OBOME - ONTOLOGY BASED OPINION MINING IN UBIPOL

Dilek Tapucu [1], Andrea Kő[2], Musab Husani[1], Ahmet Kocyigit[1], Yucel Saygın[1], Habin Lee[3]

[1]Sabancı University,Turkey [2]CUB, Hungary, [3]Brunel University, UK

[1]{dilektapucu, musabhusain, akocyigit, ysaygin, @sabanciuniv.edu},

[2]andrea.ko@uni-corvinus.hu, [3]Habin.Lee@brunel.ac.uk

Ontologies have a special role in the UBIPOL system, they help to structure the policy related context, provide conceptualization for policy domain and use in the opinion mining process. In this work we presented a system called Ontology Based Opinion Mining Engine (OBOME) for analyzing a domain-specific opinion corpus by first assisting the user with the creation of a domain ontology from the corpus. We determined the polarity of opinion on the various domain aspects. In the former step, the policy domain aspect has are identified (namely which policy category is represented by the concept). This identification is supported by the policy modelling ontology, which describe the most important policy – related classes and structure. Then the most informative documents from the corpus are extracted and asked the user to create a set of aspects and related keywords using these documents. In the latter step, we used the corpus specific ontology to model the domain and extracted aspect-polarity associations using grammatical dependencies between words. Later, summarized results are shown to the user to analyze and store. Finally, in an offline process policy modeling ontology is updated.

Keywords: Ontology, opinion mining, policy.

1 INTRODUCTION

UBIPOL system (Figure1) was designed to engage citizens in the policy-making process, which requires more informed and active citizens, who are familiar with their policy issues. They have to know more about the policy related opinions of other citizens in order to be active partner in policy making. UBIPOL system has to cope with the complexity of the specific environment of Public Administration.

Ontologies have a special role in this task, because they help to structure the policy related context, which can be applied amongst other on opinion mining process. It is also used to provide conceptualization for policy domain and they are applied to enhance opinion mining process. Before processing the text to determine its sentiment orientation, the policy domain aspect has to be identified (namely which policy category is represented by the concept). This identification is supported by the policy modelling ontology in Ubipol system, which describes the most important policy – related classes and structure. To identify those policy issues which require special attention, because of needs for improving is a key goal of policy maker. Ontology-based opinion mining is the component which will provide this support for them, through the continuous analyses of opinions and providing feedback for the decision makers and citizens as well. The main goal of this paper is to give an overview about ontologies used in Ubipol opinion mining process; present how to combine them with opinion mining to identify appropriate features of opinions and enhancing opinion classification.

Having, for an entity/product, the set of its associated aspects research in aspect based opinion mining mostly focus on extracting the set aspects from reviews, and then, for each aspect extract the set of its associated keywords. During this process, the key questions are: how the set of aspects can be obtained? How they are linguistically expressed? How they are related to each other? Which knowledge representation model can be used to better organize product features and to produce a comprehensive summary?

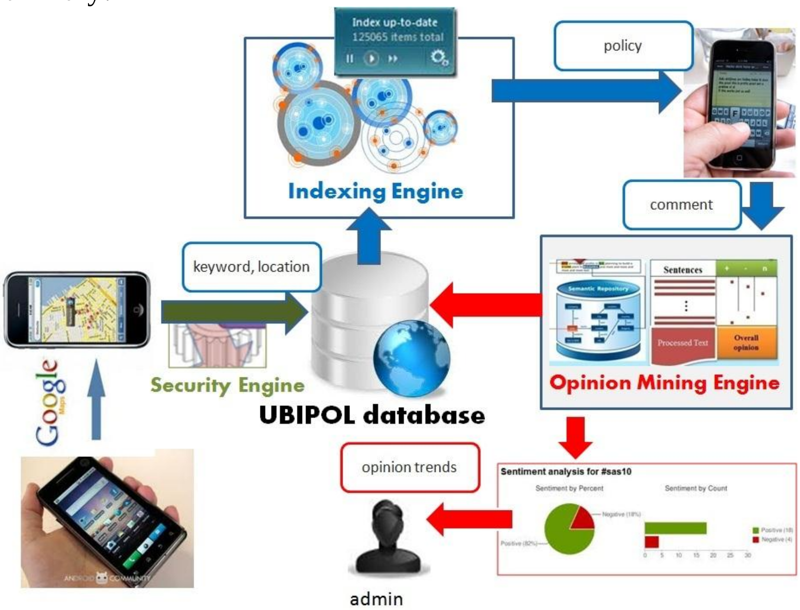

Figure1: UBIPOL Architecture

To answer these questions, we propose in this work to study the role of an ontology in aspect based opinion mining. Also Figure 2 gives Corpus Specific Domain Ontology relation with Policy Modelling Ontology. More precisely, our aim is to study how corpus based domain ontology can be used into policy modelling ontology:

- Step 1: people write comments and comments are collected under corpus
- Step 2: In a offline way aspect keyword extraction is occurred from that corpus
- Step 3: Corpus specific domain ontology is created. We show that an ontology is more suitable than a simple hierarchy where aspects are grouped using only the "is-a" relation (Carenini et al., 2005; BlairGoldensohn et al., 2008)
- Step 4: Corpus specific domain ontology is the core element for opinion mining module for the ontology based opinion mining engine. It is first used to catch aspect polarity association and second to show an aspect based scorecard summary of the review.

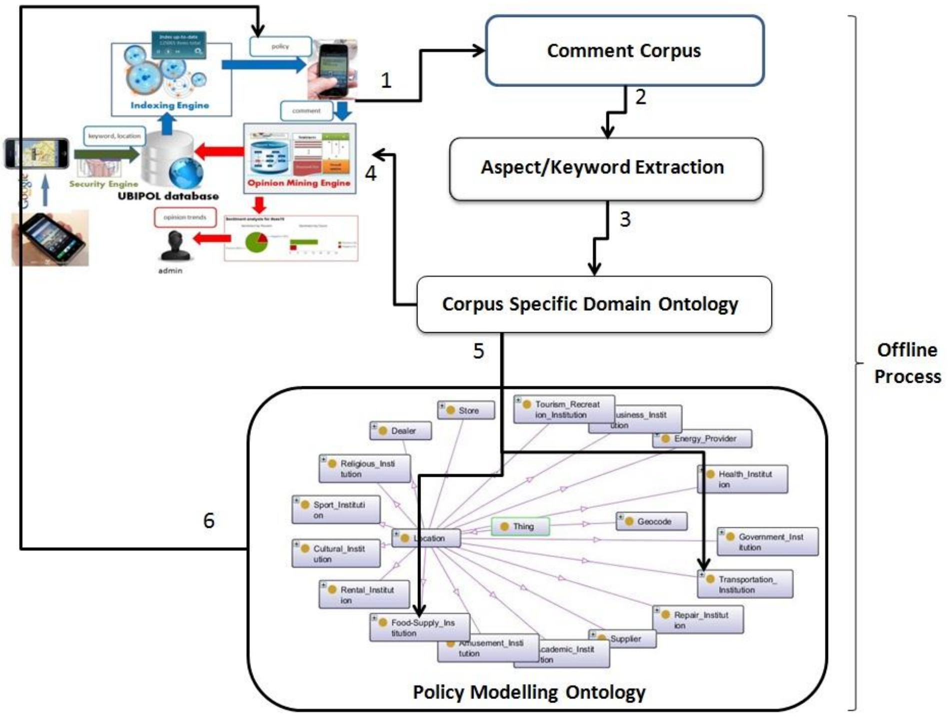

Figure2: Corpus Specific Domain Ontology relation with Policy Modelling Ontology

- Step 5: Corpus specific ontology charged to policy modelling ontology, by this way policy modelling ontology will be up to dated.

2 ONTOLOGIES ROLE IN OPINION MINING

Concept of ontology is used in many different senses and sometimes in a contradictory way. The following definition is the most cited one in the literature: "An ontology is an explicit specification of a conceptual model (conceptualisation)" (Gruber, 1993). This definition emphasizes the explicit specification, which make ontologies proper solutions for machine processing. One of the main goals of using ontology is to give a formal description of a specific domain, a task or an application. For that reason the use of ontological approach has been popular in the development of knowledge-based systems. Schreiber and his colleagues definition is based on the ontology building process in KACTUS project (Schreiber et. al. 1995): "Ontology provides the means for describing explicitly the conceptualisation behind the knowledge represented in a knowledge base." Another aspect, which is important during the discussion of ontologies is the shared specification: "Ontology is the term used to refer to the shared understanding of some domain of interest" (Uschold, Grüninger, 1996). Shared understanding has a key role from knowledge management view, because it can enhance knowledge transfer and sharing in the companies. These two features (shared understanding and explicit specification) are combined in the following definition: "An ontology is a formal explicit specification of a shared conceptualization" (Uschold, Grüninger, 2004).

The conceptual model or the conceptualisation is a kind of ideology in the wider sense; it reflects the mind of the specific domain. The ontology may appear in different forms but it has to contain the terms, terminology and semantics of the domain. It always is the appearance of collective specific domain interpretations that helps communication between the parties concerned. This common base enables the correct and successful information exchange that provides possibilities for reusability, public use and operation. In Ubipol approach, because of the domain dependency and context sensitive nature of opinion mining process we use the following definition: “Ontology is the term used to refer to the shared understanding of some domain of interest” (Uschold, Grüninger, 1996). There are diverse, known classifications of ontologies. Guarino distinguished the following categories (Guarino, 1995):

- Top-level ontology: it describes general notions that are domain; task and application independent like e.g. the space, time etc. It supports the combination and integration of the ontologies. One example is the ontology developed by Sowa (Sowa, 2000).
- Domain ontology: it contains the description of the vocabulary associated to a generic domain, according to specializing top-level ontology. Such a specific domain is e.g. the medicine, the geology, the farming, the finances that are treated irrespectively of tasks and problems, which can be correlated with the domain.
- Task ontology: it comprises the description of an activity or a task, according to the specification of the top-level ontology. Its subject is the problem solving.
- Application ontology: the most special ontology that corresponds to a specialization of the domain ontology or the task ontology for any concrete applications.

According to the categorization discussed-above, the most important dimensions used for the characterization of ontologies are the following:

- Formality: the degree of formality that is used to formulate the terminology catalogue and the definitions of words;
- Goal: for what purpose the user wants want to use the ontology;
- Domain: the nature of specific domain that is written in the ontology.

Categories of formality:

- Non-formal: explained in informal way and formulated in natural language;
- Structured informal: it is written in structured and constrained form of natural language, what increases the intelligibility and decreases the ambiguity (e.g. the text variant of the ‘Enterprise Ontology’);
- Semi-formal: description in an specification language (e.g. the Ontolingua version of the ‘Enterprise Ontology’);
- Rigorously formal, strict: determined in terms of formal semantics, theorems and proofs of such properties as consistency and completeness of theory (e.g. TOVE).

On the top level the many of the ontologies aim at a sort of reusability. According to another approach the ontology is an application dependent “intermediary language”. Based on the above, we can distinguish the next three categories of ontologies application:

- Communication: between humans - informal, unambiguous ontology can be used for these purposes.
- Cooperation: between systems - it means translation among different tools, paradigms, languages and software instruments. In this case the ontology is the basis of the data change.
- System design and analysis - the ontology can support the analysis and design of software systems with submitting a conceptual description.

One of the challenges of opinion mining is defining the objects of the study sentiments and subjectivity. Subjectivity was explained by linguists, (R. Quirk and Svartvik, 1985) defines “private state” as something that is not open to objective observation or verification. These states include emotions, opinions, and speculations, among others. In this section we review existing principles and methodologies that focus on the opinion mining process. Overall, two main families of work stand out: those that extract a simple set of aspects and others organize them into a hierarchy using taxonomies or ontologies.

2.1 Aspect Extraction

As a field of aspect extraction research, it is closely related to (or can be considered a part of) computational linguistics, natural language processing, and text mining. Identifying features and feature related keywords are important for aspect-based sentiment analysis since it is labour-intensive to manually generate domain-specific aspect keywords list. Given the abundance of input data (comments, reviews, etc.) researchers aimed to solve it by leveraging NLP techniques and classifiers. (Hatzivassiloglou et al., 1997) proposed to start with a small set of domain related keywords as a seed list. They focused solely on adjectives on the research. They scanned a comment for conjunctions and linkages between adjectives. They first found out that %77 of all conjunctions tie together adjectives of same orientation. The only exception was the linkage word but, but links two adjectives with different orientations. With this information in mind they analyzed a sentence for any linkages and identified candidate words. This paper is important since it is one of the first papers that deal with the problem of expanding aspect-keyword sets.

The other work in aspect based opinion mining is probably the one of Hu and Liu (2004) that applies association rule mining algorithm to discover product features (nouns and noun-phrases). (Hu et al., 2004) focused on extracting explicitly mentioned aspects from nouns and noun phrases. They use association mining to find frequently used nouns and noun phrases as candidates. They define frequent phrases as a phrase with a threshold of %1 frequency over entire dataset. One important concept they introduce is pure-support. Pure-support is determined by the number of sentences an aspect candidate appears as noun or a noun phrase, where there are no supersets of aspect is present. This way they can differentiate between life and battery life. Their work is important since they propose an efficient algorithm for nouns and noun phrases. (Poppescu et al. ,2005) developed a aspect-based opinion miner called OPINE. It extends Hu´s system in order to handle adjectives, adverbs, nouns and verbs. OPINE first identifies aspect keywords and then uses a Web-based point-wise mutual information system to decided if the candidate is valid or not. In order to do so, it generates discriminator phrases (eg is a scanner) and searches the web for given aspect and discriminator. If the results are above a threshold, it is counted as a valid aspect. Their claim is that they improved Hu´s system precision by %22 while losing %3 recall. (Qui et al., 2009) took a different approach. Rather than starting a seed list for aspects, the used a seed list of sentiment words. In addition, polarity analysis (positive, negative or neutral) of the document is done by assigning the dominant polarity of opinion words, it contains (usually adjectives), and regardless of polarities individually associated to each aspect. Then they defined sequences of dependencies that identify other sentiment words or aspects. One of their limitations is that they are constrained to adjective words. Since they haven’t used the same dataset with Hu and Poopescu, where is no way of comparing their performances. One other limitation is that, their method is only efficient in medium size corpora. This paper is important since it is one of the first attempts at inferring rules using dependencies in a sentence.

(Zhang et al., 2010) aims to extract noun aspect words from a given corpus. They aimed to improve on (Qui et al., 2009) by introducing two new patterns in aspect detection, namely part-of and no patterns. First pattern leverages linkages like of, is, on to determine new aspects part of a class. They use two metrics, aspect frequency and aspect relevance to asses’ candidates. In order to find aspect relevance they use a web page ranking algorithm to rank candidates. Since they did not use the same dataset as (Qui et al., 2009), there is no way to compare them.

(Zhuang et al., 2010) proposed a set of template rules to generate aspect-opinion pairs. Their template focuses on adjectives and noun phrases. They first find candidates for feature and opinion words separately. In this candidate evaluation process, they discard infrequent candidates from their set. Afterwards, they apply their template to generate opinion-aspect pairs. As they mentioned in their paper; one major drawback is that people tend to use different words to express their sentiments, and this leads to some valid candidates to be discarded as they were found frequently in the data set.

(Zhang et al., 2011) focused on noun features that express opinions. They proposed a scoring algorithm where the effect of an opinion word decreases as the distance between the opinion and the aspect word increases. They also defined simple rules for negation, but-clauses and increasing or decreasing rules. The last one is intuitive where they identified polarity words that diminish the sentiment strength of a polarity word. Words like "decease", "diminish", "prevent" and whenever one of these words is seen, the strength of the sentiment is changed.

Natural Language Processing (NLP), also known as computational linguistics, is a field of computer science that studies interactions of human languages with computers. The main goal of NLP is to enable effective human-machine communication, which could be either as spoken or written form. Several approaches (Lu et al. 2009; Popescu and Etzioni 2005, Hu and Liu 2004) attempted to identify features in the opinion text with the use of NLP-based techniques.

Sentiment prediction is another active research area. Lexicon-based sentiment prediction is very popular in the context of opinion summarization (Hu and Liu 2004; Zhuang et al. 2006). This technique generally relies on a sentiment word dictionary. The lexicon typically contains a list of positive and negative words that are used to match words in the opinion text. For identifying the opinions about aspects and their orientation, (Hu and Liu 2004) proposed a simple yet effective method based on WordNet. They start with a set of about 30 seed adjectives for each pre-defined orientation (positive and negative). Then they use the similarity and antonym relations defined in WordNet for assigning positive or negative orientations to a large set of adjectives. Thus, the orientation of an opinion about a feature was decided by the orientation of the adjective around it. (Zhuang et al. 2006) used dependency relationships to identify opinions associated with feature words.

The most common type of opinion summarization technique is aspect-based opinion summarization. Aspect-based summarization involves generating opinion summaries around a set of aspects or topics (also known as features). These aspects are usually arbitrary topics that are considered important in the text being summarized. In general, aspect-based summarization is made up of three distinct steps - aspect/feature identification, sentiment prediction, and summary generation. Some approaches, however, integrate some of the three steps into a single model.

2.2 Towards an ontology based opinion mining

The early opinion mining projects focused on certain topic oriented opinion set, where taxonomies and domain ontologies were not taken into consideration. Next feature taxonomies were applied in opinion mining process. The following works use feature taxonomies, but they have different approaches: they do not look for a "basic list" of aspects but rather a list hierarchically organized through the use of taxonomies. We recall that a taxonomy is a list of terms organized hierarchically through specialization relationship type "is a sort of". Carenini et al. (2005) use predefined taxonomies and semantic similarity measures to automatically extract classic features of a product and calculate how close to predefined concepts in the taxonomy they are. This is reviewed by the user in order to insert missing concepts in the right place while avoiding duplication.

Hu and Liu (2004) divided the mining process for three steps:

- mining product features which occurred in customers' opinions (data mining and natural language processing techniques applied in this task);
- identifying opinion sentences in each review and determine whether each opinion sentence is positive or negative;
- aggregating the results.

Liu (Liu, 2008) prepared a literature review about opinion mining. He applied an object-oriented approach in his work, which can be mapped to the ontological approach. He defined an object O (which is commented) as an entity, which can be a product, topic, person, event, or organization. It is associated with a pair, O: (T, A), where T is a hierarchy or taxonomy of components (or parts) and

sub-components of O, and A is a set of attributes of O. Each component has its own set of sub-components and attributes. He applies semantic-based approach, where in the taxonomy; the root is the object itself and each non-root node is a component or subcomponent of the object. Links between objects are part-of relationships. Each node has a set of attributes. An opinion can be expressed on any node and any attribute of the node. Semantic orientation of each opinion word is defined to indicate the direction that the word deviates from the norm for its semantic group. Each sentence in a review were classified (negative or positive), product features were mined from reviews on which the reviewers have expressed their opinions. Liu (Liu, 2008) defined semantic orientation of an opinion as a feature states whether the opinion is positive, negative or neutral. In feature-based opinion mining, first object features are identified, next opinion orientations are determined followed by grouping synonyms.
The aspect extraction phase is guided by a domain ontology, built manually (Zhao and Li, 2009), or semi-automatically (Feiguina, 2006; Cheng and Xu, 2008), which is then enriched by an automatic process of extraction / clustering of terms which corresponds to new aspect identification. They mainly use ontology to define a domain and some aspects. Then they try to extend knowledge on aspect related keywords. They also use SentiWordNet and a custom scoring algorithm to translate polarity information obtained into a floating point number. Then they take negation and conjunction words into account to calculate the polarity on a feature. Since they worked on a set of 120 reviews taken from IMDB and did not present precision or recall but simple accuracy, we cannot pass judgement on their accuracy. This paper´s main importance lies in taking domain specific information into account while doing feature based analysis.
To extract terms, Feiguina (2006) uses pattern extraction coupled to a terminology extractor trained over a set of features related to a product and identified manually in a few reviews. Same features are grouped together using semantic similarity measures. The system OMINE (Cheng and Xu, 2008) proposes a mechanism for ontology enrichment using a domain glossary which includes specific terms such as words of jargon, abbreviations and acronyms. Taking advantage of ontological techniques, OMINE achieved 10% higher recall with the same level precision on the topic extraction task, compared with the previous work. OMINE has two modules: 1) ontology-based topic extraction and 2) fine-grained polarity analysis. The first one is responsible for a generation of a uniform ontology on top of existing domain specific ones and complete lexicons of the generated ontology to identify concept-related topics. The second one acquires sentiment knowledge and generates subjective patterns to train a statistical polarity classifier. Their ontology-based topic extraction function has three subprocess:

- Offline Ontology Building
- Ontology Lexicalization
- IE-based Topic Extraction.

In offline ontology building the existing ontologies are merged to get a uniform ontology. The merge is based on the similarity of pair of concepts from two ontologies, which is determined by the numbers of same word stems used in two concepts. If there are no shared word stems, the concepts are not connected. They distinguish general and specific ontology between two ontologies in ontology building. Specific ontology is the one whose root is connected to any concept of the other ontology (if both of their roots are connected to each other, the one with less concepts is called specific ontology). The other one is called general ontology. The constraints for selection connecting concepts from a specific ontology are the next: 1) first the concept with the maximum similarity is selected; 2) if more than one concept have the same maximum similarity, the one with a maximum depth is selected and 3) if more than one concept have the same maximum depth, the first one in the current list is selected. Ontology lexicalization is responsible for managing synonyms, jargon, abbreviation, and acronyms. Head-ARGUMENT Matching or HAM was the algorithm used to retrieve related terms in a glossary for the input concept. HAM is similar to ‘headmatching’- heuristic (Cimiano, et al., 2004). IE-based

Topic Extraction identifies topics with domain concepts. Rule-based Name Entity Recognition (NER) and IE engine SProUT (Drozdzynski, et al., 2004) was applied in this process.

3 OBOME - ONTOLOGY BASED OPINION MINING IN UBIPOL SYSTEM

In this paper, we present a system called OBOME (Ontology-Based Opinion Mining Engine) for analyzing a domain-specific opinion corpus by first assisting the user with the creation of a domain ontology from the corpus and then determining the polarity of opinion on the various domain aspects. In the former step, we computationally extract the most informative documents from the corpus and ask the user to create a set of aspects and related keywords using these documents. In the latter step, we use the ontology to model the domain and extract aspect-polarity associations using grammatical dependencies between words. Finally, summarized results are shown to the user to analyze and store.

The main contributions of our tool are to provide a domain- transparent sentiment analysis method with minimal human effort, and to produce a novel way of associating sentiment polarities by analyzing dependencies between words and transferring polarity from polarity keywords to aspect keywords in a sentence.

OBOME is designed as a self-contained web application that can be deployed to any Java-based web server. This system provides an interactive interface for users to import a domain-specific opinion corpus into the application as illustrated in Fig. 1. Once imported, the corpus is analyzed to obtain the most informative corpus documents from which the user extracts domain aspects and related key-words. The ontology thus created is then utilized by the opinion mining engine to process the corpus. Finally, the results are displayed to the user for further analysis. A publicly accessible version of this system is available online at *http://ferrari.sabanciuniv.edu/obome*. In the following subsections, we will discuss the major components of the system.

Our tool has two main modules: one is for the aspect extraction (Figure3-A) and the other is for the opinion mining (Figure3-B). The inputs to the system are all comment corpus. The output is the summary of the reviews. Our contribution is to provide a domain-independent approach in a user-friendly fashion. Thus, firstly, corpus-based aspects and aspect-related keyword sets are extracted as described in aspect/keyword identification module to generate corpus specific domain ontology. Then a polarity ontology and NLP API is used to show aspect polarity association and to generate scorecard. In the following subsections, we will discuss the major components of the system.

3.1 Aspect/keyword Identification Module

This module deals with the problem of aspect and keyword extraction by corpus summarization and user annotation. We approximate aspect keywords with corpus nouns and apply a variation of the Greedy Set Cover algorithm that we developed called Eagerly Greedy Set Cover algorithm to find the minimum set of documents that cover most of the nouns in the corpus. Initially, each document is assigned a utility score based on the cumulative weight of the known aspect words (nouns) to which it can provide exclusive coverage. These documents are then displayed sequentially to the user for collecting new aspects and associated keywords. As each document is displayed, our user interface emphasizes document nouns so they can be quickly identified. The user can also interact with these emphasized nouns to mark them as aspects or keywords of an aspect. By ignoring lower utility documents as decided by a threshold which the user can set, we can significantly reduce the number of documents requiring annotation.

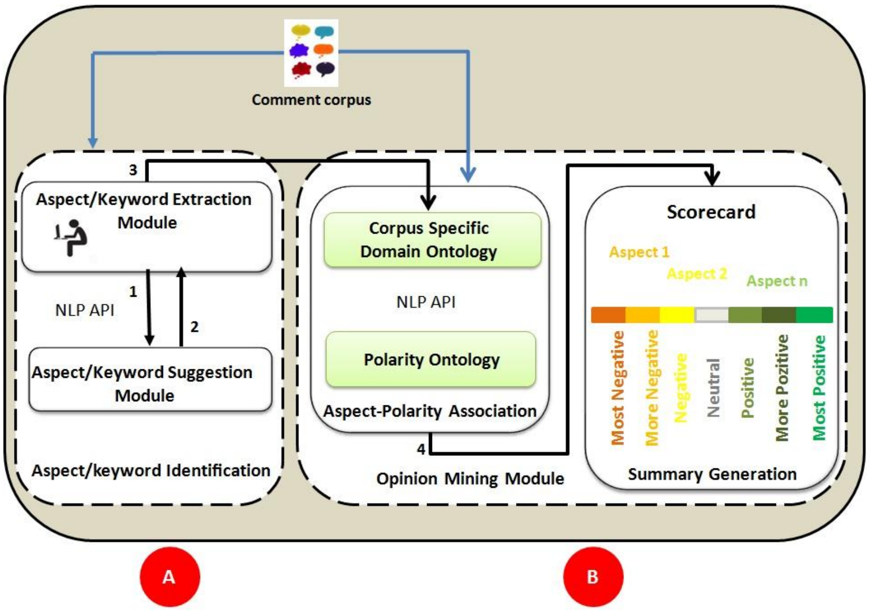

Figure3: OBOME (Ontology Based Opinion Mining Engine)

In sentiment analysis, the context of a word affects its meaning, specifically whether it has a positive or negative or even neutral orientation. Before processing for orientation, the concept that the word represents has to be known, which can be defined with in a concept ontology. In order to create a robust ontology that is useful for sentiment analysis, we have to cover a large set of comment corpus. In this process, a person (oracle) possessing domain knowledge defines all the domain concepts (aspects) and corresponding keywords from the corpus, which becomes increasingly expensive as the size of the corpus increases. Aspect extraction is a broad topic for which several methods have been used. Pre-processing the documents to get term frequencies and running a clustering algorithm are previously discovered methods for actively learning an ontology. In all of these methods, opinions were assumed to have been expressed on one and only one aspect. However, since the user is not strictly forced to cover only one aspect, each opinion document can often include several aspects. With this in mind, we propose a methodology to decrease the work load of the oracle by finding the most informative multi-aspect opinion documents from the corpus.

3.2 Opinion Mining Module

The objective of this part is to use ontology based approach for representing corpus-specific knowledge and to present aspect-based score value to summarize results in a scorecard structure. Our contribution is to provide a domain-independent approach in a user-friendly fashion. Thus, firstly, corpus-based aspects and aspect-related keyword sets are extracted as described in Sec. 4.1 to create a corpus-specific ontology. Then a polarity ontology is created from SentiWordNet. Finally, polarity-placement algorithm is used to calculate score values for each aspect. The idea of the algorithm is to get initial polarity value from the polarity ontology for any opinion word in a given comment and to transfer polarity value from polarity keywords to aspect- related keywords by using the Stanford NLP API. The reason for using the Stanford NLP API is to generate dependency tree graphs for a given sentence. After polarities are transferred on a correct token, aspect-based score value is calculated for each comment. A detailed aspect-based summary is displayed as illustrated in Fig 3-B.

3.3 Aspect Polarity Association Part

This module concentrates more on the opinions that are expressed within a review. It concerns the mining of subjective statements from texts, the identification of opinions, the estimation of opinion orientation and the extraction of arguments that relate to opinions. The identification of the opinion polarities and strength inside a text has plenty of applications and challenges.

In this module Polarity ontology and NLP API are used to discover aspect polarity association. Our approach (like others [Lu et al. 2009; Popescu and Etzioni 2005; Hu and Liu 2004b; 2004a]) attempt to identify aspects in the opinion text with the help of NLP-based techniques. Part-of-speech (POS) tagging and syntax tree parsing are very common starting points for keyword discovery. For example, as aspects/features are usually noun phrases, even basic POS tagging allow us to find candidate aspects. The annotated opinion texts are then further analyzed using data/text mining techniques. One potential problem is the practicality of these approaches is the speed of parsing or tagging it is still not 'efficient' enough for large scale processing. Also, such NLP-based techniques may not be sufficient in discovering all the aspects. This is because aspects and aspect related keywords are not always nouns, and often times they are not explicitly specified in the text. For example, the sentence, 'The hospital is small', implicitly mentions the 'size' feature, but there is no mention of the word 'size' in the sentence. This may require some domain knowledge or help from some ontological word dictionary.

Polarity ontology is created to define opinion words. This ontology contains opinion words, which are words that people use to express a positive or negative opinion. Observing that people often express their opinions of product aspects using opinion words that are located around the aspect in the sentence, we can extract opinion words from the polarity ontology using all the remaining frequent features (after pruning). In this work Sentiwordnet (www.sentiwordnet.com) that is used to create polarity ontology is a lexical resource for opinion mining. Also it is known as a lexical resource for opinion mining. The edges of the triangle represent one of the three classifications (positive, negative, and objective). A term can be located in this space as a point, representing the extent to which it belongs to each of the classifications. The score range from 0.0 to 1.0, giving a graded evaluation of opinion-related properties of the terms.

Polarity ontology in OBOME has four subclasses adj, adv, noun and verb. The reason behind this structure to take into account, different usages of a word and different polarities it can get. Under each class we have keywords as individuals associated with their respective classes. Each individual has 4 attribute: pos value, obj value, neg value and value. The first three are gathered from average of SentiWordNet results. We further classified each word as positive or negative by applying a threshold and obtained a single integer for polarity value. We take this integer value as basis on our analysis.

4 Conclusion and Future Work

In this paper we discussed ontology specifications in Ubipol, mainly from opinion mining aspects. Role of ontologies in Ubipol system were diverse. The most important application is the opinion mining related one, where we applied them for providing policy categorization and tags for the opinions. We utilized them in policy modelling for structuring places used in POIs (POI ontology) and for policy issue conceptualization (policy modelling ontology). The development process followed mainly the updated On-To-Knowledge methodology. In policy modelling ontology we completed the methodology requirement phase with text mining models to extract the policy domain related concepts from opinions, which can be further structured in the ontology development environment. In this work, we focused to study the problem of aspect-based opinion summarization of user reviews of entities. The results are then presented in a user-friendly visualization. Future work involves comparing the accuracy and the computational efficiency of our aspect extraction method with that of the other approaches. In addition, larger scale quantitative evaluation of our opinion mining method will be conducted.

5 REFERENCES

Goldensohn, B., Hannan, K., McDonald, R., Neylon T., Reis G., Reynar, J. 2008. 'Building a Sentiment Summarizer for Local Service Reviews'. WWW2008 Workshop: Natural Language Processing Challenges in the Information Explosion Era (NLPIX 2008).

Carenini, Giuseppe, Raymond T. Ng, and Ed Zwart. 2005. ' Extracting Knowledge from Evaluative Text'. In Proceedings of the 3rd international conference on Knowledge captur.

Fensel, D., van Harmelen, F., & Davies, J. 2003. 'Towards the Semantic Web – Ontology driven knowledge management'. West Sussex, England: John Wiley & Sons Ltd.

Xiwen, C., Xu, F. 2008. 'Fine-grained Opinion Topic and Polarity Identification. ' In Proceedings of the Sixth International Language Resources and Evaluation (LREC' 08), Marrakech, Morocco.

Fernandez, M., Gomez-Perez, A., & Juristo, N. 1997. 'METHONTOLOGY: From Ontological Art Towards Ontological Engineering'. AAAI-97 Spring Symposium on Ontological Engineering . Stanford University.

Fernández-López, M. 1999. 'Overview Of Methodologies For Building Ontologies'. Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5), Stockholm, Sweden, August 2, 1999.

Fernández-López, M., Gómez-Pérez, A., & Rojas, M. D. 2000. 'Ontology's Crossed Life Cycle'. Lectures Notes in Artificial Intelligence N°1937.

Gruber, T.R. 1993 'A Translation Approach to Portable Ontology Specifications'. In Knowledge Acquisition, 5(2).

Guarino N. 1995. 'Formal Ontology, Conceptual Analysis and Knowledge Representation: The Role of Formal Ontology in the Information Technology'. International Journal of Human-Computer Studies, Vol. 43, Nos. 5/6, pp. 625-640.

Hatzivassiloglou V. and McKeown K. 1997. 'Predicting the semantic orientation of adjectives'. In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, EACL '97, pages 174–181, Stroudsburg, PA, USA.

Hu M. and Liu B. 2004. 'Mining and summarizing customer reviews'. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '04, pages 168–177, New York, NY, USA.

Liu, B. 2006. 'Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data' (Data-Centric Systems and Applications). Springer-Verlag New York, Inc., Secaucus, NJ, USA

Liu, B. 2008. 'Opinion mining', http://www.cs.uic.edu/~liub/FBS/opinion-mining.pdf

Ontoweb. 2002. ' A survey on methodologies for development, maintaining, evaluating and reengineering ontologies'. OntoWeb Technical Report .

Popescu A., Etzioni O. 2005. 'Extracting product features and opinions from reviews'. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05, pages 339–346, Stroudsburg, PA, USA.

Scriptum Zrt. 2005. 'Ontológiaépítési módszertanok összehasonlítása'. Jelentés, MEO Projekt

Sowa, J.F. 2000. 'Knowledge Representation: Logical, Philosophical, and Computational Foundations', Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

Qiu,G., Liu B., Bu J., and Chen C. 2009. 'Expanding domain sentiment lexicon through double propagation'. In International Joint Conference on Artificial Intelligence (IJCAI-09), 2009.

Uschold, M. & Grüninger, M. 1996. 'Ontologies: Principles, methods and applications. Knowledge Sharing and Review', 11(2):93–155.

Uschold, M., King, M., Moralee, S., & Zorgios, Y. 1998. 'The Enterprise Ontology'. The Knowledge Engineering Review , Vol. 13:1, 31-89.

Uschold, M. & Grüninger, M. 2004. 'Ontologies and semantics for seamless connectivity'. SIGMOD Record, 33(4):58–64.Yergeau, F., Bray, T., Paoli, J., Sperberg-McQueen, J.M., Maler, E. (2004) 'Extensible Markup Language (XML) 1.0 (Third Edition)'. W3C Recommendation, 2004.

Zhang L, Liu B., Lim S.H., and O'Brien-Strain E. 2010. 'Extracting and ranking product features in opinion documents'. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 1462–1470. Association for Computational Linguistics.

Zhao L. and Li C., 2009. 'Ontology based opinion mining for movie reviews'. Knowledge Science, Engineering and Management, pages 204–214.

Zhang L. and Liu B., 2011. 'Identifying noun product features that imply opinions'. In Proceedings of the 49th Annual Meeting of the Association for Computational Lin- guistics: Human Language Technologies: short papers-Volume 2, pages 575–580. Association for Computational Linguistics.