Measuring Semantic Similarity for Bengali Tweets Using WordNet

Dwijen Rudrapal
CSE Department
NIT Agartala, India
dwijen.rudrapal@gmail.com

Amitava Das
CSE Department
IIIT, Sricity, India
amitava.santu@gmail.com

Baby Bhattacharya
CSE Department
NIT Agartala, India
babybhatt75@gmail.com

Abstract

Similarity between natural language texts, sentences in terms of meaning, known as textual entailment, is a generic problem in the area of computational linguistics. In the last few years researchers worked on various aspects of textual entailment problem, but mostly restricted to English language. Here in this paper we present a method for measuring the semantic similarity of Bengali tweets using WordNet. Moreover we defined partial textual entailment (PTE) as in real data partial entailment cases are equally prevalent with the complete/direct entailment. Although by definition entailment is a directional relationship, but here we consider entailment more as semantic similarity.

Keywords: Semantic similarity; WordNet; Synonym;

1 Introduction

Variations of natural language expression make it difficult to determine semantically equivalent sentences. The beauty of natural languages is similar meaning could be expressed in countless ways; therefore it is a very complex task to measure relatedness of natural language sentences. Morpho-Syntactic variations of similar meaning expressions are more prevalent in social media text due to its informal nature. Semantic similarity score plays important role in many Natural Language Applications (NLP) such as multi-document summarization (MDS), question answering(QA), information extraction(IE) (Bhagwani et al., 2012). Several researchers have explored numbers of semantic similarity methods mostly for English but very less for Indian languages and almost nothing for Bengali. Technically these methods can be categorized into two groups: dictionary/thesaurus-based (one such example is edge counting-based) methods and corpus-based (one such method is information theory-based) methods (Li et al., 2003). Edge counting based methods use only **semantic links** and corpus based methods combine corpus statistics with **taxonomic distances**.

The objective of this work is to design a system to measure semantic similarity score between two Bengali tweets. We adopted a lexical based method; the words are grouped into clusters in terms of their senses along with their synonyms. Our proposed method centered on analyzing shared words similarity among tweets.

Partial Textual Entailment (PTE) is defined as a bidirectional relationship among a sentence/tweet pair. It defines partial/complete meaning inference from one sentence/text from another text. We define these following 4 detailed PTE categories:

1. **Type 1**: If both the given texts are having same information and mean same, then it is a case of direct entailment and should be noted as (X=X).

2. **Type 2:** If the first/second given text has any extra information than the second/first text respectively then it is been categorized as PTE2. This type may have two variations like: (X=X+Z or X+Z=X).

3. **Type 3:** If the first given text has all the information of the second given text and has some extra information, then its 3rd variation of PTE, noted as (X+Z=X+Y).

4. **Type 4:** If both the given texts are not having common information then it is a *NOT-Entailed* case.

In all the above cases *X, Y, Z* represents a block of information in a given text.

The remainder of this paper is organized as follows. Section 2 describes corpus acquisition and annotation process, followed by section 3 introduced WordNet structure and the pre-processing step. Section 4 details experiment and evaluation setup. In the section 5 we reported performance of the baseline system. Section 6 is a discussions section on errors in results. Section 7 reviews related work and finally the section 8 concludes the paper.

2 Corpus Acquisition and Annotation

2.1 Corpus

To create Bengali tweet corpus for the proposed entailment problem we targeted tweets on specific contemporary popular topics. The rationale behind topic based tweets collection is to capture people's natural way of explaining an event using different synonymous words and varied syntactic formations while expressing the same meaning. A paid Twitter API[1] has been used for this purpose. Total 6500 Bengali tweets have been collected for the period of 2 months (August 2014-September 2014) on 25 different topics covering various domains like international and national politics, sports, natural disasters, political campaigns and elections. For example Jamayet Strike issue in Bangladesh, Cheat fund scam in Orissa and Bengal, Flood in Kashmir, Ukraine crisis, Knight Riders performance in IPL, Bi-election in West Bengal etc.

In few topics tweets were surprisingly higher, more than 2000, in some topics number of tweets were less or around 100.

2.2 Annotation and Corpus Statistics

For the manual annotation of semantic similarity among tweets, we involved two human annotators, who are native Bengali speakers but not linguist. An automatic cosine similarity method applied to same topic cluster to prune tweet pairs for the annotation from the corpus. An experimentally chosen threshold then set to create annotation pairs. Finally tweet pairs are being manually marked according to the PTE types. Annotation agreement has been measured on a small subset, randomly chosen on one topic: having 100 sentence pairs. We found the annotation consensus is of 0.86 *kappa* (Cohen J, 1960). One empirical question could be raised here that cosine similarity based pruning is a biased method, whereas empirically there are countless ways to express same meaning with different set of words (synonyms). To make sure we thoroughly analyzed our left out part of the corpora (left out after cosine pruning) and found only handful cases (3-4%) where people use different wordings altogether.

The annotation process produced a set of 804 tweet pairs, among them 350 tweet pairs were found as entailed and 454 tweets pair annotated as negative cases. The exact distribution of the

different PTE classes in the annotated data is shown in following table 1.

TWT pairs	PTE types			
	type 01	type 02	type 03	type 04
804	350 (43.5%)	94 (11.69%)	74 (9.20%)	286 (35.57%)

Table 1: Distribution of tweet pairs in PTE classes

It could be noticed that there are significant presence of PTE 2 and 3 classes in the real corpus, whereas the majority class is till the direct entailment case. Now an argument could be raised that why these negative examples i.e. PTE-04 type is so essential to include. The rationale is, these negative examples are so important because this is the exclusion set made by annotators despite of high cosine similarity value with their peers. The average cosine similarities score of the negative examples are 0.25 and for PTE-03 is **0.35**. Ranges and average cosine similarity scores on the golden set is reports in the Table 2. For example:

বৃহস্পতি ও রোববার হরতাল ডেকেছে জামায়াত
ENG: *Thursday and Sunday Jamayet called strike.*

সিরাজগঞ্জে জামায়াতের নিরুত্তাপ হরতাল
ENG: *Jamayet called strike is peacefully in Shirajganj.*

Cosine similarity: 0.516

SN	Types	Cosine Similarity	
		Ranges	Avg.
1	Entailed	> 0.70	0.70
2	Not-Entailed	< 0.70	0.35
3	PTE-type 1	> 0.70	0.70
4	PTE- type 2	0.40 - 0.69	0.46
5	PTE- type 3	0.30 - 0.39	0.35
6	PTE- type 4	< 0.30	0.25

Table 2: Ranges of cosine similarity scores

3 Bengali WordNet

WordNet is a lexical semantic network to hold semantic relations like synonyms and word-senses as the nodes of the network and relations of the synonyms and word-senses are the edges of the network. In WordNet, meaning of each word is represented by a unique word-sense and a set of its synonyms called synset. We have collected the Bengali WordNet developed by Das and Bandyopadhyay as described in (Das and Bandyopadhyay 2010), consists total 12K numbers of synsets.

3.1 Pre-Processing

Text pre-processing is a vital pre-requisite while working with noisy social media text. Pre-

538

processing involves splitting tweet into valid tokens: words and symbols, stemming, moving out stop words and part-of-speech tagging. The CMU tweet tokenizer (Gimpel et al., 2011) has been used here. Although it is primarily developed for English but also works well for other languages like Bengali. We used the Bengali stop word list, made available publicly by ISI Kolkata[2]. For the POS tagging the system developed by (Dandapat et al., 2007) has been used. Although the POS tagger is not trained on social media text and accuracy of the tagger on tweet has not been measured. This is something we would like to do next.

To trim all the surface word forms into corresponding root we developed one simple rule based Bengali Stemmer. Our stemmer concentrated on framing rules for stemming word categories like noun, verb adverb and adjectives. To frame the rules for stripping suffixes and prefixes we drew inspirations and knowledge from (Dash, 2014) and (Das and Bandyopadhyay, 2010).

3.2 Similarity Computation

We devised two kinds of similarity measurement methods for word level then accumulated those word-level similarities to sentence level.

3.2.1 Computation of Word Similarity

Study from different psychological experiments demonstrates that semantic similarity is obviously context-dependent (Medin et al., 1993), (Tversky, 1977). Meaning of a word in sentence is context-dependent, which effects semantic similarity. For example,

খাওয়ার আগে হাত ভালো করে ধুয়ে নেবে
ENG: Before the meal, wash hands properly
রিয়ানুজ এর হত্যাকান্ডে ওর ও হাত ছিল
ENG: He was also involved in the Riyanuj murder case.

Two above cited sentences have a common word "হাত/hand", but the word meaning is different in two sentences. In the first sentence "হাত" implies a part of human body and in 2nd sentence "হাত" implies association/involvement in one event.

For the semantic similarity calculation among two given words w_1 and w_2, we computed a scalar distance of these words in the meaning-spaces based on the synsets of these words extracted from the WordNet. If w_1 and w_2 both belong to same sysnset i.e.w_1 is a synonym of w_2 or vice versa, then the distance (d) between w_1 and w_2 is

0 and the semantic similarity score is 1, otherwise, the distance (d) between w_1 and w_2 is 1 and semantic similarity score is 0.

$$\text{Sim}(w_1, w_2) = \begin{cases} 1 & (if \ d = 0) \\ 0 & (if \ d = 1) \end{cases} \quad (1)$$

For example:

w_1: অভিজ্ঞ (Experienced)
w_2: পারদর্শী (Expert)

Calculated semantic similarity score is 1.

3.2.2 Sentence Similarity Computation

For the sentence level similarity calculation we performed two sets of experiments. One with fine-grained entailment PTE classes i.e. the 4 classes and the other is a binary classification task: entailed or not entailed.

To determine the semantic similarity score of two given tweets A and B, we first pre-processed the tweets as described in the section 2.2 and calculated the length of tweets. Say, x is the length of tweet A and y is the length of tweet B. Then a semantic similarity matrix R[x,y] has been developed of each pair of words w_i and w_j where i and j are the indices of words. If a word at any position in A is not available in the WordNet, we computed the word similarity based on presence of same word in B. If such a word from A gets complete word match with any word in B, then similarity score is 1 between the words else 0. For example names and abbreviations like স. পা (Samajbadi Party), বিজেপি (BJP) which are the abbreviations of political party name, are not available in WordNet. Their similarity measured based on character matching of each word in the tweets.

Every token of tweet A represents a row and every token of tweet B represents a column in the semantic similarity relative matrix R[x,y].Figure 1 1illustrates an example similarity matrix representation of two example tweets as cited below. Each cell represents the word level similarity scores. For example:

সাঈদীর আমৃত্যু কারাদণ্ড প্রদান করায় হরতাল ডেকেছে জামায়াত
ENG: Jamayet called strike on the lifetime imprisonment issue of Sighdi.
জামায়াতের বনধ চলছে, সাঈদীর আজীবন কারাদণ্ড দেওয়ার প্রতিবাদে
ENG: Jamayet's strike is going on, in protest of Sighdi's lifetime imprisonment.
Computed semantic similarity score is 0.923

	সাঈদী Sighdi	আমৃত্যু Life-Time	কারাদও Impris-onment	প্রদান Anou-nced	হরতাল Strike	জামায়াত Jama-yet
জামায়াত Jama-yet	0	0	0	0	0	1
বনধ Strike	0	0	0	0	1	0
সাঈদী Sighdi	1	0	0	0	0	0
আজীবন Life-time	0	1	0	0	0	0
কারাদও Impris-onment	0	0	1	0	0	0
দেওয়া Anno-unce	0	0	0	1	0	0
প্রতিবাদ Prot-est	0	0	0	0	0	0

Figure1: Semantic similarity matrix between tweets.

Matching weight of tweet *A* computed by summing all the row wise cell weight and Matching weight of tweet *B* computed by summing all the column wise cell weight. In above cited example matching weight of both tweet *A* and *B* is 6. Following formula is used to determine the semantic similarity score between tweet *A* and tweet *B*.

$$\text{sim}(A, B) = \frac{\sum total\ matching\ weight(A,B)}{\sum length(A,B)} \quad (2)$$

An important point is that the proposed similarity value is based on each of the individual word similarity values, so that the overall similarity always reflects the influence of each word and its senses. According to the proposed semantic similarity score formulation, similarity values ranges from 0 to 1. If all the words of tweet *A* get semantically similar to all the words in tweet *B*, score will be 1, and will be 0 if there is no match.

4   Performance

System performance has been evaluated in two folds: with the binary (entailed or not) classes and with the fine-grained PTE classes. For performance evaluation we measured similarity score of all the tweet pairs in a class. Then experimentally, we set threshold to achieve optimum accuracy for each class. Decided threshold values are reported in the table 3.

SN	PTE	Threshold Range
1	**Entailed**	> 0.75
2	**Not-entailed**	< 0.75
3	**Type 1**	> 0.75
4	**Type 2**	0.2 - 0.29
5	**Type 3**	0.3 - 0.74
6	**Type 4**	< 0.2

Table 3: Threshold values of semantic similarity for Bengali tweets.

Accuracy results of our proposed system on binary class and fine-grained classes considering the pre-set threshold values are reported in Table 4 and 5.

Types	Precision	Recall	F1
Entailed	98.23	63.42	77.08
Not-Entailed	77.85	99.11	87.2
Avg.	88.04	81.265	82.14

Table 4: Performance on binary entailment classes

PTE classes	Precision	Recall	F1
PTE- 01	98.23	63.42	77.08
PTE- 02	26.15	36.17	30.35
PTE- 03	16.54	60.81	26.01
PTE-Type 04	86.36	53.14	65.8
Avg.	56.82	53.385	49.81

Table 5: Performance on the PTE classes

We setup another experiment on English tweets to evaluate the proposed approach and for the purpose of comparison. From SemEval 2015 task 1[3], we collected POS tagged corpus of tweet pairs. We involved two human annotators and tagged 639 tweets pair according to the PTE classes. To measure inter-annotator agreement, randomly 100 tagged pairs have been chosen. We found inter-annotator agreement is 0.709. Detail distribution of the tweet pair according to PTE classes is shown in table 6.

TWT pairs	PTE types			
	type 01	type 02	type 03	type 04
639	48 (7.5%)	61 (9.5%)	83 (12.9%)	447 (69.95%)

Table 6: English tweet pairs in PTE classes

Then we applied our proposed algorithm to determine the semantic similarity using English WordNet[4] (Boyd-Graber et al., 2006). All the POS tagged tweets are pre-processed by removing stop words[5] and lemmatization (Manning et al, 2014). System performance on these English tweet pairs measured in two folds: binary classes and fine-grained PTE classes. For each fold we achieved optimum accuracy with the pre-defined threshold values as mentioned in the table 7.

SN	PTE-Type	Threshold Range
1	**Entailed**	> 0.65
2	**Not-entailed**	< 0.65
3	**Type 1**	> 0.65
4	**Type 2**	0.5 to 0.64
5	**Type 3**	0.4 to 0.49
6	**Type 4**	< 0.4

Table 7: Threshold ranges for Eng. tweets.

---

[3] http://alt.qcri.org/semeval2015/task1/

[4] http://wordnetcode.princeton.edu/standoff-files/core-wordnet.txt

[5] http://www.lextek.com/manuals/onix/stopwords1.html

Performance of the proposed system on the SemEval English tweets is reported in the Table 8 and 9.

Types	Precision	Recall	F1
Entailed	22.75	79.16	35.34
Not-Entailed	97.88	78.17	86.92
Avg.	60.32	78.67	61.13

Table 8: Performance on the binary entailment classes for English tweets

PTE classes	Precision	Recall	F1
PTE- 01	31.40	79.16	44.97
PTE- 02	14.28	16.39	15.26
PTE- 03	13.63	14.45	14.03
PTE-Type 04	94.58	66.44	78.05
Avg.	38.47	44.11	38.07

Table 9: Performance on the PTE classes for English tweets

Results on English tweets are directly comparable with (Xu et al., 2014), named as MULTIP, make use of features like string comparison, POS and topic words. The reported final accuracy was 71.5 (F-Measure), whereas feature ablation shows string + POS features achieved 49.6 (F-measure), is directly comparable with our system's result: 61.13 on binary classes, while our system is only using WordNet based lexical features. Performance degradation on fine-grained classes is quite natural NLP phenomena. Integration of POS and topic words feature into our system could be straight-forward but extracting those features for Bengali tweets, demands research endeavors as those NLP tools are unavailable presently for the language.

5 Baseline System and Performance

SN	PTE	Threshold Range	F1
1	Entailed	> 0.75	72.89
2	Not-entailed	< 0.75	84.7
3	Type 1	> 0.75	73.48
4	Type 2	0.2 - 0.29	4.60
5	Type 3	0.3 - 0.74	11.9
6	Type 4	< 0.2	75.87

Table 10: Baseline system Performance on the PTE classes for Bengali tweets.

We have developed a very basic system to categorize Bengali tweets according to the defined PTE classes. Two tweets compared using only word matching and without WordNet information. This simple method returns a similarity score among two tweets. We calculated similarity score for all the PTE class tweets and experimentally set threshold for each class to achieve highest accuracy. Threshold values for each class

and the accuracy of the system reported in table 10.

Performance of the proposed system over the baseline system shows better accuracy and also clarifying the fact that PTE recognition is more challenging than the classical unidirectional textual entailment recognition.

6 Discussion

System's poor performance on the fine-grained classes is a natural phenomenon for any NLP system. This is an ongoing work. Here in this section we are discussing on challenges related with the PTE classes.

Let us first explain why PTE classes identification is required. Common information boundary detection is essential for various applications for example multi-document summarization (MDS). A MDS needs to remove common information chunks before the aggregation.

Indeed automatic PTE detection for social media text is a challenging problem. Moreover additional NLP resources for a resource scarced language like Bengali are not well developed. Looking at the error types we decided to go for a system can take both the feature input: lexical and syntactic, but dependency parser development for Bengali tweets is a separate problem altogether.

Confusion matrix is drawn for Bengali tweets (Figure 2) and English tweets (Figure 3) to understand overlap between PTE classes and it has been observed PTE02-PTE03 are closely overlapped with each other on both the data set.

		System tagged				
Golden		PTE 01	PTE 02	PTE 03	PTE 04	Total
	PTE 01	222	03	125	0	350
	PTE 02	2	34	43	15	94
	PTE 03	2	18	45	9	74
	PTE 04	0	75	59	152	286
	Total	226	130	272	176	804

Figure 2: Confusion matrix for Bengali tweets.

		System tagged				
Golden		PTE 01	PTE 02	PTE 03	PTE 04	Total
	PTE 01	38	8	1	1	48
	PTE 02	41	10	7	3	61
	PTE 03	42	16	12	13	83
	PTE 04	46	36	68	297	447
	Total	167	70	88	314	639

Figure 3: Confusion matrix for English tweets.

7 Related Works

Automatic detection of textual entailment is a well-studied discipline, but most of the endeavors so far concentrated on English, almost no work on Indian languages especially on Bengali. There are many approaches to measure semantic similarity of words and sentences based on simple organizational schemes like Dictionary to complex organizational schemes like WordNet [Fellbaum, 2010] and ConceptNet [Liu et al., 2004]. The model proposed by [Tversky, 1977] is one of the early works in this area.

Technically these methods can be categorized into two groups: edge counting-based (or dictionary/thesaurus-based) methods and information theory-based (or corpus-based) methods (Li et al., 2003). Among two approaches, very less research work done on edge counting based method. Rada et al. (Rada, R et al., 1989), proposed a metric called distance, which determines the average minimum path length over all pair wise combinations of nodes between two subsets of nodes. Distance measure has been used to assess the conceptual distance between sets of concepts when used on a semantic net of hierarchical relations and represents the relatedness of two words

Due to the specific applications of edge counting based method like medical semantic nets (Li et al., 2003), most of the research on semantic similarity followed information theory based method (Resnik, 1993a) work is the first work on information theory based system which proposed modeled the selectional behavior of a predicate as its distributional effect on the conceptual classes of its arguments. This model experiment result suggests that many lexical relationships are better viewed in terms of underlying conceptual relationships. In a later work (Resnik, 1993b) focuses on two selectional preferences and semantic similarity as information-theoretic relationships involving conceptual classes and demonstrates the applicability of these relations to measure semantic similarity between two words. A model proposed by (Lee et al., 1993) also measured the distance of the nodes using edge weights between adjacent nodes in a graph as an estimator of semantic similarity. The work by (Richardson et al., 1994) has proposed a WordNet based scheme for Hierarchical Conceptual Graphs (HCG) to measure semantic similarity between words. System proposed by (Li et al., 2006), uses a semantic-vector approach to measure sentence similarity. Sentences are transformed into feature vectors having individual words from the sentence pair as a feature set. System proposed (Liu et al., 2008) an approach to determine sentence similarity, which takes into account both semantic information and word order. They define semantic similarity of sentence 1 relative to sentence 2 as the ratio of the sum of the word similarity weighted by information content of words in sentence 1 to the overall information content included in both sentences. The method proposed by (Liu et al., 2013) presents an information theory based approach of calculating the similarity between very short texts and sentences using WordNet, common-sense knowledge base and human intuition.

For Bengali text the work by (Sinha et al., 2012) design and develop a Bangla lexicon based on semantic similarity among Bangla words from Samsad Samarthasabdokosh. The lexicon is hierarchically organized into categories and sub-categories. The words are grouped into clusters along with their synonyms. Weighted edges between different types of words related to same or different concepts or categories exist, denoting the semantic distance between them. (Sinha et al., 2014) proposed a hierarchically organized semantic lexicon in Bangla and also a graph based edge-weighting approach to measure semantic similarity between two Bangla words.

Our work is on information theory based method rather edge counting based method. Edge counting method is expedient for particular applications with constrained taxonomies (Li et al., 2003). In this paper, our work explains an approach to determine semantic relatedness between any two tweets.

8 Conclusion and Future Work

This paper presents an initial approach to measure semantic similarity between two Bengali tweets, based on the words meanings. Bengali tweets are less noisy in nature compared to English. In general people do use less abbreviated forms ('gr8' for great), word play ('goooood' for good) and etc., but Romanization / transliterated writing and code-mixing is very much prominent in Indian social media. Even romanization of Indian languages has no writing standard. People are literally whimsical about spelling over social media; for example pyari (beloved) could be written in various phonetically similar spellings: pyaari, payari, piari, and etc. We are currently working on PTE detection on code-mixed Bengali tweets.

References

A Das and S. Bandyopadhyay, "Morphological Stemming Cluster Identification for Bangla", In Knowledge Sharing Event-1: Task 3: Morphological Analyzers and Generators, Mysore, January, 2010.

Boyd-Graber, J., Fellbaum, C., Osherson, D., and Schapire, R. (2006). "Adding dense, weighted connections to WordNet." In: Proceedings of the Third Global WordNet Meeting, Jeju Island, Korea, January 2006.

Cohen J.,"A coefficient of agreement for nominal scales". Educational and Psychological Measurement.1960; 20 (1):37–46.

D.L. Medin, R.L. Goldstone, and D. Gentner, "Respects for Similarity," Psychological Rev., vol. 100, no. 2, pp. 254-278, 1993.

Dandapat S., Sarkar S. and Basu A "Auto-matic Part-of-Speech Tagging for Bengali: An approach for Morphologically Rich Languages in a Poor Resource Scenario". In Proceedings of the Association of Computational Linguistics (ACL 2007), Prague, Czech Republic.

Das, A. and Bandyopadhyay, S. "Semanticnet-perception of human pragmatics". In Proceedings of the 2nd Workshop on Cognitive Aspects of the Lexicon, pages 2–11, Beijing, China, 2010.

Fellbaum, C. (2010). "WordNet." *Theory and Application of Ontology: Computer Applications.* New York: Springer, 231-243.

Hongzhe Liu, Pengfei Wang, "Assessing Sentence Similarity Using WordNet based Word Similarity", JOURNAL OF SOFTWARE, VOL. 8, NO. 6, JUNE 2013.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, DaniYogatama, Jeffrey Flanigan, and Noah A. Smith, "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments", In Proceedings of ACL 2011.

Lee, J., Kim, M., and Lee, Y. (1993). "Information retrieval based on conceptual distance in is-a hierarchy". Journal of documentation, 49(2):188–207.

Liu, H. and Singh, P. (2004). "Conceptnet—a practical commonsense reasoning tool-kit".BT technology journal, 22(4):211–226.

Manjira Sinha, Abhik Jana, Tirthankar Dasgupta, Anupam Basu, "A New Semantic Lexicon and Similarity Measure in Bangla", Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III), pages 171–182,COLING 2012, Mumbai, December 2012.

Manjira Sinha,Tirthankar Dasgupta, Abhik Jana, Anupam Basu, "Design and Development of a Bangla Semantic Lexicon and Semantic Similarity Measure", International Journal of Computer Applications (0975 – 8887), Volume 95– No.5, June 2014.

Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60

Niladri Sekhar Dash, "A Descriptive Study of Bengali Words" Cambridge University Press, December 2014.

Rada, R., Mili, H., Bicknell, E., and Blettner, M. (1989). "Development and application of a metric on semantic nets". Systems, Man and Cybernetics, IEEE Transactions on, 19(1):17–30.

Resnik, P. "Selection and information: a class-based approach to lexical relationships". IRCS Technical Reports Series, page 200, 1993.

Resnik, P. "Semantic classes and syntactic ambiguity". In Proc. of ARPA Workshop on Human Language Technology, pages 278–283, 1993.

Richardson, R., Smeaton, A., and Murphy, J. (1994). "Using wordnet as a knowledge base for measuring semantic similarity between words". Technical report, Technical Report Working Paper CA-1294, School of Computer Applications, Dublin City University.

Sumit Bhagwani, Shrutiranjan Satapathy, Harish Karnic, "sranjans: Semantic Textual Similarity using Maxi mal Weighted Bipartite Graph Matching", First Joint Conference on Lexical and Computational Semantics (*SEM), pages 579–585, Montr´eal, Canada, June 7-8, 2012.

Tversky, A. (1977). "Features of similarity". Psychological review, 84(4):327.

Xiao-Ying Liu, Yi-Ming Zhou, Ruo-Shi Zheng. "Measuring Semantic Similarity Within Sentences". Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 2008.

Xu, W., Ritter, A., Callison-Burch, C., Dolan, W. B., and Ji, Y. (2014). Extracting lexically divergent paraphrases from Twitter. Transactions of the Association for Computational Linguistics (TACL), 2(1).

Yuhua Li, David McLean, Zuhair A. Bandar, James D. OShea, and Keeley Crockett. "Sentence Similarity Based on Semantic Nets and Corpus Statistics". IEEE Transections on Knowledge and Data Engineering, Vol. 18, No. 8, 2006.