International Journal on Advanced Science Engineering Information Technology
Vol.7 (2017) No. 3
ISSN: 2088-5334

Representing Semantics of Text by Acquiring its Canonical Form

Mohammed Ahmed Taiye[#], Siti Sakira Kamaruddin[#], Farzana Kabir Ahmad[#]

[#]*School of Computing, College of Arts and Science, Universiti Utara Malaysia,*
06010 UUM Sintok, Sintok, Kedah, Malaysia
E-mail: tfeatslekan@gmail.com, {sakira, farzana58}@uum.edu.my

***Abstract*— Canonical form is a notion stating that related idea should have the same meaning representation. It is a notion that greatly simplifies the task by dealing with a single meaning representation for a wide range of expression. The issue in text representation is to generate a formal approach of capturing meaning or semantics in sentences. This issue includes heterogeneity and inconsistency in the text. Polysemous, synonymous, morphemes and homonymous word pose serious drawbacks when trying to capture senses in sentences. This calls for a need to capture and represent senses in order to resolve vagueness and improve understanding of senses in documents for knowledge creation purposes. We introduce a simple and straightforward method to capture the canonical form of sentences. The proposed method first identifies the canonical forms using the Word Sense Disambiguation (WSD) technique and later applies the First Order Predicate Logic (FOPL) scheme to represent the identified canonical forms. We adopted two algorithms in WSD, which are Lesk and Selectional Preference Restriction. These algorithms concentrate mainly on disambiguating senses in words, phrases, and sentences. In addition, we adopted the First Order Predicate Logic scheme to analyse argument predicate in sentences, employing the consequence logic theorem to test for satisfiability, validity, and completeness of information in sentences.**

***Keywords*— semantic; natural language processing; canonical form; first order predicate logic; word sense disambiguation**

I. INTRODUCTION

The increasing amount of data in the form of unstructured text needs to be well processed and interpreted in order to identify useful knowledge by considering its semantics. Semantics refers to the study of meaning. However, representing the semantics of text is a non-trivial task that poses many challenges. Before representing semantics, the text needed to be pre-processed and analysed [1], [2]. Pre-processing text data from documents is a crucial step in overcoming text syntactic and semantic analysis issues. Text pre-processing helps in identifying characters and words as fundamental units. These units are processed in stages from analysis to tagging text components like the morphological analysis and part-of-speech tagging in Natural Language Processing (NLP) systems [3].

NLP employs some pre-processing techniques for text input such as; Part of speech tagging, Lemmatization, Chunking, and Parsing. These techniques give meaning to texts and how they are formed [4]. Researchers compared NLP techniques with Non-NLP techniques and have concluded that NLP based techniques are better since methods based on language modelling are computationally demanding [1]. However, computational identification of relationships between words still needs to be investigated upon by mapping text inputs to a useful representation [5].

It is evident from previous works that, the application of NLP on enormous text document is incomplete and inefficient without identifying meaningful information (semantics) from the documents [6]-[9], [30]-[31]. This paper focuses on the use of NLP techniques to analyse and represent text semantics due to its vast applicability on resolving text issues [10]. Our research concentrates majorly on the semantic analysis of textual data.

A. Semantic Analysis

Semantic Analysis is a step in NLP that helps in eliciting and representing knowledge using language. Language is a very generic representation, where words are used to describe almost anything. The heterogeneity and inconsistency of language makes it difficult to detect semantics [11].

A number of practical semantic analysis systems follow compositional approach by relying on semantic grammars and extracting information from a particular source. This approach is based on the principle of compositionality; “meaning of a sentence can be composed of the meaning of its parts” i.e. Part-of-Speech of a sentence [5], [12], [13].

Statistic-based techniques such as Latent Semantic Analysis (LSA) and Probabilistic Latent Semantic Analysis (PLSA) analyse text based on contextual usage by statistical computation instead of analysing text based on its semantics features. Text semantic features help to resolve co-reference, decompound words in sentences by identifying metonymy, synonyms and polysemous words [14], [15]. The semantics of linguistic expressions can be captured in formal structures by identifying its canonical forms [5].

B. Canonical Form

Canonical form simplifies the task by dealing with a single meaning representation for a wide range of expression [5]. Reference [13] argued that there are three factors that define a canonical form in sentences; i.e. syntactic structures, semantic role and the frequency of usage. These three factors are experimentally tested in sentences to understand the level of which sentences are interpreted and understood without difficulties. However, it is important to note that canonical form lays more emphasis on how knowledge can be modelled in a concise manner by precisely identifying senses in expression.

Existing systems like the GALLEN [17] and SNOMED RT. [1] employs canonical form representation techniques to express meaning without special intervention of knowledge interpretation in medical informatics [16] & [17]. For example, in [16], a method was developed to assign responsibilities to regulate expressions. Expression collaborated will be preserved at a coherence level that is a prerequisite to distribute basic process of modelling an individual concept, from the expression in the original corpus to the canonical form in the GALLEN project. However, GALLEN and SNOMED RT. are model-driven application and are limited to medical terminologies.

Semantic representation plays a similar role with syntactic representation in the text [4]. Reference [10] proposed a framework for semantic inference at a lexical-syntactic level to solve the problem of understanding and learning the variability of natural language semantic expressions. The framework made use of an inference module to improve unsupervised acquisition of entailment rules (formal conclusions that can be drawn from NLP semantic expressions) through canonical forms, which serve as an active verb form with direct modifier [10].

Previous studies made more emphasis on text information modelling that involves building and managing the conceptual unit of discourse in the text [4], [10]. However, we are interested in identifying text canonical form not just to model textual information, but also to resolve issues related to text ambiguity and draw a formal expression on text validity, satisfiability and information completeness in documents [4]. This is the reason why we advocate the use of FOPL to represent the identified canonical forms.

Canonical form expresses a systematic relationship between word senses and the grammatical constructs found in text documents [18]. In the case of text variants sharing similar name entity, such as Mr. Jordan and Robert Jordan due to the standard English naming convention. Each name is categorized as an entity type with a canonical name as its identifier. The canonical name is the fullest least ambiguous label that can be referred to in text document [19]. To resolve ambiguity in text canonical name will be followed by its entity type and the variants linked to it as seen below.

Mr. James (PERSON) → James Bond
James Bond (PERSON) → Mr. James

II. Material and Method

To leverage more of the canonical form text identification in the document, the Word Sense Disambiguation is employed in the study. Two Word Sense Disambiguation algorithms are combined, namely Lesk and Selectional preference to efficiently reduce multiple text meaning into a single common sense and implicitly equate text input into action, agent and object forms. A detailed reason for choosing this approach is discussed in the next section.

A. Word Sense Disambiguation

Word Sense Disambiguation automatically eliminates flawed representations that resulted from incorrect meaning. It also determines the different meaning of words tagging in a text with its appropriate senses [20]. Effective Word Sense Disambiguation can improve computer application performance in text summarization, natural language understanding, machine translation and information extraction [5].

Our work focuses on using WSD to capture the senses in the text. We adopted two algorithms for WSD, which are the Lesk and Selectional preference algorithms. Lesk algorithm was leveraged because of its accuracy in short sentences and in some news stories, which reads about 50-60% accuracy. [26], gave two hypothesis on word senses. This hypothesis states that "The intended sense of the target word in a given context is semantically related to other word senses in the context. Also, semantically related words have a greater number of overlaps of their dictionary definitions". [27] adapted Lesk algorithm to use rich knowledge source from WordNet as shown in Fig. 1. The basic idea for applying Lesk algorithm is to show the relationship between generic words and a specific instance of this given words (hypernym and hyponym). As can be seen in Fig. 1 a generic word such as food is a hypernym, and a more specific kind of food items such as fruits, grains, and vegetables are hyponym. These attributes exhibit a broader idea of semantic and superordinate fields in words [29]. Lesk algorithm combines and performs these functions on all part of speech primarily focused on identifying words that are semantically related, while selectional preference restriction is used due to its accuracy and simplicity in brown corpus [4]. They were both used to identify senses in morphemes, synonymous and polysemous words which most of the time taken from a Knowledge Base (KB) or (dictionary).

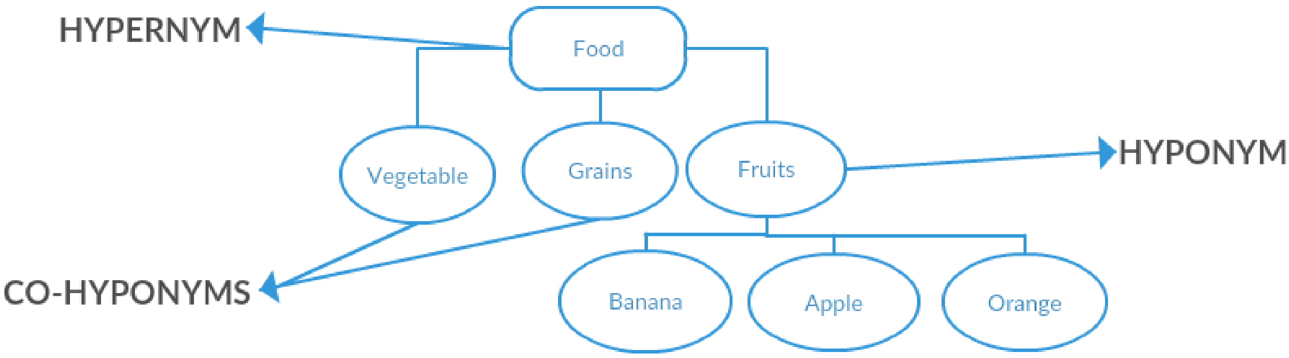

Fig. 1 Adapting lesk to wordnet

Based on this, an algorithm was developed to analyse sentence structure in order to produce a well-formed disambiguated words in sentences. The Lesk Algorithm is as follows:

Algorithm 1 Lesk Algorithm

Step 1: POS tagging of words in a sentence.

Step 2: Identify polysemous and synonymous word using the KBS (Dictionary) to derive senses

Step 3: Polysemy words are then disambiguated to evaluate senses

Step 4: FOR each polysemous and synonymous word, senses are maximized to >=3

Step 5: Polysemous words are then selected and compared with other polysemous words in each sentence.

Step 6: IF polysemous word is identified as the same word.

Step 7: THEN compare with similar words in other sentences for words to the proper disambiguated.

The Selectional Preference Algorithm is as follows:

Algorithm 2 Selectional Preference Algorithm

Step 1: POS tagging of words in sentence

Step 2: Check for polysemous word

Step 3: Check for word senses using KB

Step 4: Classify words either to be an object or action

Step 5: Compare polysemous words with other polysemous words in each sentence

The idea behind the proposed algorithms is to represent text in a standard form through which meaning can be easily interpreted. A typical example is shown in the following sentences;

"Does Maharani have a vegetarian dish?"

"Do they have vegetarian food at Maharani?"

"Are vegetarian dishes served at Maharani?"

"Does Maharani serve vegetarian fare?"

Fig. 2 illustrates how words in the sample sentences are linked based on their senses. The words are denoted as rectangles, and the senses are denoted with oval symbols. There are 3 words, i.e. Dish(*word*^{1}), Fare(*word*^{2}) and Food(*word*^{3}) and their respective senses. As can be seen in Fig. 2, ***[word^1(sense^3)=word^2(sense^2)=word^3(sense^1)]*** where ***word^1*** *= Dish,* ***word^2*** *= Fare and* ***word^3*** *= Food.*

Fig. 2 also shows how three words (i.e. Dish, Fare and Food) share at least one similar sense. The Lesk Algorithm will identify the shared senses among words and report that word 1, 2 & 3 have similar meaning because they share at least one similar sense. Most of the times, reference, and comparison of words are made through the dictionary. The Selectional preference algorithm follows this structure ***[Word → (POS→ Object)]***. For instance;

[Maharani → (Noun→ subject)],
[Vegetarian → (Noun→ Subject)],
[Served → (Adj→ predicate)].

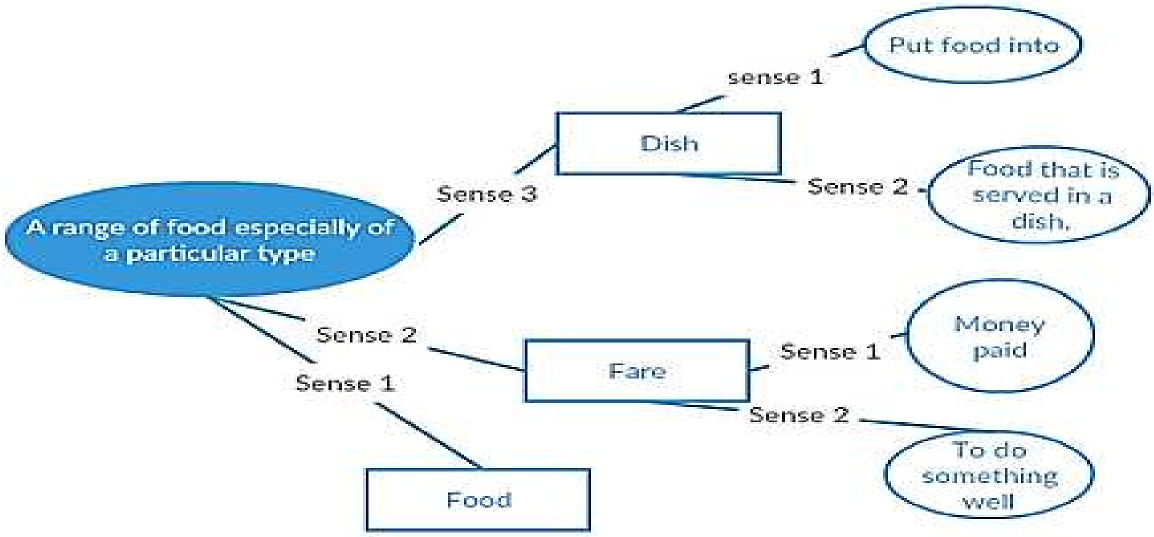

Fig. 2 Conceptual illustration of word linkage with senses

From the selected words, POS tagging of all similar words are taken into consideration. It checks for the syntactic structure of the words after it has been parsed. The syntactic structure of words using the POS tagging shows the role each word plays in a sentence (argument and predicate). Additionally, argument and predicate challenges are always non-trivial in text representation. This is why we employ FOPL to resolve this issue. The first-order logic reasoning is represented, as symbols where in our case each identified canonical form will be represented as subject and predicate.

B. First Order Predicate Logic

According to [5], FOPL is a well-defined computational and understandable knowledge representation scheme that satisfies the rules of grammatical representation in language. FOPL symbolizes reasoning of a statement or sentence that is broken down into subject and predicate. Subject's properties are modified and defined by predicates.

Reference [21] defines the link between vector spaces through lexical mapping of predicate symbols (lexical semantics) with predicate logical forms. The resulting approach was able to solve many difficult textual entailment problems that require handling complex combinations of semantic phenomena. Textual entailment is generically used to capture a broad range of inferences that are relevant from multiple applications. Another interesting research introduced the use of probabilistic logic with expressivity and automated inference provided by the logical representation to capture semantics in sentences. This research demonstrated a state-of-the-art performance in identifying semantics [11].

Inspired by [11] we employ FOPL approaches towards creating a semantic analysis of a logical proposition that enhances text inference identification in documents. It does this by syntactically analysing expressions in documents. This expression is later translated into an intermediate logical language that is human and machine-readable. FOPL leverages sentence to capture text semantics by combining the relationship between word satisfiability and tautology to justify completeness of information about an idea or discourse in corpus [5]. The logical argument must hold in all circumstances for knowledge to be created from documents.

However, the expression is said to be valid if it is true in all interpretations. To show that an expression *G* is valid we write ***G*** as [10]:

- **Satisfiability:** is said to be defined as follows: For a first order predicate logic sentence ***G*** over ***S*** is satisfiable. If there exist on ***S*** structure ***F*** such that $F \vDash \boldsymbol{G}$. The given expression $\forall x.\exists y.Q(x,y) \wedge \neg\forall u.\exists v.Q(v.u)$ is satisfiable (the representation above shows that the domain ***Q*** is the whole domain of discourse and that range *Q* is not the whole domain) [2].
- **Tautology:** FOPL sentence ***G*** over ***S*** is a tautology if $\boldsymbol{F} \rightarrow \boldsymbol{G}$ holds for every ***S*** structure of ***F***
- **The relationship between satisfiability and tautology:** This relationship leads us to the use of Logic/Semantic consequences. Let ***x*** be a set of sentences over a signature ***S*** and ***G*** be a sentence over ***S*** then ***G*** follows from ***x*** (is a semantic consequence of x). If the following implications hold for every ***S*** structure ***F***: If $\boldsymbol{F} \vDash \boldsymbol{E}$ for all $\boldsymbol{E} \in \boldsymbol{x}$, then $\boldsymbol{F} \vDash \boldsymbol{G}$.

A simple example is as follows; if my house is big and all big houses are expensive, then my house is expensive.

$$\{Big\ (myHouse),\ House\ (myHouse),\ \forall x\ ((big\ (x) \wedge house\ (x)) \rightarrow Expensive(x))\} \vDash Expensive\ (myHouse)$$

For every sentence ***G***, thus, we have complete information about a domain of discourse. Applying this to another example of polysemous words in a sentence: "Your handwriting is fabulous" can be represented as:

$$[(Handwriting \wedge writing) \rightarrow To\ write]\ [(Incredible \wedge Fantastic \wedge Fabulous\ \wedge Wonderful) \rightarrow\ To\ be\ amazing]$$

III. RESULTS AND DISCUSSION

The evaluation of the proposed work was performed using a selection of sample sentences obtained from [22] as shown in Table 1. These samples contain sentences that have similar meanings but expressed using different words and sentences that have dissimilar meaning. The sample sentences are predefined by a human as either similar or dissimilar. The purpose of the evaluation is to measure the effectiveness of the proposed method.

TABLE I
SELECTED SAMPLE SENTENCES

	Sample Sentences	**Similar/ Dissimilar**
1.	"Agriculture Secretary Luis Lorenzo told Reuters there was no damage to the vital rice crop as harvesting had just finished".	Similar
	"Agriculture Secretary Luis Lorenzo said there was no damage to the vital rice crop as harvesting had just ended".	
2.	"Taha is married to former Iraqi oil minister Amir Muhammed Rasheed, who surrendered to U.S. forces onApril 28."	Similar
	"Taha's husband, former oil minister Amer Mohammed Rashid, surrendered to U.S. forces on April 28."	
3	"Democrats and two Republicans are running for her seat and have qualified for the Feb 3 primary ballot."	Dissimilar
	"Six Democrats are vying to succeed Jacques and have qualified for the Feb 3 Primary ballot."	
4.	"Perkins will travel to Laurance today and meet with Kansas Chancellor Robert Hemenway.	Dissimilar
	"Perkins and Kansas Chancellor Robert Hemenway declined to comment Sunday night".	
5.	"On Sunday, a US soldier was killed and another injured when a munitions dump they were guarding exploded in southern Iraq.	Dissimilar
	"A soldier was killed Monday and another wounded when their convoy was ambushed in northern Iraq".	
6.	"On July 22, Moore announced he would appeal the case directly to the U.S. Supreme Court".	Similar
	"Moore of Alabama says he will appeal his case to the nation's highest court".	
7.	"I am proud that I stood against Richard Nixon, not with him,' Kerry said".	Dissimilar
	"I marched in the streets against Richard Nixon and the Vietnam War,' she said"	
8.	"The report by the independent expert committee aims to dissipate any suspicion about the Hong Kong government's handling of the SARS crisis."	Dissimilar
	"A long-awaited report on the Hong Kong government's handling of the SARS outbreak has been released	

We employ the Jaccard Similarity Index (JSI) [23] to measure the similarity between sentences. Our proposed method addresses the meaning of the sentences via pre-processing and representation scheme to compare word senses. Similar senses will be regarded as same words or phrases. According to [23], JSI is the size of the intersection of two sentences divided by the size of the union of the two sentences. It helps to identify the intersection, meeting point or similarity between two or more distinctive sentences. JSI is calculated using Equation 1.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Where A and B are two different set of sentence representation and J(A,B) computes the fraction of the intersection to the union of the sets. The results are represented with a value of "0" to"1". 0 means that the text data are completely dissimilar while 1, in contrast, means the text is similar. Words are tagged in each sentence and will be converted to figures for easy application of JSI equations.

The following is the formal description of the steps to calculate sentence similarity;

Step 1: Tagged words {Do, Does}[a] {Food, Dish, Fare}[b] {have, serve}[c] {Vegetarian}[d] {Maharani}[e]

S1: Does[a] Maharani[e] have[c] vegetarian[d] dish[b]?

S2: Do[a] they have[c] vegetarian[d] food[b] at Maharani[e]?

S3: Are vegetarian[d] dishes[b] served[c] at Maharani[e]?

S4: "Does[a] Maharani[e] serve[c] vegetarian[d] fare[b]?"

Step 2: *[(S1∩S2∩S3∩S4) / (S1US2US3US4)]*

Step 3: *[(5+5+4+5) / (20)] =0.95*

The value produced in step 3 i.e. 0.95 which are can be interpreted as similar. Hence the entire four sentence are similar.

Using human evaluation as a benchmark, we have compared the similarity of each sentence produced by the proposed method with predefined similarity check performed by a human as depicted in Table 1. In order to enable the comparison, human decisions on similar sentence were given a score of 1 and the dissimilar decision is given the score of 0. Fig. 3 shows the result of the comparison

As shown in Fig. 3, sentence similarities of the proposed method (canonical form) on the sample sentences are almost identical with the human evaluation as opposed to without canonical form. We have used the correlation coefficient to determine the relationship between the similarity scores produced with and without canonical form and the human evaluation. The correlation of the similarity score with canonical form is 96% as opposed to 31% of the scores obtained without identifying canonical forms. Therefore, Fig. 3 and the correlation coefficient value indicate that the proposed method has potentials in identifying semantics of sentences.

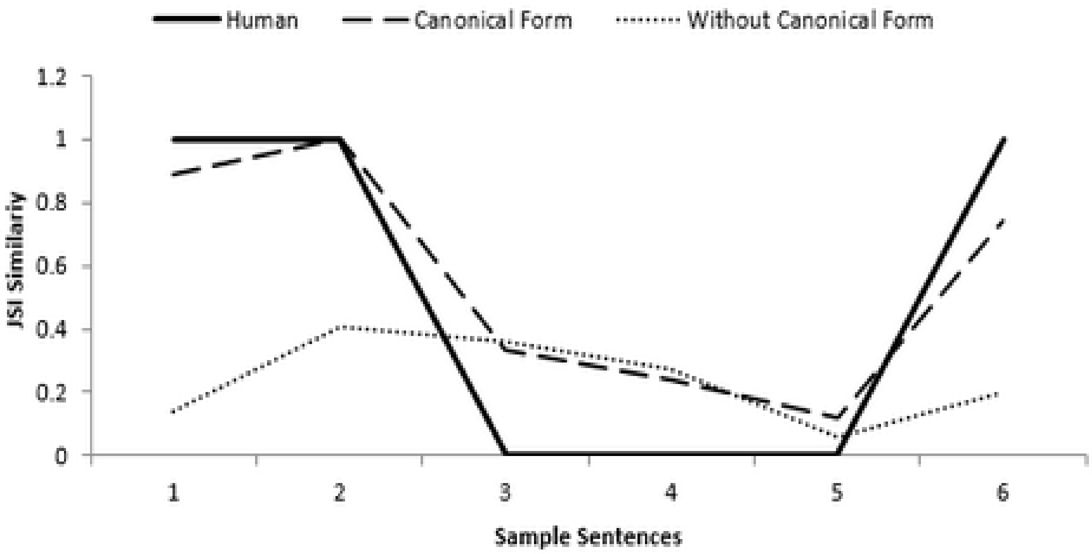

Fig. 3 Results of JSI calculations for the sample sentences

IV. Conclusion

In this paper, we have presented a canonical form representation of text employing WSD and FOPL. We proposed two algorithms in WSD i.e. Lesk and Selectional Preference Algorithm. Lesk algorithm is proposed because of its accuracy in short sentences. Selectional Preference is proposed because it can identify the senses in polysemous words. The use of semantic and logic consequences theorem in FOPL is able to satisfy the problem of complete and incomplete information in the identified canonical form.

We developed simple and easy-to-implement techniques for transforming polysemous sentences into a representation that is able to capture semantics. We performed a preliminary investigation to measure the effectiveness of the proposed method using selected sample sentences. The result of the investigation is promising.

The work reported in this paper is a work in progress towards developing method of uncovering semantic information from corpuses. Future work will concentrate on implementing the canonical form representation on larger text corpora to measure the actual performance of the proposed method.

Acknowledgement

We would like to thank the Malaysian Ministry of Higher Education for providing the funding for this research through the Exploratory Research Grant Scheme (ERGS).

References

[1] A. Classen, Q. Boucher, and P. Heymans, "A text-based approach to feature modelling: Syntax and semantics of TVL," *Sci. Comput. Program.*, vol. 76, no. 12, pp. 1130–1143, 2011.

[2] L. Jiang, H. Zhang, X. Yang, and N. Xie, "Research on Semantic Text Mining Based on Domain Ontology," pp. 336–343, 2013.

[3] V. Gurusamy, S. Kannan, "Preprocessing Techniques for Text Mining", Conference paper- October 2014

[4] R. Navigli, “Word sense disambiguation: A survey,” *ACM Comput. Surv.*, vol. 41, no. 3, p. 10, 2009.

[5] D. Jurafsky and J. H. Martin, “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition,” *Speech Lang. Process. An Introd. to Nat. Lang. Process. Comput. Linguist. Speech Recognit.*, vol. 21, pp. 0–934, 2000.

[6] E. Cambria and G. Melfi, “Semantic Outlier Detection for Affective CommonSense Reasoning and Concept-Level Sentiment Analysis,” pp. 276–281, 2015.

[7] S. S. Kamaruddin, A. R. Hamdan, A. A. Bakar, and F. Mat Nor, “Deviation detection in text using conceptual graph interchange format and error tolerance dissimilarity function,” *Intell. Data Anal.*, vol. 16, no. 3, pp. 487–511, 2012.

[8] S. S. Kamaruddin, A. A. Bakar, A. R. Hamdan, F. M. Nor, M. Z. A. Nazri, Z. A. Othman, and G. S. Hussein, “A text mining system for deviation detection in financial documents,” *Intell. Data Anal.*, vol. 19, no. s1, pp. S19–S44, 2015.

[9] F. S. Tsai, “Novelty detection for text documents using named entity recognition,” *2007 6th Int. Conf. Information, Commun. Signal Process.*, no. 3, pp. 1–5, 2007.

[10] I. Dagan, B. Dolan, B. Magnini, and D. Roth, “Recognizing textual entailment: Rational, evaluation and approaches,” *Nat. Lang. Eng.*, vol. 15, no. 4, pp. i–xvii, 2009.

[11] I. Beltagy, S. Roller, P. Cheng, K. Erk, and R. J. Mooney, “Representing Meaning with a Combination of Logical Form and Vectors,” pp. 1–44, 2015.

[12] A. Lenci, S. Montemagni, and V. Pirrelli, “The Acquisition and Representation of Word Meaning The Acquisition and Representation of Word Meaning . An Overview,” 2001.

[13] S. Gahl, L. Menn, G. Ramsberger, D. S. Jurafsky, E. Elder, M. Rewega, and L. H. Audrey, “Syntactic frame and verb bias in aphasia: Plausibility judgments of undergoer-subject sentences,” *Brain Cogn.*, vol. 53, no. 2, pp. 223–228, 2003.

[14] P. W. Foltz, “Latent Semantic Analysis for Text-Based,” *Behav. Res. Methods, Instruments Comput.*, vol. 28, no. 2, pp. 197–202, 1996.

[15] S. Hassan and R. Mihalcea, “Semantic Relatedness Using Salient Semantic Analysis.,” *Proc. 25th AAAI Conf. Artif. Intell. (AAAI 2011)*, pp. 884–889, 2011

[16] E. Galeazzi, A. R. Mori, F. Consorti, A. Errera, and P. Merialdo, “A cooperative methodology to build conceptual models in medicine,” *Stud. Health Technol. Inform.*, vol. 43, pp. 280–284, 1997.

[17] R. Baud, C. Lovis, L. Alpay, A. M. Rassinoux, J. R. Scherrer, A. Nowlan, and A.Rector, “Modelling for natural language understanding,” *Proc Annu Symp Comput Appl Med Care*, no. Journal Article, pp. 289–293, 1993.

[18] N. Wacholder and I. Berlin, “Disambiguation of Proper N a m e s in Text,” pp. 202–208.

[19] M. J. Cresswell, “Logics and languages,” in *philpapers*, Routledge, 2016.

[20] G. E. Bakx, “Machine Learning Techniques for Word Sense Disambiguation,” *Mach. Learn.*, 2006.

[21] D. Garrette, K. Erk, and R. Mooney, “A Formal Approach to Linking Logical Form and Vector-Space Lexical Semantics,” *Comput. Mean. SE - 3*, vol. 47, pp. 27–48, 2014.

[22] Li, L., Xia, H., Hu B.Y., Wang J., and Zhou, Y. M. “Measuring Sentence Similarity from Different Aspects”, Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, 2009, pp.2244 -2249.

[23] J. D. U. A.Rajaraman, J. Leskovec, *Mining Massive Data Sets Winter 2016*. Cambridge University Press, 2016.

[24] D. Jurafsky and J. H. Martin, “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition,” Speech Lang. Process. An Introd. to Nat. Lang. Process. Comput. Linguist. Speech Recognit., vol. 21, pp. 0–934, 2000.

[25] N. Wacholder and I. Berlin, “Disambiguation of Proper N a m e s in Text,” pp. 202–208.

[26] M. Lesk, “Automatic sense disambiguation using machine readable dictionaries,” Proc. 5th Annu. Int. Conf. Syst. Doc. - SIGDOC ’86, pp. 24–26, 1986.

[27] S. Banerjee, “Adapting the Lesk Algorithm for Word Sense Disambiguation to WordNet,” no. December, 2002.

[28] M. J. Cresswell, “Logics and languages,” in philpapers, Routledge, 2016.

[29] J. Weeds, D. Clarke, J. Reffin, D. Weir, and B. Keller, “Learning to Distinguish Hypernyms and Co-Hyponyms,” Proc. COLING 2014, pp. 2249–2259, 2014.

[30] F. Ramli, S.A.M. Noah, " Building an Event Ontology for Historical Domain to Support Semantic Document Retrieval", International Journal on Advanced Science, Engineering and Information Technology, Vol. 6 (2016) No. 6, pages: 1154-1160, DOI:10.18517/ijaseit.6.6.1634

[31] S. I. A. Saany,A. Mamat,A. Mustapha,L. S. Affendey and M. N. A. Rahman,"Syntax and Semantics Question Analysis Using User Modelling and Relevance Feedback," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, no. 1, pp. 329-337, 2017. [Online]. Available: http://dx.doi.org/10.18517/ijaseit.7.1.1334.