Modèle d'analyse pour l'activité en Community Management : de la ré-indexation sociale à l'organisation des connaissances en nanosciences

Sahbi Sidhom, Philippe Lambert

Modèle d'analyse pour l'activité en Community Management : de la ré-indexation sociale à l'organisation des connaissances en nanosciences

Sahbi SIDHOM [1] et Philippe LAMBERT [2]

[1] University of Lorraine & LORIA Lab., France,

[2] University of Lorraine / CNRS et NanoSciences, France

[1, 2] e-Mails: sahbi.sidhom@loria.fr, philippe.lambert@univ-lorraine.fr

Résumé :

La contribution de ce travail s'inscrit dans un domaine multidisciplinaire faisant appel au traitement automatique du langage naturel pour construire l'indexation de contenus et développer la recherche d'informations et l'organisation des connaissances dans les réseaux sociaux professionnels. En expérimentation dans le domaine des nanosciences, des valorisations ont été observées par la ré-indexation sociale au travers de nouveaux concepts dans un questionnaire d'une étude spécifique, à savoir : (i) « Quelles sont les raisons pour lesquels le répondant a adhéré au Club nanoMétrologie ? » et (ii) « Qu'est-ce qu'il attend spécifiquement d'une telle structure collaborative ? ». A l'issu des traitements et analyses du questionnaire, des recommandations en matière d'aide à la décision ont pu être proposées pour le rapprochement des activités, des projets et des acteurs associant des compétences.

Mots-clés :

indexation automatique, observations cognitives, organisation des connaissances, réindexation sociale, Community Manager, nanosciences, corpus, enquêtes d'opinion.

1. Introduction

De nos jours, la diversité des outils et applications en « veille scientifique et technique » recouvre plusieurs problématiques et ouvre de nouvelles pistes de réflexion. Ce travail de recherche s'articule autour des « traitements automatiques de la langue naturelle » (En. NLP : natural language processing), l'organisation des connaissances (En. KO : knowledge organization) et la veille en IST : informations scientifique et technique. Dans cette conjoncture, l'apport de la ré-indexation sociale a été mis en valeur : avec le web et ses usages actuels, nous montrons un intérêt qui marque son importance par la valorisation des contenus [6] dans les réseaux sociaux professionnels. Pour nous, il s'agit d'une nouvelle orientation de recherche qui donne un apport au domaine d'étude sur l'indexation automatique des ressources et qui réactualise la réflexion sur l'apport / l'action de l'usager [11], [12].

Prenons l'exemple d'un veilleur ou un professionnel de l'information qui est plongé dans une activité de synthèse/d'écriture, sans contraintes de style ou de rédaction, comme le cas d'un usager sur les réseaux sociaux. Son action préconise au moins l'existence d'une idée mentale. Dans ce contexte, le résultat (ou l'objet écrit de cette action) nous offre la réflexion première pour étudier les structures syntaxiques inhérentes. Ainsi, l'objet écrit est vu comme un ensemble cohérent d'unités linguistiques plus ou moins complexes. Chaque unité s'articule avec les autres et contribue à la réalisation d'un équilibre structurel : à la fois morphologique, syntaxique et sémantique.

Dans une autre considération, l'action d'écriture sans contraintes préconise le relâchement des règles et styles dans la langue de l'auteur : il s'agira bien de la réalisation d'un objet écrit propre à son auteur avec ses structures morpho-syntaxiques et sémantiques.

Dans une dernière considération, l'étude nous amène à examiner des corpus encapsulant les objets écrits de divers acteurs, de les analyser et de découvrir dans cette observation cognitive : « quelle est la diversité des règles employées (aussi bien les nouvelles règles véhiculées) ? » Il s'agit d'examiner les règles dans l'ordre morphologique, syntaxique et sémantique[14].

Dans cette réflexion à problématiser, le choix porté à un modèle morpho-syntaxique est essentiel et ne doit pas, d'une part, perdre de vue la qualité des résultats d'un analyseur (ou parseur) placé dans un système ouvert à l'usage[13] ; Et d'autre part, la qualité de conception (ou le formalisme d'implémentation de l'analyseur) doit rapprocher les concepts théoriques et pratiques sur l'objet d'étude : objets ou traces écrites dans les contenus (documents, IST, analyses de veille, annotations, avis, tags sociaux, etc.) et qu'on pourra étendre aux enquêtes d'opinion (ressources ouvertes sur le web, web des usages, opinions sociaux-professionnels, etc.) [16].

Le corpus utilisé récemment (après celui de l'INA) pour l'observation et la manipulation est fondé sur une enquête d'opinion associant des questions fermées et ouvertes dans le cadre de la constitution d'un Club thématique en nanosciences et nanotechnologies (nanoMétrologie). Le club nanoMétrologie est coordonné par des grands acteurs du domaine en France, comme le Laboratoire National d'Essai et de Métrologie (LNE), le Centre de Compétences NanoSciences France (C'NANO) qui rassemble l'ensemble des laboratoires du domaine. Une hypothèse faite sur la nature du texte libre véhiculé dans les questions ouvertes est qu'il ne s'agit pas de contraintes de style ou de rédaction proposées ou imposées (aux répondants de l'enquête). Pour nous, il s'agit d'un texte libre à préparer pour l'analyse automatique et l'extraction des connaissances[7].

L'objectif de ce travail est d'apporter un modèle adaptatif à l'analyse du langage dans les ressources ouvertes (corpus documentaires, corpus web, corpus web-usages et corpus d'opinions) pour la ré-indexation sociale.

Au terme de ce travail de recherche, nous prendrons appui sur l'analyse de L. Bloomfield1 (1933) qui soulignait particulièrement que : " *Ce qui concerne le sens est le point faible des études sur le langage, et le restera jusqu'à ce que nos connaissances aient avancé bien loin de leur état actuel* ".

Cet article s'articule autour de trois idées pour répondre à la réflexion problématisée, à savoir : « comment valoriser les analyses de contenus et les processus inhérents pour fonder les actions de l'usager et ses apports à l'information et à la connaissance ? ». Pour se faire, nous présentons le processus d'indexation et les apports cognitifs à l'élaboration d'une grammaire d'analyse. Puis, nous présentons le processus de réindexation pour la valorisation des contenus par les usages. En dernier, les expérimentations qui ont permis de développer la méthodologie pour diagnostiquer la structure interne dans un réseau professionnel (celui du club nanoMétrologie) et pour proposer les recommandations adaptées.

2. Indexation de contenus et apports cognitifs

En indexation, cette étude se focalise sur le discours à travers sa matérialité textuelle et s'oriente vers les unités organiques qui le compose tant sur le niveau intensionnel qu'extensionnel [14], [2] dans le langage de description [1], [8] .

[1] : Citation de Bloomfield Leonard (en 1933, *Language*, New York Press) et reproduite par Bobrow Daniel 1968, in Natural language input for a computer problem solving system, *Semantic information processing*, MIT Press, Cambridge.

Pour le corpus d'indexation, les principales étapes du processus consistent en un recueil de notices documentaires incorporant des résumés de contenus chez des fournisseurs spécialisés dans ce domaine. Notre collaboration scientifique[14] avec des spécialistes de l'INA (Institut National de l'Audiovisuel à Paris), a révélé une expérience professionnelle et des acquis qui datent depuis l'ORTF (office de radio et télévision Française) dans les années cinquante.

La construction des résumés issus des contenus documentaires est fondée, dans certains organismes spécialisés comme l'INA (la BNF et autres), selon des critères et des méthodes formelles acquises par l'expérience [3], [4], [5]. Cela permet d'assurer une régularité et une constance des traitements réalisés par les analystes de l'information. Cette richesse documentaire, une fois construite et mise à l'exploitation selon des traits attachés au contenu (ie. la capitalisation des sources d'information et des connaissances associées [19], [20]), pourra s'adapter aux diverses technologies d'exploitation et de diffusion des contenus [9], [10], [11].

2.1. Observations cognitives sur corpus

Dans les corpus de l'INA (ie. les sources de l'INAthèque et l'INAactualités), nous avons cherché à étudier la stabilité des descriptifs textuels (principalement dans les analyses de contenu). Tout particulièrement, il était question d'établir par une analyse statistique les composantes grammaticales et syntaxiques dans la phrase produite : dans les analyses de contenu (cf. Tabs. 1., 2.).

A cet effet, plusieurs situations se présentent nécessitant le repérage des syntagmes nominaux (ou SN) et éviter des structures anaphoriques, des ellipses, des syntagmes nominaux avec le déterminant zéro, etc. Ainsi, il a fallu adopter quelques règles afin d'extraire les syntagmes nominaux de façon homogène pour obtenir

des résultats statistiques cohérents dans un objectif précis : « comment établir une grammaire de réécriture représentative et qui soit fondée sur les corpus ? ». Tout en sachant que les corpus ont été développés par des professionnels, en texte libre et sans contraintes rédactionnelles.

Une manière de résoudre ces problèmes était de s'occuper seulement de l'extraction des syntagmes de surfaces « complets » sans traitement des cas anaphoriques, elliptiques ou cachés. Seuls les SN avec déterminant zéro ont été pris en compte : la facilité de remédier à ce type de problème lors de l'implémentation de l'analyseur morpho-syntaxique [14], [17].

Les structures syntaxiques qui ont subi cette étude sont les syntagmes nominaux complexes (les SN maximaux, en abrégé SN_max), les syntagmes nominaux simples ou inclus dans les SN_max (les SN inclus, en abrégé SN_inc), les syntagmes prépositionnels (SP), les expansions prépositionnelles (EP) et les phrases relatives (REL) complétives dans un SN.

2.2. Elaboration d'une grammaire cognitive

Le corpus de départ constitué de 300 notices d'analyse de contenus audiovisuels de l'INA (sur des émissions radio et télévision) nous a permis de mener à bien l'étude. Chaque notice contient au moins deux champs résumés (« chapeau » pour le résumé synthétique et « résumé » pour le résumé des descriptions détaillées) produits par les professionnels [14], [16].

Une synthèse sur les premières données statistiques obtenues montre une forme d'homogénéité dans les structures syntaxiques dans la phrase, dans les parties : résumé synthétique (chapeau) et résumé détaillé (résumé) : cf. Tab.1.

OBJET ETUDE	SN_max	SN_inc	SP	EP	REL
chapeau	2.56	4.30	4.01	0.38	0.37
résumé	2.05	4.37	3.35	0.66	0.46
phrase	**2.30**	**4.33**	**3.68**	**0.52**	**0.41**

Tab.1 : Analyses structurelles de la phrase résumé INA.

A la suite de cette première analyse purement statistique, nous avons étendu l'étude pour proposer trois modèles possibles de la construction de la phrase INA : (i) la phrase à minima de structures syntaxiques (ou phrase min.), (ii) la phrase à maxima de structures et (iii) la phrase modèle (ou phrase). L'étude permet de dévoiler les structurations suivantes : cf. Tab.2.

MODELE	SN_max	SN_inc	SP	EP	REL
phrase min.	2	4	3	0	0
phrase max.	3	5	4	1	1
phrase	**2**	**4**	**4**	**1**	**0**

Tab.2 : Modèle de la phrase résumé (INA).

En synthèse, l'analyse statistique sur le corpus a révélé une stabilité grammaticale dans les descriptifs textuels (résumés). Cette révélation grammaticale cache en réalité une stabilité de rédaction des textes par les professionnels. Nous avons observé, que les professionnels de l'INA lors de la rédaction des résumés n'ont pas de contraintes rédactionnelles, structurelles ou syntaxiques à respecter, si ce n'est d'appliquer la grille d'analyse de contenu adaptée à un type de document audiovisuel.

2.3. Organisation des structures d'analyse : la connaissance de type SN

Suite à nos observations cognitives sur corpus, nous présentons le modèle de la phrase et les règles morpho-syntaxiques associées qui ont servi à l'implémentation de l'analyseur. Ce modèle est fondé sur une grammaire cognitive du corpus INA. L'objectif donné est de faciliter l'extraction de structures nominales et les propriétés associées qui encapsule un maximum d'informations : des entités nommées formalisables pour représenter des connaissances.

La construction de la phrase (S) dans notre étude s'articule autour de trois structures fondamentales, à savoir : – une structure qui précède la phrase (ou une proposition introductive PI à S), – le syntagme nominal sujet de S (sous forme d'un SN complexe ou SN_max), – le syntagme verbal de S (ou SV), et – la phrase relative (ou REL) en option, qui reste une phrase complétive à un SN. Chacune de ces structures est identifiée en ses unités linguistiques associant son organisation morpho-syntaxique composite[14] :

$$S \rightarrow [PI] + SN + [REL_{SN}] + SV + [REL_{SV}]$$

$$[x]: \textit{ élément facultatif}$$

Le modèle de S est déterminé par une grammaire cognitive qui définit les structures suivantes :

A- Structures préfixées PI à la phrase : Proposition Introductive (PI)

PI	Exemples
1. SP, $\subset$ S	Pour les 20 ans d'AIRBUS INDUSTRIE, …
2. EP, $\subset$ S	En parallèle, …
3. EP + SP, $\subset$ 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,

B- Structures du SN dans la phrase : Syntagme Nominal (SN)

SN	Exemples
11. SN (détails sur SN dans [21])	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 sans déterminant	Psychologues et physiciens (se penchent sur leurs multiples facettes.)

C- Structures du REL dans la phrase : Phrase Relative (REL)

REL	Exemples
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.

D- Structures du SV dans la phrase : Syntagme Verbal (SV)

SV	Exemples
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…

Nous considérons que ce modèle de grammaire syntagmatique comme modèle « cognitif » de la phrase pourra nous servir à la fois comme outil d'indexation ou d'aide à la rédaction de textes et intrinsèquement à l'orientation de son usage pour la ré-indexation.

3. Ré-indexation : corpus des nanosciences et nanotechnologies

Afin de vérifier les observations cognitives présentées précédemment, nous avons appliqué notre démarche aux domaines des nanosciences et nanotechnologies au travers d'une enquête d'opinion associant des questions fermées et ouvertes. Cette enquête s'est déroulée dans le cadre d'une nouvelle structure collaborative, le Club de nanoMétrologie, mise en place par le consortium des laboratoires en nanosciences et nanotechnologies France (C'Nano) et le Laboratoire National d'Essai (LNE). Les principaux objectifs de l'exercice étaient[7] : 1°/ de dresser une cartographie des membres du Club, 2°/ de déterminer les raisons de leur inscriptions et 3°/ ce qu'ils attendaient d'une telle structure collaborative.

Une hypothèse faite sur la nature du texte (ou l'écrit des avis) : l'écrit est (i) libre, (ii) véhiculé dans les questions ouvertes et (iii) pas de contraintes de style ou de rédaction proposées ou imposées. La variété de son contenu décrit sa représentativité comme document pour valider nos choix sur la robustesse de la grammaire cognitive implémentée. Le texte libre proposé par le répondant est proposé pour l'analyse automatique afin de construire l'extraction des connaissances de type SN et ses propriétés.

Lors de l'analyse, nous avons démontré la réutilisabilité du modèle de la grammaire cognitive INA tout en changeant le contexte d'étude sur corpus. Le nouveau corpus préserve les propriétés de cette grammaire et intrinsèquement le modèle de langage implémenté initialement. Dans le corpus d'enquête d'opinion Club nanoMétrologie, les structures identifiées et leur organisation morpho-syntaxique se traduitsent par la sous-grammaire cognitive S' suivante[15], [17] :

$$S' \rightarrow [V\inf] + SN + [REL]$$
$$[x]: \textit{élément facultatif} \quad ; \text{ avec } V_{\inf} \subset \text{PI}.$$

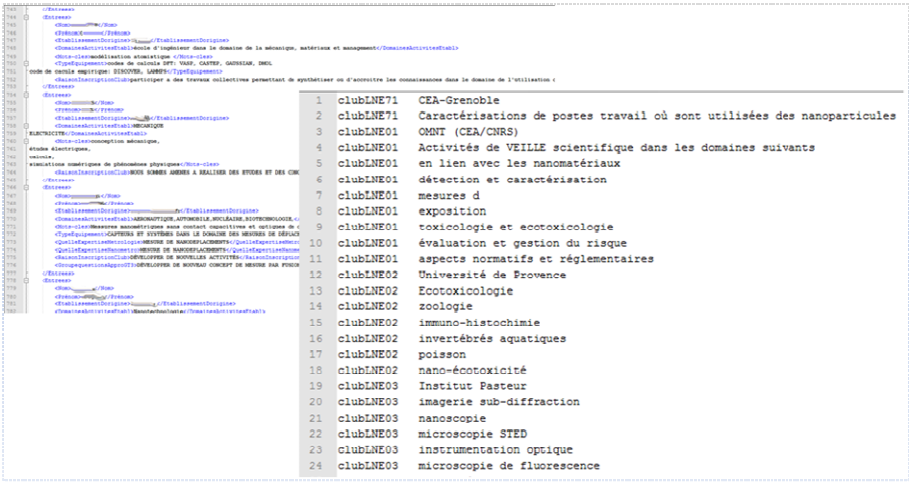
3.1. Implémentation

NooJ est un environnement linguistique développé par Max Silberztein (2005) de l'université de Franche-Comté (France) [18]. NooJ offre la possibilité de créer des grammaires locales (ie. automates finis) complètement paramétrables pour l'extraction d'informations. Les ressources de NooJ sont principalement constituées de dictionnaires (de la langue) et de graphes syntaxiques de type transducteurs à état fini, permettant le repérage d'expressions complexes, l'extraction de lemmes et l'annotation automatique de ressources textuelles [17].

La méthode retenue pour toucher le plus d'adhérents et s'assurer d'avoir un maximum de réponses par l'enquête, consistait à implémenter sous le logiciel open source LimeSurvey (*www.**limesurvey**.org*) un questionnaire en ligne sécurisé pour recueillir les réponses des adhérents. Nous nous concentrerons pour cette étude de cas sur deux principales questions ouvertes qui ont fait l'objet d'un traitement automatique, à savoir : (i) « Quelles sont les raisons pour lesquels le répondant a adhéré au Club ? » et (ii) « Qu'est-ce qu'il attend spécifiquement d'une telle structure collaborative ? ». Une centaine de répondants ont contribué aux réponses de l'enquête et spécifiquement aux deux questions ouvertes.

La méthodologie retenue pour le traitement des réponses comprend cinq phases : (i) la sélection des données-répondants, (ii) Le nettoyage des données, (iii) l'élaboration des ressources linguistiques ad-hoc, (iv) le traitement des données, (v) l'analyse des résultats.

Le système LimeSurvey permet le transfert des réponses au format csv. Pour la facilitation du traitement de la centaine de réponses obtenues, nous avons opté pour un reformatage du fichier de réponse au format XML. Ce choix correspond à deux critères spécifiques de traitement : (i) obtenir une meilleure structuration du document source permettant de retrouver plus aisément les réponses d'origine et (ii) donner la possibilité au système d'itérer sur ce fichier avec des traitements spécifiques selon les nœuds xml que nous voulons étudier dans leur particularité. Le fichier source a également été découpé en autant de textes que de réponse, constituant ainsi un corpus d'une centaine de fichiers xml. Cela a permis de retrouver aisément la source documentaire de l'extraction des granules ou lexèmes durant le processus de TALN (cf. Fig.1).



clubLNE71	CEA-Grenoble
clubLNE71	Caractérisations de postes travail où sont utilisées des nanoparticules
clubLNE01	OMNT (CEA/CNRS)
clubLNE01	Activités de VEILLE scientifique dans les domaines suivants
clubLNE01	en lien avec les nanomatériaux
clubLNE01	détection et caractérisation
clubLNE01	mesures d
clubLNE01	exposition
clubLNE01	toxicologie et ecotoxicologie
clubLNE01	évaluation et gestion du risque
clubLNE01	aspects normatifs et réglementaires
clubLNE02	Université de Provence
clubLNE02	Ecotoxicologie
clubLNE02	zoologie
clubLNE02	immuno-histochimie
clubLNE02	invertébrés aquatiques
clubLNE02	poisson
clubLNE02	nano-écotoxicité
clubLNE03	Institut Pasteur
clubLNE03	imagerie sub-diffraction
clubLNE03	nanoscopie
clubLNE03	microscopie STED
clubLNE03	instrumentation optique
clubLNE03	microscopie de fluorescence

Fig. 1 : Fichier XML des réponses au questionnaire : exemples.

Le nettoyage des données a consisté essentiellement à corriger les fautes d'orthographe et de nettoyer chaque nœud de tout caractère susceptible de créer du bruit dans les résultats d'analyse du système.

L'élaboration des structures linguistiques a été faite en deux étapes : le premier temps a consisté à créer des dictionnaires spécifiques thématiques sur les nanosciences et nanotechnologies. NooJ offre en cela des fonctionnalités utiles en créant automatiquement un dictionnaire constitué des entrées étiquetées comme inconnues (<UNKNOWN>) donnant ainsi l'opportunité de les intégrer au dictionnaire existant. Ainsi, un dictionnaire de plusieurs centaines d'entrées a pu être rapidement créé, rassemblant les thématiques du club métrologie, les techniques et les types de mesure spécifiques pour les nanosciences. Le deuxième temps a concerné la création d'automates à état fini permettant d'extraire les données spécifiques et leur reformatage pour le traitement ultérieur (cf. Fig.2).

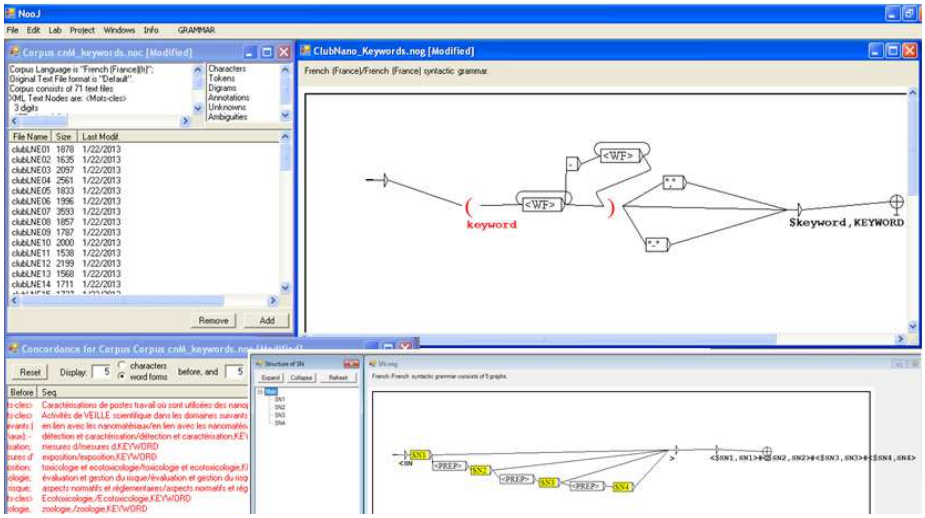


Fig. 2 : Implémentation NooJ : automate d'état fini pour l'extraction des SN en cascade.

Sous NooJ, l'utilisation des automates se fait en cascade. L'itération permet l'étiquetage des structures retenues sur plusieurs niveaux ce qui permet une extraction fine.

La phase de traitement des données a porté sur l'étiquetage des nœuds « Quelles sont les raisons de votre adhésion ? » et « Qu'attendez-vous du club ? » puis sur l'extraction des structures de type S'::= <Verbe> + <SN emboîtés>. On a ainsi obtenu une vingtaine de résultats significatifs montrant que les raisons de la participation au Club sont circonscrites dans un même champ sémantique (i.e. le réseautage) avec la majorité des réponses s'inscrivant dans une logique « *de création d'un réseau* », « *d'intégration d'une communauté* », « d'identification de Community Manager », etc. Le fait que NooJ dispose des fonctionnalités de traitement statistique, lui conférant le statut de système hybride, a permis la hiérarchisation des résultats en se basant sur l'attribution pour chaque extraction d'un indice TF-IDF. Le résultat est une série de besoins exprimés par

les participants, par ordre d'importance, correspondant au poids sémantique détecté par le système.

3.2. Principaux résultats : le Community Manager pour le club nanoMétrologie

Les principaux résultats de cette démarche tiennent de trois ordres :

Le premier consistait à traiter les réponses ouvertes d'un questionnaire dédié à l'identification des besoins en information des membres et des raisons qui les avaient poussées à devenir membres du Club de nanoMétrologie. La hiérarchisation des réponses à ces questions par l'intermédiaire de processus hybride de TALN et statistique en NooJ, ramène des résultats plutôt bons, avec peu de bruits. L'identification a ainsi pu être faite relativement facilement.

Le second objectif concerne davantage l'aspect diagnostique de notre approche, en nous permettant d'obtenir une image à un temps T de la structuration du Club de nanométrologie. Pour cela, une cartographie en réseau a été réalisée (cf. Fig.3). Cette projection permet dans un premier temps d'identifier le positionnement des acteurs par rapport à la thématique générale du Club (ou notion de centralité) puis dans un deuxième temps, d'identifier les signaux faibles, c'est-à-dire les thèmes complètement excentrés, mais qui peuvent se révéler déterminants pour l'évolution du Club. Enfin, ce graphe est également conçu dans une logique de diagnostic [7], [17] : il permettra de comparer les clubs sur une échelle temporelle à T+24 mois et de diagnostiquer quelles thématiques sont les plus fortes (nombre de création de liens entre les acteurs, les thématiques, etc.) et celles sur lesquelles un effort spécifique doit porter pour améliorer leur représentativité et tendre ainsi vers l'exhaustivité (cf. Fig.3).

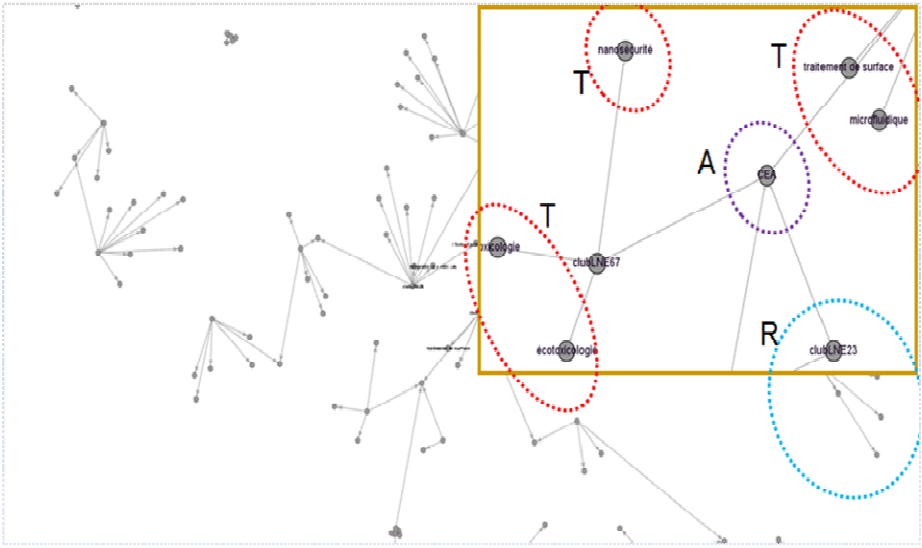


Fig. 3 : Trilogie (A : acteurs, T : thématiques, R : ressources) du Club NanoMétrologie.

4. Conclusion

Dans la synthèse des objectifs construits dans ce travail de recherche, nous avons réussi à apporter un modèle adaptatif d'un analyseur morpho-syntaxique aux ressources ouvertes pour la ré-indexation sociale. Le formalisme d'implémentation avec NooJ et les analyses automatiques nous ont beaucoup rapproché aux concepts théoriques et leur évolution pratique en accord avec la nature de l'objet d'étude : des enquêtes d'opinion et des ressources ouvertes au web usages dans le domaine des nanosciences et technologies.

Dans les résultats d'analyse, nous avons démontré une conjoncture entre les traitements automatiques de la langue et l'organisation des connaissances. Egalement, des valorisations observées par la ré-indexation sociale au travers de nouveaux concepts. Deux questions ouvertes ont fait l'objet d'un traitement automatique spécifique, à savoir : (i) « Quelles sont les raisons pour lesquels le répondant a

adhéré au Club ? » et (ii) « Qu'est-ce qu'il attend spécifiquement d'une telle structure collaborative ? ». A l'issu des traitements et analyses du questionnaire qui a concerné une centaine de répondants. Des recommandations en matière d'aide à la décision[12] ont pu être proposées pour le rapprochement des activités, des projets et des acteurs associant des compétences : ces résultats soulignent la nécessité d'une activité de Community Management. L'intérêt de cette pratique renforce la proactivité des acteurs [7], [17] ainsi que leur cohésion pour l'émergence de nouveaux projets d'appels d'offre en nano [15].

La méthodologie développée permet également de diagnostiquer la structure interne du réseau nano. La détection de la nature hétérogène du réseau peut ainsi être mise à profit pour effectuer un ré-équilibrage autour du centre de gravité de la structure (ou le noyau de cohésion du réseau).

Sur une échelle temporelle à T+24 mois, il est essentiel de diagnostiquer : « comment évoluent les thématiques les plus fortes du réseau ? » Également, « comment considérer la place des réseaux secondaires (ou les réseaux associés au noyau) ? » pour activer la création de liens entre les acteurs, les thématiques et les projets dans une démarche d'innovation.

En perspective, d'autres enquêtes compléteront ce travail sur l'identification de Community Manager en réponse à un besoin direct ou indirect de type projet. Il sera question de mettre en œuvre les valorisations attendues aux nouvelles thématiques émergentes (en signaux faibles) et les compétences existantes par les acteurs et ressources.

Bibliographie

[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] Carbonell J.G., alii. (1997). Translingual Information Retrieval: a comparative evaluation. in Proceedings IJCAI-97, Nagoya, Japan, Morgan Kaufmann, San Mateo, CA (1997).

[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] Maniez J. (1993). L'évolution des langages documentaires. Documentaliste et Sciences de l'information, 1993, vol.30, n°4-5, p.254-259.

[9] 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).

[10] Maret P., Pinon J-M., Martin D. (1994). Capitalisation of consultants' experience in document drafting. Conference Proceedings RIAO 1994, Printed by CID Paris France, p.113-118.

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] SIDHOM S. et LAMBERT, P. (2014). « Project Management in Economic Intelligence: NooJ as diagnostic Tool for Nanometrology cluster". (chapter) in Cambridge Scholars Publishing (in 2014) :NooJ for NLP. "à paraître en 2014", (sous la direction de) Svetla Koeva, Slim Mesfar, Max Silberztein (Eds.)

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

[19] Chaudiron S., Ihadjadene M. (2010), « Electronic Information Access Devices : Crossed Approaches and New Boundaries », in Information Science, London, ISTE, p. 167-189

[20] COUZINET Viviane, 2009. Complexité et document : l'hybridation des médiations dans les zones en rupture, RECIIS, Electronic journal of communication information and innovation in Health, vol.3, n°3, p. 10-16.