



## Ontology Equipped Natural Language Processing for Real World Applications

Shabina Dhuria, Dr. Harmunish Taneja  
Department of Computer Science  
DAV College, Sector-10, Chandigarh, India

**Abstract**— Natural language processing is an area of artificial intelligence that defines a set of methods and techniques used to automate the translation process between computers and humans or mediate the human-machine communication. NLP focuses on developing systems that allow computers to communicate with people using everyday language and also concerns how computational methods can aid the understanding of human language. To make this possible, NLP has to define a communication format and software able to analyze, understand and give appropriate response. For the communication level, a formal representation of the knowledge is needed that can be represented by ontology. Ontology represents knowledge in a formal model based on conceptualization. This paper has threefold objective. Firstly, it defines ontology as a concept based model to represent objects, concepts, entities and relations that exist between objects. Users' access to information and the presentation of information to users are both mediated via natural language, and the ontology used in the reasoning component is coupled with the lexicon used in the natural language component. Secondly, it throws light on ontology framework that is based on the resource description framework (i.e. defines ontology metadata) and classification of ontology languages. Thirdly, it focuses on role of ontology in natural language processing applications viz. Knowledge representation, semantic annotation and conceptual indexing along with various advantages and disadvantages.

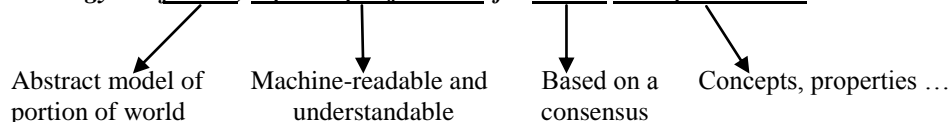
**Keywords**— Artificial Intelligence, Natural Language Processing, Ontology, Resource Description Framework, Semantic Web.

### I. INTRODUCTION

Ontology is defined as representing knowledge in a formal model based on conceptualization; knowledge area shall be understood as objects, concepts and other entities that are assumed to exist and the relations that exist among them [1]. Based on the purpose, context, coverage and usage, ontology can be general, middle or specific. Ontology defines a common vocabulary and shared understanding. Natural Language Processing (NLP) is a computational technique that helps in analysing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications [18]. It is the sub-domain of artificial intelligence concerned with the task of developing programs possessing some capability of “understanding” (i.e. transformation from one representation (the input text) to another (internal representation)), a natural language in order to achieve some specific goal. NLP purpose is to design systems having capability to make inferences and respond to user queries with correct answers. Using ontology in natural language processing is a relatively new part of artificial intelligence.

“Ontologies are formal, explicit specifications to represent the objects, concepts, and other entities in a particular system, as well as the relationships between them” [2].

**Ontology is a formal, explicit specification of a shared conceptualization.**

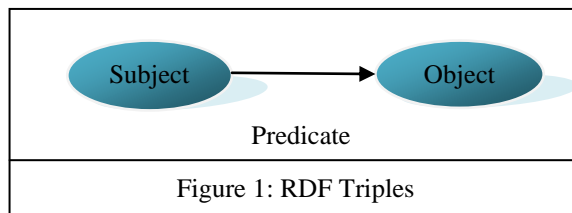


Ontology is a data structure in which symbols that represent conceptualizations are defined and manipulated by NLP software. Following can be considered among various ontologies:

- Taxonomies on the Web [4]
  - Goggle Directory
- Catalogs for on-line shopping
  - Amazon.com product catalog
- Domain-specific standard terminology
  - Unified Medical Language System (UMLS)

## II. ONTOLOGY FRAMEWORK

Ontology framework provides formalism for building ontology, without providing the contents. Resource Description Framework (RDF) is used to define ontology metadata [5]. It provides a structured metadata about the content i.e. adding semantics to a document. Information is represented in RDF by a set of triples subject-predicate-object.



In RDF triples, Subjects and objects are represented by nodes, and predicates that point from the subject to the object of the triple are represented by arcs. It is a simple data model for defining objects and their relationships thereby it offers a way to make data richer and more flexible [6]. It provides a unifying ontological syntax for defining knowledge bases. Ontological framework organizes information that is used in artificial intelligence, semantic web, biomedical informatics, and system engineering and information architecture as a form of knowledge representation [7]. It follows two steps that provide:

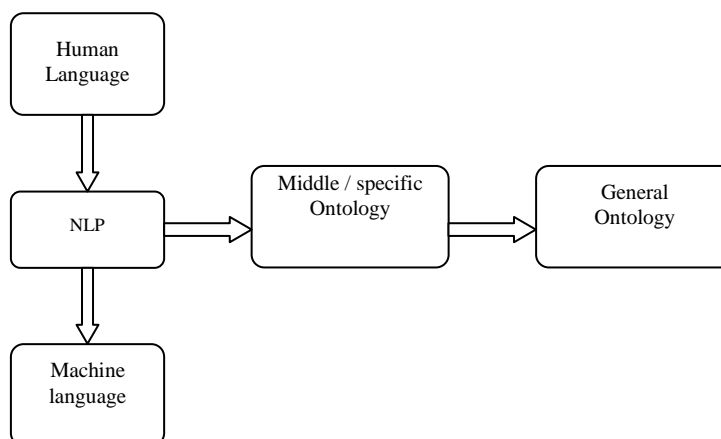
- A formal language in which the ontology can be expressed or specified, and
- Reasoning capabilities that helps in consistent ontology descriptions (i.e. free of contradictions, assuming that contradictions indicate modelling mistakes or errors) [8]

## III. ONTOLOGY LANGUAGES FOR IMPROVED NLP PERFORMANCE

Ontology languages are formal languages used to construct ontology in computer science and artificial intelligence. It allows the encoding of knowledge about specific domains and having reasoning / inference rules to support processing of knowledge. Ontology languages are declarative languages and follow generalization of frame languages [9]. These languages can be classified as logical languages, frame based languages and graph based languages as mentioned below in table2:

<b>Languages</b>	<b>Features</b>
Logical Languages	<ul style="list-style-type: none"> <li>• First order predicate logic</li> <li>• Rule based logic</li> <li>• Description logic</li> </ul>
Frame based Languages	<ul style="list-style-type: none"> <li>• Similar to relational databases</li> </ul>
Graph based Languages	<ul style="list-style-type: none"> <li>• Semantic network</li> </ul>

Ontology language defines a well-defined syntax, necessary for machine processing of ontologies. The formal semantic based description of ontology helps in answering queries based on reasoning processes. Ontology language specifies the vocabulary along with their meaning that is helpful in automated reasoning. It has sufficient expressive power to model the domain of natural language processing and to represent human knowledge. Figure 2 depicts NLP mapping to middle or specific ontology that can exploit the reasoning process for efficient computing in NLP environment [10]. Hence, it provides an efficient, powerful and understandable reasoning mechanism.



**Figure 2: Ontology Integrated NLP**

#### IV. ONTOLOGY EMBED NLP REAL WORLD APPLICATIONS

Various NLP applications such as information extraction, machine translation, sentiment analysis and question answering, require both syntactic and semantic analysis at different levels [11]. Some of these real time applications are limited by the lack of world knowledge. Creation of a domain specific ontologies help to solve this problem of world knowledge [12].

- Knowledge representation [3] (i.e. Ontology models domain knowledge that helps in natural language understanding).
- Question Answering [12] (i.e. information retrieval: to provide inexperienced users with flexible access to information, allow them to write a query in natural language thereby result in concise manner).
- Semantic annotation [13] [14] (i.e. Ontology is used as a vocabulary, classification or indexing on a collection of items).
- Semantic search [3] (i.e. Ontology is used as a query vocabulary or for query rewriting purposes).
- Configuration (i.e. Ontology defines correct configuration templates).
- Conceptual indexing [19] (i.e. Ontology provides conceptual background knowledge and semantic extraction from text).
- Query expansion (i.e. evaluates a user's input and expands the search query to match additional documents).
- Document Clustering [20] (i.e. process of grouping text documents into category groups).
- Text Categorization and Automatic Summarization.

Ontologies play pivotal roles across all parts of the framework. It provides reasoning and inference, guide content filtering (with and without inference) and tag concepts in text documents. NLP is a computational technique that helps in analysing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks [11]. NLP applied work in ontology-based information extraction for specific domains, e.g. biomedical, business intelligence. Integration of ontology and NLP results in combination of (semi-)automated knowledge acquisition with the rich knowledge representation by following these two steps:

- Ontology provides the domain model
- NLP facilitates extraction of knowledge from text documents

NLP integration with ontology having various advantages and disadvantages as follows:

Advantages	Disadvantages
<ul style="list-style-type: none"> <li>• Richer context models</li> <li>• Resolution of ambiguities</li> <li>• Improved NLP performance</li> <li>• Ontology provides the domain model</li> <li>• NLP facilitates extraction of knowledge from text documents</li> <li>• Integration of rich knowledge representation with semi-automated knowledge acquisition [15]</li> </ul>	<ul style="list-style-type: none"> <li>• Semantic gap [10]</li> <li>• Lack of ontology</li> <li>• Computational overhead</li> <li>• Compatibility issues between different approaches [16] Viz. domain model &lt;=&gt; statistical NLP</li> <li>• No unifying ontology: NLP suffers from the lack of a unifying ontology that addresses semantic as well as syntactic representation.</li> </ul>

Hence, richer context models and resolution of ambiguities improves the NLP performance. Knowledge representation and NLP moving slowly back together. Knowledge representation provides ontologies for use in NLP i.e. information extraction. NLP provides input for ontology development i.e. text mining.

#### V. CONCLUSION

Ontology is a formal model for representing information on concept basis and is a formal knowledge repository used as a resource for natural language generation tasks. Integration of NLP and domain specific ontology solves the problem of world knowledge confronted by various real time applications. Therefore by applying ontological framework, the system performance is enhanced and NLP is exploited efficiently for computing the information. Ontology framework uses resource description framework to add semantics to the documents. RDF follows a set of triples, provides a unifying ontological syntax for defining knowledge bases. Such system plays an important role in the semantic search of a web page on ontology basis.

#### REFERENCES

- [1] N. F. Noy, "Semantic integration: a survey of ontology-based approaches", *ACM SIGMOD, New York, USA*, vol. 33, issue 4, pp. 65-70, Dec. 2004
- [2] T. R. Gruber, "A translation approach to portable ontologies: Knowledge Acquisition", vol.5, issue 2, pp. 199-220, 1993.
- [3] Studer, Benjamins and Fensel, "Knowledge Engineering: Principles and Methods, Data and Knowledge Engineering", pp. 161-197, 1998.
- [4] P. Matuszek, "Artificial Intelligence Ontologies", 2008.

- [5] <http://code.google.com/p/rdfliib>
- [6] [http://en.wikibooks.org/wiki/Semantic\\_Web/The\\_Vision](http://en.wikibooks.org/wiki/Semantic_Web/The_Vision)
- [7] [http://en.wikipedia.org/wiki/Ontology\\_\(information\\_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science))
- [8] D. Estival, C. Nowak and A. Zschorn, "Towards Ontology-Based Natural Language Processing", *In Proceedings of the 4<sup>th</sup> Workshop on NLP and XML: RDF/RDFS and OWL in Language Technology*, pp. 59-66, 2004.
- [9] V. Maniraj and R. Sivakumar, "Ontology Languages – A Review", *International Journal of Computer Theory and Engineering*, vol. 2, issue 6, Dec. 2010.
- [10] <http://opennlp.sourceforge.net/>
- [11] Shabina Dhuria and Harmunish Taneja, "Natural Language Processing: A Backbone for Computational Linguistics", *National Conference on Computational Sanskrit: Issues and Challenges*, 2013, pp. 187-189, ISBN: 978-93-82529-03-3.
- [12] K. Bontcheva and B. Davis, "Natural Language Generation from Ontologies", *Springer*, pp.113-127, 2009.
- [13] P. Buitelaar, P. Cimiano and B. Magnini, "Ontology Learning from Text: Methods, Evaluation and Applications", *IOS Press*, 2005.
- [14] P. Cimiano, "Ontology Learning and Population from Text: Algorithms, Evaluation and Applications", *Springer*, 2006.
- [15] F. J. Kurfess, "Combining Ontologies and Natural Language Processing for Knowledge Retrieval", 2011.
- [16] Wimalasuriya, D. C. and Dou, "Ontology-based information extraction: An introduction and a survey of current approaches", *Journal of Information Science*, pp.306-323, 2010.
- [17] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology", Mar. 2001.
- [18] Xiaoyong Liu, "Natural Language Processing", School of Information Studies at Syracuse University.
- [19] T. Andreasen, H. Bulskov, P. A. Jensen and T. Lassen, "Conceptual Indexing of Text Using Ontologies and Lexical Resources", *8th International Conference*, Denmark, pp. 323-332, 2009.
- [20] S.C.Punitha, K Mugunthadevi and M Punithavalli, "Impact of Ontology based Approach on Document Clustering", *International Journal of Computer Applications*, vol. 22, issue 2, pp. 22–26, May 2011.