Specific ontologies for semantic indexing from natural language properties

Sahbi Sidhom, Nabil Khemiri

Specific ontologies for semantic indexing from natural language properties

Sahbi SIDHOM (LORIA/KIWI & University of Lorraine– France) & **Nabil KHEMIRI** (RIADI - Manouba University - Tunisia)

E-mails: Sahbi.Sidhom@loria.fr, Nabil.Khemiri@gmail.com

Abstract — In this paper, we present a specific ontologies based on the extraction of semantic information in text documents. Document here concerns the context of the valorization of Tunisian patrimony. As approach, we propose to represent semantic properties in document contents from heterogeneous information (multimedia) concerning by the patrimony objects. For indexing and information retrieval (IR), we develop processes based on the noun phrase (NP) properties and their semantic representations. These processes use natural language processing (NLP) to take into account the NP syntactic and semantic structures. In view of this study, the specific ontology designed has the encapsulation principle to capitalize the concept and knowledge as NP and its semantic relations.

Index Terms — ontology, indexing process, information retrieval (IR), noun phrase (NP), natural language processing (NLP), semantic relations, NP parser, NooJ platform.

I. INTRODUCTION

The goal of IR process is to find relevant information in document collections. Then, the most relevant documents are those in the information needs of the user.
IR process differs in approache by: the formulation of the request, the indexing process, the indexing terms for the user query and the matching process between the query and the indexing terms in documents.
By the digitizing industry, document can be represented in several formats: text, image, audio or video, in mixed or separate granularity (proper format, audio-visual or multimedia). Particularly, the multimedia documents are omnipresent for the public access in Internet or social Web in order to disseminate information and share ideas, knowledge and expertise.

To facilitate analysis among the various formats of documents, we have unified the processing throughout text content representations related to document sources [7]: the aim in this part is to extract, from text contents, the concepts and their informative properties.
In the research area of Sidhom (2002) [7], the concept identified for indexing, for IR process and for knowledge organization was the NP (noun phrase or nominal syntagma in French) and its semantic properties in the speech discourses: from writing production to knowledge management (KM).
In the context of this study, the use of the semantic properties and the NP relations in text structures is important: the NP refers to "*the minimal unit of speech that allows naming a person, an object, an idea, etc.*" [11].

This process "*that allows naming person, object, idea, etc.*" is important and specifically concerns the patrimony objects: paintings, textiles, manuscripts, ceramics, glass, mosaics, etc.

For valuing the patrimony objects, we need text descriptions associated to sources, with the point of views of patrimony experts: historians, restorers of patrimony objects, etc.
Thus, text descriptions about patrimony objects can initiate processes such as indexing, annotation, informational filtering, IR, KM on users or experts, etc.

In this paper, our problematic is defined on: "*How to represent the NPs and its semantic properties in the indexing and IR processes to enhance patrimony objects?*".

This problematic includes the following challenges as case studies about:
- How to index the patrimony objects?
- How to valorize the NP properties in a search problem to improve relevant information for user?
- How to generalize the process on descriptions and patrimony objects? In this case, ontologies are needed.

This paper is presented: in section 2, we present and discuss related works as a state of the art on the research domain, with a specific proposition on morpho-syntactic parsing with NooJ platform [9] in section 3. In section 4, we present the semantic of NPs and properties to formalize the IR process. Finally, we propose specific ontologies based on the inheritance of the natural properties the NPs and associated to knowledge management (KM) process.

II. RELATED WORKS

Many research studies tried to extract NP and to represent the content of a document. The complex representations like NP structures are different from single words representations: mono-terms/terms linked with artificial relations to others.
In the case of NP as term, it takes into account a set of words from their production contexts (ie. the natural language). As state of the art on this research domain, we can mention a set of research works, in semantic complexity order, like:

- Le and Chevallet (2006) [1] used a hybrid method that mixes associations between pairs of terms extracted by a statistical approach with semantic relations using a linguistic approach. The NP extraction is processed on syntactic patterns which are a set of rules, in order to concatenate grammatical categories and build NP structures. In this case, NPs are organized into syntactic dependency networks (head and expansion/modifier) by adding the statistical and semantic associations. LE combines different information sources to get a general overview of the NP and its context. This approach combines statistical information based on frequency measurements and syntactic information on the NP structures in dependency networks. The semantic information is defined through the study on relations of: synonymy, hyperonymy and causality.

- Haddad (2003) [2] has proposed a hybrid approach to use the NP in IR process. He used a morpho-syntactic parser integrated in the IOTA system [13]. NPs are represented with dependency networks in the format (head and expansion/modifier like in LE and CHEVALLET [1]). NPs as terms are represented in vectors with their respective weights calculated using a statistical ponderation (i.e. *tf-idf model*). The weight of an index term is calculated by multiplying the statistical weight with the syntactic weight. The syntactic weight of NP is determined by the sum of the syntactic term weights that compose it:

$$e_k = \begin{cases} 1 & \text{if } t_k \text{ is a substantive} \\ 0 & \text{if } t_k \text{ is a stop word} \\ 0.5 & \text{for other categories} \end{cases}$$

with:

e_k: the empirical value of the syntactic element t_k.

The author experiments had been under the vector model. He did not taking into consideration the characteristics and the dependency relations between the NPs.

- Kuramoto (1999) [3] proposed an indexing [4] and IR models with NP structures proposed by users. His study showed the relevance of NP properties and its semantic identified by users, such as defined by the research team SYDO (Research group on "SYstème DOcumentaire" in Lyon, France). In this study, the NP is defined as descriptor of the document [5, 6]. The NP recognition model has been tested on Portuguese corpora. NPs are stored in stacks linked by pointers. The query can be a set of words or a set of NPs. The IR process looks for NPs at first level and browses the NP stack to display the maximal NPs with its lower levels. The IR process proposed a query formulation restriction: it has an impact on the relevant answers.

- Sidhom (2002) [7] presented a morpho-syntactic analysis platform for automatic indexing and information retrieval. His approach was based on the NP formalization and the extraction using NLP formalism: C'ATN (Cascaded ATN: Augmented Transition Networks). He defined the semantic properties of the NP to index documents and developed the IR process applied to INA resources (INA: "Institut National de l'Audiovisuel" in Paris). Theoric and applied proposals in Sidhom researches were to build the knowledge management (KM) system with the NP concepts and its semantic properties. The NP semantic properties are:

- class relation: $NP \in N$,
- fitting relation: $NP_{max} \supseteq NP_1 \supseteq \ldots NP_n$, and
- arborescence relation: $NP_{left/g} \subseteq NP \supseteq NP_{right/d}$)

With appropriate tools, NLP analysis, morpho-syntactic tree parsing, automatic indexing using NP filtering, IR and KM based on NP concepts are developed under the platform.

In the expectations of this work, the NP conceptualization and its semantic properties can refer us to design ontology model for indexing contents and information retrieval.

In follows, we present our research work on the implementation of NPs and its properties [7], to build our parser with a new approach.

III. MORPHO-SYNTACTIC PARSER

Morpho-syntactic analysis (for French language) is a fundamental step which enables the textual analysis of contents, or representations associated to documents, in order to generate the syntactic trees.
After this step, a set of syntactic structures are extracted to supply concepts in indexing and information retrieval processes.
In the aim of our research, we chose a linguistic approach to analyze text representations in multimedia corpora about patrimony objects. The main goal in this aspect is to extract NPs and their semantic relations from syntactic trees in order to index the patrimony objects.
The NP processing chain is illustrated in the following figure (*cf.* Fig. 1).

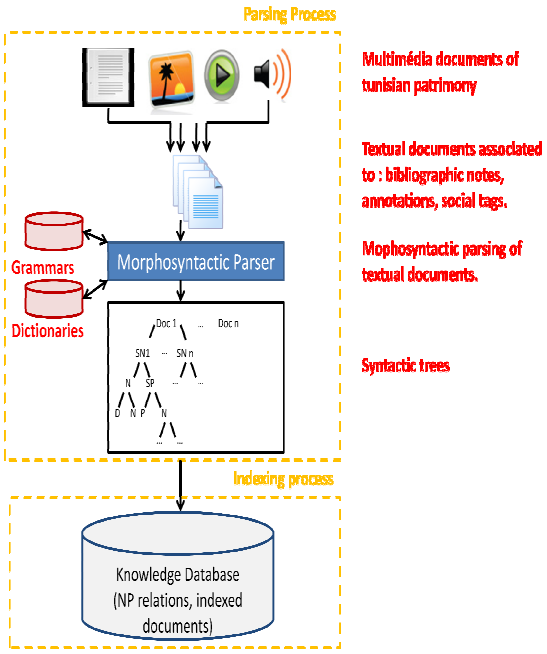


Figure 1: NP processing to index patrimony objects.

In a previous work [8], we built a corpus of Tunisian patrimony images. Documents are composed by bibliographic notes and user annotations to describe each patrimony object. In a first work, the manipulated concepts were mono-terms obtained by filtering semantic words after removing stop words in the text representation.
As an improvement of this work, we had introduced the NP analysis and its semantic properties to enrich the knowledge database (*cf.* Fig.1).

A. PRETREATMENTS

To analyze textual data that represents each multimedia document, a pretreatment phase is needed to prepare the sentence analysis. This stage involves removing punctuation marks and substitute hyphens and elided forms in the original form in the text.

- **Punctuation:**

Strong punctuations (full stop, question mark, exclamation mark and the semi-colon) allow us to cut the original text into sentences. The full stop is considered an ambiguous punctuation. In some cases, it is used in abbreviations, float numbers, email addresses, website addresses and as end of sentence marker. This process was necessary to mark the abbreviations in the text and allow the cutting of the original text into sentences.
Some sentences are too long for automatic analysis. To segment sentences into short sentences, we used some weak punctuation to extract sub-sentences nested in the main. After removal of strong punctuation, we process to all sentences. Also, we extract sentences in brackets, taking into account different levels of possible nesting. When the text is clean of any punctuation mark, we continue process to remove comma and colon, since they are considered as weak punctuation and not useful into sentence analysis.

- **Contracted forms:**

The contracted forms are composed by preposition and determiner. These forms are substituted into the text using a set of rules. For example, the contracted form in French "au" is replaced in the text by "à + le".

- **Elided forms:**

There are two types of elided forms:

- Unambiguous elided forms: they are replaced by applying a set of rules. For example in French: "d'" and "m'" are replaced by "de" and "me" in the original text.
- Ambiguous elided forms: "l'" form is ambiguous because it replaces the article "le" or "la" depending on the context. This form is processed by the parser.

- **Hyphens :**

Some hyphens are deleted according to their contexts. For example, the forms in French "-je" and "-il" are replaced with "je" and "il".

B. MORPHOLOGICAL ANALYSIS

Morphological analysis is used to segment the sentences from the pretreatment step in words. Each word is considered isolated through the inflection (gender, number, conjugation...), derivation and composition analysis. These analysis processes allows to classify each processed word by assigning a lexical nature. The analysis is done by matching the word extracted from a sentence and the dictionary database in order to recognize and classify it (lexically).
In this analysis phase, we use NooJ linguistic platform as processing environment [9]. It was created by Max Silberztein in 2005 at the Franche-Comté University (France).
NooJ can treat different forms of a word using a set of dictionaries: simple words (_delaf.nod), compound words: words formed by joining two or more simple words (_delacfn.nod) and proper names (_Prenoms.nod)…
Morphological analysis prepares the next stage about the morpho-syntactic parsing.

C. MORPHO-SYNTACTIC PARSING

Morpho-syntactic parsing allows identifying the syntactic structure of a given text and explaining the dependent relations between words that compose it. These dependencies are represented by syntax trees. Parsing is based on a set of syntax rules that builds the formal grammar.
NooJ creates a set of grammars (in finite automata) and uses hierarchical graphs. The root graph called the grammar that refers to its components as subgraphs. These grammars uses the transducer formalism to analyze the text of document collections.
To allow this process, we had created a formal grammar based on NP descriptions (*cf.* Figs. 2 and 3) and a set of cascaded graphs. Knowing that, NooJ grammars are defined as type-2 grammar (or regular grammar) according to the Chomsky classification for formal grammars.
To construct the grammar of NP descriptions we used a set of formal rules proposed by Sahbi SIDHOM [7]. These rules were requested to construct all NP graphs and to define NP grammar in NooJ linguistic environment.

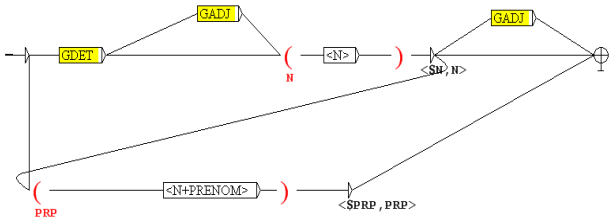


Figure 2: Simple NP graph in NooJ.

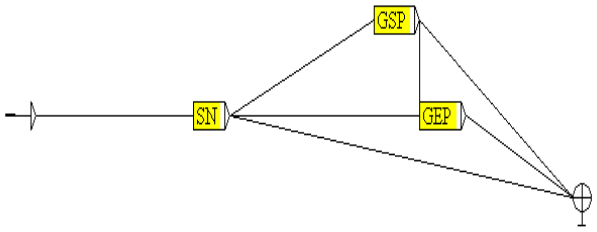


Figure 3: Complex NP graph in NooJ.

The NP definition is *"the minimum unit of speech that allows referring to an object, a person, an idea, etc."* [11]. Each NP has a head which is a noun.
The grammar covers the possible combinations to build NP structures. Knowing that NP can be:

- Simple NP (or SN), like : $SN \rightarrow N''$
- Complex NP, like: $SN \rightarrow N'' + EP' \mid N'' + SP' \mid N'' + EP' + SP'$ | etc.

With:

- *EP'*: One or more prepositional expansions
- *SP'*: One or more prepositional phrases($P + SN$).

In Example, a set of formal rules in NooJ (cf. Tab.1):

Description of graphs (A.) in NooJ Set of NP **::-** N''	
Rules	**Examples**
$N'' \rightarrow D' + N'$	une + mosaïque
$N' \rightarrow N + A'$	mosaïque + ancienne
$N \rightarrow A' + N$	splendide + mosaïque
$N \rightarrow N + EP$	vase + en christal
$N'' \rightarrow NOM\text{-}PRP$	Hannibal Barca
$N'' \rightarrow D' + N + NOM\text{-}PRP$	le + gerrier + Hannibal Barca
$D' \rightarrow D$ | ADV + P-de | …	un | beaucoup de | …
$N' \rightarrow N$	mosaïque | gerrier | …
EP $\rightarrow P + N$ | $(EP)^n$	en + christal | (en christal) à + tête + ronde
SP $\rightarrow P + N''$ | $(SP)^n$	(pendentif) dans une jarre *de Carthage* (+ dans + une + fouille archéologique à Birsa)
…	…

with:

Symbols	**Descriptions**
N'' (or NP)	Noun phrase
D'	Determiner
N	Noun
A'	*Adjectival group*
NOM-PRP	Proper noun
P	Preposition

TAB.1: SET OF NP RULES.

The morpho-syntactic analysis results are a set of syntactic trees. Trees are treated in order to filter NPs with their semantic properties and store them in the knowledge database.

IV. SEMANTIC MODEL: NP-Ontology proposed

Morpho-syntactic parsing should filter the NP structures and construct relations between its components [12]. NooJ platform produces data that contains the NPs and locations in the parsed text, annotated with the grammatical categories of each word.
In related works, several researchers used formal ontologies to represent, organize and access to the domain concepts:
ANDREASEN et al. [14, 15] proposed a new approach that uses "generative ontologies" [16] (SIABO project) for mapping the NPs extracted from documents and queries. NPs are transformed into nodes in the generative ontology. Their approach can measure distances between key concepts in texts.
ZHENG et al. [17] use NPs and semantic relations for text document clustering. The WordNet ontology has been used to improve clustering results.
BANEYX and CHARLET [18] proposed a methodology allows to build an ontology based on texts using a natural language processing (NLP) tools. Their method uses lexico-syntactic patterns to identify semantic relations (hyperonymy, synonymy…) between candidate terms (NP composed by a head and an expansion).
LASSEN and VESTSKOV TERNEY [19] proposed a method to analyze semantic relations between noun phrases that have a preposition relations using a training corpus annotated. NPs are used to construct an ontology-based hierarchical subsumption. The aim of their study is to show relations between NPs heads in the ontology by studying the relations of prepositions.
Our goal is this study is filter NP from document as descriptor and its semantic properties to supply the indexing and IR processes. Rightly, the NP has a natural organization and semantic richness through its own relations like fitting and arborescence relations in the morpho-syntactic parsing [7].

In follows, we explain different semantic relations in NP using morpho-syntactic parsing and how to direct them in indexing and IR processes.

A. NP SEMANTIC PROPERTIES

- **Class relation :**

The head of the NP (*N*) is always represented by a noun. It is used to identify the parts of speech built around the noun: it represents a class of objects having the same feature N (*cf.* Fig. 4).
The class relation is represented by: $N \in SN$.

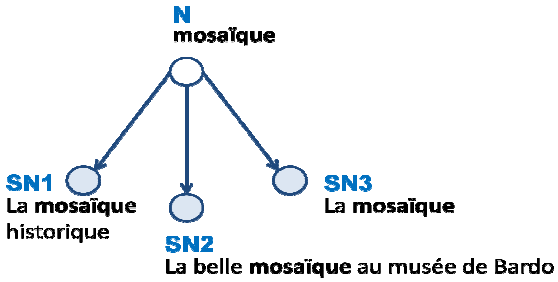


Figure 4: Class relation ($N \in SN$).

- **Fitting relation :**

The NP has fitting relations, where NP can be embedded other NPs (*cf.* Fig. 5).
The fitting relation is represented by: $SN_1 \subseteq SN_2$.

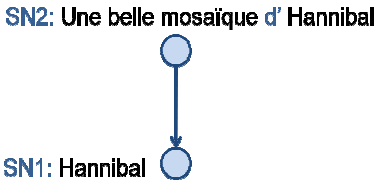


Figure 5: Fitting relation ($SN_1 \subseteq SN_2$).

- **Arborescence relation:**

The arborescence relation is marked by a semantic upsetting in the maximal NP and the result is a double fitting relation (*cf.* Fig. 6).
The arborescence relation is represented by: $SN_g \subseteq SN_{max} \supseteq SN_d$.

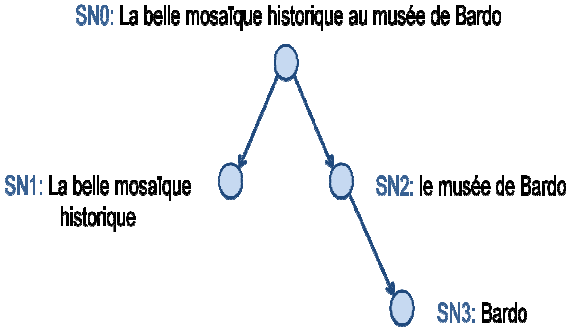


Figure 6: Arborescence relation ($SN_1 \subseteq SN_0 \supseteq SN_2$).

B. INDEXING PROCESS

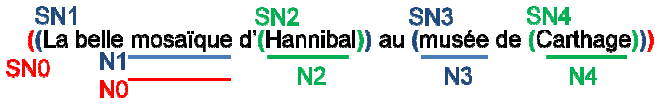
The indexing process consists of creating a formal representation to allow quick access to the information. In this step, we used the NP concepts and properties to create the index database (formal concepts) and the knowledge database that will be used in the IR process.

We illustrate our indexing process with an example:

- **Example :**

Sentence= « La belle mosaïque d'Hannibal au musée de Carthage ».

i) Sentence analysis :

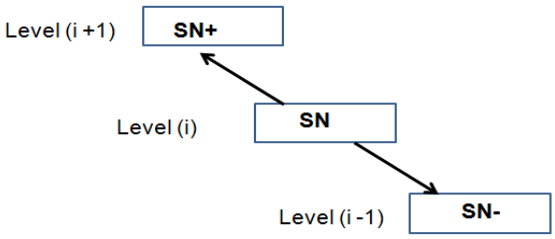


ii) sentence in NP representation:

ID	N	SN_g^-	SN^+	SN_d^-
i-007	mosaïque	La belle mosaïque d'Hannibal	La belle mosaïque d'Hannibal au musée de Carthage	le musée de Carthage
i-007	mosaïque	Hannibal	La belle mosaïque d'Hannibal	-
i-007	Hannibal	-	Hannibal	-
i-007	musée	Carthage	le musée de Carthage	-
i-007	Carthage	-	Carthage	-

iii) NP as formal representation [7]:

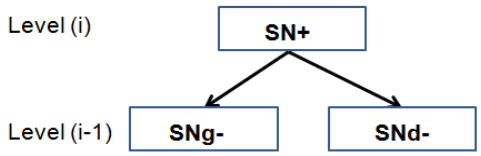
It is a linear representation of the NP and its semantic properties. An arborescence relation is a pair of linear relations: more information to be stored in this representation.



N	SN	SN^+	SN^-
N_0	SN_0	-	SN_1
N_0	SN_0	-	SN_3
N_1	SN_1	SN_0	SN_2
N_2	SN_2	SN_1	-
N_3	SN_3	SN_0	SN_4
N_4	SN_4	SN_3	-

iv) New representation of NP:

We propose an arborescent representation of NP and its semantic properties as new approach. Tree architecture has the advantage of better representation of semantic relations: less information to store.



N	SN_g^-	SN^+	SN_d^-
N_0	SN_1	SN_0	SN_3
N_1	SN_2	SN_1	-
N_2	-	SN_2	-
N_3	SN_4	SN_3	-
N_4	-	SN_4	-

C. IR PROCESS

The IR process allows satisfying a user in informational needs by providing results (semantically) near its request. Knowing that the request maybe a word, a sentence or a text representation (of needs). Queries are analyzed as text content, to extract NPs and their semantic relations. In matching process, the NPs in query are compared with stored in the KM database.

The following algorithm describes the matching between a query (Q) and the collection of documents (ΣD):

Algorithm: IR (Q, ΣD)

```
Begin
   Read(Q);
   Parse(Q);
   IF ( NP(Q) ⊆ NP(ΣD) ) Results ← ΣD' ⊆ ΣD;
   ELSE
   IF (N(Q) ⊆ N(D)) Results ← ΣD' ⊆ΣD;
   ELSE
   IF (Synonym (N(Q)) ⊆ (ΣD))   Results ← ΣD' ⊆ ΣD;
   ELSE {  Q'= Reformulate(Q);
           IR (Q', ΣD);
         }
END
```

Transition from NP representation into O(NP) ontology:
The objective of building ontology is to model knowledge in our study field about the Tunisian patrimony object. Certainly, NP semantic relations have improved relevance in concepts indexing and IR processing. These relations can be inherit into the paradigm of ontology, to enhance the knowledge domain. This domain will be formally represented by concepts and their relations. Thus, ontology formalizes concepts with interpreted semantic relations (like the NP concepts and properties).

In our work, the ontology is also thought to encapsulate different tool types: morpho-syntactic parsing for the NP extraction, indexing and to determine concepts, relations and management throw the NPs. Thus, the conceptual formalization of knowledge domain is represented by the semantic richness of NP concepts, their semantic relations and the new properties in language, as synonyms of N and NPs' similarities. With this new representation, we can calculate the semantic distance between concepts (N and NP) to classify the IR responses in relevance order between concepts.

The properties in this ontological representation O(NP) are:

- **SN[i] :** or NP,
- **N[i] :** head of NP,
- **SN_d^-[i] et SN_g^-[i] :** semantic properties obtained by fitting and arborescence relations,
- **Syn(N[i] , SN[i]) :** synonyms of the noun phrase head N[i] and similar NPs to SN[i],
- **U(N[i] , SN[i]) :** N[i] and SN[i] (as Tags) added by users (annotations).
-

Example: the noun phrase SN[i] = « une galerie »

- N[i] = «galerie»
- Syn(N[i])={corridor, couloir, loggia, péristyle, portique, vestibule}
- Syn(SN[i])={une galerie d'art moderne, une belle exposition dans une galerie d'art...}
- U(N[i])= {exposition}
- U(SN[i])={un musée, musée d'art moderne }

Each ontology O(NP) is represented as a black box which includes the NP, its semantic relations and its new properties (cf. Fig. 8):

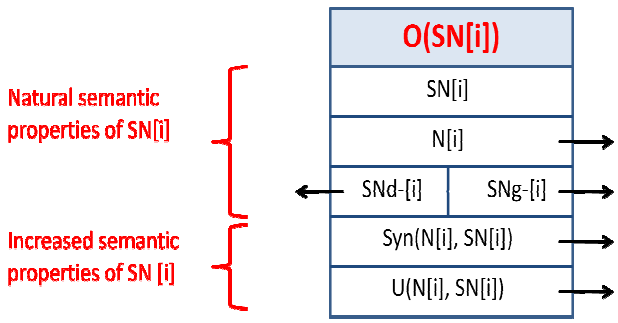


Figure 8: Structure of the NP ontology.

Each Ontology can develop relations with other ontology's connections, others concepts and properties (cf. Fig. 9).

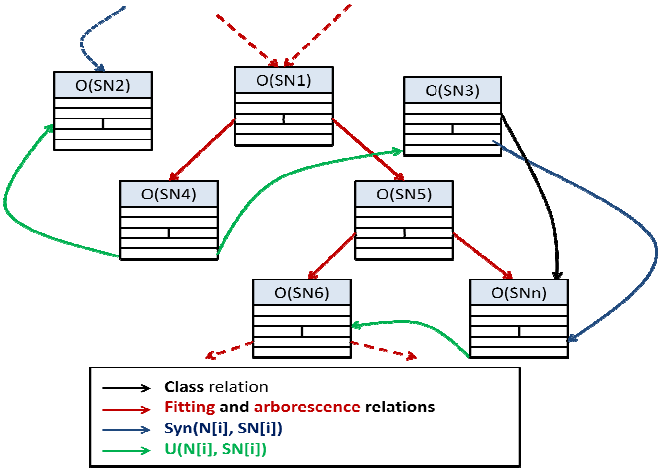


Figure 9: Relations between ontology concepts.

V. CONCLUSION AND FUTURE WORK

In this paper, we present an approach that allows using the semantic properties of the NP concept into automatic analysis applied on patrimony object representations (multimedia documents).

In a first step, we built a morpho-syntactic parser using the NooJ linguistic platform. The analytical results have brought, at first, the indexing process to store the NP concepts and its semantic relations and, at second, the IR process for research relevant information. This structure has, in part, to observe the reuse of the NP properties in the formalization of the ontology. In perspective, this formalization will be studied in IR System in order to measure the user satisfaction degree and his information needs.

About Tunisian patrimony domain, the ontology's formalization can give a new exploitation and experimentation on: - user dimension in the indexing process and - storing information needs and responses in the IR System to reuse stored results.

We can consider the important role that could play this formalization using ontological concepts based NP properties. Closer to user needs in terms of punctual information we can use the descriptive information and its variants which are built by the patrimony experts.

REFERENCES

[1] LE, Thi Hoang Diem and CHEVALLET, Jean-Pierre. (2006). Extraction et structuration des relations multi-types à partir de texte. RIVF'06, pp.53-58.

[2] HADDAD Hatem. (2003). Utilisation des Syntagmes Nominaux dans un Système de Recherche d'Information. BDA 2003.

[3] KURAMOTO Hélio. (1999). Proposition d'un Système de Recherche d'Information Assistée par Ordinateur - Avec application à la langue portugaise. Thèse de doctorat. Lyon: Université Lumière – Lyon 2, 1999.

[4] AMAR Muriel. (1997). Les fondements théoriques de l'indexation : une approche linguistique. Thèse de Doctorat en Science de l'Information et de la Communication : Université Lumière Lyon 2. 1997, p. 410.

[5] DE BRITO Marcilio. (1991). Réalisation d'un analyseur morpho-syntaxique pour la reconnaissance du syntagme nominal : Utilisation des grammaires Affixes. Thèse de Doctorat : Université Lyon 1, 1991, p. 220.

[6] METZGER Jean-Paul. (1988). Syntagmes Nominaux et Information Textuelle : reconnaissance automatique et représentation. Thèse de Doctorat d'Etat Es-Sciences : Université Claude Bernard – Lyon 1, 1988, p. 324.

[7] Sahbi 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 de l'Université Claude Bernard - Lyon I (11/03/2002).

[8] KHEMIRI Nabil, SIDHOM Sahbi, GHENIMA Malek. (2009). Capitalisation des connaissances sur l'objet image du patrimoine : acception de partage et de communication par les acteurs ». Conférence Internationale sur les Systèmes d'Information et Intelligence Economique - SIIE 2009, vol. 1, pp. 913-931.

[9] Silberztein Max. (2005). NooJ: a linguistic annotation system for corpus processing, Proceedings of HLT/EMNLP Human Language Technology Conference, 2005, pp. 10-11.

[10] Silberztein Max. (1999). Indexing large corpora with INTEX. Computer and the Humanities, n° 33-3, 1999, pp. 265-280.

[11] Lambert Ph., Sidhom S. (2010). Knowledge Extraction and Vizualisation: case study on ChroniSanté project in France, SIIE, Sousse, Tunisia, 2010.

[12] Sidhom, S., Robert, C., David A. (2005). De l'information primaire à l'information à valeur ajoutée dans le contexte numérique. Revue maghrébine de documentation et d'information, vol. 1, Tunis, 2005, Tunisie, pp. 95-118.

[13] Palmer P. (1990). Etude d'un analyseur de surface de la langue naturelle. application à l'indexation automatique des textes. Thèse de doctorat, Université Joseph Fourrier, Septembre 1990.

[14] ANDREASEN Troels and alii. (2009). SIABO - Semantic Information Access through Biomedical Ontologies, ic3k/2009 KEOD (International Conference on Knowledge Engineering and Ontology Development), pp.171-176.

[15] ANDREASEN Troels and alii. (2009). Conceptual Indexing of Text Using Ontologies and Lexical Resources, FQAS 2009 - Flexible Query Answering Systems, 2009

[16] FISCHER NILSSON Jørgen and alii. (2009). ONTOGRABBING: Extracting Information from Texts Using Generative Ontologies. FQAS 2009: pp.275-286.

[17] ZHENG Hai-Tao, KANG Bo-Yeong, KIM Hong-Gee. (2009). Exploiting noun phrases and semantic relationships for text document clustering. Information Science 179, Elsevier Science Inc., New York, USA, pp. 2249-2262.

[18] BANEYX Audrey, CHARLET Jean, JAULENT Marie-Christine. (2005). Building medical ontologies based on terminology extraction from texts: Methodological propositions. In Actes des journées « jeunes chercheurs », Paris, France, 14 Octobre 2005. Poster.

[19] LASSEN Tine, VESTSKOV TERNEY Thomas. (2006). An Ontology-Based Approach to Disambiguation of Semantic Relations. EACL-2006: 11th Conference of the European Chapter of the Association for Computational Linguistics.