



A Review of state-of-the-art Automatic Text Summarisation

Kartik Rathi, Dept. of Computer Science, SET, Sharda University, Greater Noida, India,

Saumy Raj, Dept. of Computer Science, SET, Sharda University, Greater Noida, India,

Yash Vardan Singh, Dept. of Computer Science, SET, Sharda University, Greater Noida, India,

Assoc. Prof. Sudhir Mohan, Dept. of Computer Science, SET, Sharda University, Greater Noida, India,

Abstract

Text summarisation comes under the domain of Natural Language Processing (NLP), which entails replacing a long, precise and concise text with a shorter, precise and concise one. Manual text summarising takes a lot of time, effort and money and it's even unfeasible when there's a lot of text. Much research has been conducted since the 1950s and researchers are still developing Automatic Text Summarisation (ATS) systems. In the past few years, lots of text-summarisation algorithms and approaches have been created. In most cases, summarisation algorithms simply turn the input text into a collection of vectors or tokens. The basic objective of this research is to review the different strategies used for text summarising. There are three types of ATS approaches, namely: Extractive text summarisation approach, Abstractive text summarisation approach and Hybrid text summarisation approach. The first method chooses the relevant statements out of the given input text or document & convolves those statements to create the final output as summary. The second method converts the input document into an intermedial representation before generating a summary containing phrases that differ from the originals. Both the extractive and abstractive processes are used in the hybrid method. Despite all of the methodologies presented, the produced summaries still lag behind human-authored summaries. By addressing the various components of ATS approaches, methodologies, techniques, datasets, assessment methods and future research goals, this study provides a thorough review for researchers and novices in the field of NLP.

Keywords: *Text summarisation, Abstractive, Extractive, Hybrid, Dataset.*

1. Introduction

Summarisation is the process of condensing a long piece of text into a shorter one, reducing the volume of the original text whilst maintaining important information and content significance. Because human text summarisation is indeed a stagnant and fundamentally tiresome process, automating it is becoming incredibly popular.

Text summarisation may help with a plethora of NLP tasks, such as text classification, information retrieval, legal text summarisation, main stream media summarisation and headline creation. Furthermore, the production of summaries might be embedded into these systems as a stage in the process, decreasing the size of the document.

In this era of big data, the abundance of textual data obtainable from diverse sources has increased. To be beneficial, this huge volume of data holds a plenitude of knowledge and skill that should be appropriately summed up. The increasing obtainability of documents necessitates extensive study in the domain of NLP for automatic text summarisation. It is the process of constructing a concise and vivid summary even without involvement of a human whilst maintaining the original text's meaning.

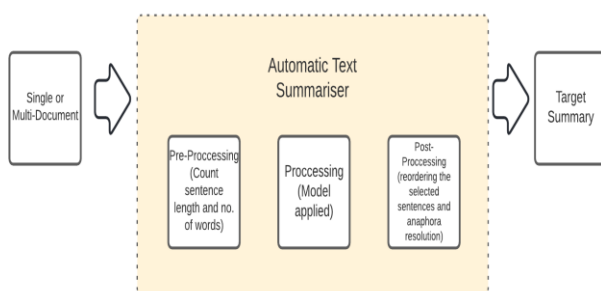
It is indeed challenging because, in an effort to create a summary of a literary piece, we usually read it in its entirety to get a clear grasp of it and then compose a summary, emphasizing its key themes. Automated text summarisation is a pretty difficult and stagnant process since computers deficit human language and cognition.

Automatic Text summarisation is a technique for compressing vast amounts of data-parallel holding up the ingenious elucidation of the data entered. Additionally, the data is structured in such a way that the reader has a thorough understanding of the huge text. People are turning to the web to obtain the information they need since the use of electronic information is growing every day. The internet maintains a significant quantity of data nowadays. People are turning to the web to obtain the information they need since the use of electronic information is growing every day. Because it is impossible for the user to read all of the data, text summarisation is used to summarize the data, which is then shown to the user so that the data may be simply understood.

Single-document and multi-document summarising systems are two types of automatic text summarisation systems. The first one creates a summary from a single document, while the second does it from a collection of documents. These are created using either an abstractive, extractive or hybrid method to summarize the text. The extractive technique generate the summary by selecting the most essential sentences from the input material. The abstractive technique transforms the text taken as input into an intermedial form before presenting a summary that includes phrases and words that vary from the original text sentences, whereas the hybrid approach combines both the approaches, that is extractive and abstractive. Section 2 elucidate the various classification for ATS. Figure 1 shows the architecture of an ATS system which includes the tasks shown below.

Figure 1: Automatic Text Summarisation System

1. Pre-Processing: Constructing an organized simulacrum of the original text by employing a variety of linguistic



approaches like stop word removal, stemming, sentence segmentation, part-of-speech tagging and tokenization and so on [1].

2. Processing: Converting an input document or text to the summary using one of the text summarisation ways, by using one or more techniques. Different types of approaches in Automatic Text Summarisation are delineated in Section 3.

3. Post-Processing: Before creating the final summary, various issues must be resolved in the generated summary sentences, such as reordering the selected sentences and anaphora resolution.

2. Classification

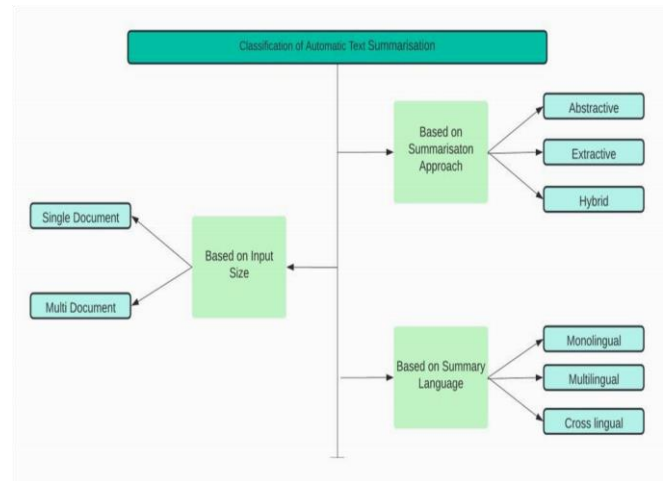


Figure 2: Automatic Text Summarisers Classification

2.1 On the basis of Input Size: The input size refers to the number of source documents used to create the target summary and it is further subdivided into two parts: 1) single document summarisation and 2) Multi document summarisation. As shown in Figure 1, SDS (Single Document Summarisation) takes single text document as input and generates a summary from it, with the goal of shortening input material while maintaining the key information. The purpose of Multi-Document Summarisation (MDS) is to decrease repeated information in the input documents by generating a summary based on a group of documents which are taken as an input. SDS is less challenging than MDS. MDS has some issues including repetition, secular relatedness, coverage, shrink ratio and so on [2, 3].

2.2 On the basis of Approach of Text Summarisation: Abstractive, extractive and hybrid are major three categories in which text summarisation is divided. The extractive text summarisation method chooses the crucial statements from the given input document provided by user and after that concatenates them into provided output summary. The document provided by the user are represented in an intermediary representation in the abstractive text summarisation technique, & the output is constructed from this. Whereas Abstractive summaries are made up of statements that are not same as the source document sentences. The extractive and abstractive processes are combined in the hybrid text summarisation methodology. Section 3 will go through these techniques in further depth.

2.3 On the basis of Summary Language: There are numerous sorts of text summarising techniques for different languages. So, a group of different languages are combined and collectively classified into major three types of categories: 1) Monolingual - When the source and destination papers are written in the same language, the summarising system is monolingual. 2) Multilingual - When the source information is written in many languages like English, French and Arabic then the summarising system is multilingual & the

summary of this is also produced in these languages. 3) Cross-Lingual - when the source content is written in one language like English and the summary is written in another like Arabic or French then the summarising system is cross lingual [4].

3. Approaches

There are 3 techniques to automatic text summarisation in general: abstractive, extractive and hybrid.

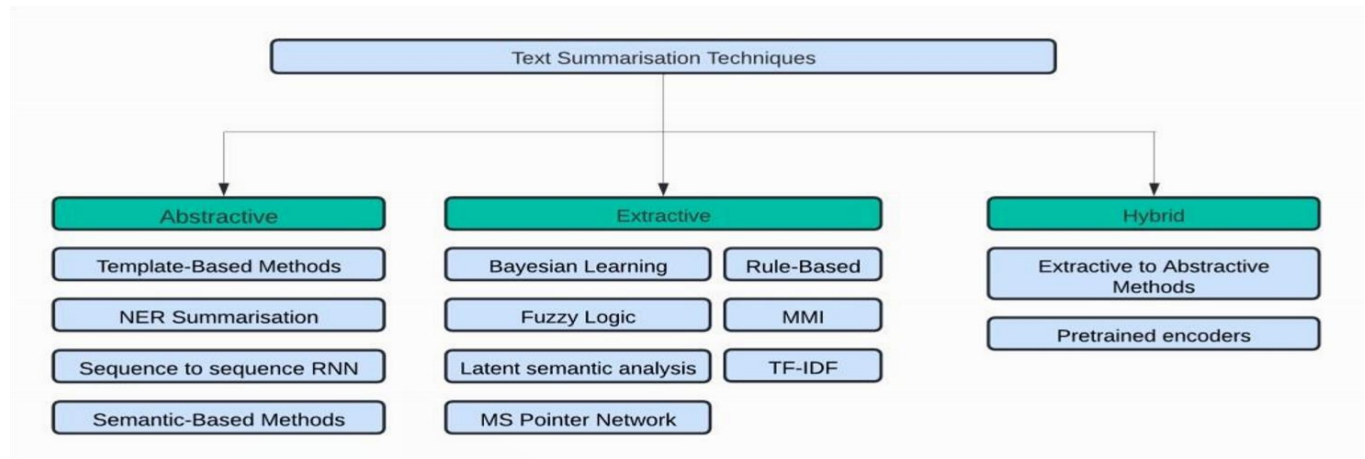


Figure 3: Techniques for automatic text summarising.

3.1 Abstractive

3.1.1 Proposal - It creates summaries that are produced by the new statements that were not there in the given input document (original copy). Abstractive text summarisation algorithms are complex in nature and complicated because they have to understand the input text, find the most relevant passages and generate syntactically correct sentences as summarisation. For hand-written regulations, such a process is practically difficult. However, recent advancement and research in AI/ML, particularly neural networks, have enabled abstractive summarisation to some extent. Furthermore, NN are the state-of-the-art in abstractive summarisation today [5].

Disadvantages: It is quite tough to provide a finest abstractive summary in practice. Abstractive summarisers which works well are difficult to create since they necessitate the usage of natural language generating technology, which is still in its growing domain. In order to produce new phrases, the abstractive technique requires a complete comprehension of the given text. The majority of abstractive summarisers generate repeated terms and with out-of-vocabulary words they are unable to handle adequately. The variety of abstract summarisers' representations limits their power. Systems can't summarise what their representations can't capture [7, 9].

3.1.2 Techniques and Methods

1. **Template-Based Methods:** Human summaries contain shared phrase forms that may be specified as templates in particular fields (e.g. meeting summaries). The abstractive summary may be generated on the basis of the given input text genre by using the information, to fill the slots in the input text in the appropriate predefined blueprint. The text samples that fill the template slots are determined using extraction rules and linguistic patterns [10, 11].

2. **NER Summarisation** – NER is an acronym for Named Entity Recognition. It is a type of method for recognizing and classifying atomic items in text into specific categories, such as people's names, organization names, places, concepts and so on. Text summarisation, question & answer, text classification and machine translation systems and in number of languages, NER has been used till date. A lot of work, advancement and research has been done in the field of NER for English, where capitalization provides a crucial indication for rules, however, Indian languages lack such qualities. This makes summarising the subject in Indian languages more difficult [12].

3. **Sequence to sequence RNN** - This concept enables sequences from a single domain to be changed into sequences from another domain. They began by describing the basic

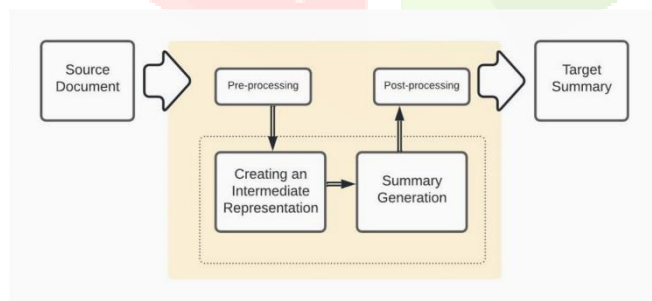


Figure 4: An abstractive text summarisation architecture

An abstractive text summariser's design is shown in Figure 4. It comprises of tasks pre-processing, processing tasks and post-processing such as 1) creating an internal semantic representation and 2) Creating a summary that would be similar-to human-generated summaries by applying natural language generation techniques [6].

Advantages: Based on paraphrasing, compression, or fusion it can employ more adjustable expressions, it provides better summaries using distinct terms that do not belong in the original text. The produced summary resembles the manual summary more closely. When opposed to extractive procedures, abstract methods can reduce the text even more [7-9].

encoder-decoder RNN, which serves as a baseline, before presenting a variety of novel summarisation models. Neural machine translation model is being depicted by this baseline model. The bidirectional GRU-RNN is being used by the encoder, whilst the unidirectional GRU-RNN is being used by the decoder with a encoder as the same hidden-state size and words are produced using attention to the tool over the source, for example: the hidden states and a soft-max layer gets more attention over the target vocabulary [13, 14].

The summarising problem, the huge vocabulary 'trick' (LVT), was also adjusted or added to this core model. This method's major goal is to lessen the size of data of the decoder's softmax layer, which is the main computational bottleneck. Furthermore, by limiting the modeling effort just on those words which are crucial with respect to a specific example, by following this type of strategy speeds up convergence.

Because a major part of the words in the summary originate from the original material, this approach is excellent for summarising [15].

4. Semantic-Based Methods – These are the ATS methods which use a semantic representation (like semantic graphs, predicate-argument structures, or information items) to build an abstractive summary from the input document(s), which is then fed into a natural language production engine developed an abstractive summariser for multi-document which 1) uses SRL to represent input documents with predicate-argument structures, 2) uses a semantic similarity measure to cluster semantically similar predicate-argument structures across the text, 3) the predicate-argument structures are ranked according to attributes that have been weighted and optimised using a Genetic Algorithm and 4) these predicate-argument structures are used to produce phrases via language generation [16, 17].

Table 1: Abstractive Text Summarisation Techniques - Advantages & Disadvantages.

Techniques	Advantages	Disadvantages
Template-Based Methods [16, 17]	Generates cohesive summaries and explanatory and the template slots may be filled by scraps gathered with the help of information extraction algorithms	Template slots need human creation of extraction rules and linguistic patterns. Lack in variation is there because of predefined templates.
NER Summarisation [12]	SpaCy library is the fastest and greatly fit for the practical applications and Flair library outperforms and also competent for experimentations.	Because of the uncertainty in the language, both the quality and constancy of the annotation are key challenges. Challenging on informal text.
Sequence to sequence RNN Summarisation [9]; [18, 19]	Suitable for the short sentences.	It needs a large volume of structured data for training. RNN-based Seq2Seq models take a long time to train and they can't capture distant dependence links for lengthy sequences.
Semantic- Based Methods [20]	Semantic Role Labelling (SRL) aids in determining the semantic link between sentences' words.	The characteristics of the semantic representation of the input text determines the quality of the summary which is constructed.

3.2 Extractive

3.2.1 Proposal - This algorithm takes bits and snippets of the input text, generally sentences, & combines them to create summary content. Most extractive summarisers follow the same two phases at a high level: First, give each sentence a score. Then choose the N phrases that get the greatest score. The way sentences are scored is the key distinction between individual extraction approaches [21].

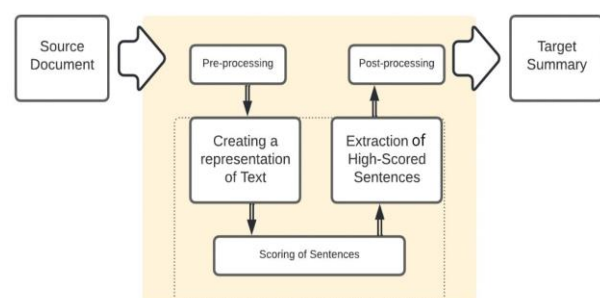


Figure 5: An extractive text summarisation architecture

An extractive text summariser's design is shown in Figure 5. It comprises of tasks as-

1) Pre-processing tasks. 2) post-processing tasks such as replacing relative temporal expression with real dates, reordering obtained paragraphs, substituting pronouns with their antecedents and so on[1]. 3) Processing tasks which includes:

- To make text analysis easier, create a proper representation of the incoming text. (For example bag of words (BOW), graphs, N-gram etc.) [2].
- Sentences scoring/ranking - Sentences are ranked depending on their representation in the input document or text [22].
- Withdrawing top-scored sentences: choosing and conjoining the essential statement from the text or input document to construct the summary [22, 23]. The length for created summary is determined by the preferred compression rate, which is limited by a length cut off or threshold that keeps the generated statement in the same sequence as the original text [9].

Advantages: The fundamental advantage of extractive approaches is that, no matter how basic the method is, it always provides syntactically accurate statements, even if they aren't helpful or grammatically perfect summaries.

Disadvantages: The extractive methodology is diametrically opposed to the way through which human specialists compose summaries. The produced extraction summary has the following flaws:

1. Some summary sentences have redundant information [7].
2. Sentences that have been extracted may be lengthier than usual [1].
3. As the extractive summaries are picked from numerous input documents, so in a multi-document system setting, temporal expressions creates conflict [1].
4. The retrieved summaries are limited with what the sentences from the actual text can predict. As a result, more detailed explanations could be out of their grasp.

3.2.2 Techniques and Methods

1. Bayesian Learning - SUMARIST, SWESUM and other automated text summary systems have been developed for the English language. However, single-syllable languages such as Vietnamese, Chinese, Japanese, Mongolian, Thai and other "native" languages of East Asia and Southeast. Many people speak single-syllable languages, which account for more than 60% of all languages spoken on the planet. As a result, processing a one-syllable language is critical. However, it is quite difficult to detect a word or phrase solely on white space and all word segmentation techniques presently do not achieve 100% accuracy. They primarily suggested a text summary approach based on the Naive Bayes algorithm and a subject word set in this research report [24, 25].

Naive Bayes categorization is used in two stages for single-syllable text: Two critical parts of the work are training and summary. It get trained using data and with the help of people to create a collection of extracted sentences in the Training phase.

2. Fuzzy logic – It is a typical model based on fuzzy logic for Automatic Text Summarisation and it takes eight features as the input for each and every sentence like (Length of sentence, Data in numerical form, Location of a sentence, Title word, Thematic words, Sentence to sentence similarity, Proper noun and Term weight) for its basic importance calculation. After extracting these eight features attributes values, it goes into a Fuzzy Inference System (FIS). Also, according to the indagation, a summary length of roughly 10% (approximately) of the real text length is appropriate and the resulting summary consists of phrases extracted in the original sequence [26].

3. Latent semantic analysis - It is a statistical-algebraic approach for detecting hidden semantic patterns in words and sentences. It's an unsupervised method that doesn't need any prior training or understanding. LSA gathers information from the context of the input material, like whether words are used collectively and whether similar ideas appear in many phrases. The presence of multiple similar phrases in the sentences indicates that they are semantically connected. Words' meanings are determined by the sentences wherein they occur and sentence meanings generally determined by word meanings. The mathematical approach of Singular Value Decomposition (SVD) is used to uncover the interconnections among phrases and SVD improves accuracy by predicting word-to-word correlations and minimising noise [27, 28].

Step 1: Forming an input matrix: An input document must be formatted as a matrix in which the sentences are represented as columns and the words/characters as rows. This way computer can easily comprehend and conduct computation on it.

Step 2: It is an algebraic method for modelling word and sentence relationships.

Step 3: Sentence selection: the key sentences are chosen using the SVD findings as well as various methods.

Following Sentence selection approaches have been used:

- LSA [29]
- SVD [28]
- Murray et al. (2005) [30]
- Cross method [31]
- Topic method [32]

4. MS Pointer Network – After a period of time, QianGu using the so-known Multi Source-Pointer technique is the next analysis received from the ML approach. This technique primarily focuses on assigning a rating to abstractive using deep learning by predicting the inaccuracy of words in the text as well as semantic inaccuracy. Basically, in this term, larger weights are assigned to words that are semantically related. The rogue is tested on the Gigaword and cable news network (CNN) datasets for this method's assessment. In compared to other ML techniques such as Sequence to sequence in addition to attention baseline, as well as Nallapati's abstractive model and the results performed quite well. The Gigaword dataset was used to test this model and it was superior to rouge-1 scoring 40.21 as shown to be, rouge-2 scoring 19.37 and rouge-L scoring 38.29. Another test is conducted using the CNN dataset, with rouge-1 scoring 39.93, rouge-2 scoring 18.21 and rouge-L scoring 37.39.

Other than all these methods, another one is loss of rouge-1 scoring measurements, which is basically contrasted with systems like baseline lead-3 given by Nallapati, with rouge-1 scoring 40.21 in the dataset of CNN. The major disadvantage of this type of model is the occurrence of recursion of the same statements in the document. By the virtue of this, it can be seen that this type of model is mainly kindred to the recursion/redundancy issue of the sentences. Qian Guo, a problem researcher, suggests adding TF-IDF or RBM to achieve a suitable or correct summaries which results in context of future study [13, 33].

5. Rule-Based – In the last ten years, this strategy has become much less prominent in the field of text summarisation. The approach's key benefit is that it can be used to a basic domain, making rule-based validation relatively straightforward. However, when utilised for a domain with a level of complexity very high, rule-based validation becomes quite difficult, so if the system is not able to identify the rules, then it cannot produce results. Aside from that, if there are more rules than are necessary, the system finds it challenging to sustain the output's performance [34].

6. Maximal marginal importance (MMI) – Current and recent ML technology studies include the Maximal Marginal Importance (MMI) approach, the PSO and a combination of other strategies such as fuzzy logic. Input is one type of document and the output is in extractive summary format. MMI produces summaries that sum up differently by determining the most unique sentences. Key sentences are chosen by taking the repetitive sentences there in the input and also by removing statements from the given input or from text-source. Techniques like PSO are used to select the least & most essential features and this fuzzy logic helps it to determine the values for the factors such as risk and ambiguity or the endurance rate can easily fluctuate. Output

was then tested and verified on database of Document Understanding Conference -2002 and then compared with different types of summaries like Sys-19, Sys-30 and MsWord summaries. Results performed more than expected with the comparison with the terms which are recall scoring 0.40 and f-measure scoring 0.422. MMI, PSO, Fuzzy are superior to different summaries like Sys-30 by the accuracy of 0.063. The main disadvantage of this method is the issue of semantic problems. This approach may be used by labelling the semantic roles in the lexical dataset and other for multi-document summarisers [35, 36].

7. TF-IDF Technique –TF-IDF approach is used in text summarising research such as [37-40]. This is from one of the algorithms that checks the link between a text and the entire collection of documents available. The major goal here is to compute the TF and IDF values [41]. Every phrase is treated as a separate document for a single input or a single type of document. The frequentness of recurrence of the word (T) in the entire single statement is used in this approach to determine how essential that word is in the input. IDF, on the other hand, is a numerical figure that represents the frequentness of the term (T) appears in a sentence. The numerical value or weightage of the word will be much more if it occurs many times in the document and also least in many other documents, one can find this by simply multiplying the TF value to the IDF value.

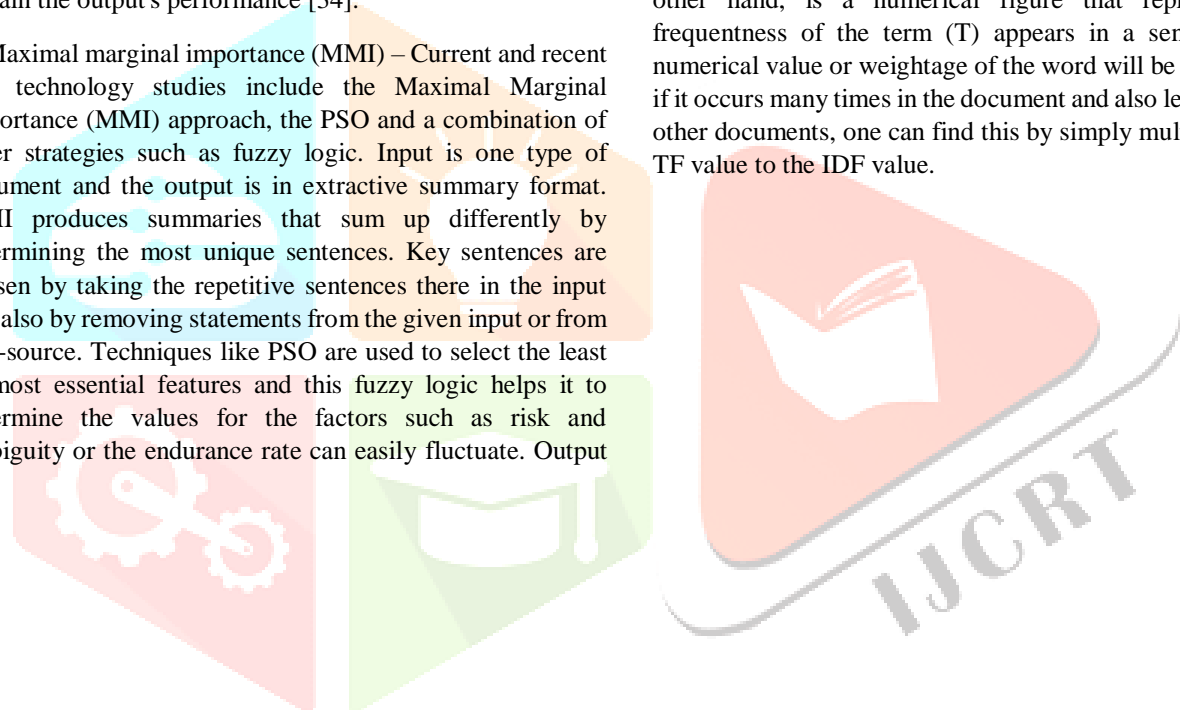


Table 2: Extractive Text Summarisation Techniques - Advantages & Disadvantages

Techniques	Advantages	Disadvantages
Bayesian Learning [24, 25]	Works efficiently for single syllable languages.	In Naive Bayes, all predictors are assumed to be independent, which is very rare case in real world. This greatly restricts the algorithm's usability in real-life scenarios.
Fuzzy logic [26]; [42, 43]	It tackles the uncertainties in the input	The duplication of the chosen statements in the summary is a negative element which might raise and negatively impact the quality of the summary.
Latent semantic analysis [4]	Creates lingual linked phrases.	SVD takes a long time to compute.
MS Pointer Network [33]	This strategy is used to give words that have semantic composites more weight.	The duplication of sentences in the summary is a flaw in this strategy.
Rule Based [34]	Simple to test and validate rule based.	If the system can't identify the rules, then no output is achieved. Due to too many rules the system's performance becomes harder to maintain.
Maximal marginal importance (MMI) [35, 36]	Create summaries with a lot of variety by focusing on the most -significant lines.	The semantic difficulty is the system's Achilles' heel.
TF-IDF Technique [37, 40]	Aids in extracting the most descriptive phrases from a document and quickly calculate the similarity of two papers.	It does not account for text location, semantics and co-occurrences across texts and so on.

3.3 Hybrid

3.3.1 Proposal - This method mainly focuses on the combination of abstractive and extractive approaches. Below Figure 6 mimics the hybrid text summariser's typical architecture. It usually includes the phases listed below:

1) Pre-processing phase. 2) Phrase extraction phase (extractive automated text summarisation) which takes out key sentences from the input document/text. 3) Create the final abstractive summary utilising abstractive approach on the collected phrases from the starting phase. 4) Post-Processing: In order to get assured that the sentences that is constructed are legitimate, certain basic rules must be devised such as: A sentence must be at least 3 words long according to sentence structure (subject + verb + object). A verb is must to appear in each and every sentence. An article (like "a", "an" and "the"), a conjunction (like "and"), a preposition (like "of") or an interrogative word (like "who") should not be used at the conclusion of the sentence [8]; [44, 45].

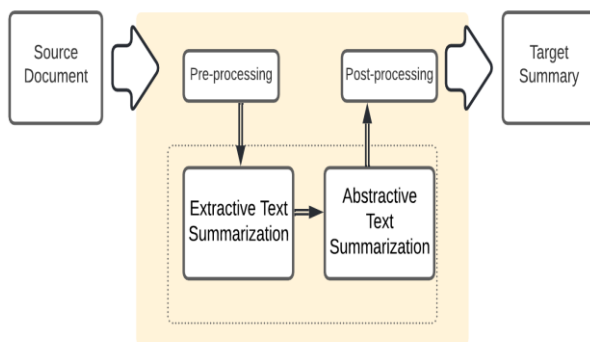


Figure 6: A hybrid text summarisation architecture

Advantages: It brings together the combined benefits of the both extractive and abstractive techniques. These two ways work together in hybrid increasing the performance of summarisation on a broad level [9].

Disadvantages: The produced summary generates a relatively low-grade abstractive summary in comparison to the pure abstractive approach since it is based on extracts (pieces of text) rather than the original text. Abstractive technique is challenging and also needs extensive use of NLP, so the researchers are engaging more on the extractive automatic text summarising strategy, which employs a variety of approaches and tactics to provide more reasoned and relevant summaries [4].

1.3.2 Techniques and methods

1. **Extractive to Abstractive Methods:** This technique goes by extracting sentences with the help of one of the extractive automatic text summarisation methods and after that applying the abstractive text summarisation techniques (one of them) to the recovered statements. Author, Wang et al. suggested the "EA-LTS" hybrid system for such challenge of summarising large texts in. The system is divided into 2 stages: 1) the extraction (cleaning/removal) phase, which applies a graph model to remove out key phrases and another

is 2) abstraction phase, which applies a pointer and attention approaches to create an encoder-decoder based on RNN and produce summaries [9].

2. **Pretrained encoder:** BERT which is an acronym for Bidirectional Encoder Representations from Transformers is a pre-trained language approach framework that has offered a quick overview of a broad range of NLP techniques and methodologies. For all types of the summarisation (extractive and abstractive), BERT can provide a full-fledged framework and architecture. It's a unique language representation model that trains differently through masked language modeling [46].

3. **Extractive** - After a neural encoder creates sentence representations, a classifier determines which statements should be used as summaries, rearranges them and adds the appropriate grammar. Different models like REFRESH (it is a learning-based system of reinforcement that has been taught by maximizing the ROUGE measure worldwide.), LATENT (Given a set of phrases, the probability of human summaries is maximized using this latent model), SUMO (it basically uses the structured attention to instigate or provide a representation of multiroot dependency tree of the material while anticipating the desired summary), NEUSUM (it is the most sophisticated extractive summarisation technique that scores and chooses sentences together) have been used for extractive summarisation [11, 23]; [46, 47].

Abstractive - In this the work is regarded/divided as a difficult sequence-to-sequence tasks. Different models like PTGEN (pointer generator network). It has a word copying feature that allows it to copy information from the original input, as well as a cover feature that maintains track of terms that have been summarised, DCA (Deep Communicating Agents) models are trained utilizing the reinforcement learning), For abstractive summarisation, DRM (deep reinforced model) is now being employed, which tackles the coverage problem by adopting an intra-attention strategy in which the decoder pays attention to previously produced words [10]; [48, 49].

Table 3: Hybrid Text Summarisation Techniques - Advantages & Disadvantages.

Techniques	Advantages	Disadvantages
Extractive to Abstractive Methods	Both approaches are used to increase the quality of the précis generated [50].	The extraction process used in the first step has considerable influence on the final outcome [51].
Pretrained encoders	It's the most effective summariser and outperforms RNN.	This strategy may not perform well when the input material is rather long and the compression ratio is quite low, since it may result in summaries that are devoid of context.

4. Text summarisation datasets

Dataset in Standard form– Authors have presented a conspectus of several corpora that have been utilized in the task of summarisation [52].

1. DUC: DUC is an acronym for Document Understanding Conference. These are the datasets, which are the most frequent and generally utilized in most text summarising analysis. There are three types of summaries in each dataset: the first one are the summaries which are manually created, second are the baseline summaries which are automatically generated and lastly the summaries supplied by challenge participants' systems, these are also automatically generated. Although these datasets are frequently used to evaluate Automatic Text Summarisation, they are insufficient for training neural network models [53].

2. SummBank Dataset: It includes 40 news clusters, human-authored non-extractive summaries, three hundred and sixty multi-document extracts, more than 2 million multi-document and single-document extracts which are made using the machine and manual methodologies [54].

3. Computer-Aided Summarisation Tool (CAST) Corpus: It includes a selection of newswire texts from the Reuters Corpus3 plus various science texts from the British National Corpus4. After signing the deal with Reuters5, the textual data of the new section of the corpus is obtainable, but the rest of it is not. There are three different sorts of information annotations are given in the corpus: sentence significance, sentence linkages and text fragments that may be extracted

from the marked statements. If a statements is not annotated, then it is considered as insignificant.

4. CNN-corpus: It can be used as information retrieval from a single document. The given source texts, word highlights and gold-standard summaries are all included. Not long ago this corpus was utilised in the competition called "DocEng'19" [55].

5. Gigaword 5: It's a well-known dataset for an abstractive summarisation studies. It has roughly 10 million articles of the English news, making it perfect for neural network training and testing. Gigaword has been chastised for summaries that simply provide the headlines [13, 53].

6. CNN/Daily Mail Corpus: It is an English-language dataset with little over 300,000 distinct news stories published by CNN and Daily Mail writers. It was first used for a passage-based question-and-answer task and afterwards it was widely used to test Automatic Text Summarisation.

To summarise, given the bulk of available data focus on the news domain, more datasets are required that 1) cover non-English languages & 2) include the diverse data domains for all language families.

The following characteristics can be seen defined for each dataset in Table 4: 1) name of the dataset, 2) language of the dataset, 3) domain of the dataset and 4) allows single-document summarisation or not 5) allows multi-document summarisation or not.

Table 4: Datasets Used for Text Summarisation.

Name of the Dataset	Language of the input document	Domain of the dataset	Allows Single Document summarisation	Allows Multi-Document summarisation
DUC 2002	English	News	✓	✓
Turkish Dataset	Turkey	News	✓	✗
XSum	English	News	✓	✗
CNN/Daily Mail news highlights dataset	English	News	✓	✗
Gigaword 5	English	News	✓	✗
SummBank	Chinese, English	News	✓	✓
CAST	English	News	✓	✗
CNN-corpus	English	News	✓	✗

5. Evaluation Metrics

Many attempts have been made to resolve the concern related to the summary evaluation during the last two decades. According to Huang et al., the objectives that needs to be addressed while creating a shortened and understandable summary:

1. Coverage of Information: the generated summary must include all of the key points from the given input material (s).
2. Importance of Information: the summary should include all of the subjects in the input material (s).The most essential topics can either be the major or central topics or user-preferred topics.
3. Redundancy of Information: can reduce the total amount of information available in the produced summary that is frequently redundant or duplicates.
4. Text Cohesion: The summary isn't merely a collection of essential but disparate sentences or words. The summary should be written in a way that is both legible and clear.

The resulting output summaries are evaluated using two different methods: The first one is intrinsic methods, human judgment is used to assess summary quality. The intrinsic assessment analyses a summary's consistency and content coverage, as well as its informativeness and the second is

extrinsic methods, it uses a task-based performance measure to assess or maintain the summary quality. The extrinsic evaluation mainly determines how useful the summaries are in certain application setting [1, 56].

To evaluate the text summarisation there are two ways: first one is manual and second is automatic. In the context of text summarising research, it is a very difficult problem to solve. To examine the quality of the Automatic Text Summarisers that created them, the automatically generated summaries must be assessed. The performance of the Automatic Text Summariser is frequently matched or compared to the other benchmarked systems, such as leading sentences from the input material or standard text summarisers like LexRank, TextRank and more [56-58].

5.1 Manual Summary Evaluation - Computer generated summaries are may be asked to assess by the human judges using the quality points listed below [56, 59]:

Readability: Evaluate the language quality of the summary generated by looking for extra spaces in its verbal structure or dangling anaphora.

Grammatical: The generated summary should not contain any improper statements or capitalization errors which conflicts the grammar norms.

Referential Clarity: The reader should be able identify the noun phrase as soon as it appears in the generated summary.

Coverage of content: The generated summary is must to encompass all of the subjects mentioned in the input material (s).

Structure and Coherence: The generated summary should be arranged properly and well-constructed. It is made up of a series of related and cohesive statements.

Non-redundancy: The generated summary should not contain any repetitions.

5.2 Automatic Summary Evaluation - Here we'll look at some of the most commonly used evaluation metrics in the literature [21].

1. Precision Score Metric: This is calculated by taking intersection of number of sentences in both the reference and candidate summaries and dividing it by the total number of sentences in the candidate summary as shown in Eq. 1.

$$\text{Recall} = \text{Sref} \cap \text{Scand} / \text{Scand} \quad (1)$$

2. Recall Score Metric: As shown in Equation 2, it is calculated by taking intersection of number of sentences in both the reference and candidate summaries and dividing it by the total number of sentences in the reference summary.

$$\text{Recall} = \text{Sref} \cap \text{Scand} / \text{Sref} \quad (2)$$

3. F-Measure Score Metric: As shown in Equation 3, F-measure is nothing but the harmonic mean of recall and precision. This is a measure that combines both the recall and precision metrics.

$$\text{F-Measure} = 2 (\text{Precision}) (\text{Recall}) / \text{Precision} + \text{Recall} \quad (3)$$

4. ROUGE Metric: It is the most trusted and widely used instrument in NLP for unmanned evaluation or assessment of the summaries, generated automatically. It basically counts the amount of overlapping units, between the candidate summaries and reference summaries. It has been shown to be useful in testing the accuracy of the model and assessing the quality of summaries and has a good correlation with human judgments [3, 9].

ROUGE evaluation has been regarded as a standard for assessing the generated summary and testing the accuracy of a summarising model since its inception, however it has the major drawback of just matching strings between the summaries without taking into account the meaning of series of words (n-grams) or single words.

The problem of human judgment is that it is subjective, with a broad range of what constitutes an "excellent" summary. This discrepancy suggests that developing an automated review and analysis system is complex and time-consuming. To decrease the expense of review, summaries created by Automatic Text Summariser are examined using automated metrics. The automated assessment measures, on the other hand, still require human effort since they rely on a testing of system-generated summaries comparing it with one or more human-created model summaries [21, 56].

6. Applications

6.1 CV or Resume Summarisation: CV summarisation will play a major role in extracting the CV document with only the required information like qualifications, marks, skills, experience, projects done and other useful information of the candidates.

6.2 News Summarisation: News blaster is basically a text summariser that assists readers in locating the most relevant news. In this system gathers, clusters, categorizes and summarises news automatically from many different sites on the daily basis

6.3 Summarisation of Scientific Papers: These publications are well-organized with a template-like structure and predictable positions of typical components in the content. To get citation information, mining the pattern of citations is one example of a way and relationships between citations, as well as summarisation approaches that recognize the material's content in both the citing and cited publications, can be employed.

6.4 Legal Documents Summarisation: To save legal professional's time, Kavila et al. presented a legal document search system that is automated. The summarising task highlights the rhetorical functions of presenting legal judgments document phrases. Based on the legal question, the search task discovers relevant historical cases. As a result, the hybrid system employs a variety of techniques, including keywords or key phrase matching applications or procedures, as well as the case-based strategy [60].

7. Challenges

7.1 Related to Text Summarisation Applications: Most of old or previous systems are focused on specific online reviews, text news and so on applications. Now is the time to concentrate on the most difficult applications, such as extended text, novel and book summaries.

7.2 Related to Multi-Document Summarisation: Redundancy, rearranging the sentences and co-reference are among the challenges that multi-document summarisation faces. Multi-document summarisation can result in improper references [4].

7.3 Related to Input Document's Length or Size: The majority of Automatic Text Summarisers are designed to handle short text documents. Existing ATS approaches may perform well when summarising small texts, but they perform poorly when summarising large texts.

7.4 Related to Languages that are supported: The majority of Automatic Text Summarisers focuses majorly on English language material. The quality of current Automatic Text Summarisers for many more languages needs to be enhanced [61].

7.5 Related to Text Summarisation Using Deep Learning: RNN in the sequence-to-sequence systems require a large-scale well organized trained data during the generating phase of summary. In actual NLP applications, the requisite training data is not always accessible. Building an Automatic Text Summariser with a very little quantity of training data by utilizing it with a classic NLP combination approach such as

syntax, grammatical, semantic analysis and so on is an interesting research issue.

8. Conclusion and future work

The main purpose of this paper is to provide the latest study, progress and research which is made in this field till date. We have discussed mainly the abstractive, extractive and hybrid summarisation techniques and their related advantages and disadvantages. Extractive summaries still hold the top of current popular trend topics in this research, even though they are far simpler than the most complicated abstractive summaries, which are quite complex. This is due to the fact that more study is required and that many questions remain unanswered in the abstractive summarisation process, which is a hurdle that researchers must overcome. It can also be shown that semantics, similarity, sentence position, sentence length, frequency, keywords and the necessity to be there are the most essential variables in making a good or clean summary. Also different datasets used for these

summarisation and the evaluation of the generated summary are the most important part of this study.

Future work in this field of textual summary research could include: i) solving problems related to feature, such as picking features to employ in data summarisation to discover the more appropriate features, uncovering new features, creating the most often utilized features, using a variety of semantic features, finding the best factors to produce coherent sentences and adding system elements. ii) Pre-processing the database problem with the right title; otherwise, POS Tagging is necessary to prevent word deletion and create tokens and this is done to distinguish word categories such as nouns, adjectives, verbs and so on. iii) Summing up the mathematical methodologies, ML and fuzzy-based is the most difficult to try in the extractive summaries. iv) We can enhance the current methodologies, such as NATSUM in some circumstances, or increase NATSUM performance by boosting compliance, by using abstractive summaries. v) Unusual datasets, such as legal papers, tourism attractions summaries and inspection documents summaries [62].

REFERENCES

- [1] V. Gupta and G. S. Lehal, "A Survey of Text Summarisation Extractive Techniques," *J. Emerg. Technol. Web Intell.*, vol. 2, no. 3, pp. 258–268, Aug. 2010, doi:10.4304/jetwi.2.3.258-268.
- [2] M. Joshi, H. Wang and S. McClean, "Dense Semantic Graph and Its Application in Single Document Summarisation," in *Emerging Ideas on Information Filtering and Retrieval*, vol. 746, C. Lai, A. Giuliani and G. Semeraro, Eds. Cham: Springer International Publishing, 2018, pp. 55–67. doi: 10.1007/978-3-319-68392-8_4.
- [3] V. K. Gupta and T. J. Siddiqui, "Multi-document summarisation using sentence clustering," in *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, Kharagpur, India, Dec. 2012, pp. 1–5. doi: 10.1109/IHCI.2012.6481826.
- [4] M. Gambhir and V. Gupta, "Recent automatic text summarisation techniques: a survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, Jan. 2017, doi: 10.1007/s10462-01694759.
- [5] N. Moratanch and S. Chitrakala, "A survey on abstractive text summarisation," in *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, Nagercoil, India, Mar. 2016, pp. 1–7. doi: 10.1109/ICCPCT.2016.7530193.
- [6] S. Chitrakala, N. Moratanch, B. Ramya, C. G. Revanth Raaj and B. Divya, "ConceptBased Extractive Text Summarisation Using Graph Modelling and Weighted Iterative Ranking," in *Emerging Research in Computing, Information, Communication and Applications*, N. R. Shetty, L. M. Patnaik, N. H. Prasad and N. Nalini, Eds. Singapore: Springer Singapore, 2018, pp. 149–160. doi: 10.1007/978-981-10-4741-1_14.
- [7] L. Hou, P. Hu and C. Bei, "Abstractive Document Summarisation via Neural Model with Joint Attention," in *Natural Language Processing and Chinese Computing*, vol. 10619, X. Huang, J. Jiang, D. Zhao, Y. Feng and Y. Hong, Eds. Cham: Springer International Publishing, 2018, pp. 329–338. doi: 10.1007/978-3-319-73618-1_28.
- [8] R. Sun, Z. Wang, Y. Ren and D. Ji, "Query-Biased Multi-document Abstractive Summarisation via Submodular Maximization Using Event Guidance," in *Web-Age Information Management*, vol. 9658, B. Cui, N. Zhang, J. Xu, X. Lian and D. Liu, Eds. Cham: Springer International Publishing, 2016, pp. 310–322. doi: 10.1007/978-3-319399379_24.
- [9] S. Wang, X. Zhao, B. Li, B. Ge and D. Tang, "Integrating Extractive and Abstractive Models for Long Text Summarisation," in *2017 IEEE International Congress on Big Data (Big Data Congress)*, Honolulu, HI, USA, Jun. 2017, pp. 305–312. doi: 10.1109/BigDataCongress.2017.46.
- [10] R. Paulus, C. Xiong and R. Socher, "A Deep Reinforced Model for Abstractive Summarisation," *ArXiv170504304 Cs*, Nov. 2017, Accessed: Mar. 05, 2022. [Online]. Available: <http://arxiv.org/abs/1705.04304>
- [11] S. Narayan, S. B. Cohen and M. Lapata, "Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarisation," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018, pp. 1797–1807. doi: 10.18653/v1/D18-1206.
- [12] P. Marek, Š. Müller, J. Konrád, P. Lorenc, J. Pichl and J. Šedivý, "Text Summarisation of Czech News Articles Using Named Entities," 2021, doi: 10.48550/ARXIV.2104.10454.
- [13] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre and B. Xiang, "Abstractive Text Summarisation using Sequence-to-sequence RNNs and Beyond," in *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, Berlin, Germany, 2016, pp. 280–290. doi: 10.18653/v1/K16-1028.
- [14] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks

- on Sequence Modeling,” ArXiv14123555 Cs, Dec. 2014, Accessed: Mar. 08, 2022. [Online]. Available: <http://arxiv.org/abs/1412.3555>
- [15] I. Gibadullin and A. Valeev, “Experiments with LVT and FRE for Transformer model,” ArXiv200412495 Cs, Apr. 2020, Accessed: Mar. 05, 2022. [Online]. Available: <http://arxiv.org/abs/2004.12495>
- [16] V. Gupta, N. Bansal and A. Sharma, “Text Summarisation for Big Data: A Comprehensive Survey,” in International Conference on Innovative Computing and Communications, vol. 56, S. Bhattacharyya, A. E. Hassanien, D. Gupta, A. Khanna and I. Pan, Eds. Singapore: Springer Singapore, 2019, pp. 503–516. doi: 10.1007/978-981-132354-6_51.
- [17] A. Khan, N. Salim and H. Farman, “Clustered genetic semantic graph approach for multi-document abstractive summarisation,” in 2016 International Conference on Intelligent Systems Engineering (ICISE), Islamabad, Pakistan, Jan. 2016, pp. 63–70. doi: 10.1109/INTELSE.2016.7475163.
- [18] T. Cai, M. Shen, H. Peng, L. Jiang and Q. Dai, “Improving Transformer with Sequential Context Representations for Abstractive Text Summarisation,” in Natural Language Processing and Chinese Computing, vol. 11838, J. Tang, M.-Y. Kan, D. Zhao, S. Li and H. Zan, Eds. Cham: Springer International Publishing, 2019, pp. 512–524. doi: 10.1007/9783030-32233-5_40.
- [19] W. Miao, G. Zhang, Y. Bai and D. Cai, “Improving Accuracy of Key Information Acquisition for Social Media Text Summarisation,” in 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS), Shenyang, China, Oct. 2019, pp. 408–415. doi: 10.1109/IUCC/DSCI/SmartCNS.2019.00094.
- [20] N. S. Ranjitha and J. S. Kallimani, “Abstractive multi-document summarisation,” in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, Sep. 2017, pp. 1690–1694. doi: 10.1109/ICACCI.2017.8126086.
- [21] N. Moratanch and S. Chitrakala, “A survey on extractive text summarisation,” in 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India, Jan. 2017, pp. 1–6. doi: 10.1109/ICCCSP.2017.7944061.
- [22] A. Nenkova and K. McKeown, “A Survey of Text Summarisation Techniques,” in Mining Text Data, C. C. Aggarwal and C. Zhai, Eds. Boston, MA: Springer US, 2012, pp. 43–76. doi: 10.1007/978-1-4614-3223-4_3.
- [23] J. Zhu, L. Zhou, H. Li, J. Zhang, Y. Zhou and C. Zong, “Augmenting Neural Sentence Summarisation Through Extractive Summarisation,” in Natural Language Processing and Chinese Computing, vol. 10619, X. Huang, J. Jiang, D. Zhao, Y. Feng and Y. Hong, Eds. Cham: Springer International Publishing, 2018, pp. 16–28. doi: 10.1007/978-3-319736181_2.
- [24] P. K. P. Mok, “Language-specific realizations of syllable structure and vowel-to-vowel coarticulation,” J. Acoust. Soc. Am., vol. 128, no. 3, p. 1346, 2010, doi: 10.1121/1.3466859.
- [25] T. Nomoto, “Bayesian learning in text summarisation,” in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05, Vancouver, British Columbia, Canada, 2005, pp. 249–256. doi: 10.3115/1220575.1220607.
- [26] A. Kumar and A. Sharma, “Systematic literature review of fuzzy logic based text summarisation,” Iran. J. Fuzzy Syst., vol. 16, no. 5, Oct. 2019, doi: 10.22111/ijfs.2019.4906.
- [27] M. G. Ozsoy, F. N. Alpaslan and I. Cicekli, “Text summarisation using Latent Semantic Analysis,” J. Inf. Sci., vol. 37, no. 4, pp. 405–417, Aug. 2011, doi: 10.1177/0165551511408848.
- [28] J. Steinberger and K. Ježek, “Text Summarisation and Singular Value Decomposition,” in Advances in Information Systems, vol. 3261, T. Yakhno, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 245–254. doi: 10.1007/978-3-540-30198-1_25.
- [29] Y. Gong and X. Liu, “Generic text summarisation using relevance measure and latent semantic analysis,” in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '01, New Orleans, Louisiana, United States, 2001, pp. 19–25. doi: 10.1145/383952.383955.
- [30] G. Murray, S. Renals and J. Carletta, “Extractive summarisation of meeting recordings,” in Interspeech 2005, Sep. 2005, pp. 593–596. doi: 10.21437/Interspeech.2005-59.
- [31] E. Linhares Pontes, S. Huet, J.-M. Torres-Moreno and A. C. Linhares, “Cross-Language Text Summarisation Using Sentence and Multi-Sentence Compression,” in Natural Language Processing and Information Systems, vol. 10859, M. Silberstein, F. Atigui, E. Kornysheva, E. Métais and F. Mezziane, Eds. Cham: Springer International Publishing, 2018, pp. 467–479. doi: 10.1007/978-3-319-91947-8_48.
- [32] N. K. Nagwani, “Summarising large text collection using topic modeling and clustering based on MapReduce framework,” J. Big Data, vol. 2, no. 1, p. 6, Dec. 2015, doi: 10.1186/s40537-015-0020-5.
- [33] Q. Guo, J. Huang, N. Xiong and P. Wang, “MS-Pointer Network: Abstractive Text Summary Based on Multi-Head Self-Attention,” IEEE Access, vol. 7, pp. 138603–138613, 2019, doi: 10.1109/ACCESS.2019.2941964.
- [34] M. A. P. Subali and C. Fatichah, “Kombinasi Metode Rule-Based dan N-Gram Stemming untuk Mengenali Stemmer Bahasa Bali,” J. Teknol. Inf. Dan Ilmu Komput., vol. 6, no. 2, p. 219, Feb. 2019, doi: 10.25126/jtiik.2019621105.
- [35] F. B. Goularte, S. M. Nassar, R. Fileto and H. Saggion, “A text summarisation method based on fuzzy rules and applicable to automated assessment,” Expert Syst. Appl., vol. 115, pp. 264–275, Jan. 2019, doi: 10.1016/j.eswa.2018.07.047.
- [36] O.-M. Foong and A. Oxley, “A hybrid PSO model in Extractive Text Summariser,” in 2011 IEEE Symposium on

- Computers & Informatics, Kuala Lumpur, Malaysia, Mar. 2011, pp. 130–134. doi: 10.1109/ISCI.2011.5958897.
- [37] School of Software, Xinjiang University, Urumqi 830008, China, R. Khan, Y. Qian and S. Naeem, “Extractive based Text Summarisation Using KMeans and TF-IDF,” *Int. J. Inf. Eng. Electron. Bus.*, vol. 11, no. 3, pp. 33–44, May 2019, doi: 10.5815/ijieeb.2019.03.05.
- [38] A. M. Azmi and N. I. Altmami, “An abstractive Arabic text summariser with user controlled granularity,” *Inf. Process. Manag.*, vol. 54, no. 6, pp. 903–921, Nov. 2018, doi: 10.1016/j.ipm.2018.06.002.
- [39] M. Yousefi-Azar and L. Hamey, “Text summarisation using unsupervised deep learning,” *Expert Syst. Appl.*, vol. 68pp. 93–105, Feb. 2017, doi: 10.1016/j.eswa.2016.10.017.
- [40] A. Alzuhair and M. Al-Dhelaan, “An Approach for Combining Multiple Weighting Schemes and Ranking Methods in Graph-Based Multi-Document Summarisation,” *IEEE Access*, vol. 7, pp. 120375–120386, 2019, doi: 10.1109/ACCESS.2019.2936832.
- [41] S. Robertson, “Understanding inverse document frequency: on theoretical arguments for IDF,” *J. Doc.*, vol. 60, no. 5, pp. 503–520, Oct. 2004, doi: 10.1108/00220410410560582.
- [42] Nasrin nazari and M. A. Mahdavi, “A survey on Automatic Text Summarisation,” *J. AI Data Min.*, no. Online First, May 2018, doi: 10.22044/jadm.2018.6139.1726.
- [43] D. Patel, S. Shah and H. Chhinkaniwala, “Fuzzy logic based multi document summarisation with improved sentence scoring and redundancy removal technique,” *Expert Syst. Appl.*, vol. 134, pp. 167–177, Nov. 2019, doi: 10.1016/j.eswa.2019.05.045.
- [44] I. K. Bhat, M. Mohd and R. Hashmy, “SumItUp: A Hybrid Single-Document Text Summariser,” in *Soft Computing: Theories and Applications*, vol. 583, M. Pant, K. Ray, T. K. Sharma, S. Rawat and A. Bandyopadhyay, Eds. Singapore: Springer Singapore, 2018, pp. 619–634. doi: 10.1007/978-981-10-5687-1_56.
- [45] E. Lloret, M. T. Romá-Ferri and M. Palomar, “COMPENDIUM: A text summarisation system for generating abstracts of research papers,” *Data Knowl. Eng.*, vol. 88, pp. 164–175, Nov. 2013, doi: 10.1016/j.datak.2013.08.005.
- [46] Y. Liu, I. Titov and M. Lapata, “Single Document Summarisation as Tree Induction,” in *Proceedings of the 2019 Conference of the North, Minneapolis, Minnesota*, 2019, pp. 1745–1755. doi: 10.18653/v1/N19-1173.
- [47] X. Zhang, M. Lapata, F. Wei and M. Zhou, “Neural Latent Extractive Document Summarisation,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018, pp. 779–784. doi: 10.18653/v1/D18-1088.
- [48] A. See, P. J. Liu and C. D. Manning, “Get To the Point: Summarisation with Pointer- Generator Networks,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, Canada, 2017, pp. 1073–1083. doi: 10.18653/v1/P17-1099.
- [49] A. Celikyilmaz, A. Bosselut, X. He and Y. Choi, “Deep Communicating Agents for Abstractive Summarisation,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana, 2018, pp. 1662–1675. doi: 10.18653/v1/N18-1150.
- [50] R. Sahba, N. Ebadi, M. Jamshidi and P. Rad, “Automatic Text Summarisation Using Customizable Fuzzy Features and Attention on the Context and Vocabulary,” in *2018 World Automation Congress (WAC)*, Stevenson, WA, Jun. 2018, pp. 1–5. doi: 10.23919/WAC.2018.8430483.
- [51] H. Li, J. Zhu, C. Ma, J. Zhang and C. Zong, “Read, Watch, Listen and Summarize: MultiModal Summarisation for Asynchronous Text, Image, Audio and Video,” *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 996–1009, May 2019, doi: 10.1109/TKDE.2018.2848260.
- [52] F. F. dos Santos, M. A. Domingues, C. V. Sundermann, V. O. de Carvalho, M. F. Moura and S. O. Rezende, “Latent association rule cluster based model to extract topics for classification and recommendation applications,” *Expert Syst. Appl.*, vol. 112, pp. 34–60, Dec. 2018, doi: 10.1016/j.eswa.2018.06.021.
- [53] H. Lin and V. Ng, “Abstractive Summarisation: A Survey of the State of the Art,” *Proc. AAAI Conf. Artif. Intell.* vol. 33, pp. 9815–9822, Jul. 2019, doi: 10.1609/aaai.v33i01.33019815.
- [54] D. R. Radev et al., “Evaluation challenges in large-scale document summarisation,” in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - ACL '03*, Sapporo, Japan, 2003, vol. 1, pp. 375–382. doi: 10.3115/1075096.1075144.
- [55] R. D. Lins et al., “The CNN-Corpus: A Large Textual Corpus for Single-Document Extractive Summarisation,” in *Proceedings of the ACM Symposium on Document Engineering 2019*, Berlin Germany, Sep. 2019, pp. 1–10. doi: 10.1145/3342558.3345388.
- [56] E. Lloret, L. Plaza and A. Aker, “The challenging task of summary evaluation: an overview,” *Lang. Resour. Eval.*, vol. 52, no. 1, pp. 101–148, Mar. 2018, doi: 10.1007/s10579017-9399-2.
- [57] D. R. Radev, H. Jing, M. Styś and D. Tam, “Centroid-based summarisation of multiple documents,” *Inf. Process. Manag.*, vol. 40, no. 6, pp. 919–938, Nov. 2004, doi: 10.1016/j.ipm.2003.10.006.
- [58] R. Mihalcea, “Graph-based ranking algorithms for sentence extraction, applied to text summarisation,” in *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions -*, Barcelona, Spain, 2004, pp. 20-es. doi: 10.3115/1219044.1219064.
- [59] I. Mani, *Automatic Summarisation*, vol. 3. Amsterdam: John Benjamins Publishing Company, 2001. doi: 10.1075/nlp.3.

[60] S. D. Kavila, V. Puli, G. S. V. Prasada Raju and R. Bandaru, "An Automatic Legal Document Summarisation and Search Using Hybrid System," in Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), vol. 199, S. C. Satapathy, S. K. Udgata and B. N. Biswal, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 229–236. doi: 10.1007/978-3-642-35314-7_27.

[61] R. Belkebir and A. Guessoum, "Concept generalization and fusion for abstractive sentence generation," Expert Syst. Appl., vol. 53, pp. 43–56, Jul. 2016, doi: 10.1016/j.eswa.2016.01.007.

[62] C. Barros, E. Lloret, E. Saquete and B. Navarro-Colorado, "NATSUM: Narrative abstractive summarisation through cross-document timeline generation," Inf. Process. Manag., vol. 56, no. 5, pp. 1775–1793, Sep. 2019, doi: 10.1016/j.ipm.2019.02.010.

