Effective Graph Based Vertex Cover Algorithm Using Multi-document Node Summarizations

Mohammad Rashid Hussain
Center for Artificial Intelligence (CAI)
King Khalid University
College of Computer Science
King Khalid University
Abha, Saudi Arabia
humohammad@kku.edu.sa

Ayman Qahmash
Center for Artificial Intelligence (CAI)
King Khalid University
College of Computer Science
King Khalid University
Abha, Saudi Arabia
a.qahmash@kku.edu.sa

Salem Alelyani
Center for Artificial Intelligence (CAI)
King Khalid University
College of Computer Science
King Khalid University
Abha, Saudi Arabia
s.alelyani@kku.edu.sa

Mohammed Saleh Alsaqer
Center for Artificial Intelligence (CAI)
King Khalid University
College of Computer Science
King Khalid University
Abha, Saudi Arabia
msalsaqr@kku.edu.sa

2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) | 978-1-6654-1974-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/CSDE50874.2020.9411596

***Abstract*— The extraction of multi-document summarization (MDS) has become the dominant technique to offer a compact form of documents that manages the opposite characteristic of the real documents. This technique finds the subset of similarity from the input documents (words/phrases/sentences) in the original text to form the summary in contrast. This paper has illustrated upon the similarity between the MDS problem with the vertex cover problem (VCP) and some important factors for the effective implementation of vertex cover algorithm (VCA) identified through a systematic review of the literature using SMART (Specific, Measurable, Assignable, Realistic, Time related) target. Further, they are categorized into five fields namely text summarization, MDS, Automatic text summarization, Sentence extraction, and Graph-based summarization and the lesson learned from these studies are discussed. The results of the research demonstrate the effectiveness of the proposed method in graph-based automatic multi-document summarization.**

Keywords*— *Multi-document summarization, Vertex cover algorithm, Maximal Marginal Relevance, Graph based summarization.

I. Introduction

The frequent growth and different approach of extraction of MDS have enhanced the usage and become the dominant technique, development of stemming algorithm mechanical translation and computational linguistic are the approaches of development of document extraction technique Reference [1]. Document summarization has no exception, due to its vast range capacity as a development model with a lot of approaches for knowledge distribution from different existing approaches; it has grown up into a widely accepted approach, inferring strategies for sentence ordering in multi-document new summarization [2]. Frequency of occurrence of the word in the sentencing measure of the importance of the word in a sentence, term weighting approaches in text retrieval (automatic) that is a part of machine learning and data mining "Automatic data summarization" i.e. a process of shorting of text documents with software [3]. The automatic data summarization can improve the process of extraction of multi-document by using Maximal Marginal Relevance (MMR), diversity-based ranking for recording documents and producing summaries [4]. It can also improve the concept of text summarization (TS) technique uses a sentence extraction technique using dynamic conference based summarization [5]. The changing and development scenarios of the technology putting more effects over researchers for higher thinking over existing technology, MDS by Automatic summarization (AS), sentence extraction, and centroid based summarization of multiple documents [6][7][8]. Using centrality TS in the pathfinder networks may be one of the solutions for getting extraction of MDS using a graphical approach for fulfilling present demand with effective techniques. The graphical concept can help in the development of automatic data summarization, which is the process of shortening text documents with software [9]. Another effective approach, NEO-CORTEX: A Performing user-oriented MDS system, and ranking sentences for extraction based summarization (EBS) using feature weight propagation in the sentence similarity network [10][11]. There are two alternative ordering strategies, chronological ordering (CO) and Majority ordering (MO) which perform for different categories of text either for a similar organization of information or different. In general MO per-form better than CO. In case of query-relevant MDS, Maximal Marginal Relevance (MMR) ranking provides maximizes redundancy. The effectiveness of the system through the MMR approach has been proposed [4]. Query relevant MDS concepts have been developed through the concept of dynamic conference based summarization. MMR rank text chunks according to their dissimilarity to one another are discussed [5]. This paper introduce the graph based automatic MDS system using VCA, vertex cover of graph represents subset of its vertices which can cover all the edges of the graph. Vertex cover graph is a set of vertices in which every edge in the graph is incident to at least one of the chosen vertices, vertex cover property is exactly to extractive MDS process that select limited set of related sentences which cover the entire concepts which is also covered through all sentences in the input documents. This paper is organized into the following different sections: In section 2, Review of related research. In section 3, the proposed vertex cover based MDS system is presented. In section 4, the investigation results and analysis. In Section 5, the conclusion is summarized.

II. REVIEW OF RELATED RESEARCH

TABLE I. REVIEW OF IMPORTANT FACTORS AND FIELDS

Important Factors	Fields	References
Summarization	Clustering algorithm	[12][13]
	Query based	[14][15]
	Genetic algorithm	[16] to [19]
	Fuzzy logic	[20] to [29]
	Evolutionary algorithm	[30]
	Graph based	[31]
	MCMR	[32]
Multi-document summarization	Clustering algorithm	[48] to [51]
	Query based	[33] to [44]
	Genetic algorithm	[45] to [47]
	Fuzzy logic	[56]
	Graph based	[53][54]
	Centroid based	[52]
	Rhetoric based	[55]
	CDDS	[57]
Automatic text summarization	Clustering algorithm	[63][64]
	Genetic algorithm	[65]
	Fuzzy logic	[61][62]
	Evolutionary algorithm	[65]
Sentence extraction	Fuzzy logic	[58][59]
	Genetic algorithm	[60]
Graph based summarization	Graph Sum	[67]
	Graph model	[68][69]
	Graph based vertex cover algorithm	Proposed one

Text Summarization refers to the approach of minimizing long pieces of text. The aim is to create coherent and fluent summary which have only the main points outlined in the document. Extraction based Summarization (EBS) technique involves putting key phrase from the source document and combining them to make a summary. The extraction is made according to the defined metric without making any changes to the text [13] [14]. An automatic procedure aimed at extraction of information from multiple text written about the same topic [33][48]. An automatic summarization is to construct a incisive and coherent summary of one or more documents, sentence extraction is used for automatic summarization of text [58] [60] [63]. Graph based summarization (GBS) entails extracting a worthwhile subset of sentences from a selection of textual documents by using a graph based model [68] to represent the correlation between pair of document terms [67]. Weighted minimum vertex cover (wMVC) graph based approach have applied greedy algorithm for selection of sentences that are well connected and cover all the contents of the sentences [69].

III. GRAPH BASED AUTOMATIC MDS SYSTEM

Summarization is a subset of data (documents, images, videos, etc.)

Automatic summarization is a process of shortening text documents and the task of generating a summary through software, which saves a lot of time for people who deal with a huge amount of textual information.

Document summarization tries to create a representative summary or abstract of the entire set of documents by finding the most informative document.

Automatic data summarization is a part of machine learning and data mining.

EBS is an approach of identification of sentences from the real text and organizes them in a way to create a coherent text which contains important information. Extractions select a subset of existing words/phrases/sentences in the original text to form the summary in contrast. Extraction is an application and systems for summarization and it is of two types (i) Generic summarization (ii) Query relevant summarization/ Query-based summarization/image collection summarization. TS technique uses a sentence extraction technique.

MDS extraction problem finds the subset of similarity from the input document in the original text to form the summary in contrast.

Extractive MDS procedure selects a limited set of related documents that cover the whole concepts which are covered by all documents in the input set of documents.

Graph-based automatic MDS system using VCA, vector cover of the graph represents a subset of its vertices which can cover all the edges of the graph.

Vector cover of a graph in a set of designated vertices such that each and every edge in the graph is incident to at least one of the appropriated vertices.

A. *Conceptual similarity between MDS and vertex cover summarization (VCS)*

- Vertex of the graph ↔Documents in the set of input documents represents
- Edge exists in between pair of vertices ↔ similarity between corresponding documents in > 0.
- Subset of vertex of graph ↔ subset of documents from the input set of documents.
- MDS problem can be transformed to a vector cover problem.

B. *MDS is performed by the steps:Document transformation, sentence extraction and Graph modeling by VCA*

- Document transformation (Feature selection):

Text categorization and automatic summarization n-grams are always used as features, where n is the number of words in sequence. Different methods are used to distinguish unimportant and important features. Discard unimportant and assign a weight to important ones. Mostly used techniques are text computed using document frequency and inverse document frequency and log-likelihood ratio. These approaches are statistical information to differentiate the word from the set of documents. Frequency occurrence of text in the documents measures how important a text in the document.

$$Weight(d_w) = t_f(d_w) * id_f(w) = \frac{t_c d_w}{|d_w|} * \log \frac{|D_N|}{|D_N(w)+1|}$$

$t_c d_w$:Number of times word w happen in document d.

$|d_w|$: Words in document d.

$|D_N|$: N Number of documents in a group of documents $D_N \to S_i = \{S_1, S_2, \ldots S_n\}$

$|D_N(w)|$: Number of documents in D_N that contains word w.

If word w reputedly occurs more often in a group of documents DN i.e. related to given topic then a set of non-relevant document D'.

- Similarity Measures (Content Selection):

In extractive summarization, either corpus-based or knowledge-based measurement can be used to identify redundancy between the textual contents.

Our proposed approach (Vertex covering) is more effective for corpus oriented metrics which always make use of vector representation.

To measure the similarity between two vectors in multidimensional space, either to use Euclidian distance or angle distance for contextual representation to check the number of common terms in the documents.

Relation between two vector context $\vec{p}$ and $\vec{q}$

$$Similarity(\vec{p}, \vec{q}) = 1/\{1 + sqrt(\sum_{i=1}^{n} (p_i - q_i)^2\}$$

The cosine similarity between two vector

$$Similarity(\vec{p}, \vec{q}) = (\vec{p} \cdot \vec{q})/(|p|\,|q|)$$

$$p \to d_i, q \to d_k, \forall i, k, \quad i \,\&\, k \to 1 \; to \; n$$

n is the total number of documents in document set D_N

Let d_i is the i^{th} document in the document set and n is the total number of document set.

$$d_i = tw_{i1}, tw_{i2}, tw_{i3} \ldots tw_{im,} \quad where, tw_{ij,} \to$$

Weight of word t_j in the document d_i and m is the number of unique stemmed words,

$$\cos(d_i, d_k) = \frac{\overrightarrow{d_i} . \overrightarrow{d_k}}{|\overrightarrow{d_i}|.|\overrightarrow{d_k}|} = \frac{\sum_{j=1}^{m} tw_{ij} , tw_{ik}}{\sqrt{\sum_{j=1}^{m} tw^2{}_{ij} * \sum_{j=1}^{m} tw^2{}_{kj}}}, \forall i, k$$

$\to 1 \; to \; n, where$ tw_{ij}, tw_{ik} indicates weight of j^{th} word in d_i and d_k document respectively.

Now, summary documents are extracted using VCA.

Document extraction through VCA (Covering of Graph):

- A set of edges that covers a graph G is said to be an edge covering or covering of G.
- Any spanning sub-graph of a graph is covering of G.
- Hamiltonian circuit of a graph is a covering of G.
- A covering is called minimal covering if removal of one edge from the covering does not cover the graph.

C. No minimum covering can contain a circuit, for we can always remove an edge from a circuit without learning any of the vertices in the circuit uncovered. Therefore minimal covering of n vertex graph can contain no more than n-1 edges.

Covering of Graph shows that, Minimal covering of n vertex graph can contain no more than n-1 edges.

Vertex cover: A vector covers of an undirected graph G= (V, E) is a sub graph of vertices v'⊆v such that if edge (u, v) is an edge of G, then either u in V or v in V' or both.

Find a vertex cover of maximum size in a given undirected graph. This optimal vertex cover is the optimization version of an NP-Complete problem.

However, it is not too hard to find a vertex cover that is near to optimal.

D. *Vertex Cover Algorithm*

Approx.-vertex cover (G: Graph)

Elected vertex ⟵ vertex with maximum document score

Summary ⟵ { } E'⟵ similarity between corresponding documents in greater than zero i.e. E (G)

$E' \leftarrow Edge\,(E_{ik}) \leftarrow$ Every pair of vertices v_i and v_k if there is similarity between the document d_i and d_k

Weight E_{ik} ⟵ similarity between documents d_i and d_k . While E' is not empty do

Let (v_i, v_k) be an arbitrary edge of E' Summary ⟵ Summary ∪ {selected vertex (v_i, v_k): Documents in the input set of documents represents

Remove from E' every edge incident on either v_i or v_k

IV. Experimental Results and Analysis

A Subset of document understanding conference (DUC) 2007 data set has been used to test the performance of proposed approach. From same topic, Multi-document written selected (DUC 2007) and 250 words are extracted from selected documents. It is an approach of identification of documents from the actual set and organizes them in a way to create a coherent set that contains important information.

A. *Extraction based summarization (EBS)*

We are looking for a short and good summary (the main point of the original content) of the retrieval process. EBS process selects a piece of text from the actual source and organizes them in a way to produce a coherent summary. The core steps of this process are the detection of content from the summary, which is called content selection subtask.

In this paper, a graph-based approach is used for content selection subtask, i.e. vertex covering algorithm for EBS.

B. Extraction based automatic summarization task

Query-Focused: A query is provided to a summarizer in addition to the source documents.

Update: To identify new piece of information in recent article.

Guided: To inspire innovative technique that actually on deeper linguistic analysis.

TABLE II. EXTRACTION BASED SUMMARIZATION PROCESS

Input	Process	Output
Documents and Query	Preprocessing (Section IV.B. a)	Preprocessed sentences
Preprocessed sentences	Feature Selection (Section III.B.1)	Set of Features
Set of features	Context representation (Section IV.B. b)	Context vectors
Context vectors	Content selection (Section III.B.2)	Selected sentences
Selected sentences	Context ordering (Section IV.B. c)	Ordered sentences
Ordered documents	Document realization (Vertex cover algorithm)	Summary

a) Preprocessing: Eliminate unused textual element for further textual data processing and perform some common routine, which is word stemming (Reducing inflected form of a word to a root form, overall features, and making the data less sparse), stop word removal (At preprocessing phase, some words are removed which are high frequency words and don't carry any particular information in the language), text segmentation (An important part of preprocessing, which divides an input document into individual textual contexts.), and Query expansion.

b) Context Representation: Data structures and algorithms are required to process the data on a computer; Vector space (VS) model is the most widely used textual representation model. Feature values are always based on the frequency distribution factor. There is a connection between the context and the meaning of large frequency terms contained in the context.

The document vector d is represented by vector di, Textual context representation as vectors make it easy to apply this technique from geometry and linear algebra. Example: In a collection of five contexts, the third and fifth context can be defined as:

$$\vec{c}_3 = (0,0,1,0,0)$$

$$\vec{c}_5 = (0,0,0,0,1)$$

The term which occur is the sum of vector of contexts

$\vec{c}_3 + \vec{c}_5 = (0,0,1,0,1)$ Similarity between the contexts.

c) As Experiment (Context Ordering):

Step 1: Combined all the selected documents to make a single document (Note: After the removal of tag and white space from sentences of each document).

Step 2: Preprocessing step, Apply stop word removal and word stemming documents.

Step 3: Transfer selected sentences to format suitable for processing.

Step 4: VCA is applied on these preprocessed sentences to extract the summary.

Step 5: Performance evaluation analysis, Recall-oriented Understanding for Gusting Evaluation (ROUGE) is used to identify automatically generated summary.

ROUGE-1.5.5 Package (ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S, and ROUGE-SU) is available to evaluate the quality of summarization.

ROUGE-L: Find the length of Longest Common Subsequence (LCS) between reference summary, systems generated and ROUGE-W: Find the weight of LCS

$$Recall(R_{LCS}) = \frac{\sum_{i=1}^{n} LCS \cup (rs_i, C)}{m}$$

$$Precision(P_{LCS}) = \frac{\sum_{i=1}^{n} LCS \cup (rs_i, C)}{n}$$

$$F-measure(F_{LCS}) = \frac{(1+\beta^2)R_{LCS}P_{LCS}}{R_{LCS}+\beta^2 P_{LCS}}$$

LCS Score=LCS ∪ (rs_i ,C), where, $\beta = \frac{P_{LCS}}{R_{LCS}}$ $and\ rs_i$ is reference sentence, and C is the system generated summary, i=1, 2… n are sentences in a reference summary containing m words.

Let $r_i = w_1w_2w_3w_4w_5$ and C contains two sentences $C_1 = w_1w_2w_7w_8\ and\ C_2 = w_1w_3w_8w_5$ then

$LCS\ of\ r_i\ and\ C_1\ is\ w_1w_2$ and

$LCS\ of\ r_i\ and\ C_2\ is\ w_1w_3w_5$

$The\ union\ LCS\ of\ r_i, C_1\ and\ C_2\ is\ w_1w_2w_3w_5$

LCS ∪ $(rs_i, C) = 4/5$

$$ROUGE-N = \frac{\sum_{s\in Refs}\sum_{N-gram\in s} Maxcount(N-gram)}{\sum_{s\in Refs}\sum_{N-gram\in s} Count(N-gram)}$$

Refs: Reference summary with sentences

N= Length of N-gram,

Max count: Maximum number of N-gram occurring in the reference summary and generated summary.

N-gram: Contiguous sequence of N items for a given sequence of text.

N: Either syllables/letters/words/base pair (Note: depends on application)

If N>1, it considers the words order in a sentence and is more accurate than cosine similarity (Note: N=1, for cosine similarity and N-gram overlap)

The evaluation method conforms to the guideline from the DUC 2007 competition. An automatic technique was used. The most recent competition which has been used for automatic evaluation in DUC 2007. The main task of this competition is the query focused MDS. DUC 2007 dataset of writing about the same topics and corresponding 250 words are extracted. DUC 2007 makes use of ROUGE evaluation of our system. ROUGE-2 and ROUGE-SU4 are used for evaluation procedure.

TABLE III. AVERAGE F-SCORE MEASURED BY ROUGE-2 AND ROUGE-SU4 OBTAINED BY THE MMR, DSDR AND VCA

Summarizer	ROUGE-2		ROUGE-SU4	
	F-Score	95% Confidence interval	F-Score	95% Confidence interval
MMR	0.0939	0.089-0.098	0.1464	0.142-0.151
DSDR	0.0952	0.091-0.100	0.1467	0.142-0.151
VCA	0.1245	0.120-0.129	0.1771	0.173-0.182

VCA on DUC 2007 dataset is compared with DSDR and MMR applied on centroid method and F-Score measured by ROUGE-2 and ROUGE-SU4.

V. CONCLUSION

An automatic graph-based MDS is a difficult task that involves different challenging subtasks. The vector cover algorithm performed well in solving the general issues of identifying useful entities based on the relation between them. Using DUC 2007 data, the experiment was conducted with MMR, DSDR, and VCA. The comparison demonstrated that VCA can deliver quite good results. Considering the success of VCA, it is certainly worth doing further research by identifying a new approach to find the edge weight in the graph which enhances the performance of the documents selected for the summary.

REFERENCES

[1] Lovins, J B. Development of a stemming algorithm. Mechanical Translation and Computational Linguistics; 1968. vol 11. p. 22-31.

[2] Barzilay R., et al. Inferring strategies for sentence ordering in multidocument news summarization. JournalofArtificial Intelligence Research. 1980. vol.17. p. 35–55

[3] Salton, et al. Term-weighting approaches in automatic text retrieval. Information processing & management;1988. p. 513-523

[4] Carbonell, et al. The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. ACM; 1998.

[5] Breck Baldwin ,et al. Dynamic coreference-based summarization. In Proceedings of the Third Conference on Empirical Methods in Natural Language Processing. Granaare DA, Spain; June 1998.

[6] Jade Goldstein, et al. MDS by Sentence Extraction. NAACL-ANLP 2000 Workshop on Automatic summarization. Seattle, Washington; 2000. vol. 4. p. 40 - 48.

[7] Inderjeet Mani. Automatic summarization, Kluwer Academic Publishers; 2004.

[8] Dragomir R. et al. Centroid-based summarization of multiple documents. Information Processing and Manageme; 2004. vol. 40. no.6. p. 919–938.

[9] Kaustubh Patil et al. SUMGRAPH: TS Using Centrality In The Pathfinder Network. International Journal on Computer Science and Information Systems; 2007. vol.2. no.1. p. 18-32.

[10] Florian Boudin, et al. NEO-CORTEX: A Performant User-Oriented MDS System.Lecture Notes in Computer Science. Springer; 2007. Vol 4394. p.551-562.

[11] Jen-Yuan Yeh ,et al. iSpreadRank: Ranking sentences for extraction-based summarization using feature weight propagation in the sentence similarity network. Expert Systems with Applications, vol. 35, no. 3, p. 1451-1462, October 2008.

[12] Abuobieda, et al, 2013, 'Differential evolution cluster-based TS methods' Proceedings of the International Conference on Computing, Electrical and Electronics Engineering, Khartoum, Sudan, pp.244-248.

[13] Mehdi Bazghandi, et al, 2012, 'Extractive Summarization Of Farsi Documents Based On PSO Clustering', International Journal of Computer Science Issues, vol. 9, no. 4, No 3,pp. 329 - 332

[14] Canan Pembe, F & Tunga Gungor 2007, 'Automated Query-biased and Structure-preserving TS on Web Documents', Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications, Istanbul.

[15] Wei Wang, et al, 'Exploring hypergraph-based semi-supervised ranking for query-oriented summarization', Information Sciences, Prediction, Control and Diagnosis using Advanced Neural Computations, vol. 237, pp. 271 - 286.

[16] Carlos N Silla, et al. 2004, 'Automatic TS with Genetic Algorithm-Based Attribute Selection', Proceedings of the 9th Ibero-American Conference on Artificial Intelligence, Lecture Notes in Computer Science, Puebla, Mexico, vol. 3315, pp. 305-314.

[17] Gunes Erkan et al. 2004, 'LexRank: Graph-based Lexical Centrality as Salience in TS', Journal of Artificial Intelligence Research, vol. 22, pp. 457-479.

[18] Harun Uguz 2011, 'A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm', Knowledge-Based Systems, vol. 24, no. 7, pp. 1024-1032.

[19] Nguyen Quang Uy, et al. 2012, 'A Study on the Use of Genetic Programming for Automatic TS', Fourth International Conference on Knowledge and Systems Engineering, Danang, pp 93-98.

[20] Inderjeet Mani 2004, Automatic summarization, Kluwer Academic Publishers, MA, USA.

[21] Kiani, A & Akbarzadeh, MR 2006, 'Automatic TS Using: Hybrid Fuzzy GA-GP', Proceedings of the IEEE International Conference on Fuzzy Systems, Vancouver, BC, pp. 977-983.

[22] Kyoomarsi, F et al. 2010, 'Extraction-Based TS Using Fuzzy Analysis', Iranian Journal of Fuzzy Systems, vol. 7, No. 3, pp. 15-32.

[23] Kyoomarsi, et al. 2008, 'Optimizing TS Based on Fuzzy Logic', Proceedings of the Seventh IEEE/ACIS International Conference on Computer and information science, Portland, OR, pp. 347-352.

[24] Ladda Suanmali, et al. 2009a, 'Sentence Features Fusion for TS Using Fuzzy Logic', Proceedings of the ninth International Conference on Hybrid Intelligent Systems, Shenyang, vol. 1, pp. 142-14.

[25] Ladda Suanmali, et al. 2009b, 'Fuzzy Logic Based Method for Improving TS', International Journal of Computer Science and Information Security, vol. 2, no. 1.

[26] Mohammed Salem Binwahlan, et al., 2009a, 'Fuzzy Swarm Based TS', Journal of Computer Science vol. 5, no. 5, pp. 338-346.

[27] Mohammed Salem Binwahlana, et al 2010, 'Fuzzy swarm diversity hybrid model for TS', Information Processing & Management, vol. 46, no. 5, pp. 571-588.

[28] Pallavi D Patil et al. 2014, 'TS Using Fuzzy Logic', International Journal of Innovative Research in Advanced Engineering, vol. 1, no. 3, pp. 42-45.

[29] Ruche S Dixit, et al. 2012, 'Improvement of TS Using Fuzzy Logic Based Method', IOSR Journal of Computer Engineering, vol. 5, no. 6, pp. 05-10.

[30] Rasim Alguliev et al. 2009, 'Evolutionary Algorithm for Extractive TS', Intelligent Information Management, vol. 1, pp.128-138.

[31] Xuan Li, et al. 2013, 'Update Summarization via Graph-Based Sentence Ranking',IEEE Transactions on Knowledge & Data Engineering, vol.25, no.5, pp. 1162-1174.

[32] Rasim M Alguliev, et al. 201 1, 'MCMR: Maximum coverage and minimum redundant TS model', Expert Systems with Applications, vol. 38, no.12, pp. 14514-14522.

[33] Chao Shen, et al. 2011, 'Learning to Rank for Query-Focused Multidocument Summarization', Proceedings of the International Conference on Data Mining, Vancouver,Canada, pp.626-634.

[34] Chowdary, et al. 2010, ‘A system for query specific coherent text MDS’, International Journal on Artificial Intelligence Tools, vol. 19 no. 5, pp. 597-626.

[35] Furu Wei, et al. 2008, ‘Query-Sensitive Mutual Reinforcement Chain and Its Application in Query-Oriented MDS’, Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, New York, USA, pp. 283-290.

[36] Jie Tang, et al. 2009, ‘Multi-topic based Queryoriented Summarization’, Proceedings of 2009 SIAM International Conference on Data Mining, Sparks, Nevada, pp. 1147-1158.

[37] Liang Ma, et al. 2008, ‘Query-Focused MDS Using Keyword Extraction’, Proceedings of the International Conference on Computer Science and Software Engineering, Wuhan, Hubei, vol. 6, pp. 20-23.

[38] LuWang, et al. 2013, ‘A Sentence Compression Based Framework to Query- Focused MDS, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, pp.1384-1394.

[39] Ramakrishna Varadarajan et al. 2005, ‘Structure-Based Query-Specific Document Summarization’, Proceedings of the 14th ACM International Conference on Information and Knowledge Management, Bremen, Germany pp. 231-232.

[40] Tingting He, et al. 2008a, ‘A New Feature-Fusion Sentence Selecting Strategy for Query-Focused MDS’, Proceedings of the International Conference on Advanced Language Processing and Web Information Technology, Liaoning, China, pp, 81 -86.

[41] Tingting He, et al. 2008b, ‘The Automated Estimation of Content-Terms for Query-Focused MDS’, Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Jinan, Shandong, China, vol. 5, pp. 580-584.

[42] Wenjuan Luo, et al. 2013, ‘Exploiting relevance, coverage, and novelty for query-focused multidocument summarization’, Knowledge-Based Systems, vol.46, pp. 33- 42.

[43] Yan Liu, et al. 2012, ‘Query-Oriented MultiDocument Summarization via Unsupervised Deep Learning’, Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, Ontario, pp. 1699 – 1705.

[44] You Ouyang, et al. 2011, ‘Applying regression models to query-focused MDS’, Information Processing & Management, vol. 47, no. 2, pp 227-237.

[45] Derong Liu, et al. 2006, ‘Multiple Documents Summarization Based on Genetic Algorithm’, Proceedings of the Third International Conference on Fuzzy Systems and Knowledge Discovery, Lecture Notes in Computer Science, Xian, China, vol. 4223, pp. 355-364.

[46] De-Xi Liu, et al. 2006a, ‘Genetic Algorithm Based MDS’, Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence, Guilin, China, Lecture Notes in Computer Science, vol. 4099, pp. 1140-1144.

[47] Lin Zhao, et al. 2009, ‘Using query expansion in graph-based approach for query-focused MDS’, Information Processing and Management, vol. 45 no. 1, pp. 35-41.

[48] De-Xi Liu, et al. 2006b, ‘ A Novel Chinese MDS Using Clustering Based Sentence Extraction’, Proceedings of the International Conference on Machine Learning and Cybernetics, Dalian, China, pp. 2592-2597.

[49] Dingding Wang, et al. 2010, ‘Document Update Summarization Using Incremental Hierarchical Clustering’(HC),Proceedings of the 19th ACM international conference on Information and knowledge management, New York, USA, pp. 279-288.

[50] Sun Park, et al. 2008, ‘Query-Based MDS Using Non-Negative Semantic Feature and NMF Clustering’, Proceedings of the Fourth International Conference on Networked Computing and Advanced Information Management, vol. 2, Gyeongju, Korea, pp. 609-614.

[51] Xiaojun, et al. 2008, ‘MDS using cluster-based link analysis’, Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, pp. 299-306.

[52] Dragomir, R, et al. 2004, ‘Centroid-based summarization of multiple documents’, Information Processing and Management, vol. 40, no. 6, pp. 919-938.

[53] Goran Glavas , et al. 2014, ‘Event graphs for information retrieval and MDS’, Expert Systems with Applications, vol. 41, no. 15, pp. 6904-6916.

[54] Zhengchen Zhang, et al. 2012, ‘Mutualreinforcement document summarization using embedded graph based sentence clustering for storytelling’, Information Processing & Management, vol. 48, no. 4, pp. 767-778.

[55] John Atkinson, et al. 2013, ‘Rhetorics-based multidocument summarization’,Expert Systems with Applications, vol. 40, pp. 4346-4352.

[56] Yogan Jaya Kumar,et al. 2014, ‘Multi document summarization based on news components using fuzzy cross-document relations’, Applied Soft Computing, vol. 21, pp. 265-279.

[57] Rasim M. Alguliev,et al. 2013, ‘ CDDS: Constraint-driven Document summarization models’, Expert Systems with Applications, vol. 40, pp. 458-465.

[58] Hsun-Hui Huang, et al. 2006, ‘Fuzzy-Rough Set Aided Sentence Extraction Summarization’, Proceedings of the International Conference on Innovative Computing ,Information and Control, Beijing, China, vol. 1, pp. 450-453.

[59] Ladda Suanmali, et al. 2009c, ‘Feature-Based Sentence Extraction Using Fuzzy Inference Rules’, Proceedings of International Conference on Signal Processing Systems, Singapore, pp. 511 - 515.

[60] Ladda Suanmali, et al. 2011, ‘Genetic Algorithm Based Sentence Extraction for Text Summarization’, International Journal of Innovative Computing, vol. 1, no. 1.

[61] Esther Hannah, et al. 2011, ‘Automatic Extractive TS Based on Fuzzy Logic: A Sentence Oriented Approach’ Proceedings of the Swarm, Evolutionary, and Memetic Computing, Lecture Notes in Computer Science Visakhapatnam, Andhra Pradesh, India, vol. 7076, pp 530-538.

[62] Ghalehtaki, RA, et al. 2014, ‘Combinational Method of Fuzzy, Particle Swarm Optimization and Cellular Learning Automata for Text Summarization’, Proceedings of the Iranian Conference on Intelligent Systems, Bam, Iran, pp. 1 -6.

[63] Jian-Hui Wang, et al. 2003, ‘Sentences clustering based automatic summarization’, Proceedings of the International Conference on Machine Learning and Cybernetics, Beijing, China, vol. 1, pp. 57-62.

[64] Zhang Pei-ying, et al. 2009, ‘Automatic text summarization based on sentences clustering and extraction’, Proceedings of the 2nd International Conference on Computer Science and Information Technology, Beijing, China, pp. 167-170.

[65] Xiaogang Ji 2008, ‘Research on the Automatic Summarization Model based on Genetic Algorithm and Mathematical Regression’, Proceedings of the International Symposium on Electronic Commerce and Security, Guangzhou, China, pp. 488-491.

[66] Rajesh Shardanand Prasad, et al. 2010, ‘Implementation and Evaluation of Evolutionary Connectionist Approaches to Automated Text Summarization’, Journal of Computer Science, vol. 6, no. 11, pp. 1366-1376.

[67] Elena Baralis, et al. 2013b, ‘GraphSum : Discovering correlations among multiple terms for graphbased summarization’, Information Sciences , vol. 249, pp. 96-109.

[68] Furu Wei, et al. 2010, ‘A documentsensitive graph model for MDS’, Knowledge and Information Systems, vol. 22, no. 2, pp. 245-259.

[69] Anand Gupta, et al 2014 “Text summarization through entailment based minimum vertex cover’ Third joint conference on lexical and computational semantics (*SEM 2014), page 75-80, Dublin Ireland.

[70] Lin C Y. ROUGE: A package for automatic evaluation summaries. In Proceedings of the workshop on text summarization branches out.

[71] DUC. http://duc.nist.gov

[72] ROUGE Toolkit. <http://berouge.com/default.aspx