Journal of Theoretical and Applied Information Technology
10th January 2014. Vol. 59 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A REVIEW ON ABSTRACTIVE SUMMARIZATION METHODS

[1]ATIF KHAN, [2]NAOMIE SALIM
[1,2]Faculty of Computing, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia
E-mail: [1]katif2@utm.my, [2]naomie@utm.my

ABSTRACT

Text summarization is the process of extracting salient information from the source text and to present that information to the user in the form of summary. It is very difficult for human beings to manually summarize large documents of text. Automatic abstractive summarization provides the required solution but it is a challenging task because it requires deeper analysis of text. In this paper, a survey on abstractive text summarization methods has been presented. Abstractive summarization methods are classified into two categories i.e. structured based approach and semantic based approach. The main idea behind these methods has been discussed. Besides the main idea, the strengths and weaknesses of each method have also been highlighted. Some open research issues in abstractive summarization have been identified and will address for future research. Finally, it is concluded from the literature studies that most of the abstractive summarization methods produces highly coherent, cohesive, information rich and less redundant summary.

Keywords: *Abstractive Summary, Sentence Fusion, Semantic Graph, Abstraction Scheme, Sentence Revision*

1. INTRODUCTION

The data on World Wide Web is growing at an exponential pace. Nowadays, people use the internet to find information through information retrieval (IR) tools such as Google, Yahoo, Bing and so on. However, with the exponential growth of information on the internet, information abstraction or summary of the retrieved results has become necessary for users. In the current era of information overload, text summarization has become an important and timely tool for user to quickly understand the large volume of information. The goal of automatic text summarization is to condense the documents into a shorter version and preserve important contents.

A document summary keeps its main content, helps user to understand and interpret large volume of text in the document and consequently reduce user's time for finding the key information in the document. Summarization, as done by humans, involves reading and understanding an article, website, or document to find the key points. The key points are then used to generate new sentences, which form the summary. For humans, generating a summary is a straightforward process but it is time consuming. Therefore, the need for automated summaries is becoming more and more apparent to automatically generate the summary and get the general idea of long textual data[1].

One of the natural questions to ask in summarization is "What are the texts that should be represented or kept in a summary?" The summary must be generated by selecting the important contents or conclusions in the original text. Finding out important information becomes a truly challenging task. Currently, the need for automatic text summarization has appeared in many areas such as news articles summary, email summary, short message news on mobile, and information summary for businessman, government officials, research, online search engines to receive the summary of pages found and so on[2].

The first effort on automatic text summarization system was made in the late 1950.This automatic summarizer selects significant sentences from the document and concatenates them together. The approach in [3] uses term frequencies to measure sentence relevance. Sentences are included in the summary if the terms in the sentences have high term frequencies.

Text summarization approaches can be broadly divided into two groups: extractive summarization and abstractive summarization[4].Extractive summarization extracts salient sentences or phrases

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

from the source documents and group them to produce a summary without changing the source text. Usually, sentences are in the same order as in the original document text. However, abstractive summarization consists of understanding the source text by using linguistic method to interpret and examine the text. The abstractive summarization aims to produce a generalized summary, conveying in information in a concise way, and usually requires advanced language generation and compression techniques[5].Early work in summarization started with single document summarization. Single document summarization produces summary of one document. As research proceeded, and due to large amount of information on web, multi document summarization emerged. Multi document summarization produces summaries from many source documents on the same topic or same event.

A distinction is made between indicative summary and informative summary based on the content of the summary. Indicative summaries are used to indicate what topics are discussed in the source text and they give brief idea of what the original text is about. On the other hand, informative summaries elaborate the topics in the source text [1].

The automatic summarization of text is a well-known task in the field of natural language processing (NLP). Significant achievements in text summarization have been obtained using sentence extraction and statistical analysis. True abstractive summarization is a dream of researchers [1]. Abstractive methods need a deeper analysis of the text. These methods have the ability to generate new sentences, which improves the focus of a summary, reduce its redundancy and keeps a good compression rate [6].

This paper presents the current state of art in the abstractive summarization methods along with the strengths and weaknesses of each method. The paper is organized as follows: Section 2 presents the definition, aim and features of automatic text summarization. Section 3 gives a comprehensive literature review of the abstractive summarization methods and these methods have been compared along three dimensions: original text representation, content selection and summary generation as shown in Table 1.Section 4 discusses the open research issues in abstractive text summarization. Discussion is given in section 5 and section 6 concludes the paper.

2. AUTOMATIC TEXT SUMMARIZATION

Text Summarization is shorter version of the original document while still preserving the main content available in the source documents. There are various definitions on text summary in the literature. According to [7] "The aim of automatic text summarization is to condense the source text by extracting its most important content that meets a user's or application needs". According to [8],"a summary is a text that is produced from one or more texts that contains a significant portion of the information text(s), and is no longer than half of the original text(s)".

Generally, a summary should be much shorter than the source text. This characteristic is defined by the compression rate, which measures the ratio of length of summary to the length of original text [4].

2.1 Text Summarization Features

Text summarization identify and extract key sentences from the source text and concatenate them to form a concsie summary . In order to identify key sentences for summary, a list features as discussed below,can be used to for selection of key sentences.

Term Frequecy: Statistics provide salient terms based on term frequency,thus salient sentences are the ones that contain the words that occur frequently[21].The score of sentences increases for each frequent word. The most common measure widely used to calculate the word frequency is TF IDF.

Location: It relies on the intuition that important sentences are located at certain position in text or in paragraph, such as beginning or end of a paragraph[9].

Cue Method: Words that would have positive or negtive effect on the respective sentence weight to indicate significance or key idea[7] such as cues: "in sumary", "in conclusion","the paper describes","significantly".

Title/Headline word: It assumes that words in title and heading of a document that occur in sentences are positvely relevant to summarization[9].

Sentence length: Short sentences express less information and therfore excluded from summary.keeping in view the size of summary,very long sentences are also not appropriate for summary[10].

Similarity: This feature determines similarity between the sentence and the rest of the document

Journal of Theoretical and Applied Information Technology
10th January 2014. Vol. 59 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sentences and similarity between the sentence and title of the document.Similarity can be calculated with linguistic knowledge or by character string overlap[9].

Proper noun: Sentences having proper nouns are considered impotant for document summary. Examples of proper nouns are: name of a person, place or organization[10] .

Proximity: The distance between text units where entities occur is a determining factor for establsihing relations between entities[10].

3. ABSTRACTIVE SUMMARIZATION METHODS

Most of the work in text summarization has focused on extractive summarization, which forms summary by selection of important sentences from the documents[11]. Statistical methods are often used to find key words and phrases [3]. Discourse structure[12] also assists in specifying the most important sentences in the document. Various machine learning techniques [13] have been applied for extracting features for salient sentences using training corpus. A few research works have addressed single and multi document abstractive summarization in academia. Different approaches and systems for single and multi document abstractive summarization have discussed in literature. In this study, we focus particularly on seven methods for abstractive summarization. First, we discuss the main idea of each method,followed by relevant research work for each method,and finally the strengths and weaknesses of each method are discussed.

Abstractive summarization techniques are broadly classified into two categories: Structured based approach and Semantic based approach. Different methods that use structured based approach are as follows: tree base method, template based method, ontology based method, lead and body phrase method and rule based method. Methods that use semantic based approach are as follows: Multimodal Semantic model, Information item based method, and semantic graph based method. Table 1 compares all the abstractive summarization methods based on the following parameters: Original text representation, Content selection and Summary generation.

3.1. Structured Based Approach

Structured based approach encodes most important information from the document(s) through cognitive schemas [6] such as templates, extraction rules and other structures such as tree, ontology, lead and body phrase structure. Different methods used this approach are discussed as follows.

3.1.1. Tree based method

This technique uses a dependency tree to represent the text/contents of a document. Different algorithms are used for content selection for summary e.g. theme intersection algorithm or algorithm that uses local alignment across pair of parsed sentences. The technique uses either a language generator or an algorithm for generation of summary. Related literature using this method is as follows.

The approach proposed in [14] automatically fuse similar sentences across news articles on the same event. The method uses language generation for producing concise summary. In this approach, first the similar sentences are preprocessed using a shallow parser and then sentences are mapped to predicate-argument structure. Next, the content planner uses theme intersection algorithm to determine common phrases by comparing the predicate-argument structures. Those phrases that convey common information are selected and ordered and some information are also added with it(temporal references, entity descriptions).Finally sentence generation phase uses FUF/SURGE language generator to combine and arrange the selected phrases into new summary sentences. The major strength of this approach is that the use of language generator significantly improved the quality of resultant summaries i.e. reducing repetitions and increasing fluency. The problem with this approach is that context of sentence was not included while capturing the intersected phrase. Context is important even if it is not a part of intersection.

In other work, sentence fusion [15] integrates information in overlapping sentences to generate a non overlapping summary sentence. In this approach, first the dependency trees are obtained by analyzing the sentences. A basis tree is set by finding the centroid of the dependency trees. It next augments the basis tree with the sub-trees in other sentences and finally prunes the predefined constituents. The limitation of this approach is that it lacks a complete model which would include an abstract representation for content selection.

3.1.2. Template based method

This technique uses a template to represent a whole document. Linguistic patterns or extraction rules are matched to identify text snippets that will be mapped into template slots. These text snippets are

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Shows A Comparative Study On Abstractive Summarization Methods Based On Text Representation, Content Selection And Summary Generation.

Author/Year	Techniques/ Methods	Text Representation	Content Selection	Summary Generation
Barzilay and McKeown, 1999	Tree based	Dependency based representation: DSYNT tree	Theme intersection algorithm	FUF/SURGE language Generator
Harabagiu and Lacatusu ,2002	Template based	Template/Frame having slots and fillers	Linguistic patterns or Extraction rules	IE based MD summarization Algorithm
Lee and Jian, 2005	Ontology based	Fuzzy ontology	Classifier	News agent
Barzilay and McKeown, 2005	Tree based	Dependency tree	Algorithm uses local alignment across pair of parsed sentences	Algorithm for reusing and altering phrases from input sentences
Tanaka and Kinoshita, 2009	Lead and Body phrase	Lead, body and supplement structure	Revision candidates (Maximum phrases of same head in lead and body sentences)	Insertion and substitution operations on phrases
Greenbacker, 2011	Multimodal Semantic model	Semantic model	Information density(ID) metric	Generation technique: Synchronous tree
Genest and Lapalme, 2011	INIT based	Abstract representation: Information item (INIT)	Ranking of generated sentences based on document frequency	NLG realizer SimpleNLG
Genest and Lapalme, 2012	Rule based	Categories and Aspects	Extraction rules	Generation patterns
Moawad and Aref ,2012	Semantic Graph based	Rich semantic graph	$\mathbf{C}_{\mathbf{weight}}$ =Average weight of each concept. $\mathbf{S}_{\mathbf{weight}}$ =Average weight of all concepts in a sentence.	Reduced semantic graph and domain ontology

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

are indicators of the summary content. Related literature using this method is as follows. The approach proposed in[16] presents a multi-document summarization system, GISTEXTER, which produces abstract summaries of multiple newswire/newspaper documents relying on the output of the CICERO Information extraction(IE) system. To extract information from multiple documents, CICERO requires a template representation of the topic for extracting information from multiple documents. In this approach, a topic is represented as a set of related concepts and implemented as a frame or template containing slots and fillers. The templates are filled with important text snippets extracted by the Information Extraction systems. These text snippets are used to generate coherent, informative multi-document summaries by using IE based multi document summarization algorithm. A significant advantage of this approach is that the generated summary is highly coherent because it relies on relevant information identified by IE system. This approach works only if the summary sentences are already present in the source documents. It cannot handle the task if multi document summarization requires information about similarities and differences across multiple documents.

3.1.3. Ontology based method

Many researchers have made effort to use ontology(knowledge base) to improve the process of summarization. Most documents on the web are domain related because they discuss the same topic or event. Each domain has its own knowledge structure and that can be better represented by ontology. Related literature using this method is discussed as follows.

The fuzzy ontology with fuzzy concepts is introduced for Chinese news summarization [17]to model uncertain information and hence can better describe the domain knowledge. In this approach, first the domain experts define the domain ontology for news events. Next, the document preprocessing phase produces the meaningful terms from the news corpus and the Chinese news dictionary. Then, term classifier classifies the meaningful terms on the basis of events of news. For each fuzzy concept of the fuzzy ontology, the fuzzy inference phase generates the membership degrees. A set of membership degrees of each fuzzy concept is associated with various events of the domain ontology. News summarization is done by news agent based on fuzzy ontology. The benefit of this approach is that it exploits fuzzy ontology to handle uncertain data that simple domain ontology cannot. This approach has several limitations. First, domain ontology, Chinese dictionary and news corpus has to be defined by a domain expert which is time consuming. Secondly, this approach is limited to Chinese news, and might not be applicable to English news.

3.1.4. Lead and body phrase method

This method is based on the operations of phrases (insertion and substitution) that have same syntactic head chunk in the lead and body sentences in order to rewrite the lead sentence. Related literature using this method is discussed as follows.

An abstractive approach proposed by [18] revise lead sentences in a news broadcast. This approach does not use the co-reference relation of noun phrases (NPs).In this approach, first the same chunks (also called triggers) are searched in the lead and body sentences. Then, maximum phrases (revision candidates) of each trigger are identified and aligned using similarity metric. Substitution of body phrase for the lead phrase takes place if body phrase has corresponding phrase in the lead and body phrase is richer in information. Insertion of body phrase into the lead sentence takes place, if a body phrase has no counterpart in the lead sentence. The potential benefit of this method is that it found semantically appropriate revisions for revising a lead sentence. This method has some weaknesses. First, Parsing errors degrade sentential completeness such as grammaticality and repetition. Secondly, it focuses on rewriting techniques, and lacks a complete model which would include an abstract representation for content selection.

3.1.5. Rule based method

In this method, the documents to be summarized are represented in terms of categories and a list of aspects. Content selection module selects the best candidate among the ones generated by information extraction rules to answer one or more aspects of a category. Finally, generation patterns are used for generation of summary sentences.

The methodology in [19] generates short and well written abstractive summaries from clusters of news articles on same event. The methodology is based on an abstraction scheme. The abstraction scheme uses a rule based information extraction module, content selection heuristics and one or more patterns for sentence generation. Each abstraction scheme deals with one theme or subcategory. In order to generate extraction rules for abstraction scheme,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

several verbs and nouns having similar meaning are determined and syntactic position of roles is also identified. The information extraction (IE) module finds several candidate rules for each aspect of the category. Based on the output of the IE module, the content selection module selects the best candidate rule for each aspect and passed it to summary generation module. This module form summary of text using generation patterns designed for each abstraction scheme. The strong point of this method is that it has a potential for creating summaries with greater information density than current state of art.

The main drawback of this methodology is that all the rules and patterns are manually written, which is tedious and time consuming.

3.2. Semantic Based Approach

In Semantic based method, semantic representation of document(s) is used to feed into natural language generation (NLG) system. This method focus on identifying noun phrases and verb phrases by processing linguistic data[20]. Different methods using this approach are discussed here.

3.2.1. Multimodal semantic model

In this method, a semantic model, which captures concepts and relationship among concepts, is built to represent the contents (text and images) of multimodal documents. The important concepts are rated based on some measure and finally the selected concepts are expressed as sentences to form summary.

In [21], a framework was proposed for generating an abstractive summary from a semantic model of a multimodal document. Multimodal document contains both text and images. The framework has three steps: In first step, a semantic model is constructed using knowledge representation based on objects (concepts) organized by ontology. In second step, informational content (concepts) is rated based on information density metric. The metric determines the relevance of concepts based on completeness of attributes, the number of relationships with other concepts and the number of expressions showing the occurrence of concept in the current document. In third step, the important concepts are expressed as sentences. The expressions observed by the parser are stored in a semantic model for expressing concepts and relationship. An important advantage of this framework is that it produces abstract summary, whose coverage is excellent because it includes salient textual and graphical content from the entire document. The limitation of this framework is that it is manually evaluated by humans. An automatic evaluation of the framework is desirable.

3.2.2. Information item based method

In this method**,** the contents of summary are generated from abstract representation of source documents, rather than from sentences of source documents. The abstract representation is Information Item, which is the smallest element of coherent information in a text.

A framework proposed in[6] for abstractive summarization took place in the context of Text Analysis Conference(TAC) 2010 for multi-document summarization of news. The framework consists of following modules: Information Item retrieval, sentence generation, sentence selection and summary generation. In Information Item (INIT) retrieval, first syntactic analysis of text is done with parser and the verb's subject and object are extracted. So, an INIT is defined as a dated and located subject–verb–object triple. In sentence generation module, a sentence is directly generated from INIT using a language generator, the NLG realizer SimpleNLG [22]. Sentence selection module ranks the sentences generated from INIT based on their average Document Frequency (DF) score. Finally, a summary generation step account for the planning stage and include dates and locations for the highly ranked generated sentences.

The major strength of this approach is that it produces short, coherent, information rich and less redundant summary. This approach has several limitations. First, many candidate information items are rejected due to the difficulty of creating meaningful and grammatical sentences from them. Secondly, linguistic quality of summaries is very low due to incorrect parses.

3.2.3. Semantic Graph Based Method

This method aims to summarize a document by creating a semantic graph called Rich Semantic Graph (RSG) for the original document, reducing the generated semantic graph, and then generating the final abstractive summary from the reduced semantic graph.

The abstractive approach proposed by[23] consists of three phases as shown in figure 1. The first Phase represents the input document semantically using Rich Semantic Graph (RSG). In RSG, the verbs and nouns of the input document are represented as graph nodes along with edges corresponding to semantic and topological relations

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

between them. The second phase reduces the generated rich semantic graph of the source document to more reduced graph using some heuristic rules. Finally, the third Phase generates the abstractive summary from the reduced rich semantic graph. This phase accepts a semantic representation in the form of RSG and generates the summarized text.

A noteworthy strength of this method is that it produces concise, coherent and less redundant and grammatically correct sentences. However this method is limited to single document abstractive summarization.

4. OPEN RESEARCH ISSUES

Automatic Text summarization has become an integral part of daily life due to the availability of large volume of information, that need to be summarized for humans so that they can read important contents in short time. We focus on abstractive summarization, which is a challenging research area because of the complexity of natural language processing. Some of the research issues in abstractive summarization methods that need to be addressed have been identified from the literature and are given as follows:

- There is no generalized framework that humans can use for abstractive summarization.
- Tree based method for abstractive summary relies on parsing and alignment of parse trees, and in this respect, its robustness is an issue.
- Future research in **tree based** and **leads & body phrase** methods for abstractive summary will need to further address grammaticality of summary sentences.
- Besides synthesizing important sentences for abstractive summary, sentence ordering in a summary is an important research issue.
- Ongoing research on abstractive summarization must still deal with issues such as scarcity of training data, appropriate integration of syntax even when the input data comes from a noisy genre, and compressions involving lexical substitution and paraphrase.
- Evaluating an abstractive summary is a difficult task because there does not exist an ideal summary for a given document or set of documents and therefore is an open research area.
- The biggest challenge for abstractive summary is the representation. Systems' capabilities are constrained by the richness of their representations and their ability to generate such structures.

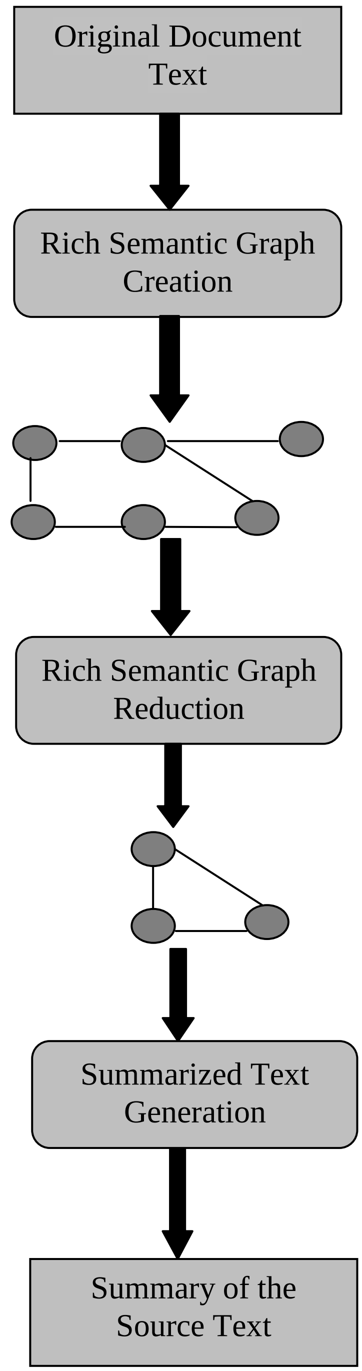

Figure 1: Semantic Graph Reduction for Abstractive Text Summarization

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. DISCUSSION

All the abstractive summarization methods are classified into two categories: structured based approach and semantic based approach.

Almost all the abstractive summarization methods under structured based approach improves the quality of summaries i.e. produces short, coherent, information rich and less redundant summary with the exception of lead and body phrase method, which produces summary with redundant sentences. However, the linguistic quality of summaries produced by all abstractive summary methods is very low. The summary sentences contain a lot of grammatical mistakes. This is due to the fact that abstractive summary methods in structured based approach do not rely on semantic representation of the original document text.

The abstractive summarization methods under semantic based approach rely on semantic representation of the original document text. These methods produce concise, information rich, coherent, and less redundant summary as well as improve the linguistic quality of summary[23] with the exception of Information item based method, in which a few summary sentences contains grammatical mistakes due to incorrect parses. A better parser will resolve this issue. The semantic based abstractive methods proposed by[23] and [6]uses rich semantic graph and abstract representation i.e. Information Item (INIT) to represent the original document. These methods produce concise, coherent, information rich and less redundant summary.

6. CONCLUSION

Automatic text summarization is an old challenge but now the research direction is leaning from extractive text summarization to abstractive text summarization. Abstractive summary methods produces highly coherent, cohesive, information rich and less redundant summary. Abstractive text summarization is a challenging area because of the complexity of natural language processing. Therefore, this study examines a review on abstractive summarization methods along with their strengths and weaknesses. The different methods are also compared based on three parameters: text representation, content selection and summary generation. Different abstractive summarization systems are explored and finally some research issues that will be addressed as future work are also identified. Certainly, this study has been adapted in a way that new researchers to the area of text summarization can get a better understanding on abstractive text summarization approaches.

ACKNOWLEDGEMENT

This research is supported by University Teknologi Malaysia.

REFRENCES:

[1] H. Saggion and T. Poibeau, "Automatic text summarization: Past, present and future," in *Multi-source, Multilingual Information Extraction and Summarization*, ed: Springer, 2013, pp. 3-21.

[2] M. Haque, *et al.*, "Literature Review of Automatic Multiple Documents Text Summarization," *International Journal of Innovation and Applied Studies*, vol. 3, pp. 121-129, 2013.

[3] H. P. Luhn, "The automatic creation of literature abstracts," *IBM Journal of Research and Development*, vol. 2, pp. 159-165, 1958.

[4] D. R. Radev, *et al.*, "Introduction to the special issue on summarization," *Computational Linguistics*, vol. 28, pp. 399-408, 2002.

[5] D. Das and A. F. Martins, "A survey on automatic text summarization," *Literature Survey for the Language and Statistics II course at CMU*, vol. 4, pp. 192-195, 2007.

[6] P.E. Genest and G. Lapalme, "Framework for abstractive summarization using text-to-text generation," in *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, 2011, pp. 64-73.

[7] H. P. Edmundson, "New methods in automatic extracting," *Journal of the ACM (JACM)*, vol. 16, pp. 264-285, 1969.

[8] E. Hovy and D. Marcu, "Automated text summarization," *The Oxford handbook of computational linguistics*, pp. 583-598, 2005.

[9] L. Suanmali, *et al.*, "Fuzzy logic based method for improving text summarization," *arXiv preprint arXiv:0906.4690*, 2009.

[10] Y. J. Kumar and N. Salim, "Automatic multi document summarization approaches," *Journal of Computer Science*, vol. 8, p. 133, 2011.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[11] J. Kupiec, *et al.*, "A trainable document summarizer," in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, 1995, pp. 68-73.

[12] K. Knight and D. Marcu, "Statistics-based summarization-step one: Sentence compression," in *Proceedings of the National Conference on Artificial Intelligence*, 2000, pp. 703-710.

[13] B. Larsen, "A trainable summarizer with knowledge acquired from robust NLP techniques," *Advances in Automatic Text Summarization,* p. 71, 1999.

[14] R. Barzilay, *et al.*, "Information fusion in the context of multi-document summarization," in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, 1999, pp. 550-557.

[15] R. Barzilay and K. R. McKeown, "Sentence fusion for multidocument news summarization," *Computational Linguistics,* vol. 31, pp. 297-328, 2005.

[16] S. M. Harabagiu and F. Lacatusu, "Generating single and multi-document summaries with gistexter," in *Document Understanding Conferences*, 2002.

[17] C.-S. Lee, *et al.*, "A fuzzy ontology and its application to news summarization," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 35, pp. 859-880, 2005.

[18] H. Tanaka, *et al.*, "Syntax-driven sentence revision for broadcast news summarization," in *Proceedings of the 2009 Workshop on Language Generation and Summarisation*, 2009, pp. 39-47.

[19] P.-E. Genest and G. Lapalme, "Fully abstractive approach to guided summarization," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, 2012, pp. 354-358.

[20] H. Saggion and G. Lapalme, "Generating indicative-informative summaries with sumUM," *Computational Linguistics,* vol. 28, pp. 497-526, 2002.

[21] C. F. Greenbacker, "Towards a framework for abstractive summarization of multimodal documents," *ACL HLT 2011*, p. 75, 2011.

[22] A. Gatt and E. Reiter, "SimpleNLG: A realisation engine for practical applications," in *Proceedings of the 12th European Workshop on Natural Language Generation*, 2009, pp. 90-93.

[23] I. F. Moawad and M. Aref, "Semantic graph reduction approach for abstractive Text Summarization," in *Computer Engineering & Systems (ICCES), 2012 Seventh International Conference on*, 2012, pp. 132-138.