A Framework Components for Natural Language Steganalysis

Roshidi Din, Azman Samsudin, and Puriwat Lertkrai

***Abstract*—In order to have a viable steganalysis method, a steganalysis framework should be developed for natural language. Thus, this paper proposes a new definition of the steganalysis view in order to detect the hidden text used on natural language. This paper also analyzes and classifies several primitive components of natural language steganalysis domain. Primitive components such as resources and techniques of the natural language processing, steganography methods and steganalysis methods are presented in steganalysis environment. Thus, the integration of all these components can be known as a framework of natural language steganalysis which in return contribute to the e-application in security environment.**

***Index Terms*—Natural language steganalysis, steganalysis methods, text steganalysis, steganography methods.**

I. Introduction

One of the concerns in the area of information security is the concept of information hiding. A survey [1] of current information hiding has shown that steganography is one of the recent important subdisciplines of information hiding (see Fig. 1).

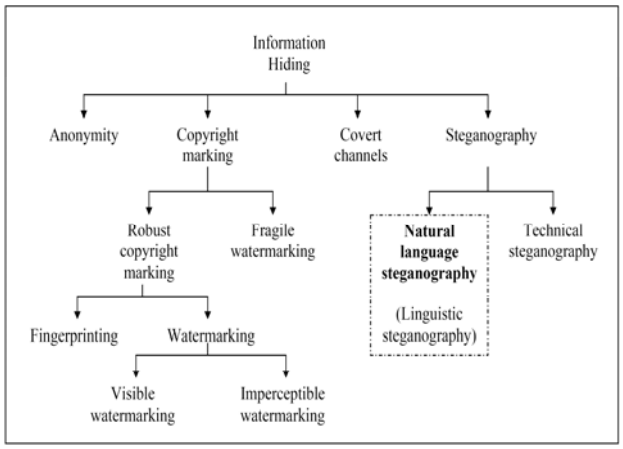

Fig. 1. Discipline areas of information hiding.

To some extent, steganography plays an important role in protecting the security of all documents over the Internet in this era of terabit networks. The goal of steganography is to avoid suspicion on the existence of the hidden messages.

In contrast, steganalysis aims to discover covert messages by rendering useless messages in a given data. The steganalyst starts by reducing a set of suspected information streams to a subset of most likely altered information streams. This is usually done with statistical analysis by using advanced statistic techniques. In contrast, steganalysis aims to discover covert messages by rendering useless messages in a given data. The steganalyst starts by reducing a set of suspected information streams to a subset of most likely altered information streams using advanced statistical analysis. Thus, steganalysis is the process of detecting steganography by looking at variances between bit patterns. In recent years, there has been an increasing interest on natural language steganalysis. This is due to research potentials within this research area particularly in measuring the undetectability of natural language steganography systems. There are two (2) reasons [2] for this:

- methods of message detection are under investigation, and
- general detection of the steganalysis has not been devised

The processes of steganography and steganalysis on natural language can be represented by *Prisoner's Problem* [3]. As show in Fig. 2, Alice is trying to send an original text *M*, within a cover text *C*, which is involving a stego key *K* through an embedding process known as *S*. The first step is applying the invertible function $e: \{M, C\} \rightarrow S$. Then, Alice can map a text *M* to a stego text *S*, using key *K* through $e(M, C) = S$. Since *S* is a stego object, Wendy will not find it suspicious, and since the function is invertible, Bob will be able to compute $e^{-1}(S) = \{M, C\}$ in order to reconstruct the original text *M* and cover text *C* with a stego key *K*. The process might use a function $d: S*C*K \rightarrow M$ to decode the stego text.

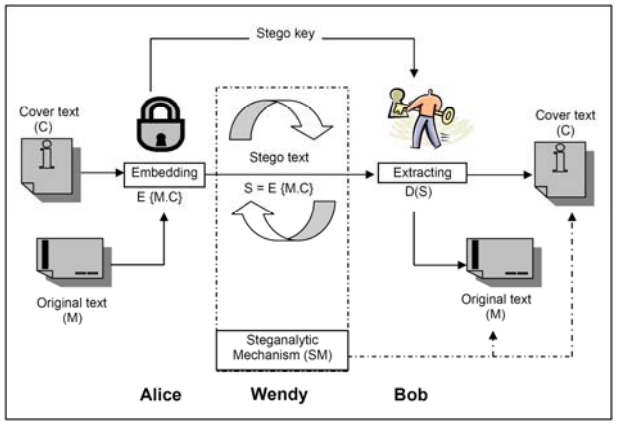

Fig. 2. A steganology processes on natural language environment.

Manuscript received June 5, 2012; revised July 5, 2012. This work was fully supported by PIBT, RIMC, Universiti Utara Malaysia (UUM) and 1001/PKOMP/842025 of USM-RU-PRGS, Universiti Sains Malaysia (USM).

Roshidi Din and Puriwat Lertkrai are with the School of Computing (SoC), CAS Universiti Utara Malaysia (UUM), Malaysia (e-mail: roshidi@uum.edu.my, puriwat_lertkrai@hotmail.com).

Azman Samsudin is with the School of Computer Sciences, Universiti Sains Malaysia (USM), 11800, Pulau Pinang, Malaysia (e-mail: azman@cs.usm.my).

We believe that the process of detecting or extracting the information on stego text *S* by Wendy can be achieved through a mechanism called *steganalytic mechanism* (*SM*). Through this mechanism, any pattern on the stego text can be identified. This mechanism was designed such that a stego text can be manipulated with or without knowing a stego key in order to detect the stego key. Based on this idea, in order to be a viable natural language steganalysis technique, a steganalysis framework has suggested to be developed [4]. Perhaps, it can work at least for a class of natural language steganography. However, the important question is: what are the components supposedly involves in this new framework? As such, the main objective of this paper is therefore to discuss the framework components for natural language steganalysis in which used in the sense of comprehensive set of techniques that allow building steganalysis methods and tools.

Clearly, the problems of developing the framework depend on prominent components such as steganography security and steganography capacity [5], [6] in natural language domain. While steganography security mainly focuses on steganography methods, steganography capacity concerns more on attacking methods of the steganography channel. Currently, steganography methods can be divided into two parts, namely text steganography methods and linguistic steganography. Meanwhile, there are several separate activities of analysis in natural language steganalysis which are divided into several methods. There are statistical attack, rhetorical attack, lexical attack, syntactical attack and semantic attack. Most of these natural language steganalysis methods work by looking at the text patterns on the carrier text. All of these prominent components that are usually used by steganalysts are presented in the following discussion in order to formalize a conceptual framework for the natural language steganalysis.

This paper is organized as follows. Section II and Section III are deals with the components of natural language steganalysis framework such as resources and techniques of the natural language processing, steganography methods and steganalysis methods. Concluding remarks are given in Section IV.

II. Natural Language Steganalysis

There are two types of detecting the information hiding inside a communication medium, namely natural language steganalysis and digital steganalysis. Natural language steganalysis is used to discover the existence of hidden message on complex linguistic structures through various methods. In relatively near future, it is suspected that steganalysis on natural language will pose as an issue. However, most of the available natural language steganalysis methods are comparatively weak compared to the digital steganalysis such as image steganalysis, audio steganalysis, and video steganalysisis which are well established [7].

Therefore, we believe that to be established and successful in steganalytic, a few components of the natural language steganalysis such as resources and techniques of the natural language processing, steganography methods and steganalysis methods should be incorporated. The integration of all these components can be presented as a conceptual framework of natural language steganalysis. Supposedly, all of these components are integrated, it is expected that a good steganalysis technique on natural language can be developed through this framework. The illustration of this proposed framework for natural language steganalysis is presented in Fig. 3.

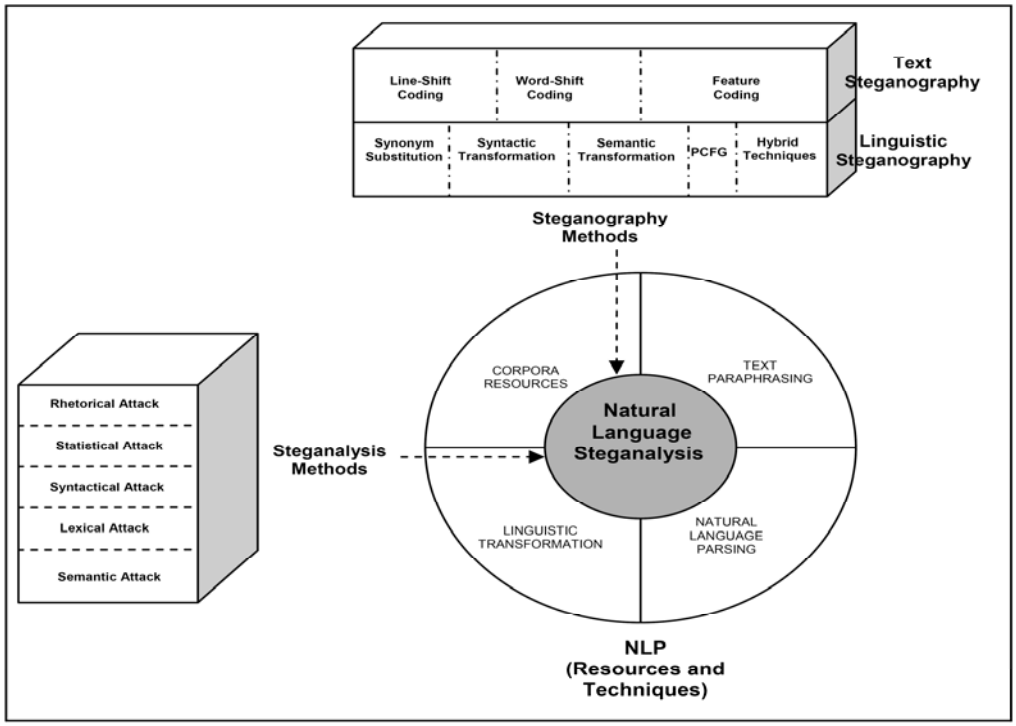

Fig. 3. A conceptual framework of natural language steganalysis.

III. Components of Natural Language Steganalysis Framework

A. Natural Language Processing Resources and Techniques

Natural Language Processing (NLP) has accumulated a great deal of fundamental knowledge in steganalysis environment. Thus, there are four (4) components [8] in NLP that can be considered during the construction of natural language steganalysis framework which are identified as;

1) *Corpora Resources*: Several numbers of electronic corpora are available on Internet that has been created for NLP research such as WordNet [9] and VerbNet [10]. These corpora resources are considered because of their availability and accessibility.
2) *Text Paraphrasing:* One of the challenges in natural language steganalysis is to paraphrase a text in order to detect the hidden message.
3) *Natural Language Parsing:* This task can be described as reprocessing the text sentences and reproducing new structure for the sentences.
4) *Linguistic Transformations:* There are three types of linguistic transformation in natural language environment which are synonym substitution, syntactic transformations, and semantic transformations.

B. Natural Language Steganography Methods

As indicated in Fig. 3, publicly available methods for information hiding within the natural language text can be grouped into two groups. The first group is called text steganography. This method is based on manipulating lines, spaces, characters, or any other features of the given message.

1) Line-shift coding [11].
2) Word-shift coding [12 – 14].
3) Feature coding [15 – 17].

The second group called linguistic steganography is based on linguistically modified cover document in order to encode the message. This is mainly re-writing the document using linguistic transformations such as synonym substitution, syntactic substitution (paraphrasing) or semantic transformation.

1) Using probabilistic context-free grammars to generate cover text [18 - 19].
2) Synonym substitutions [20 – 21].
3) Syntactic transformations [22 - 23].
4) Semantic transformations [24 – 26].
5) Hybrid techniques [27].

C. Natural Language Steganalysis Methods

Currently, there are several analysis methods in natural language steganalysis which are divided into five categories [28]: rhetorical attack, statistical attack, lexical attack, syntactical attack, and semantic attack. Most of these natural language steganalysis methods are based on the text patterns analysis of a natural language. Besides that, natural language steganalysis also tries to find a good pattern combination of expected secret messages in the natural language text itself.

1) *Rhetorical attack:* Attack uses the rhetorical structures of texts as a surface-form-based algorithm.
   - *Rhetorical parsing*: The algorithm uses information that is derived from a corpus analysis of cue phrases that provides the empirical data in order to use the mathematical foundations for rhetorical parsing algorithm [29 – 30].
   - *Decision-based approach*: Marcu [31] presented a shift-reduce rhetorical parsing algorithm based on the decision-based learning techniques [32] that learns to construct rhetorical texts structures from tagged data.
   - *Paragraph segmentation*: Idea of paragraph segmentation is to view a summarization as an extraction of important sentences from a text. Bolshakov and Gelbukh [33] assume that splitting text into paragraphs is determined by text cohesion.
2) Statistical attack
   - *Word-counting*: Several steganalysis approaches [34 – 35] have identified statistically over-represented words through enumeration. These methods are typically fast and able to handle large sequences, and can identify large number of putative motifs.
   - *Dictionary based*: Steganalysis method pioneered by Bussemaker et al. [36] called 'MobyDick' is a dictionary based approach which has introduced the concepts of dictionary and word usage frequencies for constructing sequences. MobyDick algorithm is based on a statistical mechanics model that segments the string probabilistically into words and concurrently builds a dictionary of these words.
   - *Topic identification (TID)*: Main objective of TID is to assign one or several topic labels to a flow of textual data in natural language. Labels are chosen from a set of topics fixed a priori. Bigi's et al. [37] study deals with the evaluation of TID algorithms on two kinds of textual corpora (newspaper and e-mails) using five methods which are topic unigram, cache model, TFIDF classifier, topic perplexity, and weighted model.
   - *Space characters distribution*: Based on Xin-Guang and Hui [38] idea, certain statistical parameters change in a certain range for different types of data or files. It is in fact a choice to analyze the suspicious texts on the point of statistical features. They found that, in a certain text, the space character probability grew when the quantity of embedded message rised, and the continuous space characters probability grew accordingly.
   - *A Dictionary based: wordspy*: Wang et al. [39] has developed an innovative dictionary based on motif finding algorithm for natural language steganalysis, called WordSpy. One significant feature of WordSpy is the combination of a word counting method and a statistical model.
3) Syntactical attack
   - *Key paragraph*: Fukumoto and Suzuki [40] proposed a steganalysis method based on key paragraph in multi-document summarization. Key paragraph shows how to identify differences and similarities across documents. Another study by Stein et al. [41] is a study on paragraph based extraction. Basically, paragraphs which include not only event words but also topic words are considered to be significant paragraphs.
   - *Meaning-preserving transformations*: A birthmark extraction algorithm has been proposed by Yang et al.

[42] as an effective technique used to prevent, discourage, and detect the theft of the natural language text. The proposed birthmark has achieved a relatively strong resilience against the meaning-preserving transformations. Syntactic substitution and semantic substitution have little influence to birthmarks

4) *Lexical attack* - Taskiran et al. [43] proposed a lexical steganalysis which is used as a universal steganalysis method based on language models and support vector machines (SVM) to differentiate sentences modified by a lexical steganography algorithm from unmodified sentences
5) Semantic attack
   - *Paraphrase detection*:Paraphrasing means to be able to express the same meaning in a different way. This subject has recently been receiving an increasing interest [44].
   - *First letters of words distribution*: It is based on analyzing the suspicious texts on the distribution of the first letters of given words [45]. This study assumed that the words of a text originated from a specified dictionary.

Thus, to be successful such system should consider all of these components in order to design a good steganalysis approach on natural language domain.

IV. Conclusion

The primary contribution of this paper is to present the works on natural language steganalysis. This paper also analyzes and classifies several primitive components of steganalysis framework on natural language. It is assumed that a good steganalysis technique on natural language will be produced in a near future through this proposed framework. In particular, a further improvement is expected that a computational intelligence technique will be manipulated into this proposed framework which in return would contribute to other possible applications such as e-business verification, e-document application and cyber security applications.

References

[1] F. A. P Peticolas, R. J. Anderson, and M. G. Kuhn, "Information Hiding - A Survey," in *Proceedings of the IEEE*, July 1999, vol. 87, no. 7, pp. 1062-1078.

[2] N. Johnson and S. Jajodia, "Steganalysis of images created using current steganography software," *Proc. 2nd Information Hiding Workshop*, Springer-Verlag, 1998, pp. 273-289.

[3] G. J. Simmons, *Prisoners' problem and the subliminal channel, Proc. CRYPTO83 - Advances in Cryptology*, August 1984, pp. 51-67.

[4] R. Chandramouli and N. D. Memon, "Steganography capacity: a steganalysis perspective," in *Proceeding SPIE Security and Watermarking of Multimedia Contents*, Jan 2003, vol. 5020, pp. 173 -177.

[5] R. Chandramouli, "A mathematical framework for active steganalysis," *ACM/Springer Multimedia Systems Special Issue on Multimedia Watermarking*, September 2003, vol. 9, pp. 303-311.

[6] M. Sidorov, *Hidden Markov models and steganalysis, Proceedings of the International Multimedia and security Conference*, Association for Computing Machinery (ACM), New York, USA, 2004, pp. 63-67.

[7] R. Din and A. Samsudin, "Digital Steganalysis: Computational Intelligence Approach," *International Journal of Computers*, vol. 3, Issue 1, pp. 161-170, 2009.

[8] M. Topkara, C. M. Taskiran, and E. J. Delp, "Natural language watermarking," Security, Steganography and Watermarking of Multimedia Contents VII, in *Proceedings of the SPIE*, 2005, vol. 5681, pp. 441-452.

[9] C. Fellbaum, *WordNet an electronic lexical database*, MIT Press, 1998.

[10] M. Palmer. Unified Verb Index: VerbNet. [Online]. Available: http://verbs.colorado.edu/verb-index/

[11] J. T. Brassil, L. O'Gorman, N. F. Maxemchuk, and S. H. Low, "Document marking and identification using both line and word shifting," in *Proceedings of the 14th Annual Joint Conference of the IEEE Computer and Communication Societies (INFOCOM)*, Washington DC, USA, April 1995, vol. 2, pp. 853.

[12] D. Huang and H. Yan, "Inter-word distance changes represented by sine waves for watermarking text images," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 12, p. 1237-1245, December 2001.

[13] Y. Kim, K. Moon, and I. Oh, "A text watermarking algorithm based on word classification and inter-word space statistics," in *Proceedings of the 7th International Conference on Document Analysis and Recognition (ICDAR'03)*, IEEE Computer Society Washington, USA, 2003, pp. 775 – 779.

[14] H. Yang and A. C. Kot, "Text document authentication by integrating inter character and word spaces watermarking," *IEEE International Conference on Multimedia and Expo (ICME)*, Taipei, Taiwan, June 27 -30, 2004, vol. 2, pp. 955-958.

[15] X. G. Sui and H. Luo, "A new steganography method based on hypertext," in *proc. IEEE Radio Science Conference*, Asia-Pacific, August 24-27, 2004, pp. 181-184.

[16] X. Sun, G. Luo, and H. Huang, "Component-based digital watermarking of Chinese texts," in *Proceedings of the 3rd ACM International Conference on Information Security*, Shanghai, China, ACM Press, 2004, vol. 85, pp. 76 – 81.

[17] M. H. Shirali-Shahreza and M. Shirali-Shahreza, "A new approach to Persian/Arabic text steganography," in *Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science and 1st IEEE/ACIS*, International Workshop on Component-Based Software Engineering, Software Architecture and Reuse (ICIS-COMSAR'06), July 10-12, 2006, pp. 310-315.

[18] C. Manning and H. Sch¨utze, *Foundations of statistical natural language processing*, MIT Press, 1999.

[19] P. Wayner, *Mimic functions*, Cryptologia, XVI, July 1992, issue. 3, pp. 193-214.

[20] I. A. Bolshakov and A. Gelbukh, "Synonymous paraphrasing using Wordnet and Internet. Natural Language Processing and Information Systems," in *proc. 9th International Conference on Applications of Natural Language to Information Systems (NLDB 2004), of Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, June 2004, vol. 3136, pp. 312-323.

[21] M. Topkara, G. Riccardi, D. Hakkani-Tur, and M. J. Atallah, "Natural language watermarking: challenges in building a practical system," in *Proceeding of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, San Jose, January 2006.

[22] H. Nakagawa, K. Sampei, T. Matsumoto, S. Kawaguchi, K. Makino, and I. Murase, "Text information hiding with preserved meaning – a case for Japanese documents," *Information Processing Society of Japan (IPSJ) Transaction*, vol. 42, no. 9, pp. 2339-2350, 2001.

[23] B. Murphy and C. Vogel, "The syntax of concealment: reliable methods for plain text information hiding," in *Proceeding of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, San Jose, CA, February 2007.

[24] C. Grothoff, K. Grothoff, L. Alkhutova, R. Stutsman, and M. J. Atallah, *Translation-based steganography, Proceedings of Information Hiding Workshop (IH 2005)*, Springer-Verlag, 2005, pp. 213-233.

[25] V. Chand and C. O. Orgun, "Exploiting linguistic features in lexical steganography: design and proof-of-concept implementation," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS '06)*, January 2006, vol. 6, pp. 126b.

[26] B. Macq and O. Vybornova, "A method of text watermarking using presuppositions," in *proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2007, vol. 6505, pp. 65051R.

[27] M. T. Chapman and G. I. Davida, "Hiding the hidden: a software system for concealing ciphertext as innocuous text," in *proceedings of the International Conference on Information and Communication Security, Lecture Notes in Computer Sciences 1334*, Berlin: Springer, 1997, pp. 333 - 345.

[28] K. Bennett, "Linguistic steganography: survey, analysis, and robustness concerns for hiding information in text," *CERIAS Tech*

Report 2004-13, Center for Education and Research in Information Assurance and Security, Purdue University, West Lafayette, 2004.

[29] D. Marcu, "Building up rhetorical structure trees," in *proceedings of the 13th National Conference on Artificial Intelligence (AAAI–96)*, Portland, Oregon, August 4-8, 1996, vol. 2, pp. 1069-1074.

[30] D. Marcu, "The rhetorical parsing of natural language texts," in *proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL'97/EACL'97)*, Madrid, Spain, July 7 - 12, 1997, pp. 96 – 103.

[31] D. Marcu, "A decision-based approach to rhetorical parsing," in *proceeding of the 37th annual meeting of the Association for Computational Linguistics (ACL '99)*, Maryland, June 1999, pp. 365 – 372.

[32] U. Hermjakob and R. Mooney, "Learning parse and translation decisions from examples with rich context," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL)*, 1997, pp. 482 – 489.

[33] I. A. Bolshakov and A. Gelbukh, "Text segmentation into paragraph based on local text cohesion," in *proceedings of the 4th International Conference on Text, Speech and Dialogue, Lecture Notes In Computer Science*, Springer-Verlag, London, UK, 2001, vol. 2166, pp. 158-166.

[34] J. Van Helden, B. Andre, and J. Collado-Vides, "Extracting regulatory sites from the upstream region of yeast genes by computational analysis of *Oligonucleotide* frequencies," *J. Mol. Biol.*, 1998, vol. 266, pp. 231-245.

[35] S. Sinha and M. Tompa, "A statistical method for finding transcription factor binding sites," in *proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology (ISMB 2000)*, Price Center, University of California San Diego, La Jolla, CA, August 19-23, 2000, pp. 344-354.

[36] H. J. Bussemaker, H. Li, and E. D. Siggia, "Building a dictionary for genomes: identification of presumptive regulatory sites by statistical analysis," in *Proceedings of the Natl. Academy Science USA*, 2000, pp. 10096-10100.

[37] B. Bigi, A. Brun, J. P. Haton, K. Smail, and I. Zitouni, "A comparative study of topic identification on newspaper and e-mail," in *String processing and Information Retrieval (SPIRE 2001), Proceedings of the 8th International Symposium*, November 13-15, 2001, pp. 238- 241.

[38] S. Xin-Guang and L. Hui, "A steganalysis method based on the distribution of space characters," in *proceedings of Communications, Circuits and Systems International Conference*, Guilin, Guangzi, China, June 2006, vol. 1, pp. 54 - 56.

[39] G. Wang, T. Yu, and W. Zhang, *WordSpy: identifying transcription actor binding motifs by building a dictionary and learning a grammar*, Nucleic Acids, 2005, Res 33, pp. 412-416.

[40] F. Fukumoto and Y. Suzuki, *Extracting key paragraph based on topic and event detection - towards multi-document summarization*, ANLP/NAACL Workshops, NAACL-ANLP 2000 Workshop on Automatic summarization, Seattle, Washington, 2000, vol. 4, pp. 31-39.

[41] G. C. Stein, T. Strzalkowski, and G. B. Wise, "Summarizing multiple documents using text extraction and interactive clustering," in *proceeding of the Pacific Association for Computational Linguistics*, Waterloo, Canada, August 1999, pp. 200-208.

[42] J. Yang, J. Wang, and D. Li, "Detecting the theft of natural language text using birthmark," in *proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, December 2006, pp. 699-702.

[43] C. M. Taskiran, U. Topkara, M. Topkara, and E. J. Delp, "Attacks on lexical natural language steganography systems," in *proceeding of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, San Jose, CA, February 15-19, 2006.

[44] N. Kaji, D. Kawahara, S. Kurohashi, and S. Sato, "Verb paraphrase based on case frame alignment," in *proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia*, USA, 2002, pp. 215 - 222.

[45] S. Xin-Guang, L. Hui, and Z. Zhong-Liang, "A steganalysis method based on the distribution of first letters of words," in *proc. Intelligent Information Hiding and Multimedia Signal Processing International Conference, (IIH-MSP '06)*, Pasadena, CA, USA, December 2006, pp. 369 - 372.