"Information Design" for "Weak Signal" detection and processing in Economic Intelligence: case study on Health resources

Sahbi Sidhom, Philippe Lambert

“Information Design” for “Weak Signal” detection and processing in Economic Intelligence: case study on Health resources

Sahbi SIDHOM (LORIA/KIWI & Nancy Université), **Philippe LAMBERT** (VinaLor, Nancy).

e-Mails : sahbi.sidhom@loria.fr, philippe.lambert@vinalor.fr

Abstract —

The topics of this research cover all phases of the “Information Design” applied to detect and take profit on weak signals in economic intelligence (EI) (or BI: business intelligence). The field of the information design (ID) applies the process of translating complex, unorganized or unstructured data into valuable and meaningful information. The ID’s practice requires an interdisciplinary approach which combines skills in graphic design – writing, analysis processing and editing – , human performance technology and human factors. Applied in the context of information system, it allows to end-users to easily detect implicit topics well known as “weak signals” (WS). In our approach to implement the ID, the processes used cover the development of knowledge management (KM) process in the context of the EI.

A case study concerning information monitoring on health resources is presented using ID process to outline weak signals. French and American bibliographic database corpora are applied to make the connection to multilingual concepts in health watch process.

Index Terms —

Information Design (ID), Weak Signal (WS), Natural Language Processing (NLP), Data Visualization, Information Indexing, Knowledge Management (KM), Economic Intelligence, (multilingual) concept, semantic network.

I. INTRODUCTION

November 26th. (2010), the University of California officially launched its laboratory project on the “Information Design” (ID).

The project aims to develop knowledge exchange between different actors through applications for new media platforms such as Ipads in networks or Iphone Technology. Beyond the innovative aspect of this project, we can only note that the Information Design (ID) is the projection of an important “*Prospective Approach*” in the Anglo-Saxon research world. This point is reinforced by the comparison of the scientific literature on this issue. Since the 70s, research teams have specialized on the connections’ study between the graphical representation of information and its interpretation. One of the representation techniques that tends to develop, is the “*spatial*” information across neuron networks. Especially in France, this approach has been vulgarized at first and developed by the introduction of mind-maps (ie. Mind Mapping) in education research. In recent years, this research focus has been applied to data mining from the Web (ie. Web Mining). It helps to develop new knowledge from large corpora (text themes). This technique is increasingly interested in those leaders who have watch responsibility to detect topics that could have been missed in a linear reading. In the field of EI studies, the implicit properties on corpora analysis take on the name of “weak signals” (WS) (and respectively, the explicit properties are “strong signals”).

The detection of WS allows the watcher to take better account of its environment in a dynamic sense and foresight (ie. “to prepare today for tomorrow”). However, the connection between the ID and the WS detection requires the development of complex methodological processes that represent the topic of this paper.

The first part of this paper defines the meaning of weak signals and process through a strategic approach.

The second part presents the logical in ID process that tends to present varied graph data sets while facilitating the appropriation of the “semantic” properties.

The last part is matched to a case study on “strategic” health watch process for which we used mapping and visualizing information. The study was able to bring out weak signals on corpora of scientific and technical information to qualify under the provision of ID.

II. MAPPING INFORMATION FOR DETECTING THE WEAK SIGNALS (WS)

Anticipating strategic failures is one of the most issues in EI studies. Market volatility, uncertainties on the property and economic change in a time of crisis or an organization for economic recovery are likely to announce future breaks. These breaks may be opportunities or threats in a changing world economics where the faculty of anticipation becomes a powerful strategic advantage for companies.

In 1970, I. Ansoff discussed the concept of WS in its first paper on the subject entitled “*Managing Strategic Surprise by Response to Weak Signals*” [1]. He considers the WS as

corollary of organizational factors in the company, especially due to environmental turbulence compared to the formulation of corporate strategy. In a following paper, he specified the nature of WS, by defining as "*a warning (external or internal) events and developments that are still too incomplete to allow an accurate estimate of their impact and/or to determine a full adapted response*" [2].

In what follows, we determine the theoretical framework and application to WS.

A. Theoretical framework: weak signal (WS)

Ansoff's work that followed, will give to weak signal a strong "*proactive*" value: to capture WS by the decision-maker via the channel of intuition (ie. spontaneous knowledge on the environment) to cause a request for additional information (ie. explicit formulations) on these signals.

Another contribution to the question of weak signals was made by B. Coffman who has worked on various aspects. For him, a WS can be defined as [3]:

1°/ an idea that affects the way trade and the environment in which we work;
2°/ a novelty and surprise in terms of receiving the signal;
3°/ a noise and other signals, sometimes difficult to detect among noise and other signals;
4°/ an opportunity or a threat to the organization;
5 °/ often made fun of by the "knowledge holders" or experts;
6°/ weak signal with a substantial period of time before it matures and becomes a strong signal;
7°/ therefore, this signal represents an opportunity to learn, grow and evolve.

B. Coffman also said that the WS can be of three types [4]:

(i) supra-perceptual signal,
(ii) perceptible signal but not recognized by our mental models, and
(iii) recognized signal by our mental models and by which our behavior changes.

In France, H. Lesca proposed a list of characteristics that define a WS which approximates the Ansoff logic. Namely a signal can be classified as WS if [5]:

(a) fragmentary,
(b) embedded in a mass of useless information (or noise),
(c) an apparent weak and ambiguous meanings,
(d) could not be seen,
(e) an apparent low usability, and
(f) low "palpability".

In synthesis of these definitions and presentations, we consider that a weak signal is characterized by: – a temporal discontinuity in its discovery, but also by the fact that it causes a shift – (or breaking) in the facts found by the receiver to arouse/create measurable interest in the future.
For this reason that researchers found in the notion of breaking (or the "discontinuity") in the information flow and design that any company can engage in a strategic process.
The specificity of the WS is also in its potentiality. If one considers the famous *S-curve* which describes the four phases of the product life (birth, growth, maturity and decline), we can imagine that the WS is precursor of new trend being upstream of the cycle. Hence, the importance of WS detection in a logic of competitiveness.
We note, the term "***signal***" is ambiguous, however: if one refers to the definition in the Treasury of the French language (ATILF) for the word "signal" we found: [**in French**] "*Signe convenu par lequel quelqu'un donne une information, un avertissement à quelqu'un, indique à quelqu'un le moment de faire quelque chose*" [1]. ie. [**in English**] "A sign agreement by which someone provides information, a warning to someone; someone tells the time to do something". But, precisely the opposite that constitutes the "weak signal" in strategic watch process. The transmitter of the information detected as WS do not expect the risk that competitors become aware of potentially innovative nature of the information given. The adjective term "weak" is also a semantic problem. The "weakness" of the signal is opposite to the potential of information designated by this term. Use the term "weak signal" as defined by "a high potential for breaking innovative". Then, we propose to use the term "sign" above the definition will transform the "weak signal" term to "*Latent Warning Sign*" (LWS) term for this designated information in strategic context.

B. Application framework: Knowledge discovery (KD)

For nearly a decade, several teams of researchers across the Atlantic have focused their study on the subject of Information Design (ID). The concept, however, remained latent in France where it began to show interest until recently.
However, visualize information has advantages. M.J. Eppler [6] gives six main reasons to give priority to this channel that:

1°/ motivates the receiver,
2°/ presents new perspectives,
3°/ develops memory,
4°/ encourages the learning process,
5°/ captures the attention of the receiver, and
6°/ allows to structure and coordinate the communications.

Many definitions presented the ID as an art. The art to direct information to create meaning. It is true that graphic productions go with a significant is left to creativity with formatting, colorful, animated and multiform information. However, in addition to the purely aesthetic of this approach, the ID contains intrinsically a new way of thinking about information and could be summarized as did D. Karabeg in "a new approach to information" [7]:
He explained what ID by proposing the image of a bus

—

[1] ATILF – CNRS- Université Nancy 2, Le Trésor de la Langue française, http://atilf.atilf.fr/dendien/scripts/tlfiv5/advanced.exe?8;s=66859215; [visted date: Nov. 20, 2010] :

equipped with “candle flags”. The bus represents a “modern culture” while the candles symbolize “traditional” information (ie. Informing tradition). We observe what the author means by this incongruence (ie. dysfunction) that can be surpassed by the ID. Moreover, “*modern culture*” is producing and consuming large-scale of information. The ID through the development of information technologies would act as a remedy to the problem of chronic “*infobesity*”.
On this epistemological conception of ID is being added a more technique proposed by B. Fry in his thesis entitled “*Computational Information Design*” [8]. The issue of work is to propose a methodology for data visualization and offers a comprehensive set of graphical representations to give sense to implicit relations between (connective) data.
B. Fry presents a classic seven steps in ID process to ensure the transition from data to knowledge [8]:

1°/ acquire: acquisition of data from any medium;
2°/ parse: cutting to provide a structure to data and order;
3°/ filter: filtering to select only relevant data;
4°/ mine: the search where you place the data into a mathematical context;
5°/ show: representation where it is determined, a simple representation of data can take;
6°/ refine: refining to change the simple representation to more and advanced visual renderings, and
7°/ interact: interaction by adding methods for manipulating data through visualization.

Besides the purely aesthetic information (or infographic), the ID is at the crossroads of several fields of scientific applications. We include the fields of visualization techniques to computer graphics for knowledge. Also, increasingly, the ID impact in psychology and semiotics. This is to perfect the cognitive and physiological theories of visual perception and cultural factors that come into account in the process of information visualization.
Ultimately, the ID enables the end-user, usually the expert whose skills enable him to interpret the data represented as graphs, to generate links between data and knowledge. This knowledge discovery is not the ultimate goal of the logic ID. On the contrary, since this new production aims to refocus the attention of the watcher on historical data previously unnoticed. As a result, the first information needs was altered and a new watch cycle will begin on data “unnoticed”.

In the case study, it is to illustrate the relation of ID and knowledge data discovery (KDD). The “Latent Warning Sign” (LWS) is a key component of this new product by the emergence of thematic relations that may improve the strategic watch process.

III. CASE STUDY: THE HEALTH HETEROGENEOUS RESOURCES - PROJECT« CRONISANTÉ »

In SIIE'2010, we recall our study about the working group reflection on “chronic diseases management (project 2007)”, by the High Council of Public Health (HCSP) with INIST-CNRS in France, to establish an “Information System for Decision Support” powered by a strategic watch process on the health resources [9].
The question by the HCSP was to identify how European health systems manage the problem of “chronic diseases”. The approach to this problem is based on WISP model (ie. Watcher Information and Search Problem) developed by P. Kislin (2007) [21]. This model is the extension of a watch approach to describe the information needs and help the user (decision-maker and watcher) to formulate needs.
In the context of this work, the formulation of needs has been directed against the bibliographic references obtained after consulting the business database (*cf.* III.D).
The strategic issue of this work, by the watcher side, is to formalize the declarative rules, by:

<<ISSUE>> ::= if we do NOT act on the **<OBJECT>** and knowing the state of the **<SIGNAL>**, then the risk is the expected **<HYPOTHESIS>**.

Where:
– ISSUE is defined by an OBJECT of the environment, on which it is possible to act,
– SIGNAL that prompts the decision-maker to trigger the problem,
– HYPOTHESIS which is the risk, as expected consequences, if left unchecked.

The approach in view of application is to better target the information needs of the project sponsor, the HCSP: to translate the strategic issue in a series of dimensions related to the problem with a set of indicators on information retrieval (IR) process.

One can easily imagine, given the multidisciplinary nature of the working group and the specific interests of each expert, the heterogeneity of the subject fields of the system represents a problem in the collected information and for analysis.

A. Health resources: Semantic heterogeneity

“Too much information kills information” has become the favorite expression of those responsible watchers at time of information flow overabundance. The ability to extract relevant information and quickly, while providing added-value manufactured the robustness of any watch process. This concept of added-value will be understood here as the annotation process to facilitate access to relevant information to the user [10]. Indexing and reindexing by users (ie. social tags, folksonomies, etc.) are in the list of tools for this process.
Furthermore, the quantity of information, the heterogeneity of resources and information itself are a problem well known to designers of information systems. Schematically, we speak of a dual heterogeneity that is both semantic and syntactic. The syntax for the heterogeneity of data storage formats (pdf, doc, xml, etc.), query languages and more generally across protocol data structure. The semantic heterogeneity is the differences between the interpretations of the real world inducing several terminology uses for the same reality (ontology, synonymy,

etc.) [11]. Later, we will return to this problem by relying examples on the bibliographic databases.
The added-value gained from the system use two aspects. The first is the addition of keywords or comments (ie. like social tags) by the user to information resources. This allows customization of information regarding the themes of the documentary resources. These annotations can feed up the index in system to improve the return rate for IR. The second aspect concerns the "Information Design" process [1]. Several studies in the medical sector have shown that the visual information influenced the decision-making both in strategic situation (ie. care policies) and therapeutic condition (ie. alternatives to hospitalization) [12], [13].

This logic has not only to refining the conceptual goal in the system but also to support the iterative process: – information needs – IR, – new conceptual indicators (proposing).
This process can come up from techniques of Knowledge Management (KM) and the Mind Mapping.

B. From NLP to Information Visualization

In his research, S. Sidhom (2002) has developed a Morpho-syntactic Analysis Platform for automatic indexing and information retrieval (SIMBAD) [15]. It is composed of an Indexing Kernel (ie the indexing process) that uses the noun phrases (NP) as descriptor in NL structures (ie. to extract concepts) in text documents (and opens to multimedia associated to text descriptions).

We use the definition of a noun phrase (NP) as defined by M. Le Guern (1989) [16], to place a lexicon word in the discourse of universe, *de facto*, this word is ejected in extensional logic, and gives to NP a repository status, as a reality segment associated with.

In our context, the NP appears to be the bearer of a semantic load which makes it relevant and central element to the bibliographic information analysis. Around this semantic sought that guides our analysis on corpus.

Thus, the grammar of NP recognition has three logic levels:

1°/ *Intensional level* (or natural language properties), it is represented by the level N. Words are considered free predicates as simple (ie. the noun properties) or as complex (ie. the noun properties modified by other units: adjectival units A' (ie. A'→A|Adv+A|A+Rel, etc.), expansional preposition EP (ie. EP→Prep+N', etc.), etc.;

2°/ *Intermediate level* (or taking into account the universe of discourse), it is represented by the level N'. It is the transition from the intensional to the extensional levels. Words are considered free predicates with a set construction of closed predicates to denote objects in the world (ie. N'→N+SP|N+A'|…|N);

3°/ *Extensional level* (or the NP and its complexity), it is represented by the level N''. It is the close operation using a quantifier that selects a specific element in the class N of nominal. These are the existing objects in the world, referred objects or mind constructed objects.

In this work, the morpho-syntactic grammar of the NP has been rewritten for NooJ in two levels: – firstly, the work was to reformat the linguistic resources (dictionaries and grammars) resources in our possession, – a second time, we have developed the finite state transducer of the noun phrase. Labels existing dictionaries have been harmonized to match the syntactic graph of NP (FIG. 1.2).

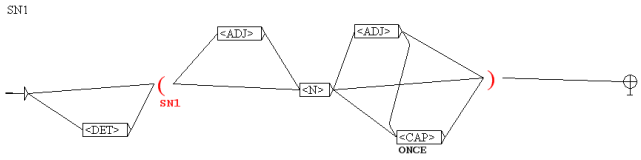

FIG.1: SYNTACTIC GRAPH OF SIMPLE NP IN NOOJ.

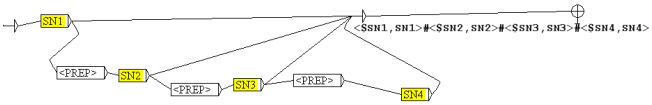

FIG.2: EMBEDDED SYNTACTIC GRAPH OF COMPLEX NP IN NOOJ.

The graph provides numbered phrases identifying the fitting relations in syntagmatic level results [9].
In a logical use of the semantic concepts (ie NP and its properties) from the bibliographic records, the results on output graphs must be operated by an end-user. This should access to information cleaned, leaving him free to evolve in the concepts from a document process: visualize information spaces feeding by heterogeneous data sources. This is increasingly in support of a economic intelligence process and in information design system [17], for "*ChroniSanté*" project.
In particular, in information watch activities, the process is a major vector for the emergence of significant associations after phases of collection, processing and analysis in a large mass of data and information.
Several solutions to information mapping software are available. The tool we used is software under GNU General Public License (GPL3) called Gephi (http://gephi.org). It allows the visualization of complex networks.

C. Corpus study and indicator valorizations

As part of our corpora construction, we mainly searched bibliographic databases via the multi-application "Webspir", a tool that was replaced start of 2009 by the platform "OvidSP" (http://www.ovid.com) with features near-equivalent but more robust for users. Three databases were selected for the constitution of corpora on bibliographic records:

« Pascal [2]», « PsycInfo[3] » et « Medline [4]».

The choice to use these three sources on health information is justified by our hope to cover as fully as possible the thematic management on chronic diseases. The basic advantage of *Pascal* database presents European references and includes records from the databases in public health (BDSP). The *Medline* database is centered on U.S. publications, as *PsycInfo*, but with broader themes in social sciences.
In synthesis and contrary to this logic which requires completeness topics, our search equations were developed to deliver results to the widest possible: first, to cover all the sub-themes on "chronic disease" and, second, to identify new sub-themes which we had not originally thought.
The browsing on the three databases has reported: 2097 references to *Pascal*, 6110 references to *Medline,* and 2177 references to *PsycInfo*. We subsequently refined our search to select only those produced between 2001 and 2009 in French. The result consists of 397 references and 303 references in deduplication pass. These corpora will be the first synthesis of our work. These results at the IR process indicate that "*chronic disease*" is a new concept in France, because of the singularity of the model in the French health system.
A second approach has motivated a second job on the database "Pubmed" as we have previously mentioned. The completion of the "*ChroniSanté*" project as "decision support system" (DSS) or SIAD in French, was faced with a semantic problem: the rendering of the term "chronic disease" as a concept purely Anglo-Saxon, which brings a series of problems in a multilingual and "Translation Terminology".

In fact, the completeness of the study involves watch process on multilingual literature to define the best concept and study what in intersects. It is in this logic that the base "Pubmed" was viewed with as a search and retrieval for the term "chronic disease" in the titles of records. The result returned is 13,222 records that were the parallel corpus in English with relation to initial multi-base.

IV. Connections to Latent Warning (or Weak Signals) in the ID process

Applying automatic analysis (NooJ) on corpora (ie. as the ID process in phases 1– acquire and 2– parse), the complex graph of NP reported 1374 concepts (ie. as the ID phase 3– filter) including the smallest concepts (ie. the lemma N) to simple or complex concepts NP (ie. levels N' + N").
The advantage of this approach is to present to users the key or primary concepts (ie. as the ID phase 4– mine) in information resources but also secondary concepts (ie the ID phase 5– Representation) that the watcher does not necessarily think in research of indicators: the translation phase of a decision problem into a IR problem in the EI context.
In this case, considering the concept of "**patient**" is central to our theme. In practice, we tend to establish our search for indicators in a **passive** acceptation with concepts: "patient monitoring [(FR) *suivi du patient*]", "patient care [(FR) *prise en charge du patient*]", "patient education [(FR) *éducation du patient*]", etc. But not in an **active** acceptation, as "patient involvement [(FR) *implication du patient*]", "active participation of patient [(FR) *participation active du patient*]", etc. (ie. as the ID phases 6– refine in iteration).
Also, the research of NPs in the titles of references highlighted ideas that apparently have no close relation with our themes, but which nevertheless appear several times in different references. In this case, on the theme of "cannabis consumption" it puts a link for "long term illnesses".
Given these observations, we took advantage of select matches the longest in the NP. This corresponds to the **fitting** relation: the concept of fitting in ($x \supseteq y$), x the longest and most informative NP; the concept fitted ($z \subseteq w$), z the shortest and the least accurate NP. This richness of meaning that emerge, allows the identification of informational collection with relevant, complex and hierarchical concepts (NP). These characteristics may go unnoticed in the linear analysis of a corpus.
Thus, the process of the corpus of bibliographic records on the platform NooJ showed satisfactory results based on the NP and its semantic properties (ie. the relations of **Tree** (T: $y \subseteq x$ and $x \supseteq z$, **Fitting** (F: $x \supseteq y \supseteq w$) and **Belongs** (B: $n \in x$/ $x \supseteq y$)) [15].

—

[2] **Pascal:** Produced by INIST-CNRS, PASCAL ® is an international and multidisciplinary database that identifies literature in Science, Technology and Medicine.

[3] **PsycInfo:** Database of the American Psychological Association (APA) provides access to journal articles (many are full text), book chapters and books, research reports and theses and dissertations in psychology and related fields (medicine, nursing, sociology, etc.) of the 19th century to today.

[4] **Medline:** Bibliographic database produced by the National Library of Medicine (NLM-USA). It covers all biomedical fields: biochemistry, biology, clinical medicine, economics, ethics, dentistry, pharmacology, psychiatry, public health, toxicology, veterinary medicine.

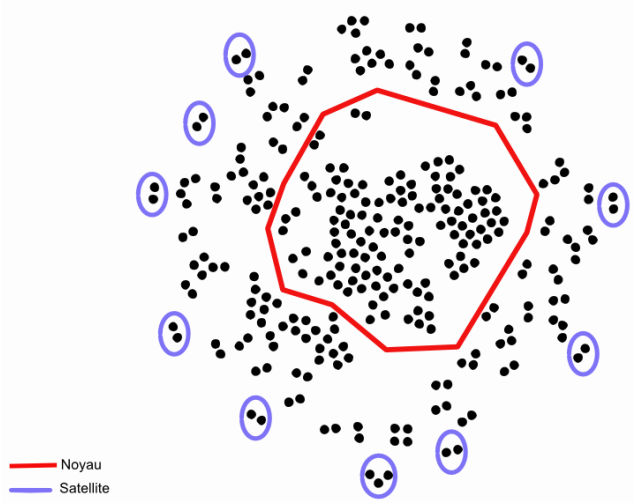

FIG.3: SEMANTIC NETWORKS BASED ON BIBLIOGRAPHIC RECORDS: NUCLEUS AND SATELLITE CONNECTIONS.

Concerning the visualization of information (ie. as the ID phase 7–interact in iteration), we tested the application with Gephi Fruchterman-Rheingold algorithm [19], on the results of extraction with NooJ. This algorithm of multi-scale force can calculate the force between two nodes and mapping complex networks. By its use, there is much emerging nucleus surrounded of satellite subsystems that can be considered non-central themes to the theme target. According to the analysis of information needs, the watcher can focus attention on these satellites nodes to be considered as "latent warning signs" themes (ie. weak signals in EI) and give them special consideration (Fig. 3).

By NooJ parsing, we present the graph results of the NP extraction in the French corpus (FIG. 4).

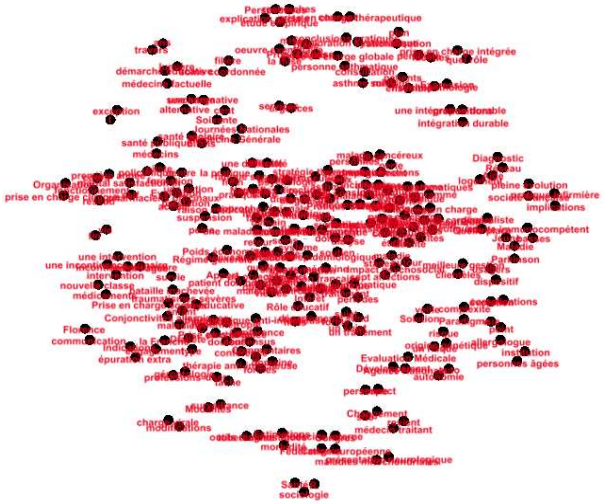

FIG.4: VISUALIZATION OF NP SEMANTIC NETWORKS.

Based on the analysis of the semantic network, we observe that the center of the graph (central nucleus) consists of terms related to the decision-making analysis: work that we completed at first in the process of EI.

Thus, for the analyzed corpora, it stands from terms such as "**coverage, care [FR:** *prise en charge*]" or "chronic diseases **[FR:** maladies chroniques**]**", new terms like: ("hepatitis C", "cardiopathy", "asthma", etc.). We also notice the relations that exist between nodes that represent the semantic connections between terms (FIG.5).

The usefulness of such "visual structure" document for a watcher, an expert or a decision-maker, is not to demonstrate. It allows presenting an interactive document likely to bring new knowledge. In such semantic logic, it allows to better understand the complex dimensions to take into account in watch or EI processes.

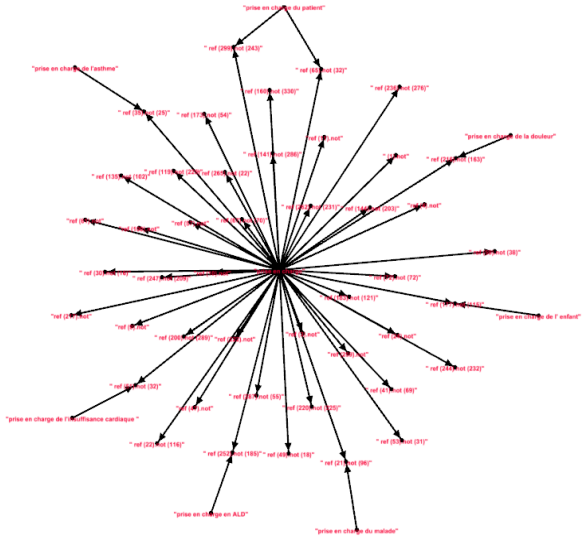

FIG.5: THE NP CORE (CENTRAL NUCLEUS) AND THEMES.

The visualization of named entities in the corpora can also be positioned relative to the documentary logic. The graph makes it possible to show relations between concepts (NP) and document references (FIG.6). In other way, we can see which resources concentrated more concepts and which one is potentially relevant.

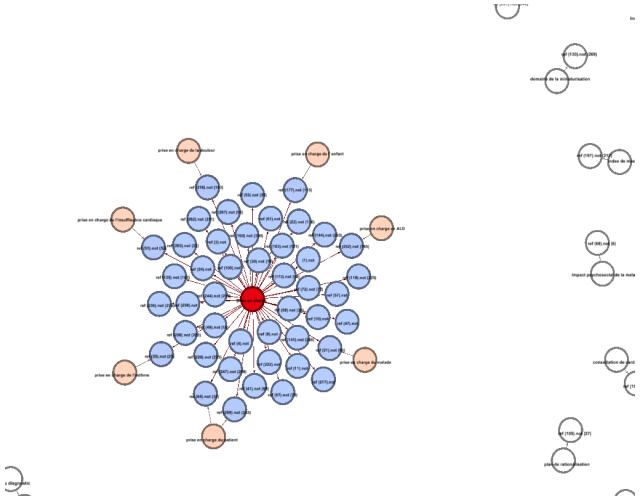

FIG.6: CONNECTION BETWEEN NP AND INFORMATION RESOURCES.

The atomic structure for the nucleus concept, "**coverage, care [FR:** *prise en charge*]" is linked to the bibliographic references in which appears. We also note the secondary concepts like: "management of asthma **[FR:** *prise en charge de l'asthme* **]**", "chronic asthma **[FR:** *asthme chronique* **]**" and other concepts appear in the peripheral area of the semantic network.

This process was applied to the second corpora in English literature, with the same Logic: the extraction of NPs based on graphs modeling [14]. Four hundred and twenty nine (429) mapped terms are returned. They are linked with their document resources. This approach requires to watcher a "proactive approach" through the search (mining) graphs proposed, in order to detect new knowledge. In the logic presented here, it may refine the concepts from another cultural sphere to watcher.

In application, the activation of a **term (T)** allows viewing **notice records (N1.. Ni)** in the connection with **subjects (S1.. Sm)** and associated **keywords (K1.. Kj).** The scenarii for identifying and refining may be multiple: as an example, from **record (Ni)** to **keywords (K1.. Kj)** or vice-versa. The possibility to link the node "notice" to the source document by hyperlink feature allows the watcher to have access to the document environment in full term of interest with any subject.

V. DISCUSSION AND CONCLUSION

This experience shows that the techniques we used require automation to achieve a state of performance, robustness and acceptable level of efficiency:

– For the "*Information Design*" [1] process defined as the "art and science to preparing information that it can be used by human with efficiency and effectiveness." In the ID process, we valued our study areas by the aspects: "graph(s)", "semantic network (s)", "project (s)" and "connection(s)" to translate into clear, immediate and appropriate information for the user. In our study, the user is often the watcher, the analyst or expert and the decision-maker.

For us, the useful information is not the increase of information quantity, but contrary the reduction of it by relevant information clusters to facilitate its reading and its appropriation [17]. This has been discussed and treated throughout this paper explicitly as the application of ID process in the context of on "chronic diseases" study (project "*ChroniSanté"*).

– For the "*Watch" process*, the information visualization extracted from the concepts of NPs is useful to actors in a strategic project in many aspects.

At first, information visualization facilitates the document indexing and contents to information system (IS), information retrieval (IR) system or decision support system (DSS). As an example, for a bibliographic record or document to analyze, to extract noun phrases in the contents may be converted into tags. This solution allows any user of this document to present the key concepts in the information database [17]. This may also encourage a new Logic of "reindexing by users": the user-tags are automatically stored; the user will add his subjective, objective and creative tags, up to added-value [20].

At second, the visualization of a semantic network (based on NPs concepts and properties) enables the production of new knowledge. Viewed nodes in a semantic network can indeed be analyzed in a working group to identify new topics related to business intelligence as convergence and divergence of represented subjects (ie. decision-maker needs). This is, to use "Humbert Lesca" logic; heuristic process to allow a collective creation of meaning [17].

At third, in watch, information and documentation processes, this visualization logic and ID process can bring out the most potentially relevant references. On this point, it should refine results by a statistical analysis: the use of bibliometric indicators [18] as the TF-IDF (term frequency-inverse document frequency).

– For the "*Economic Intelligence*", we see clearly that the usefulness of the ID process is to go beyond a simple and literal translation of IR indicators (ie. the translation phase of a decision problem into an IR problem). For complex and "multilingual" semantic search, we can take for example the conceptual differences between terms, like: "chronic disease" or "chronic disease management" and "management of chronic diseases". Thus, we can show, by semantic visualizing of these concepts, the connections with the processed information (parsing, analysis and needs information). Also, we can establish multi-level intersections between "information", "concepts" and common or different "morphemes" to get shared meanings.

Finally, on a technical level, the problem of the heterogeneous information resources, to supply the health system, can be minimized with the addition of annotations and/or linguistic processing. NLP tools (as NooJ or others) will enable the processing on multi-format sources. And for non-textual documents (multimedia), analysis can be made on the annotations associated to documents: developing a semantic homogenization, counterbalancing the syntactic heterogeneity.

VI. REFERENCES

[1] H. I. Ansoff. (1975). “Managing strategic surprise by response to weak signals”. California Management Review, vol. 18, no. 2, pp. 21–33, 1975.

[2] H. I. Ansoff. (1985). “Strategic response in turbulent environments”. Handbook of business strategy, 1985.

[3] B. Coffman. (1997).Weak Signal Research, Part I: Introduction, Journal of Transition Management, MG Taylor Corporation (1997).

[4] B. Coffman. (1997). Weak Signal Research, Part II: Information Theory, Journal of Transition Management, MG Taylor Corporation (1997).

[5] H. Lesca. (2001). “Veille stratégique: passage de la notion de signal faible à la notion de signe d’alerte précoce,” in Colloque VSST, pp. 98–105, 2001.

[6] M. J. Eppler and R. A. Burkhard. (2007). “Visual representations in knowledge management: framework and cases”. Journal of Knowledge Management, vol. 11, no. 4, pp. 112–122, 2007.

[7] D. Karabeg. (2002). “Designing information design”. Information Design Journal, vol. 11, no. 1, pp. 82–90, 2002.

[8] Benjamin Jotham Fry. (2004). Computational information design. bfa Communication Design, in Computer Science Carnegie Mellon University. Massachusetts Institute of Technology, April 2004..

[9] P. Lambert, S. Sidhom. (2010). “Knowledge extraction and visualization: case study on ChroniSanté project in France ” . in procedings SIIE'2010, Sousse (Feb. Tunisia). IHE Edition.

[10] S. Sidhom. (2008). “Approche conceptuelle par un processus d'annotation pour la représentation et la valorisation de contenus informationnels en intelligence économique (IE),” 2008. in Proceedings SIIE'08 - Systèmes d'Information et Intelligence Economique, IHE Edition vol.1, Hammamet (fév. 2009).

[11] C. H. Goh, S. Bressan, S. Madnick, and M. Siegel. (1999). “Context interchange: new features and formalisms for the intelligent integration of information,” ACM Trans. Inf. Syst., vol. 17, no. 3, pp. 270-293, 1999.

[12] L. S. Elting, C. G. Martin, S. B. Cantor, and E. B. Rubenstein. (1999). “Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures,” BMJ, vol. 318, no. 7197, pp. 1527 -1531, Juin. 1999.

[13] J. Wyatt. (1999). “Same information, different decisions: format counts,” BMJ, vol. 318, no. 7197, pp. 1501 -1502, Juin. 1999.

[14] M. Silberztein. (2005). “NooJ: a linguistic annotation system for corpus processing,” in Proceedings of HLT/EMNLP on Interactive Demonstrations, p. 11, 2005.

[15] S. Sidhom. (2002). “Plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information: de l'écrit vers la gestion des connaissances”. Thèse de doctorat, Université Claude bernard Lyon1. 2002.

[16] M. Le Guern. (1989). “Sur les relations entre terminologie et lexique,” Meta, vol. 34, no. 3, 1989.

[17] H. Lesca, S. Kriaa-Medhaffer, and A. Casagrande. (2009). “Veille stratégique: Un Facteur d’échec paradoxal largement avéré: la surinformation causée par l’Internet. Cas concrets, retours d’expérience et piste de solutions”. in Proceedings SIIE'09 - Systèmes d'Information et Intelligence Economique, IHE Edition. Hammamet (12-14 fév. 2009).

[18] G. Salton, A. Wong, and C. S. Yang. (1975). “A vector space model for automatic indexing,” Communications of the ACM, vol. 18, no. 11, pp. 613–620, 1975.

[19] S. Card, J. Mackinlay, and B. Shneiderman. (1999). Readings in Information Visualization: Using Vision to Think. 1999. Morgan Kaufmann Ed.

[20] A. Harbaoui, M. Ghenima, and S. Sidhom. (2009). “Enrichissement des contenus par la réindexation des usagers : un état de l'art sur la problématique”. in Proceedings SIIE'09 - Systèmes d'Information et Intelligence Economique, IHE Edition. Hammamet (12-14 fév. 2009).

[21] P. Kislin. (2007). “Modélisation du problème informationnel du veilleur dans la démarche d’Intelligence Économique”. Thèse de doctorat : université de Nancy2. 05 nov. 2007.