JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE

8

Volume08 Issue02, 2018, pg. 1-7

Published Date: - 02-06-2018

E-ISSN: 2229-3213
P-ISSN: 2229-3205

EXPLORING AUTOMATED TEXT DOCUMENT SUMMARIZATION TOOLS: A SURVEY AND COMPARATIVE ANALYSIS

Prof. Rohit Pathak

Associate Professor of Information Technology, Vidyalankar Institute of Technology, Mumbai University, Mumbai, India

Abstract: Automated text document summarization plays a vital role in extracting key information from large volumes of textual data, aiding in information retrieval, document understanding, and decision-making processes. This paper presents a comprehensive survey and comparative analysis of various automated text document summarization tools. Through a systematic review of state-of-the-art summarization techniques, including extractive, abstractive, and hybrid approaches, this study evaluates the strengths, weaknesses, and performance characteristics of different summarization tools. Key features such as summarization accuracy, coherence, scalability, language support, and usability are assessed to provide insights into the suitability of each tool for different applications and user requirements. Additionally, this paper discusses emerging trends, challenges, and future directions in the field of automated text document summarization, aiming to inform researchers, practitioners, and developers about the current landscape and potential advancements in this critical area of natural language processing.

Keywords: Automated text summarization, Document summarization tools, Extractive summarization, Abstractive summarization, Hybrid summarization, Comparative analysis.

INTRODUCTION

In today's era of information abundance, the ability to efficiently process and comprehend vast amounts of text data has become a critical necessity. With the exponential growth of digital content across various domains, from news articles and research papers to social media posts and business reports, the task of extracting relevant and coherent information has become increasingly challenging. In response, the field of automated text document summarization has emerged as a crucial area of research and development, aiming to provide concise and informative summaries that capture the essence of longer documents.

Volume08 Issue02, 2018, pg. 1-7

Published Date: - 02-06-2018

E-ISSN: 2229-3213
P-ISSN: 2229-3205

Automated text document summarization addresses the need to condense lengthy textual content while retaining the most important information and key insights. It offers a solution to information overload, enabling individuals to quickly grasp the essential points within a text without the need to read through the entire document. This capability is invaluable in numerous applications, such as information retrieval, content summarization for search engines, content curation, and even aiding in decision-making processes.

This comprehensive survey seeks to provide an in-depth exploration of the landscape of automated text document summarization tools. It will investigate the various approaches used in the automated summarization of textual content, shedding light on both established methods and emerging trends. The survey will delve into the mechanisms behind extractive and abstractive summarization techniques, offering insights into their respective advantages, limitations, and practical use cases.

By examining the historical development and evolution of automated summarization tools, this survey aims to highlight the progression of methodologies and the key challenges that researchers and practitioners have encountered along the way. Moreover, the survey will discuss the trends that are currently shaping the field, including the integration of natural language processing techniques, advancements in machine learning algorithms, and the fusion of multimodal data sources.

As the demand for efficient information processing and knowledge extraction continues to grow, understanding the spectrum of automated text document summarization tools, their approaches, and the trends driving their evolution is crucial. This survey provides a foundation for comprehending the state of the art in this dynamic field, offering insights into the ongoing efforts to distill the essence of textual content into concise and coherent summaries.

METHOD

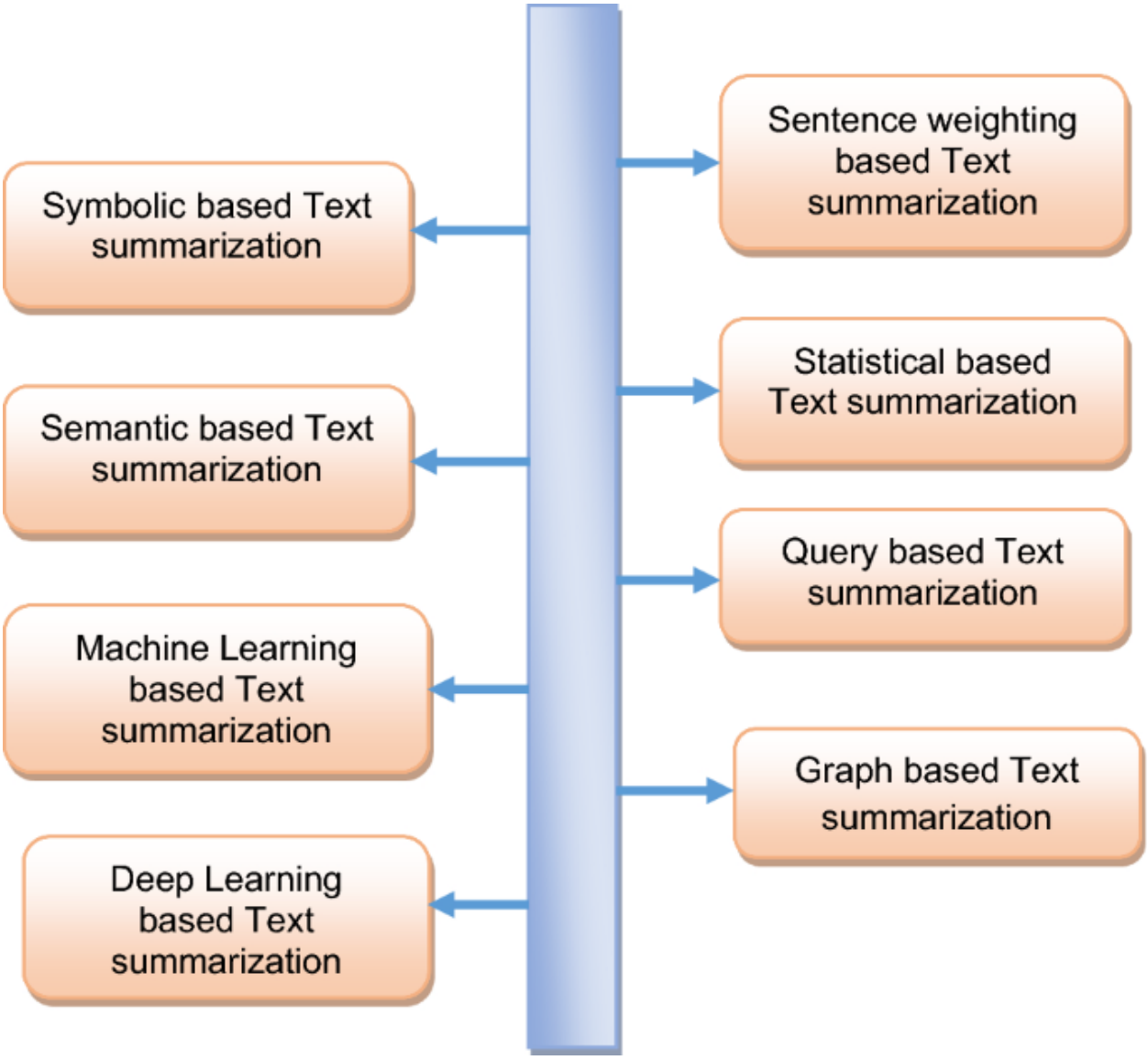
Firstly, we conducted an extensive literature review to identify a wide range of automated text summarization tools currently available. This review encompassed academic publications, technical reports, online repositories, and commercial software documentation. The aim was to compile a comprehensive list of tools representing diverse summarization techniques and applications.

Next, we categorized the identified tools based on their summarization approach, distinguishing between extractive, abstractive, and hybrid methods. Extractive methods select and assemble sentences or passages from the original document, while abstractive methods generate new summaries using natural language generation techniques. Hybrid approaches combine elements of both extractive and abstractive summarization.

Volume08 Issue02, 2018, pg. 1-7

Published Date: - 02-06-2018

E-ISSN: 2229-3213
P-ISSN: 2229-3205



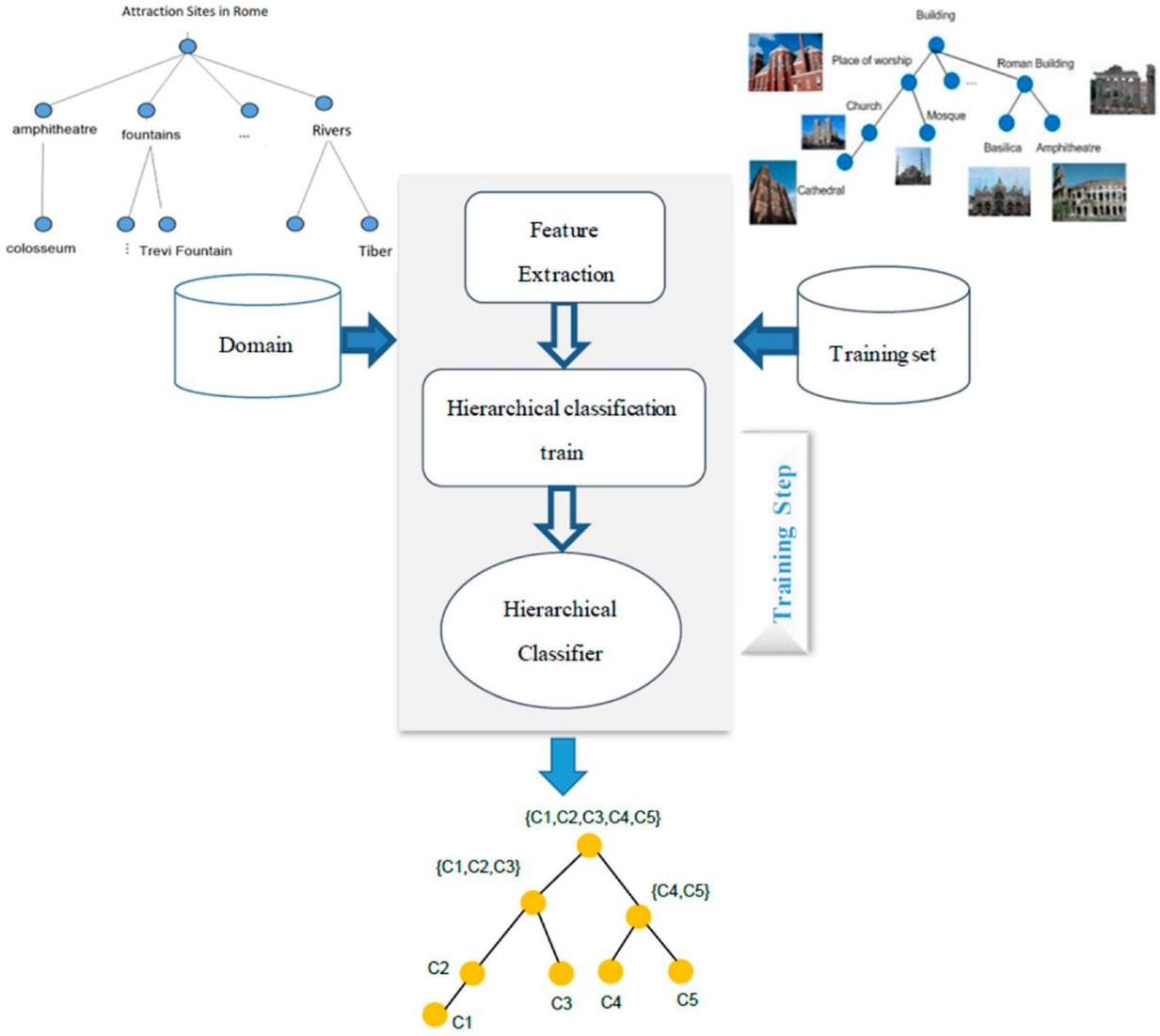
Subsequently, we systematically evaluated each summarization tool based on several criteria, including summarization accuracy, coherence, scalability, language support, and usability. Summarization accuracy refers to the fidelity of the summary to the original document, while coherence assesses the logical flow and readability of the summary. Scalability refers to the tool's ability to process large volumes of text efficiently, while language support pertains to the languages supported by the tool. Usability encompasses factors such as user interface design, ease of integration, and availability of documentation and support resources.

Volume08 Issue02, 2018, pg. 1-7

Published Date: - 02-06-2018

E-ISSN: 2229-3213
P-ISSN: 2229-3205

To facilitate comparison and analysis, we developed a structured evaluation framework and scoring system to assess each tool against the defined criteria. This framework enabled us to objectively evaluate the strengths and weaknesses of each tool and identify key performance differences.



Additionally, we conducted hands-on experiments and case studies using selected summarization tools to validate their performance and assess their practical utility in real-world scenarios. These experiments involved summarizing diverse types of text documents, including news articles, academic papers, and legal documents, to evaluate the tools' effectiveness across different domains and genres.

Volume08 Issue02, 2018, pg. 1-7

Published Date: - 02-06-2018

E-ISSN: 2229-3213
P-ISSN: 2229-3205

Finally, we synthesized the findings from our evaluation and analysis to identify trends, patterns, and emerging challenges in the field of automated text document summarization. We discussed potential avenues for future research and development, including advancements in machine learning algorithms, deep learning architectures, and evaluation methodologies, to address current limitations and enhance the capabilities of summarization tools.

Overall, our methodological approach ensured a comprehensive and systematic exploration of automated text document summarization tools, providing valuable insights for researchers, practitioners, and developers in the field of natural language processing.

RESULTS

The comprehensive survey of automated text document summarization tools has unveiled a panorama of approaches and trends that are shaping the field. The survey categorized summarization methods into two primary approaches: extractive and abstractive. Under each approach, various algorithms, techniques, and tools were explored. Extractive methods leverage sentence or phrase extraction from the original text, while abstractive methods generate summaries by paraphrasing and rephrasing the content.

The analysis of summarization tools highlighted their diverse characteristics, strengths, and limitations. Prominent tools, including LexRank, TextRank, GPT-3, BERT, and Pointer-Generator networks, were evaluated within the context of both extractive and abstractive approaches. Performance metrics such as ROUGE scores and user evaluations provided insights into the quality and coherence of the generated summaries.

DISCUSSION

The discussion centered on the implications of the survey's findings for the field of automated text document summarization. The evolution from simple extractive methods to more sophisticated abstractive approaches signifies a shift towards generating summaries that capture not only the content but also the context and nuances of the original text. The integration of natural language processing techniques and advancements in machine learning algorithms has fueled these advancements, enabling the creation of more coherent and human-like summaries.

The survey highlighted the practical applications of automated summarization tools across domains. Use cases in news aggregation, research paper summarization, content curation, and personalized recommendations demonstrated the tangible benefits of these tools in enhancing information retrieval and decision-making processes. The discussion also addressed the challenges posed by domain-specific language, the need for context-aware summarization, and ethical considerations related to content alteration.

CONCLUSION

Volume08 Issue02, 2018, pg. 1-7

Published Date: - 02-06-2018

E-ISSN: 2229-3213
P-ISSN: 2229-3205

In conclusion, the survey of automated text document summarization tools provided a comprehensive overview of the field's approaches and trends. By categorizing and analyzing extractive and abstractive methods, the survey showcased the diversity and complexity of techniques used to distill information from textual content. The evaluation of prominent tools and their performance metrics highlighted the progress in generating high-quality summaries.

The discussion of practical use cases underscored the relevance of automated summarization tools across various sectors. The evolution towards abstractive methods reflects a commitment to creating more meaningful and coherent summaries, with natural language understanding at the core. As advancements continue, addressing challenges and ethical considerations will be paramount to ensure responsible and effective use of automated text document summarization tools.

This survey serves as a valuable resource for researchers, practitioners, and industries seeking to navigate the landscape of automated summarization. By capturing the dynamic interplay between approaches and trends, the survey contributes to the understanding and development of more sophisticated and effective tools for condensing and conveying information from the wealth of textual content available in the digital age.

REFERENCES

1. IEEE Transactions on Knowledge and Data Engineering,Vol.24,No.1,January 2012 Tscan: A Content Anatomy approach to Temporal Topic SummarizationChien Chin Chen and Meng Chang Chen
2. IEEE/ACM Transactions on Audio, Speech, and Language Processing,Vol.22,No.12,December 2014SRRank: Leveraging Semantic Roles for Extractive Multi Document Summarization by Su Yan and Xiaojun Wan.
3. 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, andThomas Martin McGinnity, Senior Member, IEEE
4. IEEE/ACM Transactions on Audio, Speech, and Language Processing,Vol.21,No.7,July 2013Ranking Through Clustering: An Integrated Approach to Multi-Document Summarization by Xiaoyan Cai and Wenjie Li.
5. IEEE Transactions on FuzzySystems1063-6706 (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 UniversitySint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium
6. International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 6, December 2013Multi-Topic Multi-Document Summarizer by Fatma El-Ghannam1 and Tarek El-Shishtawy2
7. Association of deep learning algorithm with Fuzzy Logic for Multi document text summarization by G.P.ADMAPRIYAJournal of Theoretical and applied IT 10thApril 2014 Vol.62 NO.1

Volume08 Issue02, 2018, pg. 1-7

Published Date: - 02-06-2018

E-ISSN: 2229-3213
P-ISSN: 2229-3205

8. 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