Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

Graph-Based Extractive Text Summarization Models: A Systematic Review

Abdulkadir Abubakar Bichi*

*Corresponding Author, PhD Candidate, School of Computing, University Technology Malaysia, 81310 Johor Bahru, Johor, Malaysia. E-mail: engraabubakar@gmail.com

Ruhaidah Samsudin

Senior Lecturer, School of Computing, University Technology Malaysia, 81310 Johor Bahru, Johor, Malaysia. E-mail: pantea@usm.my

Rohayanti Hassan

Senior Lecturer, School of Computing, University Technology Malaysia, 81310 Johor Bahru, Johor, Malaysia. E-mail: rohayanti@utm.my

Khalil Almekhlafi

Assistant Professor, Taibah University, CBA-Yanbu, 42353, Saudi Arabia. E-mail: drkhalilalmekhlafi@gmail.com

Abstract

The volume of digital text data is continuously increasing both online and offline storage, which makes it difficult to read across documents on a particular topic and find the desired information within a possible available time. This necessitates the use of technique such as automatic text summarization. Many approaches and algorithms have been proposed for automatic text summarization including; supervised machine learning, clustering, graph-based and lexical chain, among others. This paper presents a novel systematic review of various graph-based automatic text summarization models.

Keywords: Natural Languages Processing; Text Mining; Graph approaches.

DOI: https://orcid.org/10.22059/jitm.2022.84899 Manuscript Type: Research Paper University of Tehran, Faculty of Management

Received: January 12, 2020 Accepted: March 25, 2020

Introduction

The high rate of growth of textual data both on the internet and offline storage, make it difficult and time consuming for one to read across and find the required information (Wafaa S. El-Kassas, Salama, Rafea, & Mohamed, 2021; Mojrian & Mirroshandel, 2021; Nawaz et al., 2020). The volume of indexed textual data available online is estimated to be about 4.5 billion pages and the size is continuously increasing in exponential rate (Mohamed & Oussalah, 2019). For many topics is practically impossible to go through all the content return by search engines, thus makes a research more challenging and time-consuming task to scholars (Nouf Ibrahim Altmami & Menai, 2020). This necessitate the used of computing methods to the problem, and the automatic text summarization (ATS), was found to be the most promising option (Aries, Zegour, & Hidouci, 2019; Mojrian & Mirroshandel, 2021). ATS provide fast and reliable means of acquiring knowledge and research (S. Hou & Lu, 2020). Without the use of ATS searching and studying of a particular topic from the internet will be a tedious task (W. S. El-Kassas, Salama, Rafea, & Mohamed, 2020).

A text summary is a briefer form of the original document, in which the principal information is preserved (Narayan, Cohen, & Lapata, 2018). A summary retains the idea of the articles while preserving the memory space by removing the unnecessary and repetitive parts. Readers can grasp the concept and key ideas of the document without necessarily reading the entire document. The aim of text summarization is to reduce the length of a document for easy comprehension of the content (Zamana, Shardlow, Hassan, Aljohani, & Nawaz, 2020). ATS algorithms analyses large document and generate briefer version of them (Nawaz et al., 2020).

The ATS is classified using different criteria; based on number of input files, generated output, purpose and context. Based on number of input documents, the ATS is classified into: single document and multi-document ATS. The single-document ATS generates separate summary for each input file. In the contrast, multi-document ATS generate one single summary from many related documents (Cai & Li, 2013). Based on the generated output ATS is classified into; extractive and abstractive ATS. The extractive type selects most important sentences of the document and concatenate them together to form a summary (Aker, 2013). The abstractive on the other hand, involves intense content reformatting, paraphrasing and rewriting the text in entirely different words (Yao, Wan, & Xiao, 2017). Based on purpose the ATS is classified into; query-focus and generic. In query-focus, a summary is generated based on the user biasness (Zhong, Liu, Li, & Long, 2015), usually the system considers the query words or phrases in scoring the document sentences. In contrast, the generic type covers the entire documents subtopic (Gong & Liu, 2001), and generate unbiased summary regardless of the user preference. Based on context, ATS is classified into indicative and informative. The indicative summary is less detail summary, which contains only the key outlines of the source document (Narayan, Cohen, & Lapata, 2019). Whereas the informative summary cover in depth all topics of the original text, which in most cases are enough for major analysis without referring to the original source (Vollmer, Golab, Böhm, & Srivastava, 2019).

The first model of extractive ATS has been proposed for more than 60 years (Luhn, 1958). The earlier techniques involve the use of text heuristic features like the term's frequencies (Luhn, 1958), sentences position (Baxendale, 1958), and title words (Edmundson, 1969) among others. Far along, other techniques were used for extractive ATS, including clustering method, graph method, supervised machine learning and lexical chain. The main objective of this research is to conduct a systematic review of graph-based approaches for extractive ATS, which is the first of its kind. The approach is one of the most efficient and reliable methods of automatic text summarization. A comprehensive review of the technique will help researchers to easily understand the state of the art and future directions of the approach. The remaining parts of the paper is divided into; related works, methodology, result and discussions, and conclusions

Litrecher review

A research by Wafaa S. El-Kassas et al. (2021) conducted a survey of automatic text summarization approaches, which include the proposed models, evaluation techniques and public available datasets. Similarly, a survey on extractive automatic text summarization was presented by Moratanch and Chitrakala (2017); Nazari and Mahdavi (2019); Saziyabegum and Sajja (2016), with emphasis on summarization approaches. A review of ATS approaches was conducted by Verma and Verma (2020) that outlined the challenges and strengths of common extractive summarization approaches. Elrefaiy, Abas, and Elhenawy (2018) presented a reviewed on some selected automatic text summarization techniques analyzing their strengths and weakness. And systematic review of ATS articles from 2008 to 2019 that includes methodologies, evaluation techniques and datasets was presented by Widyassari et al. (2020). Some approaches provide a survey on particular method or algorithm, Mosa, Anwar, and Hamouda (2019) present a survey on swarm intelligence approaches for ATS. Similar research by Kumar and Sharma (2019) presented a systematic review on fuzzy logic based ATS methods, deep learning-based methods (Suleiman & Awajan, 2019).

Domain specific survey was presented by Bhattacharya et al. (2019) for legal documents ATS, microblogs ATS (Dutta et al., 2019), scientific articles summarization models (Nouf Ibrahim Altmami & Menai, 2020). Abualigah, Bashabsheh, Alabool, and Shehab (2020) presented a survey of automatic text summarization of Arabic language and the survey of abstractive summarization is presented by (Gupta & Gupta, 2019; Jacquenet, Bernard, & Largeron, 2019; Lin & Ng, 2019; Tandel, Mistree, & Shah, 2019). A reviewed on automatic text summarization evaluation techniques was presented by Saziyabegum and Sajja (2017). The survey on graphical method was only found in (Y. Liu, Safavi, Dighe, & Koutra, 2018) which broadly classified the graph-based method according to the input graph into static and dynamic graph method. The emphasis of the survey is on the graph structures and fundamental graph theories rather than the ATS application. To do authors best ability this is first comprehensive systematic review on graph-based automatic text summarization approaches.

Methodology

The main goal of the study is to review the existing literature in the area of graph-based approaches for automatic text summarization. The study involves investigating the trends of progressing in the method since it first emergence about 17 years back. The published research articles on graph-based ATS from both primary and secondary sources were collected and analyzed, and finally the most relevant ones are chosen for consideration. Systematic Literature Review (SLR) approach was used in this research, based on the three phases of planning, implementation and reporting.

Research question (RQ)

RQ was prepared using PICOC as presented in Table 1. And the research questions are explained in Table 2.

Table 1. PICOC Criteria

Population	**Automatic text summarization**
Intervention	Graph based approach for automatic text summarization
Comparison	-
Outcomes	Graph based approach for automatic text summarization performance
Context	Study in computer laboratories with small and large datasets

Table 2. Research question and motivations

ID	Research question	Motivation
RQ1	What journal/conference paper about ATS	Identify the most significant publications in ATS
RQ2	What journal/conference paper about graph based ATS	Identify the most significant publications in graph-based ATS
RQ3	What graph model used in the ATS	Classify graph-based ATS approaches according to graph model
RQ4	What similarity measures used in the graph-based ATS	Identify similarity measure used in each model
RQ5	What are the strengths and weakness of the models	Identify the strengths and weakness of each graph-based algorithm
RQ6	What are the challenges of graph-based ATS	Identify the research trend in graph-based ATS

Search Strategy

The papers used in the review were collected from Scopus, Web of science databases through the University Technology Malaysia library access and Google scholar. The following string was used for searching the required papers: (graph-based) * AND (automatic text summarization OR extractive summarization OR ATS) AND (approach OR technique OR model OR method). The papers were selected in line with research questions described in Table 3, Table 4 and mind map of Figure 1.

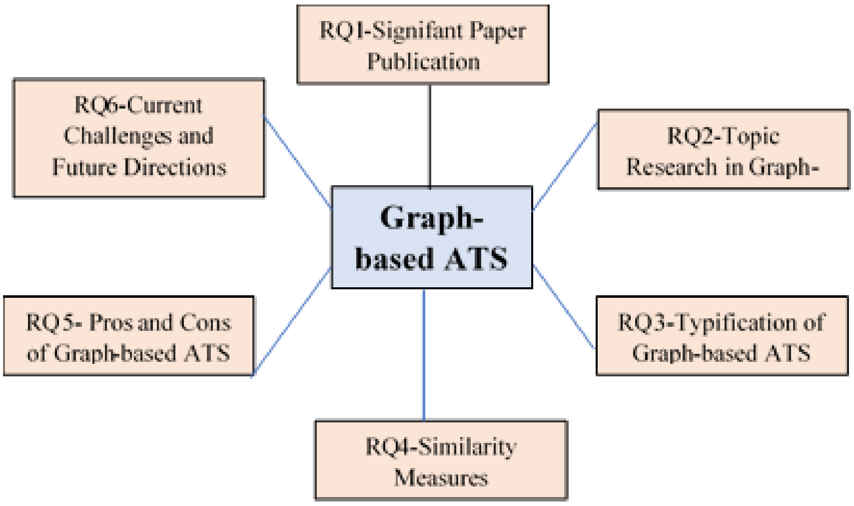

Figure. 1 Mind Map of Review Text Summarization

Table 3. Inclusion and Exclusion Criteria

Inclusion Criteria	Studies that summarize texts include topics, problems, graph model, similarity measure, from journal papers and conferences publications between 2003-2021
Exclusion criteria	Studies not written in English, discuss other topics beyond ATS, studies with unclear results

Table 4. Data Extraction

Property	Research question
Publication	QR1
Research topic trend	QR2
Classification of graph-based ATS	QR3
Similarity measures	QR4
Strengths and weakness of graph-based ATS models	QR5
Current Challenges and Future Directions	QR6

Study selection

The focuses for database searches are the publication title, abstract, and keywords. Publications papers covers journal papers and conferences papers/proceedings with an initial determination of 80% of journals and 20% of conferences written in English language. The title" Graph-based Approach for Automatic Text Summarization" was searched from Scopus, Web of Science databases and Google scholar. The over 1,202 published articles were collected from all sources and 101 references were finally selected for the review, as shown in Figure 2 and Table 5.

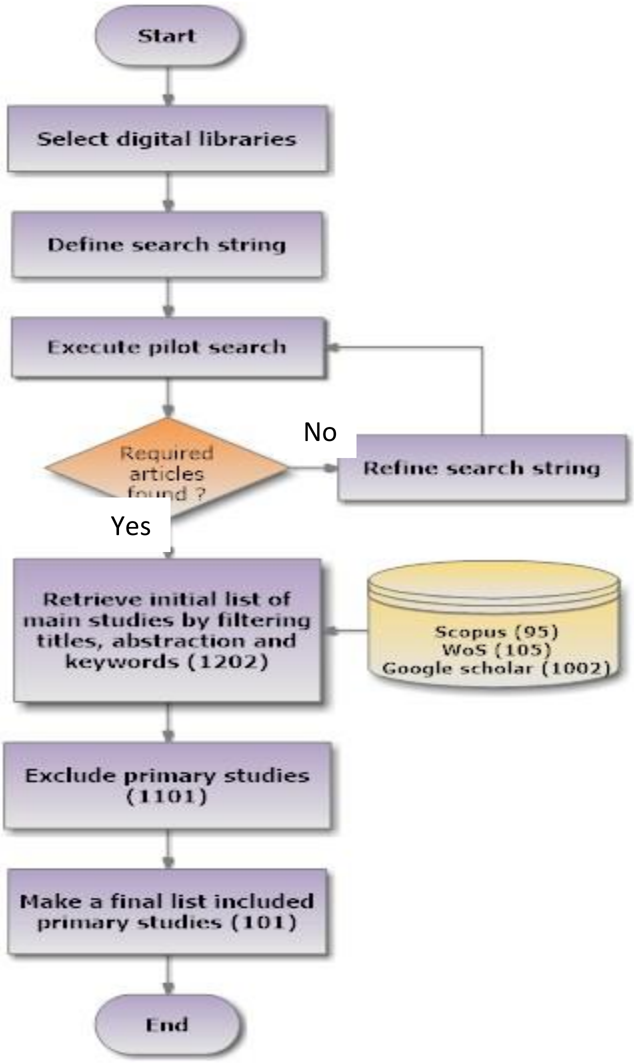

Figure 2. Search and Selection

Table 5. Research Methodology References Table

Search engine	No. Produced	Rejected at Title	Accepted at Title, Abstract read	Rejected at Abstract	Accepted at Abstract, Paper read	Accepted at Paper
Scopus	95	35	60	22	38	33
WoS	105	46	59	17	42	37
Google Scholar	1002	904	98	38	60	42
Duplicate						11
Total						101

Findings

Paper Studies

About 105 papers were collected from WoS as shown in Figure 3 and 95 paper from Scopus database, as shown in Figure 4. The remaining publications were collected from Google scholar.

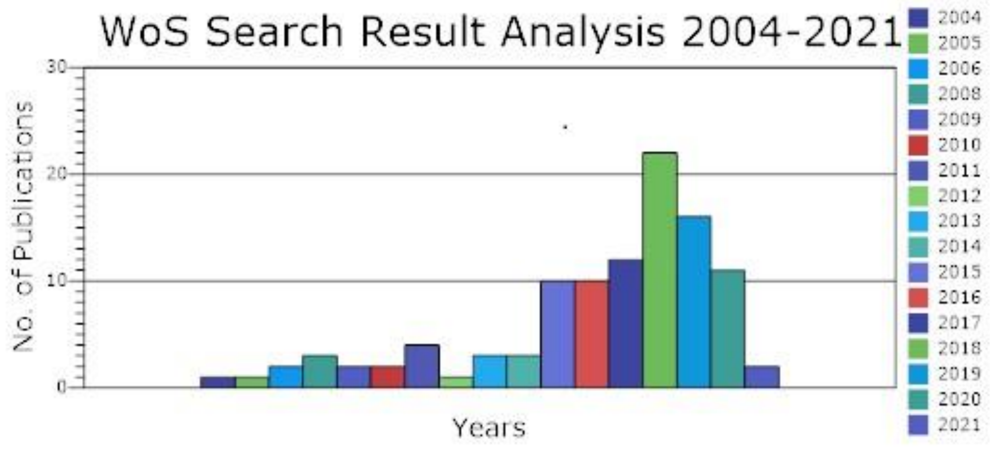

Figure 3. WoS Yearly Distribution of Publications 2004 to 2021

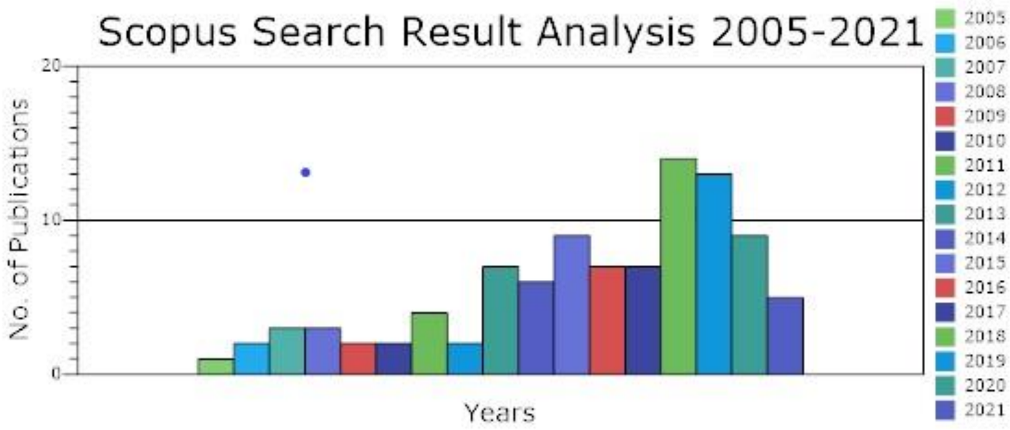

Figure 4. Scopus Yearly Distribution of Publications 2005 to 2021

Graph-based ATS Approaches

Graph-based ATS models are based on the concept of mathematical graph theory, where text sentences or words are represented with the graph vertices (Gambhir & Gupta, 2017) and the edges of the graph represent the relations between sentences. The method exploits the graph structure to establish relations between text unit and determine ranking (Gunawan, Pasaribu, Rahmat, & Budiarto, 2017; Nazari & Mahdavi, 2019). It considers the relation of sentence with all other sentences in the documents from all positions for a final ranking; therefore, produce readable and coherent summaries (Aries et al., 2019; Moratanch & Chitrakala, 2017). The approach has the advantages of language and domain independency (Nasar, Jaffry, & Malik, 2019). And like the heuristic features-based and clustering methods, the graph-based algorithms are simple to implement and the results is more efficient compare to other unsupervised methods (Al-Khassawneh, Salim, & Jarrah, 2017).

In the graph-based ATS method, sentences recommend other similar sentences and the importance of sentence depend on the importance of the sentences that recommend it. It assumes that the weights of the words are equal, so it does not consider the importance of words in the document (Fang, Mu, Deng, & Wu, 2017). The method achieves promising solution and the generated summaries are highly reliable (Mojrian & Mirroshandel, 2021). This research classified graph-based models for extractive ATS as shows in Figure 5 and details in the following subsections.

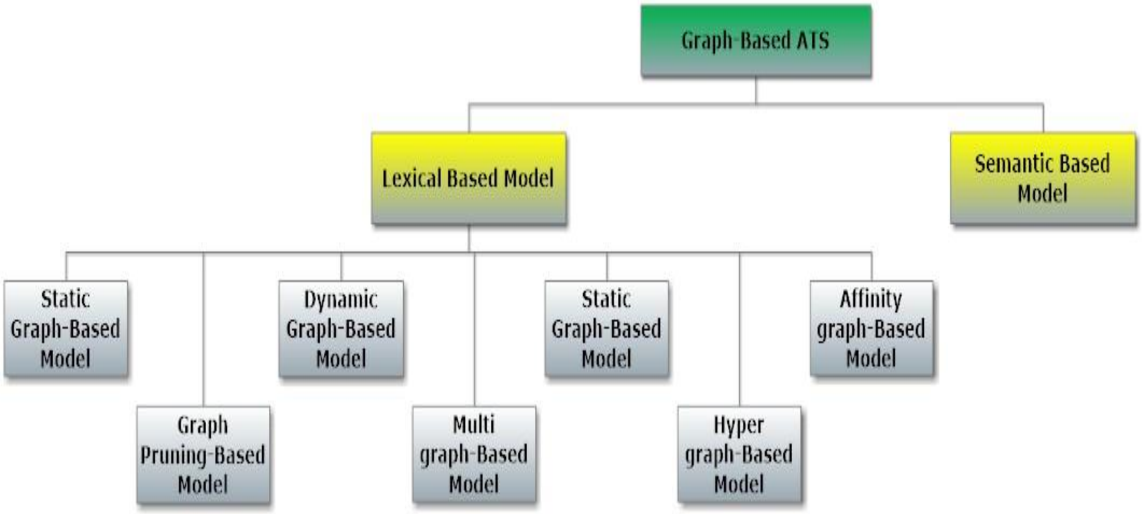

Figure 5. Graph-Based ATS Classification

A. Lexical Graph-Based Model

The lexical graph-based models consider the lexical similarities between text sentences to determine their scores and ranking. The lexical graph-based model is further classified into: static graph-based model, dynamic graph-based model, graph pruning based models, hypergraph-based Model, Affinity Graph-based Model and multigraph-based model, as discussed in the following subsections.

i. Static Graph-Based Model

The static graph-based models represent the text using undirected weighted graph. The algorithms are based on the concept of earlier graph ranking algorithms developed for other applications, such as Hyperlink-Induced Topic Search (HITS) algorithm (Kleinberg, 1999), Positional Power Function algorithm (Herings, Laan, & Talman, 2001) and PageRank algorithm (Brin & Page, 1998). Mihalcea and Tarau (2004), proposed first graph-based ATS algorithm based on the concept of PageRank algorithm called TextRank algorithm. In the algorithm, the text sentences are represented as graph vertices and the graph arcs are draw based on words overlaps between sentences. Unlike the PageRank, the TextRank algorithm uses undirected graph to represent symmetric relations between sentences and a weight wij is introduced to the edges to indicates the degree of causality between sentences i and j. Similarly, Erkan & Radev (2004) proposed another graph-based ATS algorithm based on the concept of PageRank algorithm called LexRank. But in the LexRank a cosine similarity of vectors TF-IDF is used to determine sentences similarity and it support multi-document summarization. A research by Mallick, Das, Dutta, Das, and Sarkar (2018), modified

TextRank algorithm by using inverse sentence frequency (isf) based cosine similarity to measure the pairwise similarity between sentences.

Graph-based ATS was applied for domain specific and other languages text summarization. Elbarougy, Behery, and Khatib (2019), modified PageRank algorithm for Arabic language ATS, by including the noun count of a sentence in ranking criteria. Similarly, Sikder, Hossain, and Robi (2019), modified PageRank for summarization of Bengali text. Milad Moradi, Dashti, and Samwald (2020) proposed a graph-based model for biomedical text summarization using a novel technique of representation by means of hybridizing context-sensitive and context-free embedding. Fakhrezi, Bijaksana, and Huda (2021), used TextRank for summarization of Qur’an vocabularies return by search algorithm.

Woloszyn, Machado, Wives, and Mo (2018), proposed a graph model that combined cosine-similarity with keyword-similarity for sentence scoring. The algorithm works for cross-domain extractive summarization it enables re-scoring the sentence for better performance. Natesh, Balekuttira, and Patil (2016), proposed the used of noun position in a sentence for scoring. In the approach the inverse of distance between any two nouns in a sentence is regarded as their weight, where the score of a sentence is determined by the total scores of it individual nouns. Similarly, Alzuhair and Al-Dhelaan (2019), proposed hybrid ranking algorithm, by combining PageRank algorithm with HITS algorithm using harmonic mean. Similarly, Barrios, López, Argerich, and Wachenchauzer (2016), used TextRank algorithm with BM25 ranking function, which is variation of the tf-idf function. Another research work by Mussina, Aubakirov, and Trigo (2018), proposed symmetric ranking for extractive summarization. In the approach the weight of an edge is determine by the length of longest common substring, and the total sum of weights of all edges connected to a node is the node score.

ii. Dynamic Graph-Based Model

The previously discussed ATS algorithms like TextRank and LexRank algorithms work on static graph model. Ziheng (2007), proposed the used of evolutionary graph model for ATS, the model consider the arrival of sentences into the documents. The sentences are arranged in chronological order from first to last, and modelled using a directed graph. The algorithm ranks the documents sentences by considering both their similarities with other sentences in the cluster and their similarities to the previously selected sentences in the documents using modified MMR re-ranker equation (Carbonell & Goldstein, 1998). Gallo et al. (2018), enhanced the concept of timestamps graph with time abstraction using a signal function. The method further improved the quality of scoring by selecting the best pattern and discarding the irrelevant edges. A dynamic graph concept was used for summarization of biomedical documents by modelling the input text to represent events (M. Moradi, 2018).

iii. Graph Pruning-Based Model

The graph pruning-based models of extractive summarization reduces the number of graph nodes and edges by pruning unnecessary graph edges and vertices, thus reducing the time of the graph search. K. Patil and Brazdil (2007), modified LexRank by pruning the graph using a technique of pathfinder network before applying the ranking algorithm. Miranda-Jiménez, Gelbukh, and Sidorov (2013), developed a model for single-document summarization by first pruning the graph before applying HITS to rank the sentences. Similarly, Al-Khassawneh et al. (2017), used graph triangle method for pruning graph in extractive text summarization.

More so, a research by Hark and Karcı (2020), introduced Karcı method, a graph entropy algorithm to filter out irrelevant graph vertices and select most informative sentences in each paragraph, for multi-document summarization. Likewise, the used of maximum independent set method to filter out less relevant nodes of the graph before applying the ranking algorithm for extractive generic multi-document summarization was proposed by Uçkan and Karcı (2020). The pruning graph models reduces the graph searching time but has additional time of graph pruning, thus the overall process time is not improved in the model but the accuracy of ranking and selection is better in smaller graphs.

iv. Hypergraph-based Model

Hypergraph allows one edge called hypergraph incidence to connect more than 2 vertices, thus enable more advance relations between the graph vertices. W. Wang, Li, Li, Li, and Wei (2013); W. Wang, Wei, Li, and Li (2009), proposed a model for query-focus text summarization based on the concept of hypergraph. The hypergraph model was extended for multi-document ATS using vertex-reinforced random walk (Xiong & Ji, 2016). Similarly, Lierde and Chow (2019), applied clustering technique to hypergraph model for query-focus text summarization; by first grouping the document into clusters and then construct a hypergraph for each cluster.

v. Affinity Graph-based Model

The concept of affinity graph involves grouping nodes representing similar objects from different graphs. Wan and Yang (2006), used the concept of affinity graph for multi-document summarization by utilizing both inter and intra documents diversity to determine the similarity between sentences. Another research applied random walk algorithm to affinity graph-based ATS (K. Wang, Liu, Sui, & Chang, 2017). Similarly, Hu, He, and Zhang (2015), proposed affinity model with manifold ranking and Kanitha, Mubarak, and Shanavas (2018), scores sentences using the sum of their affinity weights for extractive ATS of Malayalam language.

vi. Multigraph-based Model

Multigraph model allows more than one edges between two adjacent vertices. The number of edges indicates the strength of the connection, which is regarded as a weight of the vertex.

AlZahir, Fatima, and Cenek (2015), used multigraph graph model to represent text for extractive text summarization. In the model an edge is drawn for every two similar words in the adjacent sentences, which later represented using a symmetric matrix. W. S. El-Kassas et al. (2020) proposed a new graph-based framework for generic single-document extractive text summarization “EdgeSumm”, the approach combined the techniques of graph with statistics, semantic and centrality. The model proposed novel method of text representation in which the nouns are represented as the graph nodes and the words between nouns are the graph edges.

B. Semantic Graph-Based Model

The semantic graph-based model used a semantic similarity measure to determine relations between document sentences The method used semantic properties of the documents, such as synonymy, noun to pronouns mapping for more accurate text representation (Alami, El Adlouni, En-Nahnahi, & Meknassi, 2018; N. I. Altmami & Menai, 2018a, 2018b; Dalal & Malik, 2018; Hassan, Abdelrahman, Bahgat, & Farag, 2019; Plaza, Díaz, & Gervás, 2011) . Ullah and Al Islam (2019), utilized the idea of semantic graph for extractive text summarization by first extracting the Predicate Argument Structure (PAS) of sentences; the sematic similarity between sentences is measures using their PAS. The graph vertices in the approach are ranks using PageRank algorithm and re-rank using MMR algorithm to minimize redundancies. Sevilla, Fernández-Isabel, and Díaz (2016), proposed hybrid approach for semantic similarity graph using both knowledge source and linguistic features. Similarly, Han, Lv, Hu, Wang, and Wang (2016), used Frame-Net and word embedding to measure sematic similarity in semantic graph model for extractive text summarization. Mohamed and Oussalah (2019), introduced semantic graph-based ATS framework that support both single and multi-document generic summarization; the semantic similarity is determine using both SRL and Wikipedia knowledge. A semantic graph-based ATS model was proposed by Plaza and Díaz (2011), for summarization of biomedical documents, the text concepts were used as the graph nodes and the edges is established based on the relations between the concepts.

Discussion

The graph-based approach uses the graph structure to determine relation and ranks text sentences. The most common method to determine the degree of causality between sentences in the approach is similarity measure. The technique has been implemented for diverse type of summarizations, including single-document, multi-document, generic and query-specific. As a typical unsupervised technique, the method does not require training with annotated data, therefore less expensive to implement. The majority of the graph-based ATS algorithms do not depend on the semantic meaning of words, therefore easily applied to many languages. The method considers the relation of sentence with all other sentences in the documents from all positions for a final ranking; therefore, generate summary which are readable and coherent.

The research based the taxonomy on graph structure and classified the models into: static graph-based, dynamic graph-based, graph pruning-based, hypergraph-based, affinity graph-based, semantic graph-based, and multigraph-based models. The static graph-based models are the classical and still competitive and most commonly used models. The efficiency of an algorithm in the model is largely depends on the accuracy of the similarity calculation and ranking function. The static graph-based models are popular for their simplicity, ease of implementation and fast computation. The model has been successfully applied to both single-document and multi-document summarization and it is good in resource utilization. The dynamic graph-based model on the contrary, considers the time of sentences arrival into the document in modelling the graph and represent sentences with directed graph. The dynamic graph-based models generate summary with good readability but the models are usually led to a slow and complex graph representation. Like the static graph-based model, the approach is good for both single-document and multi-document summarization.

The graph pruning methods like triangle counting and graph entropy methods reduce the number of the graph nodes, thus improved the efficiency and accuracy of the graph search. But the technique suffered with the addition time complexity of pruning the graph. The model is good for generic extractive text summarization and the low number of the graph vertices improve the efficiency of the scoring and selection of sentences. And the model has an advantage of generating summaries with less redundancies. The resource utilization in the approach can be minimized using some implementation techniques like dynamic programming. Similarly, the affinity graph-based model improves the quality of generated summary by sourcing information from other document; but the model also has high computing time and resource utilization compares to original static graph-based model. The model exploits the technique of global voting and recommendation by considering the sentences resemblance with sentences from other documents on similar topics, thus makes the ranking process of text sentences more accurate.

The model is especially good for multi-document extractive summarization, in which many documents involve in the ranking and selection process and the generated summaries are highly informative. Likewise, the semantic graph-based models have more accurate similarity calculation, but the use of external database make the model slower and language dependent. The semantic similarity used by the model required linguistic tools and grammar of a particular language, thus make an algorithm proposed for one language very difficult to be modified for another language. On the other hand, hypergraph-based model has limited application, as it only used for query-focus summarization. But the process of determining the similarity in the model is powerful as it can group more than two sentences using hypergraph incidence. The different features of the graph-based model for extractive text summarization are analyzed in Table 6.

Table 6. Comparison of Various ATS Graph-Based Models

Model	Similarity Measure	Language Dependency	Strengths	Weakness
Static Graph-Based	lexical	no	simple implementation, fast computation, language independent	less readability
Dynamic Graph-Based	lexical	no	Coherency, good readability, language independent	additional computing time
Graph Pruning-Based	lexical	no	more accurate scoring due to small size of the graph, language independent	additional computing time
Hypergraph-Based	lexical	no	more accurate similarity calculation, language independent	applied only for query-focus summarization
Affinity Graph-Based	lexical	no	high coverage, language independent	slow computation, poor readability
Semantic Graph-Based	semantic	yes	good similarity scoring	requires external knowledge source, language dependent
Multigraph-Based	lexical	no	fast computation, language independent	Less accurate scoring

Conclusion

The field of ATS has been studied for more than 60 years, but still remain of one the most challengeable areas in natural language processing and information retrieval. There are many approaches for ATS but graph-based are prefer by many, for their less cost and language independency. The graph-based models are classified into: static graph-based, dynamic graph-based, graph pruning-based, hypergraph-based, affinity graph-based, semantic graph-based, and multigraph-based models. All the model has their pros and cons; a choice of a model depends on the human language and domain of summarization.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

American Library Association, Office of Intellectual Freedom. (2013). *Banned and challenged books*. Retrieved September 15, 2015, from http://www.ala.org/advocacy/banned/bannedbooksweek

Abualigah, L., Bashabsheh, M. Q., Alabool, H., & Shehab, M. (2020). Text summarization: A brief review. *Stud. Comput. Intell.*, *874*, 1-15.

Aker, A. (2013). Entity type modeling for multi-document summarization: generating descriptive summaries of geo-located entities. *A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy to Department of Computer Science University of Sheffield*.

Al-Khassawneh, Y. A., Salim, N., & Jarrah, M. (2017). Improving triangle-graph based text summarization using hybrid similarity function. *Indian Journal of Science and Technology, 10*(8). doi:10.17485.

Alami, N., El Adlouni, Y., En-Nahnahi, N., & Meknassi, M. (2018) Using statistical and semantic analysis for Arabic text summarization. In*: Vol. 640. Advances in Intelligent Systems and Computing* (pp. 35-50).

Altmami, N. I., & Menai, M. E. B. (2018b) Semantic graph based automatic summarization of multiple related work sections of scientific articles. In*: Vol. 11089 LNAI. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 255-259).

Altmami, N. I., & Menai, M. E. B. (2020). Automatic summarization of scientific articles: A survey. *Journal of King Saud University – Computer and Information Sciences*

AlZahir, S., Fatima, Q., & Cenek, M. (2015). New graph-based text summarization method. *Paper presented at the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM).*

Alzuhair, A., & Al-Dhelaan, M. (2019). An approach for combining multiple weighting schemes and ranking methods in graph-based multi-document summarization. *IEEE Access, 7*, 120375-120386.

Aries, A., Zegour, D. E., & Hidouci, W. K. (2019). Automatic text summarization: What has been done and what has to be done. *arXiv:1904.00688v1 [cs.CL] 1*.

Barrios, F., López, F., Argerich, L., & Wachenchauzer, R. (2016). Variations of the similarity function of textrank for automated summarization. *arXiv:1602.03606 [cs.CL]*, 65-72.

Baxendale, P. B. (1958). Machine-made index for technical literature: an experiment. *IBM J. Res. Dev., 2*(4).

Begum, N., Fattah, M., & Ren, F. (2009). Automatic text summarization using support vector machine. *International Journal of Innovative Computing, Information & Control: IJICIC, 5*(7), 1987-1996.

Bhattacharya, P., Hiware, K., Rajgaria, S., Pochhi, N., Ghosh, K., & Ghosh, S. (2019). A comparative study of summarization algorithms applied to legal case judgments. *paper presented at the advances in information retrieval, cham.*

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*, 1-7.

Cai, X., & Li, W. (2013). Ranking through clustering: an integrated approach to multi-document summarization. *IEEE Transactions on Audio Speech & Language Processing, 21*(7), 1424-1433.

Carbonell, J., & Goldstein, J. (1998). The use of mmr, diversity-based reranking for reordering documents and producing summaries. *paper presented at the 21st annual international acm sigir conference on research and development in information retrieval.*

Dalal, V., & Malik, L. (2018) Semantic graph based automatic text summarization for hindi documents using particle swarm optimization. *in: vol. 84 (pp. 284-289): springer science and business media deutschland gmbh.*

Dutta, S., Chandra, V., Mehra, K., Ghatak, S., Das, A. K., & Ghosh, S. (2019). Summarizing microblogs during emergency events: A comparison of extractive summarization algorithms. *paper presented at the international conference on emerging technologies in data mining and information security (iemis 2018), kolkata, india.*

Edmundson, H. P. (1969). New methods in automatic extracting. *Journal of the ACM*, *16*(2), 264–285.

El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2020). EdgeSumm: Graph-based framework for automatic text summarization. *Information Processing and Management*, *57*(6). doi: 10.1016/j.ipm.2020.102264

El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*.

Elbarougy, R., Behery, G., & Khatib, A. E. (2019). Extractive Arabic text summarization using modified PageRank algorithm. *Egyptian informatics journal*.

Elrefaiy, A., Abas, A. R., & Elhenawy, I. (2018). Review of recent techniques for extractive text summarization. *Journal of Theoretical and Applied. Information Technology, 96*, 7739–7759.

Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research, 22*, 457-479.

Fakhrezi, M. F., Bijaksana, M. A., & Huda, A. F. (2021). Implementation of automatic text summarization with TextRank method in the development of al-qur'an vocabulary encyclopedia. *5th international conference on computer science and computational intelligence 2020, 179*, 391–398.

Fang, C., Mu, D., Deng, Z., & Wu, Z. (2017). Word-sentence co-ranking for automatic extractive text summarization. *Expert Systems with Applications, 72,* 189-195. doi: 10.1016/j.eswa.2016.12.021

Gallo, M., Popelínský, L., & Vaculík, K. (2018). To text summarization by dynamic graph mining. *CEUR Workshop Proceedings, 2203*, 28–34.

Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: A survey. *Artificial Intelligence Review, 47*(1), 1-66. doi: https://doi.org/10.1007/s10462-016-9475-9

Gong, Y., & Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. *paper presented at the annual international acm sigir conference on research and development in information retrieval.*

Gunawan, D., Pasaribu, A., Rahmat, R. F., & Budiarto, R. (2017). Automatic text summarization for Indonesian language using textteaser. *paper presented at the iop conference series: materials science and engineering.*

Gupta, S., & Gupta, S. K. (2019). Abstractive summarization: An overview of the state of the Art. *Expert System and Application, 121*, 49–65.

Han, X., Lv, T., Hu, Z., Wang, X., & Wang, C. (2016). Text Summarization Using FrameNet-Based Semantic Graph Model *Hindawi Publishing Corporation Scientific Programming*.

Hark, C., & Karcı, A. (2020). Karcı summarization: A simple and effective approach for automatic text summarization using Karcı entropy. *Information Processing and Management, 57*(3). doi: 10.1016/j.ipm.2019.102187

Hassan, B., Abdelrahman, S. E., Bahgat, R., & Farag, I. (2019). UESTS: An unsupervised ensemble semantic textual similarity method. *IEEE Access*.

Herings, P. J., Laan, G. V. d., & Talman, D. (2001). Measuring the power of nodes in digraphs. *Technical report, Tinbergen Institute*.

Hou, L., Hu, P., & Bei, C. (2017). Abstractive document summarization via neural model with joint attention. *Paper presented at the Natural Language Processing and Chinese Computing, Dalian, China.*

Hou, S., & Lu, R. (2020). Knowledge-guided unsupervised rhetorical parsing for text summarization. *Information Systems*.

Hu, P., He, J., & Zhang, Y. (2015). Graph-based query-focused multi-document summarization using improved affinity graph. In W. M. Zhang S., Zhang Z. (eds) (Ed.), *knowledge science, engineering and management. ksem 2015. lecture notes in computer science* (Vol. 9403): Springer, Cham.

Jacquenet, F., Bernard, M., & Largeron, C. (2019). Meeting summarization, a challenge for deep learning. Paper presented at the advances in computational intelligence, Cham.

Kanitha, D. K., Mubarak, D. M. N., & Shanavas, S. A. (2018). Malayalam text summarization using graph-based method. *International Journal of Computer Science and Information Technologies, 9*(2), 40-44.

Khan, A., Salim, N., Farman, H., Khan, M., Jan, B., Ahmad, A., & Paul, A. (2018). Abstractive text summarization based on improved semantic graph approach. *International Journal of Parallel Programming, 46*(5), 992–1016. doi: https://doi.org/ 10.1007/s10766-018-0560-3

Kleinberg, J. M. (1999). Authoritative sources in a hyper linked environment. *Journal of the ACM, 46*(5), 604-632.

Kumar, A., & Sharma, A. (2019). Systematic literature review of fuzzy logic-based text summarization. *Iranian journal of fuzzy systems, 16*(5), 45-59.

Lierde, H. V., & Chow, T. W. S. (2019). Query-oriented text summarization based on hypergraph transversals. *Information processing and management, 56*.

Lin, H., & Ng, V. (2019). Abstractive summarization: a survey of the state of the art. Paper presented at the thirty-third aaai conference on artificial intelligence (AAAI-19).

Liu, Y., Safavi, T., Dighe, A., & Koutra, D. (2018). Graph summarization methods and applications: A survey. *ACM computing surveys (CSUR), 51*(3).

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM journal of research and development*, 159–165.

Mallick, C., Das, A. K., Dutta, M., Das, A. K., & Sarkar, A. (2018). Graph-based text summarization using modified TextRank. *paper presented at the soft computing in data analytics, advances in intelligent systems and computing.*

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into texts. *In proceedings of the conference on empirical methods in natural language processing*, 404-411.

Miranda-Jiménez, S., Gelbukh, A., & Sidorov, G. (2013). Summarizing conceptual graphs for automatic summarization task. In H. D. Pfeiffer, D. I. Ignatov, J. Poelmans, & N. Gadiraju (Eds.), *conceptual structures for stem research and education. Iccs 2013. Lecture notes in computer science* (Vol. 7735). Berlin, Heidelberg: Springer.

Mohamed, M., & Oussalah, M. (2019). SRL-ESA-TextSum: A text summarization approach based on semantic role labeling and explicit semantic analysis. *Information processing and management, 56*(4), 1356-1372. doi: 10.1016/j.ipm.2019.04.003

Mojrian, M., & Mirroshandel, S. A. (2021). A novel extractive multi-document text summarization system using quantum-inspired genetic algorithm: MTSQIGA. *Expert systems with applications, 171*.

Moradi, M. (2018). Frequent item sets as meaningful events in graphs for summarizing biomedical texts. *Paper presented at the 8th international conference on computer and knowledge engineering (ICCKE), 2018.*

Moradi, M., Dashti, M., & Samwald, M. (2020). Summarization of biomedical articles using domain-specific word embeddings and graph ranking. *Journal of Biomedical Informatics, 107*.

Moratanch, N., & Chitrakala, S. (2017). A survey on extractive text summarization. *Paper presented at the 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai.*

Mosa, M. A., Anwar, A. S., & Hamouda, A. (2019). A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms. *Knowledge-Based Systems*, *163*, 518–532.

Mussina, A., Aubakirov, S., & Trigo, P. (2018). Automatic document summarization based on statistical information. *Paper presented at the 7th International Conference on Data Science, Technology and Applications (DATA 2018).*

Narayan, S., Cohen, S. B., & Lapata, M. (2018). Ranking sentences for extractive summarization with reinforcement learning. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*, 1747–1759.

Narayan, S., Cohen, S. B., & Lapata, M. (2019). What is this article about? Extreme summarization with topic-aware convolutional neural networks. *Journal of Artificial Intelligence Research, 66*, 243-278.

Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management, 56*(6).

Natesh, A. A., Balekuttira, S. T., & Patil, A. P. (2016). Graph based approach for automatic text summarization. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(2), 6-9.

Nawaz, A., Bakhtyar, M., Baber, J., Ullah, I., Noor, W., & Basit, A. (2020). Extractive text summarization models for Urdu language. *Information Processing and Management*.

Nazari, N., & Mahdavi, M. A. (2019). A survey on automatic text summarization. *Journal of AI and Data Mining*, *7*, 121-135.

Patil, K., & Brazdil, P. (2007). Text summarization: using centrality in the pathfinder network. *Paper presented at the IADIS International Conference Applied Computing.*

Plaza, L., & Díaz, A. (2011). Using semantic graphs and word sense disambiguation techniques to improve text summarization. *Procesamiento de Lenguaje Natural, 47*, 97-105.

Plaza, L., Díaz, A., & Gervás, P. (2011). A semantic graph-based approach to biomedical summarization. *Artificial Intelligence in Medicine, 53*(1), 1-14. doi: 10.1016/j.artmed.2011.06.005

Saziyabegum, S., & Sajja, P. S. (2016). Literature review on extractive text summarization approaches. *International Journal of Computer Applications (0975 – 8887), 156-No12*.

Saziyabegum, S., & Sajja, P. S. (2017). Review on text summarization evaluation methods. *Indian Journal of Computer Science and Engineering (IJCSE), 8 No. 4*.

Sevilla, A. F. G., Fernández-Isabel, A., & Díaz, A. (2016) Enriched semantic graphs for extractive text summarization. In*: Vol. 9868 LNAI. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 217-226).

Sikder, R., Hossain, M. M., & Robi, F. M. R. H. (2019). Automatic text summarization for Bengali language including grammatical analysis. *International Journal of Scientific & Technology Research, 8*(6).

Suleiman, D., & Awajan, A. A. (2019). Deep learning based extractive text summarization: Approaches, datasets and evaluation measures. *Paper presented at the 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS).*

Tandel, J., Mistree, K., & Shah, P. (2019). A review on neural network based abstractive text summarization models. *Paper presented at the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT.*

Uçkan, T., & Karcı, A. (2020). Extractive multi-document text summarization based on graph independent sets. *Egyptian Informatics Journal*.

Ullah, S., & Al Islam, A. B. M. A. (2019). A framework for extractive text summarization using semantic graph-based approach. *Paper presented at the ACM International Conference Proceeding Series.*

Verma, P., & Verma, A. (2020). A review on text summarization techniques. *Journal of Scientific Research, 64*(1).

Vollmer, M., Golab, L., Böhm, K., & Srivastava, D. (2019). Informative summarization of numeric data. *Paper presented at the 31st International Conference on Scientific and Statistical Database Management (SSDBM '19), Santa Cruz, CA, USA.*

Wan, X., & Yang, J. (2006). Improved affinity graph based multi-document summarization. *Paper presented at the Human Language Technology Conference of NAACL.*

Wang, K., Liu, T., Sui, Z., & Chang, B. (2017). Affinity-Preserving Random Walk for Multi-Document Summarization. *Paper presented at the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark.*

Wang, S., Zhao, X., Li, B., Ge, B., & Tang, D. (2017). Integrating Extractive and Abstractive Models for Long Text Summarization. *Paper presented at the Proceedings - 2017 IEEE 6th International Congress on Big Data, Big Data Congress 2017.*

Wang, W., Li, S., Li, J., Li, W., & Wei, F. (2013). Exploring hypergraph-based semi-supervised ranking for query-oriented summarization. *Information Sciences, 237*, 271-286.

Wang, W., Wei, F., Li, W., & Li, S. (2009). Hypersum: hypergraph based semi-supervised sentence ranking for query-oriented summarization. *Paper presented at the 18th ACM conference on information and knowledge management*.

Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Abdul Syukur a, Affandy, A., & Setiadi, D. R. I. M. (2020). Review of automatic text summarization techniques & methods. *Journal of King Saud University – Computer and Information Sciences*.

Woloszyn, V., Machado, G. M., Wives, L. K., & Mo, J. e. P. (2018). Modeling comprehending and summarizing textual content by graphs. *arXiv:1807.00303v1 [cs.CL]*.

Xiong, S., & Ji, D. (2016). Query-focused multi-document summarization using hypergraph-based ranking. *An International Journal of Information Processing and Management, 52*(4).

Yao, J.-g., Wan, X., & Xiao, J. (2017). Recent advances in document summarization. *Knowledge & Information Systems*, 1-40.

Zamana, F., Shardlow, M., Hassan, S.-U., Aljohani, N. R., & Nawaz, R. (2020). HTSS: A novel hybrid text summarization and simplification architecture. *Information Processing and Management*.

Zhong, S.-h., Liu, Y., Li, B., & Long, J. (2015). Query-oriented unsupervised multi-document summarization via deep learning model. *Expert Systems with Applications*, 8146 -8155.

Ziheng, L. (2007). Graph-based methods for automatic text summarization. (Ph.D.), National University of Singapore,

Bibliographic information of this paper for citing:

Bichi, Abdulkadir Abubakar; Samsudin, Ruhaidah; Hassan, Rohayanti & Almekhlafi, Khalil (2022). Graph-Based Extractive Text Summarization Models: A Systematic Review. *Journal of Information Technology Management*, Special Issue, 184-202.