Extractive Summarization for Myanmar Language

Soe Soe Lwin
University of Computer Studies,Yangon
Yangon,Myanmar
soesoelwin@ucsy.edu.mm

Khin Thandar Nwet
University of Information Technology
Yangon,Myanmar
khin.thandarnwet@gmail.com

Abstract—**Due to increasing availability of online information, tools and mechanisms for automatic summarization of documents is needed. Text summarization is currently a major research topic in Natural Language Processing. There are various approaches to generate text summary. Among them, we proposed Myanmar text summarization using latent semantic analysis (LSA). Latent semantic analysis (LSA) is a technique in natural language processing,and can analyze relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. It is an unsupervised approach which does not need any traning or external knowledge. There is no LSA based sentence extraction in Myanmar language. This is the first LSA based Text Summarizer in Myanmar. This paper present generic, extractve and single-document Myanmar text summarization using latent semantic analysis. This paper compare two sentence selection methods (Steinberger and Jezek's approach and Ozay approach) of latent semantic analysis to extract important sentences. We summarize Myanmar news from Myanmar official websites such as 7day daily, iyarwaddy, etc..**

Keywords—*LSA, text summarization*

I. INTRODUCTION

With the massive amount of information, people are faced with the problem of finding relevant information efficiently and effectively. Summarization is essential technology to solve this problem. Text summarization is the creation of a shortened version of a text by a computer program. It extracts the most important information of the original text.

Summarization system can be divided into two categories: extractive and abstractive summarization. Extractive summarization extracts salient words/sentences from documents and group them to produce summary without changing [8]. Abstractive summarization examines the source text and generate the concise summary that contains some novel sentences not present in original document.

Summarization methods can be categorized according to what they generate and how they generate it. A summary can be extracted from a single document or from multiple documents. If a summary is generated from a single document, it is known as single document summarization. On the other hand, if a single summary is generated from multiple documents on the same subject, this is known as multi -document summarization. Summaries are also divided into generic summaries and query-based summaries. Generic summarization systems generate summaries containing main topics of documents. In query-based summarization, the generated summaries contain the sentences that are related to the given queries.

There are several research projects concerned with automatic text summarization for English and European languages and in Asian languages such as Chinese and Japanese. However, there is little ongoing research in Myanmar text summarization. For Myanmar Language, template driven automatic text summarization and query-based text summarization has been proposed. There is no LSA-based text summarizer in Myanmar. In this paper, generic extractive Myanmar text summarization system based on LSA is proposed.

II. RELATED WORK

Researchers have been working actively on text summarization within the Natural Language Processing (NLP) to create a better and more efficient summary. The researchers have been developing for summarization using latent semantic analysis method to achieve better summary.

W.T.Kyaw[10] used CRF (Conditional Random Field) to summarize Myanmar disaster news that is based on Information Extraction. This paper proposed multi-documents, query-focused, extractive and informative Myanmar text Summarization framework. It is not sentence extraction. It is word level summary concerned with natural disasters such as date, time, Place. Our system extracts important sentence based on LSA.

M. T. Naing proposed Summary generation system is based on semantics roles. Automatic prominal anaphora resolution in Myanmar text is used for summary generation. In sematic role labeling, argument identification and argument classification are developed using Myanmar Verb Frame Resource. The system cannot identify verbs in sentences with two main verbs [11].

Gong and Liu [5] Proposed summarization of news with the use of LSA as a way to identify the significant topics in the documents. SVD is applied to matrix A to decompose into three matrices as follows: $A = U \Sigma VT$. They proposed that the row of the matrix VT can be considered as various topics covered in the original text. And finally, they reproduce each row of matrix VT successively and extract a sentence from it which has maximum values. They select one sentence for each topic according to topic importance.

M.G. Ozsoy, I. Cicekli and F.N. Ferda Nur Alpaslan [2] explained LSA-based summarization algorithms and evaluated on Turkish and English document. While creating summaries using LSA, there are various approaches for selection of sentences. Among them, they compared five algorithms: Gong and liu, Steinberger and Jezeck, Murry, cross method and topic method. Cross method and topic method are proposed by authors of that paper. They showed that among LSA-based approaches, the cross method performs better than other approaches. They observed that cross method does not perform well in shorter documents.

J.Steinberger and K.Jezek [6] described a generic text summarization method which used the latent semantic analysis technique to identify semantically important sentences. Summarization ratio is 20%. They suggested two new evaluation methods based on LSA, which measure content similarity between an original document and its summary. These two evaluation methods are: similarity of the main topic and similarity of the term significance.

III. LATENT SEMANTIC ANALYSIS(LSA)

LSA is an algebraic method, which can analyze relations between terms and sentences of a given set of documents. LSA uses context of the input document and extracts information such as which words are used together and which common words are seen in different sentences. High number of common words among sentences indicates that the sentences are semantically related [2]. It uses SVD (Singular Value decomposition) for decomposing matrices. SVD is a numerical process, which is often used for data reduction, but also for classification, searching in documents and for text summarization [3].

There are three main steps in Latent Semantic Analysis. These steps are as follows:

- Input Matrix Creation.
- Singular Value Decomposition.
- Sentence Selection.

A. Input Matrix Creation

The input document is represented in a matrix form to perform the calculations. A matrix is created which represents the input text. The row of the matrix represents the words in the sentences and column represents the sentences of the input document. The cells of matrix represent the importance of words in sentences. There are various methods to represent importance of words. These approaches are:

Number of Occurrence: The cell is filled with frequency of the word in the sentence.
Binary Representation: If a word occurs in the sentence the cell is filled 1, otherwise the cell value is 0.
Root Type: If the root type of the word is Noun, cell value is the frequency of the word, otherwise the cell value is 0.
Term Frequency-Inverse Document Frequency:

$$TF=\frac{\text{Frequency of word i in sentence j}}{\text{Sum of frequencies of all words in sentence j}}$$

$$IDF=\log\frac{\text{Number of sentences in input text}}{\text{Number of sentences containing word I}}$$

Modified tf-idf: Cell value are first calculated with tf-idf value. If cell value is less than or equal to average TF-IDF values in the associated row, then set them zero.

B. Singular Value Decomposition

LSA process the term-sentence matrix through the algorithm called Singular Value Decomposition (SVD) [7]. SVD is a method of word co-occurrence analysis. It is based on theorem from linear algebra. The input matrix A is decomposed into three matrices U, V, Σ. It is shown in equation (1).

$$A=U \Sigma V \tag{1}$$

U= word×concept matrix

Σ=Scaling values, diagonal descending matrix

V= sentence×concept matrix

C. Sentence selection

Using the results of SVD, different algorithms use different approaches to select important sentences. Different algorithms are proposed to select important sentences from the document for summarization using the results of SVD [2]. These algorithms are Gong and lius approach [5], Steinbergerand Iezek'sApproach [6] and Ozsoy'sapproach [2]. In this paper, two methods are used to extract the importance sentence: Steinberger and Jezek's approach and Ozay approach.

IV. PROPOSED SYSTEM

This section presents our proposed Myanmar text summarization system. Figure 1 describes overall architecture of proposed system.
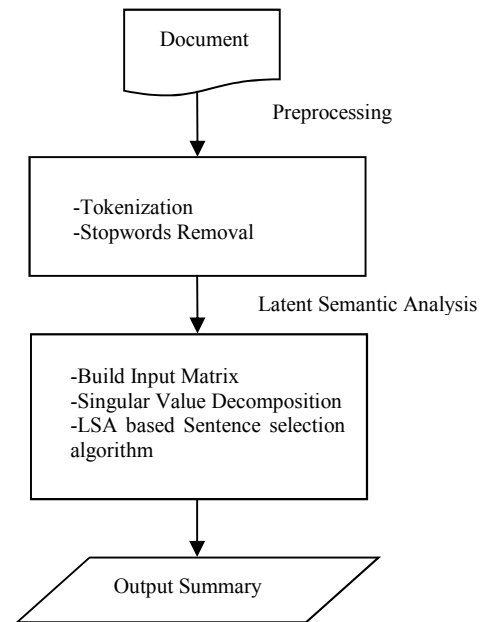

Fig. 1. proposed system for text summarization

A. Word Segmentation

Word segmentation is the process of determining word boundaries in a text. In English language, words boundaries can be easily determined by space. In Myanmar language does not have white space between words. In this paper, Word segmentation is performed Myanmar Word Segmentation using a Combined Model [9].

B. Stopwords Removal

Stopwords do not represent the main idea of the input text. Stopwords in the input document is removed. Example of Stopwords are described in fig.2.

မှ,မည့်,ကို,တွင်,မည်,ဟု,က,ပြီး,မှာ,ပါ,တော့,တယ်,ဘူး,သည်,များ,သည့်,ကာ, သာ,ဘဲ,ခွဲ,သ၊ဖြင့်,တွေ့,နဲ့,ရဲ့,အတွက်,ဖြင့်,ရန်,ထား,သော်လည်း,ရှိ

Fig.2. Example of stopwords inMyanmar

C. Term-Sentence Matrix Creation

Input matrix is created based on the source document where columns represent sentence and rows represent terms. Frequency of the terms (the terms in the figure which mean that patient, car, township, engine, officer, Nyi Nyi Zaw, say, carry, team, Monywa) in the sentence is filled in cell values of matrix. Input matrix is shown in table I.

TABLE I.   SAMPLE TERM-SENTENCE MATRIX OF DOCUMENT

	S1	S2	S3
လူနာ	1	0	1
ယာဉ်	1	1	1
မြို့နယ်	1	0	0
အင်ဂျင်	1	0	1
ဦးစီးမှူး	1	0	0
ညီညီဇော်	1	0	1
ပြော	1	0	0
သယ်ဆောင်	1	0	0
အသင်း	1	1	0
မုံရွာ	1	1	1

D. Singular Value Decomposition

For modeling the relationship among words/phrases and sentences, singular value decomposition is used. SVD is used to reduce dimension of term-by-document matrix.

E. Sentence Selection

After creating input matrix and singular value decomposition of the matrix, sentence selection is performed to generate output summary. In this paper, Steinberger and Jezek's approach and Cross Method are used for sentence selection.

Steinberger and Jezek's approach use V matrix and $\sum$ for sentence selection. Sentence length is calculated using V and $\sum$ as shown in equation 2. The dimension (n) of new space is given as a parameter. Only the concepts which indexes are less than or equal to the given dimension is used to calculate the length. To get the most important concepts, $\sum$ matrix is used as a multiplication parameter. The sentences with the highest length values are selected to be part of the summary [6].

$$\text{length} = \sqrt{\sum_{j=1}^{n} V_{ij} * \sum_{jj}} \tag{2}$$

In Ozay approach, cross method is used for sentence selection. In cross method, the first two steps of the LSA algorithm are executed, V^T matrix are used for sentence selection. The V^T matrix is preprocessed to remove the sentence that are not one of the important sentences for each concept before selecting sentence. The cell value which are less than or equal to average value is set to 0. cell values are multiplied by with the values in $\sum$ matrix. The

length of sentence vector is calculated by summing up all concepts in columns of V^T matrix, which represent the sentences. The sentence with highest length value is selected for summary [2].

V.   EXPERIMENTAL WORK

Local news and international news are used for experiment. Summarization is done by Latent semantic analysis and performance measure is shown in this section.

A. Data used for Experiment

We collected Myanmar local news and international news from Myanmar news websites [12], [13], [14]. Now, we collected 55 local News and 76 international news from these websites. We will collect more data news in the future. Table II shows data set used for experiment.

TABLE II. DATA SET USED FOR EXPERIMENT

	Local news	International News
Number of documents	55	76
Sentence per documents	275	456
Words per Document	2860	4100

B. Performance Measure and Result

Evaluation of automatic summarization in a standard and inexpensive way is a difficult task. There is a set of metrics to automatically evaluate summaries. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is the most widely used metric for evaluating automatic summarization of texts.

It determines the quality of a summary by comparing an automatically produced summary (system summary) against a set of reference summaries (human summary). ROUGE-1 refers to overlap of unigrams between the system summary and human summary. ROUGE-2 refers to the overlap of bigrams between the system and human summaries.

The table III indicates the ROUGE-2, scores on news data using Steinberger and Jezek's approach and cross Method. In this paper, ROUGE 2.0 – A Java Package for Automatic Summary Evaluation [15] is used for evaluating summaries. It can be observed that rouge score of cross method is better than the Steinberger and Jezek's approach for Myanmar language.

TABLE III.EVALUATION BY ROUGE-2 SCORES

	Rouge-2 scores		
	Precision	Recall	F-mesaure
Steinbergerand Jezek'sApproach	0.66	0.58	0.61
Cross Method	0.71	0.60	0.65

VI. CONCLUSION

Understanding a huge document without any abstract or summary is difficult and takes lot of time. This problem is solved with the automatic text summarizer. We proposed the Myanmar text Summarizer using latent semantic analysis model. We compared two sentence selection approach of latent semantic analysis (lsa). This paper mainly concentrates on single document. In future work, we will focus on better sentence selection algorithm to get better performance.

Appendix

TABLE IV. SUMMARIES OF NEWS DATA USING LSA

Input Text
ရန်ကုန် လေဆိပ် မှ ဘူတာကြီး အထိ ပြေးဆွဲမည့် YBS (Airport Shuttle) အမြန် ယာဉ် လိုင်း ကို ဇွန်လ ၂၂ ရက် တွင် စတင် ပြေး ဆွဲ မည် ဖြစ်ပြီး ယင်း နေ့ တွင် ခရီးသည်များ ကို အခမဲ့ စီးနင်း ခွင့်ပြုမည် ဟု Omni Focus ကုမ္ပဏီ မန် နေ ဂျင်း ဒါရိုက်တာ ဦးထွန်းတောက်ဝင်း က ပြောသည် ။
"YBS (Airport Shuttle) will run from Yangon airport to station, it will firstly run on June 22 and then passengers will be allowed a free ride", said U Tun Tauk Win, the director of Omni Focus company.
ရန်ကုန် လေဆိပ် မှ ပြည်လမ်း ၊ ကမ္ဘာအေး ဘုရား လမ်း အတိုင်း မောင်းနှင် ပြီး ရန်ကုန် ဘူတာကြီး တွင် ဂိတ် ဆုံး မည့် YBS (Airport Shuttle) အမြန် ယာဉ် လိုင်း ယာဉ် စီး ခ ကို ၎င်းရာ ကျပ် သတ်မှတ် ထားကြောင်း ၎င်း က ဆိုသည် ။
"The bus will run from Yangon airport to Pyay road, through KabarAye pagoda road and terminate at station and the bus fare is 500 kyats ", He said.
ပြည်တွင်း ပြည်ပ ခရီးသည်များ အများဆုံး သွားရောက် သည့် တန်ခိုးကြီး ဘုရား များ ၊ စျေးဝယ် စင်တာ ၊ ဟိုတယ် ၊ ရုပ်ရှင်ရုံ နေရာများ ကို မှတ်တိုင် အဖြစ် သတ် မှတ်ထား ကာ မှတ်တိုင် ၁၆ ခု တွင် သာ ခရီးသည် အ တင် အချ ပြုလုပ်မည် ဟု ဦးထွန်းတောက်ဝင်း က ပြောသည် ။
"The famous pagodas, shopping centre, hotels and cinemas which local and foreigners mostly visit are marked as bus stop and there are only 16 bus stop", U Tun Tauk Win said.
ခရီးဆောင် အိတ် များ ကို တန်ဆာခ မ ယူ ဘဲ တင်ဆောင် ခွင့်ပြု ပြီး သတ်မှတ်ထားသည့် မှတ်တိုင် များ ၌ သာလူ အ တင်အချ ပြုလုပ် သဖြင့် ခရီးသည်များ အနေဖြင့် အချိန်ကုန် သက်သာ ပြီး အဆင်ပြေပြေ သွားလာနိုင်မည် ဟု ၎င်း က ပြောသည် ။
"The passenger can carry luggage without paying fee and passengers can travel conveniently and save the time ", he said.
"ယာဉ်မောင်း တွေ အနေနဲ့ ဝန်ဆောင်မှု ၊ ကား ရဲ့ သန့်ရှင်းရေး က အစ ဝန်ဆောင်မှု နဲ့ ပတ်သက်ပြီးတော့ မှ ပြစ်ချက်မ ရှိအောင် လုပ်ဆောင် ထား တယ် ။ အဲ့ဒီ အတွက် ဝေဖန်မှု တစ်ခုခု ရှိခဲ့ ရင်လည်း အတတ်နိုင်ဆုံး အကောင်းဆုံး ဖြေရှင်း ပေး သွားမှာပါ" ဟု ဦးထွန်းတောက်ဝင်း က ပြောသည် ။
"We performed the best concerned with driver's service and car's cleaning service and if I had any criticism, I solve with the best possible solutions", he said.
အမြန် ယာဉ် လိုင်း ပြေး ဆွဲ ရာတွင် ကား အ စီး ရေ ၅၀ ဖြင့် ပြေး ဆွဲ ရန် မူလ က စီစဉ် ထား သော်လည်း ရပ်နား ရန် နေရာ အခက်အခဲ ရှိ သဖြင့် အ စီး ရေ ၃၀ ဖြင့် စတင် ပြေး ဆွဲ မည် ဟု ၎င်း က ဆိုသည်။
"We originally planned that 50 numbers of bus will run but the difficulty of parking problems, only 30 numbers of bus will run", he said.

Human summary

ရန်ကုန် လေဆိပ် မှ ဘူတာကြီး အထိ ပြေးဆွဲမည့် YBS (Airport Shuttle) အမြန် ယာဉ် လိုင်း ကို ဇွန်လ ၂၂ ရက် တွင် စတင် ပြေး ဆွဲ မည် ဖြစ်ပြီး ယင်း နေ့ တွင် ခရီးသည်များ ကို အခမဲ့ စီးနင်း ခွင့်ပြုမည် ဟု Omni Focus ကုမ္ပဏီ မန် နေ ဂျင်း ဒါရိုက်တာ ဦးထွန်းတောက်ဝင်း က ပြောသည် ။

"YBS (Airport Shuttle) will run from Yangon airport to station, it will firstly run on June 22 and then passengers will be allowed a free ride", said U Tun Tauk Win, the director of Omni Focus company.

ရန်ကုန် လေဆိပ် မှ ပြည်လမ်း ၊ ကမ္ဘာအေး ဘုရား လမ်း အတိုင်း မောင်းနှင် ပြီး ရန်ကုန် ဘူတာကြီး တွင် ဂိတ် ဆုံး မည့် YBS (Airport Shuttle) အမြန် ယာဉ် လိုင်း ယာဉ် စီး ခ ကို ၄၀၀ ကျပ် သတ်မှတ် ထားကြောင်း ရင်း က ဆိုသည် ။

"The bus will run from Yangon airport to Pyay road, through KabarAye pagoda road and terminate at station and the bus fare is 500 kyats ", He said.

ပြည်တွင်း ပြည်ပ ခရီးသည်များ အများဆုံး သွားရောက် သည့် တန်ခိုးကြီး ဘုရား များ ၊ ဈေးဝယ် စင်တာ ၊ ဟိုတယ် ၊ ရုပ်ရှင်ရုံ နေရာများ ကို မှတ်တိုင် အဖြစ် သတ် မှတ်ထား ကာ မှတ်တိုင် ၁၆ ခု တွင် သာ ခရီးသည် အ တင် အချ ပြုလုပ်မည် ဟု ဦးထွန်းတောက်ဝင်း က ပြောသည် ။

"The famous pagodas, shopping centre, hotels and cinemas which local and foreigners mostly visit are marked as bus stop and there are only 16 bus stop", U Tun Tauk Win said.

LSA Steinberger:

ရန်ကုန် လေဆိပ် မှ ဘူတာကြီး အထိ ပြေးဆွဲမည့် YBS (Airport Shuttle) အမြန် ယာဉ် လိုင်း ကို ဇွန်လ ၂၂ ရက် တွင် စတင် ပြေး ဆွဲ မည် ဖြစ်ပြီး ယင်း နေ့ တွင် ခရီးသည်များ ကို အခမဲ့ စီးနင်း ခွင့်ပြုမည် ဟု Omni Focus ကုမ္ပဏီ မန် နေ ဂျင်း ဒါရိုက်တာ ဦးထွန်းတောက်ဝင်း က ပြောသည် ။

"YBS (Airport Shuttle) will run from Yangon airport to station, it will firstly run on June 22 and then passengers will be allowed a free ride", said U Tun Tauk Win, the director of Omni Focus company.

"ယာဉ်မောင်း တွေ အနေနဲ့ ဝန်ဆောင်မှု ၊ ကား ရဲ့ သန့်ရှင်းရေး က အစ ဝန်ဆောင်မှု နဲ့ ပတ်သက်ပြီးတော့ မှ ပြင်ချက်မ ရှိအောင် လုပ်ဆောင် ထား တယ် ။ အဲဒီ အတွက် ဝေဖန်မှု တစ်ခုခု ရှိခဲ့ ရင်လည်း အတတ်နိုင်ဆုံး အကောင်းဆုံး ဖြေရှင်း ပေး သွားမှာ့ပါ" ဟု ဦးထွန်းတောက်ဝင်း က ပြောသည် ။

"We performed the best concerned with driver's service and car's cleaning service and if I had any criticism, I solve with the best possible solutions", he said.

အမြန် ယာဉ် လိုင်း ပြေး ဆွဲ ရာတွင် ကား အ စီး ရေ ၅၀ ဖြင့် ပြေး ဆွဲ ရန် မူလ က စီစဉ် ထား သော်လည်း ရပ်နား ရန် နေရာ အခက်အခဲ ရှိ သဖြင့် အ စီး ရေ ၃၀ ဖြင့် စတင် ပြေး ဆွဲ မည် ဟု ရင်း က ဆိုသည်။

"We originally planned that 50 numbers of bus will run but the difficulty of parking problems, only 30 numbers of bus will run", he said.

LSA Oz soy:

ရန်ကုန် လေဆိပ် မှ ဘူတာကြီး အထိ ပြေးဆွဲမည့် YBS (Airport Shuttle) အမြန် ယာဉ် လိုင်း ကို ဇွန်လ ၂၂ ရက် တွင် စတင် ပြေး ဆွဲ မည် ဖြစ်ပြီး ယင်း နေ့ တွင် ခရီးသည်များ ကို အခမဲ့ စီးနင်း ခွင့်ပြုမည် ဟု Omni Focus ကုမ္ပဏီ မန် နေ ဂျင်း ဒါရိုက်တာ ဦးထွန်းတောက်ဝင်း က ပြောသည် ။

"YBS (Airport Shuttle) will run from Yangon airport to station, it will firstly run on June 22 and then passengers will be allowed a free ride", said U Tun Tauk Win, the director of Omni Focus company.

ပြည်တွင်း ပြည်ပ ခရီးသည်များ အများဆုံး သွားရောက် သည့် တန်ခိုးကြီး ဘုရား များ ၊ ဈေးဝယ် စင်တာ ၊ ဟိုတယ် ၊ ရုပ်ရှင်ရုံ နေရာများ ကို မှတ်တိုင် အဖြစ် သတ် မှတ်ထား ကာ မှတ်တိုင် ၁၆ ခု တွင် သာ ခရီးသည် အ တင် အချ ပြုလုပ်မည် ဟု ဦးထွန်းတောက်ဝင်း က ပြောသည် ။

"The famous pagodas, shopping centre, hotels and cinemas which local and foreigners mostly visit are marked as bus stop and there are only 16 bus stop", U Tun Tauk Win said.

ခရီးဆောင် အိတ် များ ကို တန်ဆာခ မ ယူ ဘဲ တင်ဆောင် ခွင့်ပြု ပြီး သတ်မှတ်ထားသည့် မှတ်တိုင် များ ၌ သာလူ အ တင်အချ ပြုလုပ် သဖြင့် ခရီးသည်များ အနေဖြင့် အချိန်ကုန် သက်သာ ပြီး အဆင်ပြေပြေ သွားလာနိုင်မည် ဟု ရင်း က ပြောသည် ။

"The passenger can carry luggage without paying fee and passengers can travel conveniently and save the time ", he said.

VII. REFERENCES

[1] P.V.Reddy, "Text Summarization Using Latent Semantic Analysis", International Journal of Technology and Engineering Science [IJTES], Volume 1[8], pp: 1309-1313, November 2013

[2] M.G. Ozsoy, I. Cicekli, F.N. Ferda Nur Alpaslan, "Text Summarization of Turkish Texts using Latent Semantic Analysis", Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 869–876, Beijing, August 2010

[3] M. Campr, K. Ježek, "Comparative summarization via Latent Semantic Analysis", Department of Computer Science and Engineering University of West Bohemia

[4] http://calculator.vhex.net/calculator/linear algebra/singular-value-decomposition

[5] G.Yong, X.Liu, "Generic text summarization using relevance measure and latent semantic analysis". In: Proceedings of SIGIR'01 2001.

[6] J.Steinberger, K.Jezek, "Using Latent Semantic Analysis in Text Summarization and Summary Evaluation", Department of Computer Science and Engineering

[7] S.A.Babar, S.A.Thorat, "Improving Text Summarization using Fuzzy Logic & Latent Semantic Analysis", Computer Science & Engineering, Rajarambapu Institute of Technology, Sakharale, India, International Journal of Innovative Research in Advanced Engineering (IJIRAE), Volume 1 Issue 4 (May 2014)

[8] J. Kamala Geetha, N. Deepamala, "Kannada text summarization using Latent Semantic Analysis", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015

[9] W.P. Pa, N.L.Thein, "Myanmar Word Segmentation using a Combined Model", e-Case2009, Singapore, January, 2009

[10] W.T.Kyaw, N.L.Thein and H.H.Htay, "Automatic Myanmar Text Summarization System", Proceeding of the 12th International Conference on Computer Applications (ICCA 2014),Yangon, Myanmar

[11] M. T. Naing, A. Thida, "Automatic Myanmar Text Summarization System with Semantic roles", in the Proceedings of the 12th International Conference on Computer Applications (ICCA2014), Yangon, Myanmar, February2014, p. 217-223

[12] www.7daydaily.com

[13] www.iyarwaddy.com

[14] www.moi.gov.mm

[15] http://rxnlp.com/rouge-2-0/#.W7xFO2gzaUk