International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2494-9150 Vol-01, Issue 05, AUG 2015

# Survey on Automated Text Documents Summarization Tools

**[1]Mr. Sagar S. Yadav, [2]Prof. Sachin Bojewar**
*[1]Student M. E.(I.T), Vidyalankar Institute of Technology, Mumbai University, Mumbai, India.*
*[2]Associate Professor of InformationTechnology, Vidyalankar Institute of Technology, Mumbai University, Mumbai, India.*

**ABSTRACT— Text mining has become an important research field as it tries to discover valuable information from unstructured and large amount of texts. It becomes very difficult to get the relevant Information from the Unstructured and large amount of Single and Multiple Text Documents. Text Mining is an important task of Text Summarization. Automated Text Summarization is the Process of reducing the Original size of document without changing the overall meaning of the Text and achieving the relevant Information from the text documents. The goal of the Automated Text Summarization is to minimize the User's time for reading and understanding the document without disturbing the User's area of Interest. The Information Overload Problem can be easily overcome.**

***Index Terms—Information Overload, Text Documents, Text Mining, Text Summarization, Unstructured Information.***

## I. INTRODUCTION

Text Documents Summarization is gaining much importance nowadays. Recently due to the enormous growth in information on Web and Offline, need for automatic text Summarization has increased. Hence, it is usual that users trying to retrieve the multiple textual documents or information face the problem of responses of hundreds or thousands of retrieved documents on a particular topic or different topics on user's area of interest.

Also, retrieved documents have most of the redundant information. Hence, it is not easy for users to manually summarize the large number of Single or multiple documents to get related information on a particular Topic or different topics as per the need. So it is desirable to have a system that could summarize these multiple text documents. Text documents Summarization satisfies this users need by summarizing the text documents.

Also the information available on the Internet as well as Offline information to be searched on particular topic or the different topics in various Applications has been increasing to a greater extent, it has become progressively more essential to provide enhanced mechanisms to find and represent textual information effectively and efficiently. The vast amount of information available today has lead to information overload problem. Document summarization is one feasible key to handle this information overload problem. Summarization is an important activity in the analysis of a huge volume of text documents. The purpose of the text summarization is to present the main idea in a document in less space so that user will not have to waste time in reading the whole document. Text summarization is nothing, but a text that is produced from one or more documents that convey the user all the important information in the original text.

The objective and approach of summarization of documents explain the kind of summary that is generated. The summary generated reveals the salient and the relevant information of its documents. This is so because of huge growth of information on internet has led to the use of Information Retrieval systems which work on search engines.

This paper presents an idea on Automated Text Documents Summarization Tools. Summarization can be done on both the Single as well on Multiple Text Documents through Automatic Text Summarization.

## II. Text Summarization Automated Tools

Various Text Mining and Text Summarization Automated Single and Multi Document Summarizer Tools are available over the Internet with the paid User Subscription and some of them are available for Free for a stipulated time period. The Single Document Summarizer Tools are:

***1. GreatSummary***

Great Summary is a useful website which allows you to summarize articles right from your browsers. Online tool to summarize articles for free.There are two options to choose

 

from: Document Summarization or Great Summary Subscriptions. The latter creates a news feed for a webpage to track any changes made in it. For example making one for some Wikipedia page to know if some new information was added or altered. The first option allows you to put any web-page link in the URL box, and choose in how many lines you need the summary. Click "Summarize" and you will get a brief paragraph on the article, with only the main points.

**2. Ultimate Research Analyst**

The **Ultimate Research Assistant** is an Internet Search Engine and research tool using Text Mining text and multi-document summarization techniques to generate a hypertext summary of relevant content contained in the underlying documents returned by the search engine. It is based on research conducted by Andy Hoskinson and published in the IEEE Computer journal in November 2005. Essentially, the Ultimate Research Assistant tool automates the process of analyzing search results, extracting and compiling relevant data, performing related searches, and organizing the knowledge contained in the dozens of relevant pages the typical search engine returns.

The Ultimate Research Analyst is available for Free upto limited Sources or Applications and after its limit User needs to Sign Up for the Service available with a payment of mode.

The Ultimate Research Assistant works in the following way:

1.Return search results from Google, Yahoo, or other search engine for a given search term/topic.

2.Mine "concepts" (keyphrases) and associated text summaries from the underlying search result documents using an information extraction engine.

Organize the concepts and associated text summaries into a taxonomy using a clustering engine; Generate a "hypertext knowledge base" (essentially, a multi-document summary) from the taxonomy. This knowledge base includes several visualizations including tag clouds, histograms showing concept popularity and a mind map view.

**3. SMMRY** (pronounced *SMUR-EE*) was created in 2009 by Amir Elmaani is an Online Free Document Summarizer. Its API are also available for the Use. SMMRY is accessible by an API. Developers are able to implement SMMRY into applications that may require a summary of a webpage or variable. Developers need to generate an API key to access this feature. SMMRY's mission is to provide an efficient manner of understanding text, which is done primarily by reducing the text to only the most important sentences.

SMMRY accomplishes its mission by:

• Ranking sentences by importance using he core algorithm.
• Reorganizing the summary to focus on a topic; by selection of a keyword.
• Removing transition phrases.
• Removing unnecessary clauses.
• Removing excessive examples.

**4. Free Summarizer**

Free Summarizer allows you to copy a passage of text and paste it in the text box. You can then choose the number of sentences in which the summary should be made, and input your email to which the summary will be emailed. This is an Online Free Single document Summarizer Tool available.

**5. Topic Marks**

TopicMarks is one of the most advanced summarizing tools found on the Internet. It makes use of a complex algorithm to summarize text documents and display only the most important points. To make the most effective use of TopicMarks, uploads all files that are linked together in order to get the best summary. TopicMarks also allows you to upload or use files from Evernote and Google Docs and Google Reader. However, to make use of the service, you first need to make an account on the website. Once you have an account, you can summarize and save summaries on their cloud-managed servers.

**6. Wiki Summarizer**

Wiki Summarizer is a website specially made for summarizing articles found on Wikipedia. You can search Wikipedia right from the website and choose the article you wish to summarize. Or, you can enter the keywords or link to the article you want to summarize and get the most vital points extracted from the article and displayed to you. The user-friendly design really makes it easy for anyone to use the service.

## III. Five Existing Multi Document Summarization Systems presented as pilot ones

We reviewed these systems in addition to other systems related to the same domain.

**1. MEAD** is a News tracking and summarizing system based on a Centroid-based approach. It works by clustering documents describing an event. Clusters are chosen to include two or more documents ordered by time. The system then constructs cluster centroids consisting of words best describing that cluster. Each sentence within the cluster is scored based on its similarity of centroid words, and n sentences are selected to best represent the cluster. The process is repeated for each cluster, and the extracted sentences are ordered based on news time stamps to constitute the summary. In other work, MEAD was extended to allow the system to collect clusters either by user input keywords, or keywords extracted from example news. El_Haj et. Al extended the MEAD concept to include the Arabic language. They used different parameter settings for

 

extracting sentences including the cluster size and the selection model.

**2. CIASSY** is an event-focused, query-based multi-document summarization. It summarizes Arabic documents by translating documents into English to extract the summary. The system uses language dependent trimming rules to focus on important parts of the sentences. Sentences are then ranked by their inclusion of "signature words" which occurs significantly more than expected in the document.

**3. MMR-MD** is a query-relevant MDS, in which each passage (sentence) is ranked by summing its cosine similarity to user input query, its similarity to other passages within cluster, its inclusion of specific word types (such as named entities) and its time stamp sequence; the latest of which is preferred. To achieve diversity, the ranking score of a passage is penalized by its cosine similarity to previously selected passages. It is penalized also if it is part of previously selected clusters and documents.

**4. Lexical Chain** method disambiguates word senses using shallow syntactic analysis, part-of speech tagger, and WorldNet thesaurus. The system organizes semantically related noun words as lexical chains. Sentences corresponding to the strongest lexical chains are then extracted to construct the summary.

**5. CIST**: Liu et. Al introduces an extractive multi-document summarization method for 7 languages including Arabic. The method constructs a hierarchical tree structure of candidate sentences based on hierarchical topic model of hierarchical Latent Dirichlet Allocation (hLDA). Each sentence is represented by a path in the tree, and each path can be shared by many sentences. The assumption is that sentences sharing the same path should be more similar to each other because they share the same topics. Also, if a path includes a title sentence, then candidate sentences on that path are more likely to be included in the generated summary. The system extracts the sentences with the highest scores to include into the summary.

## IV. Some of the Software available for Automated Text Summarization

**1. Intellexer Summarizer:**

Intellexer Summarizer Pro is a professional desktop application for high speed text summarization. It ensures outstanding quality of summaries and work process improvement.

Requirements:

License**:** Shareware

Category: File and Disk

Developer: Effective Soft Ltd.

Downloads: 269

Price: $399.95

A Traial Version is available at the Intellexer Summarizer Webpage .

**2. NATURAL LANGUAGE PROCESSING V.4**

LJParser Nature Language Processing Introduction: LJParser is a developing platform for web search and mining. It is a middleware by LING-JOIN Software, which is well known for over ten years of expertise in natural language understanding and web search. LJParser Nature Language Processing software provides powerful modules including precise search for multiple language, new words detection, text summarization, keywords extraction and more which are detailed below Features:

-Chinese New Words Detection Identify the new words and phrases which are not listed in the dictionary.

-Chinese Lexical Analysis Make Chinese word segmentation, part of speech tagging and recognition of Chinese named entities, including persons, locations and organizations.

- Text Clustering Finding a structure in a collection of unlabeled texts.

- Text classification and filtering Dealing with information classification, resume classification, mail classification, office document classification.

-Text Summarization Automatically summarizing a text by computers.

- Keywords Identification Extract the feature word on behalf of the essence of the article.

**2.Miraplacid Text**

Miraplacid Text Driver extracts text from any printable documents. Accounting professionals, Medical Insurance companies, Health care providers and many others use it to extract text from all kind of document formats they receive and import it into their systems.

License: Shareware

Category: Document Management

Developer: Miraplacid

Trial length:15 Day Trial

Downloads**:**249

Size:5.25 Mb

Price: $49.95

**3.Essential Summarizer**

This is a Business Multilingual Text Summarization Software, Essential Summarizer, offers a more efficient method of text reading.

Your business requires the reading of a considerable number of documents: we have the solution to reduce up to 80% of your time reading and get the gist of the content. It is very easy to use, the software produces a summary of the text of your documents and web pages as well as a synthesis of the entire contents of a folder on your computer with just one click.

 

Also, through the "Search to Summarize" functionality you can have real-time summaries of your search results on top search engines like Google, Bing, Yahoo and Baidu as well as Wikipedia. Through years of research, the method developed by Mining Essential can reproduce a summary consistent with the original meaning through a sophisticated linguistic analysis of text. Efficient both in local and foreign languages, Essential Summarizer is associated with a language translation module. You will get the summary of the content in 21 different languages. Free trial is available directly online. Further Services are available on Paid basis.

## V. Some of the API's available for Text Summarization

| | | |
|---|---|---|
| 1. | Stremor Automated Summary and Abstract Generator | Language Heuristics goes a step beyond Natural Language Processing to extract intent from text. Summaries are created through extraction, but maintain readability by keeping sentence dependencies intact |
| 2. | Text-Processing | Sentiment analysis, stemming and lemmatization, part-of-speech tagging and chunking, phrase extraction and named entity recognition |
| 3 | Text Processing | The WebKnox text processing API lets you process (natural) language texts. You can detect the textâ€™s language, the quality of the writing, find entity mentions, tag part-of-speech, extract dates, extract locations, or determine the sentiment of the text. |
| 4 | nlpTools | Text processing framework to analyse Natural Language. It is especially focused on text classification and sentiment analysis of online news media (general-purpose, multiple topics). |
| 5 | DuckDuckGo | DuckDuckGo Zero-click Info includes topic summaries, categories, disambiguation, official sites, !bang redirects, definitions and more. You can use this API for many things, e.g. define people, places, things, words and concepts; provides direct links to other services (via !bang syntax); list related topics |
| 6 | Textalytics Text Classification | Automatic multilingual text classification according to pre-established categories defined in a model. The algorithm used combines statistic classification with rule-based filtering, which allows to obtain a high degree of precision for very different environments. |

## VI. Summary of Automated Text summarization Tools / Software Availability and their features

| Tools/Software | Features | Availability |
|---|---|---|
| 1.Great Summary | Summarizes to get a brief Paragraph on the article, with only the main features. | Online Open Source Freely available Single Document Summarizer Tool |
| 2.Ultimate Research analyst | An Internet Search engine and Research Tool to generate a Summary widely used in News Media | Online Free Single Document Summarizer Tool available freely upto limited access for further services requires Payment of mode |
| 3.SMMRY | Accesible by an API. Efficient Summary is achieved by generating an API key. | Online free Single Document Summarizer Tool. |
| 4.Free Summarizer | Number of sentences should be choosen by the User for generating the necessary number of sentences for Summarization. | Single Document Free Summarizer Tool. |
| 5.Topic Marks | Most Advanced Summarizing Tool found on the Internet; makes use of complex algorithm to Summarize Text Documents and display the only Important sentences. | Need to make an account on the Website. Limited only for a free trial. |
| 6.Wiki Summarizer | Summarizes articles found on Wiki. Single Document Summarizer. | Open Source Freely available. |
| 7.MEAD | News Tracking and Summarizing Multi Document Summarizing System based on Centroid based approach of Clustering of Documents. | Freely available upto limited access. |
| 8.CLASSY | Event focussed MDS, query based. Summarizes Arabic documents by translating the documents into English to extract the Summary. | Limited for Free Trial Version. |
| 9.MMR-MD | Query relevant MDS. | Limited for Free Trial Version. |
| 10.Lexical Chain | MDS Uses Syntactic Analysis. The System organizes semantically related Noun words as lexical chains. | Needs to sign up on the website .Free Trail version is available upto limited access. |
| 11.Cist | Extractive MDS for languages including Arabic. Constructs Hierarchical Tree Structure Model. | Requires payment of mode upto 50 $ can be downloaded from the website cist.com . |
| 12.Intellexer Summarizer | Professional software for high speed Text summarization | Licensed Version of Shareware available for $400.Developed by Effective Software Ltd. |
| 13.NLP V.4 | Professional Software provides keyword extraction, summarization. | Available for Free Trial version. Software can be downloaded from NLP.com at $100. |
| 14.Mirplacid Text | Text driver extracting Software used by Medical, health and Insurance Companies. | Licensed version of Shareware developed by Mirplacid for $50. |
| 15.Essential Summarizer | Produces real Time Summaries through sophisticated linguistic analysis of text. | Free Trail version available at the website. Further requires payment of mode at $175 from the Website. |

Thus these are the different Online Tools, Software and API's available for Automated Text Document Summarization.

## VII. CONCLUSION

We know that information Technology offers Large amount of data to Users when Users want to search information of Data. When User wants to search information that presents on Internet or in database, the Resultant data may be huge that is in turn quite large and may miss track the User from what they actually want. Thus there is a need to have such a System which compresses the huge information and prepares the Summarized Result without changing the overall Objective of User's Research.

 

In this Paper We presented a brief Summary on automated Text summarization Tools which generates Text Summary of meaningful sentences from the Input documents. Automated Text Summarization Tools explained above proved to be significant to generate accurate, relevant and efficient Text summary.

Thus instead of spending extra time on reading large amount of related data the Text Summarizer performs well in generating effective Summary for the Input documents.

## VIII. REFERENCES

[01] IEEE Transactions on Knowledge and Data Engineering,Vol.24,No.1,January 2012 Tscan: A Content Anatomy approach to Temporal Topic Summarization Chien Chin Chen and Meng Chang Chen.

[02] IEEE/ACM Transactions on Audio, Speech, and Language Processing,Vol.22,No.12,December 2014 SRRank: Leveraging Semantic Roles for Extractive Multi Document Summarization by Su Yan and Xiaojun Wan.

[03] IEEE Transactions on Knowledge and Data Engineering,Vol.25,No.8,August 2013 A Context-Based Word Indexing Model for Document Summarization by Pawan Goyal, Laxmidhar Behera, Senior Member, IEEE, and Thomas Martin McGinnity, Senior Member, IEEE.

[04] IEEE/ACM Transactions on Audio, Speech, and Language Processing,Vol.21,No.7,July 2013 Ranking Through Clustering: An Integrated Approach to Multi-Document Summarization by Xiaoyan Cai and Wenjie Li.

[05]IEEE Transactions on Fuzzy Systems 1063-6706 (c) 2013 IEEE. Using data merging techniques for generating multi-document summarizations by Daan Van Britsom, Antoon Bronselaer, Department of Telecommunications and Information Processing, Ghent University Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium.

[06] International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 6, December 2013 Multi-Topic Multi-Document Summarizer by Fatma El-Ghannam**1** and Tarek El-Shishtawy**2.**

[07] Association of deep learning algorithm with Fuzzy Logic for Multi document text summarization by G.P.ADMAPRIYA Journal of Theoretical and applied IT 10th April 2014 Vol.62 NO.1.

[08] TECHNIA – International Journal of Computing Science and Communication Technologies, VOL. 2, NO. 1, July 2009. (ISSN 0974-3375) Sentence Clustering-based Summarization of Multiple Text Documents by Kamal Sarkar.

[09] International Science Conferences, ACM, Jgateplus, ebsco, ijit libraries.

[10] A Hybrid Approach for Extractive Document Summarization Using Machine Learning and Clustering Technique. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014

[11]A Survey on Automated Text Summarization and Mining Methodologies. International Journal of Scientific Research Engineering &Technology (IJSRET) Volume 2 Issue 8 pp 512-517 November 2013 www.ijsret.org ISSN 2278 –0882 IJSRET @ 2013K.Divya Computer Science Engineering, Medicaps Institute of Technology and Management, Indore.

[12]Han J. and M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann publishers, 2nd Edition.

 