FOM Hochschule für Oekonomie & Management

University Centre Essen

Group Project

in the study course Applied Project II

on the Topic

Global Market Analysis of Software Tools/ Suppliers

by

Agnishwar Das

Gourav Chandra

Sankha Chakraborty

Tamal Chakraborty

Professor:	Prof. Dr. Dirk Stein
Matriculation number:	547965 550465 550491 557644
Delivery date:	2021-08-22

Table of Contents

List of Figures

Table Directory

1. Introduction

1.1 Importance of the topic

Books can be represented as one of the oldest forms of written communication which are being used since thousands. In the past few decades, large fraction of the electronic documents are increasing exponentially in form of web pages, news articles, scientific reports, electronic books, and others. As a result, data is growing rapidly in every domain such as news, social media, banking, health, education, etc. Consequently, it has become increasing difficult and time consuming for researchers and professionals to get an abstract idea of larger documents. As a result, , there is a need of automatic summarizer due to the excessiveness of data in form of books and documents. Automatic summarizer is capable to summarize such vast information of the original document, especially textual data without losing any critical purposes.

Automatic summarization is the process of shortening a set of data computationally, to create a subset, that represents the most important or relevant information within the original content. In addition to text, images and videos can also be summarized. Text summarization finds the most informative sentences in a document. Some of the advantages of Automated Summarization are –

- Summaries reduce reading time.
- When researching documents, summaries make the selection process easier.
- Automatic summarization improves the effectiveness of indexing.
- Automatic summarization algorithms are less biased than human summarizers.
- Personalized summaries are useful in question-answering systems as they provide personalized information.

In this paper, we focus to illustrate a transparent global market analysis on several software tools and services for automated summary provided by different companies across the globe.

1.2 Problem Statement

There is no transparent market overview, and it is challenging to find enough information on the software tools/ suppliers who provide automated book summary.

1.3 Research Question

1. How to achieve global market analysis for automated book summaries ?
2. Who are the leading market providers for automated book summaries and what are the differences among the features of their software or services ?

1.4 Planned Output

2 To identify the global software/ tools suppliers.

3 To make a comparison among the services of different market providers.

4 To achieve a brief and transparent market analysis on automated book summary.

2. Theoretical Foundation

2.1 Significance of Automated Book Summaries

Textual information as a digital document quickly gathers the massive amounts of data. Most of this huge amount of data is unstructured, unhindered and has not been organized into traditional databases. Documents processing is a brief task because of the lack of standards. So, it has become very difficult to implement automatic text summarization by compressing the original text. Summaries are the rightest way of decreasing the length of an oversized document. In books or any large documents, abstracts and table of content are the ways of representing concentrated form of the document. Automated book summarization is the process by which computer program creates a shortened and accurate version of text.[1] The literature provides many definitions of automated book summary. One of the definition states that the summarization of a document is reduced, accurate and representation of the text that looks for to supply the exact idea of its contents. The main objective is to give information about and provide special access to the source documents. The product of this process contains the most importance and significant points from the original text. Search engine like Google use this automated book summarization to produce key phrase extractions in search results. Summarization is automatic when it is generated by software or an algorithm. Information which has been rejected during the summarization process is not considered relevant. Determining the relevance of the information included in documents is one of the main challenges of automated book summarization. The main reasons why we use the automated book summary are[2] –

1. Summaries reduce the length of the texts
2. Automated book summary improves the effectiveness of indexing
3. Summaries reduce reading time
4. Summaries make the selection process easier while choosing the researching documents.
5. Summarization improves the adequacy of ordering.

[1] Cp. D. Das and A. F. Martins, "A survey on automatic text summarization", 2007, pp. 192-195

[2] Cp. Tas, Oguzhan, Farzad Kiyani. "A survey automatic text summarization", 2007, pp. 205-213.

6. Summarization calculations are less one-sided than human summarizers.
7. Personalized summaries are useful in question-answering systems because they provide the personal information.
8. Using the automated book summary system enables commercial abstract services to increase the number of texts they are able to process.

An automated book summarizer must overcome some challenges which are[3] –

1. Calculating which sentences are the most important.
2. Making the summary readable, cohesive, and more accurate.
3. Minimize the number of references.

2.2 Nature and Type of Automated Summarization

In the late 1950 and early 60 there were there were few studies and experimentations which indicated that automated text summarization was feasible on computer. In early days, the methods which were developed mostly relied on surface level ideas, such as sentence placement and word frequency counts, and focused mostly on extracting the information rather than abstract formation. A few decades have passed since then and the increasing number of books and emerging presence of online texts in corporate sectors, especially in web formats have revived the interest of scientists in automated text summarization.[4]

With the advancements in Natural Language Processing (NLP) and enormous development in computer memory and speed in the past few decades have helped the researchers to develop more sophisticated and precise techniques with encouraging results. The United States had invested relatively small sums of money in research in the late 1990s over the course of a few years including commercial efforts at Microsoft, Lexis-Nexis, Oracle, SRA, and TextWise and efforts at Carnegie Mellon University, New Mexico State University, The University of Pennsylvania, and The University of Southern California.[5] As a result, number of viable products were produced along with a number of innovations that promise further improvements. There have also been a number of workshops, a collection of books, and several tutorials that attest to the growing research interest in automating the process of text summarization.[6]

[3] Cp. Mahak, Gupta. "Automatic text summarization techniques", 2017, pp. 1-66.
[4] Cp. H. P. Luhn, "The Automatic Creation of Literature Abstracts," 1958, pp. 159-165.
[5] Cp. Gambhir, Mahak, and Vishal Gupta. "Recent automatic text summarization techniques", 2017, pp. 1-66.
[6] Cp. Aone, C., M.E. Okurowski, J. Gorlinsky, B. Larsen. "A Scalable Summarization System using Robust NLP.", 1998, no page number

The fact remains, however, if the various systems are examined and their achievements are considered, an immanent similarity can be observed by means of their narrowness of focus, and large number of unknown factors surrounding the problem. For example, what does a summary consist of and what precisely is the nature of a summary ? Although, there are numerous definitions and opinions around this field of study but there is no general consensus to it. After analyzing comparative literature study and document analysis, it can be defined that – A summary is a text that is derived from one or more texts and includes some of the same material from the original texts while being no longer than half the total length of the original. To have a transparent and concise picture, we have followed and extended previous research works by identifying the following aspects where at least three major classes of characteristics can be attributed to any summary –

1. **Input : Characteristics of the Source**

- *Source size : single-document vs. multi-document* – Single-document summaries are based on a single source of input. Although information obtained from other texts may be utilized in the summarization process itself. A multi-document summary is a single text that accounts for the content of multiple input texts, and it is usually used only when those texts are relevant.
- *Specificity : domain-specific vs. general* – If the input texts all refer to a single domain, it may be useful to use area-specific summary methods, to focus on specific contents and formats. A domain-specific summary is derived from input texts whose subject areas concerns to a particular domain and several techniques such as uncertainty of terms, grammar usage, specialized formatting may be applied for the process. On the other side A general type of summary can be derived from input texts in any domain and requires no such premises.
- *Genre* – Newspaper articles and editorials, long and short stories, non-fiction books, business reports are among the typical input genres.
- *Scale* – The scale of the summary may vary from books to paragraphs depending on the length.

2. **Output : Characteristics of the Summary**

- Derivation : Extract vs. abstract – An extract consists of single words or whole paragraphs. Despite the fact that content is extracted from original document and includes key phrases

or important sentences, the extracted content remains intact. An abstract is essentially a newly generated text generated by analyzing the input texts through some type of computer programs. In abstract methods, a semantic representation of the original content is built and used as the basis for a summary that goes closer to what the mind would express. Although such transformation requires modern techniques like natural language processing which is computationally much more complex than extraction, and precise knowledge of the original domain is also needed in such case.

- Coherence: Fluent vs. Disfluent – Fluent summaries contain complete, grammatical sentences with relevant details and dependencies based on systematic structure. Unlike a coherent summary, a disfluent one is fragmented, consisting of words or parts of sentences that cannot be put together into meaningful sentences or paragraphs.
- Partiality: Neutral vs. Evaluative – When the input material is influenced by opinions, this characteristic pertains specifically. Neutral summaries, no matter how partial or impartial, summarize the content of input texts. Performing an evaluative summary requires some bias on the part of the system, whether explicitly through statements of opinion or implicitly through the inclusion of one biases and omission of others.
- Conventionality: Fixed vs. Floating – Fixed situation summaries are created for specific uses, audiences, and situations. Therefore, it can be formatted and highlighted according to appropriate conventions. Floating summaries can be designed and demonstrated in different settings and for different audiences for various purposes, so that it cannot be assumed to have fixed conventions.

3. Purpose: Characteristics of the Summary Usage Audience

- Orientation: Generic vs. query-oriented – A generic summary presents the author's perspective on the input texts, giving equal weight to all major themes in it. In a query-oriented or user-oriented summary, specific themes or aspects of the text are highlighted in response to a user's desire to learn more about just those topics. This may be done explicitly by highlighting pertinent themes, or implicitly by leaving out subjects that don't align with the user's interests.
- Usage: Indicative vs. Informative – The indicative summary describes only the main subject matter of the input texts without describing their contents. An informative summary can provide insight into the topic of the input text, but not necessarily the information

contained inside. Informative summaries reflect and describe parts of the text and can serve as a summary of the main points.

- Expansiveness – Readers are assumed to have prior knowledge in general for background summaries. In a just-the-news summary, the main points are just the new or principal developments, with the reader assuming enough background to interpret them accordingly.

2.3 Statistics in Automated Summarization

2.3.1 History of Statistical Approach

- In 1958, H. P. Luhn began seriously investigating automatic summarization based on statistics. Based on the relative significance of word occurrences in a scientific article, he marked entire sentences for extraction into an abstract using an IBM 704 data processor.[7] His algorithm helped to remove stop words, grouped words which were similar , and calculated frequency of word usage. In order to segment sentences, significant words and their distance from other significant words were assessed. In terms of total segments, frequencies were normalized by total segments. A routine that Luhn devised is the precursor of stemming. Using letter-to-letter correspondence, similar words were compared and if six or more letters did not coincide, they were considered distinct terms. The occurrence of error was about 5% in his method.

- In the late 1960s, Edmundson developed the idea of extracting summary information from text using text features.[8] Among those features there were – title words that appear in subtitles and titles; the frequency of keyword occurrences in the document, placement of sentences in first and last paragraph, paragraphs under subheadings, different categories of cue words, summary words, additional positive words, words associated with negative stigma, and null words that are irrelevant. Although, he didn't include stop words. According to Edmundson, the total weight of the sentence denoted by the sum of the individual word weights –

  Total Weight = Titles + Keywords + Locations + Cues.

- Pollock and Zamora followed Edmundson in the mid-1970s by identifying summary sentences from a corpus of narrow subjects at the Chemical Abstracts Service (CAS) primarily by cue words and phrases. Weights were adjusted so as to decrease the effects of

[7] Cp. H. P. Luhn, "The Automatic Creation of Literature Abstracts, 1958, pp. 15-21.
[8] Cp. H. P. Edmundson, "New Methods in Automatic Extracting,", 1969, pp. 264-285.

too many positive or negative terms on summary length. Additionally, the tests demonstrated the power of domain knowledge by referencing a controlled vocabulary of standard abbreviations and terminologies used in chemistry, both during cue phrase analysis and when creating summaries.[9]

- Even as automatic indexing was being studied extensively during the late 1970s and 1980s, research into text summarization slowed precipitously. However, statistical techniques based on IR generated remarkable interest due to their ease of implementation, to use Moens' term.[10] Additional to the proximity analysis inspired by Edmundson, other research areas currently include statistical analysis of clustering terms, structure of the text, discourse, and also training algorithms that can analyze abstracts generated by human to predict when a text summary is automated, how likely are certain source text statements to appear. These approaches reflect different viewpoints as we draw closer to understanding full-text. These research initiatives illustrate key IR techniques and approaches in text extraction.

- In 1980, Salton and his colleagues at Cornell University researched term weighting and automatic indexing and were pioneers in automatic summarization. An array of term-and segment-normalization techniques were used in Salton's experiment (1992, 1994, 1997) to predict closely related segments within a document and compare them with other documents, with the automatic linking of links generated when the similarities were great.[11] Furthermore, the analysis of similarities between the paragraphs within a document can reveal relatedness (finding relevance of the theme elsewhere) or unrelatedness (suggesting a departure from the original theme and starting of a new topic). When aggregated, intra-document results indicated an overall text structure while requiring no complex linguistic terminology. Also, at the time of retrieval, these internal links were comparable with queries and an extracted summary. As a result, automated summarization was hugely progressed which now provided to tailor specific need of users. Afterwards, Salton compared automatic text structuring and summarization with manual abstract production. In his analysis, he found that the abstract generated from the statistical model overlapped

[9] Cp. J. J. Pollock, A. Zamora, "Automatic Abstracting Research at Chemical Abstracts Service, 1975, pp. 226-232.
[10] Cp. Marie-Francine Moens, Automatic Indexing and Abstracting of Document Texts, 2000. pp. 144.
[11] Cp. Maristella, Crestani, Melucci, "Use of Information Retrieval Techniques for the Automatic Construction of Hypertext,", 1997, pp. 135

with about 45.6% of the text, a number that was almost exactly the same (45.81%) as human made text extraction.[12] In similar studies, to shed light on types of hyperlinks in a document collection of heterogeneous type, Allan evaluated the tf-idf weights and visualization techniques for graphical representation to identify their tangential, summary, expansion, comparison, contrast and equivalent links.[13]

- In the field of machine learning, Kupiec and others used bayesian statistics, which allowed recalculations as learning progressed.[14] Using frequencies of text elements in the source text as input, these probabilities measure how likely it is that a given sentence from a source text will appear in the summary. An accumulation of 188 full text/summary pairs were examined for sentence features including the length and of the sentence, cue phrases, keywords for the location of the sentence, and specific names. Various types of matches were identified from the analysis – direct matches were found in cases where summary sentence and original sentence were similar, or direct join where two source sentences were adjoined into a single summary sentence. At a 25% compression of the source text, the machine summaries overlapped 84% with the manual summaries which was double than Edmundson's citation of 42% overlapping at the same compression ratio.[15] In the end, a combination of location, cue phrase, and sentence length led to be the optimum feature for best results.

- Other researchers have drawn inspiration from Kupiec's work, including experiments in Korea that showed the Bayesian method was language independent.[16] It was found that the combination of cue words, sentence location, and title words was the most effective. To discover the best location features for sentence extraction, Hovy and Lin developed the "Optimal Position Policy," which identifies the places in the source text that are most likely to yield sentences which are worthy for abstracts.[17] Hovy and Lin leveraged Miller's WordNet database[18] in order to provide a rudimentary interpretation of selected sentences chosen by a topic-identification routine. Based on about 13,000

[12] Cp. Gerard Salton and others, "Automatic Text Structuring and Summarization,", 1997, pp. 193-207.
[13] Cp. James Allan, "Building Hypertext Using Information Retrieval,", 1997, pp. 145-159.
[14] Cp. Julian Kupiec, Jan Pedersen, and Francine Chen, "A Trainable Document Summarizer", 1995, pp. 68-73.
[15] Cp. Kupiec, Pedersen, and Chen, pp. 68-73.
[16] Cp. Sung Hyon Myaeng and Dong-Hyun Jang, "Development and Evaluation of a Statistically Based Document Summarization System 68-73.
[17] Cp. Eduard Hovy and Chin-Yew Lin, "Automated Text Summarization in SUMMARIST 81-94.
[18] Cp. G. Miller, "WordNet: A Lexical Database for English1995 39-41.

document/summary/keyword triples about technology industry announcements, they developed topic-rich keywords by applying tf.idf weighting algorithms and similarity coefficients. The keyword lists were used to develop ranked lists of sentence positions containing topical terms. A number of collections had varying topics/sentence positions; in the test set, the title was the most effective, then the sentence at the beginning of paragraph two. On contrary, an analysis of 30,000 Wall Street Journal articles found that the title was the most meaningful element, followed by the first paragraph's first sentence.[19]

2.3.2 Statistical Techniques

Statistical techniques and methods can use the statistical characteristics of sentences (such as title, location, term frequency) to summarize documents, assign weights to keywords, and then calculate sentence scores and choose the sentence with the highest score in the abstract.[20] Significance of a sentence can be decided by several methods such as –

- **Title method**

This method suggests that the sentences that appear in the title are considered more important and final summary is more likely to include the title sentence. The sentence score is calculated as the total number of words that are used most frequently between the sentence and the title. If the document does not contain header information, the title method will be invalid.[21]

- **Location method**

Whether the text appears at the beginning, middle, or end of a paragraph, or in a prominent part of the document (such as a conclusion or introduction), the text is weighted according to its position. As a result, last few sentences and the sentences from conclusion are entitled to be more important and those should be included in summary. Hovy & Lin and Edmundson used this method as we have already discussed in the previous section. The location method is based on the following intuition – sentences at the beginning and end of the paragraph, bold texts usually contain important insight for the summarization.[22]

---

[19] Cp. Allan, J. (1997). "Building hypertext using information retrieval", pp. 145-159.
[20] Cp. ibid
[21] Cp. Hovy, E., & Lin, C. Y. (1999). "Automated text summarization in SUMMARIST", pp. 81-94.
[22] Cp. Sreejith C, Sruthimol, P C Reghuraj, "Box Item Generation from News Articles Based Paragraph Ranking using Vector Space Model", 2014, no page number

- **tf-idf method**

The term frequency-inverse document frequency is a statistical method based on numeric which demonstrates the necessity of a word in a document. It is often utilized as a weighting factor while retrieving the information also and also used for text mining. tf-idf is broadly used for the filtering process of stop words in text summarization and also for the classification applications. The value of tf-idf grows proportionally with the number of appearances for a single word in the document. Search engines utilize the tf–idf weighting factor essentially by means of a core tool for scoring and ranking any document according to the relevance of query. Reverse document frequency is a measurement factor of whether the term is common or unusual for all documents. It is deducted by dividing the total number of documents by the number of documents containing the specific word.[23]

- **Cue word method**

In this method, certain positive and negative weight is assigned to a phrase or word according to the significance of its usage. Cue phrases usually vary by genre. The sentences which consists of such cue phrases are supposed to be included in the final summary. It can be assumed that such cue phrases or cue words provide a contextual scenario to identify the most important sentences. As a result, the summary obtained from this method is a set of cue words and the location where they are contained.[24]

- **LSA Method**

LSA method is known as Latent Semantic Analysis. Words or phrases which have semantic relationships in related contexts are considered in the same singular space, even though they do not have common words. This methodology can be used to efficiently extract the main content-sentences and subjective-words from the original document. The advantage of utilizing LSA vectors for summary generation instead of using the word vectors is that semantic relationships expressed by human are automatically identified. On contrary, applying word vectors without the LSA requires intensive designing to find out conceptual relations.[25][26]

---

[23] Cp. Allan, J. "Building hypertext using information retrieval. Information Processing & Management", 1997, pp. 145-159.
[24] Cp. Bagga, A. and B. Baldwin. "Entity-Based Cross-Document Cross-Referencing using the Vector Space Model", 1998, pp. 79-86.
[25] Cp. Lin, C-Y. and E.H. Hovy. "Automatic Text Categorization: A Concept-Based Approach", 1998, no page no.
[26] Cp. Naresh Kumar Nagwani ,Dr. Shrish Verma, "A Frequent Term and Semantic Similarity based Single Document Text Summarization Algorithm"

2.3.3 Linguistics Approach

Linguistics is a scientific study of language that combines both semantics and pragmatics. Semantics involves the systematic examination of the words and their meanings and Pragmatics includes an understanding of how context contributes to meaning. A linguistic approach aims to find the most important ideas by analyzing a set of words to determine their underlying connections. Although, linguistic approaches face challenges while using linguistic analysis tools of high quality (such as discourse parser) and linguistic resources (Lexical Chain, Word Net, Context Vector Space etc).[27]

- **Lexical chain**

Morris & Hirst were the first to introduce the concept of lexical chains. The main purpose of lexical chains is to exploit the cohesion between an arbitrary number of words. The lexical chains of a document can be calculated by grouping together words that have semantic links.[28] Word relationships such as identities, synonyms, and hypernyms/hyponyms can connect words and lead to their being grouped together lexically. The lexical chains are computed by grouping noun instances according to the previous relations. A lexical chain should be created by arranging words in the most powerful and longest way possible and there are several challenges when it comes to determining which word instance belongs to which lexical chain.[29]

- **Word Net**

WordNet is an English lexical database accessible through the Internet. It organizes the English words into a set of synonyms called a sys-net. Besides providing a short definition of each sys-net, WordNet also provides semantic relationships between them. The WordNet database also serves as a thesaurus and many systems use it to identify word relationships. Approximately 118,000 word forms are contained in Word Net. LexSum produces a lexical chain from Word Net for a summarization system.[30]

---

[27] Cp. Saggion, Horacio, Guy Lapalme. "Generating indicative-informative summaries with sumum.", 2002, pp. 497-526.
[28] Cp. Silber, H. Gregory, Kathleen F. McCoy. "Efficiently computed lexical chains", 2002, pp. 487-496.
[29]Cp. A.R.Kulkarni1, S.S.Apte2, "An Automatic Text Summarization Using Lexical Cohesion And Correlation Of Sentences ", 2014, no page number
[30] Cp. Kedar, Sarma, Sarma, Loiwal, Mehta, Ramakrishnan, Bhattacharyya. "Generic Text Summarization Using WordNet", 2004, no page number

- **Graph theory**

By combining the graph theory with the structure of the text and its relationship to the sentences of the document, we can compute the relationship between the sentences. Throughout the document, sentences are represented by nodes. The margins between nodes are viewed as connections between sentences which are related by similarity. In lexical representations of two sentences, the overlap is measured by the number of shared tokens between them. Different comparison criteria are developed to assess similarities between two sentences. In order to process a summary, only sentences with the highest scores are included. Based on the entire graph, the significance of various vertex elements is calculated iteratively. TextRank is an algorithm which utilizes a graph-based algorithm in the summarization process.[31]

- **Clustering**

Clustering involves grouping and clustering similar data in a document to serve as a summary. Summarization results are influenced not only by the sentence features, but also by the degree of similarity between sentences. Zhang Pei-ying and LI Cun-ho developed one of the sentence clustering methods. The K-means method is used to determine the number of clusters. By clustering the sentences, it identifies the topic sentences, and by extracting those, creates an extractive summary.[32]

While Linguistics play a crucial role in automated summarization, the approaches to implement it tend to be more difficult, while statistical approaches are more successful, though not without limitations. As there are thousands of books and documents in over thousands of languages in the world, the significance of automated summaries to be provided in different languages also seem to be an increasing demand in the market. To provide summarization in several languages, the integration of translation within the technical procedures of summarization is necessary. Up to now, we have limited resource on linguistic or computational studies in different aspects of summaries. It may be fruitful to conduct further research on the typology of summaries, Text analysis can be done by linguists as well as computational linguists in order to generate summaries that follow some or all of the characteristics listed before.

[31] Cp. Doran, William, Nicola Stokes, Joe Carthy, and John Dunnion. "Comparing lexical chain-based summarisation approaches using an extrinsic evaluation.", 2004, no page number
[32] Cp. Divya, S., & Reghuraj, P., “Eigenvector based approach for sentence ranking in news summarization”, 2014, no page number

Understanding the different types of summaries will facilitate the development of techniques and systems that will better serve the numerous purposes of summarization.

2.4 Evaluation Matrices

The evaluation matrix was formed keeping the purpose of use and the type of users in mind. The purpose of evaluation matrix was intended to weight the criteria according to the main user and to have a main objective to evaluate the application or connected object. Using an evaluation matrix is a method of objectively evaluating a series of options based on various criteria. Evaluation matrices are another way to encapsulate many standards as targets. The best practice of Evaluation Matrix according are as follows:

- Information of the user (for both description and the consent).
- Healthcare Content of the user (content design initialization, standardization, generative content and interpretive content).
- Technical content (technical design and flow of data).
- Security (including cybersecurity).
- Reliability (bargaining the confidentiality).
- Usage (integration, import, design).

All the criteria are ranked and the evaluation is done by weighting the most important elements. Two levels of scoring matrix can be used when the criteria must be absolutely met. Since requirements are generally written in ‘shall’ or ‘should’, the mandatory standards used in the first level assessment are expressed as ‘shall’ requirements, and the ‘should’ requirements are preferably used in the second level assessment level. When setting standards, the team should make every effort not to consider the options being considered to help ensure that the standards are not biased. If no criteria have been established, the design team must seek to reflect the current problem under review. To ensure that the criteria are produced without prejudice, the team should make every effort not to reflect on the choices under consideration when developing the criteria. Each criterion should be well worded so that everyone participating in the review process can understand it consistently. Each criterion should be unique from the others, and each should be double-checked to verify that none of them are incompatible.[33]

[33] Saggion H., D. Radev, S. Teufel, and W. Lam. 2002. Meta-evaluation of summaries in a crosslingual environment using content-based metrics. In *Proceedings of COLING-2002*, Taipei, Taiwan.

Desirable criteria are those that should be satisfied to some extent in order to reap some advantage. Satisfaction with these criteria is ideal, although some may be compromised in favor of others if the end outcome allows for a net gain. The evaluation team is free to rate desired factors as they see fit. In general, the higher the value ascribed to a criterion, the more important or valuable it is. A scale of 1 to 10 or a "high, medium, low" scale with numeric representations for each level (e.g., high=5, medium=3, low=1) can be utilized. When assessing the relevance of criteria, it is common to consider all criteria to be somewhat essential. As a result, a good rule to follow is that each rating should be utilized at least once and no more than three times. A single criterion can be presented as both a necessary and a desirable criterion for evaluation purposes.

2.4.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a computerized method used for text analysis that is founded on a set of ideas as well as a set of technology. And, because this is a very dynamic area of study and growth, there isn't a single agreed-upon description that would please everyone, but there are several elements that any informed person would include in their definition. Natural Language Processing is a set of theoretically based computer approaches for evaluating and modeling naturally occurring texts at one or more levels of linguistic analysis in order to achieve human-like language processing for a variety of activities and applications.[34] Several aspects of this concept can be expanded upon. For starters, the nebulous concept of a 'spectrum of computational approaches' is required since there are several methods or strategies from which to pick when performing a certain sort of language analysis. Naturalistic literature can be written in any language, style, or genre. Oral or written texts are both acceptable. The only stipulation is that they are written in a language that humans use to communicate with one another. Furthermore, the text being analyzed should not be produced especially for the sake of the study, but rather obtained from real use.[35]

The concept of 'levels of linguistic analysis' refers to the fact that when people generate or comprehend language, many forms of language processing are known to be at work. Humans are considered to use all of these levels on a regular basis since each level transmits distinct

[34] Cp. Chowdhury, Gobinda G. "Natural language processing", 2003, pp. 51-89.
[35] Cp. Nenkova, Ani, and Kathleen McKeown. "A survey of text summarization techniques", 2012, pp. 43-76.

sorts of meaning.[36] However, different NLP systems use different degrees of language analysis, or combinations of levels of linguistic analysis, as evidenced by the variations in NLP applications. NLP is regarded a discipline within Artificial Intelligence, according to the term ‘human-like language processing' (AI). While NLP's complete genealogy is dependent on a number of different disciplines, it is fair to consider it an AI discipline because it aspires for human-like performance. As mentioned above, the objective of NLP is to "achieve human-like language processing."[37] The term "processing" was chosen with care and should not be substituted with "understanding." Because, while the discipline of NLP was once known as Natural Language Understanding (NLU) in the early days of AI, it is now widely acknowledged that, while the aim of NLP is genuine NLU, it has yet to be achieved. A complete NLU system would be able to do the following:

1. Take an input text and translate it into a different language.
2. Respond to questions regarding the text's substance.
3. Use the text to make assumptions

While NLP has made significant progress in achieving these three objectives, the fact that NLP systems cannot draw conclusions from text on their own means that NLU remains the aim of NLP. The ‘levels of language' approach is the most explanatory technique for describing what actually happens within a Natural Language Processing system.[38] This is also known as the synchronic model of language, and it differs from the previous sequential model, which proposes that the levels of human language processing occur in a precise sequential order. Language processing, according to psycholinguistic research, is significantly more dynamic, as the levels can interact in a number of ways. Introspection indicates that we regularly employ knowledge gained at a higher level of processing to aid in a lower level of analysis. Because of this, the following level descriptions will be provided in order. The important point here is that every level of language conveys meaning, and since humans have been demonstrated to use all levels of language to achieve comprehension, the more capable an NLP system is, the more levels of language it will employ.[39]

---

[36] Cp. Li, Liuqing, Jack Geissinger, William A. Ingram, Edward A. Fox. "Natural language processing through big data text summarization", 2020, no page number
[37] Cp. Tas, Oguzhan, and Farzad Kiyani. "A survey automatic text summarization" ,2007, pp. 205-213.
[38] Cp. Saggion, Horacio, Kalina Bontcheva, Hamish Cunningham. "Robust generic and query-based summarization", 2003, no page number
[39]Cp. Zhang, Haoyu, Jianjun Xu, and Ji Wang. "Pretraining-based natural language generation for text summarization", 2019, no page number

While natural language processing (NLP) is a relatively new field of research and application in comparison to other information technology approaches, there have been enough successes to suggest that NLP-based information access technologies will continue to be a major area of research and development in information systems now and in the future.[40] Most of the websites and mobile application has some kind of Natural Language Processing in-build into them for summary generation. It's quite popular and multi-purpose because the same resources can be harvested again into other application such as speech generation and recognition. So far, the focus of natural language processing technology has been to automate how to target short articles. However, we are witnessing a change in recent times. Projects such as Gutenberg (http://www.gutenberg.org), Google Books Search (http://books.google.com) or the Million Books Plan (http://www.archive.org/details/millionbooks) are making more and more books available in electronic format. Likewise, most of the recently published books are often available. This implies that the need for language processing techniques to be capable of handling very large documents such as books is becoming necessary. Although a lot of research has been done on the task of text summarization, most of the work has focused on short summaries, especially news. However, different books are of different length and genre, so different summarizing skills are required for automating the process. In fact, the simple application of state-of-the-art aggregation tools can lead to poor results.[41] This is not surprising, because these systems are specifically developed to summarize brief informational documents.

2.4.2 Recall, Coverage and Retention

In the study of summarization, memories at different compression rates have been used to assess the efficiency of automated systems in recalling important materials from original document.[42] Assume we have a system overview Ss and a model overview Sm. The number of sentences in Ss is Ns, the number of sentences in Sm is Nm, and the total number of sentences in both Ss and Sm is Na. Hence Recall is defined as (Na/Nm). Applying this technique directly without modification is not appropriate because –

- Multiple system units contribute to multiple model units.
- Exact overlap between S_S and S_m does not seem to occur often.

[40] Cp. Kota, Chowdary, Reddy, Prasanna. "Text Summarization Using Natural Language Processing", 2021, pp. 535-547.
[41] Cp. Zhang, Haoyu, Jianjun Xu, and Ji Wang. "Pretraining-based natural language generation for text summarization", 2019
[42]Cp. Melamed, I. Dan, Ryan Green, Joseph Turian. "Precision and recall of machine translation", 2003, pp. 61-63.

- Judgment of overlapping is not binary.

The Compression Ratio is therefore calculated by dividing the length of the model summary by the length of the original document. Recall at threshold t, Recall, can be defined as – (Number of Model Units marked at or above *t* / Total number of Model Units in the model summary). Instead of thresholds, coverage C can be utilized as the coverage score – 1 for all, 3/4 for most, 1/2 for some, 1/4 for scarcely any, and 0 for none. To avoid confusion with information retrieval recall, it can be redefined as weighted retention and can be described it as follows – {(Number of Model Units marked * *C)* / Total number of Model Units in the model summary}[43]

2.4.3 Text Segmentation

The large theme change typical of large documents is a significant difference between short documents and long documents. Although short documents usually focus on one topic at a time, longer documents (such as books) sometimes cover multiple topics. Therefore, the abstract is supposed to include information that covers all the key elements of the document's topic, not just the general aspects that apply to the entire document. Therefore, a system for summarizing large texts should extract basic ideas from the topics of all documents. If the topic boundaries are known before the summarization stage, this work will be easier to complete. This is where text segmentation comes in. It tries to detect subject transitions and thus divide the document into smaller parts. Please note that although some books in our database have access to chapter restrictions, this is not always the case because other volumes do not have clear chapter restrictions. We chose not to use chapter borders, but to use automatic text segmentation methods to ensure uniform processing of the entire data set.[44]

Although various text segmentation systems have been developed so far, most applications choose to use graph-based segmentation algorithms and standardized slices, which have proven to be superior to other technologies. Simply put, the segmentation method first models the text as a graph, sentences as nodes, and intersecting similarities as weighted edges. Cosine similarity is used to calculate sentence similarity, and the smoothing factor is combined with the term count in adjacent sentences. A modification of the tf.idf metric is used to weight the words, where the document is evenly divided into blocks for tf.idf calculations. In this technique, two parameters must be set – the word length of the block that approximates the

[43] Cp. ibid
[44] Cp. Over, Paul. "Evaluation of generic news text summarization systems", 2003, no page number

sentence and the cut-off value for drawing borders between nodes.[45] We cannot use the same settings as pointed out in the document because the technique was originally used for oral lecture segmentation. Once the text is divided into several parts, a separate summary for each fragment is generated, and then the final summary is created by combining the sentences in the summary of each fragment. In other words, the application selects one sentence from each paragraph summary at a time until we reach the required book summary length, starting with the sentence ranking list provided by the summary algorithm for each paragraph.[46]

2.4.4 ROUGE System

ROUGE is an acronym that stands for Recall-Oriented Understudy for evaluation. It provides metrics to determine the quality of an automated summary by comparing it to other summaries written by people in order to understand the efficiency. The measurements count the amount of overlap units such as n-grams, word sequences, and word pairs that exist between the computer-generated summaries to be assessed and the ideal summaries created by humans. Human evaluations of several quality criteria, including as coherence, conciseness, grammaticality, readability, and substance, have traditionally been used to evaluate summarization. Even a basic manual review of summaries on a wide scale based on a few linguistic quality questions and content coverage, such as that used in the Document Understanding Conference (DUC) (Over and Yen, 2003), would take over 3,000 hours of human work.[47]

This is both costly and difficult to do on a regular basis. As a result, the question of how to automatically assess summaries has gotten a lot of interest in the summarizing research community in recent years. For example, offered three content-based evaluation techniques for comparing summary similarity. Cosine similarity, unit overlap (i.e. unigram or bigram), and longest common subsequence are the methods used. They did not, however, illustrate how the outcomes of these automated evaluation systems correspond with human judgements. Lin and Hovy (2003) demonstrated that techniques comparable to BLEU, such as BLEU (Papineni et al., 2001), may be used to evaluate machine translation, n-gram co-occurrence statistics, for example, might be used to assess summaries. We introduce ROUGE, a software for automated

[45] Cp. Horacio, Radev, Teufel, Lam. "Meta-evaluation of summaries in a cross-lingual environment", 2002, no page number
[46]Cp. Dragomir, Teufel, Saggion, Lam, Blitzer, Celebi, Qi, Drabek, Liu. "Evaluation of text summarization", 2002, no page number
[47] Cp. ibid

assessment of sum- maries and their evaluations, in this work. Re-call-Oriented Understudy for Gisting Evaluation (ROUGE) is an acronym for Re-call-Oriented Understudy for Gisting Evaluation. It provides many automated assessment methods for comparing summary similarity.[48]

3. Research Design

Although, this piece of research can be used by practitioners and academic users, however, in order to gain a comprehensive understanding of the specific factors driving towards change, academic research on these issues should be examined and analyzed thoroughly. While the use of technology has become more mainstream and well recognized within the profession, the effects of blockchain and artificial intelligence remain somewhat of a mystery, both for practitioners and researchers alike. In order to obtain a more comprehensive understanding of how the profession will evolve over time, one should first analyze, understand, and comprehend what said technologies are, as well as how they differ from current market choices. We have stormed through the internet and finally shortlisted some of the literature work by previous researchers and identified the key findings out of them.

3.1.Literature Analysis

Table 1: Literature Research & Key Findings

Title	**Key Finding**	**Author**	**Publisher**
Towards Automated Related Work Summarization	1. Related Work Summarization system is a prototype, which takes in set of keywords arranged in a hierarchical fashion that describes a target paper's topics. 2. Initial results show an improvement over generic multi-document summarization baselines in human evaluation.	Cong Duy Vu Hoang and Min-Yen Kan	National University of Singapore
Relevance Vector Machine Optimization in Automatic Text Summarization	1. Relevance Vector Machine (RVM) algorithm in automatic text summarization 2. the correlation coefficient can be used to determine the order of extraction features	K E Dewi, E Rainarli	Informatics Engineering Department, Universitas

[48] Cp. ibid

			Komputer Indonesia
Automatic Summarization using Terminological and Semantic Resources	1. automatic summarization of specialized texts combining terminological and semantic resources: a term extractor and an ontology. 2. the highest score are chosen to take part of the final summary. 3. evaluate the algorithm with ROUGE, comparing the resulting summaries with the summaries of other summarizers.	Jorge Vivaldi, Iria da Cunha, Juan-Manuel Torres-Moreno and Patricia Velazquez-Morales.	Institut Universitari de Linguistica Aplicada (UPF),
Explorations in Automatic Book Summarization	1. address this gap and explore the problem of book summarization. 2. introduce a new data set specifically designed for the evaluation of systems for book summarization. 3. systems developed for the summarization of short documents	Rada Mihalcea and Hakan Ceylan	Department of Computer Science University of North Texas.
Automated Summarization Evaluation (ASE) Using Natural Language Processing Tools	1.Introduce an automated summarization evaluation (ASE) model that depends strictly on features of the source text or the summary, allowing for a purely text-based model of quality. 2. This model effectively classifies summaries as either low or high quality with an accuracy above 80%.	Scott A. Crossley,Minkyung Kim, Laura Allen and Danielle Mc Namara	Artificial Intelligence in Education, 2019, Volume 11625

A Neural Attention Model for Sentence Summarization	1. a fully data-driven approach to abstractive sentence summarization. 2. this method utilizes a local attention-based model that generates each word of the summary conditioned on the input sentence. 3. The model shows significant performance gains on the DUC-2004 shared task compared with several strong baselines.	Alexander M. Rush, Sumit Chopra and Jason Weston	Facebook AI Research
Using Latent Semantic Analysis in Text Summarization and Summary Evaluation	1. two new evaluation methods based on LSA, which measure content similarity between an original document and its summary. 2. compare seven summarizers by a classical content-based evaluator and by the two new LSA evaluator 3. an influence of summary length on its quality from the angle of the three mentioned evaluation methods	Josef Steinberger, Karel Ježek	Department of Computer Science and Engineering, Univerzitní 22, CZ-306 14 Pilsen, Czech Republic
Neural Latent Extractive Document Summarization	1. Extractive summarization models require sentence-level labels 2. aims to automatically rewrite a document into a shorter version while retaining its most important content 3. pro- posed model can indeed improve over a strong ex- tractive model while application of the compression model to the output of our extractive system leads to inferior output.	Xingxing Zhang Mirella Lapata Furu Wei Ming Zhou	Microsoft Research Asia, Beijing, China
Extractive Document Summarization Based on Convolutional Neural Networks	1. adapt the original CNN model to address a regression process for sentence ranking. 2. . Pre-trained word vectors are used to enhance the performance of the model. 3. evaluate the proposed method on the DUC 2002 and 2004 datasets covering single and	Yong Zhang, Yong Zhang and Mahardhika Pratama	Nanyang Technological University Singapore and department of Computer Science and

	multi-document summarization tasks, respectively.		IT La Trobe University Melbourne, Victoria 3086, Australia
The model shows significant performance gains on the DUC-2004 shared task compared with several strong baselines.	1. a global encoding framework, which controls the information flow from the encoder to the decoder based on the global information of the source context. 2. Experiments on the LCSTS and Giga word show that our model outperforms the baselines, and the analysis shows that it is able to reduce repetition in the generated summaries, and it is more robust to inputs of different lengths, compared with the conventional seq2seq model.	Junyang Lin, Xu Sun, Shuming Ma, Qi Su	MOE Key Lab of Computational Linguistics, School of EECS, Peking University School of Foreign Languages, Peking University
Extractive Text Summarization using Neural Networks	1. a fully data-driven approach using feedforward neural networks for single document summarization. 2. train and evaluate the model on standard DUC 2002 dataset 3. The proposed model is scalable and is able to produce the summary of arbitrarily sized documents.	Aakash Sinha, Abhishek Yadav and Akshay Gahlot	Department of Computer Science and Engineering
Manual and Automatic Evaluation of Summaries	1. manual and automatic evaluations of summaries using data from the Document Understanding Conference (DUC-2001) 2.accumulative n-gram overlap scores between system and human summaries.	Chin-Yew Lin and Eduard Hovy	USC Information Sciences Institute

Abstractive Summarization on Dynamically Changing Text	1.Abstractive summarization is there is an excessive amount of specialize in generating good results with respect to a specific sentence and insufficient on the corpus of text containing thousands of such sentences. 2. A transformer model is used to get individual sentence summaries of respected review text 3. used a mix of Universal Sentence Encoder, statistical methods and graph reduction algorithm to pick the foremost relevant sentences to best represent the full text. 4. results show that even by increasing the degree of contraction of the text corpus like particularly large text corpus, the identical accuracy is achieved.	Rahul, Pranay Rawat , Vivek, Amaan Elahi	Department of Computer Science and Engineering
New Alignment Methods for Discriminative Book Summarization	1. two new methods based on hidden Markov models, generally targeted to the present problems and demonstrate gains on an extractive book summarization task. 2. all sentences are aligned to a sentence within the summary are assigned a label of 1 (appearing in summary) and 0 otherwise (not appearing in summary).	David Bamman and Noah A. Smith	School of Computer Science Carnegie Mellon University
Automatic Summary Evaluation without Human Models	1.KL and Jensen-Shannon divergence , Cosine similarity, unigram and multinomial models of text 2.Results can be used to rank participating systems very similarly to manual human model based evaluations	Annie Louis and Ani Nenkova	University of Pennsylvania Philadelphia, PA 19104, USA

	3.Jensen-Shannon divergence conducts to correlation as high as 0.9 with manual evaluations.		
Generative Adversarial Network for Abstractive Text Summarizatio	1.Train a generative model as an agent of reinforcement learning, which takes the raw text as input and predicts the abstractive summarization and a discriminative model to differentiate the generated summary from the bottom truth summary. 2.Model evaluates moderate ROUGE scores with the state of the art methods on CNN/Daily Mail dataset. 3. The model is ready to come up with more abstractive, readable and diverse summaries	Linqing Liu, Yao Lu, Min Yang, Qiang Qu, Jia Zhu, Hongyan Li	Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Alberta Machine Intelligence Institute
Multi-document summarization of news articles using an event-based framework	1.an event based framework is presented for integrating and organizing information extracted from different articles. 2. a tree view interface was implemented for showing a multi document summary based on the framework. 3.all the human subjects preferred the framework based summaries. 4.multi document summary of news article can choose the proposed event-based framework.	Shiyan Ou, Christopher S.G. Khoo and Dion H. Goh	Division of Information Studies, School of Communication and Information, Nanyang Technological University, Singapore
Text summarization using Latent Semantic Analysis	1. Different LSA-based summarization algorithms are explained between them two of which are presented by the author of this paper. 2.Their performances are compared using their ROUGE scores.	Makbule Gulcin Ozsoy and Ferda Nur Alpaslan and	Department of Computer Engineering, Middle East Technical

	3.One of the algorithms produces the most effective scores and both algorithms perform equally well.	Ilyas Cicekli	University, Turkey. Department of Computer Engineering, Hacettepe University, Turkey.
Compressive Cross-Language Text Summarization	1.Cross-language text summarization produces a summary in a language, which is not similar to the language of the source documents. 2. Analyzed other NLP tasks like word encoding representation, semantic similarity, sentence and multi sentence compression for producing more fixed and instructive cross lingual summaries. 3. A neural network model that mixes recurrent and convolutional neural networks to estimate the semantic similarity of a pair of sentences (or texts) based on the local and the general contexts of words. 4.This model predicts better similarity scores than baselines by analyzing better the local and also the general meanings of words and multi-word expressions. 5. A multi-sentence compression method that compresses similar sentences by fusing them in correct and short compressions that contain the most important information of those similar sentences.	Elvys Linhares Pontes	ACADÉMIE D'AIX-MARSEILLE UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

Automated Summarization Evaluation with Basic Elements	1.a framework called BE Package within which various implementations of automated summary content evaluation methods are often housed and compared. 2. Implement a particular evaluation method using very small units of content, called Basic Elements. 3. This method is tested on DUC 2003, 2004, and 2005 systems and produces excellent correlations with human judgments.	Eduard Hovy, Chin-Yew Lin, Liang Zhou, and Junichi Fukumoto	Information Sciences Institute ,University of Southern California
Recent automatic text summarization techniques: a survey	1.an inclusive survey of extractive approaches of latest text summarization developed within a period of 10 years. 2. abstractive and multilingual text summarization approaches 3. intrinsic still as extrinsic both the methods of summary evaluation are described very well together with text summarization evaluation conferences and workshops. 4. evaluation results of extractive summarization approaches are presented on some shared DUC datasets.	Mahak Gambhirand Vishal Gupta	University Institute of Engineering and Technology, Panjab University, Chandigarh, India
CaseSummarizer: A System for Automated Summarization of Legal Texts	1.CaseSummarizer is a toll for automated text summarization of legal documents that uses standard summary methods supported word frequency augmented with additional domain specific knowledge. 2. Summaries are provided through an informative interface with abbreviations,	Seth Polsley, Pooja Jhunjhunwala, Ruihong Huang	Department of Computer Science and Engineering, Texas A&M University

	significance heat maps, and other flexible controls. 3. Evaluate the summary using ROUGE and human scoring against several other summarization systems including summary text and feedback provided by domain experts.		
EVALUATION MEASURES FOR TEXT SUMMARIZATION	1. a new evaluation measure for assessing the standard of a summary. 2. Latent Semantic Analysis (LSA) which may capture the main topics of a document. 3. a high correlation between human rankings and also the LSA-based evaluation measure. 4.gives more precise results, using a standard ROUGE measure.	Josef Steinberger, Karel Jezek	Department of Computer Science and Engineering, University of West Bohemia in Pilsen, Czech Republic
Multi-document text summarization - A survey	1.Cluster based approach, Topic based approach, Lexical Chains approach 2. LexRank takes all the sentences within the cluster under consideration while constructing the similarity graph which incorporates unnecessary sentences moreover. 3. the gist of every document by extracting few sentences from each of them and apply LexRank algorithm, the complexity yet because the density of the resultant graph is reduced by a greater amount.	Amol Tandel, Brijesh Modi, Priyasha Gupta, Shreya Wagle and Mrs. Sujata Khedkar	Dept. Of Computer Engineering V.E.S.I.T. Chembur, Mumbai, India

Source: Adapted from various Literature Research

4. Research Results

In our research we tried to collect as much information as we can from the company website and from other sources about the companies who provide the Automated book summaries or text summaries. These are the some of the organization's name from which we get the information about what kind of summarization they provide and how long they are in this business, which technology they are using, how much market share they hold and what are they going to bring in the market.

4.1. Software & Service Provider:

These are some Software or service provider all over the world which we have shortlisted –

- Getabstract
- Smmry
- Soundview
- Sassbook.com
- Readingraphics
- Tools 4 noobs
- Spinbot.com
- Text Summarization
- Resoomer
- Spin Rewriter
- Article Builder
- WordAI
- Article Forge
- Summarizebot

4.2. Companies and their Working Procedures:

In this topic we gathered all the information available out there regarding their services, features, working procedure, pros & cons, etc.

Summarizebot:

Summarizebot is specialized in information extraction, structuring and analyzing. Summarize bot uses technologies like machine learning, natural language processing, artificial intelligence and blockchain to simplify complex and long process and automate which was done by human intelligence previously. They provide abstractive summarization, predictive analytics and question answering systems for fintech industry. According to their website they developed and using advanced analytical artificial intelligence algorithm to provide service for financial, automotive, publishing, finance, media, legal, consulting, hospitality, healthcare, pharmaceutical and other industries.

Table 2: Overview of SummarizeBot

Products	**Features**	**Technology**	**Plan & Price**					
			Plan Type	**Price**	**Period**	**Request**	**API/ Minute**	**Max File Size**
Summarization Sentiment Analysis News Aggregation Fake News Detection Linguistic Processor Audio Summarization Semantic Search Emotion/Mood Analysis Named Entity Detection Intent Analysis Short Text Language Detection Articles Search Keywords Extraction Article Extraction Good/Bad News Analysis	Summary generation News Summaries Keywords Extraction Key Fragments List Vary Summary Size Save results	Machine Learning Artificial Intelligence Blockchain	Free Standard Custom	$0 $179 UL	14 30 UL	5000 120k UL	5 20 UL	3 10 UL

Comment Extraction								
Face Detection								
Image Recognation								
Language Detection								
Video Identification								

Source: Adapted from Summarizebot.com

Pros:

- No Character Limit
- Work with Slack and Facebook
- URL & File import and Export option available
- Custom Summary setting
- Multi Language Support
- No Advertisement

Cons:

- No Web Version

Smmry:

Smmry is using the same pronunciation Summary but with a different spelling. Smmry provide the text summarization by using their core algorithm. They use Natural Language processing model to summarize the text, but they did not mention what type of algorithm and how does it work. The procedure of summarization is first they rank each sentence according to their importance using their core algorithm. Then the summarization par comes in where core algorithm recognizes the summary by focusing on the topic and the keywords. Then it removes the transition Phrases, unnecessary clauses, and excessive examples from the text. The core algorithm summarizes the text in several steps, which includes:

1) Associate all the words with grammatical counterparts.

2) Calculate the occurrence of words in the text.

3) Identify the popularity of each word and assign them in order.

4) Detects periods which represent the ending statement of sentences.(e.g "Mr." does not).

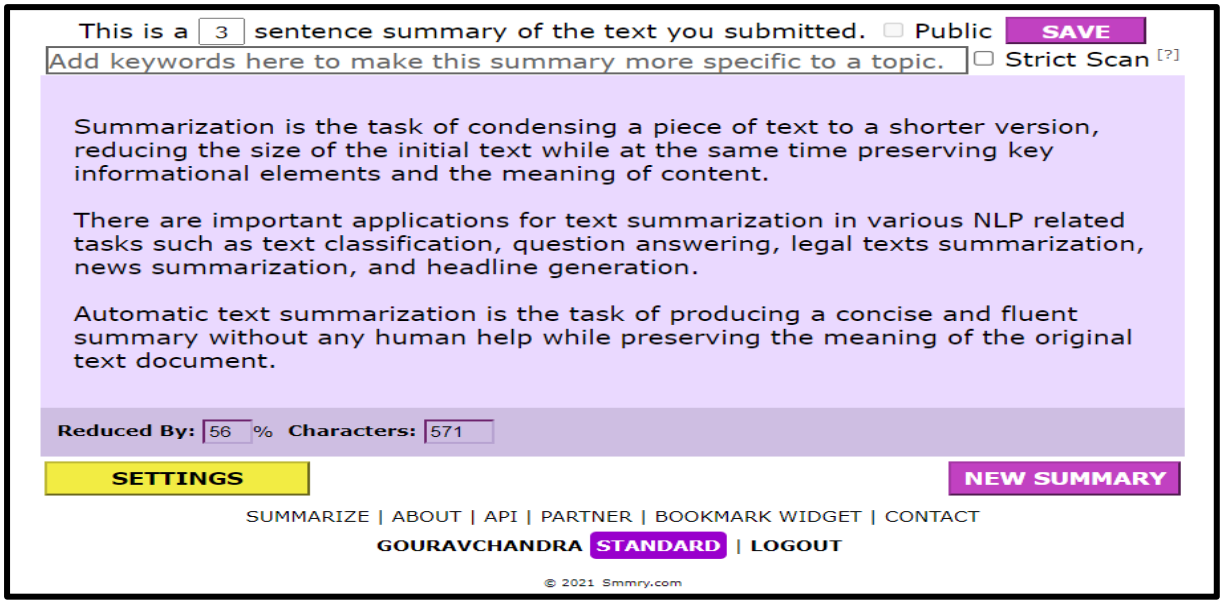
5) Split all the text into individual sentences.

6) Rank all the sentences by the occurrence of the words.

7) Return X in chronological order of highly ranked sentences.

Summary is free service on the website with some limitation like user can paste the text and upload the file or paste website URL but cannot export the summarized text. They only providing access to save the summarization within the website with a standard account. Although there are no word limit but the features of summarization is is also limited. After summarization users can define the following terms to get their desired summarization like reduced % of text, count of characters and count of sentences. It also provide keyword summarization, which means after summarized text, user can choose the specific keywords and get the summarized results of that specific keywords.

Figure 1: Smmry result:

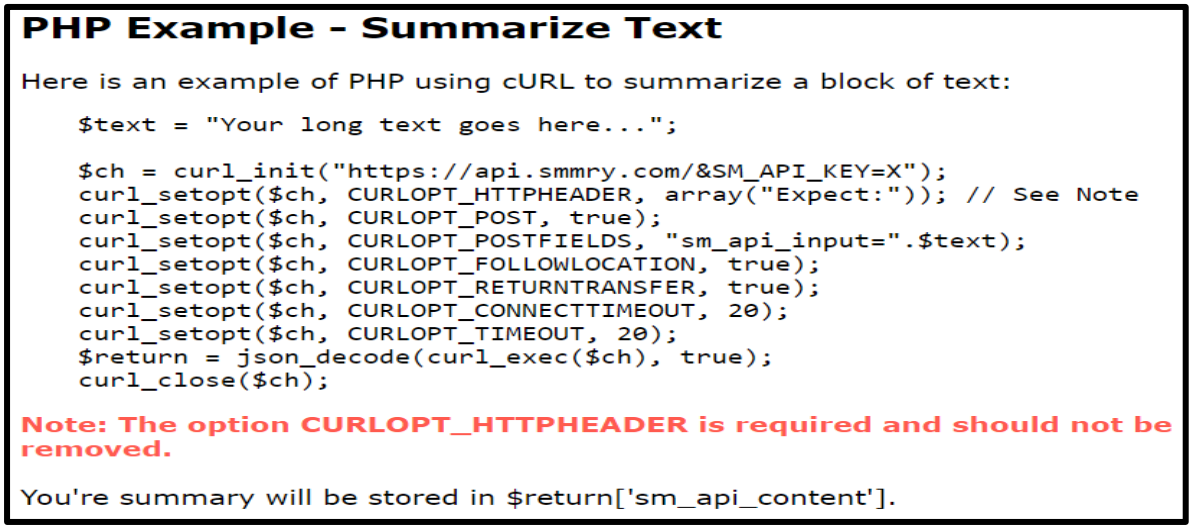


Source: Smmry.com

Smmry also provide API access to registered and premium partners where they can use the API keys to do the summarization directly in their systems. They provide two plans for the API token, first one is free in which they provide 100 free API request daily in which 10 second buffer time between the API request. Second plan is Premium one where they provide unlimited API request with no buffer time. In the premium features, with 1 credit point user can access 500 characters of reused text summarization returned by API and that will cost

$0.001. An additional 8 credit will cost for every new text summarization. Reuse of same text summarization will be cheaper.

Figure 2: Sample algorithm of smmry.



PHP Example – Summarize Text

Here is an example of PHP using cURL to summarize a block of text:

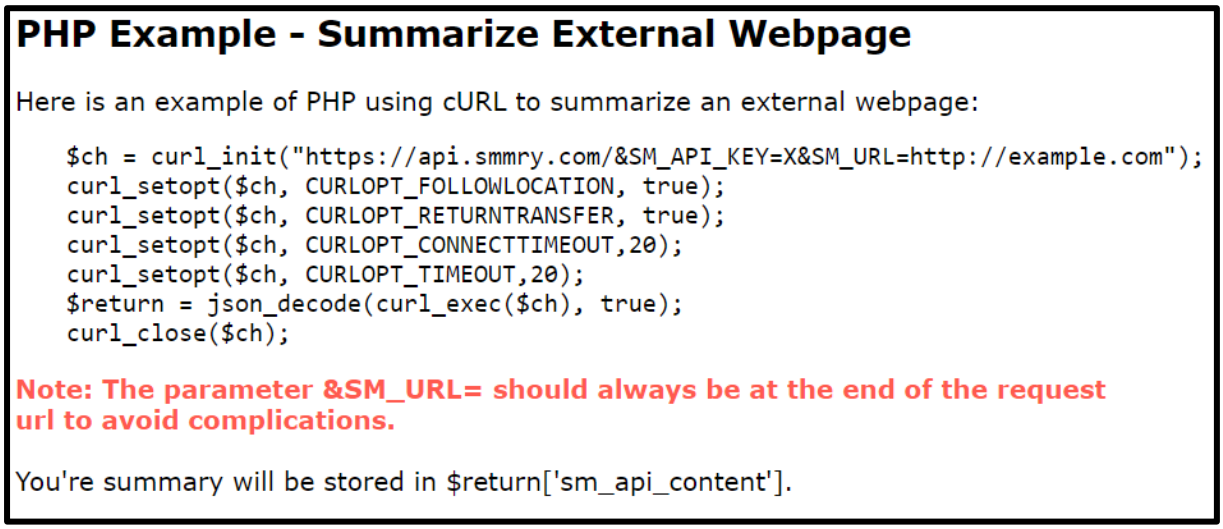
```
$text = "Your long text goes here...";

$ch = curl_init("https://api.smmry.com/&SM_API_KEY=X");
curl_setopt($ch, CURLOPT_HTTPHEADER, array("Expect:")); // See Note
curl_setopt($ch, CURLOPT_POST, true);
curl_setopt($ch, CURLOPT_POSTFIELDS, "sm_api_input=".$text);
curl_setopt($ch, CURLOPT_FOLLOWLOCATION, true);
curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);
curl_setopt($ch, CURLOPT_CONNECTTIMEOUT, 20);
curl_setopt($ch, CURLOPT_TIMEOUT, 20);
$return = json_decode(curl_exec($ch), true);
curl_close($ch);
```

Note: The option CURLOPT_HTTPHEADER is required and should not be removed.

You're summary will be stored in $return['sm_api_content'].

Source: smmry.com

Figure 3: Sample algorithm of smmry2



PHP Example - Summarize External Webpage

Here is an example of PHP using cURL to summarize an external webpage:

```
$ch = curl_init("https://api.smmry.com/&SM_API_KEY=X&SM_URL=http://example.com");
curl_setopt($ch, CURLOPT_FOLLOWLOCATION, true);
curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);
curl_setopt($ch, CURLOPT_CONNECTTIMEOUT,20);
curl_setopt($ch, CURLOPT_TIMEOUT,20);
$return = json_decode(curl_exec($ch), true);
curl_close($ch);
```

Note: The parameter &SM_URL= should always be at the end of the request url to avoid complications.

You're summary will be stored in $return['sm_api_content'].

Source: smmry.com

Table 3: A short overview of smmry:

Products	Features	Tech.	Plan & Price				Platforms
			Plan Type	Price	API Request	Max File Size	
Text Summarization	1) Associate words with their grammatical counterparts. 2) Calculate the occurrence of each word in the text. 3) Assign each word with points depending on their popularity. 4) Detect which periods represent the end of a sentence. 5) Split up the text into individual sentences. 6) Rank sentences by the sum of their words' points. 7) Return X of the most highly ranked sentences in chronological order.	Core Algorithm	Free Full	$0 $0.00 1/Credit	100/Day UL/Day	not available 500 characters /Credit	Browser Android Web iOS

Source: smmry.com

Pros:

- Simple Interface
- Custom Summary Size
- Additional Useful Tools
- Files can be imported from URL or Local Storage
- No Ads

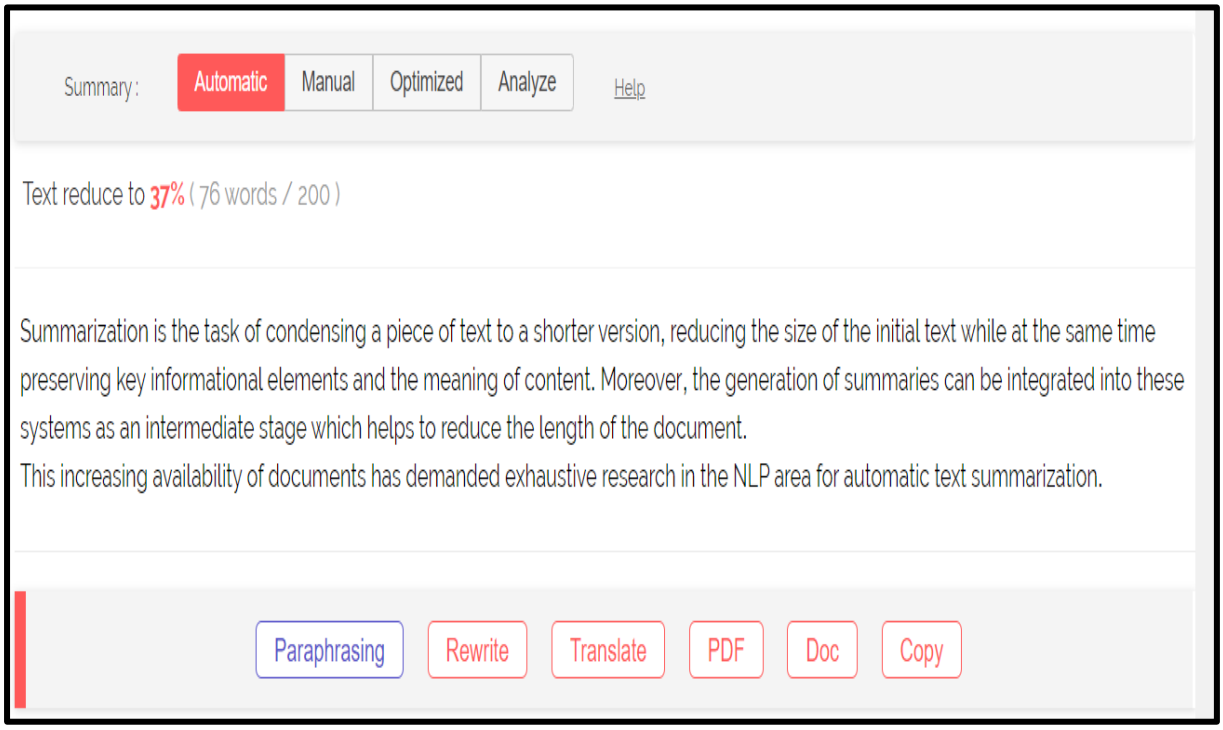
Cons:

- No File Export option
- Not many features available to categorize the summarization
- No multiple Language support

Resoomer:

Reesomer is France based organization who provide text summarization tool for argumentative text. It provides summarization for scientific text, history text, and other articles. It supports multiple language which gives an advantage to this tool. Although they did not revel about their technology or model they are using to summarize the text but they are pretty clear with the fact that users can get good results only if the text is argumentative. They also have preety good partners who use their service like Journalism.co.uk, BrainBuxa, GEEKHEBDO, and JOOBLE. Resoomer provides numerous settings to get the desired results like Automatic summarization, manual summarization, optimized summarization, and analyze text.

Figure 4: An example of text summarization by Resoomer:



Source: Resoomer.com

Resoomer do not provide API tokens, they only provide browser extension. Although Resoomer is popular among the students, teachers, editors and it provides goods results, but it still has some pros and cons.

Table 4: Simple overview of Resoomer:

Products	**Features**	**Plan & Price**				**Platforms**
		Plan Type	**Price**	**Period**	**Max File Size**	
Text Summarization	Automatic Mannual Optimized Analyzed	Free	0		40000 Characters	Web Web Extension
		Premium	4.90 Euro	30	200000	
			12.90 Euro	90	200000	
			39.90 Euro	365	200000	

Source: Adapted from resoomer.com

Pros:

- Multi Language Support
- No Registration
- Export to PDF/DOC
- Custom Summary
- URL Import Support
- 40000 characters limit in free version

Cons:

- Argumentative texts only
- No file import option in free version
- Ads is Free Version
- NO API

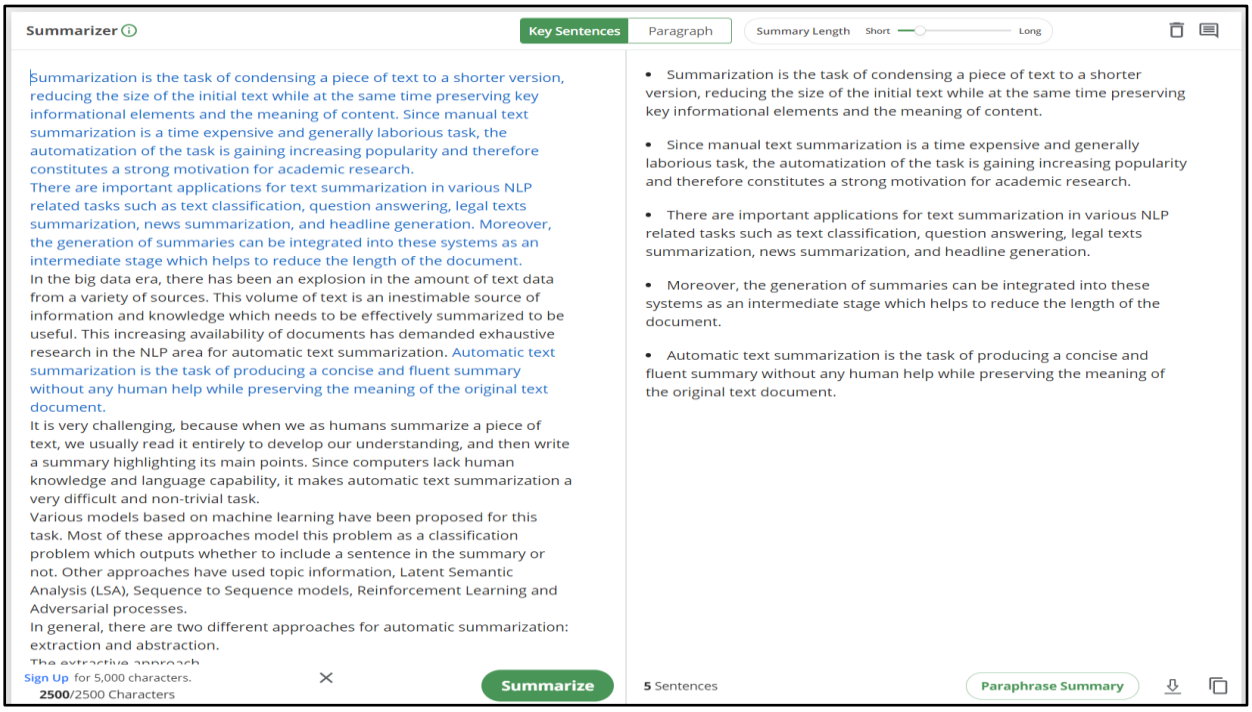
QuillBot:

Quillbot is a low-cost, high-quality paraphrasing tool. Its primary goal is to repurpose text while keeping the original meaning intact. Rohit Gupta, Anil Jason, and David Silin co-founded the company in 2017. Since then, they've continued to improve the product's quality by introducing additional features. It claims, it can rephrase any text with perfect language, stye and tone for any occasion. Quillbot provides 4 different types of services, which are paraphraser, Grammar Checker, summarizer, and citation generator. As we are only talking about summarization so let's focus on the summarization process of QuillBot. In text summarization they don't provide a wide variety of settings. They provide Key sentences, paragraph, and summary length of the text. An example of QuillBot Summarizer –

Key Sentences:

This mode provide the most important sentences from the paragraph. Moreover, it also gives the control to choose the quantity of sentences and the length of the sentence.

Figure 5: Key Sentences:

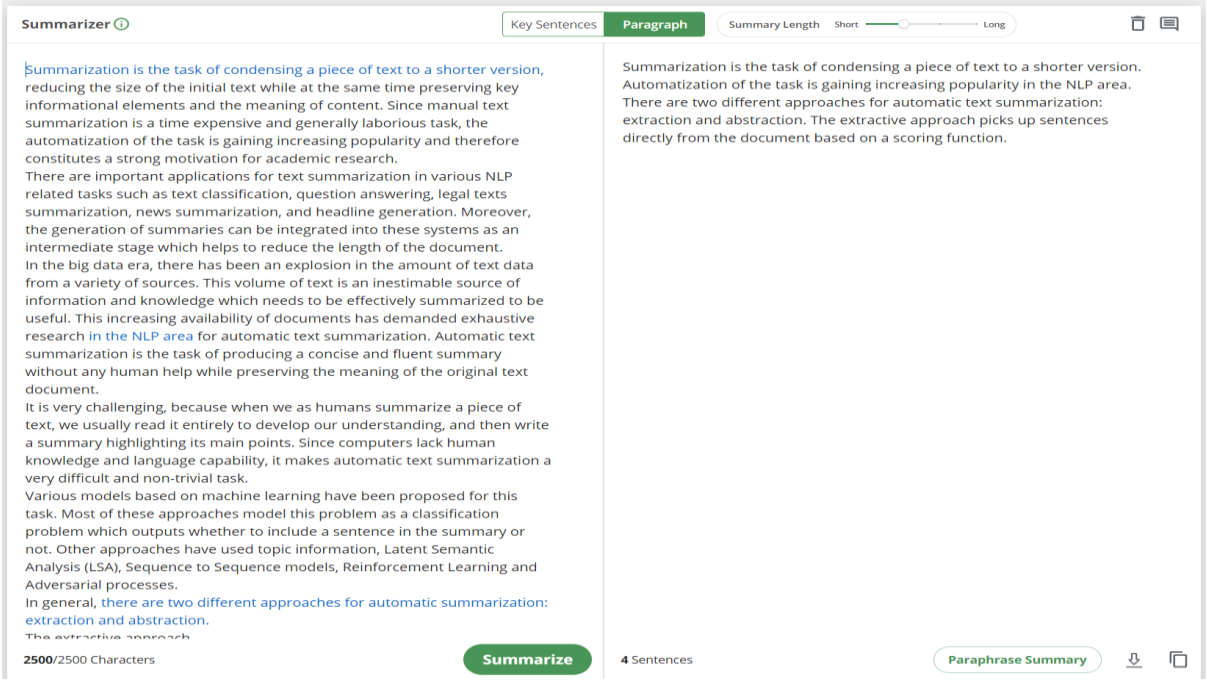


Source: Quillbot.com

Paragraph:

This mode summarizes the entire paragraph without losing the meaning of the natural sentences. It also gives the control to choose the summary length.

Figure 6: Paragraph:



Source: Quillbot.com

QuillBot's AI uses Natural language processing model to extract the important information from the paraphrase while retaining the original context. Quillbot also provide API which helps users to integrate into their own application. QuillBot API pricing are:

Figure 7: Quillbot's API Pricing

API Plans	Price/month	Characters
Pro	$20	100000 Monthly Quota
Ultra	$150	1000000 Monthly Quota
Mega	$1300	10000000 Monthly Quota

Source: Quillbot.cm

QuillBot's Modes:

QuillBot comes with seven different settings. All of these modalities are in charge of the paraphrasing and quality of the content generated.

Standard, Fluency, and Creative are the only modes available on the free plan. User must purchase the premium version to gain access to Creative+, Formal, Shorten, and Expand modes.

Standard: This is the default mode of Quillbot. It make changes in text without loosing the natural meaning.
Fluency: This mode makes the least changes into text and create new text that sounds more natural and grammatically correct. Moreover, this Fluency mode will keep the Word Flipper setting as lowest as possible.
Creative: It is a creative one. This mode focuses on more changes in the text and try to create a text which is bit different from the original.
Creative+: This is the enhanced version of Creative Mode. This mode does more intuitive changes by understanding certain things such as common sentences and return in a better way.
Formal: If users want to write in a business or academic context, then this mode is very useful for users. This mode modifies text in a way that appropriates for formal audience.
Shorten: If users want to reduce the length of the text, then this mode will be helpful for users. Because "Shorten Mode" will shorten the text without changing the meaning of the text.
Expand: If user want to increase the text length or overall word count than this mode will help. This mode will try to add more words in the text as possible to increase the word count.

Table 5: Sample Overview of QuillBot:

Products	**Features**	**Tech.**	**Plan & Price**		**Platform**
Paraphraser Grammar Checker Summarizer Citation generator	Synonyms Key Sentences Summary Length	Natural Language Processing	Free	• 5000 summarized charac. • 700 parapharaser character • sentence process at a time • synonymos options • 3 writing modes • 1 freeze words or phrase	• Website • Microsoft word plug in • Chrome Extenion • Google Doc Extention

				• Chrome and doc extention	
			Premium	• 25000 Summarizer character limit • 10000 Paraphraser character limit • 15 Sentences processed at once • 4 Synonyms options • 7 Writing modes • Unlimited Freeze Words and phrases • Compare Modes (only on Desktop) • See longest unchanged words • Google Chrome & Doc Extensions	

Source: Quillbot.com

Pros

- Simple and easy to use
- Affordable premium plans
- Forever Free plan is available
- Provides a developer API
- It is web-based tool
- Also, it offers Summarizer and grammar checker tool within a single platform.
- Word flipper and word freezing functionality offered by them.

Cons

- The free plan and premium plan both have characters limitations.

Blinkist:

Blinkist is a mobile software that condenses and summarizes the main points of the best nonfiction books into short, memorable words. It has 4,500-title library includes everything from nonfiction classics to popular guides and hot new releases. Each title is written carefully by highly qualified authors and made available to the user as a short text and audio title based on scientific results. In audio summarization there are team of editors and experts who summarize the words and then professional voice actors record the audio.

There are no solid proof of technology or the algorithm used by Blinkist for their book and audio summarization but as per the information shared by Blinkist website, they are more focused on editors and experts teams which leads us to the conclusion that they uses human summarization and also algorithmic summarization.

Blinkist does not provide Summarization tools to the users. It has a wide library of pre summarized book in the form of text and audio. They provide service on mobile Application and the website. Even though they provide only premium version but they also have a 7 days trial period. In the premium version they have 2 plans, One is monthly subscription, which cost 12.99 Euro and 2nd one is Annual subscription, which cost 79.99 Euro. Blinkist is currently one of the cheapest service provider in the market.

There are still some pros and cons with the services they provide, which are:

Pros of blinkist app:

- The book summaries are complete, which is one of the benefits of using the Blinkist app. Users don't just get a few key ideas or concepts from a book; users get them all.
- It's very simple to consume information and learn. The app is easy to use and navigate, and a child could learn how to use the main features, such as reading and listening, in minutes.
- Remembering facts from summaries is much easier with the use of highlighters. This is an underappreciated feature that can have a significant impact on how much information users remember, but users must use it to reap the benefits.
- Blinkist provides a seamless audio experience. Users press a button, the music starts playing, and users listen. It's ideal for learning during long workouts, work commutes, or while cooking or cleaning.

Cons. Of Blinkist App:

- The book is almost completely devoid of humor and most emotion. Some authors make it a point to include a summary that captures some of the book's unique spirit. Blinkist doesn't; it's more focused on facts.
- Reading a lot of summaries in a row can become tedious. That's mostly due to the lack of plot, which is why I only recommend reading 1-2 summaries per day.
- Some summaries are excessively brief in comparison to the book they represent. I've seen summaries with only three or four blinks, but the book itself was quite long. I'm sure they'll leave out a lot of important information.
- Individual books are not subjected to scrutiny. None of the books are reviewed, critiqued, ranked, or fact-checked by the Blinkist team. That means users're still in charge of these tasks.

Table 6: Sample Overview of Blinkist:

Product	Features	Technology	Price	Platforms
Book Summarization	Text Summarization	Not Found	12.99 Euro Monthly	Web Site
	Audio Summarization		79.99 Euro Annually	Mobile Application

Source: blinkist.com

Tools4Noobs:

Tools4Noobs is an India based text summarization tools. It is a user-friendly article summarizer with a variety of options. The Threshold function can be used to limit the number of phrases based on their significance or to condense the summary to a certain length. It is also possible to see or highlight the significant terms in the text from this location. User can either type in their own text or provide the software a URL to summarize.

Table 7: Sample Overview of Tools4Noobs

Products	Features	Technology	Plan & Price
Text Summarization	Threshold/No. Of Lines Minimum Sentence Length Minimum word Length Show Sentence Relevance Show Best Words Keywords highlighting Show Sentences	PHP Script	Free

Source: tools4noobs.com

Pros:

- There are no character limit
- Absolutely free, no charges.
- No registration required
- No Advertisements
- Lots of other tools available
- URL import available
- Custom Summary Size

Cons:

- No file import and export option.
- Not many languages available

Split Brain Summary Tools:

Split Brain Summary Tool is a useful software for quickly summarizing books and articles in a wide range of languages. Users can make a handful of sentences in one of thirty-nine languages on their article! The summarization ratio can also cause a difference in summaries. User can customize the density of paraphrase by changing it from 5% to 80%. There's also the option of using a URL instead of text. There is, however, no way to import a file or export the

result to PDF, DOC, or any other common format. The website is free of advertisements and offers a variety of additional educational resources.

Pros:

- No character limit
- No advertisements
- URL Import available
- Support multiple languages
- Custom Summary Ratio
- No registration required

Cons:

- No file import and export support

TextSummarization:

Text Summarization is offers by Mashape. Text Summarizing API is based on advanced Natural Language Processing and Machine Learning technologies, and it belongs to automatic text summarization. It may be used to summarize text from a URL or document provided by the user. It allows users to get the summarize text by clicking on Summarize now but user can decide only 1 option select number of sentences for the summarization. It does not provide and other custom settings. Alternatively, users can sign up for our Text Summarization API's free plan on Mashape. The Text Summarization API, which is based on the Mashape Platform, can be used in any environment that can make HTTP queries, such as Java/JVM/Android, Node.js, PHP, Python, Objective-C/iOS, Ruby, .NET.

Pros:

- No character limit
- Summary size vary in Number of sentences
- Works with URL
- No registration Required
- Simple Interface
- Provide API

Cons:

- Not much option available for the summarization
- No file import and export option
- Advertisements on the website

Table 8: Sample Overview of TextSummarization

Products	Features	Technology	Plan & Price
Text Summarization	No of Summarized Sentence	Natural Language Processing Machine Learning	Free

Source: Textsummarization.com

GetAbstract:

getAbstract is one of the oldest book summary apps, having been founded in 1999. So far, they hold the largest collection of summaries. They not only provide book summaries, but also summaries of instructive articles, reports, and videos such as TED presentations. While reading book summaries has its benefits and drawbacks, today we'll focus on choosing the finest book summary app for you if you're already convinced that book summaries apps can be a valuable addition to your learning arsenal.

getAbstract is also not open about their technologies like what kind of algorithms they are using for the book summarization and what new technologies they are migrating, whether they are providing human summarization and algorithmic summarization. Get abstract has a huge library of 20000 books of different genre where user have to spend roughly about 10 minutes on every summarization. They provide summarization in the form of audio and text. Get abstract also provide analysis in the summaries which is very sharp and insightful. One of the good things about Getabstract is they have channelized their books and users can select their interest of genre to find the summarized books easily. In the summary section they have provided several categories by which users will get all the information about the books and authors before reading the summaries, like:

- Recommendation: This section is about the book and what expectation users can set before reading.
- Takeaways: This section highlight the key takeaways from the books, like important sentences.
- Summary: this section gives the Summary of the books.
- About the author: They provide little information about the author after every summary, which helps interested users about the authors.
- More on this topic: This section gives similar topic or summaries.

- Audio Features: This is the section where users can listen to the summary. Most of the summary has audio feature.
- Rating, Liking, commenting: After every summary Rating option where users can give their feedback about the summary and the experience.

The online, iOS, and Android versions of getAbstract are also available. The app service is not free however it allows users for 3 days trial period. The app's interface is straightforward and simple to use. It provides everything for clarity and usability. They have two plans for the users:
The Starter plan: This is the starter pack where user gets 500 summaries which cost $99.
The Pro Plan: This plan comes with more than 20000 summaries for $299 or Euro per year.
They also have Enterprise plan and the student plan. In the enterprise plan they have every feature of other plans. Additionally, they have added Customized Enterprise portal, API Access, Custom SSO & LMS integration, Dedicated learning consultant, and advance reporting. The pricing tier for the enterprise depends on the company so every organization need to contact them before they purchase any service. In the student plan they have special offers, which are Student starter and student pro. In student starter plan they provide 5000 Summaries for free and in the student pro plan they provide more than 20000 summaries $59 or Euro Annually. They also have gift plan option where one can gift any plan, but they have to pay as per the plan.

5. Projection of Results:

5.1. Technical Analysis and Comparison

Above mentioned organization has one thing similar; they are in this market for at least 5 years. Although they are not quite open about their technologies and business models, but we can get an idea that most of them are using Natural language processing for the text summarization process. Some of them mentioned that they are using advance AI algorithm to get the most correct and suitable results and even they mentioned that they are using Blockchain technologies to improve the model accuracy and better results like SummarizeBot but they are not clear about how they are integrating blockchain technology with the Machine learning and AI algorithm to improve the accuracy. So, we tried to dig down about the process and we found out some interesting facts:
Machine learning and blockchain is two major discussed terms in today’s era. To get a good results and high accuracy machine learning model need a lot of data. The more data we train the high accuracy we will get. It is not possible to train few data sets in a model and based on

that accuracy we predict the results. Blockchain is the ledger of data. Blockchain technology is being used in distributing structure and weights of trained neural network which is based on tensor matrix theory. It is the process of huge replication in Edge AI computing unit production. Blockchain technology can be used to update multiple identical Neural Networks of swarm drones in order to adapt to new situations or increase accuracy, among other things. The same data set, on the other hand, can be distributed via blockchain to distributed Cloud AI or hierarchies of neural hypernetworks clusters for training and classifications. However, there are some drawbacks, like Blockchain is designed to be immutable, whatever information is entered becomes immutable. However, there is no mechanism in place to verify whether the information is correct. As a result, if there is incorrect information, it becomes immutable. I believe AI (Deep Learning, Neural Intelligence) can be used to determine the validity/correctness of data, which can then be sent to a blockchain application to make the information permanent. So, if we use both technologies, it will serve better.

To understand how machine learning and blockchain can work together and how can blockchain be used to approve the accuracy of the model we first need to understand how they work individually and how can we integrate them.

Working mechanism of Machine Learning:

Machine learning makes use of massive amounts of data to identify underlying patterns in datasets. This enables the creation of more efficient systems that can learn from their mistakes in the past. Speech recognition software, fitness tracking hardware, self-driving cars, and other machine learning systems are examples. When exposed to new information, machine learning software adapts. For example, if an email's spam detection software detects a spam email, subsequent communication from the same sender is likely to be classified as spam as well. This type of filtering is performed by machine learning, which is based on models developed using datasets with high-quality information that has not been tampered with.

Working Mechanism of Blockchain:

The Blockchain is the most secure database in the world, storing permanent, unalterable, time-stamped instances of data in a decentralized ledger. To assign digital ownership to entities on the blockchain, blockchain technology employs digital signatures and a consensus mechanism. As a result, a highly decentralized ledger that is highly resistant to censorship by governments or other malicious actors is created. As a result, blockchain is an excellent candidate for storing sensitive information that should not be altered in any way. Machine learning operates on the

"Garbage In, Garbage Out" principle. This means that if the data used to build a prediction model was tainted in any way, the resulting model would be useless as well.

Integrating Machine Learning with blockchain:

To build models for accurate prediction, machine learning relies on massive amounts of data. A significant portion of the cost of obtaining this data is incurred in collecting, organizing, and auditing the data for accuracy. This is an area where blockchain technology can make a significant difference. Data can be transferred directly and reliably from its point of origin using smart contracts. For example: A machine learning model for self-driving trucks, for example, would necessitate hundreds of Terabytes of actual truck driving data. Traditionally, all data such as driving speeds, fuel consumption, breaks, and so on would be collected using various trackers. It would then be sent to a processing facility, where auditors would sift through the data to ensure its authenticity before sending it to data scientists to be processed. Smart contracts, on the other hand, have the potential to significantly improve the entire process by utilizing digital signatures. We could program smart contracts to directly send data from the truck driver to data scientists who would use the data to build machine learning models by using blockchains to ensure the security and ownership of the collected data. This means that the combination of blockchain technology and machine learning is a game changer for self-driving research because it has the potential to create a marketplace for data for research.

Now that we have understood that how blockchain can be used to improve the model accuracy so we can say that SummarizeBot might be using the blockchain technology in the same way to improve their model. Rest of the organization is still using the natural language processing model for the summarization (unknown for getabstract and blinkst). So, we did a comparison among the data we found and tried to rank them according to their technology, services, accessibility and ease of access.

Table 9: Comparison of Different Organizations:

Organization	**Services**			**Offering Tools**	**Technology**	**Multi Language**	**Free/ Premium**
	Text Summarization	**Book Summarization**	**Audio Summarization**				
SummarizeBot	Yes	No	Yes	Yes	ML AI Blockchain	Yes	Premium

Smmry	Yes	Yes	No	Yes	NLP	No	Both
Resoomer	Yes	No	No	Yes	Not Found	Yes	Both
QuillBot	Yes	Yes	No	Yes	NLP	Yes	Both
Blinkist	No	Yes	Yes	No	Not found	Yes	Premium
Tools4Nobs	Yes	No	No	Yes	NLP	No	Free
SplitBrain	Yes	No	No	Yes	PHP Script	Yes	Free
Text Summarization	Yes	No	No	Yes	NLP ML	No	Free
GetAbstract	No	Yes	Yes	No	Not found	Yes	Premium

Source: Adapted from above mentioned data

According to the information we have gathered we can say that summarizeBot is in the top of the list for using latest technologies and improving their models (Personal Judgment), QuillBot and summarizeBot are providing lots of different types of services apart from the summarization which helps users to get everything in a single place. QuillBot has a good accessibility and ease of access with the tools, Resoomer and SplitBrain are providing lots of custom setting for summarization which helps users to get result in different ways.

5.2. Short Overview of Key findings:

After analyzing the information, we have found that there are only a few companies who are the global leading service provider for Automated Book Summaries. They are providing service Using new technologies like AI or Blockchain.

Although most of them are not transparent with their technology and the process they are using.

Almost all companies are providing summary service for different types of text but only Reesomer is open about that they are providing summary for argumentative text only.

No company are open about their training data or training method by which we can understand and have an idea of what kind of books or text will be good to summarize and have a good result by using their service.

5.3. Limitations:

- Although we have listed out a few companies for Automated Book Summaries, but there are not much information available information on their web sites about the technologies they used.
- We have also tried to contact a few companies but didn't get any response yet.
- Moreover, what we have understood from our market analysis is that, until and unless we become a customer we cannot get into details about their technology.

6. Conclusion and Outlook

Although we have tried to gather as much information as we can about the companies, but there are only one source which we can trust is company website. There are some reviews and key findings are available in other website but we cannot be 100% sure of their results. So, we tried to be within the boundaries as of now. The data we have gathered, from that we have understood that most of the companies are using Natural Language processing model and some of them are using the latest technology. So, there is a good scope that we can build a new model which will compare different algorithm and based on the accuracy we can get the final results. Moreover, in future we can gather more data and do analysis to find out more about different companies who are providing services and software for Text Summaries.

We are also hoping to contact the marketing department of the big companies like summarization bot and then get some information about it so that we can understand their technology and find out some flaws which will help us to build a more effective Automated Book Summaries.

7. Bibliography:

Agosti, Crestani, MelucciInformation Processing & Management332 133-144.Allan, James. "Building Hypertext Using Information Retrieval." Information Processing & Management 33, no. 2 (March, 1997): 145-159.

Berger, A. L., & Mittal, V. O. (2000, July). "OCELOT: a system for summarizing web pages." In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 144-151).Borko, H., and C. L. Bernier. Abstracting Concepts and Methods. San Diego: Academic Press, 1975.

Carbonell, J. G., and J. Goldstein. "The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries." In 21st International ACM SIGIR Conference Research and Development in Information Retrieval, 335-336. New York: ACM Press, 1998.

Edmundson, H. P. "New Methods in Automatic Extracting." Journal of the Association for Computing Machinery 16, no. 2 (1969): 264-285.

Feldman, Susan. "NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval." Online 23, no. 36 (May, 1999): 62-72.

Hahn, Udo, and Inderjeet Mani. "The Challenges of Automatic Summarization." Computer 33, no. 11 (November, 2000): 29-37.

Hovy, Eduard, and Chin-Yew Lin. "Automated Text Summarization in Summarist." In Advances in Automatic Text Summarization, eds. Inderjeet Mani and Mark Maybury, 81-94. Cambridge, Mass.: The MIT Press, 1999.

Kupiec, Julian, Jan Pedersen, and Francine Chen. "A Trainable Document Summarizer." 18th International ACM SIGIR Conference on Research and Development in Information Retrieval, 68-73. New York: ACM Press, 1995.

Luhn, H. P. "The Automatic Creation of Literature Abstracts." IBM Journal of Research and Development 2, no. 2 (1958): 159-165.

Mani, Inderjeet. Automatic Summarization. Amsterdam: John Benjamins Publishing Co., 2001.

Mani, Inderjeet, and Eric Bloedorn. "Summarizing Similarities and Differences among Related Documents." Information Retrieval, no. 1/2 (1999): 35-67.

Mani, Inderjeet, and Mark Maybury. Advances in Automatic Text Summarization. Cambridge, Mass.: The MIT Press, 1999.

Maybury, Mark. "News on Demand." Communications of the ACM 43, no. 2 (February 2000): 32-34.

Miller, G. "WordNet: A Lexical Database for English." Communications of the ACM 38, no. 11 (1995): 39-41.

Moens, Marie-Francine. Automatic Indexing and Abstracting of Document Texts. Boston: Kluwer Academic Publishers, 2000.

Mueller, Erik T. “Prospects for In-Depth Story Understanding by Computer.” Massachusetts Institute of Technology, 1999; last updated 1999, <http://xenia.media.mit.edu/~mueller/papers/storyund.html> (Date accessed: Jan. 23, 2003).

Myaeng, Sung Hyon, and Dong-Hyun Jang. “Development and Evaluation of a Statistically Based Document Summarization System.” In Advances in Automatic Text Summarization, eds. Inderjeet Mani and Mark Maybury, 61-70. Cambridge, Mass.: The MIT Press, 1999.

Paice, C.D. “Constructing Literature Abstracts by Computer: Techniques and Prospects.”Information Processing & Management 26, no. 1 (1990): 171-186.

Piller, Charles. “Five Reasons to Hope: New Technologies That May Help Silicon ValleyRise Again.” Los Angeles Times Magazine, March 9, 2003, 30-32.

Pinto, Maria, and C. Galvez. “Paradigms for Abstracting Systems.” Journal of InformationScience 25, no. 5 (1999): 365-380.

Radev, Dragomir R., Eduard Hovy, and Kathleen McKeown. “Introduction to the SpecialIssue on Summarization.” Computational Linguistics 28, no. 4 (2002): 399-408.

Salton, Gerard, Amit Singhal, Mandar Mitra, and Chris Buckley. “Automatic TextStructuring and Summarization.” Information Processing & Management 33, no. 2 (March 1997): 193-207.

Sparck Jones, Karen. “Automatic Summarizing: Factors and Directions.” In Advances in Automatic Text Summarization, eds. Inderjeet Mani and Mark Maybury, 2-12. Cambridge, Mass.: MIT Press, 1999.

Tas, Oguzhan, and Farzad Kiyani. "A survey automatic text summarization. " Press Academia Procedia 5, no. 1 (2007): 205-213.

Gambhir, Mahak, and Vishal Gupta. "Recent automatic text summarization techniques: a survey." Artificial Intelligence Review 47, no. 1 (2017): 1-66.

van Halteren, Hans. "New feature sets for summarization by sentence extraction." IEEE Intelligent Systems 18, no. 4 (2003): 34-42.

Edmundson, Harold P. "New methods in automatic extracting." Journal of the ACM (JACM) 16, no. 2(1969): 264-285.

Salton, Gerard, Amit Singhal, Mandar Mitra, and Chris Buckley. "Automatic text structuring and summarization." Information processing & management 33, no. 2 (1997): 193-207.

Allan, James. "Building hypertext using information retrieval." Information Processing & Management 33, no. 2 (1997): 145-159.

Kupiec, Julian, Jan Pedersen, and Francine Chen. "A trainable document summarizer." In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 68-73. 1995.

Eduard Hovy and Chin-Yew Lin, “Hovy, Eduard, and Chin-Yew Lin. "Automated text summarization in SUMMARIST." Advances in automatic text summarization 14 (1999): 81-94. in Advances in Automatic Text Summarization, in Mani and Maybury, 81-94.

Das, Dipanjan, and A. F. Martins. A survey on automatic text summarization. Literature survey for the course Language and Statistics II. Technical report, Carnegie Mellon University, 2007.vol. 4, pp. 192-195, 2007.

Nagwani, Naresh Kumar, and Shrish Verma. "A frequent term and semantic similarity based single document text summarization algorithm." International Journal of Computer Applications 17, no. 2 (2011): 36-40.

Munot, Nikita, and Sharvari S. Govilkar. "Comparative study of text summarization methods." *International Journal of Computer Applications* 102, no. 12 (2014).

Lin, Chin-Yew, and Eduard Hovy. "Automated multi-document summarization in neats." In Proceedings of the Human Language Technology Conference (HLT2002), pp. 23-27. San Diego, CA, USA, 2002.

Zhang, Haoyu, Jianjun Xu, and Ji Wang. "Pretraining-based natural language generation for text summarization." *arXiv preprint arXiv:1902.09243* (2019).

Nenkova, Ani, and Kathleen McKeown. "A survey of text summarization techniques." In Mining text data, pp. 43-76. Springer, Boston, MA, 2012.

Chowdhury, Gobinda G. "Natural language processing." Annual review of information science and technology 37, no. 1 (2003): 51-89.

Li, Liuqing, Jack Geissinger, William A. Ingram, and Edward A. Fox. "Teaching natural language processing through big data text summarization with problem-based learning." (2020).

Prudhvi, Kota, A. Bharath Chowdary, P. Subba Rami Reddy, and P. Lakshmi Prasanna. "Text Summarization Using Natural Language Processing." In Intelligent System Design, pp. 535-547. Springer, Singapore, 2021.

Gambhir, Mahak, and Vishal Gupta. "Recent automatic text summarization techniques: a survey." Artificial Intelligence Review 47, no. 1 (2017): 1-66.

Eduard Hovy and Chin Yew Lin, “Automated text summarization in SUMMARIST”, MIT Press, 1999, pages 81–94.

Aone, Chinatsu, Mary Ellen Okurowski, James Gorlinsky, and Bjornar Larsen. "A scalable summarization system using robust NLP." In *Intelligent Scalable Text Summarization*. 1997.

Bagga, A. and B. Baldwin. 1998. Entity-Based Cross-Document Cross-Referencing using the Vector Space Model. In Proceedings of COLING/ACL, 79-86. Montreal, Canada.

Lin, C. Y., and E. H. Hovy. "Automatic Text Categorization: A Concept-Based Approach." (1998).

Saranyamol, C. S., and L. Sindhu. "A survey on automatic text summarization." International Journal of Computer Science and Information Technologies 5, no. 6 (2014): 7889-7893.

Divya, S., and P. Reghuraj. "Eigenvector based approach for sentence ranking in news summarization." IJCLNLP, April (2014).

Saggion, Horacio, and Guy Lapalme. "Generating indicative-informative summaries with sumum." Computational linguistics 28, no. 4 (2002): 497-526.

Bellare, Kedar, Anish Das Sarma, Atish Das Sarma, Navneet Loiwal, Vaibhav Mehta, Ganesh Ramakrishnan, and Pushpak Bhattacharyya. "Generic Text Summarization Using WordNet." In LREC. 2004.

Silber, H. Gregory, and Kathleen F. McCoy. "Efficiently computed lexical chains as an intermediate representation for automatic text summarization." Computational Linguistics 28, no. 4 (2002): 487-496.

Lin, Chin-Yew, and F. J. Och. "Looking for a few good metrics: ROUGE and its evaluation." In Ntcir Workshop. 2004.

Melamed, I. Dan, Ryan Green, and Joseph Turian. "Precision and recall of machine translation." In Companion Volume of the Proceedings of HLT-NAACL 2003-Short Papers, pp. 61-63. 2003.

Over, Paul. "An introduction to DUC 2003: Intrinsic evaluation of generic news text summarization systems." In Proceedings of Document Understanding Conference 2003. 2003.

Radev, Dragomir, Simone Teufel, Horacio Saggion, Wai Lam, John Blitzer, Arda Celebi, Hong Qi, Elliott Drabek, and Danyu Liu. "Evaluation of text summarization in a cross-lingual information retrieval framework." Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, Tech. Rep 6 (2002).

Saggion, Horacio, Dragomir Radev, Simone Teufel, and Wai Lam. "Meta-evaluation of summaries in a cross-lingual environment using content-based metrics." In COLING 2002: The 19th International Conference on Computational Linguistics. 2002.