



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

Ontology Based Information Retrieval on Twitter Data for Automobile Domain

Priya Gupta

Faculty of Engineering and Technology, CSE, MRIU at Faridabad - Haryana, India

ABSTRACT: Corpus Data refers to as a collection of huge datasets. Sentiment analysis contributed to a popular research area for twitter. The sentiment analysis done without feature extraction fails to give the deep result about the users opinion but, features of the domain are extracted by building ontology which helps in getting the refined sentiment analysis. Ontology means a formal, explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of some world phenomena. Ontology is used for knowledge sharing and reuse. It improves information organization, management and understanding. In this paper, we have used ontology to analyse the tweets to increase augmentation and efficiency of sentiments which is obtained using naïve Bayesian algorithm. The work is done in five stages. In first stage, the tweets are extracted from Twitter4J and stored in a repository. Then sentences are extracted one by one. Sentences extracted are simplified by removing stop words and redundant words. In Second stage, the words left in the sentences are used for sense matching using WordNet-an online semantic dictionary. WordNet dictionary is used to extract features from tweets. In Third stage, Ontology is being generated by using java customized code. Crawler is being designed next to get the details about the automobile domain. The data is stored in text manner. In fourth stage, Mapping of data is done which includes mapping of ontology with the crawler data, together with ontology validation. In fifth stage, Analysis of tweets is done using ontology by applying naïve Bayesian algorithm and comparison of automobile is done which one is better and what all are the attributes that other automobile does not fall into this category.

KEYWORDS: Sentiment analysis, Ontology, Tweets, Twitter4J

I. INTRODUCTION

Corpus Data is referred to as a group of huge datasets having a large quantity of information. Corpus Data is produced from varied sources like social networking sites including Facebook, Twitter etc, and therefore the data which is generated are in varied formats like structured, semi-structured or unstructured format. Twitter is considered as a common research area for sentiment analysis. For a variety of domains it offers various advantages. The sentiment analysis prepared without extraction of features fails to provide the deep result containing the user's opinion. Features of the domain can be mined by developing ontology that helps in obtaining the refined sentiment analysis. Ontology is referred as a formal, explicit specification of a shared conceptualization. Conceptualization is an abstract model of approximately world phenomena. The relationship among ontology concepts and their relations needs to be explicitly defined. Benefits obtained from ontology are bridging the linguistic gap between the information, applications, data processing algorithms, and data processing results. To provide data processing algorithms with a priori information that helps in guiding the mining process. And to have a formal way of expressing the data mining flow, starting from data pre-processing to mining results

The need of using ontology includes sharing the common understanding for the structure of knowledge between people and software agents. Also, ontology helps in enabling reuse of the domain knowledge. Lastly, to form domain assumptions explicit

There are various tools available for ontology creation. Figure 1 shows the comparison between tools on various different features like extensibility and ontology storage, tool's interoperability: in that Import Format, Export Format and Merging features are discussed. Tool's inference services which include Inference Engine, Exception Handling and Consistency Checking. Tools' usability discusses Collaboration with other tools, Ontology Library and Visualization. Overview of Tools' versioning and collaborative work support on the basis of Versioning and collaboration. Different



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

tools provide documentation of ontology, import/export of ontology in different formats, visual view for ontologies, ontology libraries as well as attached inference engines. It is quite clear Ontology development is an ad-hoc approach. Among several viable alternatives, one need to find which one would work better for the projected task that can easily and effectively be maintained and expressed. Though foundation of ontology is logic but it is a model of reality and the concepts in the ontology must reflect this reality.

Figure 1: Compression of Tools

Features	Apollo	IsaViz	Protégé	Swoop
Extensibility	No	No	Via Plugin	No
Ontology Storage	Files	Files	Files and DBMS	Files
Import Format	OCML	XSLT, RDF(S), OIL, DAML+OIL, OWL	XML, RDF(S), XML Schema and OWL	RDF(S), OIL, DAML
Export Format	OCML	XSLT, RDF(S), OIL, DAML+OIL, OWL	XML, RDF(S), XML Schema, Java, HTML	RDF(S), OIL, DAML
Merging	No	No	Via ANCHORPROMPT plug-in	No
Inference Engine	No	Yes	With PAL	No
Exception Handling	No	No	No	Yes
Consistency Checking	Yes	Via type inheritance and detection of cycles in hierarchies	Via plug-ins like FACT and PAL	Only checks writing mistakes
Collaboration with other tools	No	No	No	No
Ontology Library	Yes	No	Yes	No
Visualization	No	Via plug-ins like GraphViz	Via GraphViz	No
Versioning	No	No	Yes	No
Collaboration	No	No	No	No

For Ontology development[8], effective tools are a central need. To conclude, there are various open source ontology tools like Protégé[7], there are ontology tools that demand learning/knowing a specific language i.e. SWOOP and there are ontology tools that are more graphic i.e. IsaViz. Other tools are Web-based application i.e. Apollo and SWOOP or follow a methodology i.e. Protégé and SWOOP. Therefore, protégé is most convenient ontology development tool offered by Stanford.

II. LITERATURE SURVEY

In [1] authors have proposed the deployment of ontology-based techniques for determining the subjects discussed in tweets and breaking down each tweet into a set of aspects relevant to the subject. Their result is the assignment of a sentiment score to each distinct aspect. A baseline scenario has also been presented that deals with the domain of a popular product (smartphones) and results in comparatively evaluating the distinct features of each model series. In [2] authors have proposed custom sentiment analysis tool for twitter that increases the performance by increasing the overall scoring of tweets compared to the third party result, with the deployment of the ontology the subjects discussed in tweets, are analysed and the score is assigned to the feature of laptop and also the opinion words helps in identifying the mood of people and scores the tweets to the best. In [3] authors have focussed on SPARQL i.e. a W3C standard, which is a built in querying tool available with Protégé. Though it supports for RDF and RDF graphs, it still works with OWL ontology. Their paper presented the purpose of using SPARQL on the OWL ontology to verify whether it is capable enough to query OWL ontology. Their results so obtained were a proof of their support. In [4] authors have developed a sample ontology which was tested by executing semantic queries against it. All the nodes as well as each instance metadata could be searched through the SPARQL query. According to them, the main advantages of using ontology are reusability and semantically searchable. Also, in the era of Semantic Web, the ontologies have become a powerful tool for knowledge sharing and it also supports the semantic interoperability among heterogeneous distributed systems. In [5] authors has aimed to create a process of sentiment analysis based on ontologies in the automobile domain. The process aims at making a social media analysis, identifying feelings and opinions about brands and vehicle parts. The proposed process is capable of generating information that answers questions such as: "In the opinion of the customer, which car is better: Corsa or Palio? Which one is more beautiful? Which engine is stronger?"

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

To answer these questions by comparison, one can show a general view reflected on different social networks, indicating, for example, that for a given vehicle, a certain percentage of responses are considered positive, while for others, the percentage is considered negative. In [6] authors have focused on file metadata and file content metadata can be extracted from the Application files and folders using API's. The extracted components can be stored in the Hadoop Distributed File System along with the application environment. Extracted metadata will be in XML format. XML deals with syntactic level and the Web Ontology Language (OWL) supports semantic level for the representation of domain knowledge using classes, properties and instances. This paper converts the data model elements of XML to OWL Ontology that implements the mapping the standard XML technology XSLT.

III. EXISTING APPROACH

A set of tweets is given as an input about a particular domain and for every feature found in the tweet the sentiment score is assigned. Below are the three steps involved in this methodology [2]:

- (i) Creation of domain ontology for the particular problem that is under consideration.
- (ii) Extraction of the relevant tweets to the features of the domain from twitter.
- (iii) To do the sentiment analysis for the retrieved tweets.

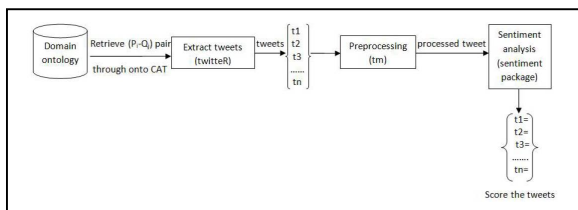


Figure 2: Architecture of Existing approach

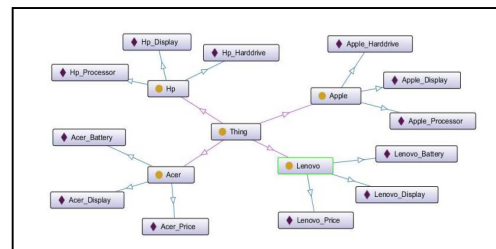


Figure 3: Ontology of Laptop

Figure 2, shows the architecture of existing approach in which first of all domain ontology is created using protégé 4.122 and OWL language for the ontology creation. Figure 3, shows the ontology of laptop created using protégé. After the creation of the ontology, the querying of the object-attribute pair is done in DL query or SPARQL but as the approach is using the implementation language as R so the onto CAT package of R is used for querying which is a Java library depending on OWL API and Java package of R. Formal concept analysis is used for the easy representation of the ontology and from the objects and attributes collection ontology can be easily derived. For the ontology in figure 3, $C = \{\text{Domain}\}; \{\text{Laptop}\}$, $P = \{\text{set of objects}\}; \{\text{Lenovo, HP, Apple, Acer}\}$ and $Q = \{\text{set of attributes}\}; \{\text{Display, Processor, Screen size, Hard drive}\}$. Secondly, Extraction of Tweets based on the Features Identified For extraction, consumer key and consumer secret key, access token, access secret token are granted to user which a user can to implement the authentication of the application. "twitterR" package available in R is referred as twitter REST API which is used for retrieving the tweets based upon the given set of keywords. They passed the search term as "lenovo+battery", the tweets that contain "Lenovo" and "battery" are extracted and also the tweets relevant to the Lenovo's battery are retrieved. Lastly, Sentiment Analysis is done. Below is the sequence of steps involved in Sentiment Analysis.

- (i) The data retrieved needs to be cleaned which also known as data preprocessing. The cleaning of data can be easily done by using the text mining package available in R. Cleaning includes, removal of the punctuations, numbers, '@' symbol and removal of URLs (the tweets starting with http://). Removal of only the '#' hash tag from the sentence.
- (ii) Lemmatization plus the tokenization of tweets are also done by using packages available in R.
- (iii) Sentiment package available in R is used for classifying the polarity of the tweets and it uses the inbuilt emotion dataset for approximately classifying the emotions into six categories like anger, joy, disgust, fear, sadness and surprise. The opinion lexicon dictionary consists of the positive opinion words list and also the negative opinion words list for finding the opinion words in a sentence and scores it accordingly.
- (iv) After the sentiment is analyzed for each feature the histogram is generated for easy visualization of the sentiment score and also four csv files are created which holds the opinion, scores, plot and stack.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

- (v) The frequent words available in the tweets are found. Term frequency is defined as a ratio of how many times a word occurred in the tweets to the total number of words in the tweets and it is calculated as a normalized frequency.
- (vi) And finally a word cloud is created for the visualization of the important words.

GAPS IN EXISTING APPROACH

1. Negation Handling: Negation plays a vital role in altering the polarity of the associated adjective and hence the polarity of the text. Negation words in general, includes not, neither, nor. For example, "No wonder the restaurant is good" and "Not only the food was yummy, the interior and service was also excellent". The use of pure language processing fails to tackle negations completely.

2. Domain Generalization: Certain words exhibit different polarities when used in different domains. For example, "The movie was inspired from a Tollywood movie" (negative orientation) "I got inspired from the novel"(positive orientation) Here, the word 'inspired' exhibits two different polarities for two different contexts. Sentiment Analysis is usually carried out targeting a specific domain and this has depicted very good results of accuracy. But a generalized sentiment analyzer still remains a challenge because of difference in the meaning of a word/sentence in different domains.

3. Pronoun Resolution: The object for which the opinion is expressed is usually texted in the first sentence of the complete text or could be spotted in the early portions of the sentence. The other portions may also contain opinionated data about the object but the object may have been referenced by a pronoun like it, this, that, he, she etc. To resolve the pronoun i.e. for what it has been used for is a complex task.

4. Language Problem: People, organisations and software systems must communicate between and among themselves. However, due to different needs and background context, there can be widely varying viewpoints and assumptions regarding what is essentially the same subject matter Each uses different jargon. The consequent lack of a shared understanding leads to poor communication within and between these people and their organisations In the context of building an IT system, this lack of a shared understanding leads to difficulties in identifying requirements and thus in the defining of a specification of the system.

5. Mapping Slangs: Slangs are usually the informal short forms of original words often used in online texting. For E.g. gr8 is a slang used for opinionated word great, 5n or fyn for fine

6. Hashtag Segmentation: For handling hashtags, we looked for the existence of the positive or negative words in the hashtag. But there can be some cases where it may not work correctly. For example, "#thisisnotgood".

IV. PROPOSED APPROACH

The proposed approach uses an ontology which add features and meaning to the tweets and WordNet which is an online semantic dictionary, electronic lexical database of nouns, verbs, adverbs, and adjectives. WordNet helps in extraction features from tweets. In our proposed approach we are considering automobile domain. The proposed approach has the following steps:-

1. Tweet Extraction: Tweets are extracted from Twitter4J. Twitter4J is referred as an unofficial library of Java for the Twitter API. With Twitter4J, you can easily integrate your Java application with the Twitter service.

2. Tweet Repository: A repository is a central place in which an aggregation of data is kept and maintained in an organized way. The repository is made using SQL server management.

3. Sentence segmentation: Sentence segmentation is the process of determining the longer processing units consisting of one or more words. This task involves identifying sentence boundaries between words in different sentences. Since most written languages have punctuation marks which occur at sentence boundaries, sentence segmentation is frequently referred to as sentence boundary detection, sentence boundary disambiguation, or sentence boundary recognition. All of these terms describes the same task i.e. determining how a text should be divided into sentences so as to use for further processing.

4. Keyword extraction: Keyword extraction is a subtask of information extraction. The goal of terminology extraction is to automatically extract relevant terms from a given set of tweets.

5. Sense matching : Overlap value is obtained by matching the sense of the words with the help of WordNet.

6. Ontology: Ontology is created which involves relation between the tokens. Ontology helps in adding meaning to the tweets and relevant analysis can be obtained by removing redundant information and irrelevant information from the



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

tweets and augmented analysis is obtained. Figure 4, shows the ontology design created using java customized code. Customized ontology example is shown below. In our case, C= {Domain}; {Automobile}
P = {set of objects}; {Maruti, Hyundai, Tata, Mahindra,Honda,Volkswagen,Chevrolet,Ford,Toyota,Nissan} and Q = {set of attributes}; {Engine Capacity, Torque, HorsePower, GearBox, FuelEfficiency, Milage, Petrol, Diesel }.

```

<?xml version = "1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="AutoOntology" type="MyType1"></xsd:element>
  <xsd:complexType name="MyType1">
    <xsd:sequence>
      <xsd:element name="Manufacturer" type="MyType2" minOccurs="1" maxOccurs="unbounded" />
    </xsd:sequence>
  </xsd:complexType>
  <xsd:simpleType name="MyType3">
    <xsd:restriction base="xsd:string">
      <xsd:enumeration value="Maruti"></xsd:enumeration>
      <xsd:enumeration value="Hyundai"></xsd:enumeration>
    </xsd:restriction>
  </xsd:simpleType>
  <xsd:complexType name="MyType2">
    <xsd:sequence>
      <xsd:element name="ModelName" type="xsd:string"></xsd:element>
      <xsd:element name="Engine" type="MyType4"></xsd:element>
      <xsd:element name="GearBox" type="xsd:positiveInteger"></xsd:element>
      <xsd:element name="FuelEfficiencyInLtrs" type="xsd:positiveInteger"></xsd:element>
      <xsd:element name="MilageKMPL" type="xsd:positiveInteger"></xsd:element>
      <xsd:element name="Petrol" type="xsd:boolean"></xsd:element>
      <xsd:element name="Diesel" type="xsd:boolean"></xsd:element>
      <xsd:attribute name="CompanyName" type="MyType3" />
      <xsd:attribute name="Version" id="ID" type="xsd:string"></xsd:attribute>
      <xsd:attribute name="Price" type="xsd:decimal"></xsd:attribute>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="MyType4">
    <xsd:all>
      <xsd:element name="Capacity" type="xsd:string"></xsd:element>
      <xsd:element name="HorsePower" type="xsd:string"></xsd:element>
    </xsd:all>
  </xsd:complexType>
</xsd:schema>

```

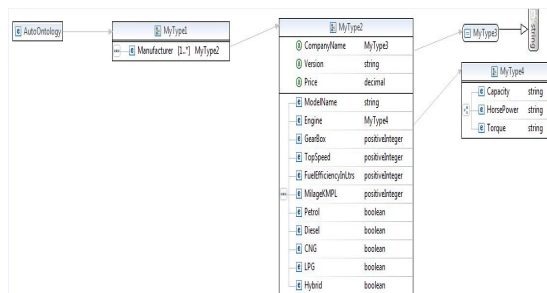


Figure 4: Ontology Design

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

7. Ontology Mapping and Validation: The conversion process requires at the most three steps for XML instance data and at least one step for XML Schema only. To process XML Schema only, the model of the ontology is created along with classes and properties. XML to OWL conversion uses classes namely owl:Class that will be emerged from xsd:complexType as well as xsd:elements according to the following rules mentioned below: for this case, an element in the source XML tree is always be a leaf, containing only of a literal and no other attributes. This element has to be mapped to an owl:DatatypeProperty having a domain class that denotes the adjacent element. Constraints like xsd:minOccurs and xsd:maxOccurs, that are used to map the equivalent cardinality constraints in OWL, owl:minCardinality and owl:maxCardinality are also contained in XML Schema.

8. Comparative Analysis: In this stage, comparative analysis is done using naïve Bayesian classification algorithm. We are comparing which vehicle is better over other and what are the other attributes that the other vehicle does not fall into this category.

The above discussed architecture is shown in figure 6. The flow of the proposed approach is shown in the figure7.

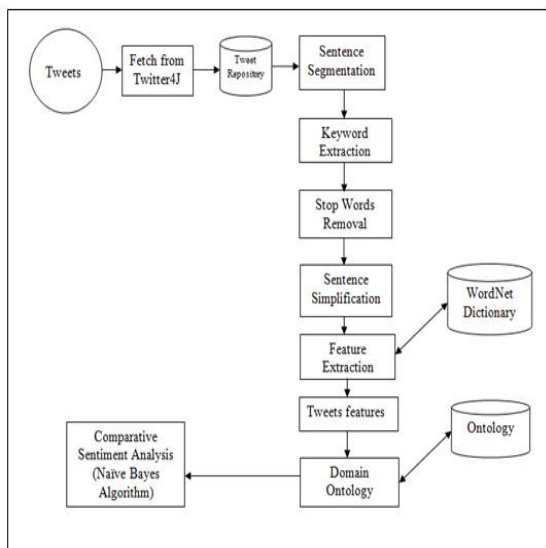


Figure 5: Architecture of proposed Approach

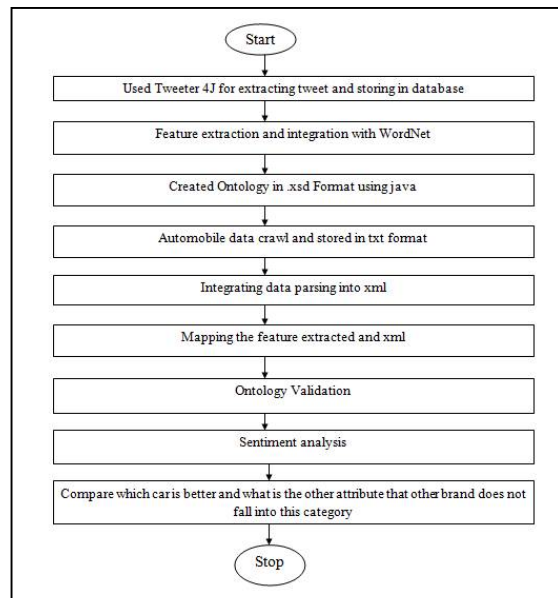


Figure 6: Flow chat of proposed approach

V. SIMULATION RESULTS

Figure 7, shows the graphical user interface in which we select the manufacturer of automobile. In Figure 8, Comparison table is achieved. Figure 9 shows comparison of cars on the bases of on the bases of tweets considering its features and opinions. In figure 10, we found that according to users sentiments in form of tweets, Nissan micra is positive in terms of prise, having average engine capacity and decent milage. Figure 11, show comparison with Nissan Sunny, in which we analysed that it is having decent price, good engine and decent milage. Figure 12, shows comparison of Chevrolet tavera, in which we found that it is having affordable price, good engine capacity, good fuel efficiency and low milage. Figure 13, shows the comparison with Chevrolet beat which is having good price, good engine efficiency and average milage. Figure 14, shows the comparison of Volkswagen ameo, which is convenient for users to buy in terms of price, decent engine, fuel efficiency and milage. Figure 15, shows the analytical comparison of proposed and existing approaches on bases of parameters like flexibility, easy to understand, code reusability, augmentation and usability.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

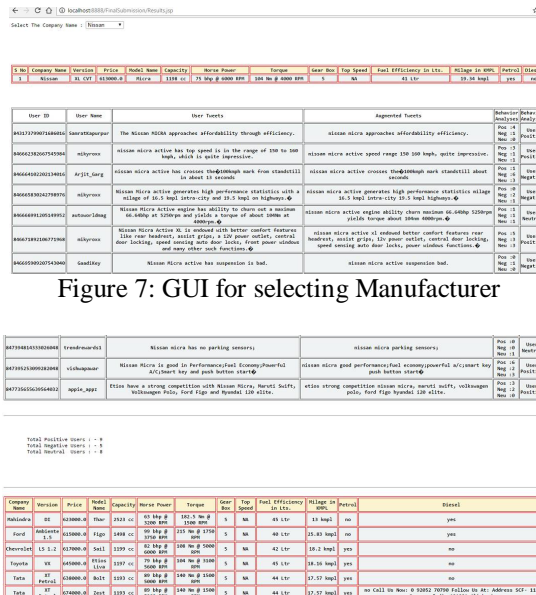


Figure 7: GUI for selecting Manufacturer

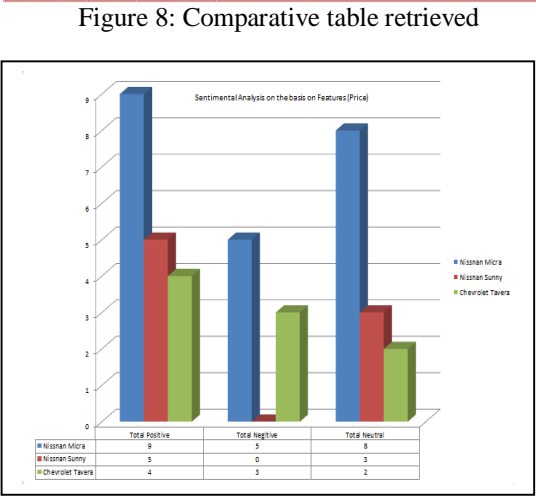


Figure 9: Comparison of cars



Figure 10: Comparison with Nissan Micra



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Nissan	XE	8E+05	Sunny	1498 cc	98 bhp @ 6000 RPM	134 Nm @ 4000 RPM	5	41 Ltr	16.35 kmpl	yes	no
Car in Comparison with Nissan Sunny											
Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Maruti	VDI	8E+05	Breza	1289 cc	89 bhp @ 4000 RPM	200 Nm @ 1750 RPM	5	48 Ltr	24.3 kmpl	no	yes
Chevrolet	Neo	8E+05	Tavera	2499 cc	80 bhp @ 3300 RPM	186 Nm @ 1800 RPM	5	55 Ltr	13.7 kmpl	no	yes
Maruti	Alpha 1.3	9E+05	Baleno	1248 cc	74 bhp @ 4000 RPM	190 Nm @ 2000 RPM	5	37 Ltr	27.39 kmpl	no	yes
Maruti	ZDI	8E+05	Dzire	1248 cc	74 bhp @ 4000 RPM	190 Nm @ 2000 RPM	5	42 Ltr	26.59 kmpl	no	yes
Hundai	SX 1.1	8E+05	Xcent	1120 cc	71 bhp @ 4000 RPM	180 Nm @ 1750 RPM	5	43 Ltr	24.4 kmpl	no	yes
Hundai	Sportz 1.4	8E+05	i20 Elite	1396 cc	89 bhp @ 4000 RPM	220 Nm @ 1500 RPM	6	45 Ltr	22.54 kmpl	no	yes
Hundai	SX 1.2	9E+05	i20 Active	1197 cc	82 bhp @ 6000 RPM	115 Nm @ 4500 RPM	5	45 Ltr	17.19 kmpl	yes	no
Hundai	VTVt 1.4	8E+05	veena	1396 cc	106 bhp @ 6300 RPM	135 Nm @ 5000 RPM	5	43 Ltr	17.4 kmpl	yes	no
Honda	S Petrol	9E+05	Mobilio	1497 cc	110 bhp @ 6600 RPM	145 Nm @ 4600 RPM	5	42 Ltr	17.3 kmpl	yes	no
Honda	S Petrol	9E+05	City	1497 cc	117 bhp @ 6600 RPM	145 Nm @ 4600 RPM	5	40 Ltr	17.4 kmpl	yes	no

Figure 11: Comparison with Nissan Sunny

Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Chevrolet	Neo	814000	Tavera	2499 cc	80 bhp @ 3300 RPM	186 Nm @ 1800 RPM	5	55 Ltr	13.7 kmpl	no	yes
Car in Comparison with Chevrolet Tavera											
Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Maruti	VDI	816000	Breza	1289 cc	89 bhp @ 4000 RPM	200 Nm @ 1750 RPM	5	48 Ltr	24.3 kmpl	no	yes
Maruti	Zdi	913000	Glz	1373 cc	91 bhp @ 6000 RPM	130 Nm @ 4000 RPM	5	43 Ltr	20.7 kmpl	yes	no
Maruti	Alpha 1.3	863000	Baleno	1248 cc	74 bhp @ 4000 RPM	190 Nm @ 2000 RPM	5	37 Ltr	27.39 kmpl	no	yes
Maruti	ZDI	824000	Dzire	1248 cc	74 bhp @ 4000 RPM	190 Nm @ 2000 RPM	5	42 Ltr	26.59 kmpl	no	yes
Hundai	SX 1.1	821000	Xcent	1120 cc	71 bhp @ 4000 RPM	180 Nm @ 1750 RPM	5	43 Ltr	24.4 kmpl	no	yes
Hundai	Sportz 1.4	816000	i20 Elite	1396 cc	89 bhp @ 4000 RPM	220 Nm @ 1500 RPM	6	45 Ltr	22.54 kmpl	no	yes
Hundai	SX 1.2	854000	i20 Active	1197 cc	82 bhp @ 6000 RPM	115 Nm @ 4500 RPM	5	45 Ltr	17.19 kmpl	yes	no
Honda	S Petrol	865000	Mobilio	1497 cc	110 bhp @ 6600 RPM	145 Nm @ 4600 RPM	5	42 Ltr	17.3 kmpl	yes	no
Honda	S Petrol	867000	City	1497 cc	117 bhp @ 6600 RPM	145 Nm @ 4600 RPM	5	40 Ltr	17.4 kmpl	yes	no
Honda	E Petrol	912000	BR-V	1497 cc	117 bhp @ 6600 RPM	145 Nm @ 4600 RPM	6	42 Ltr	15.4 kmpl	yes	no

Figure 12: Comparison with Chevrolet Tavera

Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Chevrolet	PS Petrol	409000	Beat	1199 cc	79 bhp @ 6200 RPM	108 Nm @ 4400 RPM	5	35 Ltr	18.6 kmpl	yes	no
Car in Comparison with Chevrolet Beat											
Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Nissan	XL	467000	Micra Active	1198 cc	67 bhp @ 5000 RPM	104 Nm @ 4000 RPM	5	41 Ltr	19.49 kmpl	yes	no

Figure 13: Comparison with Chevrolet Beat

Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Volkswagen	Trendline 1.2L	544000	Ameo	1198 cc	74 bhp @ 5400 RPM	110 Nm @ 3750 RPM	5	45 Ltr	17.83 kmpl	yes	no
Car in Comparison with Volkswagen Ameo											
Company Name	Version	Price	Model Name	Capacity	Horse Power	Torque	Gear Box	Fuel Efficiency in Ltr.	Milage in KMPL	Petrol	Diesel
Mahindra	DI	623000	Thar	2523 cc	63 bhp @ 3200 RPM	182.5 Nm @ 1500 RPM	5	45 Ltr	13 kmpl	no	yes
Mahindra	K4 D	600000	KUV 100	1198 cc	77 bhp @ 3750 RPM	150 Nm @ 1750 RPM	5	35 Ltr	25.32 kmpl	no	yes
Ford	Ambiente 1.5	615000	Figo	1498 cc	99 bhp @ 3750 RPM	215 Nm @ 1750 RPM	5	40 Ltr	25.83 kmpl	no	yes
Ford	Ti-VCT 1.2	601000	Aspire Trend	1196 cc	87 bhp @ 6300 RPM	112 Nm @ 4000 RPM	5	42 Ltr	18.26 kmpl	yes	no
Chevrolet	LS 1.2	637000	Sail	1199 cc	82 bhp @ 6000 RPM	108 Nm @ 5000 RPM	5	42 Ltr	18.2 kmpl	yes	no
Volkswagen	Trendline	551000	Polo	1198 cc	74 bhp @ 5400 RPM	110 Nm @ 3750 RPM	5	45 Ltr	16.47 kmpl	yes	no
Nissan	XL CVT	613000	Micra	1198 cc	75 bhp @ 6000 RPM	104 Nm @ 4000 RPM	5	41 Ltr	19.34 kmpl	yes	no
Tata	XT Petrol	638000	Bolt	1193 cc	89 bhp @ 5000 RPM	140 Nm @ 1500 RPM	5	44 Ltr	17.57 kmpl	yes	no

Figure 14: Comparison with Volkswagen Ameo



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Parameter	Proposed Approach	Existing Approach
1. Flexibility	XML is flexible in representing many different kinds of information from diverse sources.	SPARQL uses only RDF data.
2. Easy to Understand	XQuery is derived from an XML query language called Quilt, including XPath 1.0, XQL, XML-QL, SQL, and OQL.	All relationships are of a fixed size and data lives in a single graph, SPARQL does not require explicit joins to specify the relationship between differently structured data.
3. Code Reusability	The purpose of the making ontology in JAVA is to achieve the code reusability for software development.	Ontology developed using tools does not provide code reusability.
4. Augmentation	The ontology is enriched with synonyms and hyponyms of the detected attributes using WordNet.	Sentiments done without ontology are not enriched with synonyms and hyponyms.
5. Usability	The information in ontology is structured as trees of objects that relate to each other as relationships. Each object may have one or more terms that can be used to identify references to that particular object on the sentences extracted.	The information of making relations is difficult without using ontology.

Figure 15: Comparison of proposed approach and existing approach

VI. CONCLUSION AND FUTURE SCOPE

In our proposed approach, we have concluded that ontology used to analyses the tweets to increase augmentation and efficiency of sentiments which is obtained using naïve Bayesian algorithm is working properly. Ontology used is specification of shared conceptualization and it helps in finding hidden relationships and adds meaning to our knowledge. The main components of ontology based analysis system are tweets extraction from Twitter4J, Tweet Repository, keyword extraction, sense matching and ontology, Crawling automobile data, mapping ontology and crawler data and analysis.

Our approach has some limitations including negation handling, language problem and hashtag segmentation, For example, “#thisisnotgood”.

In future, work can be carried forward in developing a fully automated ontology. Also, hash tag, negation handling, etc can be worked upon.

REFERENCES

1. Efstratios Kontopoulos and Christos Berberidis, “Expert Systems with Applications”, www.elsevier.com/locate/eswa, 4065–4074, 2015.
2. K. Vithiya Ruba and D. Venkatesan, “Building a Custom Sentiment Analysis Tool based on an Ontology for Twitter Posts”, *Indian Journal of Science and Technology*, Vol 8(13), July 2015.
3. Ms. Swaminarayan Priya R. et al, “A Comprehensive study of Query Languages for Semantic Web and retrieval of data from University Ontology Using SPARQL”, *International Journal of Information and Computing Technology* ISSN: 0976 – 5999.
4. Ms. Swaminarayan Priya R. et al, “Knowledge Representation of Published Articles in Semantic Web using Upper Ontology”, *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 2, Issue 8, August 2012.
5. R. Baracho, G. Silva, and L. Ferreira, “Sentiment Analysis in Social Networks: a Study on Vehicles” *ONTOBRAS-MOST, CEUR Workshop Proceedings*, volume 938, page 132-143, 2012.
6. Gopinath Ganapathy and S. Sagayaraj, “Automatic Ontology Creation by Extracting Metadata from the Source code”, *Global Journal of Computer Science and Technology*, Vol.10, Issue 14, Ver.1.0, November 2010.
7. <http://protege.stanford.edu>
8. Bhaskar Kapoor and Savita Sharma, “A Comparative Study Ontology Building Tools for Semantic Web Applications”, *International journal of Web & Semantic Technology (IJWesT)* Vol.1, Num.3, July 2010.