ISSN 1870-4069

Comparative Study of Text Summarization Approaches

Amina Merniz, Anja Habacha Chaibi, Henda Hajjami Ben Ghezala

University of Manouba,
RIADI Laboratory, National School of Computer Science (ENSI),
Tunisia

{anja.habacha,henda.benghezala}@ensi.rnu.tn
amina.merniz@ensi-uma.tn

Abstract. Since 1958, automatic text summarization has been an interesting field that emerged to be more effective considering the exponential growth of information on the web. An intensive research on the matter of this subject has been conducted last year. Text summarization is the process of generating concise summary of the original text by reserving the main ideas and the relevant information. The paper presents the different types of automatic text summarization namely: (single document or multi-document), (generic or request oriented), (Abstractive or Extractive), (monolingual , multilingual or cross-lingual), in addition to the classification of techniques used to produce a summary, including limits and advantages of the related works in each category has been covered. A comparative study of recent text summarization systems has been introduced, which demonstrates that most of the studies focused on extractive text summarization, however summaries in the Arabic are still limited. In addition the thematic aspect was not taken into consideration. The objective of the paper is to help researchers to concentrate on those limitations to decide their future directions.

Keywords: Automatic text summarization, extractive, abstractive, monolingual, multilingual.

1 Introduction

In the last years, we live in a huge increase in the amount of data, which allows the data science to explore them also to extract knowledge. This continuous growth of information on the web requires developers to search for a way to retrieve information even to provide accurate and complete idea of document content. Hence they are oriented towards the construction of automatic document text summarization started with the generation of single document text summarization Luhn (1958) [1], then passed to multi-document text summarization [5, 32]. Several techniques have been generated in this area until now. An automatic text summarization should be relevant, concise, and shorter than the original text. For multi-document text summarization there are some issues to be avoided such as redundancy, meaning, sentence order, etc.

Thereby making automatic text summarization still challenging task. The remainder of paper is presented as follow: Section 2 defines different types of text summarization. Section 3 describes classification of related works according to the techniques used in text summarization with limits and advantages of each related work. Section 4 discusses the comparative study of these approaches. Finally, the conclusion is presented in Section 5.

2 Types of Text Summarization

We define summary as one of the crucial tasks in Natural language processing. Which consists in abbreviating one or several texts in a shorter version, by reserving only the main ideas and the most relevant information. In order to facilitate the task of reading to the user, as well as help him to obtain a clear, meaningful, and accurate view about the content of the source text. We classify a summary depending on the following categories: based on documents numbers (single or multi-document), based on summary usage (generic or query request), based on characteristics of summary as text (abstractive or extractive), based on language (monolingual, multilingual or cross-lingual).

Based on the number of documents: single document (**S**) it takes only a single document as input to summarize and also generates a single document. Multi-document (**M**) it takes as input a collection of documents, produces the summary then output a single document. Based on summary usage: generic summary (**G**) The summary produced based purely on the text source. Query-oriented (**O**) the summary is guided by a user's request.

In that case, the system must first locate on the set of documents the passages involved with the user request, then produces the summary [2]. Based on the characteristics of summary as text: abstractive summary (**A**) the term abstractive is used to describe a summary, that requires a detailed study of the text. It can synthesize a short version of the original sentence even to add new vocabulary or a new sentence not included in the final source text.

The objective thus of the abstractive summary is to reduce the redundancy and improve the compression ratio [3]. Extractive summary (**E**) Most studies have based on the construction of extractive summary because it avoids the problem of generating the text which is always regarded as(considered) a very complex task. It is based on statistical analysis to assign scores to original text sentences to extract the most frequents ones to produce the summary [3].

Based on language: Monolingual (**ML**) the language of the source document and the target document are identical. Multilingual (**MUL**) The source document is in different languages (English, Arabic, French or others), the summary is also generated in these languages. Cross-lingual (**Cross**): The source document is in the language while the generated summary is in another language different from that of the source text.

3 Techniques of Text Summarization

3.1 Statistical Based Approaches

This approach is simple, it consists of extracting keywords from text documents. Based on statistical features such as TF (Term Frequency), TF-IDF (Term Frequency-Inverse Document Frequency), POK (Position of Keyword), and others. Its principle is to compute a score characterizing the importance of each textual unit (sentence, paragraph, ...), then retain units that have a score above a certain threshold, until reaching the rate or percentage usually defined by the user [25]. Fukumoto (2004) [5], proposed a summary system which automatically classifies documents into three types: “one topic type”, “multi-topic type”, or “others”.

Both single, also multi-document summary, in the first case, it is based on TF-IDF and sentence position in the document to assign weight to sentences and extract those with higher scores and eliminate non-important sentences. In the case of multi-document summarization, it exploits and reapplies the technique used for the single-document summarization for each document then generates the final summary. This system uses a simple strategy to create the summary.

The limit of this work is that the implementation involves some system bugs in the classification mechanism. Ouyang et al. (2009) [6], create a new hierarchical tree representation of words based on the most frequent terms at the top of the hierarchy. They estimate the words sense on the tree and extract the sentences able to integrate various objectives of multi-document to generate a relevant summary. The benefit of the generated system is that the idea of incorporating goals of multi-document summary into one framework is meaningful.

The disadvantages of the system are that it fails to create better summaries on some document sets, the constructed hierarchical tree can’t always represent the concepts for some documents sets, another problem is that the two constraints used in the algorithm of tree construction are still not correct in the real data. Gupta et al. (2012) [7], present a statistical text summarization approach using the kernel of the original text. The kernel-based system, called Kernel- Sum (KERNEL SUMMarizer), uses the kernel as a guide or guideline for identifying and selecting segments of text to be included in the summary.

The determination of kernel sentences relies on simple statistical methods namely: kernel Key based on the identification of keywords, and kernelTFISF based on the inverse distribution of sentences in the source text. The size of the summary can be specified by the user when calling the tool. The proposed approach includes the kernel. The extracts convey well the main idea of the source texts. The inconvenient of the proposed summary is that many authors have stressed the need to convey the main idea and justification of the results in the automatic summary.

El-Haj et al. (2013) [24], presented a general, extractive, multilingual summary for both types of summary single as well as multiple. They used the same pipeline processing for both summaries. A log-likelihood score is computed according to dataset words. Log-likelihood helps to identify words that are frequently unexpected. Summaries are constructed by selecting the highest sentences sum of the log-likelihood scores of their words.

In this work, single-document summaries, in English and Arabic, have worked particularly well in the automatic evaluation. They are ranked first and second respectively. The limitations consist the low scores of automated assessment of Arabic and English multi-document summaries due to two main reasons. First, they treated all the multiple documents as one big document. Second, they did not work to eliminate redundancy. Finally, the log likelihood score could be improved by the inclusion of a scatter score or weight to examine the regularity of the spread of each word in all documents. Bhatia et al. (2015) [25], proposed system for single document summarization.

The source document is converted to a Universal Networking Language (UNL) document by the En-Conversion process. The textual summary algorithm is applied to the UNL document. This summary document is provided to EUGENE for conversion to a different natural language. The proposed algorithm consists of deleting the relations that are not important to minimize the complexity, Then calculate the score of each UNL sentence. Each UNL sentence consists of words and universal connection ships. The sentence score is the total of universal word weights, the weight of these words is computed using TF-IDF.

Then sentences are chosen according to their calculated score and the document size, the phrases having the high score are included in the summary. After deleting the redundancy in the selected sentences, they combine those who have the same universal word as the head to form a complex and meaningful sentence. Finally, the summary obtained is processed again before it is passed to DeConverter for further processing. The limitation of the system is that the algorithm has not been applied to the multilingual document, which makes its performance weaker and unjustified.

3.2 Machine Learning Based Approaches

Approaches of machine learning are organized into classes namely : supervised, unsupervised, or semi-supervised. The first category has a set of documents and their corresponding references generated by humans. In summarization sentence is labeled as correct when it belongs to the reference summary, or as incorrect in the opposite case, using a collection of learning documents the most popular techniques used in this category are: Nave Bayes classification, mathematical regression, neural networks.

In the second class, systems do not involve training data. They produce the final summary based only on the target documents. They determine the hidden structure in unlabeled data. Thus, they are suitable for all newly observed data without any advanced modification. Such systems are guided by heuristic rules to select very relevant sentences to be included to the summary, Clustering, Markov's model, etc. are examples of unsupervised techniques, for the last category, it requires labeling or/and unlabeled data to generate an appropriate classifier.

Agarwal et al. (2011) [9], the summary system is called Scisumm, it applies Texttiling algorithm to segment the input documents. Then generate the labeled clusters from the segments, using an algorithm based on clustering and various terms. The groups are classified according to relevance to the query generated by the user. This is achieved by a ranking module that uses the cosine similarity between the question and the centroid of each cluster.

Finally, the blade of summary generation display the groups obtained. The approach is based on clustering which is extremely fast in execution time, so it is relatively efficient regarding space required gives ends, a set of many names is generated for each cluster which a comprehensible cluster description. The inconvenient of the system is that it creates summaries of scientific articles. The order of groups based on relevance must be improved to make the set of observations of the cited articles more diverse. Li et al. (2007) [10], extract sentences from the original text and reorganize them into a summary taking into consideration the query.

The process of extracting sentences depends on various characteristics; the SVR (Vector Support Regression) approach is used to combine these characteristics. We note that the Vector Support Regression (SVR) template is used to connect features and mark sentences automatically, the appropriate lexical and syntactic characteristics are adopted and SVR appropriately assigns the weight parameters. The limitations consist that many sentence simplification and reorganization methods are not introduced, the performance in responsiveness evaluations is not good.

Yong et al. (2011) [11], based on neural networks to generate summary in three steps: A text preprocessing subsystem, a keyword extraction subsystem, a summary production subsystem. The benefit of the proposed system is that the competitive network architecture of the system is carefully designed as it directly affects the suitable system's output. In contrast, the system does not cover all kinds of documents such as: legal documents, and financial, also the considered system generates the summary without compromising readability.

Nallapati et al. (2016) [12], present SummaRuNNer Recurrent Neural Network Encoder-Decoder based on sequence model for text summarization. In which they propose models for construction summary problems such as: capturing keywords, modeling rare or invisible keywords, and the capture of document structure hierarchical. This article offers several solutions for text summarization problems many of these solutions have contributed to improved performance.

While the model is misinterpreting the semantics of the text, capturing the meaning of complex sentences remains a weakness for this model, the same phrase is often repeated in summary. Fejer and Omar (2014) [26], propose single and multi-document Arabic text summarization, based on clustering algorithms namely: the agglomerative hierarchical classification algorithm with (single link and complete link) and k-means. Then extract key phrases from each cluster, reorganize and classify them.

The similarity algorithms, namely the cosine similarity and the Jaccard coefficient, are used to select a sentence from each set of similar sentences ignoring other sentences. These sentences will generate the final summary. The disadvantages of the system consist that single-document summarization results are efficient compared to multi-document summarization, the hierarchical algorithm applied is less efficient than other clustering techniques.

3.3 Linear Programming Based Approaches

McDonald was the first who introduce linear programming in automatic document summarization [33]. McDonald's principle is to maximize an objective function satisfying the selection criteria and penalizing the redundancy between the selected units. The calculation is effected using a set of constraints according to the summary system. Alguliyev et al. (2011) [13], present text summarization approach modeled as a linear programming problem called MCMR. This model tries to optimize relevance, redundancy and summary length. The approach applies to both singles as well as multi-document. The system discovers key phrases in the given document (s), it covers the main content of the input document (s). Also, it reduces redundancy. But models result depends directly on the optimization algorithm. Banerjee et al. (2015) [14], propose an abstractive summarizes begins with identifying the most important document in the multi-document set.

Each sentence in the most important document is initialized in clusters distinct, each sentence in the other documents are assigned to the cluster that has the highest similarity to that sentence. A word graph is generated from the sentences in each cluster; several paths can be extracted from each word graph, in which case they choose the shortest K-paths. To make the most informative and linguistically well-formed sentences, they use an ILP (Integer Linear Programming) based approach, including only the paths that maximize the content of information and linguistic quality of summary.

The ILP system can combine the knowledge of different sentences and present a readable summary; this approach can generate informative summaries by maximizing the selection of content from several sentence clusters, they integrate the linguistic quality and the information to select the coherent sentences in the final summary using this ILP-based approach. The limit in the linguistic variety of the reviews generated is not yet improved; the summary produces incoherent grammatical sentences. Berg-Kirkpatrick et al. (2011) [27], they are learning a model of extract sentences and compression to multi-document summarization.

Their model marks the candidate summaries according to a combined linear model whose features take into consideration the types of n-grams in the summary and the compressions used. Inference in their model can be expressed as ILP and resolved in a reasonable time. The originality of this system that it can use or create reasonable sentences of average length. The Limit is the joint extraction and compression system produce short and less productive sentences. Oliveira et al. (2016) [28], propose an ILP concept-based approach for single document summarization.

Such an approach maximizes the coverage of the important concepts, in summary, avoiding redundancy, and taking into consideration informativeness and readability aspects of the generated summary. The readability of the output summary is improved by incorporating into the ILP model a specific constraint concerning the resolution of dangling co-references and speech analysis. The limit is the inclusion of consistency constraints in ILP models to eliminate co-references and discourse analysis has decreased system performance.

Zhang et al. (2016) [34], propose abstractive cross-lingual summarization framework. The source documents are translated using machine translation system. The approach generates bilingual concepts represented using bilingual elements of source predict, argument structures (PAS) and their target counterparts. In order to maximize the translation quality of (PAS) elements in the final summary, a linear programming algorithm has been applied. The proposed framework can combine relevant information from different original sentences by fusing PAS sharing the same concept.

But the system produces the summary sentence by blending several original sentences and this may violate the correct word ordering.

3.4 Graph Based Approaches

The text is presented using a graph where the vertices are the textual units (concepts or sentences) while the arcs represent the adjacency or semantic relations between nodes. The most popular approaches based graph are: HITS and Google Page Rank. Wan (2008) [15], examines the impact of the document on summarization performance using the document importance and sentence- to-document correlation based on graph ranking process. This approach assumes that the sentences related to an important document and are strongly correlated with this document will be selected in summary.

In order to incorporate the document-level information and the sentence- document relationship, they proposed two-layer links graph models including both sentences and documents. Results show the robustness of the proposed model. The Limit is that The parameters of the proposed model can deteriorate the summary performance. Zhang et al. (2005) [16], generate a new method based on the hub-authority framework. That unites the text content with some cues and explores the subtopics using graph-based sentence ranking algorithm to produce the expected output.

The advantage of the approach is a useful graph-ranking schema in multi-document generic text summarization. Also, top-ranked hub words can be used as keywords to identify document topics in particular applications. The limit of the proposed system is that it's difficult to determine the subtopics. Patil and Brazdil (2007) [17], propose a system called (Summgraph). In this system, the text is presented as a graph where nodes represent the document sentences while the weights on links present the intra-sentence dissimilarity.

This system is focused on the use of the Pfnet concept (PathFinder Network Scaling) to calculate the importance of a sentence in the text. The limitation consists the inability of the model to improve performance. Thakkar et al. (2010) [18], present graph-based methods for text summarization. A score is computed for nodes using graph ranking algorithms: HITS and Page Rank. The algorithm of the shortest path has also been applied to generate the summary.

It is easy to implement, language independent, and it makes an efficient outline by including the most significant parts of the original text. The inconvenient is that Considering the shortest paths for choosing summary sentences may not be enough. Heu et al. (2015) [29], propose "FoDoSu" multi-document text summarization system. Word analysis words frequency and analyzes semantic words importance by exploiting the Flickr tag clusters. The contribution of the word in a document is calculated using the HITS algorithm.

The module of sentence analysis generates the summary by selecting only the highest ranked sentences. The FoDoSu system performs with a low calculation cost during the phase of semantic analysis of the words in the document, FoDoSu effectively selects the sentences containing significant words by exploiting the HITS algorithm with tag clusters during document summaries. FoDoSu analyzes the proper names. The disadvantages consist methods used for semantic analysis of words that are difficult to analyze, such as proper nouns and newly coined words are insufficient.

3.5 Lexical Cohesion Based Approaches

It is primarily based on the cohesion relations between words, such as a lexical chain (LC), a score of a lexical chain of a word (LCS), WordNet (WN), etc. Chen et al. (2005) [19], study the use of lexical chains as a model of several documents written in Chinese to generate an easy and indicative summary. The algorithm of calculating lexical chains using HowNet knowledge database is changed to improve performance and adapt to Chinese compression.

Based on the semantic analysis, the algorithm can eliminate redundant similarities and maintain differences in information content between multi-documents. The approach has excellent performance in capturing meanings and topics of multi-documents, the plan is highly domain- independent, even though its power has illustrated mainly for news-wire texts. The presented system can be used in daily web texts. However the extracted sentences are modified by the algorithm, the summary generated may not be as flexible, concise and coherent.

AL-Khawaldeh and Samawi (2015) [20], develop an LCEAS system in four phases, namely: pre-processing of the text (elimination of stop words, ...), word sense disambiguation: identify a real sense of words. Lexical cohesion based segmentation: distinguish important from non-important sentences in the text. And text-based segmentation for summarization: decide whether the sentence sense is derived from another sentence. The advantage of the system is the elimination of poor sentences in lexical cohesion with word sense disambiguation (WSD), and eliminating of redundant phrases.

The limit is that in the sentence analysis process semantic relationships are insufficient. Azmi and Al-Thanyyan (2012) [21], generate extractive Arabic summary. They create the first summary based on RST (Rhetorical Structure Theory), where they assign a score to each of first summary sentences to produce the final review. The system can let the user define the maximum size of the summary in the form of some words, the percentage of the original or number of sentences although the algorithm failed to generate a review for the document of size 12 and less. Saxena and Saxena (2016) [30], present extractive summary based on the lexical chain approach, wordnet knowledge database is used to identify semantic relations among words, also to create the lexical chains.

These chains will be classified to distinguish the strong ones based on the sense of the lexical chain in a document, the semantic relation between word and the lexical chain, and The utility of each lexical chain to specify its contribution in the report. The selection of sentences included to summary depends on strong lexical chains. The method is significant in extracting relevant or accurate sentences. The Limit is that the processing time which is a little more. The proposed approach takes some time for generating the summary after the complete procedure.

3.6 Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) Based Approaches

This approach uses algebraic theories namely the matrix, the transposition of the matrix, etc. The most used algorithms for constructing a text summarization based on this approach are LSA (Latent Semantic Analysis), and SDD (Semi-Discrete Decomposition). Wang and Ma (2013) [22], propose LSA-based Text summarization algorithm that Combines term description and sentence description for each topic then select most relevant sentences which include the terms that can best represent the topic. For each subject, several sentences are chosen for the summary, which allows to fully expressing the issue. The limitation consists of the sense of sentences.

Xiong and Luo (2014) [23], propose a novel method for evaluating a subset of sentences based on their capacity to reproduce word projections on singular vectors. The algorithm is computationally efficient. The inconvenient of the system is that it needs another method to produce a better summary. Conroy et al. (2013) [31], present methods for multilingual document summarization.

For term weighting of multilingual single document summarization, three approaches were presented namely: global entropy, the logarithm of frequency, and a personalized variant of TextRank. For multi-document summarization three algorithms were applied: well-known LSA, the more recent latent Dirichlet allocation (LDA) and a new interval non-bounded matrix factorization method (IB- NMF). The evaluation of single document summarization indicates that the approach significantly outperformed the baseline system.

In contrast, the LDA method for term weight was the weak- est of the three and therefore did not improve the performance of the system. Demirci et al. (2017) [32], produce an extractive multi-document summary for Turkish news. The news was collected from various web sources using the JSOUP and RSS Feed frameworks. The collected contents were ranked according to their cosine similarity score as the elements of similar topics.

An LSA based algorithm is applied to identify relations between concepts and sentences then select just the most important. The proposed approach is competitive for short texts. The limits consist some of words observed in the sentences which affect the scoring during LSA. The phrase that has more words considered the most important. The sentence length impacts its importance. Therefore results indicate that performance decreases for long texts.

4 Discussions

Several techniques have been utilized for automatic text summarization task as specified in Table 1. In addition, most of these studies focused on Extractive text summarization, only(Banerjee et al 2015) [14], (Zhang et al. 2016) [34], proposed an abstractive text summarization based on linear programming approach, as well as (Nallapati et al. 2017) [12], generated an abstractive summary using Machine Learning approach.

The comparative study also shows that the majority of proposed systems are monolingual systems in particular in English language, except (Azmi et al. 2012) [21], (Fejer et al. 2014) [26], (AL- Khawaldeh et al. 2015) [20], presented an Arabic text summarization, also one works in Japanese and another in Turkish were covered by (Fukumoto 2004) [5], and(Demirci et al. 2017) [32] respectively. We also noted that there is a considerable lack of multilingual text summarization (Conroy et al. 2013) [31], the only who produced a summary in nine languages, then (Gupta 2012)

Table 1. Recent automatic text summarization.

Authors	Summarization approach	**S**	**M**	**G**	**O**	**E**	**A**	**ML**	**MUL**	**Cross**
Fukumoto 2004 [5]	Statistical	×	×	×		×		× Japanese		
Yong et al 2006 [11]	Machine Learning	×		×		×		× English		
Zhang 2005 [16]	Graph		×	×		×		× English		
Chen et al 2005 [19]	Lexical cohesion		×	×		×		× Chinese		
Li et al 2007 [10]	Machine learning		×		×	×		× English		
Patil et al 2007 [17]	Graph	×		×		×		× English		
Wan 2008 [15]	Graph		×	×				× English		
Ouyang 2009 [6]	Statistical		×		×	×		× English		
Thakkar et al 2010 [18]	Graph	×		×		×		× English		
Agarwal et al 2011 [9]	Machine Learning		×		×	×		× English		
Alguliyev et al 2011 [13]	Linear Programming	×	×	×		×		× English		
Berg-Kirkpatrick et al 2011 [27]	Linear Programming		×	×		×		× English		
Gupta 2012 [7]	Statistical	×		×		×			× Portuguese	
									Brazilian	
									English	
Azmi et al 2012 [21]	Lexical Cohesion	×		×		×		× Arabic		
Wang et al 2013 [22]	LSA	×	×	×		×		× English		
Conroy et al 2013 [31]	LSA, LDA	×	×	×		×			× English	
									Arabic	
									French	
									Czech	
									Greek	
									Hebrew	
									Hindi	
									Spanish	
									Romanian	
									Chinese	
El hadj et al 2013 [24]	Statistical	×	×	×		×			× English	
									Arabic	
Fejer et al 2014 [26]	Machine Learning	×	×	×		×		× Arabic		
×iong et al 2014 [23]	LSA		×	×		×		× English		
Banerjee et al 2015 [14]	Linear Programming		×	×			×	× English		
AL-Khawaldeh et al 2015 [20]	Lexical Cohesion	×		×		×		× Arabic		
Bhatia et al 2015 [25]	Statistical	×		×		×			×	
Heu et al 2015 [29]	Graph		×	×		×		×		
Nallapati et al 2017 [12]	Machine Learning	×		×			×	× English		
Zhang et al 2016 [34]	Linear Programming		×	×			×			×
Saxena and Saxena 2016 [30]	Lexical Cohesion	×		×		×		×		
Oliveira et al 2016 [28]	Linear Programming	×		×		×		×		
Demirci et al 2017 [32]	LSA		×	×		×		× Turkish		

[7], and (El hadj et al. 2013) [24], introduced a multilingual summary included the English language. We also observed that the system of (Conroy et al. 2013) [31] explored the thematic and semantic aspect of text to generate the summary. However, most researchers did not take into count thematic aspect of the text as specified in Table 1. Due to the diversity of performance evaluation metrics of summary: rouge, precision, recall, f-measure. Some works used rouge with its variants [13, 23] while others utilized Precision, Recall and F-measure [32]. In addition these studies used several corpus to evaluate results of the summaries generated: DUC 2002 [23, 22, 17, 15], DUC 2004 [16, 12, 22], DUC 2005 [10, 13, 14], DUC 2006 [10], DUC2007 [10,

10], TSC3 [5], EASC [20]. The comparison of evaluation performance of summary is not significant since the studies did not use the same metrics and corpus.

5 Conclusion

This article presented a survey of text summarization techniques, the investigation has introduced different types of summary. Then a classification of great techniques used in this field has also covered in this paper, with a list of related works that included the limits and advantages of each proposed method in each category. Which is considered a significant contribution to this paper. The research studies have been compared in a tabular form, the purpose of the paper is to help researchers to have a clear view about necessary information of text summarization task, also to choose their appropriate frameworks based on this article.

In the future work, we are going to propose a novel approach for abstractive multi-document multilingual text summarization in particular (Arabic, French and English), with the consideration of the thematic aspect of the text based on additional external semantic resources. We are also going to study the comparison of performance evaluation metrics of text summarization with more details.

References

1. Luhn, H. P.: The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development, vol. 2 , pp. 159–165 (1958)
2. Chaar, N., Sana, L.: Filtrage d'information pour la Construction de Résumés Multi-Documents Guidée par le Profil Utilisateur: le Système REDUIT. Université de Marne-la-Vallée (2004)
3. Khan, A., Salim, N., Kumar, Y. J.: A Framework for Multi-Document Abstractive Summarization Based on Semantic Role Labelling. Applied Soft Computing. Elsevier, vol. 30, pp. 737–747 (2015)
4. Bharti, S. K., Babu, K. S.: Automatic Keyword Extraction for Text Summarization: a Survey (2017)
5. Fukumoto, J., Sugimura, T.: Multi-Document Summarization Using Document Set Type Classification. Proceedings of NTCIR (2004)
6. Ouyang, Y., Li, W., Lu, Q.: An Integrated Multi-Document Summarization Approach Based on Word Hierarchical Representation. Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Association for Computational Linguistics, pp. 113–116 (2009)
7. Gupta, V., Priya, C., Sohan, G., Anita, B., Shobha, K.: An Statistical Tool for Multi-Document Summarization. International Journal of Scientific and Research Publications, vol. 2 (2012)
8. Gambhir, M., Gupta, V.: Recent Automatic Text Summarization Techniques: a Survey. Artificial Intelligence Review. Springer, vol. 1, pp. 1–66 (2017)
9. Agarwal, N., Gvr, K., Reddy, R. S., Rosé, C. P.: Towards Multi-Document Summarization of Scientific Articles: Making Interesting Comparisons with SciSumm. Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages, Association for Computational Linguistics, pp. 8–15 (2011)
10. Li, S., Ouyang, Y., Wang, W., Sun, B.: Multi-Document Summarization using Support Vector Regression. Proceedings of DUC, Citeseer (2007)

11. Yong, S. P., Abidin, A. I., Chen, Y. Y.: A Neural-Based Text Summarization System. Transactions on Information and Communication Technologies, WIT Press, vol. 47 (2006)
12. Nallapati, R., Zhai, F., Zhou, B.: SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents, AAAI, pp. 3075–3081 (2017)
13. Alguliyev, R. M., Aliguliyev, R. M., Hajirahimova, M. S., Mehdiyev, C. A.: MCMR: Maximum Coverage and Minimum Redundant Text Summarization Model. Expert Systems with Applications. Elsevier, vol. 38, pp. 14514–14522 (2011)
14. Banerjee, S., Mitra, P., Sugiyama, K., Multi-Document Abstractive Summarization Using ILP Based Multi-Sentence Compression, pp. 1208–1214 (2015)
15. Wan, X.: An Exploration of Document Impact on Graph-Based Multi-Document Summarization. Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, pp. 755–762 (2008)
16. Zhang, J., Sun, L., Zhou, Q.: A Cue-Based Hub-Authority Approach for Multi-Document Text Summarization. Natural Language Processing and Knowledge Engineering, 2005, IEEE NLP-KE'05. Proceedings of 2005 IEEE International Conference on, pp. 642–645 (2005)
17. Patil, K., Brazdil, P.: Text Summarization: Using Centrality in the Pathfinder Network. Int. J. Comput. Sci. Inform. Syst. vol. 2, pp. 18–32 (2007)
18. Thakkar, K. S., Dharaskar, R. V., Chandak, M. B.: Graph-Based Algorithms for Text Summarization. Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference, pp. 516–519 (2010)
19. Chen, Y. M., Wang, X. L., Liu, B. Q.: Multi-Document Summarization Based on Lexical Chains. Machine Learning and Cybernetics, 2005, Proceedings of 2005 International Conference, vol. 3, pp. 1937–1942 (2005)
20. Al-Khawaldeh, F., Samawi, V.: Lexical Cohesion and Entailment Based Segmentation for Arabic Text Summarization (lceas). The World of Computer Science and Information Technology Journal (WSCIT), vol. 5, pp. 51–60 (2015)
21. Azmi, A. M., Al-Thanyyan, S.: A Text Summarizer for Arabic. Computer Speech & Language, vol. 26, pp. 260–273 (2012)
22. Wang, Y., Ma, J.: A Comprehensive Method for Text Summarization Based on Latent Semantic Analysis. Natural Language Processing and Chinese Computing, pp. 394–401 (2013)
23. Xiong, S., Luo, Y.: A New Approach for Multi-Document Summarization Based on Latent Semantic Analysis. Computational Intelligence and Design (ISCID), 2014 Seventh International Symposium. IEEE, vol. 1, pp. 177–180 (2014)
24. El-Haj, M., Rayson, P.: Using a Keyness Metric for Single and Multi Document Summarisation. Proceedings of the MultiLing 2013 Workshop on Multilingual Multi-document Summarization, pp. 64–71 (2013)
25. Bhatia, P., others.: Multilingual Text Summarization with UNL. Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances, pp. 740–745 (2015)
26. Fejer, H. N., Omar, N.: Automatic Arabic Text Summarization using Clustering and Keyphrase Extraction. Information Technology and Multimedia (ICIMU), 2014 International Conference on IEEE, pp. 293–298 (2014)
27. Berg-Kirkpatrick, T., Gillick, D., Klein, D.: Jointly learning to Extract and Compress. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, vol. 1, pp. 481–490 (2011)
28. Oliveira, H. Lima, R., Lins, R. D., Freitas, F., Riss, M., Simske, S. J.: A Concept-Based Integer Linear Programming Approach for Single-Document Summarization. Intelligent Systems (BRACIS), 2016 5th Brazilian Conference. IEEE, pp. 403–408 (2016)

29. Heu, J., Asim, I., Lee, D.: FoDoSu: Multi-Document Summarization Exploiting Semantic Analysis Based on Social Folksonomy. Information Processing and Management. Elsevier, vol. 1, pp. 212–225 (2015)
30. Saxena, S., Saxena, A.: An Efficient Method Based on Lexical Chains for Automatic Text Summarization. International Journal of Computer Applications, Foundation of Computer Science, vol. 144 (2016)
31. Conroy, J., Davis, S.T., Kubina, J., Liu, Y.K., O'leary, D.P., Schlesinger, J.D.: Multilingual Summarization: Dimensionality Reduction and a Step Towards Optimal Term Coverage. Proceedings of the MultiLing 2013 Workshop on Multilingual Multi-Document Summarization, Foundation of Computer Science, pp. 55–63 (2013)
32. Demirci, F., Karabudak, E., İlgen, B.: Multi-document summarization for Turkish news. Artificial Intelligence and Data Processing Symposium (IDAP), 2017 International, pp. 1–5 (2017)
33. McDonald, R.: A Study of Global Inference Algorithms in Multi-Document Summarization. European Conference on Information Retrieval, pp. 557–564 (2007)
34. Zhang, J., Zhou, Y., Zong, C.: Abstractive Cross-Language Summarization via Translation Model Enhanced Predicate Argument Structure Fusing. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), vol. 24, pp. 1842–1853 (2016)