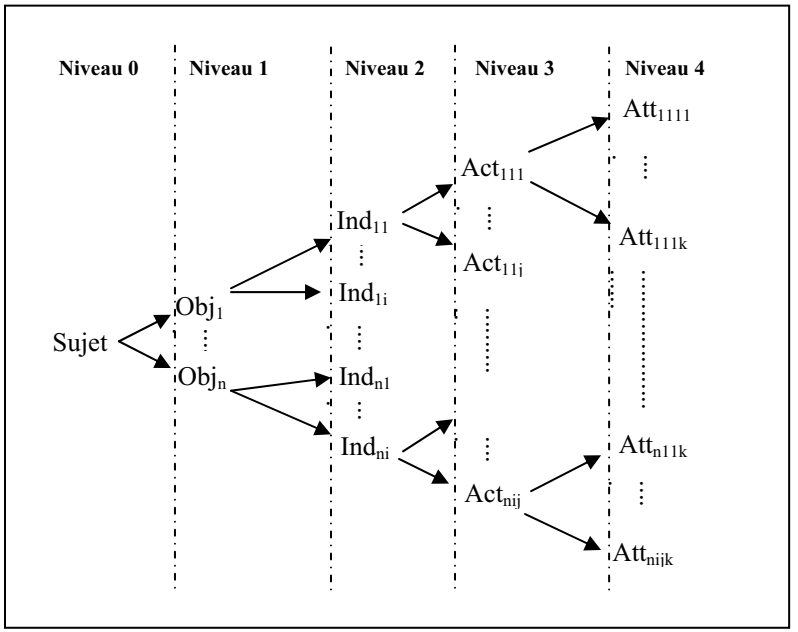

Figure 1 : Hiérarchie des besoins

Exemple

Dans l'Objectif d'une évaluation de la recherche scientifique (Les collaborations, La production scientifiques, Les thématiques de recherches).
Analyser Les collaborations entre auteurs ***En Fonction*** du nom de l'Auteur de l'article, et ***en Fonction*** des co-Auteurs.
Analyser Les collaborations entre organismes ***En Fonction*** du nom de l'Auteur de l'article, et ***en Fonction*** des Organismes de recherches des auteurs.
Analyser Les collaborations entre pays ***En Fonction*** du nom de l'Auteur de l'article, et ***En Fonction*** du Pays.
Analyser L'évolution des collaborations entre pays ***En Fonction*** du nom de l'Auteur de l'article, ***En Fonction*** du Pays et ***En Fonction*** de la Date de publication.
Analyser Le nombre de publication ***En Fonction*** du nom de l'*Auteur* de l'article.
Analyser L'évolution du nombre de publication ***En Fonction*** du nom de l'*Auteur* de l'article ***En Fonction*** de la Date de publication.
Analyser L'évolution des thématiques de recherches des auteurs ***En Fonction*** du nom de l'*Auteur* de l'article ***En Fonction*** des Descripteurs de l'article ***En Fonction*** de Date ***Pour*** la période 1999-2004

$$B_s = < S_A^s, Obj_A^s, Ind_A^s, Act_A^s, Att_A^s >$$

S_A^s = « évaluation de la recherche scientifique »
Obj_A^s = < 'Les collaborations', 'La production scientifiques', 'Les thématiques de recherches'>
Ind_A^s = {< '*Les collaborations*', < 'Les collaborations entre auteurs', 'Les collaborations entre organismes', 'Les collaborations entre pays'>>, < '*La production scientifiques'*, < 'Le nombre de publication', 'L'évolution du nombre de publication'>>, < '*Les thématiques de recherches*', <'L'évolution des thématiques de recherches des auteurs'>>}
Act_A^s = {< '*Les collaborations entre auteurs'*, < Auteur >>, < '*Les collaborations entre organismes*', < Auteur, Organismes >>, < '*Les collaborations entre pays'*, < Auteur, Pays >>, < '*Le nombre de publication'*, < Auteur >>, < '*L'évolution du nombre de publication'*, < Auteur, Date >>, <'*L'évolution des thématiques de recherches des auteurs'*, < Auteur, Descripteurs, Date >>}
Att_A^s = << Auteur, Descripteurs, Date >, < 1999, 2000, 2001, 2002, 2003, 2004>>

2.2 La collecte et le traitement des données

2.2.1 Identification des sources et collecte des données

Cette étape consiste à sélectionner les données sur lesquelles va porter l'analyse. Il s'agit de restreindre l'espace d'information à explorer dans un but de retenir que les informations pertinentes par rapport aux besoins informationnels de l'analyse.
Nous décomposons cette étape en trois phases :

Phase 1 : Collecte d'informations à partir des sources de données, La collecte d'information se base sur la sélection des données qui est généralement réalisée par l'interrogation d'une source de données spécifique ou d'un ensemble de sources de données. Ainsi cette phase consiste à construire le corpus source à partir des corpus de données collectées et d'identifier les structures associées dans le cas d'utilisation de plusieurs sources.

Phase 2 : Extraction et structuration du corpus source, La deuxième phase permet l'uniformisation du corpus source, par un processus d'homogénéisation des données et de leurs structures afin d'avoir une vue unifié des données collecté que nous appelons corpus structuré.

Phase 3 : Représentation multidimensionnel du corpus structuré, La troisième phase, définie la représentation du corpus structuré sous forme multidimensionnelle, afin de construire un corpus multidimensionnel qui synthétise toutes les informations utiles à l'analyse sous forme relationnelle.

2.2.2 Collecte d'informations à partir des sources de données

Elle consiste à la mise en relation des besoins identifiés pour l'analyse et les sources de données existantes. Cette mise en relation permet de traduire chaque élément de l'arborescence des besoins en un ensemble de mots clés, qui vont faciliter l'identification et la collecte de données.

Ainsi, notre approche de collecte de données à partir des sources de données est décomposée en deux étapes :

- La première étape est la recherche d'information. Cette étape repose sur les Systèmes de Recherche d'Information (SRI) qui intègrent un ensemble de modèles et de processus permettant de sélectionner des informations pertinentes en réponse aux besoins identifiés. Le processus de recherche d'information consiste à mettre en correspondance les besoins identifiés sous forme d'un ensemble de mots clés (requête) avec l'ensemble des descripteurs de la collection de documents (ou de pages web). Ce processus permet de restituer une liste de documents triés selon leur pertinence vis-à-vis de la requête. Cette liste nous l'appellerons corpus de données collectées. Ce corpus correspond donc à un ensemble de données déconnectées du SRI qui a permis de les obtenir et auquel est associée une structure. Remarque : cette étape n'est pas détaillée dans notre proposition, puisque nous nous intéressons seulement à son résultat (corpus de données collectées).
- La deuxième étape, consiste à identifier et à rassembler les différents corpus collectés afin de définir le corpus source.

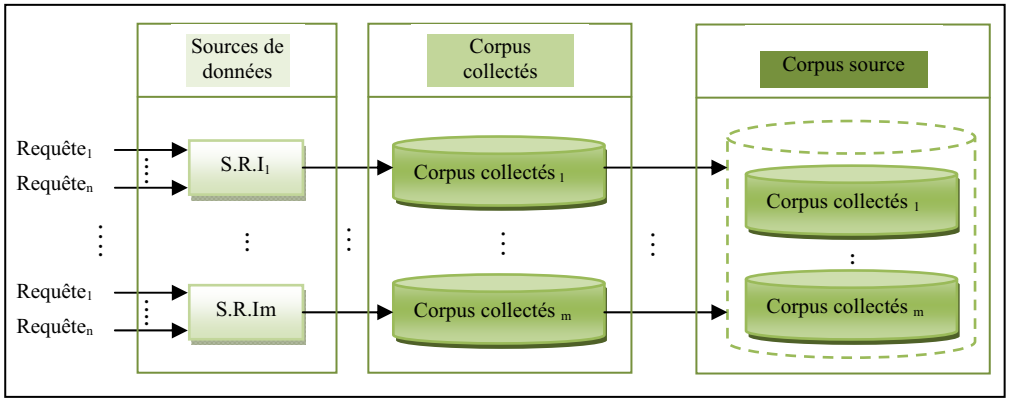

Figure 2 : Processus de recherche

Dans ce qui suit, nous présentons les sources de données que nous retenons dans le cadre de nos travaux et nous proposons une formalisation de la structure des corpus de données collectées et du corpus source.

2.2.2.1 SOURCES DE DONNEES

La caractéristique des sources de données traitées sont des sources de données textuelles. Ces sources représentent une collection de documents électroniques qui sont définis comme un ensemble d'informations organisées et représentées selon les choix de l'auteur. Nous pouvons associés à un document plusieurs vues [5] :

- Une vue décrivant le **contenu** du document,
- Une vue décrivant la **structure physique** du document, qui concerne la présentation physique du document et ces supports de visualisations,
- Une vue décrivant la **structure logique** du document, qui permet d'identifier les granules d'information d'un document et définir un découpage de l'information d'un point de vue hiérarchique.

La définition du concept de structure logique à fait apparaitre trois types de documents : les documents non structurés (ne contenant pas d'informations de structure) ; les documents semi-structurés (contenant peu d'information sur la structure du document) ; et les documents structurés (contenant l'ensemble des informations sur la structure du document).

Les sources que nous retenons dans le contexte de nos travaux, sont uniquement des sources électroniques et qui contiennent seulement des documents de type structurés ou peu structurés, ainsi nous définissons deux types de sources de données :

- ***Sources de données structurées*** : les bases bibliographiques, les bases de brevets… (aussi bien sur serveurs en ligne que sur cd-rom),
- ***Sources de données peu structurées*** : Flux RSS, Site web, Page web, Traces de connexions, Brevets, Groupes de discussions, Presse en ligne

2.2.2.2 Structure du corpus de données source

La structure des sources de données est définie à partir des vues de la structure logique et son contenu associées aux documents. Elle est représentée sous forme d'un ensemble de champs qui peuvent contenir une ou plusieurs valeurs associées. Ces valeurs renseignent sur le type, la nature et la localisation de toutes les informations élémentaires que chaque champ peut contenir [13] dans un document. Cette structure permet une description synthétique des documents de la source.

Nous décrivons le corpus source et sa structure à partir de la structure logique et le contenu de la source de données associé (structure du corpus collecté). Ils sont construits à partir d'une ou plusieurs sources de données. Nous associons chaque corpus collecté à sa source de donnée.

Soit la structure $Structure^S$ de la source de données S définie comme suit :

$$Structure^S = < Champ_1^S, \ldots, Champ_i^S, \ldots, Champ_j^S >$$

Avec

$Champ_i^S$: représente le champ i de la structure logique des documents de la source S.

Le corpus collecté $Corpus_{Collecté}^{S}$ issu de la source *S* est défini alors comme suit :

$Corpus_{Collecté}^{S} = < \{ valeur_1^1, \ldots, valeur_1^i, \ldots valeur_1^j \}, \ldots, \{ valeur_k^1, \ldots, valeur_l^i \ldots, valeur_m^j \} >$

Avec

- $\{ valeur_1^1, \ldots, valeur_k^1 \}$: représente l'ensemble des valeurs associés au $Champ_1^S$
- $\{ valeur_1^i, \ldots, valeur_l^i \}$: représente l'ensemble des valeurs associés au $Champ_i^S$
- $\{ valeur_1^j, \ldots, valeur_m^j \}$: représente l'ensemble des valeurs associés au $Champ_j^S$
- $\{ valeur_k^1, \ldots, valeur_l^i \ldots, valeur_m^j \}$: représente l'ensemble des valeurs associées à un document du corpus collecté.

Le corpus source $Corpus_A^{Source}$ associé à une analyse A est défini comme suit :

$$Corpus_A^{Source} = \{< Nom_o^S, Structure_o^S, Corpus_{oCollecté}^S >\}, o \in \{1..N\}$$

Nom_o^S correspond au nom de la source o.

Exemple

Soient deux corpus collectés issus respectivement des sources de données Pascal[1] et Factiva[2].

- Le document du corpus collecté de la source Pascal est présenté comme suit :

TI : Combining mining and visualization tools to discover the geographic structure of a domain
AU: MOTHE-Josiane; CHRISMENT-Claude; DKAKI-Taoufiq; DOUSSET-Bernard; KAROUACH-Said
AF: Institut de Recherche en Informatique de Toulouse, 118 Route de Narbonne, 31062 Toulouse, France
CF: *Workshop on Geographic Information Retrieval (GIR)
SO: Computers-environment-and-urban-systems. 2006; 30 (4) : 460-484
PY: 2006
CP: United-Kingdom
DE: Geographic-information-system; Information-retrieval; Congress-; Data-mining

La structure associée à la source est : $Structure^{Pascal} = <$ TI, AU, AF, CF, SO, PY, CP, DE $>$

[1] Pacsal : Base de données bibliographiques en ligne multidisciplinaire répartie entre Biomed : - Médecine - Pharmacologie - Psychologie et Scitech : - Sciences de l'ingénieur - Chimie - Physique - Sciences relatives à la terre, l'océan, l'espace ainsi que les sciences en général

[2] Factiva : Base de données en ligne de presse et d'information économique.

Et le corpus collecté associé : $Corpus_{Collecté}^{Pascal}$ = < {Combining mining and visualization tools to discover the geographic structure of a domain, MOTHE-Josiane; CHRISMENT-Claude; DKAKI-Taoufiq; DOUSSET-Bernard; KAROUACH-Said , Institut de Recherche en Informatique de Toulouse, 118 Route de Narbonne, 31062 Toulouse, France,, Workshop on Geographic Information Retrieval (GIR), Computers-environment-and-urban-systems. 2006; 30 (4) : 460-484,2006, United-Kingdom , Geographic-information-system; Information-retrieval; Congress-; Data-mining }>

- Le document du corpus collecté de la source Factiva est présenté comme suit :

SE	Emploi ET formation
HD	La veille stratégique pour anticiper
BY	François Courvoisier
WC	549 mots
PD	12 janvier 2007
SN	Le Temps
LA	Français

La structure associée à la source est : $Structure^{Factiva}$ = < SE, HD, BY, WC, PD, SN, LA >

Et le corpus collecté associé : $Corpus_{Collecté}^{Factival}$ = < { Emploi ET formation, La veille stratégique pour anticiper François Courvoisier , 549 mots ,12 janvier 2007, Le Temps , Français } >

Le corpus source $Corpus_A^{Soure}$ associé à une analyse A, construit à partir des sources de données Pascal et Factiva est :

$$Corpus_A^{Source} = \{< \text{Pascal}, Structure^{Pascal}, Corpus_{Collecté}^{Pascal} >, < \text{Factiva}, Structure^{Factiva}, Corpus_{Collecté}^{Factiva} >\}$$

2.2.3 Extraction et structuration des données du corpus

Différent travaux traitent de l'harmonisation des bases de données hétérogènes [2][1][23][19][8]. Ces travaux concernent la standardisation de la requête sur des bases de données relationnelles. L'objectif de cette étape est différent puisque elle consiste à définir [8] une uniformisation logique des données sélectionnées, sans uniformisation des bases elles mêmes.

Dans le contexte de notre proposition, nous nous basons sur les principes d'extraction d'informations définis par [8] [5] qui permettent d'extraire des informations prédéfinies à partir de documents qui peuvent être hétérogènes mais pour lesquels la localisation de l'information à extraire est balisée. Ces solutions permettent de:

- Définir un descripteur générique de documents de structure hétérogène,
- Gérer les différents conflits syntaxiques,
- Gérer les différents conflits sémantiques.

Le descripteur générique des documents issus du corpus de données source hétérogènes correspond à une représentation structurée prédéfinie de l'ensemble des documents. Sa création se base sur la prise en compte des descripteurs de formats spécifiques des différents corpus sources pour générer un descripteur générique commun (Voir Figure) dans le cas d'utilisation de plusieurs bases sources.

2.2.3.1 Descripteurs de formats spécifiques au corpus

Le descripteur de format spécifique décrit de manière complète le corpus auquel il est associé. [8] ont définis les descripteurs de format spécifiques par les composantes locales suivantes :

- La structure,
- Les règles d'extraction qui permettent de décrire la façon dont les informations utiles seront extraites,
- Les transformateurs sémantiques qui permettent de s'affranchir des éventuels conflits liés aux relations sémantiques (synonymies, …) entre les valeurs prises par un même attribut.

2.2.3.2 Descripteurs génériques

Soient (E1, E2,…, En), n corpus de données qui représentent le résultat de l'interrogation de n bases différentes. Pour chaque corpus n issus d'une base n, nous associons un descripteur de format spécifique que nous le définissons par sa structure et ses règles d'extraction. L'objectif de cette étape et de dériver à partir de l'ensemble de ses descripteurs de format spécifique un descripteur générique commun à toute les bases.

Corpus structuré

Nous définissons, $Structure_{Global}^{Extrac}$ la structure globale d'extraction associée au corpus source :

$$Structure_{Global}^{Extrac} = < Chp_{1G}^{Extrac}, \ldots, Chp_{iG}^{Extrac}, \ldots, Chp_{jG}^{Extrac} >,$$

Avec

- Chp_{iG}^{Extrac} : correspond à l'élément i de la structure d'extraction global,

Le corpus global $Corpus_{Global}$ issu du corpus source est défini comme suit :

$$Corpus_{Global} = < \{ valeur_1^{1G}, \ldots, valeur_1^{iG}, \ldots, valeur_1^{jG} \}, \ldots, \{ valeur_k^{1G}, \ldots, valeur_l^{iG}, \ldots, valeur_m^{jG} \} >$$

Avec

- $\{valeur_1^{1G},\ldots, valeur_k^{1G}\}$: représente l'ensemble des valeurs associés à l'élément d'extraction Chp_{1G}^{Extrac}.
- $\{ valeur_1^{iG}, \ldots, valeur_l^{iG}\}$: représente l'ensemble des valeurs associés à l'élément d'extraction Chp_{iG}^{Extrac}.
- $\{ valeur_1^{jG}, \ldots,valeur_m^{jG}\}$: représente l'ensemble des valeurs associés à l'élément d'extraction Chp_{jG}^{Extrac}.
- $\{ valeur_k^{1G}, \ldots, valeur_l^{iG}, \ldots, valeur_m^{jG}\}$: représente es valeurs associée à un document du corpus structuré.

Le corpus structuré $Corpus_A^{Structué}$ associé à l'analyse A est défini alors comme suit :

$$Corpus_A^{Structué} = \{< Structure_{Global}^{Extrac}, Corpus_{Global} >\}$$

2.2.4 Représentation multidimensionnelle du corpus

L'objectif de cette structure est de permettre l'étude de l'évolution des interactions entre variables afin de réaliser des projections dans l'avenir, qui sont essentielles pour la prise de décisions stratégiques. Notre proposition consiste à définir une structure unique de données intermédiaires entre informations brutes et pré-connaissances déduites, sous la forme d'un entrepôt de données générique, qui ne contiendra que des pré-connaissances sous forme de relations évolutives. Cette structure de corpus servira de support pour l'application des différentes fonctions de découverte de connaissances.

La structure du corpus multidimensionnel repose sur une modélisation à trois dimensions. Cette dernière permet de définir les différentes relations de dépendances entre les éléments de la structure d'extraction du corpus structuré (les variables du corpus) avec la prise en compte de la structure temporelle (la variable temporelle) (voir figure).

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

$$Structure_{Global}^{Extrac} = <\text{N° Doc, Date, Auteur, Revue, Pays, Mots C, Organisme}>$$

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.à.d. l'élément « Date ».

Principe

Notre but est d'identifier toutes les relations de dépendances existantes dans le corpus entre les différentes variables de l'étude (voir Figure). Ces relations sont définies par des matrices de co-occurrences. Ces matrices indiquent la présence simultanée des modalités de deux variables qualitatives dans un document.
Nous adoptons ces matrices en y rajoutant une troisième variable comme suit :

- Les deux premières variables sont les variables qualitatives associées au corpus multidimensionnel,
- Et la troisième variable est toujours la variable temporelle (Data, année, …) associée au corpus.

Ainsi, la matrice de co-occurrence consiste à indiquer la présence des modalités de ces trois variables dans un document (structure trois dimensions). Nous nommons cette matrice « Cube ».
Le cube permet de regrouper les relations existantes dans un corpus en périodes. Nous identifions deux types de forme de cube :

- Cube sous forme de matrice 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 asymétrique : dans le cas où nous considérons la présence des modalités de deux variables distinctes et la variable temporelle dans un document.

Pour un corpus structuré dont la structure d'extraction est définie comme suit :

$$Structure_{Global}^{Extrac} = < Chp_{1G}^{Extrac}, \ldots, Chp_{iG}^{Extrac}, \ldots, Chp_{jG}^{Extrac} >,$$

Avec

Chp_{iG}^{Extrac} : correspond à l'élément i de la structure d'extraction global,

Nous définissons les types de matrices à trois dimensions entre les éléments de la structure comme suit :

Dimensions	Chp_{1G}^{Extrac}	…	Chp_{iG}^{Extrac}	…	Chp_{iG}^{Extrac}
Chp_{1G}^{Extrac}	Symétrique	…	Asymétrique	…	Asymétrique
⋮	⋮	⋮	⋮	⋮	⋮
Chp_{iG}^{Extrac}	Asymétrique	…	Symétrique	…	Asymétrique
⋮	⋮	⋮	⋮	⋮	⋮
Chp_{iG}^{Extrac}	Asymétrique	…	Asymétrique	…	Symétrique

Figure 2 :Types de Matrice

Tableau à trois dimensions

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 $\{ x_1, \ldots, x_p \}$ de la variable X.

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

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

Matrice symétrique

La matrice asymétrique est une matrice à r lignes et s colonnes qui a pour élément générique le nombre n_{ijk} d'individus tel que $x_i \in \chi$ et $x_j \in \chi$ et $t_k \in \tau$

X x X x 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}
⋮		⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮
x_i		n_{i11}	…	n_{i1r}	.....	n_{ij1}	…	n_{ijr}	.....	n_{ip1}	…	n_{ipr}
⋮		⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮
x_p		n_{p11}	…	n_{p1r}	.....	n_{pj1}	…	n_{pjr}	.....	n_{pp1}	…	n_{ppr}

Figure 3 :Matrice symétrique

Matrice asymétrique

La matrice asymétrique est une matrice à r lignes et s 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 x Y x 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}
⋮		⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮	⋮
x_i		n_{i11}	…	n_{i1r}	.....	n_{ij1}	…	n_{ijr}	.....	n_{iq1}	…	n_{iqr}

Figure 4 :Matrice asymétrique

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 garderons que les cases dont les valeurs sont supérieur ou égales à un.

Remarque : Nous considérons dans le corpus multidimensionnel chaque variable comme dimension, leurs modalités et les valeurs du cube comme attributs.
Nous définissons le corpus multidimensionnel associé au cube comme suit :

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

$$SC_{M3D} = \{< Dim_i, Dim_{ij}, Dim_T, Nb_{Doc}^{ijT} >\}$$

Le corpus multidimensionnel C_{M3D} définie comme suit :

$$C_{M3D} = \{< Att_x^i, Att_y^{ij}, Att_z^{ijT}, Att_o^{ijT} >\}$$

Avec

- $Att_x^i \in \boldsymbol{D_i}$ l'ensemble des attributs $\{Att_1^i, \ldots, Att_p^i\}$ de la Dimension i « Dim_i »,
- $Att_y^{ij} \in \boldsymbol{D_j}$ l'ensemble des attributs $\{Att_1^j, \ldots, Att_q^j\}$ de la Dimension j associé à la dimension i «Dim_{ij} »,
- $Att_z^{ijT} \in \boldsymbol{D_T}$ l'ensemble des attributs $\{Att_1^T, \ldots, Att_r^T\}$ de la Dimension temps « Dim_T »,
- $Att_o^{ijT} \in \boldsymbol{Nb_D}$ l'ensemble des attributs $\{Att_1^{Nb}, \ldots, Att_l^{Nb}\}$ du nombre de documents ou les trois dimensions apparaissent simultanément.

$$Att_o^{ijT} = \boldsymbol{D_i} \times \boldsymbol{D_j} \times \boldsymbol{D_T} [Att_x^i, Att_y^{ij}, Att_z^{ijT}] \text{ et } Att_o^{ijT} >= 1$$

2.3 Analyse

L'interrogation s'effectue selon deux approches :
- Soit en utilisant des requêtes prédéfinis, qui représentent une synthèse du savoir faire sur l'utilisation stratégique du relationnel,
- Soit par des requêtes formulées par l'utilisateur avec l'assistance d'un outil d'aide à la formulation de requête. Cette étape, est en cours d'experimentation, nous l'introduisons dans nos perspectives proches de nos travaux.

Les requêtes prédéfinies sont des requêtes qui sont figées et où l'utilisateur se contente de choisir les variables qu'il veut analyser et les filtres à appliquer.
Exemple 1 :

- Corpus analysé (ensemble de brevets)
- Métrique (cooccurrences)
- Question posée (évolution du top 10 des entreprises sur les 3 dernières périodes).
- Support (matrice de cooccurrence 2D : Entreprises X Temps)
- Requête choisie (évolution d'une variable)
- Paramétrage (Dimension: Entreprises, Filtre : Top 10 + Définition des 3 dernière périodes)

Exemple 2 :

- Corpus analysé (ensemble de brevets)
- Métrique (cooccurrences)
- Question posée (évolution de la stratégie des co-dépôts de brevets d'un panel d'entreprises sur 2 périodes).
- Support (matrice de cooccurrence 3D : Entreprises X Entreprises X Temps)
- Requête choisie (évolution des relations dans une variable)
- Paramétrage (Dimension : Entreprises, Filtre : le Panel + Définition des 2 périodes)

2.4 La restitution et interprétation des résultats

Les fonctions de « reporting » sont essentielles pour réussir la présentation d'un travail de veille et pour convaincre les décideurs par un document lisible, pertinent et concis. Outre les grands classiques (histogrammes 2 et 3D, camemberts, boîtes à pattes, droite de régression, zoom de matrices,…), nous comptons intégrer des techniques de visualisation propres à chaque type de requête comme (histogrammes d'évolution 2D et 3D, histogrammes comparatifs ou cumulatifs 2D et 3D, pondération par des données externes, cartes géographiques, étoiles ou flocons, graphes relationnels, classifications, transitivités, …). Cet ensemble de possibilités doit permettre à chacun de trouver les bons réglages pour découvrir puis communiquer l'information stratégique ciblée à intégrer dans son rapport d'analyse personnalisé.

Conclusion

Au cours de cet article nous nous sommes intéressés à la création de la connaissance par le processus d'IE. Dans ce contexte (IE), une grande part de l'information à portée stratégique vient du relationnel et la pertinence des connaissances extraites dépend très souvent de la prise en compte de l'évolution des données mais aussi de celles de leurs interactions. Nous avons définis une modélisation de ces relations sous forme relationnelle que nous appelons ' pré-connaissance' et qui nous sert de base pour la création de la connaissance. Nos perspectives, sont de continuer nos expérimentations sur les différents types de relations (matrice) afin de pouvoir proposer un modèle unifié pour générer et organiser les données sous forme relationnelle que nous appelons « pré-connaissance » et delà extraire des connaissances implicites dont le contenu et la mise en forme sont adaptés à des décideurs non spécialistes des domaines de l'extraction des connaissances.

3 Bibliographie

[1] **BARAL C.** (1991). *Combining multiple knowledge bases*, IEEE transactions on knowledge and data engineering, 208-231.

[2] **BATINI C.,** (1986). *A competitive analysis of methodologies for databases schema integration*, ACM computing surveys, 323-364.

[3] **BESSON B., POSSIN J-C.,** (1996), *Du Renseignement à l'Intelligence Economique*, Dunod, Paris.

[4] **BOURNOIS F., ROMANI P-J.,** (2000), *L'IE et stratégique dans les entreprises Françaises*, Edition Economica, Paris.

[5] **CHRISMENT C.,** (1997). *Extraction et synthèse de connaissances à partir de bases de données hétérogènes*, Ingénierie des systèmes d'information , 367-400.

[6] **CLERC P.,** (1997), *IE : enjeux et perspectives, dans rapport mondial sur l'information*, chapitre 22, Accessible sur http://www.unesco.org.webworld/wirerpt

[7] **COLLETIS G.,** (1997), *IE : vers un nouveau concept en analyse économique ?* Revue d'Intelligence Economique, n°1.

[8] **DKAKI T.,** (1996). *Extraction et synthèse de connaissances à partir de bases de données hétérogènes*. Inforsid, 287-308.

[9] **GUILHON A.,** (2003), *Le processus d'IE et l'identité de la PME, l'IE dans la PME : Visions éparses, paradoxes et manifestations*, Editions Economica.

[10] **LEVET J. L.,** (2001), *IE : mode de pensée, mode d'action*, Economica, Paris.

[11] **MARTINET B., MARTI Ph.,** (1995), *L'IE : comment donner de la valeur concurrentielle à l'information*, Editions d'organisation, Paris.

[12] **MARTRE H.,** (1994), *IE et stratégie des entreprises*, Œuvre Collective du Commissariat au Plan, Paris, la documentation Française.

[13] **MOTHE J.,** (1994). *Modèle connexionniste pour la recherche d'information, expansion dirigée de requête et apprentissage*. Toulouse: Thèse, Université Paul Sabatier.

[14] **MOTHE J.,** (2000). *Recherche et exploration d'information: Découverte des connaissances pour l'accès à l'information*. Toulouse: Université Paul Sabatier.

[15] **OUBRICH M.,** (2003a), *Le processus d'IE : transformer l'information en connaissance*, Acte de colloque de l'Association Information et Management, Grenoble.

[16] **OUBRICH M., GUILHON A.,** (2003b), *L'IE, un processus de création de connaissance : pour une reconsidération du rôle de l'apprentissage organisationnel*, Acte de colloque Médiation et Ingénierie des Connaissances, Marseille.

[17] **OUBRICH M.,** (2005), *La création des connaissances dans un processus d'IE*, Université de la Méditerranée (Aix-Marseille II), phd.

[18] **PATUREL R.,** (1998), *Panorama général et synthétique des thèses françaises en management stratégique*, Actes de la journée FNEGE «Recherche en Gestion», 23 Octobre.

[19] **REDDY M. P.,** (1994). *A methodology for integration of heterogeneous databases*. IEEE transactions on knowledge and data engineering, 920-933.

[20] **REVELLI C.,** (1998), *Intelligence stratégique sur Internet, comment développer efficacement des activités de veille et de recherche sur les réseaux*, Dunod, Paris.

[21] **SALLES M., CLERMONT Ph., DOUSSET B.,** (2000). *Une méthode de conception de système d'IE*. Communication au colloque IDMME'2000, Montréal.

[22] **TOURNIER R.,** (2007), *Analyse en ligne (OLAP) de documents*, Toulouse: Thèse, Université Paul Sabatier.

[23] **GOTTHARD W.,** (1992). *System-guided viex integration for object-oriented databases*, IEEE transactions on knowledge and data engineering, 1-22.