Towards an Architecture of Ontological Components for the Semantic Web

Nesrine Ben Mustapha[1], Marie-Aude Aufaure[2], and Hajer Baazhaoui-Zghal[1]

[1] Riadi Lab., ENSI Campus Universitaire de la Manouba, 2010 Tunis, Tunisie
{nesrine.benmustapha, hajer.baazaouizghal} @riadi.rnu.tn

[2] Supelec, Computer Science Department, Plateau du Moulon,
91 192 Gif sur Yvette, France
Marie-Aude.Aufaure@Supelec.fr

Abstract. This paper presents an architecture of ontological components for the Semantic Web. Many methods and methodologies can be found in the literature. Generally, they are dedicated to particular data types like text, semi-structured data, relational data, etc. Our work deals with web pages. We first study the state of the art of methodologies defined to learn ontologies from texts. Then, our architecture of ontological components for the Semantic web is defined, in order to improve knowledge discovery from the web. At least, we detail the incremental construction of the domain ontology component and we present the general conceptual model associated to it.

1 Introduction

The volume of available information on the web is growing exponentially. Consequently, integration of heterogeneous data sources and information retrieval become more and more complex. Adding a semantic dimension to web pages is a response to this problem and is known as the semantic web [1]. Ontologies can be seen as a fundamental part of the semantic web. They can be defined as an explicit, formal specification of a shared conceptualization [2]. Indeed, their use allows facilitating web information retrieval, domain knowledge sharing as well as knowledge integration. Meanwhile, building ontology manually is a long and tedious task. Many approaches for learning ontologies can be found in the literature. Section 2 synthesizes such ontology learning methodologies. We present our architecture of ontological components for the semantic web, integrated in a customizable ontology building environment, in section 3. At least, we conclude and give some perspectives for this work.

2 Ontology building methodologies

Methodologies for ontology building can be classified according to the use or not of a-priori knowledge (such as thesaurus, existing ontologies, etc.) and to learning methods. The first ones were dedicated to enterprise ontology development [3] [4] and manually built. Then, methodologies for building ontologies from scratch were developed. They do not use a-priori knowledge. An example of such a methodology is OntoKnowledge [5] which proposes a set of generic techniques, methods and principles for each process (feasibility study, initialization, refinement, evaluation and maintenance). Some research work is dedicated to collaborative ontology building such as CO4 [6] and (KA)2 [7]. Another research area deals with ontology reengineering [8]. Learning methodologies can be distinguished according to their input data type: texts, dictionaries [9], knowledge bases [10], relational [11], [12] and semi-structured data [13], [14], [15].

In the following section, we focus on the general dimensions implied in ontology learning. Section 2.3 deals with ontology learning from texts, including web pages.

2.1 General dimensions implied in ontology learning

The existing methods can be distinguished according to the following criteria: learning sources, the type of ontology to build, techniques used to extract concepts, relationships and axioms, and existing tools. The most recent methodologies generally use a priori knowledge such as thesaurus, minimal ontology, other existing ontologies, etc. Each one proposes different techniques to extract concepts and relationships, but not axioms. These axioms can represent constraints but also inferential domain knowledge. As for instance extraction, we can find techniques based on first order logic [32], on Bayesian learning [33], etc. We have to capitalize the results obtained by the different methods and to characterize existing techniques, their properties and how we can combine them. The objective of this section is to synthesize the characteristics of these methods in order to expose our problematic and to argue our choices.

Learning ontologies is a process requiring at least the following development stages:

- Knowledge sources preparation (textual corpus, collection of web documents), eventually using a priori knowledge (ontology with a high-level abstraction, taxonomy, thesaurus, etc.),
- Data sources preprocessing,
- Concepts and relationships learning,
- Ontology evaluation and validation (generally done by experts)..

The ontology is built according to the following dimensions:

- Input type (data sources, a priori knowledge existence or not, …),
- Tasks involved for preprocessing : simple text linguistic analysis, document classification, text labeling using lexico-syntactic patterns, disambiguating, etc.,
- Learned elements : concepts, relationships, axioms, instances, thematic roles,
- Learning methods characteristics: supervised or not, classification, clustering, rules, linguistic, hybrid,
- Automation level: manual, semi-automatic, automatic, cooperative,

- Characteristics of the ontology to build: structure, representation language, coverage,
- Usage of the ontology and users' needs [16].

2.2 Learning ontologies from texts

The proposed approaches can be classified according to the technique used, namely linguistic, lexico-syntactic patterns extraction, clustering or classification, and hybrid ones. The input data is constituted by linguistic resources like a list of terms and relationships. The huge volume of these data, their quality and relevance has to be taken into account by filtering methods. These methods can be guided by an expert (knowledge acquisition from texts) or automatic (text mining, learning).

Linguistic-based techniques include lexical, syntactic and semantic texts analysis. The objective is to extract a conceptual model of a domain. We can quote two main methodologies defined by [16] and [17]. The methodology defined by [18] intends to extract knowledge from technical documents. Two hypothesis are given by the authors: the first one states that the ontology designer has a good knowledge of the application domain and can determine the relevant terms, while the second one states that he also have a precise idea of the ontology usage. This methodology analyses a corpus with tools appropriate for natural language automatic processing and linguistic techniques and extracts terms and relationships. The normalization step concerns the mapping between natural language and a formal language. The semantic interpretation of the texts is managed by usage and expertise. Semantic relationships are obtained from lexical relationships, and the concepts hierarchy is built using the semantic relationships. The formalization step automatically translates the ontology into a given format like RDF, OWL, etc.

In these linguistic approaches, lexico-syntactic patterns are manually defined by linguists. Some research work has been proposed to automatically extract lexico-syntactic patterns. [19] starts from an existing ontology and extract a set of pairs of concepts linked by relationships, in order to learn hyponymy relationships and produce lexico-syntactic patterns. These ones are used to discover other relationships, based on the learned patterns, between the concepts of the existing ontology. This approach is used to extend an existing lexical ontology. [20] proposes to combine the previous approach with contextual signatures to improve the classification of new concepts. The KAT system (Knowledge acquisition from Texts) [21] includes four steps: learning new concepts, classification, learning relationships and ontology evaluation. Concept classification consists in analyzing words that appears in the expression associated to a candidate concept: [word, seed concept], where "word" can be a noun or an adjective. This classification states that the concept [word, seed concept] subsumes the seed concept that is equivalent to add a hyponymy relationship.

These techniques, based on lexico-syntatic patterns learning, lead to good results for learning hyponymy relationships. In the meantime, some problems appear like terms polysemy or errors produced that are dependant from the corpus. The use of classification techniques like hierarchical or conceptual clustering is a way to solve these problems. The methodology proposed by [22] consists in classifying documents into collections related to words sense, using a labeled corpus and Wordnet. Then for each collection, the relative frequencies are extracted and compared to the other collections. Topic signatures are computed and compared to discover shared words. This methodology is dedicated to enrich concepts of existing ontologies by analyzing web texts. Other methods [23] [24] combine linguistic techniques and clustering to build or extend an ontology.

Some research work is done to study the distribution of words in texts to improve concepts clustering by the way of new similarity measures. DOODLE II, an extension of DOODLE [25], is an environment for the rapid development of domain ontologies. It is based on the analysis of lexical co-occurrences and the construction of a multi-dimensional space of words [26]. This approach extracts taxonomic relationships using Wordnet and non taxonomic relationships learning by searching association rules and extracting pairs of similar concepts using the words multidimensional space.

2.3 Web-based ontology learning

Our main objective is to define an approach to build ontologies for the semantic web. This kind of ontology, closely linked to the web usage, has to integrate the dynamic aspects of the web. In this section, we present some approaches defined specifically for the web.

Many propositions have been done to enrich an existing ontology using web documents [22][27]. However, these approaches are not specifically dedicated to web knowledge extraction.

The approach proposed by [28] attempts to reduce the terminological and conceptual confusion between members of a virtual community. Concepts and relationships are learned from a set of web sites using the Ontolearn tool. The main steps are: the terminology extraction from web sites and web documents data warehouse, the semantic interpretation of terms and the identification of taxonomic relationships.

Some approaches transform html pages into hierarchical semantic structured encoded in XML, taking into account html regularities [29].

Finally, we can also point out some approaches only dedicated to ontology construction from web pages without using any a priori knowledge.

The approach described in [30] is based on the following steps: (1) extract some keywords representative of the domain, (2) find a collection of web sites related to the previous keywords (using for example Google), (3) exhaustive analysis of each web site, (4) the analyzer searches the initial keywords in a web site and finds the preceding and following words; these words are candidates to be a concept, (5) for each selected concept, a statistical analysis is performed based on the number of occurrences of this word in the web sites and at last, (6) for each concept extracted using a window around the initial keyword, a new keyword is defined and the algorithm recursively iterates.

In [31], a method is proposed to extract domain ontology from web sites without using a priori knowledge. This approach takes benefit from the web pages structure and defines a contextual hierarchy. The data preprocessing is an important step to define the more relevant terms to classify. Weights are associated to the terms according to their position in this conceptual hierarchy. Then, these terms are automatically classified and concepts are extracted.

3 Ontological components for the Semantic Web

Starting from the state of the art in ontology learning, we propose a hybrid approach to build domain ontology; our objective is to increase the capability of this ontology to specify and extract web knowledge in order to contribute to the semantic web. Analyzing the web content is a difficult task relative to relevance, redundancies and incoherencies of web structures and information. Moreover, semantic similarity measures highly depends on the quality of data, and the complexity of algorithms such as conceptual clustering increase with the volume of data. For these reasons, proposing an approach to build automatically an ontology still remains utopian.

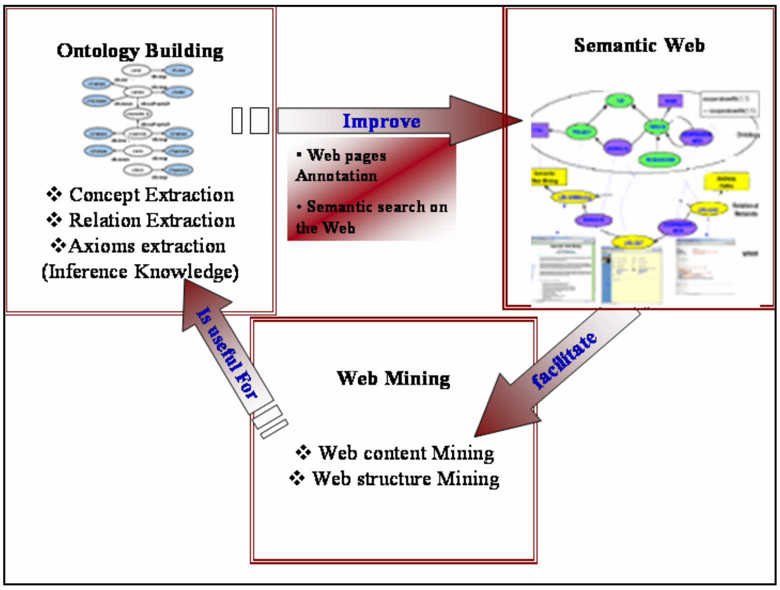

Fig. 1. The cyclic relation between web mining, semantic web and ontology (extracted from [34]

Our approach is based on the cyclic relation between web mining, semantic web and ontology building as stated in [34] and resumed in figure 1. Our proposal is based on the following statements: (1) satisfy the fact that the ontology is useful to specify and extract knowledge from the web, (2) link the semantic content within the web documents structure, and (3) combine linguistic and learning techniques taking into account the scalability and the evolution of the ontology. Our ontology is produced using web mining techniques. We mainly focus on web content and web structure mining. Building this ontology leads us to solve two main problems. The first one is relative to the heterogeneity of web documents structure while the second one is more technical and concerns technical choices to extract concepts, relationships and axioms as well as the selection of learning sources and the scalability. We propose an architecture of ontological components to represent the domain knowledge, the web sites structure and a set of services. These ontological components (figure 3) are integrated into a customizable ontology building environment (figure 2).

3.1. Architecture

Learning ontologies from web sites is a complex task because web pages can contain more images, hypertext and frames than text. Learning concepts is a task that needs texts able to explicitly specify the properties of a particular domain. A positive point in the context of learning ontologies from web pages is that web sites structure can be exploited by web mining techniques.
Starting from the state of the art (section 2.2), we can say that no learning method to extract concepts and relationships is better (in most cases, the ontology evaluation is manually done). For these reason, we propose a customizable ontology building environment as depicted in figure 2. The customization takes into consideration the general dimensions defined in section 2.1.

In this environment, we propose a set of interdependent ontologies to build a Web knowledge base on a particular domain. These ontologies are related to the content, structure and services semantics. Such environment is composed by the following modules: (1) Learning data sources module, (2) Ontological components enrichment module, (3) Linguistic module, (4), Ontological components editor. Theses modules uses resources such as: data warehouse in XML format, linguistic resources (Wordnet, general ontologies, thesaurus, patterns collection), web knowledge bases, constituted by a set of web documents, their structure and associated services, as depicted in figure 3. We distinguish two types of actors in the environment, namely, software actors as the miner agents and human actors as the linguist expert, the domain experts and the system adminstrators. The functionnalities provided by this system are inspired from the web discovery processus, starting from the data sources pretreatment to the knowledge discovery, including the datawarehouse building. Besides, this environment intends to relate domain ontologies, ontologies of services and web structure ontologies in order to build and enrich web knowledge bases of the domain.

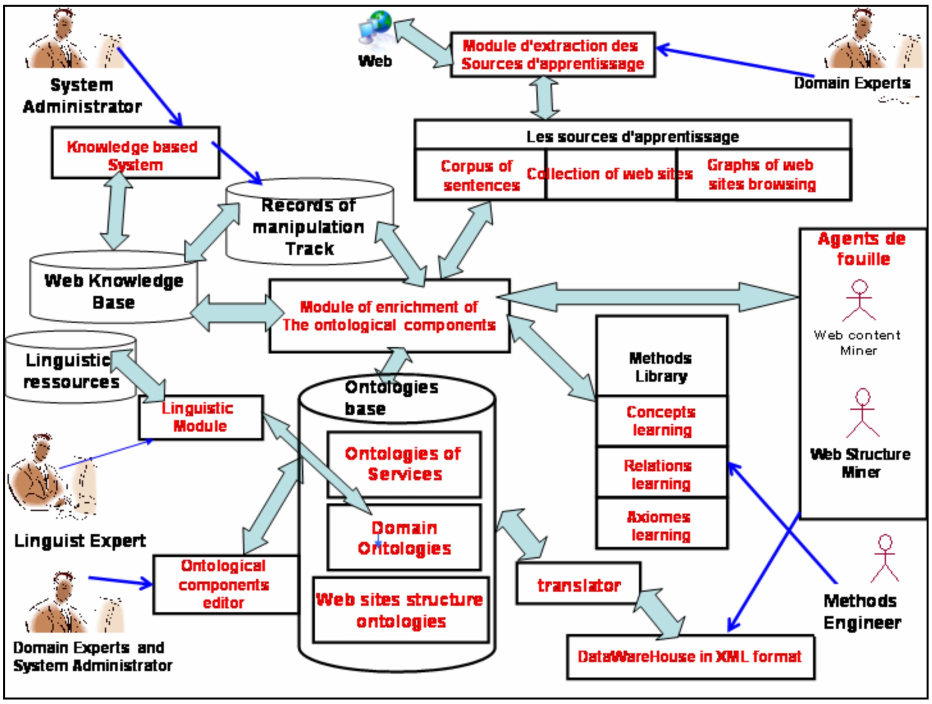

Fig. 2. Customizable ontology building environment

We distinguish three ontologies, namely a generic ontology of web sites structures, a domain ontology and a service ontology.

The generic ontology of web sites structure contains a set of concepts and relationships allowing a common structure description of HTML, XML and DTD web pages. This ontology enables to learn axioms that specify the semantic of web documents patterns. The main objective is to ease structure web mining knowing that the results can help to populate the domain ontology. The ontology of Web sites structure is useful to improve the extraction of the concepts and the relevant relations of the domain by studying semantics of the various markup elements of the languages HTML or XML. We can find in different sites the same contents but presented by different manners which indicates the degree of importance granted to some information according to the context. The second perspective is to translate a semantic relation (e.g. “part-of” relationship, etc.) into a set of adequate markup elements (e.g. a relationship between a concept and instances can be represented by a list, a markup “Small” before a word can express that this word is less important than the previous one). Mining the structure allows us to extract some regularity from which we can define an axiom. Indeed, the axioms of the structure ontology help in elaborating the mapping of HTML markups to the semantic relations inferred. Such an ontology may also adjust the gap existing between the physical and logical structure of a document Web.

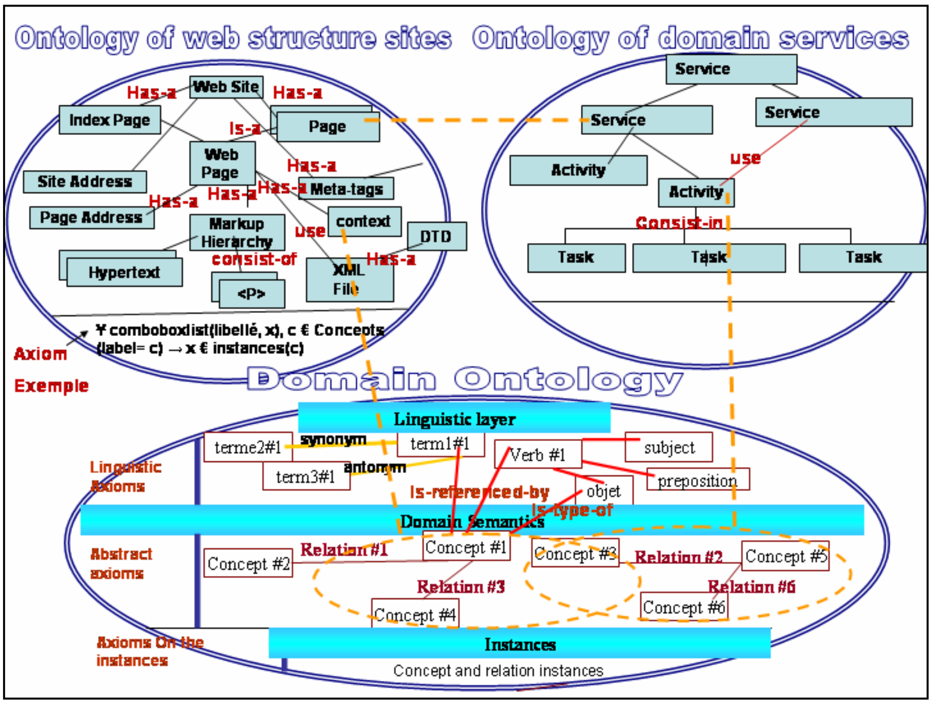

Fig. 3. Architecture of Ontological Components for the Semantic web

The domain ontology is divided into three layers according to their level of abstraction. The first level is a lexical one: this layer specifies high-level lexical knowledge which can help discovering lexico-syntactic patterns. The lexical knowledge covers the concepts, relationships and general axioms of the central layer of our domain ontology. The last layer is more operational and, additionally to concepts and relationships instances, also contains a set of axioms specifying domain knowledge. These axioms are incrementally enhanced by web content and web structure mining.

The ontology of services is defined starting from the concept of task ontology [35]. In our web context, we speak of web services instead of tasks. This ontology specifies the domain services and will be useful to map web knowledge into a set of interdependent services. This is built from the central layer of our domain ontology in order to constitute a set of domain services, and web pages constitute the instances of these services. This ontology is a macroscopic view of the domain and is hierarchically structured: the upper level is the root service while the leaves are elementary tasks for which a triplet “concept-relation-concept” belonging to the domain ontology is associated. These three ontological components are interdependent where the axioms included in an ontology are used to enhance another ontology component. As an example, if we consider the axiom defined on the structure ontology in figure 3, we can say that if the label of a “combobox” is a concept of the domain ontology then all the elements of this structure are instances of this concept. This axiom is used to populate the ontology domain. Meanwhile, these ontologies differ from their use. The domain ontology is used to specify the domain knowledge. The service ontology specifies the common services that can be solicited by web users and can be attached to several ontologies defined on subparts of the domain. As we said previously, the axioms of the structure ontology are used to extract instances of the domain ontology.

The structure of our domain ontology is more complex than usual domain ontologies. This particularity is oultlined in figure 4.

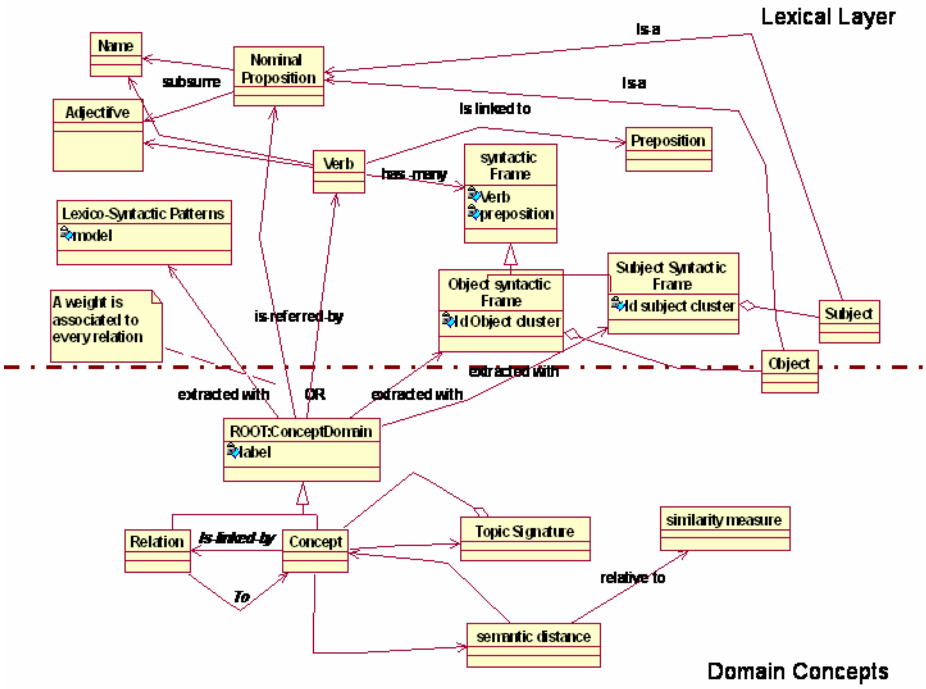

Fig. 4. Domain ontology abstract conceptual model

This figure shows the abstract concepts from which the learned concepts and relationships will be inserted. Confidence weight according to the learning technique used, are associated to concepts and relationships in this ontology. This abstract model is divided into two levels. The linguistic level specifies how a concept was extracted using linguistic techniques, namely the lexicosyntactic patterns [19, 20] or the syntactic frame learning techniques [24]. The domain concepts and relationships are derived from the root of the ontology. They are referenced by verbs or nominal groups. Concepts are extracted by learned lexicosyntaxiques patterns. Such concepts can also be a frequent object or a subject of a syntactic frame. A syntactic frame is a triplet (< subject > < verb > < Object >). The extracted concepts will be weighted by the frequency of the learned patterns in the corpus. Relationships are extracted if a syntactic frame is frequent between two concepts having a close semantic distance. The similarity measures are extensible. We distinguish several semantic distances according to the chosen measure. We quote for example the cosine between two vectors of concepts in the word space and the measure of the closest neighbors between two concepts in the graph of the concepts. As an example, a concept can be learned using a lexico-syntactic pattern. In this case, the weight associated to this concept is relative to the appearance frequency of the concept satisfying the pattern. These weights are updated at each step dedicated to the ontology enhancement. The semantic distance in the conceptual model is related to the similarity measure between two concepts. This measure is computed from a multidimensional space of words.

3.2. Building the domain ontology

In this section, we focus on the domain ontology extraction. Our strategy (figure 5) is based on three steps. The first one is the initialization step. The second one is an incremental learning process based on linguistic and statistic techniques. The last one is a learning step based on web structure mining. We now define them.

The initialization is based on the following steps:

- The design and manual building of a minimal ontology related to the domain; this construction is based on concepts and relationships of Wordnet,
- Composition of concepts and relationships learning sources:
    - Web search of documents related to our domain using the concepts defined in the minimal ontology as requests,
    - Classification of these web documents,
    - Composition of a textual corpus containing a set of phrases in which we can find at least one concept of the minimal domain ontology,
    - Composition of a corpus of HTML and XML documents indexed by their URL.

Each iteration of the second stage includes two steps. The first one (Procedure A) is defined by the following tasks:

- Enrichment of the ontology with new concepts extracted from semi-structured data found in the web pages (XML, DTD, tables),
- Construction of a word space [36] based on the concepts of the minimal domain ontology,
- Lexico-syntactic patterns learning based on the method defined in [20] (by combining lexico-syntactic patterns and topic signatures); these patterns are related to non taxonomic relationships between the concepts of the minimal ontology,
- Lexico-syntactic patterns learning to extract synonymy, hyponymy and part-of relationships (lexical layer of the domain ontology),
- Similarity matrix building: this matrix allows computing the similarity between pairs of concepts found in the multidimensional space word.

The second step (Procedure B) consists in:

- Update the textual corpus and the web documents collection by searching them according to the concepts defined in the minimal ontology,
- New concepts and non taxonomic relationships extraction by the application of lexico-syntactic patterns,

- Attribution of a weight for each extracted relationship relative to the frequency of the relationships that apply the lexico-syntactic pattern,
- Update the minimal ontology.

Each iteration can be validated by the domain expert. This process is incremental: we repeat the procedures A and B until we do not want to integrate new data.

The last stage consists in an enrichment of the structure ontology and an extraction of structure patterns for each relationship of the domain ontology. The objective is to ease instances extraction using the tagged structure of web pages.

The implementation of this strategy is still in progress.

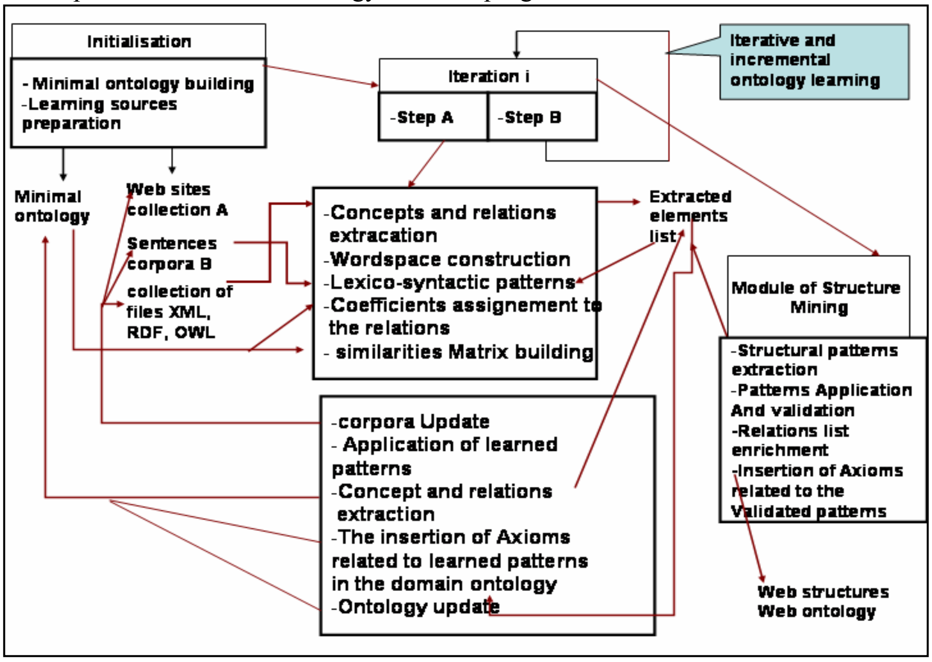

Fig. 5. Learning domain ontology strategy

We have realized a little case study to identify the main characteristics for learning ontology from web sites. From this case study, we have concluded that structure mining techniques are useful for the extraction of non taxonomic and sub-part of relationships as well as for the construction of services ontology. Building a words space is also useful to compute the similarity between concepts but highly depends on web sites corpus. The techniques used are dependant from the learning sources. For this reason, defining a flexible environment should help satisfying the personalization of ontology learning.

4 Conclusion and perspectives

The methodologies for building ontologies described in this paper are fairly complementary. Indeed, methodologies defined to build ontologies from scratch are oriented towards ontological engineering and ontology life-cycle and are based on information systems development methodologies. Learning methodologies try to give a response to the time-consuming manual ontology building task. Learning techniques can be either numeric or symbolic. They have been exploited to semi-automate some foundation tasks such as concept hierarchy building, taxonomic relationships extraction, non taxonomic relationships learning, etc. All these research works constitute a methodological toolbox which can be used to semi-automate ontology construction. We have to take into account previous experiences and to solve cited problems. Let us now outline our contributions in the field of ontology learning and building. Firstly, we conceived an ontological architecture based on a semantic triplet, namely, semantics of the contents, the structure and the services of a domain. So, we take into account both the content and the structure, and a set of services are based upon the domain ontology concepts. The second point is that all the ontologies defined in our architecture contain the weightings of their concepts, relationships and axioms according to their sources. These ontologies could be handled by the inference engine. Concepts and the relationships can be represented by facts and predicates. Axioms are the inference rules allowing the insertion of new facts and predicates.

At least, lexico-syntactic patterns will be stored in the linguistic layer of the domain ontology, so that such patterns will allow the specification of the axioms. These ones define the existing types of relationships, in order to identify such relationships in the web pages. In, our case, axioms are the basis of the incremental enrichment of the ontology. Moreover, the data sources are incrementally updated to satisfy a good semantic coverage of the ontology.

Our perspective is to use semantic web mining techniques and to restructure web pages in order to implement an adaptive web based on the semantic structure, content and services. Our framework presented in this paper is based on an ontological components architecture integrated into a customizable ontology construction environment. We first focus on the automation of the domain ontology construction. Then, we will implement the other ontological components. The final goal is to build a knowledge web base on a specific domain. From our case study, we can conclude that we must take into account various learning sources (like on-line dictionaries) and structure regularities in web sites to go further in the implementation.

References

1. Berners-Lee, T, Hendler, J, Lassila, O.: The Semantic Web, Scientific American (2001)

2. Gruber, T.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, special issue on Formal Ontology in Conceptual Analysis and Knowledge Representation. Eds, Guarino, N. & Poli , R. (1993)

3. Grüninger, M., Fox M.S.: Methodology for the design and evaluation of ontologies. IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada (1995)

4. Uschold, M., King, M.: Towards a Methodology for Building Ontologies. IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing.,Ed. D. Skuce (1995) 6.1-6.10.

5. Staab, S., Schnurr, H.-P., Studer, R., Sure, Y.: Knowledge Processes and Ontologies. In: IEEE Intelligent Systems 16(1), January/Febrary 2001, Special Issue on Knowledge Management

6. Euzenat, J.: Building consensual knowledge bases: context and architecture, Proceedings of 2nd international conference on *building and sharing very large-scale knowledge bases*, Enschede, IOS press, Amsterdam (1995)

7. Decker, S., Erdmann, M., Fensel, D. Studer, R.: Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In: Semantic Issues in Multimedia Systems, Proceedings of DS-8. Kluwer Academic Publisher, Boston, 351-369 (1999)

8. Gòmez-Pérez, A., Rojas, M.D.: Ontological Reengineering and Reuse. European Workshop on Knowledge Acquisition, Modeling and Management (EKAW), Lecture Notes in Artificial Intelligence LNAI 1621 Springer-Verlag, 139-156, eds., Fensel D. & Studer R. (1999)

9. Jannink, J.: Thesaurus Entry Extraction from an On-line Dictionary. In: Proceedings of Fusion '99, Sunnyvale CA (1999)

10. Suryanto, H., Compton, P.: Discovery of Ontologies from Knowledge Bases. Proceedings of the First International Conference on Knowledge Capture, The Association for Computing Machinery, New York, USA, pp171-178 (2001)

11. Rubin D.L., Hewett, M., Oliver, D.E., Klein, T.E, Altman, R.B.: Automatic data acquisition into ontologies from pharmacogenetics relational data sources using declarative object definitions and XML. In: Proceedings of the Pacific Symposium on Biology, Lihue, HI (2002)

12. Stojanovic, L., Stojanovic, N., Volz, R.: Migrating data-intensive Web Sites into the Semantic Web. Proceedings of the 17th ACM symposium on applied computing (SAC), ACM Press, (2002)

13. Deitel,, A., C. Faron,, C., Dieng, R.: Learning ontologies from RDF annotations. In: Proceedings of the IJCAI Workshop in Ontology Learning, Seattle (2001)

14. Papatheodrou, C., Vassiliou, A., Simon, B.: Discovery of Ontologies for Learning Resources Using Word-based Clustering, ED MEDIA 2002, Copyright by AACE, Reprinted, Denver, USA (2002)

15. Volz, R., Oberle, D., Staab, S., Studer, R.: OntoLiFT Prototype. IST Project 2001-33052 WonderWeb Deliverable 11 2003)

16. Aussenac-Gilles, N., Biébow, B, Szulman, S.: Revisiting Ontology Design: A Methodology Based on Corpus Analysis. In: 12th International Conference in Knowledge Engineering and Knowledge Management (EKAW). Juan-Les-Pins, France (2002)

17. Nobécourt, J.: A method to build formal ontologies from text. In: EKAW-2000 Workshop on ontologies and text, Juan-Les-Pins, France (2000)

18. Aussenac-Gilles, N., Biébow, B., Szulman, S.: Corpus Analysis For Conceptual Modelling. Workshop on Ontologies and Text, Knowledge Engineering and Knowledge Management, 12th International Conference EKAW, Juan-les-pins, France (2000)

19. Hearst, M.A.: Automated Discovery of WordNet Relations. In WordNet: In C Fellbaum ed. "*Wordnet An Electronic Lexical Database*". MIT Press, Cambridge, MA, 132-152 (1998)

20. Alfonseca, E., Manandhar, S.: Improving an Ontology Refinement Method with Hyponymy Patterns. Language Resources and Evaluation (LREC-), Las Palmas, Spain (2002)

21. Moldovan, D. I., Girju, R.: Domain-Specic Knowledge Acquisition and Classication using WordNet, In: proceeding of FLAIRS2000 Conference, Orlando (2000)

22. Agirre, E., Ansa, O., Hovy, E., Martinez, D.: Enriching very large ontologies using the WWW. In: Proceedings of the Workshop on Ontology Construction of the European Conference of AI (2000)

23. Khan, L., Luo, F.: Ontology Construction for Information Selection. In: Proc. of 14th IEEE International Conference on Tools with Artificial Intelligence, Washington DC,122-127 (2002)

24. Faure, D., Nédellec, C., Rouveirol, C.: Acquisition of Semantic Knowledge using Machine learning methods: The System ASIUM. Technical Report ICS-TR-88-16, LRI Paris-Sud University, January (1998)

25. Sekiuchi, R., Aoki, C., Kurematsu, M., Yamaguchi, T.: DODDLE: A Domain Ontology Rapid Development Environment. PRICAI98 (1998)

26. Hearst, M., Schütze, H.: Customizing a Lexicon to Better Suit a Computational Task. Proc. of the Workshop on Extracting Lexical Knowledge (1993)

27. Faatz, A., Steinmetz, R.: Ontology enrichment with texts from the WWW. Semantic Web Mining 2nd Workshop at ECML/PKDD-2002, Helsinki, Finland (2002)

28. Navigli, R., Velardi, P.: Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites. Computational Linguistics, MIT press (2004)

29. Davulcu, H., Vadrevu, S., Nagarajan, S.: OntoMiner: Bootstrapping and Populating Ontologies from Domain Specific Websites. Proceedings of the First International Workshop on Semantic Web and Databases (SWDB 2003), Berlin (2003)

30. Sanchez, D., Moreno, A.: Automatic generation of taxonomies from the WWW. In: proceedings of the 5^{th} International Conference on Practical Aspects of Knowledge Management (PAKM 2004). LNAI, Vol. 3336. Vienna, Austria (2004)

31. Karoui, L., Aufaure, M.-A., Bennacer, N.: Ontology Discovery from Web Pages : Application to Tourism. Workshop on Knowledge Discovery and Ontologies (KDO), co-located with ECML/PKDD, Pisa, Italy, Sept. 2004, pp. 115-120 (2004)

32. Junker, M., Sintek, M., Rinck , M.: Learning for Text Categorization and Information Extraction with ILP. In: J. Cussens (eds.), Proceedings of the 1st Workshop on Learning Language in Logic, Bled, Slovenia, 84-93 (1999)

33. Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., Slattery, S.: Learning to construct knowledge bases from the World Wide Web. Artificial Intelligence, (2000)

34. Berendt, B., Hotho, A., Stumme, G.: Towards semantic web mining. In: International Semantic Web Conference, volume 2342 of Lecture Notes in Computer Science, Springer, 264–278 (2002)

35. Gomez-Perez, A., Fernandez-Lopez, M., Corcho, O.: Ontological Engineering. Springer (2003)

36. Yamaguchi, T.: Acquiring Conceptual Relationships from Domain-Specific Texts. Proceedings of the Second Workshop on Ontology Learning OL'2001 Seattle, USA, August 4 (2001)

37. Maedche, A., Volz, R.: The Text-To-Onto Ontology Extraction and Maintenance Environment. Proceedings of the ICDM Workshop on integrating data mining and knowledge management, San Jose, California, USA (2001)

38. Maedche, A., Staab, S.: Ontology Learning for the Semantic Web. IEEE Intelligent Systems,Special Issue on the Semantic Web, (2001)