

# L'intelligence économique par l'analyse relationnelle

Ilhème Ghalamallah

ghalamal@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

Dans le contexte de l'intelligence économique (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. Dans cet article nous proposons une nouvelle approche originale d'analyse pour l'IE. Cette approche repose sur l'extraction des connaissances par l'analyse des données relationnelles évolutives. Notre but est de proposer un système 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.

# 1 Introduction

L'hyper compétition, la globalisation des marchés, l'innovation technologique constituent des facteurs de multiplication des risques qui rendent la progression des entreprises de plus en plus délicate à une échelle économique mondiale. Ces difficultés d'écoulement d'un environnement complexe, incertain et turbulent. Avec la mondialisation, l'entreprise doit s'ouvrir vers l'extérieur et donc s'informer tout en gardant la maîtrise de son environnement. Déjà, lors du colloque de Ile Rousse de 1995, Jo LINK-PEZET et all. ont souligné le passage, à forte implication stratégique, d'une économie de production à une économie d'environnement et à une économie de marché (création et développement d'alliances, fusions, acquisitions). Dans un univers où l'information devient pléthorique et multiforme, il est nécessaire d'adopter une attitude "anticipatrice" par une exploitation permanente des informations vitales: c'est ce que permettent les systèmes d'information pour l'aide à la décision. La réussite d'une entreprise dépendra donc en grande partie de sa capacité à gérer la collecte, le traitement et la diffusion de l'information à des fins stratégiques. D'où l'importance d'une démarche qui intègre à tous les niveaux la prise en compte des informations externes.

Selon le rapport Martre [1], l'efficacité d'une démarche stratégique repose sur le déploiement de véritables dispositifs d'intelligence économique qui instituent la gestion stratégique de l'information comme l'un des leviers majeurs au service de la performance économique et de l'emploi. L'intelligence économique devient un outil à part entière de connaissance et de compréhension permanente des réalités des marchés, des techniques et des modes de pensée des concurrents, de leur culture, de leurs intentions et de leur capacité à les mettre en œuvre.

Dans le cadre de cet article nous proposons une approche originale d'analyse pour l'intelligence économique. Cette originalité 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 ce contexte, nous présenterons un système d'aide à la décision dans la perspective d'automatiser la mise en ligne de l'information relationnelle (pré-connaissance) et de proposer des outils d'analyse et de navigation orientés intelligence économique. Ce système permet d'effectuer des analyses stratégiques sur des corpus d'information textuelle issus des sources les plus diverses et donne la possibilité aux décideurs de mener et gérer par eux même la collecte, le traitement et la diffusion de l'information stratégique sans la participation d'un expert du domaine.

## 2 Intelligence économique

La notion d'intelligence économique que nous considérons, est une démarche d'anticipation et de projection dans le futur, une démarche de mise en évidence de liens unissant différents acteurs dans un même secteur d'activités. L'intelligence économique 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 à un niveau individuel et collectif. La coordination des actions dans le cas d'une action commune requiert une forte capacité à saisir les variations et les réactions environnementales à chaque étape d'exécution d'une stratégie afin de repérer les facteurs de changement et réévaluer la stratégie et en tenir compte.

Dans le contexte de nos travaux, nous retenons la notion de l'intelligence économique telle qu'elle a été définie par Henri Martre c.a.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" [1].

## 3 Proposition

Les outils d'intelligence économique permettent à des organismes de maîtriser leur organisation interne et de comprendre leur environnement par l'acquisition, le classement, l'analyse, l'interprétation et l'exploitation systématiques d'information [2]. Carvalho et Ferreira [3] définissent deux types d'outils d'intelligence

économique. Le premier type concerne les systèmes qui permettent une manipulation massive des données opérationnelles et d'en extraire les connaissances. Parmi ces systèmes nous retrouvons, les systèmes d'aide à la décision interactifs, les systèmes d'information exécutifs, traitement en ligne analytique (OLAP), entrepôts de données et des systèmes d'extraction de données. Ils se basent sur des systèmes de gestion de base de données et sont utilisés dans le but d'indiquer les tendances et les modèles qui sont cachés dans leurs bases de données opérationnelles volumineuses [4]. Le deuxième type d'outil, appelé aussi outil d'intelligence concurrentielle, ces outils consistent à systématiser la collecte et l'analyse de l'information à partir d'un environnement concurrentiel pour l'aide à la prise de décision stratégique. Notre proposition se situe au niveau du deuxième type, ou l'information est principalement issue des sources publiques les plus diverses et qui sont présentes sur le web tel que: les bases en ligne, les Cd, le Web visible et invisible, les news, les brevets, la presse, les traces de connexions aux sites, les bases internes...etc.

Notre proposition s'intègre au processus d'intelligence économique tel qu'il a été défini dans le rapport Martre. Cette définition englobe quatre étapes à savoir : (1) formulation de besoin, (2) collecte d'information, (3) analyse, (4) diffusion des connaissances (Voir Figure 1.).

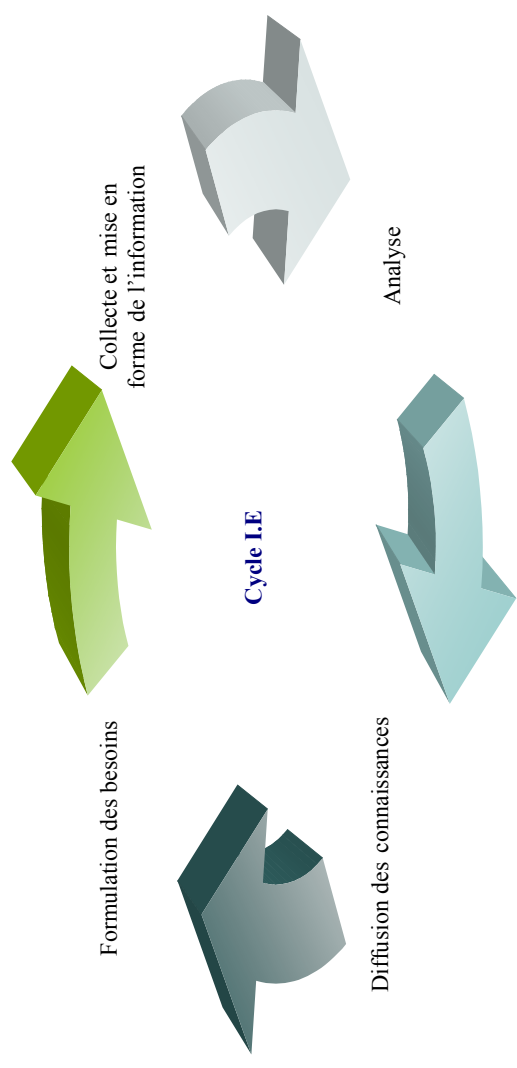
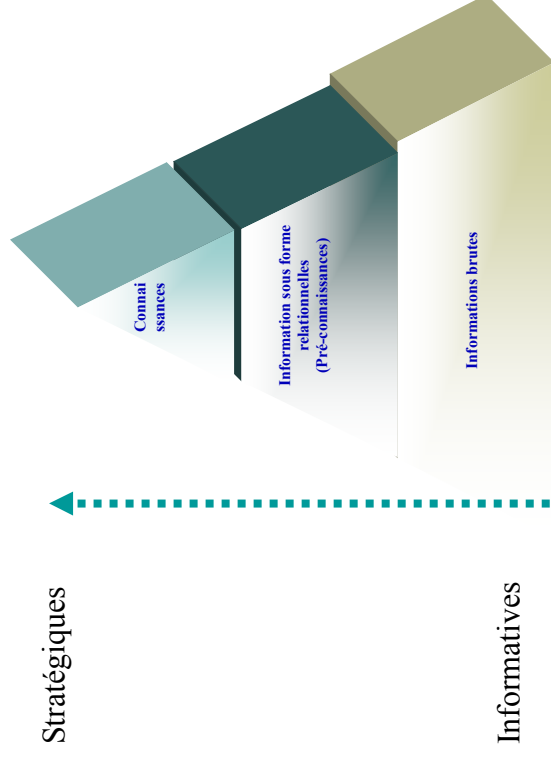


Figure 1.: Cycle de l'IE

[3], s'intéresse aux étapes 2, 3 et 4 du cycle de l'IE, qui sont les seules étapes qui peuvent être automatisées en utilisant les technologies de l'information. Dans le cadre de notre proposition, notre prototype couvre les trois traitements du cycle I.E à savoir : Collecte d'information, Requêtage, Diffusion des connaissances.

Nous définissons l'évolution et le traitement des données dans notre processus par trois étapes importantes à savoir : la collecte des données brutes, transformation de données brutes en informations sous forme relationnelles ce que nous appelons pré-connaissance, puis analyse des pré-connaissances pour l'extraction des connaissances. La figure qui suit représente la pyramide des données qui couvre le passage des données à but informatives à des données à but stratégiques.



**Figure2.: Pyramide des données**

Dans ce qui suit, nous allons définir les différentes étapes du cycle de notre approche en précisant les étapes que couvre notre prototype.

### 3.1 Formulation du besoin

Initialement, cette phase permet aux décideurs d'exprimer les besoins dont ils ont conscience et donc d'identifier et de hiérarchiser les cibles sur lesquelles ils doivent s'informer, tout cela pour piloter au mieux la ou les stratégies qu'ils ont retenues. Mais leurs besoins exprimés sont parfois non pertinents ou difficiles à satisfaire. Le but de l'IE et de leur faire connaître l'espace des informations potentiellement disponibles et surtout celles qui leurs seraient utiles : adéquation entre désirs et besoins réels. Le cycle de l'IE permet progressivement de faire converger la demande consciente du décideur vers ses besoins réels. D'après une étude réalisée auprès de 70 PME de la région midi Pyrénées, les besoins en informations exprimés par les dirigeants concernent, dans l'ordre, le marché, (clients, fournisseurs), l'identification des concurrents, et la recherche de méthodes de détection de nouveaux marchés.

### 3.2 Collecte d'information potentiellement utile à l'analyse

Tout l'art du collecteur d'informations réside dans la mise en relation entre les besoins des utilisateurs de l'intelligence économique et les recueils, les techniques, liées à la récolte des informations. Ainsi, les auteurs montrent que savoir où collecter l'information désirée, sous entend de connaître les caractéristiques de l'information. La collecte est une des phases des plus importantes des activités de veille dans les pratiques d'intelligence économique. La masse informationnelle immense se caractérise également par une pluralité de formats, de lieu d'obtention, etc.

Cette étape est complètement traitée par le prototype développé. La collecte d'information se base sur des sources les plus diverses comme: les bases en ligne, les Cd, le Web visible et invisible, les news, les brevets, la presse, les traces de connexions aux sites, les bases internes.... Nous nous intéressons seulement aux ressources de type textuelles. Le processus de collecte et de préparation de l'information en vue de son analyse est défini par les étapes suivantes :

#### Identification des sources

Ce traitement consiste à identifier les sources qui sont susceptibles de contenir l'information recherchée. Sur des serveurs comme Dialog, qui proposent des centaines de bases métiers, une indexation globale facilite cette tâche. Sur le web visible les moteurs de recherche offrent une aide similaire. Dans tous les autres cas, seule une bonne connaissance de l'offre permet de mener à bien cette phase dite de « sourcing ».

#### Identification de l'information

Cette étape permet de rechercher l'information et d'identifier celle qui est utile dans les sources sélectionnées pour alimenter l'analyse. Cette recherche se base principalement sur des mots clés et sur le moteur de recherche d'information intégré à chaque système utilisé.

**Remarque :** Dans le cadre de nos travaux de recherches et le point d'évolution de nos recherches, nous supposons que les deux étapes précédentes (Identification des sources, Identification de l'information) nous fournissent des résultats pertinents. Notre principal travail se base sur les étapes qui suivent (Structuration de l'information et homogénéisation, Matrice 3D, Génération de la base) et qui sont axés sur l'extraction de connaissances.

#### Structuration de l'information et homogénéisation

Des métadonnées spécifiques à chaque base sont établies afin de décrire au mieux et sous forme standardisée le format natif des documents. Pour chaque base structurée ou semi-structurée on définit son descripteur de format spécifique qui permet de l'interfacier avec notre plate forme de traitement de l'information. Voici un exemple d'un extrait de descripteur générique de la base SCI (web of science).

Nom	Abréviation	Champ	Visible
Numéro	NO	UT	True
Type_pub	PT	PT	False
Auteur_1	AL	AU	True
Auteur_C	AC	AU	True
Titre	TI	TI	True
Journal	JN	SO	True

Figure3.: Descripteur de la base SCI

Il reste à résoudre les problèmes des synonymies, du filtrage et du choix de la granularité pour chaque variable. Ceci peut être réalisé grâce aux notions d'équivalence et d'inclusion entre certains des items identifiés. Exemples :

- Equivalence : Jean Dupont -> J. Dupont (regroupement des synonymies)
- Inclusions : Toulouse -> France -> UJE -> Europe (changement de granularité)

#### Matrice 3D ou liste de relations évolutives

Ces structures de données, qui sont équivalentes, permettent de mettre deux à deux en relation les différentes variables de l'analyse en intégrant systématiquement la variable temporelle. Notre système propose un module de génération de ces structures (matrices pour Téralogie et listes pour Xplor). Les différents types de relations (appelées pré-connaissances) qui sont générés sont :

- Liens non orientés : Présence / absence, Contingence, Co-occurrence, Co-occurrence pondérée, Adjacence, ...
- Liens orientés : Citation, Acquisition, Action, Inclusion, ...

Pour le moment, les matrices que nous avons expérimentés dans notre prototype sont les matrices de Co-occurrence et Présence / absence. Dans la figure3. Nous présentons la visualisation d'une matrice de co-occurrence sous forme de réseaux de liens.

### Génération de la base

Un ensemble d'outils intégrés à la plate-forme Téralogie permet de générer automatiquement la base de données des relations (pré-connaissances). Il est possible de sélectionner les variables, les filtres (positifs ou négatifs), les équivalences (synonymes et/ou granularités) ainsi que les matrices de croisement.

Un premier outil génère la table des variables figure4. et celles des items, un second génère les listes des relations Figure5. instanciées correspondant aux matrices choisies (table des relations).

	Base	Titre	Date	Descipteur	Mot_clef	Auteur	Organisme	Adresse	Ville	Pays	Frq	q <sub>00</sub>	Liste de mots	Synonymes	Charge?
Base	: BA										1	0	X-BA	BA	
Titre	: TI										1	0	X-TI	TI	
Date	: DP										1	0			
Descipteur	: DE										1	0	X-DE	DE	
Mot_clef	: MH										1	0	X-MH	MH	
Auteur	: AU										1	0	X-AU	AU	
Organisme	: OR										1	0	X-OR	OR	
Adresse	: AD										1	0	X-AD	AD	
Ville	: VI										1	0	X-VI	VI	
Pays	: PA										1	0	X-PA	PA	

Figure4.: Choix et génération des variables et des items

	Base	Titre	Date	Descipteur	Mot_clef	Auteur	Organisme	Adresse	Ville	Pays	Frq	q <sub>00</sub>	Filtres	Synonymes
Base	: BA										1	0	X-BA	BA
Titre	: TI										1	0	X-TI	TI
Date	: DP										1	0		
Descipteur	: DE										1	0	X-DE	DE
Mot_clef	: MH										1	0	X-MH	MH
Auteur	: AU										1	0	X-AU	AU
Organisme	: OR										1	0	X-OR	OR
Adresse	: AD										1	0	X-AD	AD
Ville	: VI										1	0	X-VI	VI
Pays	: PA										1	0	X-PA	PA
Reference	:										1	0		

Figure5.: Choix et génération des relations

### 3.3 Requêtage

L'interrogation s'effectue selon deux approches :

- Soit en utilisant des requêtes prédéfinies, 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, n'a pas été encore expérimentée, 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 (co-occurrences)
  - Question posée (évolution du top 10 des entreprises sur les 3 dernières périodes).
  - Support (matrice de co-occurrence 2D : Entreprises X Temps)
  - Requête choisie (évolution d'une variable)
  - Paramétrage (Variable : Entreprises, Filtre : Top 10 + Définition des 3 dernière périodes)
- Exemple 2 :
- Corpus analysé (ensemble de brevets)
  - Métrique (co-occurrences)
  - Question posée (évolution de la stratégie des co-dépôts de brevets d'un panel d'entreprises sur 2 périodes).
  - Support (matrice de co-occurrence 3D : Entreprises X Entreprises X Temps)
  - Requête choisie (évolution des relations dans une variable)
  - Paramétrage (Variable : Entreprises, Filtre : le Panel + Définition des 2 périodes)

### 3.4 Diffusion de l'information pertinente

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.

Dans la Figure6., nous présentons les différentes expérimentations faites avec le prototype développé :

Le graphe numéro 1 représente la fonction Soleil qui permet d'interagir avec toutes les dimensions d'un élément ou un groupe d'éléments. Elle représente les relations existantes entre la variable à analysé et celles qui lui sont liées ; dans notre exemple nous avons un groupe d'auteurs et les différentes dimensions qui lui sont liés tel que Date, Publications, Affiliations ...etc. Ainsi l'utilisateur pourra sélectionner la dimension liée aux auteurs qui veut étudier à une période donnée.

Le graphe numéro 2 représente la matrice de co-occurrence Auteurs-Auteurs à une période donnée. Elle fait apparaître l'ensemble des collaborations, leurs structures, les personnes qui les animent ainsi que les différentes équipes isolées du domaine.

Le graphe numéro 3, permet de faire apparaître les collaborations internationales entre auteurs et indirectement entre Pays à une période donnée.

Le graphe numéro 4, représente une représentation en ligne des auteurs filtrés sous forme de camembert à une période donnée.

Les graphes numéro 5, 6,7, ceux sont les différentes représentations de l'évolution de la productivité scientifique de chaque auteur du groupe filtré et de détecter les auteurs qui produisent le plus.



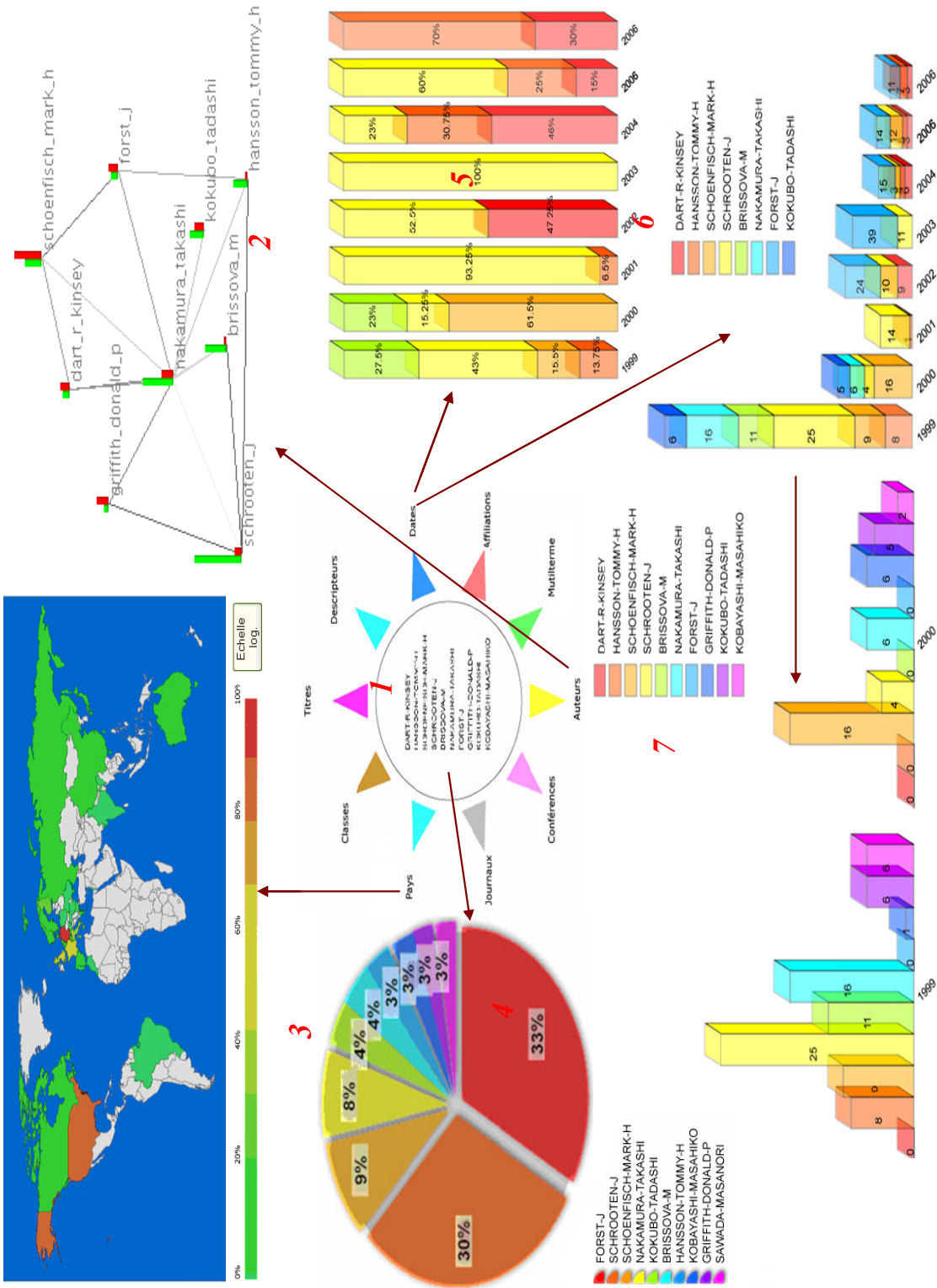


Figure 6.: Sorties graphiques du prototype développé

## 4 Conclusion

Avec ce prototype, nous avons validé quelques propositions à savoir la collecte des sources puisque nous arrivons à couvrir diverses sources d'information électronique disponibles à l'heure actuelle, à savoir : les bases documentaires en ligne ou sur CD/rom (comme Medline, Inspec, Current contents, Biosis, Pascal, Sci, Chemical abstract, ...), les pages web, les news groups, les traces connexions aux sites, la presse, les brevets (Esp@cenet, Uspto, Derwent, Inpi, ...), les dépêches d'agences, le non structuré et bien entendu toutes les bases structurées. Comme le relationnel est la principale source d'information stratégique, l'organisation des données que nous proposons va permettre aux utilisateurs de naviguer dans leurs analyses par des techniques qu'ils maîtrisent maintenant très bien (InterNet, les statistiques descriptives, le filtrage, les fonctions de reporting). Comme la génération d'une nouvelle base, sur un sujet ciblé, ne prend que quelques minutes, un décideur, même distant, pourra être informé très rapidement sur les évolutions récentes de son environnement et sur les stratégies induites par les ruptures ou les structures émergentes mises à jour. Enfin, le coût de cette approche peut être limité : de nombreuses sources sont gratuites, un analyste n'est plus nécessaire et le serveur est partagé par de nombreux utilisateurs.

## 5 Bibliographie

- [1] MARTRE H., "Intelligence économique et stratégie des entreprises", travaux de groupe dirigé par Henri Martre pour le commissariat Général du Plan, La documentation Française, 1994.
- [2] CHUNG, W, CHEN, H JAY, F, NUNAMAKER JR. " Business Intelligence Explorer: A Knowledge Map Framework for Discovering Business Intelligence on the Web". Department of Management Information Systems, Eller College of Business and Public Administration, The University of Arizona. (2003).
- [3] CARVALHO, R. AND FERREIRA, M. "Using information technology to support knowledge conversion processes," Information Research, vol. 7. (2001).
- [4] CHOO, C. W. "The Knowing Organization". Oxford: Oxford University Press. (1998).