Lexico-syntactic and Semantic Patterns for Extracting Knowledge from Persian Texts

Mehnroush Shamsfard
Faculty of Electrical and Computer Engineering
Shahid Beheshti University,
Tehran, Iran

***Abstract*—In recent years, knowledge extraction from texts is focused to overcome to bottleneck of building ontologies for semantic web. Pattern based, template driven and linguistic methods are among the successful approaches to extract ontological knowledge from raw texts. This paper introduces some lexico-syntactic and semantic patterns and templates for extracting conceptual knowledge from Persian texts. The described patterns are general and domain/ application independent and work at sentence level. They are used to extract taxonomic and non-taxonomic relations and axioms from phrases and sentences. Among the introduced patterns and templates, semantic patterns are language independent and although linguistic (lexico-syntactic) patterns are introduced for Persian language, they could easily be adopted to other languages such as English.**

This paper will first have a brief overview on linguistic and template driven methods to discover ontological knowledge from texts. Then the templates for Persian simple sentences will be introduced. The extracted relations may be hyponymy, meronymy, attribute/value, part-of, equivalency, etc. The introduced templates exploit semantic information to not only extract the instances of predefined relations (the concepts related by them), but also define new relations. In each case some examples of the experimental results will make the patterns clear.

Keywords- knowledge extraction; linguistic patterns; taxonomic relations; ontology learning; Persian text processing.

I. INTRODUCTION

Nowadays ontologies are playing a major role in computer science and information technology communities. They are used in wide range of fields such as natural language processing, information extraction and retrieval, semantic web, search engines, e-commerce, e-learning, knowledge engineering, multi agent systems, enterprise engineering and medical or geographic information systems.

The major problems in building ontologies are the bottleneck of knowledge acquisition and time-consuming construction of various ontologies for various domains/applications. Meanwhile moving toward automation of ontology construction is a solution [1]. One of the main sources to learn ontologies from, are texts.

Ontology learning from texts which is enabled by extracting conceptual knowledge from texts means to find ontology elements (ontels) such as concepts, relations between concepts, instances and axioms from natural language texts [2]. Knowledge extraction from texts can be done using non-symbolic (statistical or probabilistic) or symbolic (logical, linguistic based and template driven) methods and heuristics may be used to facilitate each one [1]. There are also Hybrid approaches, which combine two or more of the above and employ their benefits and eliminate their limitations. Linguistic and (as a subset of it) template driven approaches are two among other symbolic methods which are mentioned in this section because of their relevancy to the paper.

Linguistic methods such as morpho-syntactic analysis [3], syntactic analysis (in ASIUM [4]), lexico-syntactic pattern parsing [5], semantic processing (in Hasti [1]) and text understanding (in SynDiKATe [6]) are used to extract ontological knowledge from natural language texts. They are mostly language dependent and usually perform the preprocessing on the input text to extract essential knowledge to build ontologies.

In Keyword/ pattern/ template matching methods -which are a subset of linguistic approaches- the input (usually the text) will be searched for predefined keywords, templates or patterns which indicate some relations of interest, e.g. the taxonomic relation.

Next section discusses some related works and compares them with our approach. Then the introduced templates will be described in section 3. The experimental results (section 4) will show the application of our templates on some sample texts.

II. RELATED WORK

There are various types of templates, syntactic or semantic and general or special purpose to extract various ontology elements. As the primary work on pattern matching we can mention the one done by Hearst [7]. In her paper, she introduced some lexico-syntactic patterns in the form of regular expressions to extract hyponymy/ hyperonymy relations from English texts. The followings are Hearst's six patterns which all imply that every NPs in the pattern have hyponymy (ISA) relations to the NPH.

(1) NPH such as {NP,}* {and | or} NP

(2) Such NPH as {NP,}* { or | and} NP

(3) NP {,NP}* {,} or other NPH

(4) NP {,NP}* {,} and other NPH

(5) NPH {,} including {NP,}* {or | and} NP

(6) NPH {,} especially {NP,}* {or | and} NP

Symbolic interpretation rules in [8] use grammatical patterns to map syntactic dependencies onto the semantic relations such as hyperonymy, possession, location, modality, causality, agentivity, etc. The grammatical patterns (markers) indicate syntactic relators (subject, direct object, and prepositions), morpho-syntactic categories of the two related words (verb and noun), and presence or absence of determinant in the Complement.

Another work is done by Sundblad [9] in which some linguistic patterns are used to extract hyponymy and meronymy relations from question corpora such as the followings:

Who is/was X? (X isa person),

What is the location of X?(X isa location),

What is/was the X of Y? (X has Y)

Heyer and colleagues [10] proposed two patterns as follows:

-"(profession) ? (last name)" implies that the unknown category ? is in fact a first name (e.g. actress Julia Roberts) and

-"(class name) like ?" implies that the unknown category ? is in fact an instance name. (e.g. metals like nickel, arsenic and lead).

The patterns may be general and application/domain neutral [7, 9] or specific to a domain or application [3].On the other hand patterns may be manually defined [8, 9] or may be extracted (semi) automatically such as in PROMETHEE [5] and AutoSlog-TG [11].

In this paper we introduce some lexico-syntactic and semantic templates for extracting conceptual knowledge from texts. The templates are developed for simple Persian but can be used with minor changes for English too. The main differences between our work and previous ones such as Hearst's and Sundblad's are listed here:

- Semantic versus lexico-syntactic templates: we have developed semantic templates which uses not only lexemes, POS tags and syntactic categories but also semantic similarities and constraints. So there are some differences in the type of essential knowledge for evoking the pattern and also in the coverage area. Semantic templates may need more background knowledge, more computation and so may be slower but they cover a wider area with more precision and so are more reliable.
- Existing relations versus new defined relations: In most of the previous works there are some predefined relations (especially hypernymy) for which the knowledge extractor find the instances. In other words they find concepts which are related together by some predefined relations. Exploiting our templates we can extract new relations from text before finding an instance of it
- Extending templates for both Persian and English: In this work we have adapted some existing English patterns to cover Persian and also showed their changes to become semantic templates. We have also defined some new semantic templates for both Persian and English.

III. INTRODUCING LINGUISTIC PATTERNS FOR PERSIAN

We have defined two types of patterns or templates for extracting knowledge from simple Persian sentences: syntactic templates and semantic templates.

Syntactic templates are the functional attachments to the grammar, which help the system, make sentence structures. Sentence structures (called SSTs) indicate the thematic roles of various constituents in the sentence. To indicate the thematic roles, the system uses some subcategorization information coded at verb entries in the lexicon. If the entry does not denote the role of an NP after (or before) a specific preposition then the system assigns the default (or possible) role(s) for the complement according to its preposition by using the syntactic templates. The templates show the default role(s) to be assigned after each preposition in combination with a verb. In some cases which more than a role is assigned to a constituent, further processes may resolve the ambiguity, otherwise all possible solutions will be returned.

Some examples of these templates are shown below. The assigned thematic roles are shown within <> in the grammatical rules.

(1) S → np <<agent>> vp-a

(2) S → np <<patient>> vp-p

Vp-a → vp1 | vp2

(3) vp1 → intrans-verb <<action>>

(4) vp2 → …| np <<patient>> 'ra '

{'az' np <<source>>}

{'beh' np <<destination>>}

{'ba' << instrument>>}

trans-verba <<action>> | …

The first and second templates show that the subject of an active verb is usually the agent and the subject of a passive verb is usually the patient. Template (4) says that the direct object of an active transitive verb is the patient and the complement following the preposition 'az' (from) is the source, the np following 'ba' (with/by) is the instrument and the one following 'beh' (to) is the destination all by default. These defaults can be defined the same for all verbs (like for the direct object) or be defined for different verb categories separately. For example we can define a template showing that the complement following preposition 'beh' for verbs which semantically show an abstract transfer (e.g. give, sell) has the beneficiary role while these complements coming with verbs denoting a physical movement has the destination role.

Semantic templates -which are mainly focused in this paper-, are used to convert SSTs to ontels by extracting taxonomic and non-taxonomic relations and axioms. They denote the semantic relations between different roles in a sentence. The proposed semantic templates are divided into four categories [12]:

- Noun phrase patterns,
- Copular patterns,
- Verbal patterns,
- Axiom patterns.

These categories are described in more details in the following:

A. Noun phrase Patterns

Noun phrase patterns are defined to discover relations between different parts of a noun phrase. They are used to extract hyponymy, meronymy attribute-value and possession relations. They include adaptations of Hearst' patterns for Persian, the exception template and the modification and genitive templates to extract relationships between the head and modifiers of a noun phrase.

All of Hearst's patterns can be adapted for Persian by replacing English lexemes to Persian ones (e.g translating 'including" to 'shamel' and 'such as' to 'menjomleh' or 'manand', …). To make more general and language semi-dependent templates we translated English lexemes to their corresponding concepts instead of Persian lexemes. For example we changed the lexeme 'such as' in the first template to <similarity-cue> concept and listed English words 'such as' and 'like' and Persian words 'manand', hamchon', 'mesl', 'menjomleh' and 'az jomleh' under this concept. We call these, language semi-dependent as they are independent to language lexemes but dependant to language grammar so they are sem-syntactic instead of lexico-syntactic patterns. The mentioned language-semi-dependant template will be shown as follows:

NPH <Similarity-Cue> NP-List

We also defined some new templates in the same way to make them ready to use for multi languages. As an example the exception template implies the hyponymy relations as follows:

<Univ-Q> NPH <exception-Cue> NP-List

In which <Univ-Q> concept is lexicalized by a universal quantifier marker and can be 'all' or 'every' in English and 'hameh', or 'har' in Persian, <exception-Cue> is the word denoting an exception such as 'except' or 'excluding' in English or 'joz' or 'bejoz' in Persian and NP-List is a list of Nps separated by comma, and, or.

The above template implies that

(ISA NPi NPH) (for all Npi in the Np-list)

For example the phrase 'all birds except penguin' implies that penguin ISA bird.

Other noun phrase templates relate the head of a noun phrase to its modifiers which (in Persian) may be adjective or other NP. So we have two cases:

(i) If the modifier is an adjective then it should be related to the head as a value to an attribute of head with (has_prop H T V) in which H is the head of NP, T is the modifier attribute and V is the modifier value and means that the attribute 'T' for object 'H' has the value 'V' and according to the system's knowledge T may be known or unknown. As an example, consider the simple phrase (the red apple). Applying the modification template on it will result in the following ontel:

(Has-prop apple-0 atx-0 red-0) if red is an unknown adjective or

(Has-prop apple-0 color red-0) if red is known as a color.

In other words all syntactic constituents recognized by

NP → Noun -e Adj (in Persian) or

NP → Adj Noun (in English)

Should be translated into

(Has-prop Noun ATT Adj)

(ii) If the modifier is an NP, it may have various relations with the head according to its meanings. For example "ketab -e daniel" (Daniel's book) indicates a possession relation between Daniel and the book, while "dar -e otaq" (the door of the room) indicates a genitive case (part-of relation) and "angoshtar -e tala" (the gold ring) shows that the ring is made of gold and so makes an expressive relation and "derakht –e sib" (the apple tree) means that the tree brings apple as the fruit or apple is the name of the tree and so makes a descriptive relation. All of these combinations are shown in the same way in Persian while there are some signs in English for some of them which simplify the recognition of the relation type. In other words identifying the relation between the head and the NP modifier without having enough semantic knowledge about the constituents is not easy. For English we can introduce a simple template to make a possession or part-of (partitive) relation between two NPs connected by "'s" or "of". But in Persian we do not have any explicit sign for genitives so possessive, expressive, descriptive, partitive and infinitive genitives and relations can be embedded in a phrase in the same way. The only sign for such combinations which is shared among all-called Ezafe construction- is a vowel 'e' which is not written in the text but read (I show it by -e in the examples),

On the other hand in some English phrases the first part ends with a preposition when participating in a combination (e.g. waiting for, worried about …). In these cases the connection will be made by the preposition. There is no such case in Persian too and again the connection is made just by Ezafe. "Negaran -e to" (Worried about you) and "Montazar -e bahar" (waiting for spring) are some examples.

As determining the type of genitive combinations in Persian is very difficult and needs some ontological knowledge, at this phase we have assumed a 'has' relation as a default for the genitive type. This default which is extracted from some

statistical study of our sample texts shows both possessive and partitive relationships.

B. Copular Patterns

Copular patterns are exploited to discover hyponym-hypernym, meronym-holonym, and attribute-value relations from copular sentences. A modal (copular) sentence has a noun phrase as its subject, a predicate and a copula verb. The predicate can be one of the followings types (in each case the predicate is written in italic):

(i) A noun phrase such as "Josef mard e bozorg-i ast." (Josef is a great man.) or "baradar -e bozorg -e Maryam doost -e khoob -e hamkar -e jadid -e Daniel ast" (Maryam's old brother is the good friend of Daniel's new colleague),

(ii) An adjective phrase like "Rang e sib sorkh bood." (The color of apple was red) or "lebas -e boland -e Sara tamiz va ziba ast." (Sara's long dress is clean and nice.),

(iii) A prepositional phrase such as "In angoshtar az tala -e sefid ast." (This ring is of white gold) or "ketab -e dastan -e Maryam rooy -e miz -e gerd ast."(Maryam's story book is on the round table).

We have defined some templates to extract knowledge from each of the above cases. Some of them are introduced here using the following grammar:

S -> Subject Predicate Copula

Subject -> NP

Predicate -> Adj* | NP | PP

NP-> Head Adj* NP_modifier*

Adjective predicates

If the predicate is an adjective phrase then three cases may occur:

(i) The head of subject is a property-name and the head of the adjective phrase in the predicate is an instance of it. In this case we extract:

(has-prop <subject.modifier.head> <subject.head> <predicate>)

For example in the sentence "Rang -e machin sabz ast" (Machine's color is green), the head of subject (color) is a property and the head of predicate (green) is an instance of it. So we extract (has-prop machine-0 color-0 green-0)

(ii) The head of subject is a property-name but the adjective in the predicate is not an instance of it. In this case we extract:

(has-prop <subject.modifier.head> <subj.head>)

In such cases we also apply the NP-adjective template on the head of subject and the adjective in the predicate. For example in the sentence "Rang -e machin ziba ast." (Machine's color is nice), the head of subject (color) is a property but the predicate (nice) is not an instance of it. So we imply

(has-prop machine-0 color-0)

(has-prop color-0 atx-0 nice-0)

(iii) Otherwise if the subject is not a property and there is no semantic taxonomic relation between the subject and the predicate we just apply the NP adjective modifier template on the subject and predicate. For example the sentence "machin tamiz ast." (Machine is clean) implies (has-prop machine-0 atx-0 clean-0).

NP predicates

When the predicate is an Np, usually a sub-class relation will be held between the head of the subject and the head of the predicate (e.g. 'The food is beef' or 'horse is an animal'). But determining the direction of this taxonomic relation is not easy. This can be seen in the examples too. In the first one beef is a kind of food and in the second one horse is a kind of animal. We solved this problem as follows:

In general in cases which the predicate is a noun phrase and we know a few about features of the subject and predicate, a co-reference (equality) relation will be held between the heads of the subject and the predicate as follows.

(equal <subject.head> <predicate.head>)

Then new relations will be extracted according to some co-reference rules. Three main co-reference rules are introduced as follows:

Rule-1: (=> (and (HAS a b) (= b c))
(IS-B-OF a c))
(for defining new relations)

Rule-2: (=> (and (= a b) (instance-of a c) (instance-of b c))
(merge a b))

Rule-3: (=> (and (= a b) (one-of b c))
(isa a c))

As an example the sentence "Daniel baradar -e Miriam ast" (Daniel is Miriam's brother) implies (has Miriam brother-0) from NP-genitive-template and (= Daniel brother-0) from copula template for NP-predicates. Now after applying first co-reference rule we will have a new defined relation (is-brother-of) in the extracted knowledge: (Is-brother-of Miriam Daniel).

Another example is sentence "khane -e sabz –e Miriam sara -e bozorg –e Daniel ast" (Miriam's green home is Daniel's big house) which after applying multiple rules will imply:

(and (instance-of home-0 Home)
(has Miriam home-0)
(has Daniel home-0)
(has-prop home-0 color green-0)
(has-prop home-0 size big-0))

As an example for rule 3 consider sentence " John yeki az daneshjouyan ast" (John is one of the students) which implies in:

(isa John Student).

PP predicates- In these cases, the preposition type indicates the relation between the head of subject and the head of predicate. This relation can be spatial, temporal, kind-of, member-of, etc. For example in the sentence "ketab roo –e miz ast" (The book is on the table) there is a spatial (location) relation between the head of subject (book) and the head of predicate (table) and we extract:

(Location-of book table).

C. *Verbal Patterns*

Verbal patterns are used to relate the concepts of thematic roles to the concept of the verb, in non-copular sentences. They are mainly used to extract non-taxonomic relations between thematic roles in a sentence both with themselves and with the verb. The main relation between a concept of a role and the concept of the verb is the role (such as agent, patient, beneficiary …) and relations between various roles depend on the verb. For example in a simple sentence "Maryam sib ra khord" (Maryam ate the apple) following ontels may be extracted by verbal templates:

(agent Maryam eat-0)

(patient apple-0 eat-0)

(Eater apple-0 Maryam)

(Eats Maryam apple-0)

As it can be seen, the first two relations are the thematic roles relating the verb with the role owner, and the second two are non-taxonomic relations between various roles in the sentence (here between agent and patient). By these templates again like copulas, we can learn instances of existing relations (agent, patient, beneficiary, location …) or define new relations (such as eater, eats in, eats by, eaten by …).

To define new relations we have to label non-taxonomic relations between various thematic roles of a verb by linguistic rules. For example in English, the relation from the agent of a verb to its patient will be labeled as '<verb>+s', from patient to agent '<past participle of verb>+ by', from agent to location '<verb> + at', from patient to location/time '<verb+en>+ at', so on. These derivations for Persian are '<verb _stem+andeh>' (or '<present verb>') for agent to patient, '<present-verb + dar>' for agent to location and so on. As an example processing the sentence 'printer prints documents' will result in creating a "prints" relation from printer to documents and a "printed by" relation from documents to printer.

D. *Axiom Patterns*

Axiom patterns are defined to extract axioms from conditional or quantified sentences in the following sequence:

- Separate the antecedent and consequent parts of the sentence. (In compound sentences, these are the subordinate clause and the base sentence respectively).
- Complete the incomplete and necessary roles of each part according to the other part using heuristics for resolving referential and omission ambiguities,
- Create ontels for each part separately and name them antecedent ontels and consequent ontels respectively,
- Make an implication relation (□) between the antecedent ontels and the con-sequent ontels.
- Convert the unbound instances to variables in the axiom,
- Extract and add facts to the axiom to denote the implicit features of the implication such as the temporal, spatial or causality relations.

Here is an example. Applying the above sequence on the sentence "har kasi sam bokhorad mimirad" (anybody who eats poison will die) will result in the following ontels:

```
(=> ( and  (instance-of ?h0 Human)
     (instance-of ?p0 Poison)
     (instance-of ?e0 Eat)
     (instance-of ?t1 Time)
     (Agent ?h0 ?e0)
     (Patient ?p0 ?e0)
     (Time ?e0 ?t1))
     ( and (instance-of ?die-0 Die)
          (Theme ?x ?die-0)
     (instance-of ?t2 Time)
     (Time ?die-0 ?t2)
     (After ?t2 ?t1)))
```

IV. EXPERIMENTAL RESULTS

We implemented the ideas about the introduced patterns and tested them on simple Persian texts. We performed two kinds of tests: first a standalone test to see the results of applying templates on texts and second using templates in an automatic ontology learning system called Hasti [1]. In the standalone tests, we tested our templates on test cases in different domains and complexities, from children's books to newspaper articles. Tests showed that children's books have more matches with the proposed templates due to their simpler structure and syntax. According to our test cases %90 of noun phrases' structures in children's books (for age 4-7 years) matches our NP templates while this percentage for newspaper articles is %70.4. It means that only %10 of the noun phrases in children's books and %29.6 in articles do not have the supposed NP structure and can't use our NP templates to extract knowledge. For sentence templates (copular and verbal) the match percentage in children's books is %85.6 and in newspaper articles is %19.8. The reason is obvious: articles contain more complex sentences than children's books.

Table 1 shows the results of testing our 20 patterns on children's story books (simple texts). In this table, matching rate is the percentage of sentences matching with at least one pattern in the test text. The precision measure shows the

percentage of correct extracted relations among all extracted relations.

TABLE I. THE RESULTS OF TESTING OUR 20 PATTERNS ON SIMPLE TEXTS

Pattern	Matching Rate	Precision
Hypernym	%5	%80
Copular	%30	%90
Verbal	%60	%30
Adjectives	%41	%100
Modifiers	%59	%61
Co-references	%10	%100

Some of the test results are described in more details here.

Syntactic templates - According to our test cases the average precision for determining thematic roles and building SSTs by using default values is %62.5 and without using default values is %47.5. Thus the introduced syntactic templates increase the precision.

NP- Hypernym Extraction Templates - As it was described before, we have 7 patterns to extract Hypernym relations between concepts, 6 patterns for the sem-syntactic version of Hearst's work and one for the exception template. In all matched cases we got correct results from 6 of these patterns. The only pattern we had some problems with was the sem-syntactic and Persian version of the first template of Hearst. In this template we used 'mesl' or 'manand' which are good substitutions for 'such as' and 'like'. There are some examples in which although the text matches the template but the extracted knowledge is not a correct one. For example in sentences such as 'John manand –e ali be madreseh raft.' (John went to school like Ali) the template does not work while in a syntactically similar sentence such as 'keshvarha-yee manand –e Iran, Iraq and Afqhanestan …' (Countries such as Iran, Iraq and Afghanestan …) the template works properly. We defined some heuristics to overcome this problem. They reduced the error rate but did not remove it totally.

Modification Templates- Figure (1) shows a diagram denoting the distribution of various modification types in a sample of about 400 compound Persian phrases in newspaper articles.

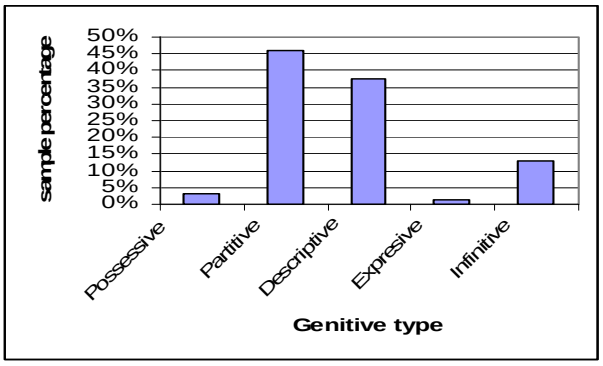

Figure 1. Distribution of genitive types.

The distribution shows that accepting the 'has' relation between the modifier and the head in genitive cases is a good choice for more than %50 of cases.

Copular Templates - Tests showed that the templates for adjective and prepositional phrase predicates are always correct for matched sentences. For Np predicates, in %75.6 of matched test cases the head of subject ISA the head of predicate. In the rest of the cases this relation is reverse.

Verbal Templates - The main errors in testing verbal templates are due to the lack of case frame information for various verbs. If we complete the lexicon by sub-categorization frames of verbs, the process of finding thematic roles and relate their concepts should be done with higher accuracy.

V. CONCLUSION

In this paper I introduced some (lexico|sem)-syntactic and semantic templates to discover new relations and /or extract related concepts by the predefined relations from simple Persian texts. The developed semantic templates use not only lexemes, POS tags and syntactic categories but also semantic similarities and constraints. Most of the templates can be used for other languages such as English with no modifications and some of them need some adaptations to be used by other languages. The templates can extract various types of taxonomic and non-taxonomic relations and axioms. Hyponymy, meronymy, thematic roles, attribute-values, possession and time relations are some of predefined relations which their instances will be found.

Completing the set of templates, enhancing the modifiers templates to distinguish various kinds of NP modification and applying more tests on large corpora are among the future works to complete the current research.

REFERENCES

[1] M. Shamsfard, A.A. Barforoush, Learning Ontologies from texts, International Journal of Human-Computer Studies 60 (2004) 17-63.

[2] M. Shamsfard, Designing the ontology learning model: Protoyping in a Persian text understanding System, Ph.D. Dissertation, AmirKabir University of Technology, Tehran, Iran (2003).

[3] H. Assadi, Knowledge Acquisition from Texts: Using an Automatic Clustering Method Based on Noun-Modifier Relationship, Proceedings of the 35th annual meeting of the association for computational linguistics (ACL'97), (Spain, 1997).

[4] D. Faure, C. Nedellec, A Corpus-based Conceptual Clustering Method for Verb Frames and Ontology Acquisition, in LREC Workshop on Adapting Lexical and Corpus Resources to Sublanguages and Applications, (Granada, Spain, 1998).

[5] M. Finkelstein-Landau, E. Morin, Extracting Semantic Relationships between Terms: Supervised vs. Unsupervised Methods, in Proceedings of International workshop on Ontological Engineering on the Global Information Infrastructure, (Dagstuhl-Castle, Germany, 1999), pp. 71-80.

[6] U. Hahn, M. Romacker, The SYNDIKATE Text Knowledge Base Generator, Data and Knowledge Analysis, 35(2) (2000) 137-159.

[7] M.A. Hearst, Automatic Acquisition of Hyponyms from Large Text Corpora, Proceedings of the Fourteenth International Conference on Computational Linguistics, (Nantes, France, 1992).

[8] P. Gamallo, M. Gonzalez, A. Agustini, G. Lopes, and V.S. de Lima, Mapping Syntactic Dependencies onto Semantic Relations. Proceedings of OLT'2002, (Lyon, France, 2002).

[9] H. Sundblad, Automatic Acquisition of Hyponyms and Meronyms from Question Corpora, Proceedings of the ECAI 2002 Workshop on Machine Learning and Natural Language Processing for Ontology Engineering (OLT'2002), (2002).

[10] G. Heyer, M. Läuter, U. Quasthoff, T. Wittig, C. Wolff, Learning Relations using Collocations, Proc. of IJCAI'2001 Workshop on Ontology Learning (OL'2001), (Seattle, USA, 2001).

[11] E. Riloff, Automatically Generating Extraction Patterns from Untagged Text. Proceedings of the Thirteenth Conference on Artificial Intelligence, (AAAI Press/ MIT Press, 1996). Pp. 1044-1049.

[12] M. Shamsfard, Introducing Linguistic and Semantic Templates for Knowledge Extraction from Texts, Workshop on Ontologies in Text Technology: Approaches to Extract Semantic Knowledge from Structured Information, (Germany, 2007).

Dr. Mehrnoush Shamsfard has been with Shahid Beheshti University, Tehran, Iran.as an Assistant Professor since 2003. She is the founder and the head of Natural Language Processing research laboratory at the Electrical and Computer Engineering Department of of Shahid Beheshti University. Her research interests include natural language processing, ontology engineering, text mining, semantic web and machine learning.