Project Management within Economic Intelligence: using NooJ as Diagnostic Tool for nanometrology cluster

Sahbi Sidhom, Philippe Lambert

NooJ 2013
international conference NooJ in June 3-5, 2013 in Sarrbrucken – Germany.

Project Management within Economic Intelligence: using NooJ as Diagnostic Tool for nanometrology cluster

Sahbi Sidhom [1], Philippe Lambert [2]

1 : Sahbi.sidhom@loria.fr
KIWI (LORIA)
INRIA – CNRS : UMR7503 – Université de Lorraine (France)

2 : Philippe.lambert@univ-lorraine.fr
Institut Jean Lamour / C'Nano Grand Est (CNRS)
Université de Lorraine (France)

ABSTRACT

This paper presents the methodology of monitoring/watch process (WP) and its complexity in economic intelligence (EI) context study. The examination of the theoretical foundation for knowledge representation through corpora studies is tested and then applied in application framework with the subject of nonoscience and nanotechnology (in France). The implementation was on NooJ platform. The methodology applied includes five phases: data selection, cleaning, linguistic resources development, NooJ processing and result analysis. The main results of this paper presents the connection of the WP methodology through a specific case study in nanometrology, which allows the diagnosis of the internal structure and the creation of links between actors, themes and projects towards the innovation.

Processing with NooJ was, firstly, to mine on actor activities linked to information needs in order to get the better management of the project and, secondly, to determine the professional classes of each actor in the nanometrology cluster.

KEYWORDS

Economic intelligence (EI) : study field; monitoring/watch process (WP); decision-problem; nanometrology cluster; knowledge organization (KO); Community Management; nano-science;

--
NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).

1. Introduction

In recent years, our research focus has been applied to data mining from the Web (i.e. Web Mining and Semantic Web). That helps to develop new knowledge from large corpora (text contents in bibliographic records). The developed techniques are increasingly interesting for leaders who have watch responsibility to detect topics that could have been missed in a linear reading. In the field of "Economic Intelligence" (EI) studies, the implicit properties on corpus analysis take on the name of "Weak Signals" (WS) [19], [20]. The detection of WS allows watchers and business actors in company to take into account the environment indicators in a dynamic sense and foresight (i.e. "prepare today for tomorrow"). However, the connection between EI and WS detection requires the development of complex processes in the context of monitoring/watrch process (WP). This process represents the topic of this paper.

This paper presents in section 2. the methodology of WP and its complexity in EI context. In section 3., the examination of the theoretical foundation for knowledge representation through corpora studies. In section 4., The application framework is on the subject of nonoscience and nanotechnology in France. The implementation was on NooJ platform. The methodology applied includes five phases (data selection, cleaning, linguistic resources development, processing and result analysis). Finally, we present the connection of the methodology through a specific case study in nanometrology, which allows the diagnosis of the internal structure and the creation of links between actors, themes and projects towards the innovation.

2. Methodology of WP and its complexity

In study field, Economic Intelligence (EI) is a concept that is the subject of multiple definitions. Undoubtedly, these are due to the confusion on the word "intelligence" (i.e. the "intelligence and espionage" in English, but "knowledge and expertise (or know-how)" in Latin meaning and certainly in French). B. Carayon (2003) as a visionary in the field in France explained the concept by: "Sometimes equated with economic espionage, sometimes traditional methods of strategy information processing for the benefit of the few companies, (…)" [4]. However, most experts in Europe and particularly in France seem to agree on the definition of B. Carayon (2003). For this author, doing Economic Intelligence, is trying to master, i.e. seek and protect the critical information necessary to perform an economic activity in a business firm [4]. The key verbs of the Economic Intelligence, as actions to do, are: inform, anticipate and defend interests.

The coordination in actions are commonly admitted by EI practitioners as: "search", "processing" and "distribution" for "exploitation". Starting from both experiences and theoretical researches in EI, the topics of monitoring/watrch process (WP) and knowledge management (KM) [8] are joint with the use of information processing [18]. In methodology approach NooJ is used to add semantic methods to the EI system allowing a gain of performance in information retrieval, in decision-problem elicitation towards information monitoring.

First, we should distinguish the two processes involved in the EI and WP. In monitoring/watch process, the functionality of the collection, analysis and processing to

--

NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).

sharing information are integrated into continuous and cyclical approaches by strictly informational needs. Requirements explained in the functional protocol specifications (FPS) receive answers explained in terms of information synthesis, document resources and informational sources. In no way, the deliverable of FPS (output WP) will suggest other that information in order to preserve the objectivity of WP product and its actors (watcher) in response to the need.

Downstream of the monitoring process, strategy, expertise and decision under other processes are not in the role of watcher and inherent in the standby process. In upstream, the decision-maker of WP product is not able to disclose strategic issues, preferences or projections decision to preserve the objectivity of watchers and not bias the responses in FPS.

In monitoring/watch methodology, raw information when it is collected, analyzed and shared in a defined context will produce value in contrast to the opinion of analysts. When this information is valued in a strategic context by aligning the issues with the needs of the decision-maker, the value-information will necessarily be of premises (ie. facts or truths) in the decision process: these facts will be used in rules that govern the decision-making. In this regard, a decision rule is the situation of facts by logical inference.

3. knowledge representation in corpora

Our study focuses on discourse through its textual materiality and is moving towards organic units that compose both extensional and intensional levels in language description.

Theoretical orientations in computational linguistics leave to specialists, as computational-linguists, a large number of decisions that we cannot hold in our developed recommendations [14],[15]: our orientation determines only the theoretical basis of the approach described in the research work of S. Sidhom [16], [17]. This research concerned the descriptor status for indexing model [9],[10] versus the noun phrase (or NP) status in natural language model. The real challenge for the transition from language model to indexing [3] model is based on the extraction of morpho-syntactic structures through NP: its semantic, its formal representations and its properties.

This fitting of the semantic search in formal representations and properties in NPs, develops the founding architecture of this work by drawing a bridge between information source unstructured contents [5],[6] in knowledge representation highly structured and processed in the system [11],[12],[13].

For the aspect formalization [1],[2], we present the different morpho-syntactic rules, applied to French language. These grammar rules were used in the implementation of the parser, in NooJ platform, and related to the concept of cognitive grammar [7] in INA corpus:

Tab.1 - Introductive Proposal (PI) in sentence:

PI	Examples
1. SP, $\subset$ S	Pour les 20 ans d'AIRBUS INDUSTRIE, …

--

NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).

2. EP, ⊂ S	En parallèle, …
3. EP + SP, ⊂ S	En direct depuis l'observatoire de Meudon, … En compagnie de Marianne GRUNBERG-MANAGO, …
4. PPas + SP, ⊂ S	Embarqués à bord de l'astrolabe depuis l'extrême sud de l'Australie, …
5. PPré + SN, ⊂ S	Proposant un voyage à travers les sites industriels de France, …
6. Prép(en) + SNdat, ⊂ S	En juin 1986,
7. Prép(en) + PPré + SP, ⊂ S	En passant [par (la littérature)], …
8. Conj, ⊂ S	Cependant, …
9. Conj + Adv + SP, ⊂ S	Car contrairement aux Américains, …
10. « en » + PPrés, ⊂ S	En vaccinant,

Tab.2 - NP or Nominal Syntagma (SN) in sentence:

SN	**Examples**
11. SN (details in ref. [17])	Le lac … ⊃SN le lac dans le nouveau Québec (…)⊃SN
12. EP ⊂ SN	Une équipe ⊃de tournage … Un avion Hercule ⊃de transport stratégique
13. SP ⊂ SN	La présence ⊃d'un lac …
14. { SN, SP, EP} ⊂ SN	L'utilisation ⊃d'images de synthèse …
15. REL (relative explicative) ⊂ SN	La présence d'un lac ⊃qui se serait formé suite à la chute d'une météorite …
Exceptions :	
16. SN^{∇}= SN without determiner	Psychologues et physiciens (se penchent sur leurs multiples facettes.)

--
NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).

Tab.3 - Relative Sentence (REL) in sentence:

REL	Examples
17. /REL = Prel + SN/ ⊂ SN	…, qui + son père, …
18. /REL = Prel + SV/ ⊂ SN	… qui + se serait formé suite à la chute d'une météorite …
19. /REL = Prel + S/ ⊂ SN	a) … qu' + il a réalisé sur le même sujet en 1973. b) … dont + le pouvoir suggestif déborde largement le cadre du bâtiment lui-même.

Tab.4 - Verbal Syntagma (SV) in sentence :

SV	Examples
20. V + (Prép + V-inf)+ SP	…**est** (de récupérer) de la matière cosmique
21. V + (Prép + V-inf)+ SN	…**sont montrées** (pour comprendre) les difficultés techniques et économiques
22. V + (V-inf) + SN	…**a pu** (rencontrer) AIRBUS INDUSTRIE
23. V + (V-inf) + (Prép + V-inf)+ SN	…**devait** (permettre) (d'identifier) le sexe
24. V + (PPrés) + SN	…**a suivi** (durant) trois semaines les activités d'une équipe
25. V + SN	…**sont** le reflet de notre société
26. V + SP	… **est réservée** aux avions Hercule
27. V + {SN, SP, EP, PV}	… **essaie** d'expliquer le mystère de l'étoile de Bethléem
28. V + (Adv) + SN	…**explique** (comment) les pays européens exportent des armes
29. V + (Adv) + V	…**sont** intimement **liées** …**est** ainsi **développé**
30. V + (Adv) + (Prép + V-inf)+ SN	…**s'attache** (plus) (à expliquer) la course du côté soviétique
31. V + /EP/ + SN	…**démontre** /en particulier/ la polititique de la France à ce sujet
32. V + /Conj/ + SN	…**poursuit** /donc/ cette balade à la fois historique, sociologique et architecturale
33. V	Ce chien **mord**

--

34. V + (Adj) + SP	…**furent** (découvertes) en 1988
35. V + {Adv, Adj}	(il) **résout** scientifiquement…

The sentence structure (S) in our study is based on three fundamental structures, namely: - the structure that precedes the sentence (PI, $\subset$ S), after - the subject of noun phrase of S (in the form of complex-NP or SN_max $\subset$ S), after - the verbal sentence of S (SV $\subset$ S), and - the relative structure (REL) that can complete SN or SV in option ([REL $\subset$ SN|SV] $\subset$ S). Each of these structures is identified in its component units and its morpho-syntactic organization.

In application context, NooJ languistic environment, developed by Max Silberztein (2005), [21], [22] was applied to our corpus. NooJ processing with our 35 rules enables the identification of complex expressions, extracting lemmas and automatic annotation on text resources in corpora [16], [17].

In the fields of nanoscience and nanotechnology, a new corpus was used to overcome the observations and manipulations: is was made from a public opinion survey combining closed and open issues in the field. The opinion survey aims to observe members of a new collaborative structure, The Nanometrology Cluster in France, (or “Club nanoMétrologie”, club-nanometrologie.fr), set up by the consortium of laboratories in nanoscience and nanotechnology in France (NanoSciences France - C'Nano) and the National Laboratory of Metrology and Testing (LNE). The main hypothesis made about the nature of the text from the opinion survey: it is (i) free text, (ii) carried in open issues and (iii) there are no constraints in style or editorial proposed or imposed. The variability in its contents should describe its representation as a document to validate our choices on the robustness of cognitive grammar implemented. In the process, it is a free text to submit for the automatic analysis to build the knowledge representation: NP extraction and properties.

4. Main results and discussions

The survey methodology used to reach the most members (in nano domain) and ensure to have a maximum responses, was to implement it under the open-source software LimeSurvey (www.limesurvey.org) as online survey to collect responses adherents with secure. The case study focuses on two main problems in open questions and have been processed in NooJ. The expressed problems were: (i) "What are the reasons for which the respondent has joined the Cluster? ", and (ii)" What specifically expected in such collaborative structure? ".

For the treatments, the methodology includes five phases, which are: (i) data selection, (ii) data cleaning, (iii) development of ad-hoc linguistic resources, (iv) data processing, and (v) results analysis. All phases were implemented on NooJ platform.

--
NooJ’2013: 1st. proposal (S. Sidhom & P. Lambert).

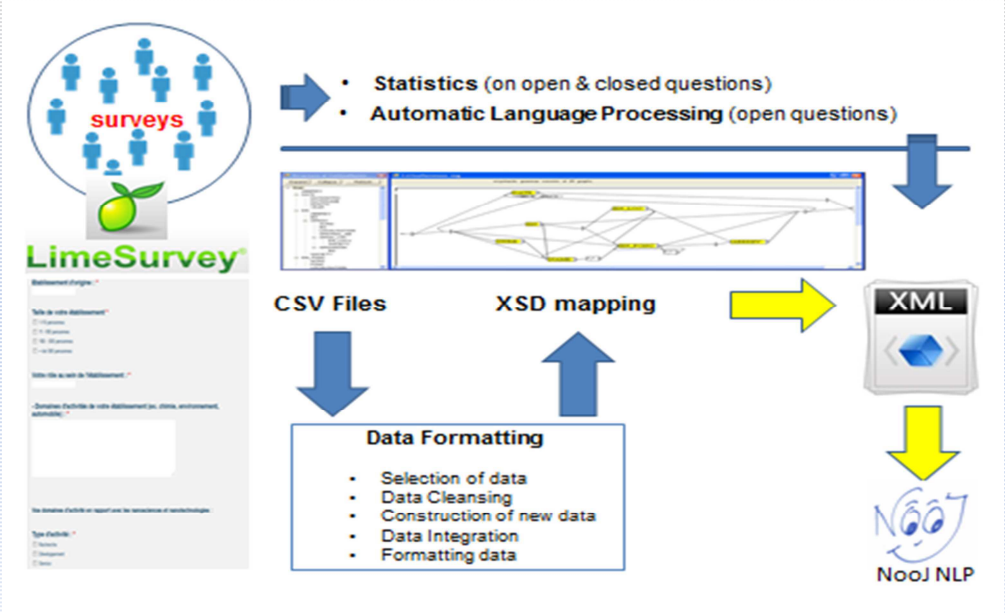

Fig. 1: methodology in five phases with NooJ processing.

The LimeSurvey system allows the downloading of responses into csv format (cf. Fig.1). To facilitate the processing on hundreds of responses, we decided to reformat the responses file into XML format (cf. Fig.2). This choice corresponds to two specific treatment criteria: (i) obtaining a better structuring of the source document to find easily the original answers, and (ii) giving the system to iterate over the file with specific treatments as XML nodes to study in their particularities. The source file has also been divided into as many texts as responses, thus constituting a corpus of a hundred xml files. This has allowed to easily find the source document for extraction units or tokens (NPs and structure properties) during the NLP process with NooJ.

Data cleaning consisted primarily to correct spelling mistakes and clean each node of any nature that may create noise in the results of the system analysis.

1	clubLNE71	CEA-Grenoble
2	clubLNE71	Caractérisations de postes travail où sont utilisées des nanoparticules
3	clubLNE01	OMNT (CEA/CNRS)
4	clubLNE01	Activités de VEILLE scientifique dans les domaines suivants
5	clubLNE01	en lien avec les nanomatériaux
6	clubLNE01	détection et caractérisation
7	clubLNE01	mesures d
8	clubLNE01	exposition
9	clubLNE01	toxicologie et ecotoxicologie
10	clubLNE01	évaluation et gestion du risque
11	clubLNE01	aspects normatifs et réglementaires
12	clubLNE02	Université de Provence
13	clubLNE02	Ecotoxicologie
14	clubLNE02	zoologie
15	clubLNE02	immuno-histochimie
16	clubLNE02	invertébrés aquatiques
17	clubLNE02	poisson
18	clubLNE02	nano-écotoxicité
19	clubLNE03	Institut Pasteur
20	clubLNE03	imagerie sub-diffraction
21	clubLNE03	nanoscopie
22	clubLNE03	microscopie STED
23	clubLNE03	instrumentation optique
24	clubLNE03	microscopie de fluorescence

Fig. 2: instance of XML file answers and NooJ processing in the survey.

--

NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).

The development of linguistic structures was done in two stages: the first stage was to create specific thematic dictionaries in nanosciences and nanotechnologies. NooJ offers useful features by creating automatic dictionary consisting of entries labeled as unknown (<Unknown>), thus giving the opportunity to integrate into the existing dictionary. Thus, a dictionary with hundreds of tokens has been quickly created, bringing together the themes of the nanometrology cluster techniques and specific types of measurement for nanosciences. The second stage concerned the creation of finite state automata to extract specific data and reformatting for further NLP processing (cf. Fig. 3).

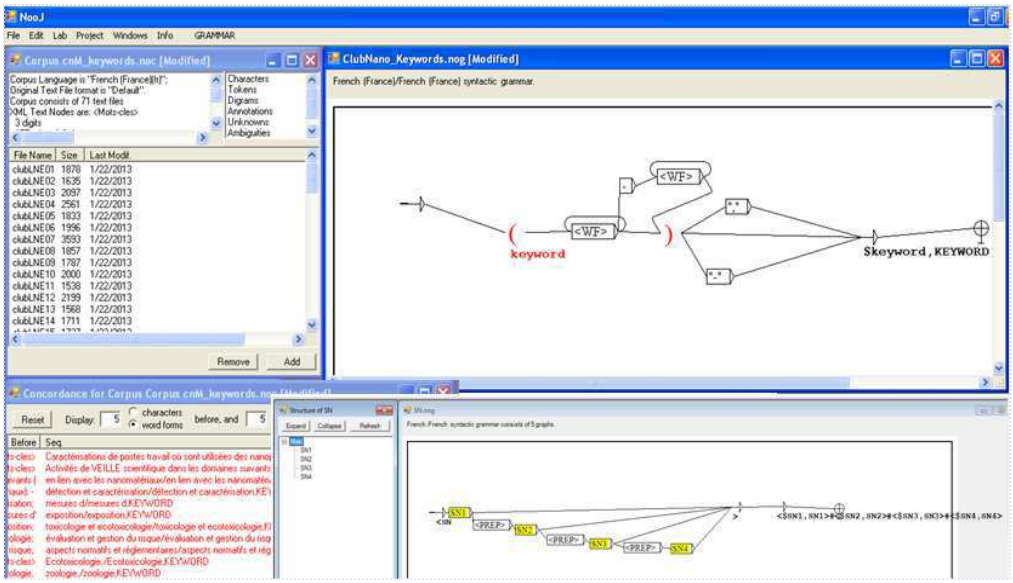

Fig. 3: instance of finite state automaton for NP extracting in cascade.

Under NooJ, the use of automata is built in cascaded approach. Iteration permits labeling of selected structures at several levels so into thin extractions.

In the phase of data processing, the work has focused on labeling nodes "What are the reasons for your membership?", and "What do you expect from Nanometrology Cluster?". We have obtained twenty significant results showing that the reasons for participation in the cluster are linked in the same semantic field (ie. networking). The majority of responses are in logic part of: " creating network ", "community integration", "identification of community Manager ", etc. NooJ has the functionality of statistical processing, giving it the status of hybrid system, it allowed the ranking results based on the allocation for each NP extraction of TF-IDF index. The result is a collection of needs expressed by the participants, with order of importance, corresponding to the semantic weight detected by the NooJ parsing.

5. Conclusion

In synthesis, the objectives built in this research were able to bring an adaptive model using a morpho-syntactic parser for social re-indexing resources. The formalism and implementation in NooJ was in line with the nature of the study object: opinion surveys in nanosciences and nanotechnologies.

--

NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).

In the analysis, we demonstrate a link context between "natural language processing" (NLP) and knowledge organization (KO). The valuations were observed by social re-indexing through new concepts. Open questions in nano survey have been the subject of a specific automatic processing, to observe: (i) "What are the reasons for which the respondent has joined the Cluster?", and (ii) "What specifically expected such in collaborative structure?". For decision support, we had proposed the harmonization of the activities, projects and actors combining skills: these results highlight the need for Community Management (CM) activity. This kind of activity give advantage to enhance proactive actors [8], [19] and the emergence of new offers in nano projects by the combination of activities and skills [8].

The developed methodology also allows the diagnosis of the internal structure of the network in Nanometrology Cluster. Detection of the heterogeneous nature of the network can be exploited to carry out the re-balancing around the center of gravity in the structure and then to enhance the cohesion network in nano domain.

REFERENCES

[1] Bachimont B. (1999). La documentation au coeur du processus de production. in Dossier de l'Audiovisuel, Janvier-Février 1999, n°83, INA-Publications, p.38-39.

[2] Bachimont B. (2004). Signes formels et computation numérique : entre intuition et formalisme. In H. Schramm, L. Schwarte & J. Lazardzig (Eds.), Instrumente in Kunst und Wissenschaft - Zur Architektonik kultureller Grenzen im 17. Jahrhundert. Berlin: Walter de Gruyter Verlag.

[3] Browne G. (1996). Automatic indexing and abstracting. in Indexing in Electronic Age Conference, Robertson, NSW 20-21 April 1996, Australian Society of Indexers, 8p.

[4] Carayon B. (2003). Intelligence économique, compéttivité, cohésion sociale. Rapport au Premier minsitère, La Documentation Française, Juillet 2003, Paris.

[5] Guimier-Sorbets A-M. (1993). Des textes aux images : accès aux informations multimédias par le langage naturel. Documentaliste – Sciences de l'information, 1993, vol.30, n°3, p.127-134.

[6] Régimbeau G. (1998). Accès thématiques aux d'art contemporaines dans les banques de données. in Documentaliste Sciences de l'Information, Volume 35, n°1, janvier 1998, p.15-23.

[7] Lambert P., Sidhom S. (2011). Problématique de la veille informationnelle en contexte interculturel : étude de cas d'un processus d'identification d'experts vietnamiens". in Proceedings : ISKO-Maghreb'11 – Concept and Tools for Knowledge Management (KM). ESCE-University of la Manouba Edition. Hammamet (Tunisia) May. 2011.

[8] Lambert P., Sidhom, S. (2010). Vers le Design d'information pour valoriser les résultats d'une veille sur les maladies chroniques. in Proceedings: Journée d'étude sur la "Mutualisation des ressources documentaires : Hétérogénéité des ressources et accessibilité dans un espace collaboratif." ELICO - Université Jean Moulin Lyon3, 05/11/2010 Lyon (France).

[9] Harbaoui A., Ghenima M., Sidhom S. (2009). Enrichissement des contenus par la réindexation des usagers : un état de l'art sur la problématique. in 2nd International Conference on Information Systems and Economic Intelligence - SIIE 2009 vol.1 (2009) pp. 932-942 IHE Edition Tunis.

--

NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).

[10] VAN SLYPE G. (1987). Les langages d'indexation : conception, construction et utilisation. in dans les systèmes documentaires Paris : Editions d'organisation, 1987. 277 p. - (Systèmes d'information et de documentation).

[11] Calmet J., Maret P. (2013). Toward a trust model for knowledge-based communities. WIMS 2013: 47.

[12] Vercouter L., Maret P. (2012). Introducing Web Intelligence for communities. Web Intelligence and Agent Systems 10(1): 91-92 (2012).

[13] Stan J., Do V-H., Maret P. (2011). Semantic User Interaction Profiles for Better People Recommendation. ASONAM 2011: 434-437.

[14] Mseddi R., Sidhom S., Ghenima M., Ben Ghezala H. (2011). From information to decision: information management methodology in decisional process. in in Proceedings SIIE'2011 : Information Systems and Economic Intelligence (SIIE'2011) vol.1 (2011) pp.219-226, IGA Edition. Marrakech (Morocco) Feb. 2011.

[15] B. Bertin, V. Scuturici, J.M. Pinon, E. Risler. (2012). CarbonDB : a Semantic Life Cycle Inventory Database. in Conference on Information and Knowledge Management (CIKM) 2012, Maui, Hawaï.2012.

[16] Sidhom S., Hassoun M., Bouché R. (1999). Cognitive grammar for indexing and writing. ISKO-España Conference Proceedings, 22-24 april 1999 Granada, p.11-16.

[17] Sidhom S. (2002). Plateforme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information: de l'écrit vers la gestion des connaissances. Thèse de doctorat de l'Université Claude Bernard Lyon1. France. Mars 2002. p.247.

[18] Sidhom S., Ghenima M., Lambert P. (2010). Systèmes d'information et Intelligence économique : enjeux et perspectives. in Proceedings IEMA-4, 4ème Colloque International sur l'Intelligence Économique et le Knowledge Management - vol.1 (17/05/2010) Alger. (Sidhom S. comme conférencier invité).

[19] Sidhom S., and Lambert P. (2011). "Information Design for Weak Signal detection and processing in Economic Intelligence: case study on Health resources". in Proceedings SIIE'11: Information Systems and Economic Intelligence, IGA Edition. Marrakech (Morocco) Feb. 2011.

[20] Sidhom S. (2013). Conjoncture des processus d'indexation et de gestion des connaissances : vers la réindexation par les usages. in Didactiques et métiers de l'humain et de la relation : nouveaux espaces et dispositifs en question. (direction de Frisch M.), ID Collection L'Harmattan. pp.85-125. Paris, 2013.

[21] Donabédian A., Khaskarian V., Silberztein M. (2013). NooJ Computational Devices. In Formalising Natural Languages with NooJ. Eds. Cambridge Scholars Publishing: Cambridge.

[22] Silberztein M., Anaïd Donabédian. (2013). Formalising Natural Languages with NooJ: selected papers from the NooJ 2012 International Conference. Cambridge Scholars Publishing: Cambridge.

--

NooJ'2013: 1st. proposal (S. Sidhom & P. Lambert).