

# X-Plor : Système d'analyse relationnelle en ligne

Ihème Ghalamallah

Institut de Recherche en Informatique de Toulouse, IRIT-SIG  
Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 9  
ghalamal@irit.fr

## 1 Introduction

Dans le contexte de la veille stratégique (VS), Tétralogie (Dousset, 2003) est un outil particulièrement bien adapté aux analyses macroscopiques, il permet en effet de dégager les signaux forts, les signaux faibles et les tendances à partir d'un corpus de documents collectés sur un sujet précis. L'information élaborée, qui en est issue, représente une synthèse obtenue par diverses méthodes d'analyse de données et diffusée via des visualisations graphiques. Mais à l'issue des très nombreuses analyses stratégiques que nous avons déjà réalisé avec ce logiciel, il est apparu que les utilisateurs finaux des analyses produites veulent, en complément de l'aspect global et stratégique (connaissances générales), des zooms plus précis sur certains points et ce afin de satisfaire leurs besoins spécifiques (connaissances ciblées) d'informations sur des éléments qu'ils ont déjà identifiés (concurrence, marchés, nouveaux produits ou procédés, partenaires potentiels, ...) ou afin d'en découvrir d'autres. A posteriori, de nombreux experts ou décideurs sont demandeurs de plus de détails dans l'approche des éléments constituant traditionnellement leur environnement. Notamment, pour tout ce qui concerne leur vocabulaire spécifique, les acteurs qu'ils côtoient, les marchés qu'ils convoitent, les alliances qu'ils projettent.

Nous proposons donc de compléter nos analyses macroscopiques par 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 (IE). Ce système permet d'effectuer des analyses stratégiques sur des corpus d'information textuelle issus 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, et donne la possibilité aux décideurs de mener par eux même leurs investigations sans la participation d'un analyste ou d'un expert. L'interaction (Système - décideur) intervient au niveau du choix des variables, de l'extraction et de la sélection des données utiles, du choix des analyses à déployer et de la visualisation des résultats.

## 2 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 (Chung, 2002). Carvalho et Ferreira (2001) 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 (Choo, 1998). 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, où 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 de Martre « l'IE est l'ensemble des actions coordonnées de recherche, de traitement, de distribution et de protection de l'information utile aux acteurs économiques en vue de la mise en œuvre de leurs stratégies individuelles et collectives ». Cette définition englobe quatre étapes à savoir: (1) formulation de besoin, (2) collecte d'information, (3) analyse, (4) diffusion des connaissances.

Chung, (2002), s'intéresse aux étapes 2, 3 et 4 du cycle de l'IE, qui sont les seules étapes qui peuvent être automatisé en utilisant les technologies de l'information. Dans le cadre de notre proposition, notre prototype couvre les trois traitements du cycle IE à savoir :

- Collecte d'information,
- Requêtage,
- Diffusion des connaissances.

### 3 Conclusion

Avec le prototype Xplor que nous avons développé qui couvre les trois étapes du cycle de l'IE, 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.

### References

- Carvalho, R. and Ferreira, M. (2001). "Using information technology to support knowledge conversion processes," Information Research, vol. 7.
- Choo, C. W. (1998). *The Knowing Organization*. Oxford: Oxford University Press.
- Chung, W, Chen, H Jay, F, Nunamaker Jr. (2003). *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.
- Chrisment, C., T.Dkaki, J.Mothe, B.Dousset (1997). *Extraction et synthèse de connaissances à partir de bases de données hétérogènes*, Ingénierie des Systèmes d'Information, vol. 5, n°3, pp. 367-400.
- Dousset, B. (2003). *Intégration des méthodes interactives de découverte de connaissances pour la veille stratégique*, Mémoire de habilitation à diriger les recherches, Université Paul Sabatier, Toulouse.
- Frank S.C. Tseng, Annie Y.H. Chou (2005). *The concept of document warehousing for multi-dimensional modelling of textual-based business intelligence*, National Science Council, TAIWAN.
- Sosson, D. , M. Vassard, B. Dousset (2001). *Portail pour la navigation en ligne dans les analyses stratégiques*. *Veille stratégique, scientifique et technologique : VSST01*, Vol 1, pp 347-358, (Barcelone, Espagne).