# XPLOR EVERYWHERE – A TOOL FOR COMPETITIVE INTELLIGENCE ON THE WEB AND MOBILE

**Anass EL HADDADI (*, **), Bernard DOUSSET (*), Ilham BERRADA (**)**
anass.el-haddadi@irit.fr ; dousset@irit.fr ; iberrada@ensias.ma
(*) IRIT, SIG, Université Paul Sabatier, route de Narbonne, 31062 Toulouse cedex 04
(**) ENSIAS, Equipe Al BIRONI, Université Mohamed V – Souissi, B.P. 713 AGDAL, Rabat - Maroc

**Mots clés :**
Téléphonie Mobile 3G, Intelligence Economique, découverte de connaissance, Ontologie.

**Keywords:**
Mobile Technology, Competitive Intelligence, Knowledge discovery, Ontology.

**Palabras clave:**
Telefonía Móvil, inteligencia competitiva, concimiento discovry, ontologia.

## Abstract:

Competitive intelligence (CI) is the set of coordinated researches, treatments and distribution of useful information to stakeholders' to-wards action and decision making. In order to enable users to search, monitor, validate and rebroadcast strategic information, we provide our new tools (Xplor EveryWhere), which can be helpful for them in their executive travels. Firstly, XEW is based on a specific domain ontology describing the actors' networks, and secondly, it is auto-secure managed. In this paper, we focus on the architecture of the CI system – XEW to describe our approach and the way we treat different data-sources (Patents, Paper, etc.).

# 1 Introduction

Nowadays, companies are faced with external risk factors linked with an increased competition market place – we know that markets are extremely dynamic and unpredictable: new competitors, mergers and acquisition, sharp price cuts, rapid changes in consumption patterns and values, weak brands and their reputation…

CI is a discipline to better anticipate risks and identify opportunities. Fifteen years after the canonical definition proposed by Martre [1], CI is still a concept with unstable borders. The last few years have seen a multiple definition of CI: from definitions oriented mapping process practice of CI, strategic vision of the CI to others including the concepts of knowledge management, collective learning and cooperation [2].

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

There are several types of sources to watch (the scientific databases, the patents databases, the media, press, blogs, RSS, Internet, intranet, forums...). This heterogeneity of sources is a real problem for CI process, particularly when you analysis a topic from different points of view (science, technology, news, etc...). Hence the need to propose an approach, to resolve this heterogeneity [3] [4]. Multivariate techniques are currently well controlled for all available quantitative data on information systems or business administration, on condition that DBM be suitable, the database schema be adapted and data be of the highest quality (homogeneous, current, complete...). It is always possible to extract the relevant data to a database custom built for multidimensional searches, conducting, all adjustments necessary to this operation. Consequently, textual data from all electronic sources is difficult to implement: data sources have different formats or are even unstructured, they are distributed, heterogeneous and encountered many particular / singular cases. To standardize the multidimensional search text data from all sources, we propose a unified structure for storing all relationship items encountered in documents analyzed. This method allows the construction of tridimensional analysis between variable and a time. Since 2001[5], a first tool Xplor has been suggested to upload this type of structure in client server, to perform a custom search tools through various graphic restitution of the results. All text data are then in the same structure and are therefore common tools of interactive investigation. An improved version of Xplor appeared in 2007[6].

In order to enable users to search, monitor, validate and rebroadcast strategic information, we provide our new tools (Xplor EveryWhere), which can be helpful for them in their executive travels. Firstly, XEW is based on a specific domain ontology describing the actors' networks, and secondly, it is auto-secure managed.

# 2 Literature view

## 2.1 Competitive intelligence

CI is not an innovative practice created exclusively for the development of technology and information society. Each country, depending on its history, its cultural and religious past, its location, its natural or industrial resources, has led, over the centuries, a process that can be called CI.

A broad definition of CI is a set of coordinated actions of research, treatment and distribution of useful information to stakeholders to enable the action and decision making. More, CI is a legal business practice, as opposed to industrial espionage which is illegal. Generally, the focus is on the external business environment [8, 9].

The concept of CI is now in a emerging phase, but nevertheless it is still necessary to produce a reference framework based on a multidisciplinary approach [10, 11]. In our research team, we coordinate the process of CI around three concepts namely strategic analysis, environmental scanning and information system. Fig. 1 shows a hierarchy illustrating the coordination of these three concepts.

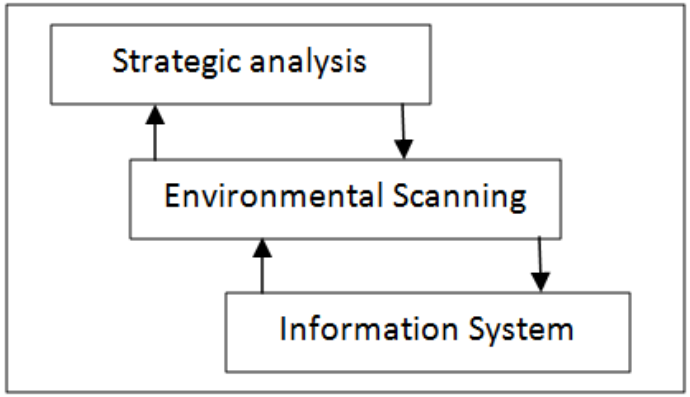



*Figure 1- Coordination of concepts related to CI [12]*

- Strategic analysis: reveals the informational needs of the company on its environment in order to reduce the uncertainty and to facilitate decision making.
- The environmental scanning: will guide and steer the process of collecting, processing, analyzing and disseminating information to meet the expressed needs.
- The information system: is used to support the various activities from data collection to data analysis and reporting.

## 2.2 Competitive Intelligence process

The CI system integrates processes of strategic analysis and environmental scanning based on information systems (cf. Figure 2).

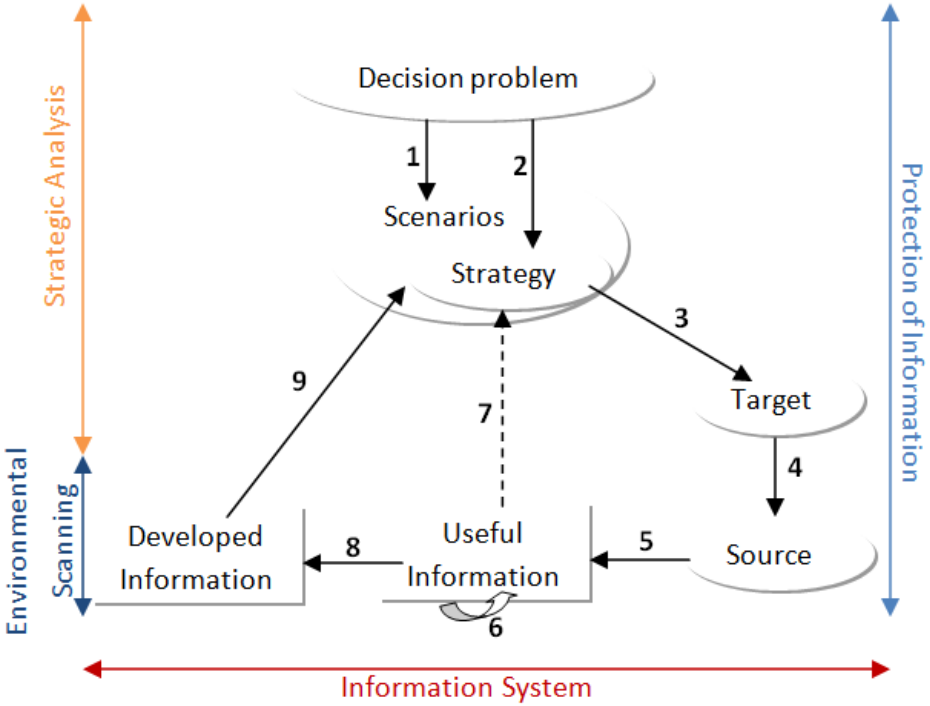



*Figure 2- Competitive Intelligence process*

1. Develop,
2. Choose,
3. Identify & Prioritize,
4. Identify & select,
5. Collect & Evaluate,
6. Organize & Remember,
7. Validate & Stream,
8. Analyze & Interpret,
9. Validate & Spread

The four essential steps of the intelligence cycle are included in this schematic presentation:

- understanding the need,
- researching and gathering information,
- processing information,
- disseminating information.

## 2.3 Source of information for CI

The activity of identifying sources of information consist in list any formal and informal sources that may contain information useful to the theme of the analysis. In the context of CI, sources of information can be of two types:

## 2.4 Formal sources

The information is called formal when it is published in printed papers, computer, microfilm ... It can be formal or not, but it is in all cases of information directly accessible and usable. This type of sources corresponds to white information. Formal sources are composed mainly of the press, television, radio, books, data banks and CD-ROM, patents, legal information, studies conducted by public or private providers, Internet. These sources have the advantage of being reliable and fairly comprehensive, easy to access.

In the context of environmental scanning, the most viewed databases are scientific, technological, and regulatory. Among the most interesting databases, we can cite Factiva[1] in the economic field, Inspec[2] in the physical field and Kompass[3] Europe oriented business, which represents a Pascal multidisciplinary source, PubMed[4] specializes in the medical field as an example.


*Figure 2 - A formal source for CI*

## 2.5 Informal sources

The informal information consists of all non-formal and not directly available sources. It is therefore necessary to undertake steps direct from owners assumed that information. This type corresponds to gray information. These sources can be exhibition, vendors, seminars, conventions, and clubs: we exchange

[1] www.global.factiva.com
[2] http://web.ebscohost.com
[3] http://fr.kompass.com/ip
[4] http://www.ncbi.nlm.nih.gov

information, communicate there. The information flow is often of great strategic value, such as those of competitors, commercial and financial communications, newsletters, etc... And the internal sources of the company representing over 80% of such information as a decision-maker, personal sites, research studies conducted by a group students or PhD student.

However the great problem is the analysis of multi-data sources. So we developed an original approach to treat this heterogeneity of information sources [3] [4]. This originality is based on specific description appropriate for each source (top-level metadata) and a generic description (second-level metadata). We can simultaneously handle heterogeneous body of information in their native format.

## 2.6 Evaluation of competitive intelligence tools

"Fuld & Company" is a global leading CI company specialized in analyzing markets and competitive intelligence. Since 1998, this company has always valued the technological tools dedicated to CI. The final evaluation report of "Fuld & Company" (cf. Intelligence Software Report 2008-2009) focuses on tools for advanced information processing used in the context of CI approach. It is based on self-evaluation of 480 companies in the world.

The authors differentiate CI tools and Business Intelligence (BI) tools. The BI tools focus on data warehousing and analyzing of quantitative data which are exclusively internal to the company's. However, the CI tools concern the collection and analysis of qualitative data (internal and external to the company).

| Evaluation of CI Tools 2008-2009 | Expression of needs | Collection of information | | Processing and analysis | Distribution | Average |
|---|---|---|---|---|---|---|
| | | **Formal** | **Informal** | | | |
| **Brimstome** | 4 | 3 | 3 | 4 | 3 | 3,4 |
| **Cipher** | 5 | 3 | 4 | 3 | 4 | 3,8 |
| **Comergence** | 5 | 3 | 5 | 4 | 5 | 3,8 |
| **Comintell** | 3 | 3 | 3 | 2 | 2 | 2,6 |
| **Cymfony** | 4 | 5 | 5 | 4 | 5 | 4,6 |
| **Global Intelligence** | 3 | 2 | 3 | 2 | 2 | 2,4 |
| **KB Intelligence** | 1 | 2 | 1 | 1 | 1 | 1,2 |
| **Netro-City** | 1 | 3 | 4 | 3 | 3 | 2,8 |
| **QL2 Software** | 3 | 3 | 3 | 2 | 1 | 2,4 |
| **Strategy Software** | 5 | 3 | 5 | 2 | 4 | 3,8 |
| **Temis** | 3 | 2 | 2 | 2 | 2 | 2,2 |
| **Traction Software** | 3 | 4 | 3 | 1 | 4 | 3 |
| **Wincite** | 5 | 3 | 2 | 2 | 4 | 3,2 |

*Table 1 - Evaluation of CI tools 2008-2009 [12]*

The assessment tool is based on over 120 criterias (functions) covering the five stages of the information processing cycle in the context of CI, namely: "Expression of needs, collection of published information, and collection of primary information, processing and analysis, broadcast".

Thirteen tools have been selected among the 480 identified. They are tools that cover a large part of the CI process stages. Each tool is then evaluated according to criterias associated with these five steps. Each step is rated on a scale of 1 to 5: “Poor, Right, Good, Very good, excellent”. The report employs the comparative tools used by the traditional process of information processing (Table 1).

The tools tested do not optimally cover all cycle stages. The percentage of the cycle’s coverage of the cycle is on average between 50% and 80%. These tools are essentially based on techniques for finding web information using intelligent agents. There are only four of them (QL2 Software, Temis, Cymfony, Traction Software) which are based text Mining and two (Brimestone, Cymfony and Netro City) which are based on visualization techniques and reporting. Our goal is to take into consideration evolutive criterias to improve our CI system.

# 3 Xplor EveryWhere System Architecture

The information is the raw for CI process. However, most sources of information contain information potentially useful to the decision (section 2.3). Therefore, companies need systems to extract and analyse information to match entrepreneurial prospective on specific areas: aerospace, pharmaceutical… (e.g. identifying technology suppliers, competitors, identify merger and acquisitions of companies appearing in the business press).

Our contribution consists in the proposed of an information system adapted to the needs of the CI approach. The objective of this system is to provide a reference methodology in order to collect, process and analyse information. Our system will observe and analyse the information environment of a decision problem in all its dimensions. The approach combines two methods: knowledge discovering in text (KDT) and environmental scanning.

The dynamic aspect is vital to any analysis in the context of CI. These dynamics include continuous monitoring of the business environment in order to detect its changes and developments. The proposed information system, based on an exploratory multivariate analysis model: “the relational aspect and the time dimension”, which we call Xplor [12], plus the mobility which we call Xplor EveryWhere . It is based on extracting knowledge from textual data by analyzing relational data and their evolution. This model allows time specification to situate events, strategies and actions as well as in:

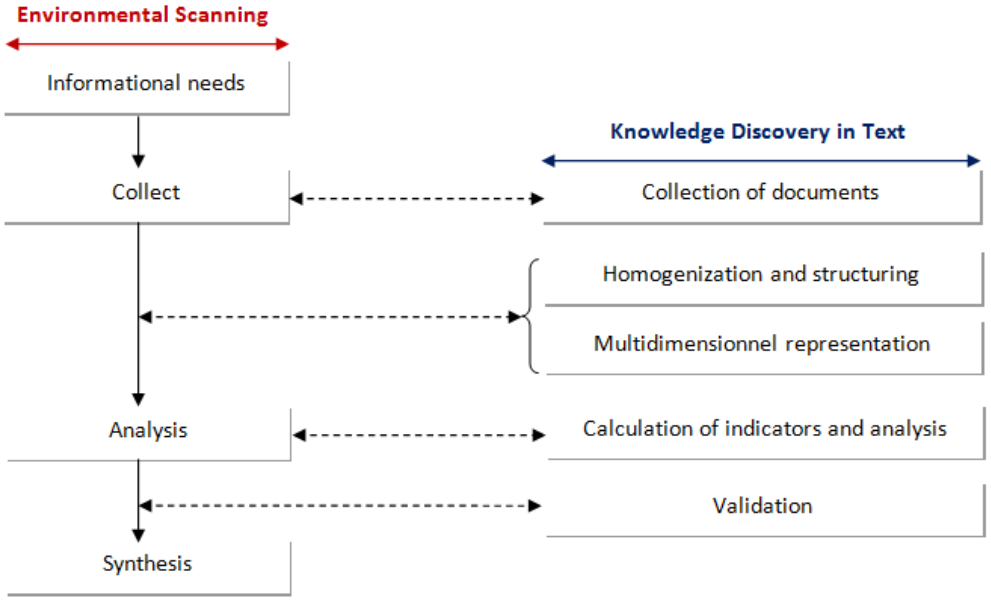



*Figure 3 - Coupling of Environmental scanning and KDT [12]*

- The past by reconstructing the chronology.
- The present-oriented time.
- The future advance for everything concerning the successive organizations of a network, such as partnerships, alliances, mergers, acquisitions, co-citations, co-signatures, co-occurrences of all kinds.

## 3.1 System architecture

The architecture of our platform consists in five main services:

- Monitoring Service, which is to collect information (scientific database, Patents database, RSS, Blogs, etc.).
- Homogenization and structuring Service, through our tool "TETRALOGIE"[3].
- Administration Service, based on a specific domains ontology describing the actors' networks, and is auto-secure managed.
- Analysis Service
- Reporting Service.

### 3.1.1 Monitoring Service

The first phase of the environmental scanning begins by selecting a source of information formally or informally. Then, depending on the source and the need for the user, the request is made, the collection of documents returned by the source will be analyzed. If the results are considered relevant by the user, then conduct the business of collecting such documents or the request will be reformulated. In the case where the target is the entire corpus, validate and stores.

### 3.1.2 Homogenization and structuring Service

In the proposed approach, we assume that the documents contained in the corpus can be targeted from heterogeneous sources. The heterogeneity of sources can consist in the format, language, etc... It is therefore to solve problems of semantic type (name clash or attribute type, no value ...), but also structural type (untagged documents, HTML) or syntactic. To solve this problem, our approach is based on the principles of extracting information defined by [3]. This approach allows extracting predefined information from text documents, where the location of information to be extracted is marked or separated by strings. This solution allows us to:

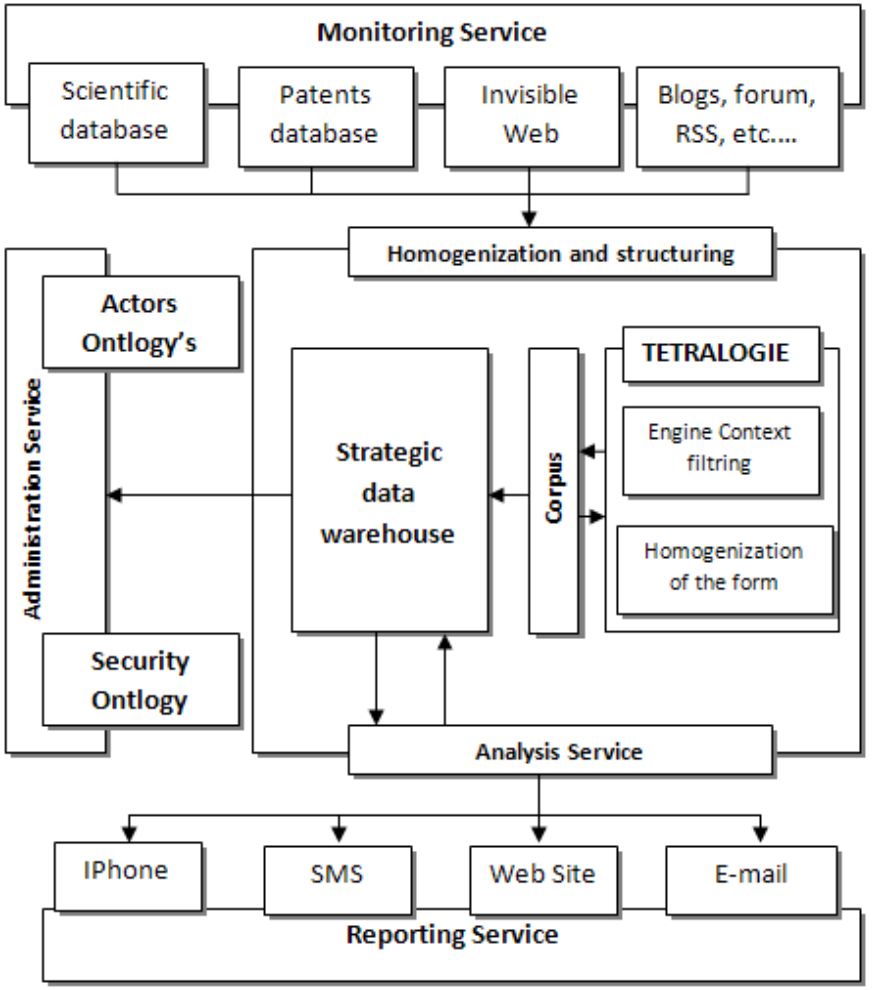



*Figure 4 - Xplor EveryWhere Architecture*

- Define a unified view of documents in the corpus target
- Manage the various conflicts such as: “semantic, syntactic, inclusion and specificity”.

The unified view of the corpus corresponds to a structured, predefined, logical representation of collections sets. Its definition is based on the consideration of specific and generic format descriptors (cf figure 6) [4].

### 3.1.3 Administration service

The need for security and management (administration service) in a CI platform raises ham the strategic nature of information conveyed. Such security cannot be considered as an additional option that a CI platform can provide just in order to be distinguished from one another. Especially as the leak of this information is not the result of inherent weaknesses in corporate computer systems, but above all it is an organizational issue.

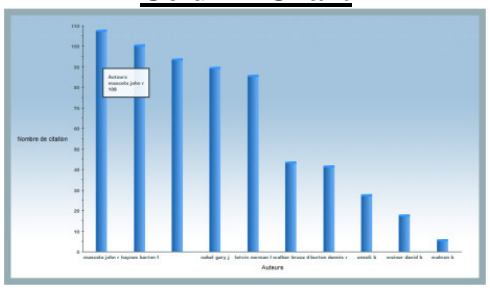
So the issue is how to protect both data (CI as a product) and treatments (CI as a process). Thus, we propose a data access control model and treatment in CI platforms [2010] where all tasks are only executed by authorized users and, thus enhance the privacy related with manipulation of data during the lifecycle of a CI process.

### 3.1.4 Reporting Service

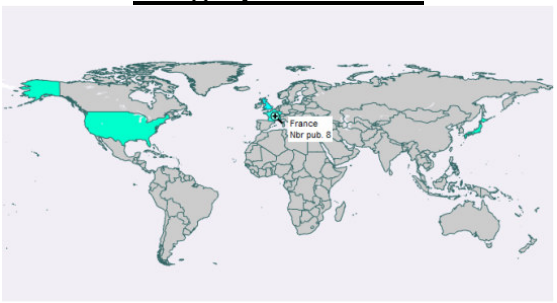
The « reporting » is the last and very important service to accomplish by CI process. In this level we propose 4 types of services: Phone Service, SMS Service, Web Site Service and E-mail Service. With these different services, it is possible to accede to strategic information.

In order to ease the navigability on the strategic information, we intend to integrate specific visualization technique to each type of request like: evolutionary histograms, geographical charts, social networks, profile networks, semantic networks and the international networks (cf. figure 7).

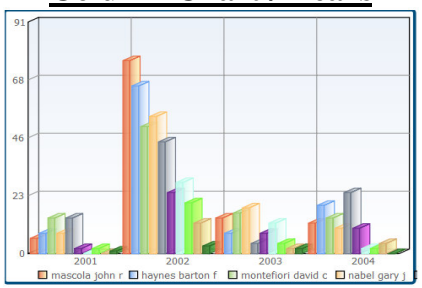
**Column Chart**



**Geographical Chart**



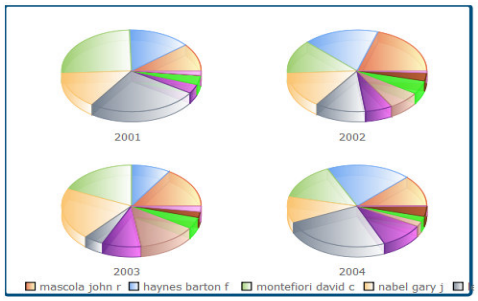

**Column Chart / Years**

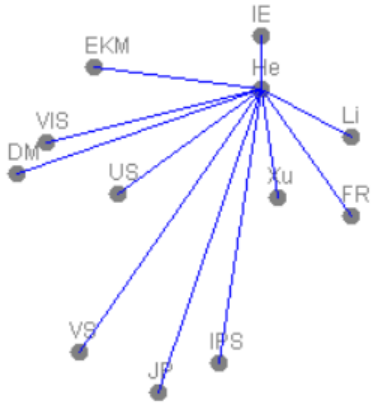



**Pie Chart / Years**

**Profil Network**

**Line Chart / Years**



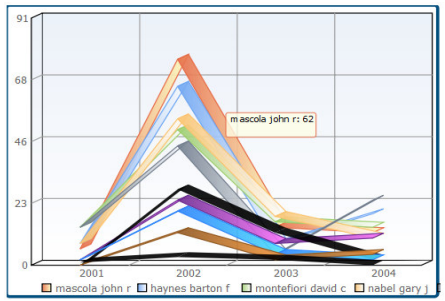




*Figure 5 - Various representation of reporting*

With the possibility to navigate between different types of networks figure 8:

- The Social networks: we have relationship between the different authors, inventors, research teams, companies, etc… And the evolution of relations.
- The semantic networks: we have the relationship between keywords in a domain, the evolution of research topics.
- The international networks: we have the international collaboration between countries.

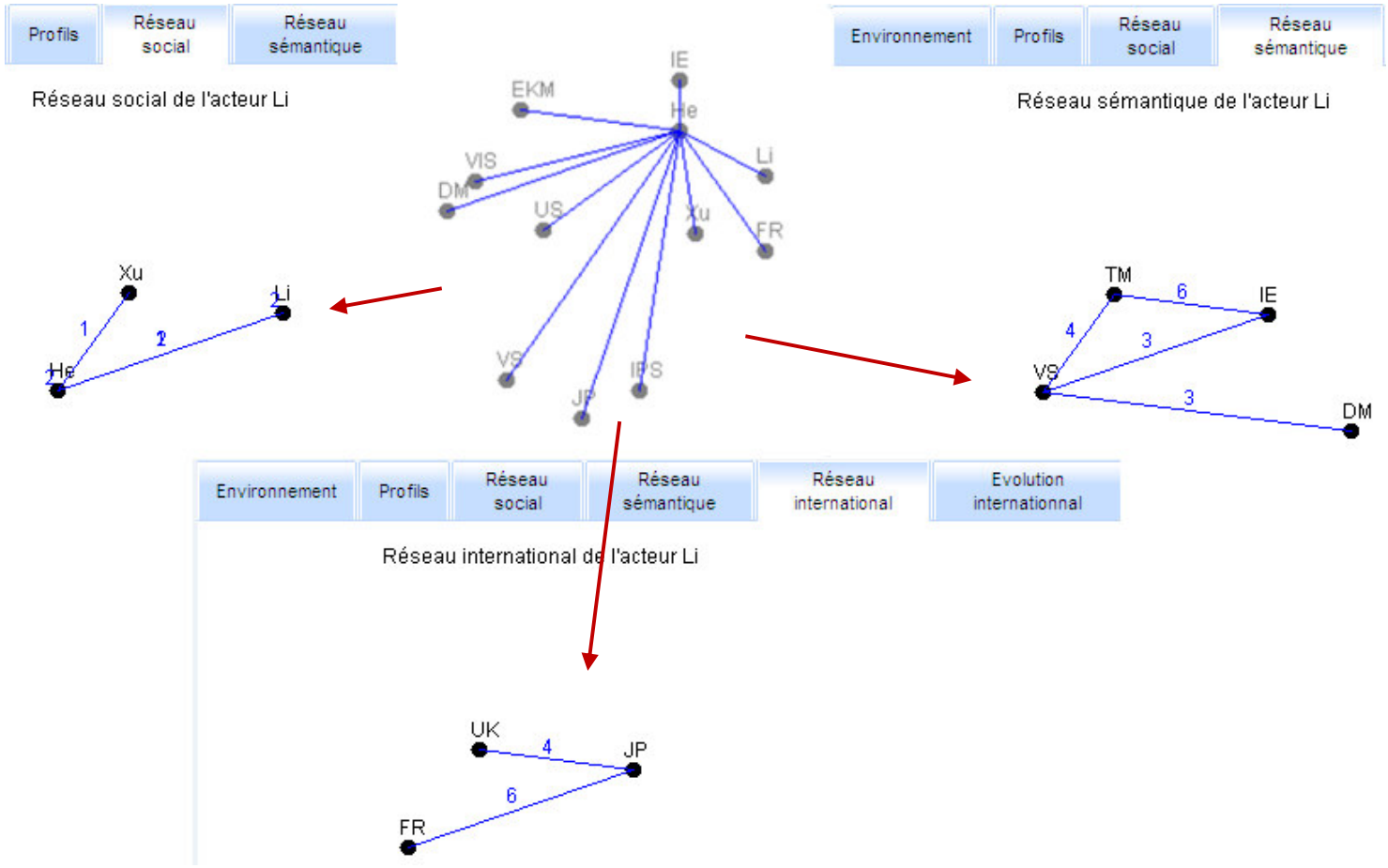



*Figure 6 - The possibility to navigate between different types of networks*

# 4 Conclusion

In this paper, we present a new competitive intelligence system tool, dedicated to cover all stages of discovery, extraction and data management. Without for getting the criteria proposed by “Fuld & company”, nowadays we get to handle any type of information (formal or informal). And with Xplor EveryWhere we get completed the reporting service, especially the aspect of mobility. Lastly with this system, it’s possible to:

- View updated information as we have access to our strategic database server in real-time, itself fed daily by watchmen.
- Raise information "field" at trade shows, customer visits or after meetings.
- Request information specific emergency will be posted by the watchmen.

With the evolution of technology, such a portal for mobile will enable us to increase efficiency and responsiveness because at any moment, it is possible to access all strategic information by markers itself can be information back very quickly "field" which may possibly trigger other strategic analysis.
However, one major problem is the protection of strategic and sensitive data. Thus, to obtain a satisfactory level of security, it is necessary to know the vulnerabilities inherent in this type of network, and develop a specific ontology for security management. That's the subject of our current research, before deployment our platform online.

# 5 References


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

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

[3] **DOUSSET B.** , Intégration de méthodes interactives de découverte de connaissances pour la veille stratégique. Habilitation à diriger des recherches, Université Paul Sabatier, Toulouse, (2003).

[4] **EL HADDADI A., DOUSSET B., BERRADA I., LOUBIER I.,** Les multi-sources dans un contexte d'Intelligence Economique , EGC 2010, P A1-125 A1-136, HAMMAMET – TUNISIE 26 AU 29 JANVIER (2010).

[5] **SOSSON D., M. VASSARD, DOUSSET B.,** Portail pour la navigation dans les analyses stratégique , VSST'01, Vol 1, pp 347-358, Barcelone – Espagne, (2001).

[6] **GHALAMALLAH I, GRIMEH A, DOUSSET B**, Processing data stream by relational analysis, EUROPEAN WORKSHOP ON DATA STREAM ANALYSIS, March, 14-16, CASERTA, ITALY, N° 36 , pp 67-70 (2007).

[7] **EL HADDADI A, DOUSSET B, BERRADA I., and KASSOU I,** Construction d'une ontologie de domaine fondée sur le text mining, colloque Veille Stratégique et technologique, 2009.

[8] **HAAG S.,** Management Information Systems for the Information Age , Third Edition. McGraw-Hill Ryerson, (2006)

[9] **GILAD B.,** The Future of Competitive Intelligence, Contest for the Profession's Soul , Competitive Intelligence Magazine, 11(5), 22, (2008).

[10] **FAVIER L.,** recherche et application d'une méthodologie d'analyse de l'information pour l'intelligence économique application à un centre technique du secteur de la plasturgie , Thèse de doctorat de l'université Lumière Lyon II, France, (1998).

[11] **ALABDULSALAM M., PATUREL R., ,** Outil permettant aux pme/pmi l'accès a la demarched'intelligence économique , XVème Conférence Internationale de Management StratégiqueCIMS, Suisse, (2006).

[12] **GHALAMALLAH I ,** Proposition d'un modèle d'analyse exploratoire multidimensionnelle dans un contexte d'intelligence economique, doctorat de l'université de toulouse, 18 décembre (2009).

[13] **MARON M., KUHNS J. ,** On relevance, probabilistic indexing and information retrienval , Journal of the Association for Computing Machinery, pages 216-244, (1960).

[14] **SALTON G.,** A comparison between manual and automatic indexing methods. Journal of American Documentation. Pages 61-71, (1971)..

[15] **ROCCHIO J. ,** Relevance feedback in information retrieval. Prentice Hall Inc., Englewwod Cliffs, NJ, (1971).

[16] **SALTON G., MCGILL M.,** Introduction to modern information retrevial. McGrawHill Int. Book Co (1984).

[17] **KLEINBERG J.M.,** Authoritative sources in a hyperlinked environment. Journal of the ACM, pages 604-632, (1999).

[18] **ROBERSTON S.E.,** The probability ranking principle in IR. Journal of Documentation. Pages 294-304. (1977).

[19] **MOTHE J. ,** Modèle connexionniste pour la recherche d'informations – Expansion dirigée de requêtes et apprentissage. Thèse de doctorat, Université Paul Sabatier, Toulouse, (1994).

[20] **BOUGHANEM M., CHRISMENT C., MOTHE J., SOULE-DUPUY C., TAMINE L.** , Connectionist and genetic approaches to perform IR. F. Crestani and G. Pasi edi tors. Soft computing in information retrieval, techniques and applications. Physica Verlag, pages 173-198, Heidelberg. (2000).

[21] **SAUVAGNAT K. ,** Modèle flexible pour la recherche d'information dans le corpus de documents semi-structurés. Thèse de doctorat, Université Paul Sabatier , Toulouse, (2005).