International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Ontology learning on product reviews to extract aspects and opinions

SUNIL BHUTADA, B.SHIVANI, C. SWEETY, E. NIKHIL BHARGAVA, K.DHANVI

***Abstract:** More number of online product reviews, into e-commerce database, from time to time on a daily basis are produced. In order to analyze huge number of reviews for aspects and opinions is a complex task. This is because these reviews that are produced from time to time are not properly structured and there is a lot of fancying in the literature. This often makes the language, unstructerd and thus makes it difficult to analyze. 'NLP', which means the Natural language processing and Ontology learning techniques are used to automate these tasks. The semantic gap (gap between written reviews and the actual knowledge) was observed when aspects and opinions are extracted through these techniques. The original Ontology learning (OL) reduces this gap. Maximum number of aspects and opinions extraction is estimated using OL. These aspects and opinions can be found individually or in pairs..*

***Keywords:** Association Rule Mining, Syntactic Analysis, Apiori Algorithm, Combination Approach, Resource Description Framework, Lexico Syntactic Pattern, Natural Language Processing(NLP).*

I. INTRODUCTION

In the new era, every product and service is taken into consideration only after reviewing about the product. Review is an evaluation of products, books, services provide, performances and so on. It can be done in various ways like orally, manually and through online. Online Reviews play a significant role for the customers to shop anything they need. Based on the reviews the customer decides if the certain product and service can be bought or not. So, it is important for the market to maintain the standards in order to sustain. The requirement to analyze is to create a link between analyzed reviews and ontology. Ontology Engineering is basically forward Ontology. Forward Ontology is converting the concepts of data into individual terms. The concepts are taken as the basic input and through step by step process it is converted to terms for understanding. There are various tools and models to build this. Semantic Gap is a gap between the written content and the actual knowledge. The basic data and the processed ontology will not entirely build the relations between concepts and terms. In ontology Engineering, while analyzing the online reviews there is no unique way to represent the reviews.

There is no rule or constraint followed while analyzing. The concept used here is "WHAT THEY FEEL IS WHAT THEY WRITE". There is no compulsion on grammar which makes retrieval complicated because analyzation happens grammatically on the basis of grammar used. Ontology Learning means building concepts from terms.
Conversion of terms to concepts is the basic idea of Ontology Learning. It is the reverse of Ontology Engineering. It is the solution to the semantic gap that is created by Ontology Engineering.
The reviews are collected from the online e-commerce platforms where the customers give their respective feedbacks. These reviews are pre-processed by removing the stop-words, unnecessary punctuation marks, etc. Key words are known, after which they undergo through the parts of speech tagging.
OI model (underlying vocabulary for building ontology); Text corpus (preprocessed product reviews collection as input); Terms extraction; Association Extraction (frequent term combinations are extracted and framed as rules);
Pruner (a step used in post preprocessing phase);Taxonomy builder (approach identifies concepts and sub-concepts from extracted terms); Instance extraction (the concepts and sub-concepts and the remaining terms are implicitly extracted as instances); Relation learning (applying restrictions on learned concepts and associations); Ontology Comparison (evaluating the obtained ontology for having better knowledge model); Final learned Ontology.

II. RELATED WORKS

Here the collected reviews are implemented on Ontology Learning, in a procedural manner.
The raw data i.e., reviews are taken as input. This is called as Text Corpus [1]. Pre -Processing is done using certain techniques, which are linguistic, such as parts of speech tagging, lemmatization, named entity recognition and parsing. But here we use only parts of speech tagging. Pre-Processing makes the data more easy and understandable. This gives concepts (of domain) and terms are accurate. POS tagging [1] is assigning each token to its corresponding token such as noun, verb, adjective, adverb; Verbs can be events or states or actions, beliefs or attitudes, Nouns which are classes; Adjectives are class attribute values; Adverbs describes the manner. The representation of noun is _nn, adjective is _jj adverb is _rb, verb is _vb. According to the above representation, every word in the corpus is given the POS tagging.

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
Sunil Bhutada*, Professor, Dept. of Information Technology, Sreenidhi Institute of Science and Technology (Autonomous), Hyderabad
B.Shivani, Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology (Autonomous), Hyderabad
C.Sweety, Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology (Autonomous), Hyderabad
E.Nikhil Bhargava, Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology (Autonomous), Hyderabad
K.Dhanvi, Student, Dept. of Information Technology, Sreenidhi Institute of Science and Technology (Autonomous), Hyderabad

The Term Extraction should be started with an OI model and a corpus to which the extracted terms must be added. These added terms are concepts. This is the step to be done after pre-processing of the data [2]. Term Extraction can be done using absolute frequency, entropy, c value and so on. The Relation Extraction utilizes affiliation principles to find hopeful connections between terms in a content corpus [3]. The relations are recorded and arranged which can be added as property or pecking order to a comparing OI-Model by the client. Association rule mining is an efficient method to describe relations among variables in large datasets [4]. Association rule mining aims to find the rules which enable to predict the instance of specific information based on the instance of the other information present in the document. Association rules are calculated from information which is made up of two or more modules. If rules are built by analyzing all the possible information, many rules can be formed that consist of runty meaning.

$$\text{support (X)} = \frac{\text{The number of transactions in which X appears}}{\text{The total number of transactions}}$$

$$\text{Confidence (X}\longrightarrow\text{Y)} = \frac{\text{Support (X}\cup\text{Y)}}{\text{support (X)}}$$

The methodology utilized to compare learned ontology and referenced ontology is Reference Ontology Based. The limitations could be as per the following:

1. Lexical Recall: Comparision of dictionaries between the OI-Models. It assesses the proportion of the OI-Model proficiency in vocabulary and language.
2. Lexical Precision: The exactness or precision of the OI-models and their terms and ideas are to be assessed.
3. Taxonomic Overlap: To distinguish and assess the taxonomical (order) covers in the framed philosophy.
4. Relational Overlap: To distinguish and assess the social (correspondence) covers in the picked up cosmology to play out the comparision between the OI-Models.
5. Sister Overlap: To distinguish and assess the comparable terms extricated from a similar idea covers in the picked up philosophy.

III. PROPOSED METHODOLOGY

Flowchart

The procedure of Ontology learning begins by extricating terms and their equivalent words from content. At that point relating terms and equivalent words are consolidated to frame ideas. From that point forward, ordered and non-ordered relations between these ideas are found. At long last, maxim schemata are instantiated and general sayings are removed from unstructured content. This entire procedure is known as philosophy learning layer cake.

SYNONYMS: In this layer, as terms are separated in upper layer and their equivalent words are to be removed in this layer.

•CONCEPTS: In this layer, Interpretation of information and extraction of terms to distinguish critical ideas. Idea Extraction is absolutely founded on content in a corpus. Connection between the ideas can be distinguished.

•CONCEPT HIERARCHY: In this layer, ideas separated can be utilized to shape chain of command level. An idea chain of command characterizes an arrangement of terms from a lot of low-level ideas to more elevated amount, progressively broad ideas.

•RELATIONS: In this layer, Relationship extraction is the undertaking of removing semantic connections from content. Separated connections more often than not happen between at least two substances of a particular sort and fall into various semantic classifications. The relations are recorded and arranged by certainty or support and can be added as property or pecking order to a comparing OI-Model.

•RELATION HIERARCHY: In this layer, relations extricated can be utilized to frame a pecking order. The utilization of progressive connections is the essential component that recognizes a scientific classification or from the other one. Various leveled connections depend on degrees or dimensions of super-ordination and subjection, where the super-ordinate term speaks to a class or a whole, and subordinate terms allude to its individuals or parts.

•AXIOMS SCHEMETA: A maxim mapping is a recipe in the meta-language of an aphoristic framework, in which at least one schematic factors show up. These factors, which are meta-linguistic builds, represent any term or sub equation of the framework, which could possibly be required to fulfill certain conditions.

•GENERAL AXIOMS: A maxim is an explanation that is taken to be valid, to fill in as a reason or beginning stage for further thinking and contentions.

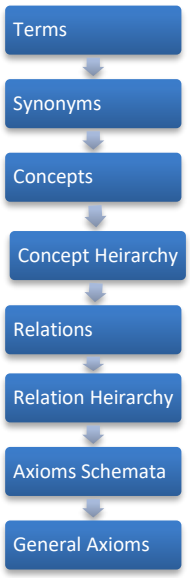


Figure 1: Flowchart

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

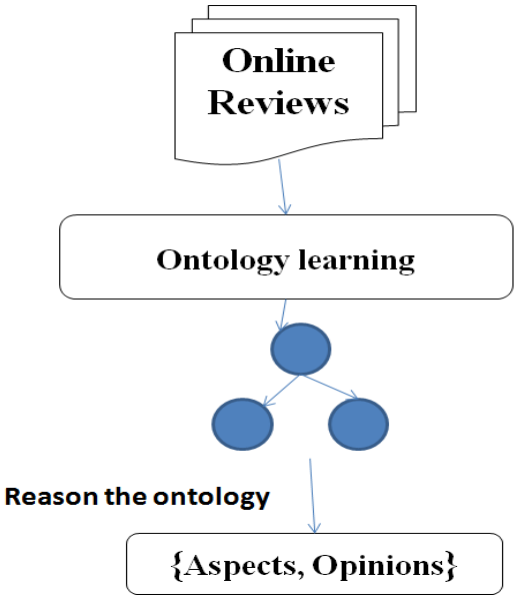


Figure 2: Block Diagram / Working model

The working model represents the following steps:
The reviews are collected from the online e-commerce platforms where the customers give their respective feedbacks. These reviews are pre-processed by removing the stop-words, unnecessary punctuation marks, etc. Key words are known, after which they undergo through the parts of speech tagging.
The corpus is loaded along with the keywords and the ontology learning process takes place in the following order:
OI model → underlying vocabulary for building ontology
Text Corpus → preprocessed product reviews collected as input
Terms Extraction
Association Extraction → frequent term combinations are extracted and framed as rules
Pruner → a step used in the post preprocessing phase
Taxonomy Builder → approach identifies concepts and sub-concepts from extracted terms
Instance Extraction → the concepts and sub-concepts and remaining terms are implicitly extracted as instances
Relation Extraction→ applying restrictions on learned concepts and associations
Ontology Comparision→ evaluating the obtained ontology for having better knowledge model
Final Learned Ontology
Datasets are gathering of related arrangements of data that is made out of independent components however could be controlled as a unit by PC. An informational index is composed into some kind of information structures.
1) Datasets are gathered from Amazon Online Reviews where clients give their individual inputs.
2) The gathered audits are pre-handled by expelling the stop-words, superfluous accentuation marks and so forth.
3) Key words are known, after which they experience through the grammatical forms labeling.
The corpus is stacked alongside the catchphrases and the philosophy learning process happens.
After the last philosophy is found out, these are made to experience different inquiries where it can recover and control the information put away in the RDF (Resource Description Framework) design.

Datasets are put away in documentaries.

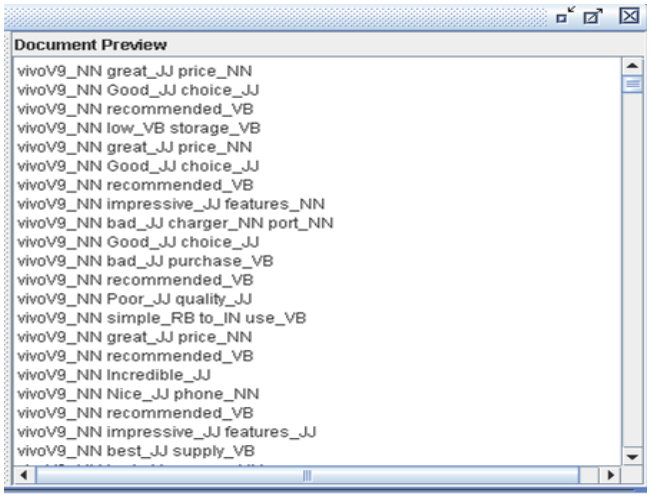


Figure 3: An example dataset for Implementation

IV. IMPLEMENTATION

Term Extraction

The main aim of term extraction is to emphasize on linguistic patterns. The linguistic patterns can be defined as grammatical rules or constraints that govern and allow their users to use the language in a proper and understandable way. Considering the linguist's perspective, Grammar is not just collection of rules, but rather it is defined as a set of blueprints which helps in guiding and governing the users in producing meaningful, comprehensible and predictable sentences. The linguistic pattern can be taken as and how the implementation of the terms extracted demand, although a default linguistic pattern is provided. This is also called as Linguistics for knowledge extraction. For our project implementation the linguistic pattern considered is (NN)*(JJ)*(NN) +

This is because the dataset we have extracted demands the linguistic pattern to be in this way, as in, the form of a noun followed by an adjective which is followed by a noun again.

Since our datasets are concerned with the product reviews therefore, every term starts with the name of the product which is a noun and then followed by the description or the review of that product which is an adjective and then the end would be a noun again.

The considered linguistic pattern is then applied to the process of term extraction where the extracted terms are based on the given linguistic pattern thereby, making the term extraction semantically genuine.

Concept Extraction

To extract the concepts, we either can use the seed words in the extraction process to retrieve domain-specific words from given datasets and make use of them from the corpus to retrieve the concepts for the construction of ontology or we can include the seed words manually in a text file to make the extraction process easy. However, it is recommended that the use of lexicons is also *feasible for the extracting concepts.*

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

However, here we emphasize on the impact of wordnet on extraction of concepts. The word-net concerning to our ontology learning process can be considered as the datasets and corpus taken by the user, whatever the concepts are extracted are directly related to the selected wordnet. Thus, impact the extraction of concepts.

Relation Extraction

The relation extraction is done to infer the underlying relations between the extracted terms and concepts from the previous modules. The relation extraction mainly emphasizes on the association rule mining to obtain those relations between the extracted terms and concepts. As a technique of relation extraction, lexico-syntactic patterns are put into use. It is an approach with respect to rule-based method that plays a vital role in the extraction phases of taxonomic and non-taxonomic relation which is concerned to ontology learning. In order to extract relations, the algorithm used here utilizes regular expressions. This type of approach which is concerned to rule-based is quite useful in the extraction of is–a relationship which is used to extract taxonomic relationships. On the other par, patterns like 'NP is-a-part of NP' can be used in the extraction of non-taxonomic relationships. Besides, Semantic lexicons are known for their knowledge resources under the domain of ontology which plays an important role at various levels of ontology learning. In this module, the Semantic lexicons can be used in the extraction taxonomic and non-taxonomic relations

Although, the emphasis on association rule mining analysis of learned association rules for concept hierarchies and object properties plays another significant role.

Evaluation

Evaluating the quality and genuineness of ontology acquisition is a very important aspect as it is responsible in allowing the researchers to evaluate the correctness at lexical level, Coverage at conceptual level, wellness at taxonomical level and adequacy at non-taxonomical level of yielded ontologies.

Evaluation of acquisition securing creates possibility to re-characterize and re-design the whole ontology learning process in the event of surprising resultant ontologies, which don't fit with the particular necessities of a client. As examined before, ontology learning is a staggered procedure so this makes the assessment procedure of ontology extraction really hard.

Thinking about the multifaceted nature of assessing space ontologies, innumerable assessment systems have been proposed in the recent years and this zone is still under constant advancement. All proposed systems fall under one of these classifications, which are commonly characterized based on sort of target ontologies and motivation behind assessment.

Different types of evaluation techniques are

- Golden standard-based evaluation
- Application-based evaluation
- Data-driven evaluation
- Human evaluation

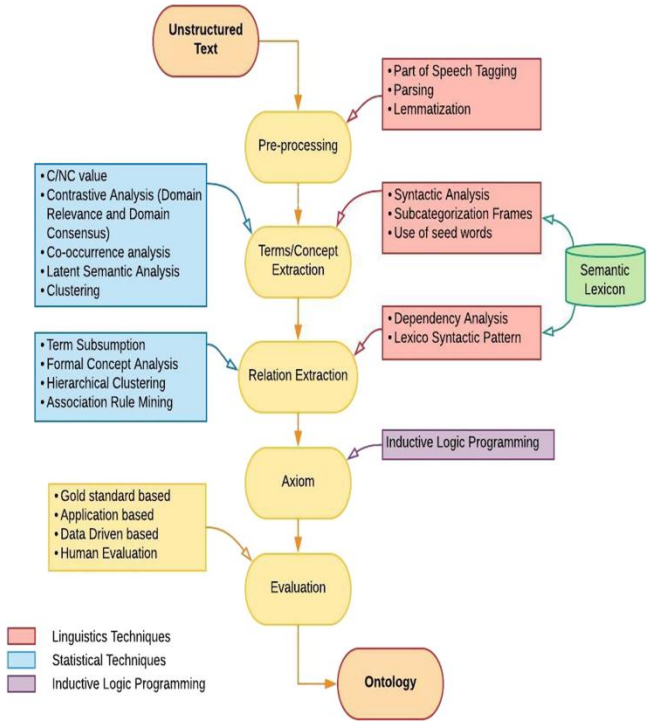


Figure 4: Ontology Learning Approach (from reference [1])

Human evaluation

Human evaluation of ontologies is by and large dependent on characterizing and detailing different choice criteria for the choice of best ontology from a predetermined arrangement of competitor ontologies. A numerical score is allotted in the wake of assessing ontology against every rule.

V. RESULT ANALYSIS

Preprocessed reviews:

vivo_NN nice_JJ phone_NN
vivo_NN nice_JJ upgrade_VB
vivo_NN easy_JJ setup_VB
vivo_NN works_VB good_JJ
vivo_NN great_JJ phone_NN
vivo_NN very _JJ unhappy_VB
vivo_NN very_JJ dissapointed_VB
vivo_NN damaged_VB goods_NN
vivo_NN very_JJ pleased_VB
vivo_NN worst_JJ experience_VB
vivo_NN good_JJ phone_NN
vivo_NN really_RB appreciate_VB
vivo_NN bad_JJ recommendation_VB
vivo_NN worst_JJ phone_NN

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Terms extracted=20
(Afer term extraction)

Terms	Frequency	TFIDF	Entropy	C-Val
Worst_jj	2	2.996	1	-13.338
Work vb	1	2.996	1	-13.338
Good jj_	2	2.996	1	-13.338
Vivo jj	12	2.996	1	-13.338
setup nn	1	2.996	1	-13.338
Nice jj	1	2.996	1	-13.338

Table 1: Term Extraction results

Association Extraction

premise	Conclusion	support	confidence	Abs-freq
Vivo nn	Phone nn	0.15	0.214	3
Phone nn	Vivo nn	0.15	0.75	3
Vivo nn	Very jj	0.1	0.143	2
Very jj	Vivo nn	0.1	1	2
Vivo nn	Bad jj	0.05	1	1
Vivo nn	Good jj	0.05	0.071	1

Table 2: Association Extraction results

Pruning

Terms	Frequency	Cumilative Frequency
Worst_jj	3	3
Work vb	1	1
Good jj_	3	3
Vivo jj	12	12
setup nn	1	1
Nice jj	1	1

Table 3: Pruning results

Evaluation

Lexical Recall	0.65217
Lexical Precision	0.9375
Taxonomic Overlap	0.9375
Relational Overlap	0
Sister Overlap	0.625

Table 4: Evaluation Measure after comparing ontologies

It is observed from this table that, 0.65217value of Lexical Recall; 0.9375 value of Lexical Precision; 0.9375 value of Taxonomic Overlap; 0.625 value of Sister Overlap, inferred from the learned ontology, which was implemented by human evaluation, reference based ontology comparision. These results indicate that ontology learning for product reviews identifies the maximum number of aspects and opinions along with acquisition of more domain knowledge.

Visualisation of learned Ontology

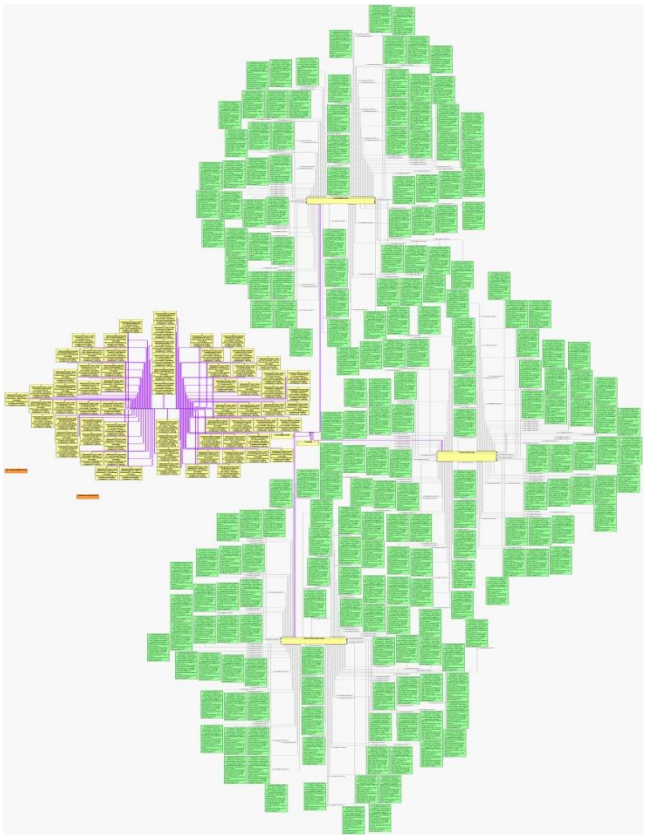


Figure 5: Graphical representation of the Learned Ontology (ASPECTS – yellow **and OPINIONS -** green**)**

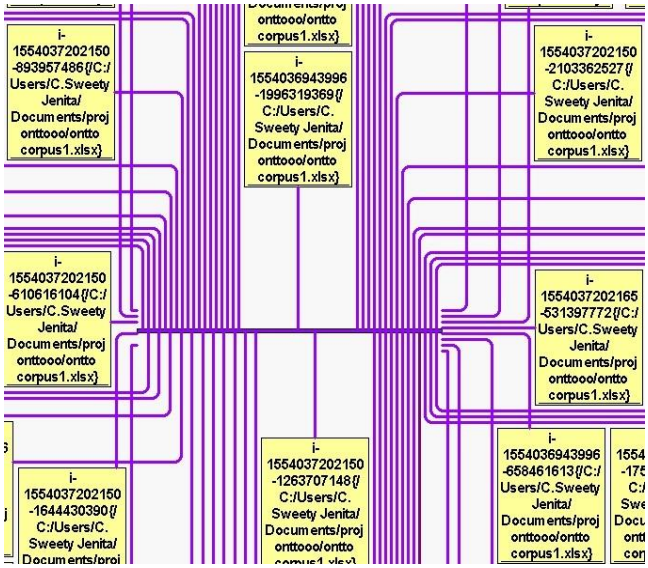


Figure 6: The aspects that are learned from the ontology (a part from the whole aspects in Figure 5)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

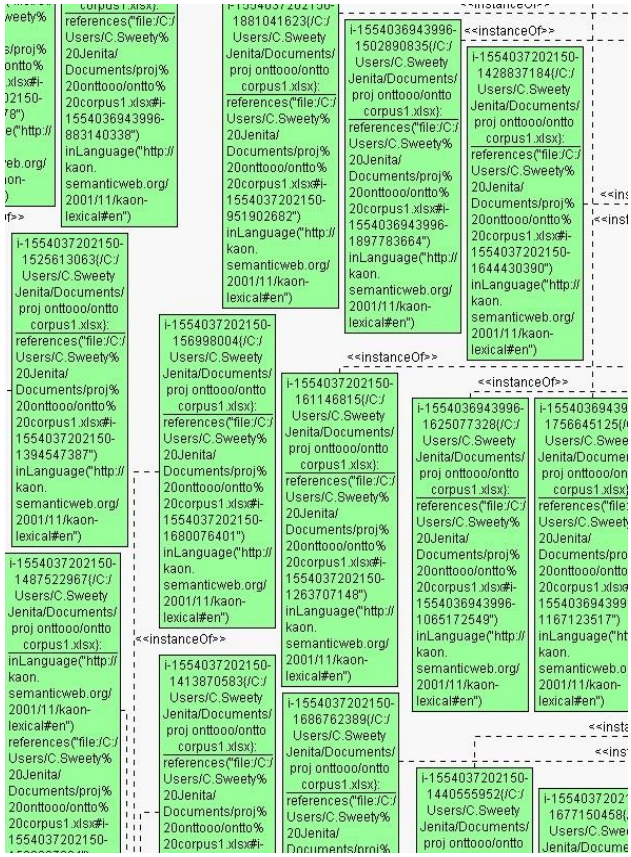


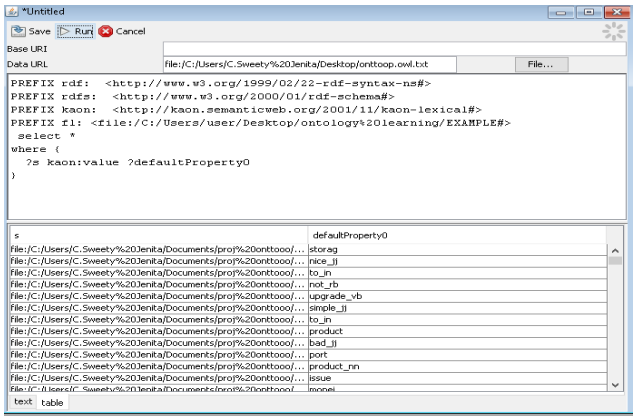
Figure 7: The opinions that are learned from the ontology (a part from the whole opinions in Figure 5**)**

VI. Testing of the Learned Ontology

Following sparQL queries are executed on the learned Ontology

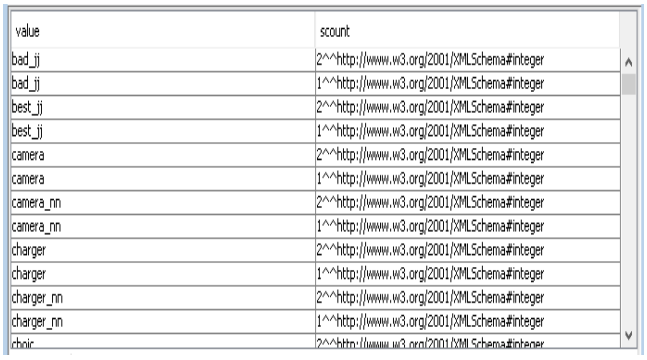
1)

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX kaon:  <http://kaon.semanticweb.org/2001/11/kaon-lexical#>
PREFIXfl:
<file:/C:/Users/user/Desktop/ontology%20learning/EXAMPLE#>
 select *
where {
  ?s kaon:value ?defaultProperty0
}
```



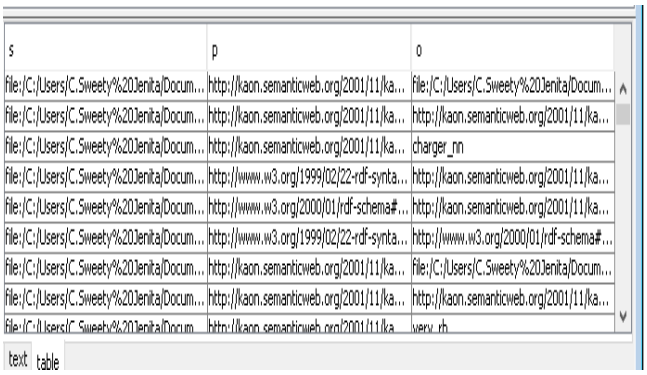
2)

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX kaon:  <http://kaon.semanticweb.org/2001/11/kaon-lexical#>
select ?value (count(?Synonym) as ?scount)
where {
  {?Synonym kaon:value ?value}
 }
GROUP BY ?value ORDER BY ?scount
```



}3)

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX kaon:  <http://kaon.semanticweb.org/2001/11/kaon-lexical#>
PREFIXfl:
<file:/C:/Users/user/Desktop/ontology%20learning/EXAMPLE#>
 select *
where {
  ?s ?p ?o
```



VII. CONCLUSION AND FUTURE WORK

Ontology learning is a multidisciplinary task that separates significant terms, characteristics and relations from unstructured content by obtaining procedures from various domains like text classification, natural language processing. Natural language processing has different bottlenecks, for example, grammatical feature labeling, relation extraction from unstructured data. From results examined in the segment entitled Linguistics for pre-processing, it tends to be reasoned that methods like PoS labeling and parsing can lead toward the advancement of better ontologies. With the progression in NLP methods,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

improved PoS taggers and parsers are being acquainted that needs with be blended into ontology learning frameworks for better execution. In content arrangement, specialists are growing new calculations to choose profoundly discriminative highlights among the classes.

Challenge 1: Existence of semantic gap between the knowledge and text which leads to absence of unique method of representation of reviews.

Proposed Solution: Our very own ontology learning, which uses the reverse methodology for implementing the ontology, eliminates this gap. Since we use the approach for extracting concepts from the pre-processed terms, which gives us the chance to carefully eliminate the unnecessary terms and which also allows us to only include domain specific terms instead of the complete wordnet.

Challenge 2: "WHAT THEY FEEL IS, WHAT THEY WRITE"

This concept leads to absence of rules and constraints, over the language or the literature that the customer uses to write reviews. The fancying of the literature leads to a bizarre and a social chaos. This analysis happens through grammar, which makes the retrieval complicated.

Proposed Solution:

The process of pre-processing, term extraction, pruning and evaluation makes sure that the reviews collected are semantic and syntactic. Thus, leads to a better retrieval and analysis.

In future, a better algorithm for building and learning concept hierarchy and object property will be developed, also the reasoned aspects and opinions are further analyzed towards providing case-based recommendation systems.

ACKNOWLEDGEMENT

We would like to thank our Research Guide Dr. Sunil Bhutada, Professor in Department of Information Technology, SNIST, Hyderabad for the continuous support and valuable suggestions throughout the work carried. Authors are also grateful to the reviewer for the renovation of the manuscript. We would also like to thank the Department of Information Technology, for providing us with the facility for carrying out the simulation.

REFERENCES

1. Muhammad Nabeel Asim, Muhammad Wasim, Muhammad Usman Ghani Khan, Waqar Mahmood and Hafiza Mahnoor Abbasi, A survey of ontology learning techniques and applications, a REVIEW by DATABASE, 2018.
2. Paul Buitelaar and Phillip Cimiano, Ontology learning and Population: Bridging the Gap between Text and Knowledge, IOS Press, 2008.
3. Phillip Cimiano, Ontology learning and population from Text algorithms, Evaluation and applications, First Edition, Springer Science+Business Media, LLC, 2006.
4. Texttoonto user manual, 2003.
5. Online source: https://www.w3.org/RDF/Validator/
6. Online source: https://www.cognitum.eu/semantics/FluentEditor/
7. Online source: ttps://www.w3.org/TR/rdf-sparql-query/

AUTHORS PROFILE

Sunil Bhutada is presently working as a Professor in IT at Sreenidhi Institute of Science and Technology, Hyderabad. His areas of intrest are Data Mining and Text Mining, Topic Modeling, Network Security Cryptography, Computer Networks. He worked as the Program Committee Member of few prestigious conferences like Sixth International Conference on Network Communications Security (NCS - 2014), The Sixth International Conference on Wireless Mobile Networks (WiMoNe - 2014), Sydeney, Australia. He has published many research papers in the elite areas of Information Technology.

B.Shivani is the graduate student from the department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad. Her areas of interests are Web & Data Mining, Natural Language Processing.

C.Sweety is the graduate student from the department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad. Her areas of interests are Web & Data Mining, Natural Language Processing.

E.Nikhil Bhargava is the graduate student from the department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad. His areas of interests are Information Security, Data mining and text mining.

.

K.Dhanvi is the graduate student from the department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad. Her areas of interests are Natural Language Processing, , Data mining and text mining.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)