# A Survey on Paraphrase Detection and Generation Techniques

[1]*Ravinder Mohan Jindal,* [2]*Vijay Rana,* [3]*Sanjeev Sharma*
[1]*Research Scholar,* [2,3]*Assistant Professor*
[1,2]*Department of Computer Science and Applications,* [3]*Department of Computer Applications*
[1,2]*Sant Baba Bhag Singh University, Jalandhar,*[3]*DAV University, Jalandhar*

## Abstract
Whenever "the same thing," need to be expressed using different ways or by various alternatives an automated paraphrase generation mechanism would be useful. One reason why paraphrase generation systems have been difficult to build is because paraphrases are hard to define. Although the strict interpretation of the term "paraphrase" is quite narrow because it requires exactly identical meaning, in linguistics literature paraphrases are most often characterized by an approximate equivalence of semantics across sentences or phrases. This paper presents a survey of paraphrase generation techniques for Indian and foreign languages.

**Keywords**: Paraphrasing, Sentence simplification, Sentence fusion, Sentence compression.

## Introduction
Paraphrase, as a general concept, is a very capricious thing; at one end, paraphrase is the relation between fragments of text having identical meanings, as under the transformational generative view, and at the other end there is the view that no text is interchangeable with any other. Each point in the spectrum ranging between these two views is a valid one, and in order to focus on a specific point it is necessary to examine paraphrase under a particular context and application.

## Existing work
Various authors have worked on detection and generation of paraphrases for various languages. The techniques for generating paraphrases have been classified in four categories. These are sentence compression, multi sentence fusion, sentence paraphrasing and sentence simplification.

## Sentence compression
Sentence compression means to rewrites a sentence into a shorter form without changing the semantics of the sentence. (Knight and Marcu, 2000 [1] used statistics based approach and developed noisy channel based model and decision tree based model and concluded that the performance of the noisy-channel compression algorithm degrades smoothly, while the performance of the decision-based algorithm drops sharply. Cohn and Lapata, 2008 [2] proposed a novel grammar extraction method, uses a language model for coherent output, and can be easily tuned to a wide range of compression specific loss functions. Further authors present a discriminative tree-to-tree transduction model that can naturally account for structural and lexical mismatches; Filippova and Strube, 2008 [3] presented a dependency tree based unsupervised model for sentence compression in which sentence can be shorten by removing subtrees. Pitler, 2010 [4] discussed and analyzed various compression techniques; Filippova et al. 2015 [5] used LSTM approach for deletion-based sentence compression and when compared with human this model outperforms the baseline and achieved 4.5 in readability and 3.8 in in formativeness.; Toutanova et al., 2016[6]).

## Sentence fusion:
Sentence fusion means fusing two or more sentences into single. The fused sentences must have overlapped information content. The aim of the fusion is to preserve common information and delete irrelevant details (McKeown et al., 2010[7]; Filippova, 2010 [8]; Thadani and McKeown, 2013 [9]).

## Sentence Paraphrasing:
Sentence paraphrasing aims to rewrite a sentence without changing its meaning (Dras, 1999 [10]; Barzilay and McKeown, 2001 [11]; Bannard and Callison-Burch, 2005 [12]; Wubben et al., 2010 [13]; Mallinson et al., 2017 [14]).

## Sentence Simplification:

Sentence (or text) simplification means reducing the difficulty of text. This is mainly done by replacing the difficult words present in a sentence with their easier synonyms (Siddharthan et al., 2004 [15]; Zhu et al., 2010 [16]; Woodsend and Lapata, 2011[17]; Wubben et al., 2012 [18]; Narayan and Gardent, 2014 [19]; Xu et al., 2015[20]; Narayan and Gardent, 2016[21]; Zhang and Lapata, 2017 [22]).

Other than above mentioned techniques, different concepts used by some researchers are mentioned in the following section:

## An approach to treat numerical information in the text simplification process [23]

In this research article, Susana Bautista, Raquel Herva´s, Pablo Gerva´s and Javier Rojo explained a technique to process numerical information present in the large complex sentences. This processing of numerical information is then used for simplification of sentences. The developed system, when tested, gave a precision of 0.94, a recall of 0.93 and an F-measure of 0.93. Further, the linguistic accuracy of the output was analyzed and the results showed that 83.56 % (almost 84 %) of the complex sentences having numerical expressions, when simplified, were correct.

## Corpus-based Sentence Deletion and Split Decisions for Spanish Text Simplification [24]

Sanja Štajner, Biljana Drndarevi´c, and Horacio Saggion in this research article, explained the process of automatic simplification of complex texts in Spanish. A corpus based technique was used by the author. In corpus based technique, a corpus containing original and manually simplified news articles was undertaken. Three basic features i.e. POS frequencies, syntactic information and text complexity measures were used for classification of complex and simplified text. At the end of the experiment, it is observed that the simplified and un-simplified sentences can be classified with an overall F-measure up to 0.92.

## Syntactic Sentence Simplification for French [25]

In this research article, Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat and Thomas Franc¸ois discussed a corpus based approach to syntactically simplify the complex French text in to simple one. This research was done to make text easier to understand by simplifying those complex syntactic structures that hinder reading. Simplification rules were developed using two parallel corpora. The system was tested on 202 sentences taken from informative corpora and the results showed that about 80% of the simplified sentences produced by this system were accurate.

## Text Simplification for Reading Assistance: A Project Note [26]

Kentaro Inui et al. described an ongoing research project on text simplification for Japanese language. Paraphrasing was used for syntactic simplification of complex sentences. In order to evaluate the effect of paraphrases on the readability of sentence, a questionnaire containing complex sentences along with a pair of paraphrases was developed. The readability of this pair of paraphrases when classified using SVMbased classifier gave a precision 95% and recall 89%.

## A Semantic Relevance Based Neural Network for Text Summarization and Text Simplification[27]

In this research work, Shuming Ma, Atsushi Fujita, Tetsuro Takahashi and Ryu Iida have shown their efforts to improve semantic relevance between source texts and simplified texts for text summarization and text simplification. Authors introduce a Semantic Relevance Based neural model to encourage high semantic similarity between texts and summaries. Author conducted the experiments on two corpus, namely PWKP (Parallel Wikipedia Simplification Corpus) and EW-SEW (English Wikipedia and Simple English Wikipedia) and got 48.26 BLEU (Bilingual Evaluation Under Study) score on PWKP (Parallel Wikipedia Simplification Corpus), and 88.97 BLEU score on EW-SEW (English Wikipedia and Simple English Wikipedia).

## Sentence Simplification with Deep Reinforcement Learning [28]

In this research, Zhang and Lapata developed a reinforcement learning-based text simplification system called DRESS (Deep Reinforcement Sentence Simplification). DRESS can jointly model the simplicity, grammaticality and semantic fidelity to the input. Author also proposed a lexical simplification component that further boosts the performance of DRESS. Author find that reinforcement learning offers a great means to inject

prior knowledge to the simplification task. When tested, DRESS (Deep Reinforcement Sentence Simplification) scores lower on FKGL (Flesch-Kincaid Grade Level index) and higher on SARI (system output against references and against the input sentence), which indicates that the model has indeed learned to optimize the reward function which includes SARI. When DRESS was integrated with lexical simplification (DRESS-LS), it yields better BLEU, but slightly worse FKGL and SARI.

## Splitting Complex English Sentences [29]

John Lee, J. Buddhika and K. Pathirage Don in this paper, applied parsing technology for syntactic simplification of English sentences. Author mainly focused on the identification of that text spans which can be removed from a complex sentence. Author used a dataset of sentences that exhibit simplification based on coordination, subordination, punctuation/parataxis, adjectival clauses, participial phrases, and appositive phrases. Author trained a decision tree with features derived from text span length, POS tags and dependency relations, and showed that it significantly outperforms a parser. Among the different constructs, the proposed system performed best for punctuation/parataxis, with precision at 0.92 and recall at 0.95.

## Text Simplification for Language Learners: A Corpus Analysis [30]

Lee and Ostendorf performed a detailed analysis of the corpus of news articles and abridged versions written by a literacy organization. This analysis was done to learn what kinds of changes people make when simplifying texts for the language learners. From the corpus of original and abridged news articles, author concluded that the simplified sentences contain fewer adverbs and coordinating conjunctions. Author's analyses about the original and abridged articles showed the importance of syntactic features (in addition to sentence length) for taking decisions about sentence splitting.

## A Novel Approach to Paraphrase Hindi Sentences using Natural Language Processing [31]

Nandini Sethi, Prateek Agrawal, Vishu Madaan and Sanjay Kumar Singh in this paper, discussed an approach for reframing the Hindi sentences to generate paraphrases. The tool developed by the author generates paraphrases for Hindi sentences only. Proposed approach was based upon the generation of alternate sentences by replacing synonym and antonyms in an input sentence. These alternate sentences were then combined to generate paraphrase sentence. The accuracy of the system was not mentioned by the author.

## Split and Rephrase [32]

Shashi Narayany Claire Gardentz Shay B. Coheny Anastasia Shimorinaz in this research article proposed a split and rephrase technique for simplification of complex sentences in English language. Each complex sentence was split into at least two simple sentences by converting the relative clause and subordinate clause into main clause. For split and replace task, four different models were proposed by author. These models includes: HYBRIDSIMPL Model, SEQ2SEQ, MULTISEQ2SEQ model, SPLIT-MULTISEQ2SEQ and SPLIT-SEQ2SEQ model. Besides these models, author also created WebSplit (Collection of complex sentences along with their corresponding simple sentences) for training and testing of these models. On evaluating these models, the BLUE score obtained was 39.97 for HYBRIDSIMPLE, 48.92 for SEQ2SEQ, 42.18 for MULTISEQ2SEQ, 77.27 for SPLIT-MULTISEQ2SEQ and 78.77 for SPLIT-SEQ2SEQ model.

## Paraphrase Generation as Monolingual Translation: Data and Evaluation [33]

In this paper, Sander Wubben, Antal van den Bosch, Emiel Krahmer presented a technique for generating paraphrases using monolingual corpus. For generation of monolingual corpus, author used Google News and a standard Phrase-Based Machine Translation (PBMT) framework. The developed system was tested using automatic score metrics i.e. BLEU, METEOR and ROUGE scores. Author observed that these automatic scores correlate with human depending upon the edit distance. At higher edit distances, correlation of BLEU with human judgments is 0.78.

## Paraphrasing with Bilingual Parallel Corpora [34]

In this paper, Colin Bannard and Chris Callison-Burch proposed a method for generating paraphrases using bilingual parallel corpora. Author showed that using alignment techniques from phrase based statistical machine translation; paraphrases in one language can be identified using a phrase in another language as a pivot. The system was evaluated using a set of manual word alignments. This system was further improved with

paraphrases extracted from automatic alignments and hence increased the accuracy of the paraphrases extracted from the automatic alignments from 48.9% to 57% without language model re-ranking, and further to 61.9% when language model re-ranking was included.

## Text Simplification as Tree Labeling [35]

Joachim Bingel and Anders Søgaard presented a Conditional Random Field over Dependency based model for text simplification and paraphrasing. The model was developed with objective to delete or paraphrase entire subtrees in dependency graphs so as to avoid ungrammatical output. This model used three-fold parallel monolingual corpus to learn paraphrases and deletions. The developed model reached readability scores comparable to word-based compression approaches across a range of metrics and human judgements while maintaining more of the important information. Further human evaluation showed that the proposed approach lead to readability figures that were comparable to previous state-of-the-art approaches to the more basic sentence compression task, and better than previous work on joint compression and paraphrasing.

## Unsupervised Sentence Simplification Using Deep Semantics [36]

Shashi Narayan and Claire Gardent presented an unsupervised technique for sentence simplification and this technique was based upon deep semantics. All modules used for lexical simplification, syntactic simplification and sentence compression were unsupervised. Authors compared this unsupervised technique with four state-of-the-art supervised systems developed by Zhu, Woodsend, Wooben and Narayan, and manual examination of the results indicates that unsupervised technique achieves good simplicity rates through both deletion and sentence splitting. Moreover, it is observed that the average word length of simplified sentences using unsupervised technique is smaller i.e. 26.22 than for Wubben (28.25) and Woodsend (28.10); comparable with Narayan (26.19) and higher only than Zhu (24.21). It is further experimentally tested that unsupervised, semantic-based splitting strategy allows for a high number (49%) of good quality (2.37 score) sentence splits as compare to supervised systems.

## Summarization beyond sentence extraction: A probabilistic approach to sentence compression [37]

In this paper, Kevin Knight and Daniel Marcu presented corpus based sentence compression algorithms using noisy channel and decision tree approach. The main feature of the algorithm was that the compression generated were grammatical correct and it retained the most important pieces of information. For training the system, Authors used the Ziff–Davis corpus, a collection of newspaper articles announcing computer products. Authors evaluated results against manual compressions and a simple baseline. The results showed that the decision-based algorithm was most aggressive: on average, it compressed sentences to about half of their original size. The compressed sentences produced by both algorithms are more “grammatical” and contain more important words than the sentences produced by the baseline model.

## Paraphrasing exercises and training for Chinese to English consecutive interpreting [38]

Andrew K.F. Cheung, in this research explained his experiment in which he used paraphrased version of the text for machine translation system. He concluded that that Chinese to English translation performance may improve when interpreting is based on a paraphrased version.

## Plagiarism and Paraphrasing Criteria of College and University Professors [39]

In this paper, an experiment is performed by Miguel Roig to check the views of University professors about the rewritten of the text. Results showed that 44% of the sample of professors considers a sample rewritten version as not being a case of plagiarism. The average plagiarism scores obtained by respondents from each sample were analyzed with a one-way between-subjects analysis of variance (ANOVA), but no statistically significant differences were detected, $F(5, 195) = 0.98$, $p = .43$. A one-way between-subjects ANOVA carried out on plagiarism scores of the various academic groupings failed to reach statistical significance, $F(4, 170) = 1.99$, $p = .10$.

## Using Internet based paraphrasing tools: Original work, patch writing or facilitated plagiarism? [40]

In this paper, Ann M. Rogerson et. al. highlight the existence, development, use and detection of use of Internet based paraphrasing tools. Author mentioned that a subsequent quick Google search highlighted the broad range

and availability of online paraphrasing tools which offer free 'services' to paraphrase large sections of text ranging from sentences, paragraphs, whole articles, book chapters or previously written assignments. Two sites are compared to demonstrate the types of differences that exist in the quality of the output from certain paraphrasing algorithms, and the present poor performance of online originality checking services such as Turnitin® to identify and link material processed via machine based paraphrasing tools.

## Challenges in Argumentation and Paraphrasing Among Beginning Students in Educational Sciences [41]

In this research article author discussed various challenges in argumentation and paraphrasing among beginning students in educational science. Author aimed to identify difficulties in writing at the beginning of educational science programmers in the Finnish Open University by analyzing the students' written argumentation and use of sources at the textual level. The results obtained showed that many students began their educational studies with weak writing competencies. While many of the problems were directly related to students' failure to explain the ideas in their sources in their own words, some problems pertained to other aspects, such as the inability to construct convincing arguments.

## Extraction of Drug-Drug Interaction from Literature through Detecting Linguistic-based Negation and Clause Dependency [42]

In this paper, author mainly focuses on simplification of complex sentences to extract Drug-Drug interaction (DDI) for medical literature. Complex sentences were identified by using a kernel based method. Three types of kernels were Global Context Kernel, Local Context Kernel and Subtree kernel. Dependent clauses were identified by using two feature categories.

## Legal Question Answering Using Paraphrasing and Entailment Analysis [43]

In this research article, author describe a system that exploits paraphrasing and sentence-level analysis of queries and legal statutes. Author evaluated this system using the training data from the competition on legal information extraction/entailment (COLIEE)-2016. SVM model was built by heuristic selection of attributes which provides the basis for making a decision on the yes/no questions. Proposed model shows better mean average precision than TF-IDF.

## A Novel Approach to Paraphrase Hindi Sentences using Natural Language Processing [44]

In this article author has discussed sentence reframing technique using NLP. Author proposed a system to convert the existing sentence in different form by remaining the semantic or meaning same. This will helpful in converting the complex sentence into simpler one. System proposed by author mainly deals with Hindi Sentences and its different forms. It takes a sentence as input and produces another sentence without changing its semantic after applying synonyms and antonyms replacement. Further work can be used to change a complex sentence in simplified form. Author implemented this system in java. Wamp Server was used as database.

## Paraphrasing Strategy Instruction for Struggling Readers [45]

This study investigated the effects of the TRAP (Think before you read, Read a paragraph, Ask myself, "What was this paragraph mostly about?" and Put it into my own words) paraphrasing strategy taught using the Self-Regulated Strategy Development model. Participants were seven middle school students identified as fluent readers who experienced difficulty with comprehension. All instruction for the participants was in pairs or small groups. Results indicate that the TRAP strategy increased reading comprehension as measured by the percentage of text recall and short-answer questions. Ideas for future research and implications are discussed.

## Maulik: A Plagiarism Detection Tool for Hindi Documents [46]

In this paper, author discussed an automated plagiarism detection software tool called Maulik. This tool detects plagiarism in Hindi documents. Maulik divides the text into n-grams and then matches it with the text present in repository as well as with documents present online. Author claimed **s**imilarity score of 96.3 which is higher as compared to the existing Hindi plagiarism detection tools such as Plagiarism checker, Plagiarism finder, Plagiarisma, Dupli checker, Quetext.

## Aligning Texts and Knowledge Bases with Semantic Sentence Simplification [47]

In this paper, author presents an approach to build a dataset of triples aligned with equivalent sentences written in natural language. Author performed three main steps: First, target sentences are annotated automatically with knowledge base (KB) concepts and instances. The triples linking these elements in the KB are extracted as candidate facts to be aligned with the annotated sentence. Second, use textual mentions referring to the subject and object of these facts to semantically simplify the target sentence via crowdsourcing. Third, the sentences provided by different contributors are post-processed to keep only the most relevant simplifications for the alignment with KB facts. Datasets developed by author contain 1,050 sentences aligned with 1,885 triples.

## Conclusion:

From above survey, it is concluded that paraphrasing can be performed using various techniques including rule based, statistical based (supervised and Unsupervised), by deleting unnecessary information and by replacing synonyms etc. It is further concluded that not much work has been done for Indian languages except Hindi language. Therefore, existing paraphrasing techniques can be applied on Indian languages for detection and generation of paraphrases.

## References:

[1]. Kevin Knight and Daniel Marcu. 2000. Statisticsbased summarization-step one: Sentence compression. In Proceedings of AAAI-IAAI.

[2]. Trevor Cohn and Mirella Lapata. 2008. Sentence compression beyond word deletion. In Proceedings of COLING.

[3]. Katja Filippova and Michael Strube. 2008. Dependency tree based sentence compression. In Proceedings of INLG

[4]. Emily Pitler. 2010. Methods for sentence compression. Technical report, University of Pennsylvania.

[5]. Katja Filippova, Enrique Alfonseca, Carlos Colmenares, Lukasz Kaiser, and Oriol Vinyals. 2015. Sentence compression by deletion with LSTMs. In Proceedings of EMNLP.

[6]. Kristina Toutanova, Chris Brockett, Ke M. Tran, and Saleema Amershi. 2016. A dataset and evaluation metrics for abstractive compression of sentences and short paragraphs. In Proceedings of EMNLP.

[7]. Kathleen McKeown, Sara Rosenthal, Kapil Thadani, and Coleman Moore. 2010. Time-efficient creation of an accurate sentence fusion corpus. In Proceedings of NAACL-HLT.

[8]. Katja Filippova. 2010. Multi-sentence compression: Finding shortest paths in word graphs. In Proceedings of COLING.

[9]. Kathleen McKeown, Sara Rosenthal, Kapil Thadani, and Coleman Moore. 2010. Time-efficient creation of an accurate sentence fusion corpus. In Proceedings of NAACL-HLT.

[10]. Mark Dras. 1999. Tree adjoining grammar and the reluctant paraphrasing of text. Ph.D. thesis, Macquarie University, Australia

[11]. Regina Barzilay and Kathleen R McKeown. 2001. Extracting paraphrases from a parallel corpus. In Proceedings of ACL.

[12]. Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In Proceedings of ACL.

[13]. Sander Wubben, Antal Van Den Bosch, and Emiel Krahmer. 2010. Paraphrase generation as monolingual translation: Data and evaluation. In Proceedings of INLG.

[14]. Jonathan Mallinson, Rico Sennrich, and Mirella Lapata. 2017. Paraphrasing revisited with neural machine translation. In Proceedings of EACL.

[15]. Advaith Siddharthan. 2010. Complex lexico-syntactic reformulation of sentences using typed dependency representations. In Proceedings of INLG.

[16]. Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In Proceedings of COLING.

[17]. Kristian Woodsend and Mirella Lapata. 2011. Learning to simplify sentences with quasi-synchronous grammar and integer programming. In Proceedings of EMNLP.

[18]. Sander Wubben, Antal van den Bosch, and Emiel Krahmer. 2012. Sentence simplification by monolingual machine translation. In Proceedings of ACL.

[19]. Shashi Narayan and Claire Gardent. 2014. Hybrid simplification using deep semantics and machine translation. In Proceedings of ACL.

[20]. Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in current text simplification research: New data can help. Transactions of the Association for Computational Linguistics, 3:283–297.

[21]. Kristina Toutanova, Chris Brockett, Ke M. Tran, and Saleema Amershi. 2016. A dataset and evaluation metrics for abstractive compression of sentences and short paragraphs. In Proceedings of EMNLP

[22]. Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In Proceedings of EMNLP.

[23]. Bautista, Susana, et al. "An approach to treat numerical information in the text simplification process." Universal Access in the Information Society 16.1 (2017): 85-102.

[24]. Štajner, Sanja, Biljana Drndarevic, and Horacio Saggion. "Corpus-based sentence deletion and split decisions for Spanish text simplification." Computación y Sistemas 17.2 (2013).

[25]. CENTAL, ILC. "Syntactic sentence simplification for French." Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)@ EACL. 2014.

[26]. Inui, Kentaro, et al. "Text simplification for reading assistance: a project note." Proceedings of the second international workshop on Paraphrasing-Volume 16. Association for Computational Linguistics, 2003.

[27]. Ma, Shuming, and Xu Sun. "A semantic relevance based neural network for text summarization and text simplification." arXiv preprint arXiv:1710.02318 (2017).

[28]. Zhang, Xingxing, and Mirella Lapata. "Sentence Simplification with Deep Reinforcement Learning." arXiv preprint arXiv:1703.10931 (2017).

[29]. Lee, John, and J. Buddhika K. Pathirage Don. "Splitting Complex English Sentences." Proceedings of the 15th International Conference on Parsing Technologies. 2017.

[30]. Petersen, Sarah E., and Mari Ostendorf. "Text simplification for language learners: a corpus analysis." Workshop on Speech and Language Technology in Education. 2007.

[31]. Sethi, Nandini, et al. "A novel Approach to Paraphrase Hindi sentences using Natural language Processing." Indian Journal of Science and Technology 9.28 (2016).

[32]. Narayan, Shashi, et al. "Split and rephrase." arXiv preprint arXiv:1707.06971 (2017).

[33]. Wubben, Sander, Antal Van Den Bosch, and Emiel Krahmer. "Paraphrase generation as monolingual translation: Data and evaluation." Proceedings of the 6th International Natural Language Generation Conference. Association for Computational Linguistics, 2010.

[34]. Callison-Burch, C., and C. Bannard. "Paraphrasing with bilingual parallel corpora." Proceedings of 43th Annual Meeting of the Association for Computational Linguistics. 2005.

[35]. Bingel, Joachim, and Anders Søgaard. "Text simplification as tree labeling." The 54th Annual Meeting of the Association for Computational Linguistics. 2016.

[36]. Narayan, Shashi, and Claire Gardent. "Unsupervised sentence simplification using deep semantics." arXiv preprint arXiv:1507.08452 (2015).
[37]. Knight, Kevin, and Daniel Marcu. "Summarization beyond sentence extraction: A probabilistic approach to sentence compression." Artificial Intelligence 139.1 (2002): 91-107.
[38]. Cheung, Andrew KF. "Paraphrasing exercises and training for Chinese to English consecutive interpreting." FORUM. Revue internationaled'interprétation et de traduction/International Journal of Interpretation and Translation. Vol. 14. No. 1. John Benjamins Publishing Company, 2016.Roig, Miguel. "Plagiarism and paraphrasing criteria of college and university professors." Ethics & Behavior 11.3 (2001): 307-323.
[39]. Roig, Miguel. "Plagiarism and paraphrasing criteria of college and university professors." Ethics & Behavior 11.3 (2001): 307-323.
[40]. Rogerson, Ann M., and Grace McCarthy. "Using Internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism?." International Journal for Educational Integrity 13.1 (2017): 2.
[41]. Hyytinen, Heidi, Erika Löfström, and Sari Lindblom-Ylänne. "Challenges in argumentation and paraphrasing among beginning students in educational sciences." Scandinavian Journal of Educational Research 61.4 (2017): 411-429.
[42]. Bokharaeian, B., and A. Diaz. "Extraction of Drug-Drug Interaction from Literature through Detecting Linguistic-based Negation and Clause Dependency." Journal of AI and Data Mining 4.2 (2016): 203-212.
[43]. Kim, Mi-Young, et al. "Legal Question Answering Using Paraphrasing and Entailment Analysis." Tenth International Workshop on Juris-informatics (JURISIN). 2016.
[44]. Sethi, Nandini, et al. "A novel Approach to Paraphrase Hindi sentences using Natural language Processing." Indian Journal of Science and Technology 9.28 (2016).
[45]. Hagaman, Jessica L., Kathryn J. Casey, and Robert Reid. "Paraphrasing strategy instruction for struggling readers." Preventing School Failure: Alternative Education for Children and Youth 60.1 (2016): 43-52.
[46]. Garg, Urvashi, and Vishal Goyal. "Maulik: A Plagiarism Detection Tool for Hindi Documents." Indian Journal of Science and Technology 9.12 (2016).
[47]. Mrabet, Yassine, et al. "Aligning texts and knowledge bases with semantic sentence simplification." (2016): 29-36.