International Journal of Research Studies in Science, Engineering and Technology
Volume 7, Issue 12, 2020, PP 18-27
ISSN 2349-476X

Automatic Domain Ontology Learning From Unstructured Afaan Oromo Text

SADIK ABAS[1*], MELKAMU BEYENE[2]

[1]*Information Technology Department, College of Informatics, Bule Hora University, Ethiopia*

[2]*Information science department, Addis Ababa University, Addis Ababa, Ethiopia*

***Corresponding Author:** SADIK ABAS, Department of Information Technology, College of Informatics, Bule Hora University, Ethiopia.*

ABSTRACT

Automatic ontology learning from unstructured text is an essential and challenging task in Natural Language Processing (NLP), particularly for resource scarce language like Afaan Oromo (AO). Ontologies simplify automatic processing and understanding of textual resources, and also it could be an effective way to improve the searching process and exploiting information on the web. The use of ontology is that it provides a standard for the vocabulary used in a specific domain and relations. For Automatic ontology learning from unstructured AO text on genealogy of Oromo domain four main steps used. First steps is the pre-processing one which includes tasks like Tokenization, Normalization, Stop-word removal and Pos tag. Secondly after the preprocessing is performed candidate term extraction is performed in order to extract the correct candidate term for ontology learning as ontological concepts by using JAPE grammar rule. Thirdly between the extracted candidate terms, relation is extracted using Rule-based and machine learning for relation extraction. Lastly construct a set of transformation rule in order to identify appropriate elements for ontology construction form candidate term and relations. To practically test the performance of the methods used the researcher's uses human expert review and protege application. The human review or (domain expert review) for extracting candidate term outperforms Precision of 93.76%, Recall of 96.9% - and F-measure of 95.3%. The reasoner shows that class/subclass is well consistent. The result shows that Afaan Oromoo ontology learner can be used as a starting point for future researches related to Ontologies and Ontology learning from Afaan Oromo text.

Keywords: *Ontology, Ontology learning, Candidate, taxonomy relation, relation extraction*

INTRODUCTION

After the vision of semantic web was broadcasted, ontology becomes asynonym for the solution to many problems concerning the fact that computers do not understand human language. If there were an ontology and every document were marked up with it and we have agents/ computers that would understand the ontology, then computers / agents would finally be able to process our queries in correct way.

Ontologies simplify automatic processing and understanding of textual resources. It has been applied to a number of different domains such as bio-medicine [6], education [7] and software engineering [8]. For the above mentioned and other applications different methodologies are proposed by different researchers for the design and building of ontologies. Manual ontology construction process is time consuming and not effective one. Hence, the design and development of methods and software tools to support automatic ontology construction is an important research topic, which is known as ontology learning[1]. The process of defining and instantiating aknowledge base is referred to as knowledge markup or ontology population, whereas (semi-)automatic support in ontology development is usually referred to as ontology learning.

Ontology learning is concerned with knowledge acquisition and in the context of this research more specifically with knowledge acquisition from text. Obviously, much of the work in this area builds on the large body of work in this direction within NLP, AI, and machine learning.

Ontology learning is a challenging area in its own right, with a lot of innovating research to which also this thesis hopes to contribute specifically for Afaan Oromo. Some of the ontology learning methodology and tools are designed for different languages such as Spanish [2], Amharic [4] and also for English [5].

But adopting the methodology and tools used as they are to Afaan Oromo is too difficult because Afaan Oromo have its own unique grammar [16], and syntax. Therefore designing methods and tools that work with Afaan Oromo is must, which is the objective this research.

Ontologies serve as a means for knowledge representation and are capable of expressing a set of entities, their relationships, constraints, axioms and the vocabulary of a given domain. However, the manual construction of ontologies is an expensive and time consuming task because the professionals required for this task (i.e. a domain specialist and a knowledge engineer) usually are highly specialized.

This difficulty in capturing the knowledge required by knowledge based systems is becoming very common and is labeled as "knowledge acquisition bottleneck".

Hence, fast and cheap ontology development is crucial for the success of knowledge based applications and the Semantic Web. This has been designed and implemented for English [5], Spanish [2] and Amharic [4].

Afaan Oromo has its own unique grammar [16], syntax, character (qubee) representation and statement formation. The other problem is that for Afaan Oromo there is no sufficient resources means that the structured sources of data or semi-structured sources of data such as dbpedia, Wikipedia and Afaan Oromo WorldNet is not sufficiently available as required. These reasons make it clear that ontology learning from Afaan Oromo text involves reasonably different methods and tools compared to English or Spanish or Amharic text. But no such tools exist for Afaan Oromo text.

Related Work

We reviews a number of research works about automatic ontology learning from English, Amharic, Spanish and Afaan Oromo text and Named Entity Recognition (NER). However, little or no attention has been paid to Afaan Oromo text.

[37], proposed an approach for NER for Afaan Oromo language using hybrid approach, which contains machine leaning and rule-based approach and The experiment result shows that the performance of their system as 84.12% Precision, 81.21% Recall and 82.52% F-Score.

[36], proposed the system that recognize NE by using Conditional Random Field (CRF) algorithm which was machine learning approach for Afaan Oromo language and experiment result shows that the performance of their system as 77.41% Recall, 75.80% Precision and 76.60% F1-measure.

[47] Proposes an approach based on ontology learning and natural language processing for automatic construction of expressive ontologies, specifically in OWL DL from English text. The approach constructs the expressive ontology correctly in more than half of the cases; and does not construct the expressive ontology in more than half of the cases.

[48] Proposes a system for information extraction from plain text in form of RDF triples. The approaches is capable of identifying grammatical structure (syntactic) of an input sentence and analyze its semantic to generate meaningful RDF triples of information, by Stanford and Senna tools for translating plain text documents into a machine-readable format.

[49] proposed system that can learn the ontology automatically from unstructured Amharic text by extracting single-word and multi-word concepts and deduce the relationships exist between the domain concepts in their corpus. [22] Develop a system that automatically builds ontology from Vietnamese texts using cascades of annotation based grammars. Gate is used to implement the system. The system includes two components: syntactic analysis and ontology construction, the syntactic analysis is to detect noun phrases and relation phrases from input documents, and then identify candidate phrases representing classes, individuals, relationships and properties, subsequently, the ontology extraction component uses Text2Onto [44] to generate the output ontology. [49] Presents a technique for automatically constructing ontology from a given lecture notes. This system extracts the concepts using Term Frequency Inverse Document Frequency (TF-IDF) weighting scheme and then determines the associations among concepts using apriori algorithm. Evaluating the system is performed by comparing the results with the dependencies determined by an expert in the subject area.

Methodology

The aim of this study was to design Automatic Domain Ontology Learning from unstructured Afan Oromo text. To this end, we come up with architecture of the system, based on which different techniques and methods are used. Our proposed system for ontology learning from unstructured text is divided into four modules or stages, which is shown in the following figure.1 ststage, Pre-processing stage is the stage where different NLP processing techniques is performed. For our case we use shallow data processing. Some of the NLP processing techniques in our work include, Tokenization, Normalization, Removing stop word, POS tagging to make the data suitable for the next stage as input. In order to perform the pre-processing task we are using GATE Framework. 2ndstage, Candidate term extraction (term extraction) it can be performed by simply extracting the candidate terms by using the JAPE and Gazetteer developed. 3rdstage, Relation Extraction between candidate terms is extracted for the purpose of generating Triple statements in subject-predicate-object format. It is performed by using machine learning and patterns. 4th stage, annotates terms and relations, then transform the annotated terms in to Ontological Concepts and Relationships and lastly construct the ontology for the given domain.

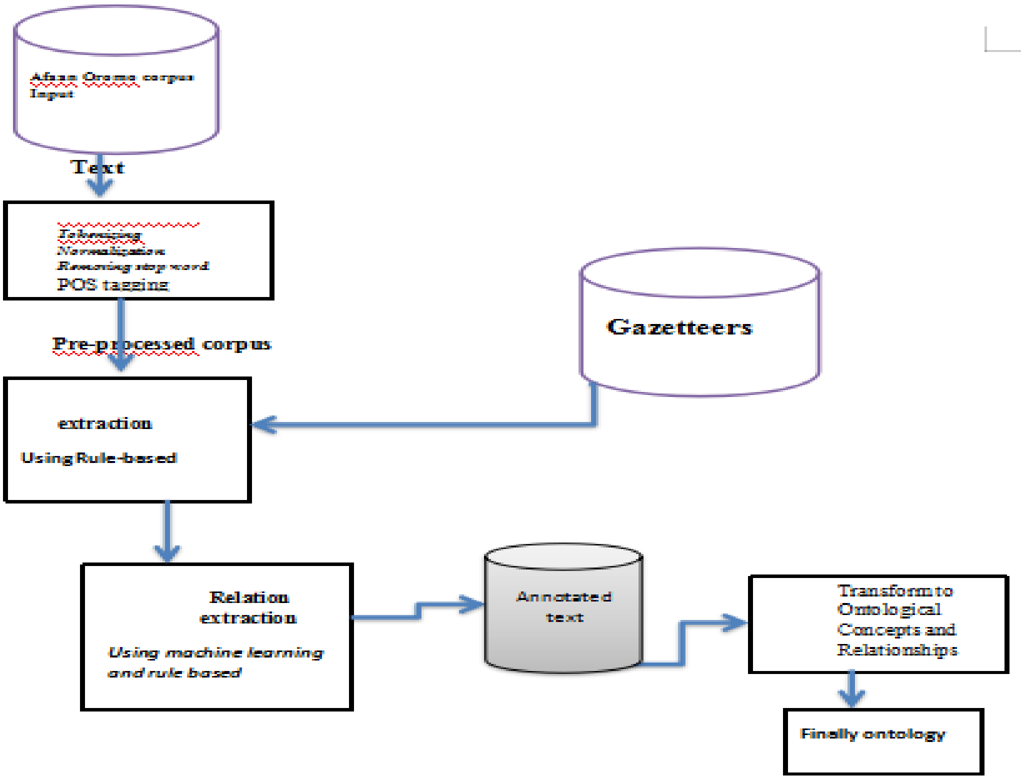


Figure4: Proposed Architecture for Ontology learning from text

To conduct the procedure with a better performance, the process working of this study has been consisted the following sub sections.

Pre-Processing

Preprocessing is the task of ensuring the corpus a suitable state for the information extraction methods to be effective, and annotating the corpus with linguistic features needed by certain information methods. The sub process of preprocessing, each one can perform specific function to make data suitable for the next stage and to construct ontology.

Preparing Corpus

Corpus preparation is one of the most required steps in our approach. Shortly corpus or corpora is a collection of documents in specific domain. We have collected documents that are nearly related the domain genealogy of Oromo (Hidda Latiinsa Oromoo). We collect the document from book like, Seenaa Oromoo Hanga Jaarraa 16ffaa, website and also by making interview with some oromo scholars. Interviewing is required for checking our datasets collected from different sources.

Candidate Term Extraction (Term Extraction)

Term extraction is an automatic method of analyzing text in order to identify candidates which fulfill the criteria for terms to be a candidate term. One of the best ways is through NER. A named entity (NE) is a word or phrase that contains the name of: person, location, an organization, date, amounts of money, number, percent, nationality, product wars, and quantity.

For example, in the sentence “Ilmaan Baarentuu Oromiyaa fi naannoo biroo keessatti baayyinaan argamu” contains two named entities “Baareentuu” which is person and “Oromiyaa” which is location. To extract the candidate term we proposed a simple integration between lexical resources with some rule based system.

For our proposed term extraction, we use Gazetteers as linguistic knowledge where they are able to detect complex entities, and then we enhance detected entities with rules.

Relation Extraction

The relation to be extracted is the taxonomic relations rather than non-taxonomic relations. The relations that we are going to extract are relations between entities, such as parents (families). Extraction rules capture taxonomic relations by identifying specific lexical elements in a text, such as keywords, although extraction rules can be defined following regular expression, like GATE’s Java Annotation Patterns Engine (JAPE) [52].A hyponymy relation between two entities NP0 and NP1 refer to membership relations in the form NP0 is a (child of) NP1, where (child of) is one of the taxonomic relation categories such as “Ilma”, dhale, dhalche, dhalfate, guddifate ,kan, damee or “ka”.

Transforming To Ontological Concepts and Relationships

Now that concepts and taxonomic relations have been identified, it is possible to produce an explicit representation in ontological form. The representation of the knowledge using instances extracted and annotated from text is important task in ontology construction. We formulate conceptual classes, instances and their relationships to represent this information using existing Resource Description Framework (RDF). The motivation behind RDF representation is it enables the possibility of complex querying on the extracted information. We represent knowledge as subject-predicate-object triples. Subjects are resources and extracted as concepts, predicates are taxonomic relations, and objects are resources and extracted as concepts.

Implementation and Experimental Results

In this paper, an experiment is made to construct Domain Ontology from unstructured Afaan Oromo text. The main goal of this chapter is exploring the possibilities of learning domain ontology system from unstructured Afaan Oromo text by using hybrid techniques. In this paper, the researcher describes the source of data as well as the techniques that have been used in preprocessing the data. The system training parameters and processes and learning algorithm are also presented. Test results are then presented and discussed.

Remove Stop-Word

While developing the rule it is easy to handle the stop-words simply by adding to the rule the following command if it is not required in our process.

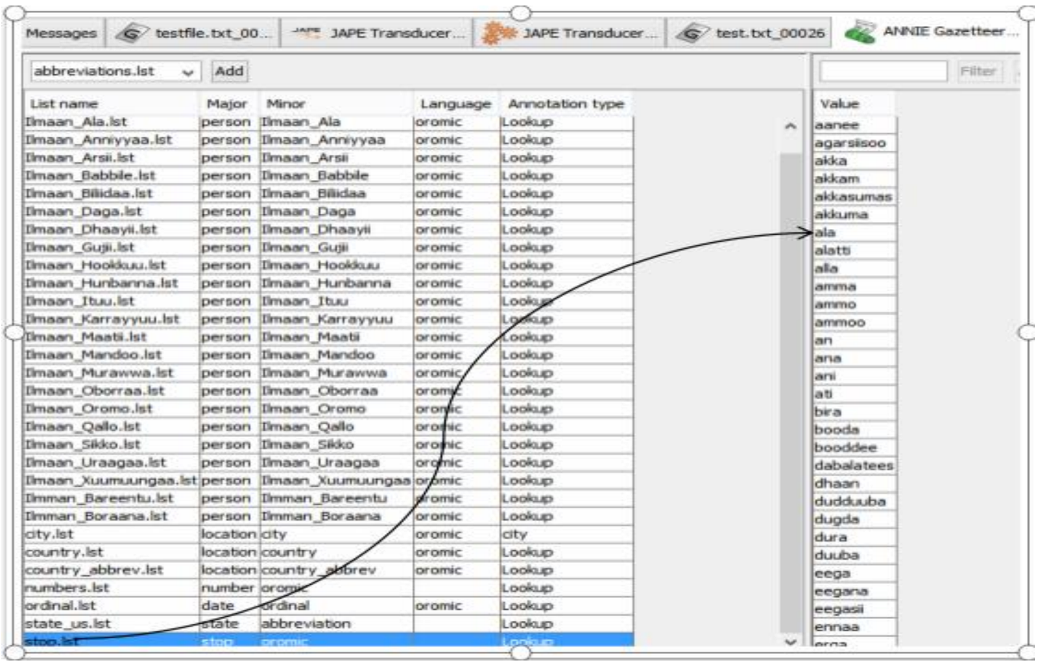
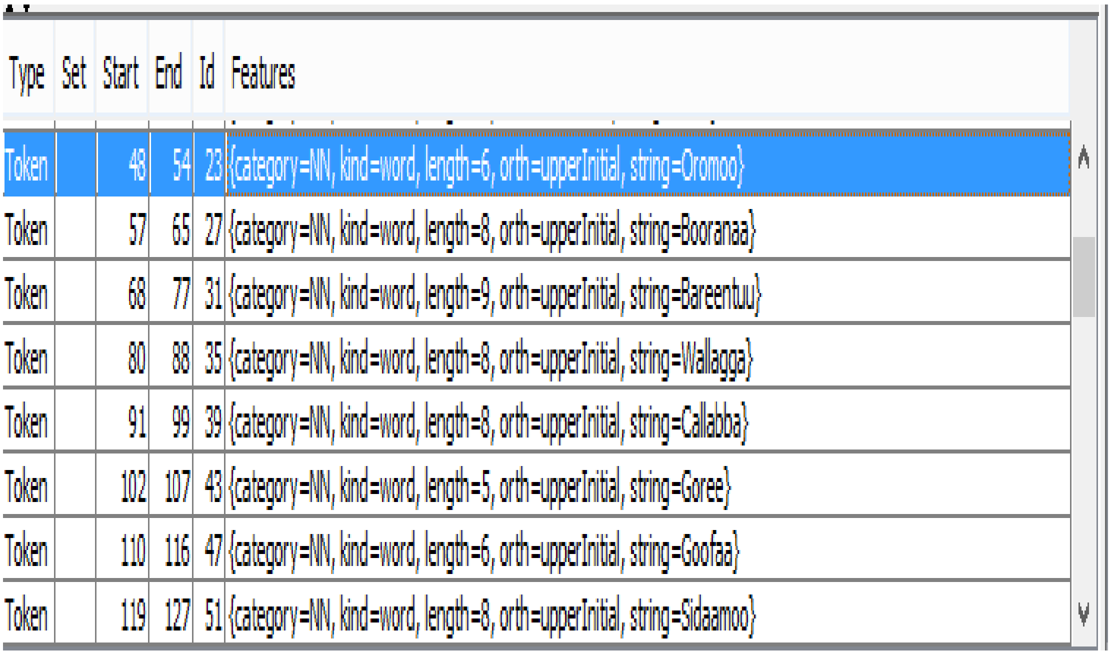


Figure8. *stop_word sample*

POS Tagging



Type	Set	Start	End	Id	Features
Token		48	54	23	{category=NN, kind=word, length=6, orth=upperInitial, string=Oromoo}
Token		57	65	27	{category=NN, kind=word, length=8, orth=upperInitial, string=Booranaa}
Token		68	77	31	{category=NN, kind=word, length=9, orth=upperInitial, string=Bareentuu}
Token		80	88	35	{category=NN, kind=word, length=8, orth=upperInitial, string=Wallagga}
Token		91	99	39	{category=NN, kind=word, length=8, orth=upperInitial, string=Callabba}
Token		102	107	43	{category=NN, kind=word, length=5, orth=upperInitial, string=Goree}
Token		110	116	47	{category=NN, kind=word, length=6, orth=upperInitial, string=Goofaa}
Token		119	127	51	{category=NN, kind=word, length=8, orth=upperInitial, string=Sidaamoo}

Candidate Term Extraction (Term Extraction)

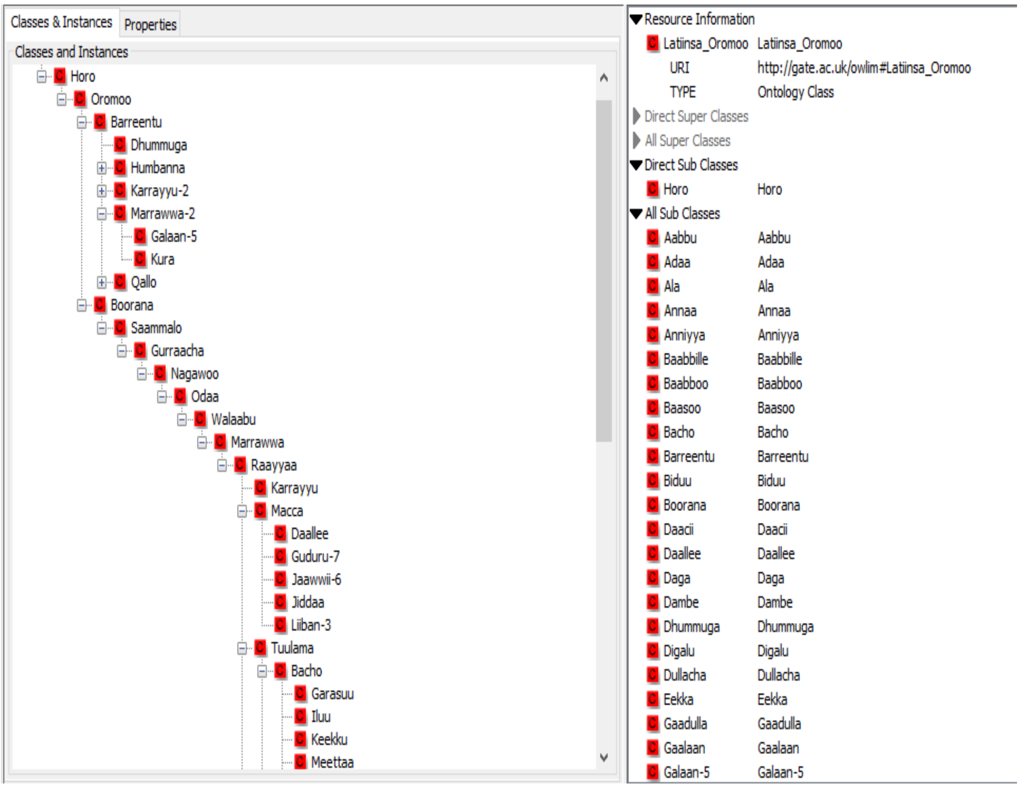
Type	Set	Start	End	Id	Features
Candidate		116	124	243	{String= Wallagga, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#ILMMAN_BORAANA, classes=oromo, maj
Candidate		127	135	244	{String= Callabba, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Pers
Candidate		138	143	245	{String= Goree, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Person
Candidate		146	152	246	{String= Goofaa, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Perso
Candidate		155	163	247	{String= Sidaamoo, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Per
Candidate		166	170	248	{String= Arii, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Person, m
Candidate		173	178	249	{String= Dacee, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Person
Candidate		181	187	250	{String= Garrii, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Person,
Candidate		190	195	251	{String= Guraa, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Person
Candidate		198	207	252	{String= Giriirraa, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Perso
Candidate		210	220	253	{String= Naayroobii, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Pe
Candidate		223	228	254	{String= Gujii, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#RANGE, classes=oromo, majorType=Person, r
Candidate		257	267	255	{String= Xummuungaa, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#ILMMAN_BAREENTU, classes=oromo,
Candidate		281	290	256	{String= Karrayyuu, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#ILMMAN_BAREENTU, classes=oromo, m
Candidate		311	316	257	{String= Arsii, class=http://gate.ac.uk/owlim#ILMA, classM=http://gate.ac.uk/owlim#ILMAAN_XUUMUUNGAA, classes=oromo, majo

Relation Extraction

```
phase: Taxonomic_Relation
Input: Token
options: control = appelt
Rule: Is_a
(
{Token.string == "ka" }|
{Token.string == "kan" }|
{Token.string == "ilmaa" }|
{Token.string=="ijoollee"}|
{Token.string=="damee"}|
{Token.string == "ilmaan"}
):imaani
-->
:imaani{
gate.AnnotationSet predi = (gate.AnnotationSet) bindings.get("imaani");
gate.FeatureMap features = Factory.newFeatureMap();
features.put("rule","Is_a");
outputAS.add(predi.firstNode(),predi.lastNode(),"Taxonomic_Relation",features);
}
//-------------------------------------------------------------------
Rule: Has_a
(
{Token.string=="dhale"}|
{Token.string == "dhalche" }|
{Token.string == "dhalfate" }|
{Token.string == "guddifate" }|
{Token.string=="ijoollee"}|
{Token.string=="Abbaan"}
):hasilman
-->
{
gate.AnnotationSet has = (gate.AnnotationSet) bindings.get("hasilman");
gate.FeatureMap features = Factory.newFeatureMap();
features.put("rule","Has_a");
outputAS.add(has.firstNode(),has.lastNode(),"Taxonomic_Relation",features);
}
```

Transforming To Ontological Concepts and Relationships

In order to transform the candidate term and relations to ontological concepts and relations we developed a JAPE rules.



Automatic ontology building evaluation is a hard task because it is sometimes difficult to find triples for a given document or set of documents and usually depends on human experts. To the knowledge of the researcher, there is no previous research made regarding Afan Oromo ontology learning with which a comparison can be made.

The performance of modules in ontology learning system depends on different factors, such as: The corpus used the performance of NLP tools used and the nature of the language itself:

However in order to see performance of proposed approach of automatic domain ontology learning from unstructured Afan Oromo text of genealogy of Oromo domain; we used two methods for performance evaluation: human expert review and protégé application.

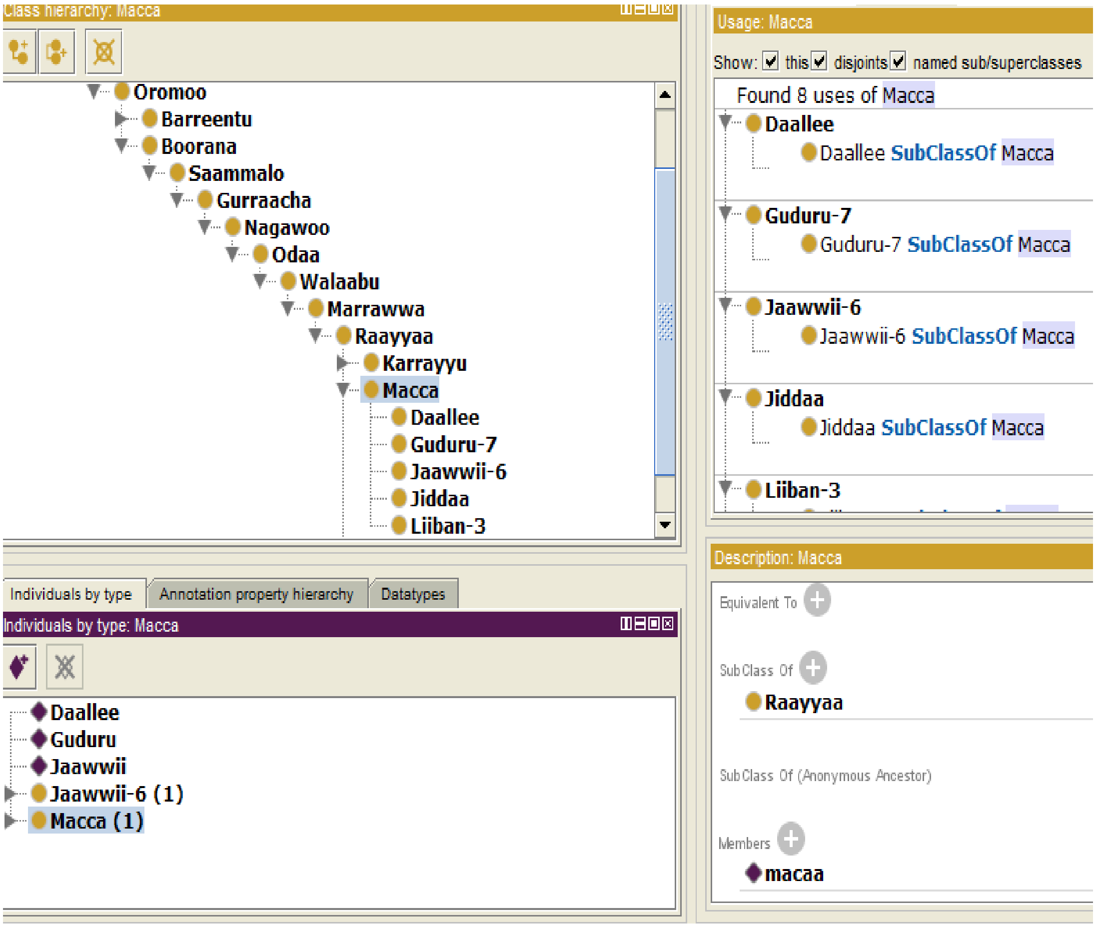


Figure18. *The ontology asserted hierarchies without any inconsistencies.*

Conclusions And Recommendation

This research focused on methods for automatic ontology learning from unstructured Afan Oromo texts and we proposed different methods to be applied at different layers in ontology development process. Starting from developing a JAPE Grammar rule we used in our process. Pre-processing tasks is performed first on the corpus then Extracting all possible concepts that are believed to describe a domain and then the concept hierarchy or taxonomy of concepts is constructed , finally the ontology is learned.

For evaluation purposes, the three common effective measures were used; Recall, Precision and F-measure. The results of annotation achieve satisfactory results for all terms and taxonomic relation extractions. Recall of 96.9%, precision of 93.76% and f-measure of 95.3%, for extracting candidate terms. And if we use the protégé for the evaluation of whether or not one class is a subclass of another class, it shows the result in the figure 18 above. Using the approach overcomes the problem of the manual construct ontology from Afan Oromo documents. This means saving time and overcome difficulties with the manual process.

During the course of this research, we able to find some portions that require further improvements that we recommend as future work. Some of them are as follows:-

In future work, we recommend linguists to enhance the gazetteers and explore the possibility of improving the system with adding more lists. There is also a space for improving the grammatical rules implemented within the rule-based component through analyzing the hybrid system's output in a way to automate the enhancement process.

Our work is limited to recognizing concepts and the relations among them (Taxonomic), but the ontology could be much improved if different attributes and instances of those concepts and relations are included. And also based on the type data used the non-Taxonomic should also included

The theoretically discussed advantages of ontology can also be shown by developing a system that uses an ontology input and prove the advantages.

Extending the rules that are used in extracting taxonomic relations to be used as a basis for more specific relations and properties at the level of OWL such as symmetric/ asymmetric, cardinality, disjointness to name a few.

References

[1] Biemann, Chris, Ontology learning from text: A survey of methods, LDV forum, Vol. 20.No. 2, 2005.

[2] Ochoa, José, Maria Luisa Hernandez-Alcaraz, Rafael Valencia and Rodrigo Martinez, A Semantic Role-Based Approach for Ontology Learning from Spanish Texts, International Symposium on Distributed Computing and Artificial Intelligence, Springer Berlin/ Heidelberg, 2011.

[3] Mamo Abebe , Esubalew Alemneh Ontology Development for Anemic Pregnant Women, 5th International Conference on the Advancement of Science and Technology (ICAST-2017),Bahir Dar University

[4] Berhanu mengiste Automatic Ontology Learning from Unstructured Amharic Text, Addis Ababa, 2013

[5] Maedche, Alexander, and Steffen Staab, Ontology learning for the semantic web, Intelligent Systems, IEEE 16.2 (2001): 72-79.

[6] García-Sánchez, Francisco, Jesualdo Tomas, Rafael Valecia, Juan Miguel and Rodrigo Martinez, Combining Semantic Web technologies with Multi-Agent Systems for integrated access to biological resources, Journal of biomedical informatics 41.5 (2008): 848.

[7] Happel, Hans-Jörg, and Stefan Seedorf, Applications of ontologies in software engineering, Proc. of Workshop on Sematic Web Enabled Software Engineering(SWESE) on the ISWC, 2006.

[8] Hashim, Fatimah, Gazi Mahabubul Alam, and Saedah Siraj, Information and communication technology for participatory based decision-making-E-management for administrative efficiency in Higher Education, Int. J. Phys. Sci 5.4 (2010): 383-392.

[9] Buitelaar, Paul, Philipp Cimiano, and Bernardo Magnini, Ontology learning from text: methods, evaluation and applications, Vol. 123, Ios PressInc, 2005.

[10] Khurshid Ahmad and Lee Gillam Automatic Ontology Extraction from Unstructured Texts,

[11] Buitelaar, Paul, Daniel Olejnik, and Michael Sintek, A protégé plug-in for ontology extraction from text based on linguistic analysis, The Semantic Web: Research and Applications (2004): 31 -44.

[12] Sabot, M. Wroe, C., Goble, C., and Mishne, G., (2005). 'Learning Domain Ontologies for Web Service Descriptions: an Experiment in Bioinformatics', in proceedings of the 14th International World Wide Web Conference (WWW2005), Chiba, Japan, 10-14.

[13] Davulcu, H., Vadrevu, S., Nagarajan, S., and Ramakrishnan, I.V, (2003). 'Onto Miner: Bootstrapping and Populating Ontologies from Domain Specific Web Sites'. In IEEE Intelligent Systems, 18(5), 2003.

[14] Hu, H., Du, X. , Liu D, and Ouyang, J.H., (2004). 'Ontology Learning using WorldNet Lexicon, Proc. of the International Conference on Computational Methods (ICCM 2004), Singapore.

[15] Li, M., Du, X., and Wang, S., (2005). 'Learning ontology from relational database', in Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou.

[16] DebelaTesfaye, "A rule-based Afaan Oromo Grammar Checker", International Journal of Advanced Computer Science and Applications, Vol. 2, No. 8, 2011

[17] Getachew Mamo, Automatic Part Of Speech Tagging for Oromo Language, Master"s Thesis, School of Graduate studies, Addis Ababa University, 2009.

[18] Tesfaye Guta, Oromo search engine, Master's Thesis, Addis Ababa University, Departement of Computer Science, 2010

[19] Mylanguage.org, Oromo Numbers, 2011. [Online]. Available: http:// www. mylanguages

.org/learn_ oromo.php. [Accessed: 12-Jan-2012].

[20] Wakshum Mekonnen, "Development of a Stemming Algorithm for Oromoo Text," Master's thesis, Addis Ababa University, School of Information Studies, 2000.

[21] Simon Ager, Oromo Language, Online edition, 2012. [Online].Available: www. sas.upenn.edu/African_Studies/Hornet/_Oromo_19777.html.[Accessed:25-Feb 2012].

[22] Nguyen, D. Q., Nguyen, D. Q., Ma, K. T., & Pham, S. B. (2011). Automatic Ontology Construction from Vietnamese Text. Paper presented at the Natural Language Processing andKnowledge Engineering (NLP-KE), 2011 7th International Conference on.

[23] Thakker, D., Osman, T., & Lakin, P. (2009). Gate Jape Grammar Tutorial Nottingham Trent University, UK, Phil Lakin, UK, Version, 1.

[24] Buitelaar, Paul, Daniel Olejnik, and Michael Sintek, A protégé plug-in for ontology extraction from text based on linguistic analysis, The Semantic Web: Research and Applications (2004): 31-44.

[25] Nakagawa, Hiroshi, Experimental evaluation of ranking and selection methods in term extraction, Bourigault D, L'Homme M.-C, Jacquemin C.(éd.), Recent advances in computational terminology, John Benjamins Publishing Company, Amsterdam (2001): 303-326.

[26] Miller, George A, WordNet: a lexical database for English, Communications of the ACM 38.11 (1995): 39-41.

[27] Powers, David MW, Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation, School of Informatics and Engineering, Flinders University, Adelaide, Australia, Tech. Rep, SIE-07-001 (2007).

[28] Nenadić, Goran, Irena Spasić, and Sophia Ananiadou, Automatic discovery of term similarities using pattern mining, COLING-02 on COMPUTERM 2002: second international workshop on computational terminology-Volume 14, Association for Computational Linguistics, 2002.

[29] Powers, David MW, Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation, School of Informatics and Engineering, Flinders University, Adelaide, Australia, Tech. Rep, SIE-07-001 (2007).

[30] Daniel Jurafsky & James H.Martin, Speech and Language Processing: An introduction to speech recognition, natural language processing, and computational linguistics, October 12, 2007.

[31] Pandit, S. (2010). Ontology-Guided Extraction of Structured Information Unstructured Text: Identifying and capturing complex relationships.

[32] Bontcheva, K., V. Talban, D. Maynard, and H. Cunningham (2004). Evolving GATE to meet new challenges in language engineering. Natural Language Engineering 10(3/4), 349–373

[33] H. Cunningham, D. Maynard, K. Bontcheva and V. Tablan, Gate, A framework and graphical development environment for robust natural language processing tools and applications, in Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics' 2002.

[34] De Sitter A., Calders, T. and Daelemans, W., A formal framework for evaluation of information extraction, 2004.

[35] Tilahun Gamta, The Oromo language and the latin alphabet, Journal of Oromo Studies, 1992.

[36] Mandefro Legesse, Named Entity Recognition for Afaan Oromo, Master‟s Thesis, School of Graduate studies, Addis Ababa University, 2012.

[37] Abdi Sani, Afaan Oromo Named Entity Recognition Using Hybrid Approach, Master‟s Thesis, School of Graduate studies, Addis Ababa University, 2015.

[38] Tim Berners-Lee, James Hendler, and Ora Lassila, The semantic web. Scientific American, pages 29–37, 2001.

[39] Blomqvist, 2005

[40] Ahmed, Z, Domain Specific Information Extraction for Semantic Annotation. (Unpublished Master Thesis), Charles University, 2009.

[41] Nakashole, Weikum, & Suchanek, 2012

[42] (Cimiano & Volker, 2005)Learning concept hierarchies from text with a guided agglomerative clustering algorithm.

[43] (Shamsfard & Barforoush, 2003)

[44] (Corcho et al., 2003)[

[45] Champin, 2001

[46] (McGuinness & Van Harmelen, 2004).

[47] Azevedo et. al. (Ribeiro de Azevedo et al., 2014)

[48] Hassanzadeh (Hassanzadeh, 2013)

[49] Gantayat (Gantayat, 2011)

[50] (El-Khair, 2006).

[51] Sekine, S. and Ranchhod, E., Named Entities Recognition, classification and use, Special issue of Lingvisticæ Investigationes, 30(1): 3-26, 2007.

[52] (Cunningham et al., 2009).

Citation: *SADIK ABAS, MELKAMU BEYENE, "Automatic Domain Ontology Learning From Unstructured Afaan Oromo Text", International Journal of Research Studies in Science, Engineering and Technology. 2020; 7(12):18-27*