Modeling the NLP Research Domain using Ontologies

An Ontology Representation of NLP Concepts from a Research Perspective

Auhood Abdullah Alfaries
Information Technology Department
King Saud University
Riyadh, Saudi Arabia
aalfaries@ksu.edu.sa

Randa Hamad Aljably
Computer Science Department
Shaqra University
Druma, Saudi Arabia
raljebly@su.edu.sa

Muna Saleh Al-Razgan
Information Technology Department
King Saud University
Riyadh, Saudi Arabia
malrazgan@ksu.edu.sa

***Abstract*—Natural language processing is an active research filed that will benefit greatly from a shared understanding of its concept terms. Modeling this research domain using an ontological semantic representation that defines and links its terms will definitely benefit the research community. This paper aims to present an ontology model for NLP. This ontological model will serve as a tool in analyzing and understanding the knowledge of the domain and to help practitioners in extracting and aggregating information and then generating an explicit formal and reusable representation of the NLP concepts along with their interrelationships.**

Keywords—NLP; ontologies style; knowledge domain; concepts

I. Introduction

Natural language processing (NLP) is a major area of artificial intelligence research. It involves using computers to recognize and understand human's natural language [1], [2]. NLP began in the 1950s as the intersection of artificial intelligence and computational linguistics. It serves as a field of application and interaction of a number of other traditional AI areas [3]. This field of science focused on limiting the computer's need for human experience about the knowledge in linguistics or statistical algorithms.

Furthermore, it's internally connected to text mining and corpus linguistics [4] but focuses on knowledge representation, logical reasoning, and constraint satisfaction. In the last decade, NLP has made a dramatic shift in statistical methods applications, such as machine learning and data mining. This paved the road for learning and optimization methods that constitute the core of modern AI and NLP, most notably genetic algorithms and neural networks [1].

Originally NLP was distinct from Text Information Retrieval (IR), Natural language generation (NLG) and Language Engineering (LE) [5], [6]. Although these three fields combine their basic techniques like document clustering, filtering, new event detection. NLP are distinguished itself by developing applications to extract meaning from text, represent knowledge, reason logic and satisfy constraints [5], [7].

However, this field of science has proven to be particularly beneficial for business companies such as Google and Yahoo in using an astronomical amount of resources to perform NLP tasks. According to the research center at Google, NLP has been used to solve problems in multiple languages at web scale. Similarly, Yahoo has integrated NLP into their sports products to makes it easy for football fans to search through stats and find only those relevant to making picks and monitoring their team.

Additional examples of NLP multi lingual applications include Sentence Segmentation, Deep Analytics, Named Entity Extraction, Part-of-Speech (POS), Tagging, Word Sense Disambiguation, Chunking, Information Extraction, Semantic Role Labeling, and Dependency Analysis [2], [6], [8], [9]. These applications use various techniques such as stemming, part of speech tagging, compound recognition, de-compounding, chunking, word sense disambiguation and others.

Despite the advances in NLP research; real world natural language systems and applications still cannot perform high level tasks. More research is needed for commonsense reasoning or extract knowledge in a general and robust manner. Accordingly, NLP goal has not been achieved in automatic understanding natural language [4], [5]. Recently the research area focused on using superficial yet powerful techniques instead of unrestricted knowledge and reasoning capabilities.

An emerging collaboration between Semantic Web, language processing and standards improve resource reusability and interoperability through development of standards for encoding annotation data [8]. In our research, we aim to focus on standards based on a unitedly recognized NLP components. These standards will illuminate the inconsistency in NLP presentation, documentation and ongoing research. Especially when different practitioners and researchers refer to different texts on NLP theories and principles and find that their own knowledge base and terminology is different from each other.

Secondly, we intend to address the issue of lack of shared understanding concerning the nature of NLP by creating an ontological model used by researchers, students and experts to share and annotate information in their works [10].

Given that ontology includes machine interpretable definitions of basic concepts in the domain and relations among them. It defines a common vocabulary for researchers who need to share information in a domain. Therefore, we will introduce an ontology model to provide solid ground for common understanding of the structure of NLP. We will introduce an ontological model representing the definition of NLP's common vocabulary, concepts, interconnected objects and complex structure [11]. This model will assure that domain assumptions are explicit and will separate domain knowledge from operational knowledge and further on will support the design and development of the future NLP software systems [3], [12], [13].

This assumed support can be accomplished through providence of general guidance to the structure of successfully designed NLP systems alongside their optimal results. Secondly the literature of the results of different assembly of different components within these systems such as the combination of different approaches in algorithms with different models. Additionally, providing suggestions for approaches that have not yet been done, and a simple prediction for their results based on the state of the art ones.

The purpose of this research is to develop an ontology models alongside an NLP development environment, it will provide reusable implementations for the rapidly changing components and allow an implementation of complex field applications and systems. Such an ontology model will be a key element in the learning process and will contribute to eLearning and enhance the students' performance and comprehension once its deployed into the LMS [14]. To create adaptive e-learning environments and reusable educational resources, we intend to develop dynamic applications such as questions' bank that help students test and evaluate their knowledge level and project idea enumeration application. This is accomplished by building a lexicon model to relate ontology concepts, relations and instances to lexical items and their occurrences in texts.

The paper is organized as follows: Section 2 outlines literature on related works, Section 3 lists the contribution point, and illustrates the proposed methodology, Section 4 evaluates the proposed methodology, Section 5 discusses the results and finally Section 6 is the conclusion and future work.

II. Research Background

Ontologies play an important role in information organization, domain concepts and relationships in many fields. Ontologies are applied in various fields such as artificial intelligence, Semantic Web, biomedical informatics, natural language processing, and information architecture. In this work we will illustrate examples upper level and lower level ontologies, as discussed in the research.

A. Upper Level Ontologies

The first example of ontology is Suggested Upper Merged Ontology (SUMO). SUMO is an open source upper-level ontology that acts as foundation ontology for many specific domains through defining general purpose terms and concepts. SUMO created by merging publicly available ontological content into a single, comprehensive, and cohesive structure [15], [16]. SUMO can be applied in NLP applications, semantic search and semantic web. It can be used to create ontologies for specific domains such as financial transactions and the Quality of Service (QoS) [16].

Secondly, Basic Formal Ontology (BFO) supports scientific research ontologies. It consists of several ontologies categorized into two categories that represent some distinct perspective of reality: ontologies of type SNAP (snapshot) and ontologies of type SPAN [17], [18]. BFO is applied to the biomedical domain [19].

The third example BabelNet is constructed automatically. It is a very large multilingual semantic network and ontology, which is generated by a mapping between Wikipedia and WordNet. Wikipedia provides large amounts of semantic relations. Thus when they overlap, the two resources complement the information about the same named entities or concepts [20]. Another similar application that uses this approach is the YAGO ontology which is built from extracted facts and semantics from WordNet and Wikipedia [21].

Fourth is Cyc, the largest existing common-sense knowledge base. Its ontology makes a heavy use of higher-order logic constructs such as a context system. Many of these higher-order constructs are believed to be key to Cyc's ability to represent common-sense knowledge and reason efficiently [22]. Cyc has been used in the domains of natural language processing. In particular, for the tasks of word sense disambiguation and question answering and network risk assessment [17].

Fifth work is GATE's ontology model consists of a class hierarchy. At the top is a taxonomy class which is capable of representing taxonomies of concepts, instances, and inheritance between them. At the next level is ontology classes which can represent also properties, i.e., relate concepts to other concepts or instances. Properties can have cardinality restrictions and be symmetric and/or transitive. There are also methods that provide access to their sub- and super-properties and inverse properties. The property model distinguishes between object (relating two concepts) and datatype properties (relating a concept and a datatype such as string or number) [8].

Sixth is YAMATO. It covers issues which exiting upper ontologies fail to explain such as quality and quantity, ontology of representation, and distinction between processes and events [23], [24].

B. Lower Level Ontologies

The first example in this category is Dublin Core. A language that contains a set of terms to describe resources, such as Resources has title: "The Picture of Dorian Gray" or Resources has date created: 1890. It has two levels: The Element Set which consists of fifteen elements. The Qualifiers included two classes: encoding schema to parse rules that aid in the interpretation of an elements value and element refinements to make a meaning more specific [25].

Secondly is FOAF (Friend of a Friend). A project that describes persons and how they are relative to other people and to objects. FOAF links persons and information, which help the merchants to gain search about their user's intentions. It identifies users and their interest for the web site to show the users the recommendation that interest [26].

Thirdly is Music Ontology. An ontology based on FOAF. It expresses tastes of music consumers and provides a vocabulary to link music related data on the web [27], [28].

C. Hybrid Ontologies

On the other hand, we present an ontology that includes upper level and a lower level ontology which is Gellish English Dictionary. It is a public domain standard data and knowledge representation language and ontology. It includes taxonomy and a universal database implementation method [27]. Gellish is not the only example of open source; Foundational Model of Anatomy (FMA) is open source that references ontology for the human anatomy discipline. It is concerned with the representation of types or classes and relationships for the symbolic representation of the phenotypic structure of the human body [28].

III. PROPOSED NLP ONTOLOGY METHODOLOGY

The proposed methodology will build an ontology using the Stanford University Top – Down ontology creation strategy. Our methodology populates ontology with concepts and their instances discovered in texts [8], [29]-[34].

Our NLP ontology is organized by a subset of NLP concepts representing NLP domain knowledge. These concepts represent the vocabulary of basic NLP terms and their meanings. The reason for the usage of concepts rather than words is to facilitate the understanding of common NLP knowledge that may have different explanations in different contexts and resources which can make the meaning of terms confusing and ambiguous, the ontology is created and implemented using OWL 2 language and Protégé tool.

Ontology construction is the core phase, which involves the creation of ontology framework, and then it will evolve to reach the final ontology. Each phase has input, output, participants and activities. This methodology is organized into four steps as shown below, see Fig. 1.

Step 1: Collect important terms and vocabulary, taxonomy, and relations

The NLP ontology necessarily embodies part of a world view with respect to the domain. This world view is often conceived as a set of key concepts. Obviously there is no formal definition of what a key concept is. However, we used some selection criteria to elect and choose n umber of concepts.

The first criteria is the density of a term. It choses concepts that are information rich in an ontological sense. The second criteria is the number of occurrence. Those concepts that have higher occurrence rates indicate a strong involvement within the NLP field. The third criteria is generality, as the NLP domain expands rapidly we highlighted concepts that were general rather than specific to certain techniques or systems. These three criteria's yield a sample of 20 concepts in which it is seemed reasonable to begin with in the ontology construction and subject to expansion.

We created a list of terms that represent concepts, relations among the terms, any properties that the concepts may have. Our method for determining the domain list of informal terms is adapted from text books and research articles literature. We then arranged the concepts into classes and slots. Table 1 displays a small sample of the general terms that exist in NLP domain.

Step 2: Define formal concepts, concept hierarchy

In this process, we proceed by considering how to identify what the important concepts and ideas are in a domain of interest thus limiting the scope of the ontology.

There are several approaches to develop the class hierarchy (is_a relationship between concepts) such as top down, bottom-up and middle-out development approaches. The proposed methodology will follow the first approach, as shown in Table 2 and Fig. 1.

The information for these concepts, their domains ,values, hierarchy and description as in Tables 1 to 4 have been withdrawn from previous related research in NLP [9], [35]-[55]. In addition, researches in [56]-[62] were used in deducing Object and Data properties found in Tables 3 and 4.

TABLE I. SMALL SAMPLE OF ONTOLOGY KEY TERMS

General	NLP specific	NLP specific
Science	System	Language Translation
Natural language	Algorithm	Language processing
Theory	Model	Rule
Method	Application	Meaning
Approach	NLP Task	NLP challenges

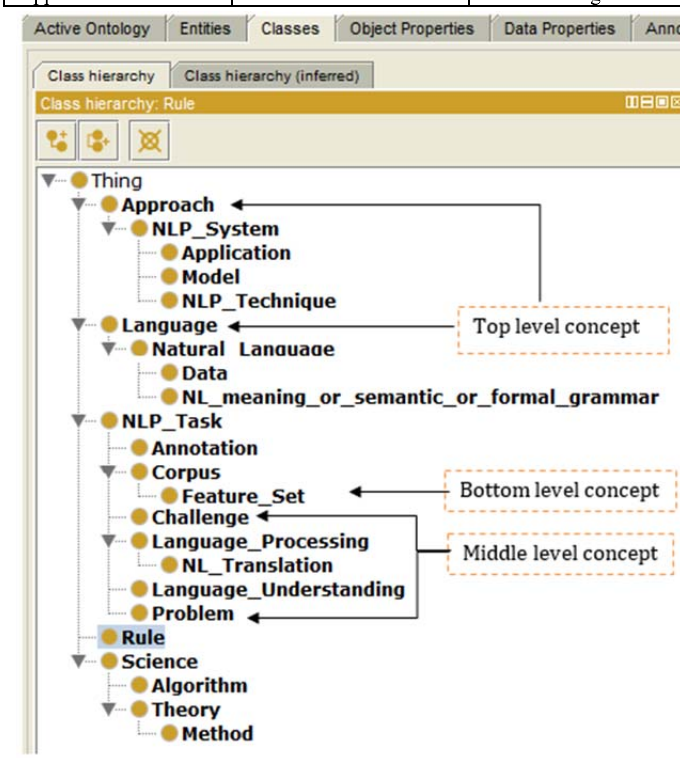

Fig. 1. Class hierarchy.

Step 3: Define the properties

Here we described the classes' internal structure in which we defined the data properties of the classes and object properties between classes and define the relationships between properties. (i.e. is_a properties hierarchy).

Tables 2 and 3 describe some of these properties in some detail as domain and range of each property. The first column denoted the object properties between concepts, in other words the name of the classes relationship, the second and third columns show the domain (subject concept) and the range (object concept) and the description denotes the proper way to read the property and its relation to other properties such as being a sub property of another property.

TABLE II. CLASS HIERARCHY

Class notation	**Super classes**	**Subclasses**
Approach	None	NLP system
NLP system	Approach	Application Model Technique
Application	NLP system	None
Model	NLP system	None
NLP Technique	NLP system	None
Language	None	Natural language
Natural language	Language	Data. NL meaning. NL Semantics. Formal Grammar.
Annotation	NLP Task	None
Data	Natural language	None
NL meaning	Natural language	None
NL Semantics	Natural language	None
Formal Grammar	Natural language	None
Corpus building	NLP Task	Feature Set
Algorithm	Science	None
Science	None	Algorithm. Theory.
NLP Task	None	Challenge Language Processing Language understanding Problem Corpus building Annotation
Language Processing	NLP Task	NL Translation
NL Translation	Language Processing	None
Rule	None	None
Theory	Science	Method

Step 4: Define the facets on the slots (constraints)

Here we defined different value types for the data properties (i.e. string, numeric, date/time, other concept, Boolean, enumerated …). Slots are placeholders for values associated with instances of class. To determine the set of slots for an instance, the class precedence list for the instance's class is examined in order from most specific to most general. A class is more specific than its super classes. Table 4 describes some of these properties in some detail as domain, range of each property and description of the property or if it’s related to other properties.

TABLE III. SAMPLE OF ONTOLOGY OBJECT PROPERTIES

Object Properties	**Domain**	**Range**	**Description**
A branch of	Science	Sciences	A science that is a subfield of another one
Super field of	Science	Sciences	A science that is a super field of another one
Concerned with	Science	Natural language	Science that studies NL
Involves	Science	Algorithm	Algorithm is part of Science
Is –a	Natural language	other languages	NL is a Sub class of other languages
Is –a	Natural language	other languages	NL is a Super class of other languages
Translated by	Natural language	NLP System	NLP systems translate NL
Concerned with	Challenges	Science	Challenge connected to NLP field of science
Facing	Challenges	NL	Challenge facing NL
Evolves into	Challenges	NLP system	Challenge evolves into NLP system
Over comes	NLP task	Challenges	NLP task overcome challenges
introduces	NLP task	NLP Model + NLP system	NLP task introduces NLP models and NLP system
develop for	Applications	NLP Task	Applications are developed to accomplish task
Accomplishes	NLP system	NLP task	NLP system accomplishes an NLP task
Produce	NLP system	Result	NLP system produces annotation results
Faces	NLP system	challenge	NLP system faces challenges
Based on	NLP system	Algorithm	NLP system uses algorithms
Produces	Algorithm	NLP system	An algorithm produces and NLP system
presents input to	model	(ML) algorithm	A model presents input to ML algorithm
Facilitates experiments on	Algorithm	Data sets	An algorithm facilitates experiments on Datasets
provided for	features	ML algorithm	Features are provided for ML algorithm
Executed on	Algorithm	Corpora	The algorithm is executed on a corpora
Capture linguistic patterns of	Model	Data	Model captures (relationships, properties) of data
Provide info	Model	System	Model provides semantic ,meaning, synonyms information to systems
Minimizes risk of	NLP Techniques	Model over fitting	Techniques seek to minimize the risk of overfitting the model
Defines	Theory	Language	Theory defines a language
Used in	Method	NLP application	A method is used in an NLP application

TABLE IV. SAMPLE OF ONTOLOGY DATA PROPERTIES

Data properties	Domain	Range	Description
Automatic annotation	Corpus	Boolean	Corpus Automatic annotation feature
Ambiguity	Natural language	Literal	Natural language ambiguity attribute
Area of study	Science	Literal	Science Area of interest
Complexity level	Challenge	Literal	The complexity level of the challenge
Data type	Data	Literal	Data type of the Data
Structure	Data	Literal	Sub property of the Data type pf Data.
Origin	Natural language	Literal	The Origin (Source) of a Natural language
Development Data	NLP Task	Literal	The Development Data used in the NLP Task
Difficulty	NL Translation	Literal	The level of difficulty of the translating NL
Form	NL	Literal	The form of the NL
History	Science	Literal	The milestones and literature of a science
Impact	Science	Literal	The impact of the Science on the world
Intention	NL translation	Literal	The intention of the NL translation
NLP Task	Corpus	Literal	The language of the Data in the Corpus
SVM Model	Model	Literal	Is the model considered a SVM model
Uneven margins	Model	Literal	Sub property of SVM model
Size	Corpus	Double	The size of the corpus
Evaluation metric	NLP system	Literal	A sub property of system name
Beginning	Science	Date time	The Beginning date of the science
Formality	Natural language	Boolean	The formality level of the Natural language
Input type	Feature set	Literal	The Input type of the Feature set
Model Type	Model	Literal	The model Type of the Model
Processes	NL translation	Literal	The Process included in the NL translation
Challenge category	Challenge	Literal	The category of the NLP challenge
Cause	challenge	Literal	Sub property of Challenge category

Step 5: Define axioms for complex concepts and create instances

Some complex concepts and properties can be created from existing ones by means supported by the ontology language. We created individual instances of classes adapted from real world, in our case.

Step 6: Select implementation language and tool

We generated the ontology using OWL 2 language and Protégé 4.2 tool. Fig. 2 and 3 are Protégé snapshots to show the built ontology's classes.

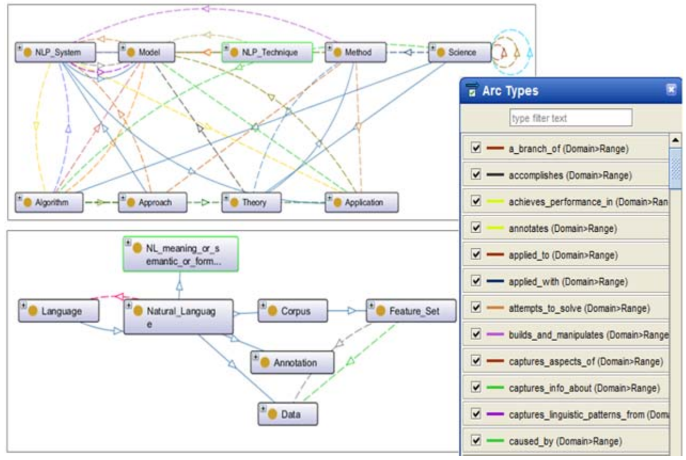

Fig. 2. Part of the NLP ontology graph [63], [64].

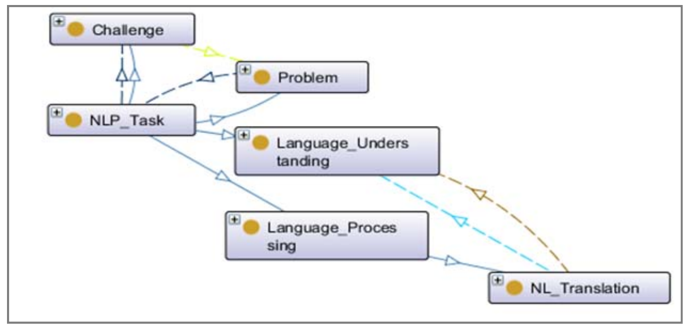

Fig. 3. Part of the NLP ontology graph [64].

Fig. 4 and 5 represent NLP ontology model in two different forms.

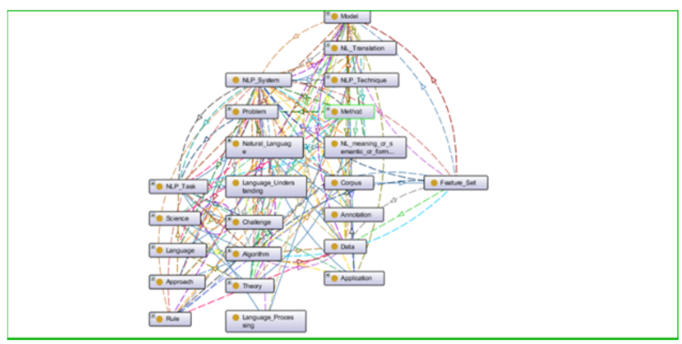

Fig. 4. Grid view of the Entire NLP ontology graph.

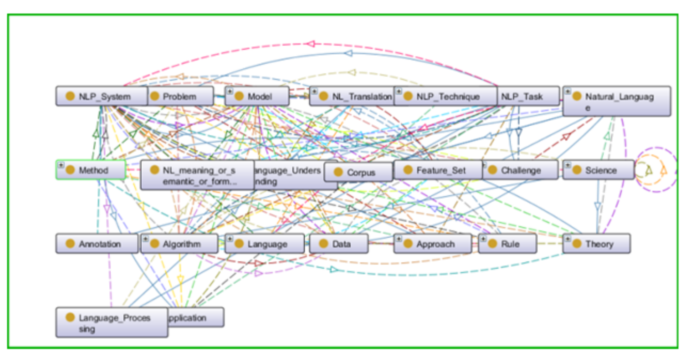

Fig. 5. Spring view of the Entire NLP ontology graph.

Fig. 6, 7 and 8 are Protégé snapshots to show the built ontology's object and data type properties and instances of the classes.

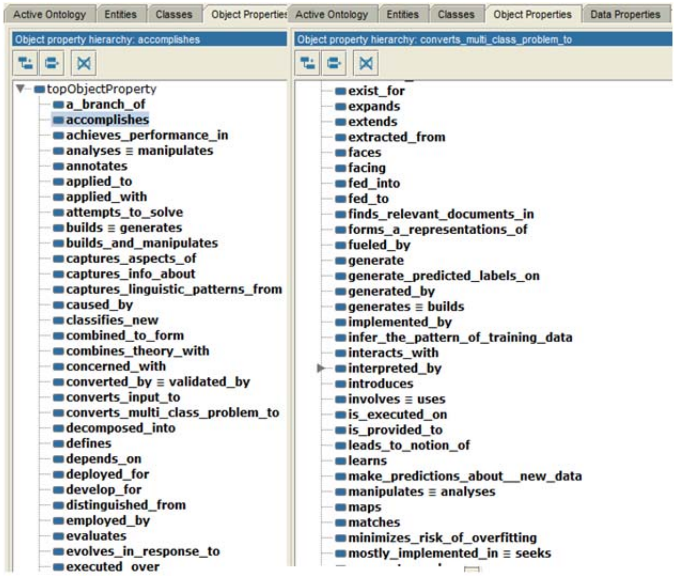

Fig. 6. Object properties [63], [64].

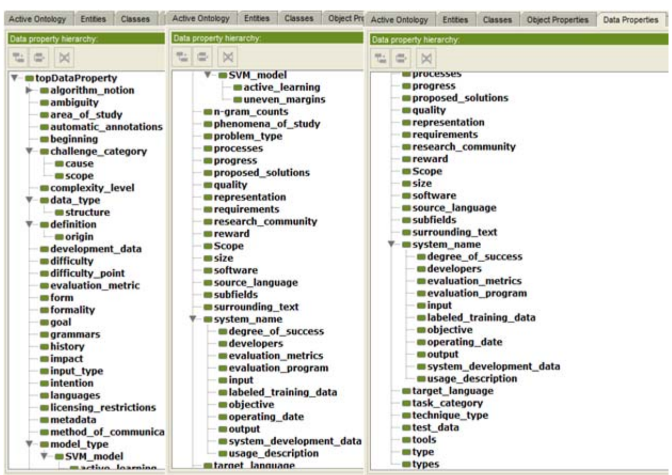

Fig. 7. NLP ontology data properties.

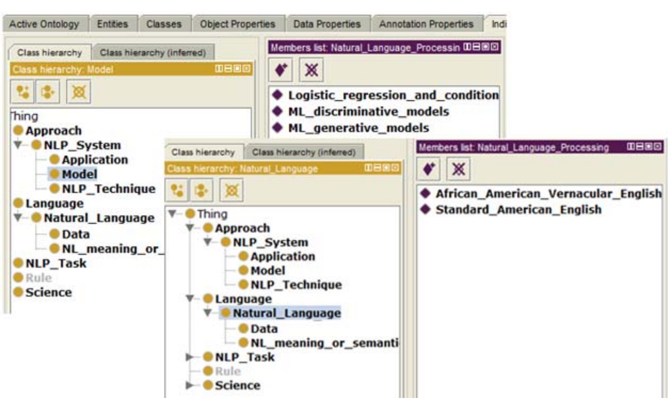

Fig. 8. NLP ontology instances for Model and NL concepts.

IV. Evaluating the NLP Ontology Model

Ontologies are a fundamental data structure for conceptualizing a domain, but since these tools are relatively easy to build, evaluation criteria are called upon for ontology users, designers and engineers for choosing and evaluating them.

In this paper, we chose peer review as a formal method to find errors of omissions, contradictions, flaws in logic, inconsistencies, and in-efficiency and to improve the quality of the ontology under consideration [66]. According to Karl E. Wiegers [67] his approach manages to provide value added input from reviewers with different technical backgrounds, experience, and expertise and professional growth to participants by giving them an opportunity to look at different development or maintenance methodologies and approaches.

Features of evaluation [65]-[70] are:

a) The data level: Including lexical terms and vocabulary, the evaluation scans the various knowledge documents composing the data and focusing on the chosen concepts and instances, assessing the vocabulary of each concept. In this paper, we provided the NLP experts with a body of natural language text and compare their output (extracted concepts, definition and is-a relation).

b) Taxonomy level: Here the evaluator focuses on each is-a relation, the reviewers can also compare the ontology to a human provided standard, to evaluate an approach for automatically extracting a set of lexons, i.e. triples of the form ⟨term1, role, term2⟩.

c) Semantic relations level: Precision and recall measures are used to evaluate relations other than is-a between concepts. We will mainly follow a human manual standard in comparison to the relational level of our ontology standard, supported by a list of statistical relevant terms. Then we will assess the philosophical notions [67].

The reviewers were provided with the ontology tables listing:

- The Class Hierarchy.
- The Object and Data properties.
- The Class Instances.
- Evaluation Forms 1 and 2 as shown in Table 5 and List 1.

TABLE V. Sample of Evaluation Form 1

Class notation	**Super classes**	**Subclasses**	**Reviewer opinion**	
Science	None	Algorithm Theory	Agree	Disagree
Natural language	Language	Data NL meaning NL Semantics Formal Grammar	Agree	
Algorithm	Science	None	Agree	Disagree
NLP Task	None	Corpus building Annotation Challenge Language Processing Language under-standing Problem	Agree	Disagree
Language Processing	NLP Task	NL Translation	Agree	Disagree

List 1: The evaluation form included these evaluation questions:

1) Does the modelled concepts and properties in the proposed ontology correctly represents entities in the world being modelled?

2) Does the modelled ontology correctly represent hierarchical and taxonomical concepts?

3) Do the concept attributes represented in the ontology clearly compel with in reality what is part of the object and what is not?

4) Does all instances of the subclass are necessarily instances of the super-class?

5) Are all the relations between class (A) and class (B) a part-of or subclass relation in the real world as demonstrated in this ontology?

6) Does the modelled concepts and relations in ontology provides a common understanding of a domain for parties to agree or commit to?

7) Does the ontology have a potential of fulfilling its main purpose (provides a common understanding of a domain for parties to agree or commit)?

8) Is the ontology structured in a way that has an ability for further development?

V. Results

We have applied peer review as an evaluation methodology to a subset of an ontology in the NLP domain. This methodology examines the usage of the sub-super relation between classes in an ontology using formal notions (rigidity, unity, identity and correctness) to capture various characteristics of classes, and constraints upon those Meta properties. Based on the results we were able to correct all inconsistencies. The presented ontology model is the result of modification and refinement according to the reviewers and NLP specialists' answers and remarks. However, it was noted that the resulting ontology could be further improved.

VI. Conclusion

In this paper, we have proposed, analyzed and evaluated an ontology model that represents part of the NLP domain. We have provided a graphical notation of the modeled NLP ontology. Once the ontology model has achieved describing NLP concepts declaratively, stating explicitly what the class hierarchy is and to which class's individuals belong; there are many future improvements that can be added to the model like introducing new concepts, and changing its conceptualization. We intend to deploy this model in a Universities LMS system to measure the improvements in the management, distribution and retrieval of the learning material and answering advanced search queries and creating complete cycles between the classes hierarchy based on the history of research in NLP, and for a supportive system capable of making decisions based on the structural properties of each class.

Acknowledgement

This research was partially supported by IWAN Research Group. We thank our colleagues from Information Technology Department who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

References

[1] Alexander Gelbukh, Natural language processing: Perspective of CIC-IPN. Keynote talk. 2nd International Conference on Advances in Computing, Communications and Informatics, ICACCI-2013, technically co-sponsored by IEEE Communications Society, Mysore, India, August 22–25, 2013.available at: http://www.gelbukh.com/

[2] Collins, Michael. "Machine learning methods in natural language processing." Tutorial presented at COLT (2003).

[3] Evans, Noah, Masayuki Asahara, and Yuji Matsumoto. "Cocytus: parallel NLP over disparate data." (2008).

[4] Steven Bird, Ewan klein ,Edward Loper, published by Oreilly Media, "Natural Language Processing with Python", 2009,USA.

[5] Brants, Thorsten. "Natural Language Processing in Information Retrieval." CLIN. 2003.

[6] Cunningham, Hamish, Yorick Wilks, and Robert J. Gaizauskas. "New methods, current trends and software infrastructure for NLP." arXiv preprint cmp-lg/9607025 (1996).

[7] Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. "Natural language processing: an introduction." Journal of the American Medical Informatics Association 18.5 (2011): 544-551.

[8] Bontcheva, Kalina, et al. "Evolving GATE to meet new challenges in language engineering." Natural Language Engineering 10.3-4 (2004): 349-373.

[9] Adler, Harry. Handbook of NLP: A manual for professional communicators. Gower Publishing, Ltd., 2002.

[10] Natalya F. Noy and Deborah L. McGuinness," Ontology Development 101: A Guide to Creating Your First Ontology", Stanford University, Stanford,2013.

[11] Somers, Harold L., Jun-ichi Tsujii, and Danny Jones. "Machine Translation without a source text." Proceedings of the 13th conference on Computational linguistics-Volume 3. Association for Computational Linguistics, 1990.

[12] Trim, Craig. "NLP-Driven Ontology Modeling (Language Processing)". Ibm.com. N.p., 2012. Web. 12 Mar. 2016.

[13] Clark, Alexander, Chris Fox, and Shalom Lappin, eds. The handbook of computational linguistics and natural language processing. John Wiley & Sons, 2013.

[14] Monachesi, P., et al. "What ontologies can do for eLearning." Proceedings of IMCL2008,(2008).

[15] Niles, I & Pease A. (2001) "Towards A Standard Upper Ontology." In Proceedings of FOIS 2001, October 17-19, Ogunquit, Maine, USA

[16] Pease, A.; Niles, I. & Li, J. (2002), The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications, in 'In Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web' .

[17] Mascardi, V.; Cordì, V. & Rosso, P. (2008), A Comparison of Upper Ontologies (Technical Report DISI-TR-06-21). Dipartimento di Informatica e Scienze dell'Informazione (DISI), Universitŕ degli Studi di Genova, Via Dodecaneso

[18] Bittner, Thomas & Smith, Barry (2003). Granular Spatio-Temporal Ontologies. AAAI Symposium:12-17.

[19] Smith, Barry ; Grenon, Pierre & Goldberg, Louis (2004). Biodynamic Ontology: Applying BFO in the Biomedical Domain. Studies in Health and Technology Informatics 102:20–38.

[20] Roberto Navigli and Simone Paolo Ponzetto. (2010). BabelNet: building a very large multilingual semantic network. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 216-225.

[21] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. WWW '07: Proceedings of the 16th International World Wide Web Conference, Banff, Canada, page 697-706. (2007)

[22] Ramachandran D, Reagan P, Goolsbey K. (2005). First-Orderized ResearchCyc: Expressivity and Efficiency in a Common-Sense Ontology. Papers from the AAAI Workshop on Contexts and Ontologies: Theory, Practice and Applications.

[23] Cox, David N. ; Lanoue, Thomas J. , "AUTOMATED METHOD AND TOOL FOR DOCUMENTING A TRANSFORMER DESIGN" , Patent No. 7721241.

[24] Pooshfam, H.; Abdullah, R. "A Proposed Method for Brain Medical Image Registration by Hierarchical Clustering Algorithm", Modelling & Simulation, 2009. AMS '09. Third Asia International Conference on, On page(s): 315 – 319

[25] Jiehan Zhou; Rautiainen, M.; Ylianttila, M.; Foulonneau, M.; Blandin, P. "Metamodeling for Community Coordinated Multimedia and Experience on Metamodel-Driven Content Annotation Service Prototype", Congress on Services Part II, 2008. SERVICES-2. IEEE, on page(s): 88 – 95.

[26] Jun Liu; Ying-xin Xie; Yin Liu; Feng-yu Wang; Mao-mao Wu "Regional energy management system considering demand response", Control and Decision Conference (2014 CCDC), The 26th Chinese, on page(s): 5119 – 5123.

[27] Y. Raimond, S. Abdallah, M. Sandler, F. Giasson, The Music Ontology Proceedings of the International Conference on Music Information Retrieval (ISMIR), 2007

[28] Wardhana, A.T.A.; Nugroho, H.T. "Combining FOAF and music ontology for music concerts recommendation on facebook application".

[29] Arvidsson, Fredrik, and Annika Flycht-Eriksson. "Ontologies i." PDF). http://www. ida. liu. se/janma/SemWeb/Slides/ontologies1. pdf. Retrieved 26 (2008).

[30] Alatrish, Emhimed Salem, Dusan Tosic, and Nikola Milenkovic. "Building ontologies for different natural languages." Comput. Sci. Inf. Syst. 11.2 (2014): 623-644.

[31] Motik, Boris, et al. "Modeling Ontologies using OWL, Description Graphs, and Rules." OWLED. 2008.

[32] Funk, Adam, et al. "Ontological integration of information extracted from multiple sources." extraction 4 (2007): 6.

[33] Lehmann, Jens, and Johanna Völker, eds. Perspectives on Ontology Learning. Vol. 18. IOS Press, 2014.

[34] S.Staab and R.Studer (ed.), Handbook on ontologies , International Handbooks on Information Systems , DOI 10.1007/978-3-540-92673-3,Springer-Verlag Berlin Heidelberg 2009.

[35] Alex Penev ,Raymond Wong, "Shallow NLP techniques for Internet Search", Published in: Proceeding ACSC '06 Proceedings of the 29th Australasian Computer Science Conference - Volume 48 ,Pages 167-176 Australian Computer Society, Inc. Darlinghurst, Australia, 2006.

[36] Damljanović, D., Agatonović, M., Cunningham, H. and Bontcheva, K. (2013). "Improving habitability of natural language interfaces for querying ontologies with feedback and clarification dialogues". Web Semantics: Science, Services and Agents on the World Wide Web, 19, pp.1-21.

[37] Afreen, Hina, and Imran Sarwar Bajwa. "A Framework for Automated Object Oriented Analysis of Natural Language Software Specifications." International Journal of Software Engineering and Its Applications 6.2 (2011): 15-22.

[38] Karttunen, Lauri and Kemal Oflazer. "Introduction To The Special Issue On Finite-State Methods In NLP". Computational Linguistics 26.1 (2000): 1-2. Web.

[39] "NLP sub-problems: application to clinical text".

[40] University of Sheffield Natural Language Group Information Extraction: the GATE pipeline. 2011. http://www.gate.ac.uk/ie (accessed 1 Jun 2011).

[41] Wilson, Bill. "The Natural Language Processing Dictionary". Cse.unsw.edu.au. N.p., 2012. Web. 9 Oct. 2015.

[42] Ide, Nancy and Keith Suderman. "Open American National Corpus | Open Data For Language Research And Education". Anc.org. N.p., 2016. Web. 9 Sept. 2015.

[43] "Santa Barbara Corpus of Spoken American English (SBCSAE)| Department Of Linguistics - UC Santa Barbara". Linguistics.ucsb.edu. N.p., 2016. Web. 9 June 2016, available at: http://www.linguistics.ucsb.edu/research/santa-barbara-corpus.

[44] MIT press journals, Computational Linguistics is published by the MIT Press on behalf of The Association for Computational Linguistics.

[45] Michael Gamon,et al " Using Contextual Speller Techniques and Language Modeling for ESL Error Correction" , Proceedings of IJCNLP, Hyderabad, India.2008,available at: http://research.microsoft.com/en-us/groups/nlp/.

[46] Hristea, Florentina T. "Statistical Natural Language Processing." International Encyclopedia of Statistical Science. Springer Berlin Heidelberg, 2011. 1452-1453.

[47] Li, Yaoyong, Kalina Bontcheva, and Hamish Cunningham. "Adapting SVM for data sparseness and imbalance: a case study in information extraction." Natural Language Engineering 15.02 (2009): 241-271.

[48] McCoy, Kathleen, et al. "Applying natural language processing techniques to augmentative communication systems." Proceedings of the 13th conference on Computational linguistics-Volume 3. Association for Computational Linguistics, 1990.

[49] Crocker, Matthew W., Martin J. Pickering, and Charles Clifton, eds. Architectures and mechanisms for language processing. Cambridge University Press, 2000.

[50] Shapiro, Stuart Charles, and Lucja M. Iwánska, eds. Natural language processing and knowledge representation: language for knowledge and knowledge for language. Aaai Press, 2000.

[51] Agrawal, Avinash J., and O. G. Kakde. "Semantic analysis of natural language queries using domain ontology for information access from database." International Journal of Intelligent Systems and Applications 5.12 (2013): 81.

[52] Rangel, Rodolfo A. Pazos, et al. "A domain independent natural language interface to databases capable of processing complex queries." Mexican International Conference on Artificial Intelligence. Springer Berlin Heidelberg, 2005.

[53] Karttunen, Lauri, and Kemal Oflazer. "Introduction to the special issue on finite-state methods in NLP." Computational Linguistics 26.1 (2000): 1-2.

[54] Coleman, John. Introducing speech and language processing. Cambridge university press, 2005.

[55] Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python. " O'Reilly Media, Inc.", 2009.

[56] Shinde, Subhash K., Varunakshi Bhojane, and Pranita Mahajan. "Nlp based object oriented analysis and design from requirement specification." International Journal of Computer Applications 47.21 (2012).

[57] Maynard, Diana, Yaoyong Li, and Wim Peters. "Nlp techniques for term extraction and ontology population." Proceeding of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge. 2008.

[58] Covington, Michael A. Natural language processing for Prolog programmers. Englewood Cliffs (NJ): Prentice hall, 2013.

[59] Jiehan Zhou; Rautiainen, M.; Ylianttila, M.; Foulonneau, M.; Blandin, P. "Metamodeling for Community Coordinated Multimedia and Experience on Metamodel-Driven Content Annotation Service Prototype", Congress on Services Part II, 2008. SERVICES-2. IEEE, on page(s): 88 – 95.

[60] Penev, Alex, and Raymond Wong. "Shallow NLP techniques for internet search." Proceedings of the 29th Australasian Computer Science Conference-Volume 48. Australian Computer Society, Inc., 2006.

[61] Kilgarriff, Adam. "Thesauruses for natural language processing." Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003 International Conference on. IEEE, 2003.

[62] Bontcheva, Kalina, et al. "TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text." RANLP. 2013.

[63] Indurkhya, Nitin and Fred Damerau, "Handbook of Natural Language Processing" (second edition) (eds., 2010), Chapman & Hall/CRC.

[64] Yaoyong Li, Kalina Bontcheva, Hamish Cunningham, "Adapting SVM for Data Sparseness and Imbalance: A Case Study on Information Extraction", Natural Language Engineering 1 (1): 1–32. Printed in the United Kingdom , 1998 Cambridge University Press.

[65] Yu, Jonathan, J. Thom, and Audrey Tam. "Evaluating an ontology with OntoClean." Proceedings of 10th Australasian Document Computing Symposium (ADCS). 2005.

[66] YuThom and Tam, Proceedings of the 10th Australasian Document Computing Symposium, Sydney, Australia, December 12, 2005.

[67] Karl E. Wiegers, "Peer Reviews in Software: A Practical Guide" .2001, Addison Wesley information technology series

[68] Brank, Grobelnik and Mladenić ,"A Survey Of Ontology Evaluation Techniques", Conference on Data Mining and Data Warehouses (SiKDD) 2005.Ljubljana, Slovenia.

[69] Guarino , Welty , "Common." Evaluating Ontological Decisions With Onto clean" ",(2002), Vol. 45, No. 2 COMMUNICATIONS OF THE ACM.

[70] Pak and L. Zhou, "A Framework for Ontology Evaluation",2009. The Eight Workshop on E-Business, Dec. 15, Phoenix, AZ.