The American Journal of Applied sciences
(ISSN – 2689-0992)
VOLUME 05 ISSUE 12 Pages: 1-5
SJIF IMPACT FACTOR (2020: 5. 276) (2021: 5. 634) (2022: 6. 176) (2023: 7. 361)
OCLC – 1121105553
Publisher: The USA Journals

Journal Website: https://theamericanjournals.com/index.php/tajas

Research Article

SURVEY OF AUTOMATED TEXT DOCUMENT SUMMARIZATION TOOLS: APPROACHES AND TRENDS

Submission Date: November 22, 2023, **Accepted Date:** November 26, 2023,
Published Date: December 01, 2023 |
Crossref doi: https://doi.org/10.37547/tajas/Volume05Issue12-01

Prof. Ravi Borde
Associate Professor of Information Technology, Vidyalankar Institute of Technology, Mumbai University, Mumbai, India

ABSTRACT

This survey presents an extensive exploration of automated text document summarization tools, focusing on the diverse approaches and emerging trends in this field. With the proliferation of digital information, the need to extract key insights from large volumes of textual content has become increasingly vital. This study surveys various methods employed in automated summarization, including extractive and abstractive techniques, along with their strengths, limitations, and real-world applications. By analyzing the evolution of these tools, the survey highlights the current trends, challenges, and future directions in automated text document summarization.

KEYWORDS

Automated text summarization, document summarization tools, extractive summarization, abstractive summarization, natural language processing, information extraction, text mining, machine learning, trends, challenges.

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

The American Journal of Applied sciences
(ISSN – 2689-0992)
VOLUME 05 ISSUE 12 Pages: 1-5
SJIF IMPACT FACTOR (2020: 5. 276) (2021: 5. 634) (2022: 6. 176) (2023: 7. 361)
OCLC – 1121105553
Publisher: The USA Journals

development, aiming to provide concise and informative summaries that capture the essence of longer documents.

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

Literature Review:

Conduct an extensive review of scholarly articles, research papers, conference proceedings, and relevant books on automated text document summarization tools.

Identify key concepts, approaches, and trends in the field of text summarization.

Classification of Summarization Approaches:

Categorize summarization methods into two main approaches: extractive and abstractive.

Subdivide each approach into subcategories based on algorithms, techniques, and underlying principles.

Data Collection:

Collect a diverse range of example summaries produced by various automated text document summarization tools.

Include summaries from different domains such as news articles, academic papers, social media content, and legal documents.

Analysis of Summarization Tools:

Analyze the characteristics, strengths, and limitations of prominent automated text document summarization tools within each approach.

The American Journal of Applied sciences
(ISSN – 2689-0992)
VOLUME 05 ISSUE 12 Pages: 1-5
SJIF IMPACT FACTOR (2020: 5. 276) (2021: 5. 634) (2022: 6. 176) (2023: 7. 361)
OCLC – 1121105553
Publisher: The USA Journals

Investigate their performance metrics, such as ROUGE scores, F1 scores, and user evaluations.

Identification of Trends:

Identify and analyze trends driving the evolution of automated summarization tools, such as advancements in natural language processing techniques and machine learning algorithms.

Explore the integration of multimodal data sources, contextual understanding, and summarization personalization.

Case Studies and Use Cases:

Present case studies showcasing the practical applications of automated text document summarization tools in real-world scenarios.

Analyze the impact of these tools on enhancing information retrieval, content curation, and decision-making processes.

Comparative Analysis:

Conduct a comparative analysis of extractive and abstractive summarization techniques, highlighting their strengths and weaknesses.

Discuss scenarios where one approach might be more suitable than the other.

Challenges and Future Directions:

Identify the challenges and limitations faced by current automated text document summarization tools.

Discuss ongoing research efforts and potential directions for future advancements, such as better handling of domain-specific language, improving coherence in abstractive summaries, and addressing ethical concerns.

Synthesis and Discussion:

Synthesize the findings from the analysis of summarization approaches, trends, case studies, and challenges.

Discuss the implications of these findings for researchers, practitioners, and industries utilizing automated text document summarization tools.

By following this methodological approach, the survey aims to provide a comprehensive and insightful exploration of automated text document summarization tools, encompassing their approaches, trends, practical use cases, and future directions.

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.

The American Journal of Applied sciences
(ISSN – 2689-0992)
VOLUME 05 ISSUE 12 Pages: 1-5
SJIF IMPACT FACTOR (2020: 5. 276) (2021: 5. 634) (2022: 6. 176) (2023: 7. 361)
OCLC – 1121105553
Publisher: The USA Journals

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

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

The American Journal of Applied sciences
(ISSN – 2689-0992)
VOLUME 05 ISSUE 12 Pages: 1-5
SJIF IMPACT FACTOR (2020: 5. 276) (2021: 5. 634) (2022: 6. 176) (2023: 7. 361)
OCLC – 1121105553
Publisher: The USA Journals

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
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
9. International Science Conferences, ACM, Jgateplus, ebsco, ijit libraries.
10. A Hybrid Approach for Extractive Document Summarization Using Machine Learning andClustering Technique(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014