A Survey on Machine Learning Based Keyphrase Generation in Natural Language Processing

Jimmy Jose[1*] and P. Beaulah Soundarabai[1]

[1] CHRIST (Deemed to be University), Department of Computer Science, Bengaluru - 560029, India

***CORRESPONDING AUTHOR:**

Jimmy Jose
Email: jimmy.jose@res.christuniversity.in

ISSN : 2382-5359(Online), 1994-1412(Print)

DOI:
https://doi.org/10.3126/njst.v22i2.85238

Date of Submission: 8 Feb, 2023

Date of Acceptance: 17 Mar, 2024

ABSTRACT

With the exponential increase of available textual data, transforming the natural language content into potential information becomes crucial to assist the prominent application domains. In Natural Language Processing (NLP) applications, keyphrase generation has recently become an increasingly prominent research topic. Even though numerous advancements are realized in the current keyphrase generation models, the proliferation of neural network models profoundly impacted natural language generation to a new era. Over the past several years, various studies on keyphrase extraction and generation have been developed that deliver significant contributions to the current state of keyphrase generation research. The researchers confront understanding the deep insights for further developments from the conventional keyphrase generation research while adopting the deep neural network models. Hence, this work notably focuses on studying the keyphrase generations with the impact of the exploding deep learning methods. This survey briefly introduces keyphrase generation in NLP and reviews the recent abstractive methods using deep learning for meaningful keyphrase generation that achieves state-of-the-art performance. By determining and discussing the shortcomings in the previous machine learning and deep learning-based keyphrase generation, this work facilitates strong groundwork for understanding the recent developments in keyphrase generation. Also, it discusses the research challenges in keyphrase generation from the perspective of both text mining and deep learning models. Thus, this study lays the foundation for the researchers to develop potential solutions to resolve the research constraints in keyphrase generation.

Keywords: Natural Language Generation, Machine Learning, Keyphrase Extraction, Keyphrase Generation, Deep Learning.

1. INTRODUCTION

In recent years, the rapid growth of textual data and the use of deep learning has propelled the field of Natural Language Processing (NLP). NLP is considered a potential field of Artificial Intelligence (AI) to solve computational models and processes to analyze natural languages for real-world problems automatically (Lauriola et al., 2022, Landolt et al., 2021). To be concise, NLP is a technology concerned with interactions between natural human languages and computing devices. Owing to the ubiquitous human-computer interaction, NLP spread its applications as given by Kalyanathaya et al., 2019 in various fields ranging from text summarization and text mining to web search engines, voice assistants, machine translation, robotics, and speech recognition (El-Kassas et al., 2021).

Text generation, formally categorized as Natural Language Generation (NLG), has become a challenging task in NLP that produces a natural language text from the non-linguistic representation of information (Gatt and Krahmer, 2018).Keyphrases are phrases that provide highly-condensed information, which facilitates improving the utilization efficiency in a document. There are two main methods involved in automatic keyphrase generation methods: extraction and abstraction.Several researchers have proposed various keyphrase generation methods and keyphrase extraction methods for the implementation of various NLP tasks (Parida et al., 2021).

Deep learning has significantly impressed numerous areas of NLP involving Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs) proposed by Michelucci, 2022, and several other potential deep neural network architectures (Patel, 2020). Different from these conventional models, deep neural architectures are Generative Adversarial Networks (GAN) and Variational Encoder, which are also used for text generation models, but these models possess a different way of training the network (Lu et al., 2018). Researchers have developed text generation models using Pre-trained Language Models (PLM) to learn semantics and contextual representations of universal languages, which are trained on the large-scale text corpus and tuned for various downstream NLP applications (Li et al., 2021).

Several papers presented keyphrase generation models in NLP. This paper presents a survey of NLP with the aim of deep insights into keyphrase generation using deep learning approaches to assist the researchers in further improving the performance of the keyphrase generation methods. The organization of this paper is as follows. The summary of NLP and its significance with various categories of applications are discussed in Section 2. Section 3 describes the role of keyphrase extraction and Keyphrase generation in NLP and the different keyphrase extraction methods and keyphrase generation. Section 4 reviews various machine learning and potential deep learning-based keyphrase generation methods. Section 5 presents research challenges in keyphrase generation, and Section 6 concludes this work.

2. NATURAL LANGUAGE PROCESSING

The powerful NLP is one aspect of Artificial Intelligence that empowers machines to analyze and comprehend natural language based on computer algorithms and programs. With the advent of natural language content through online and offline processes, NLP plays a crucial role in various application areas by designing different information extraction processes with text mining procedures. It is a potential tool for many organizations to identify powerful insights from comprehending human-generated large data corpus.

2.1 Significance of NLP Approaches

With the emergence of web-based technologies, textual data has had tremendous growth in the form of unstructured or semi-structured data (Hariri et al., 2019). This unstructured or semi-structured data makes searching and analyzing key content information more complex. Thus, it demands automatic information filtering and summarization and the field of NLP has made great progress to resolve this problem (Khurana et al., 2022). NLP components include Natural Language Understanding (NLU), NLG, Speech Processing, Text Mining, and Text Analytics. Among the NLP tasks, Text mining, and Text Analytics, have seen unprecedented growth. Text Mining has gained significant attention in NLP applications, which deal with discovering and extracting a pattern from unstructured natural language text data, and it facilitates classifying documents or clustering the documents into categories (Sebastião et al., 2022).

2.2 Tasks of NLP

Generally, a sentence in natural language is based on rules, syntax, and semantics, whereas symbols are used to convey the information contained in the text sentence. NLP concerns the machine's ability to understand and generate natural language. NLP methods strongly rely on language linguistics, which studies language, including meaning, grammar, structure, and phrases. Part-of-speech tagging (POS), Stemming, Lemmatization, and Stopword Removal are the generally applied NLP tasks (Lourdusamy and Abraham, 2018).

2.3 Categories of NLP

The broad categorization of NLP includes NLU and NLG, which evolves the task of generating and classifying the text.

2.3.1 The Natural Language Understanding

Natural Language Understanding (NLU) tries to interpret meaning by dealing with common human errors, including mispronunciations or transported letters or words (Otter et al., 2020). To understand and represent the natural language in potential representation, NLP applies different levels of linguistic analysis that assist people in extracting meaning from text or spoken language. The different stages of analysis in processing natural language involve Morphological analysis, Lexical analysis, Syntactic analysis, Semantic analysis, Discourse analysis, and Pragmatic analysis.

2.3.2 Natural Language Generation

Natural Language Generation (NLG) is categorized as a sub-field of NLP aiming at generating a text understandable to human language (Dong et al., 2021). It involves the analysis of the versatility of the language constructs. Based on the context, NLG systems decide to generate the natural language. The instances of NLG are text-to-text generation and data-to-text generation. Text abbreviation, Text translation, and Text expansion are the categorization of text-to-text generation. The task of generating text based on image or video is an application of data-to-text generation. Because of its extremely challenging and promising application prospects, NLG has recently gained more attention among researchers.

3. ROLE OF KEYPHRASE EXTRACTION AND GENERATION IN NLP

Due to the increasing usage of keyphrases, Keyphrase extraction and Keyphrase generation play a significant role in the field of NLP. For a single document, keyphrases serve as a condensed summary, such as an article, enabling the reader to decide quickly whether the article is interesting. The keyphrase prediction model recommends an automatically generated set of representative phrases related to the main topics inherent in a document. Keyphrase prediction is broadly categorized into extraction and generation methods (Nasar et al., 2019).

3.1 Methods in Keyphrase Extraction

Keyphrase extraction is a fundamental step for various tasks of NLP, including document classification, document clustering, and text summarization. Automatic Keyphrase extraction methods determine a group of phrases, also known as 'candidate phrases which are scored and ranked (Alami, et al., 2020). By employing the keyphrase extraction method, the best phrases are selected as a document's keyphrases. Ranked extracted keyphrases are presented in two sets: Single document keyphrases or document collection keyphrases. Keyphrases extraction methods are widely used in NLP, text mining, document mining, information retrieval, and document collection.

3.2 Methods in Keyphrase Generation

Keyphrase generation gives a set of keyphrases in a document, which is an important task in NLP. The main task is to predict keyphrases from the source text automatically. The automatic keyphrase generation model combines the keyphrase extraction and keyphrase generation processes. The domain-independent keyphrase generation approach incorporates the unsupervised keyphrase extraction phase and the keyphrase inference phase to describe and classify the text (DeNart et al., 2014). Several automatic keyphrase extraction methods focused on extracting present keyphrases, the phrases appearing exactly in the source text. Ignoring the semantics and the underlying context of the content is an issue and cannot generate absent keyphrases that do not appear in the source text (Almutiry, 2020). Several researchers have presented sequential keyphrase generation models to resolve this constraint and generate both the present and absent keyphrases. Supervised and unsupervised methods are the categorization of Keyphrase Generation methods.

Supervised Method: Supervised Keyphrase Generation approaches consider keyphrase generation as a classification problem (Meng et al., 2020). The supervised approach is a learning model trained based on potential features of labelled keyphrases and classifies a candidate keyphrase and not a candidate keyphrase. Moreover, it transforms the keyphrase extraction task into a classification or regression problem.

Unsupervised Method: Domain-independent algorithm, also called an unsupervised algorithm, applies to documents in different domains without needing controlled vocabularies or prior knowledge. In contrast to the supervised approaches, the unsupervised approaches consider keyphrase extraction as a ranking that employs techniques such as clustering as proposed by Nair et al., 2021 or graph-based ranking to determine the keyphrases in the document.

4. IMPACT OF MACHINE LEARNING AND DEEP LEARNING MODELS IN NLP

NLP offers potential means of gaining access to information and facilitates many natural language applications and information retrieval tasks, including text summarization, opinion mining, text categorization, and document clustering (Ibrahim et al. 2019). The major process of machine learning and deep learning models in NLP is categorized into text preprocessing, text representation, model training, and model evaluation. Due to the increased amount of textual information on the web, summarizing the entire information or document along with the main points in the NLP task is critical, so keyphrase extraction and generation methods are essential.

4.1 Review on Machine Learning Approaches in Keyphrase Generation

Different Machine learning approaches include supervised, unsupervised, and reinforcement learning in Keyphrase Generation. The ranking-based keyphrase generation approaches develop machine learning models that determine the keyphrases based on the previously generated key phrases which are relevant and then rank them. The ranking phrase candidates often utilized supervised and unsupervised machine learning features with manually defined features, including **TF-IDF** and **PageRank**. However, these manually annotated features do not determine the significance of each word in the document, leading to inaccurate prediction of the semantics relation among the words underlying the document content. **Naive Bayes**, a supervised machine learning model, employs TF-IDF measurement to predict and classify whether or not the candidate key phrase is the best phrase. Traditional word representation methods, such as a bag of words and one-hot encoding, do not contain semantic information.

To resolve this constraint, the word distribution method as proposed by Le et al., 2014 employs a neural network and maps each word vector to a shorter word vector using the word embedding method. The sentence embedding method is **Sent2Vec**, and the word embedding method is **Word2Vec**. The two classical embedding methods in which Word2Vec maps the entire document and Sent2Vec maps the sentences to vectors that utilize n-gram features to generate the sentence embeddings.

By employing the labelled training knowledge, Supervised machine learning models can automatically generate this mapping process and generates the target text. Hence, supervised classifiers are trained on documents annotated with keyphrases which are useful for determining whether a candidate phrase is a key phrase. **Naive Bayes, bagged C4.5, SVM**, and **maximum entropy** are the different supervised machine learning models widely used to predict and generate keyphrases. Unsupervised machine learning approaches consider keyphrases as a ranking problem, and it performs keyphrase extraction and generation without prior knowledge.

4.2 Review on Deep Learning Approaches in Keyphrase Generation

This section reviews several existing keyphrase generation approaches using different deep learning models. Most of the existing machine learning-based keyphrase generation methods fail to predict the absent keyphrases and fail to capture the text's semantic meaning.

The research work by Meng et al., 2017proposed the deep keyphrase generation model, **CopyRNN**, by utilizing RNN with copy mechanism and predicting the keyphrases without considering the presence or absence of the keyphrases in the text. Initially, it extracts the semantic and syntactic features using an encoder model with an RNN model for semantic understanding. In subsequence, the copy mechanism in the RNN model predicts the significant word based on the attention of positional and syntactic information in the context, and consequently, it enables the RNN to predict the *OutOfVocabulary* words in the source text. Finally, the decoder model in the RNN learns the significant features obtained from the copy mechanism and effectively generates target keyphrases in a text. Even though this work considers the semantic meaning, it fails to predict correlation among the target keyphrases.

The **Title-Guided Network (TG-Net)** model by Chen et al., 2019 utilizes the input as source context with title and generates the keyphrases based on the analysis of the title-influenced words in the document using the encoder-decoder architecture. It provides equal importance to the document's title while extracting the keyphrases from its main body. Initially, the encoder model in the TG-Net learns the input as source context input and title input and generates the title-guided context representation through the sequence of three embedding layers. Then, the attention-based decoder model in the TG-Net learns the highly summarized information extracted from the encoder module and effectively generates the target keyphrases of the text. Even though seq2seq models have achieved superior performance on keyphrase generation task, seq2seq models are inappropriate for large-scale unlabeled samples with the lack of labeled information or training knowledge.

To address this issue, the researchers suggested a semi-supervised learning approach by Ye et al., 2018 that utilizes labelled and unlabeled data to generate the keyphrases. It initially tags the synthetic keyphrases extracted from the unsupervised keyphrase extraction methods with the unlabeled documents and combines the tagged unlabeled samples with the labelled samples. Then, by jointly learning, keyphrases and titles of the document are generated. The seq2seq model fails to predict the Out Of Vocabulary (OOV) words without considering the correlation among the generated keyphrases.

To handle this, the **seq2seq RNN** model incorporates a copy mechanism, attention mechanism, and coverage mechanismas proposed by Zhang et al., 2018for predicting the key phrases and generating the highly correlated keyphrases based on semantic information. The proposed model constrains the OOV issue through the copy mechanism and generates the highly correlated keyphrases based on the analysis of the relation between the generated keyphrases with the assistance of the coverage mechanism in the model.

To resolve the constraint of generating too few keyphrases, researchers presented a **reinforcement learning approach** by Chan et al., 2019, to generate adequatekey phrases from the basis of the adaptive reward function. It incorporates a Wikipedia knowledge base to detect name variations of the keyphrases and generates accurate keyphrases.

To automatically extract multiple keyphrases from Twitter-like sites, the research work by Zhang et al., 2016 proposed a **joint-layer RNN** model which combines keyword ranking, keyphrase generation, and keyphrase ranking. The joint-layer RNN model comprises two hidden layers and two output layers for processing the keyword ranking and keyphrase generation task.

The research work by Zhao et al., 2019 presented a parallel Seq2Seq model called **ParaNet** with the coverage attention mechanism to address the generation of overlapping keyphrases issue. It employs a word encoder and tag encoder in the ParaNet on the source side and a multi-task learning with POS tags for words on the target side to generate key phrases. The encoder in the ParaNet learns the semantic representation of the words in the source text. While the decoder in the ParaNet model comprises a keyphrase decoder that generates the keyphrases, and the POS tag decoder predicts the POS tags of words in keyphrases. Moreover, this work enhances the keyphrase generation performance through linguistic constraints of keyphrases in the Seq2Seq network and multi-task learning framework.

The **Sentence Selective Network (SenSeNet)** model by Luo et al., 2020 avoids the constraints of generating keyphrases from unimportant sentences using end-to-end training through the straight-through estimator and sentence selection module through weak supervision. Initially, it extracts the features by employing the bi-GRU model and subsequently selects the significant sentences through CNN in the sentence selection module, which learns the significant features and generates the sentence representation. Finally, the work generates the keyphrases based on estimated sentences related to keyphrases obtained from the selection model.

The research by Zhu et al., 2020 employed an encoder-decoder model with a copy mechanism called **CopyNet** to generate keyphrases byanalyzing semantic and significant information in the source text. Initially, this work automatically builds the keyphrase semantic web to extract more related information for source text and to improve the keyphrase generation performance. Based on the attention mechanism, it extracted the source text's potential information with the semantic web's assistance and effectively generated the keyphrases of the source document.

Table 1 compares the conventional keyphrase generation approaches with their advantages and limitations.

Table 1: Comparison of the Conventional Keyphrase Generation Approaches

Author Name and Year	Learning Model	Objective	Algorithm	Advantages	Limitations
Le et al. (2014)	Machine learning	To produce the distributed representation of the input sequence with variable length	Unsupervised model, Paragraph Vector	A neural network maps the word vector to a shorter word vector from training.	Lacks to support the generation of the keyphrases with parsing on complex sentences.
Menget al. (2017)	Deep learning	To analyze the semantic meaning of the keyphrase generation using a copy attention mechanism	Recurrent Neural Network and Copy RNN	It predicts the absent keyphrases	Needs to focus on discontinuity in the sentences
Chen, et al. (2019)	Deep learning - Title Guided Network	To generate keyphrases in a document through title-influenced words.	Encoder-Decoder architecture	It provides the importance of the document's title while extracting keyphrases from the source document.	Fail to rely on simple statistics or rules to yield good results.
Ye et.al (2018)	Deep learning	This model uses labeled and unlabeled data for generating the keyphrases.	Semi-supervised learning	It generates keyphrases for new data.	Lack of nonlinear uncertainties problem
Zhang and Xiao (2018)	Deep learning	To generate OOV words and highly correlated keyphrases throughpotential information of the document.	Seq2Seq model	It enhances the information with the help of GRU	Gold-standard keyphrases are inadequate to validate the keyphrase generation
Chan et al. (2019)	Deep learning	To generate accurate and adequate keyphrases from large text documents.	Reinforcement learning	It resolves the generation of too few key phrases	Lack of keyphrase boundary leads to ineffective adaptive modeling
Zhang et al. (2016)	Deep learning	To generate accurate keyphrases from a large collection of documents such as twitter sites.	Recurrent Neural Network	It combines keyword and context information.	Fails to focus on the dependencies in the labels

Zhao and Zhang(2019)	Deep learning	To resolve the overlapping keyphrase generation.	Parallel seq2seq network (paraNet)	It utilizes syntactic constraints to prevent overlapping keyphrase generation.	Lack of determining the number of keyphrases leads to the over-generate of keyphrases
Luo, et al.(2020)	Deep learning	To generate the efficient metadata sentence from the documents.	seq2seq network	It resolves the discontinuity problem during keyphrase generation	Fails to predict the absent keyphrase
Zhuet al.(2020)	Deep Learning	To generate keyphrases with an attention mechanism	Encoder-decoder framework	It handles Out-Of-Vocabulary(OOV) problems through semantic web	Lack of keyphrase boundary leads to inaccurate keyphrase generation

5. RESEARCH CHALLENGES IN KEYPHRASE GENERATION

In the NLP, keyphrase generation models have several research gaps, which are presented as follows.

- Lack of analyzing the document's semantic meaning while applying the machine learning methods alone for keyphrase generation results in inaccurate key phrases.
- Supervised and unsupervised machine learning methods fail to predict the absent key phrases during the keyphrase generation process.
- A unique challenge in keyphrase generation is generating multiple target keyphrases during the keyphrase prediction process.
- Models predict inaccurate or incomplete keyphrases if the input sequence data is inaccurate or incomplete.
- Existing Keyphrase generation methods neglect correlation among multiple keyphrases, thus resulting in duplication and coverage issues.
- Overlapping keyphrases have been occurring while applying existing deep learning models, resulting in inaccurate key phrases.
- The neural seq2seq model fails to generate the keyphrases for new data when learning the patterns from the labelled data
- Without adaptively tuning the deep learning model and utilizing the external knowledge source, the keyphrase generation method confronts the generation of the absent keyphrases and the generation of present keyphrases.
- Existing generative models predict multiple keyphrases and the number of keyphrases from the input document but fail to generate accurate and too few keyphrases.

6. CONCLUSION

This survey presented the keyphrase generation methods and techniques for large-size documents in various NLP applications. It enhanced the automatic keyphrase generation methods using prior knowledge and data to predict semantic meaning and absent keyphrases. Keyphrase generation also applied semi-supervised learning algorithms to predict the present and absent keyphrases simultaneously. In addition, it utilized deep reinforcement learning with an adaptive reward to predict the semantic meaning of the absent keyphrases. By discussing the pros and cons, this work enables better directions for future research work in keyphrase generation models.

REFERENCES

AlamiMerrouni, Zakariae, BouchraFrikh, and BrahimOuhbi., 2020. Automatic keyphrase extraction: a survey and trends. Journal of Intelligent Information Systems, Vol. 54, No. 2, pp.391-424. DOI: 10.1007/s10844-019-00558-9

Almutiry,Omar.,2021.AutomaticKeyPhraseExtraction. IEEE, International Conference on Engineering and Emerging Technologies (ICEET), pp. 1-7. DOI: 10.1109/ICEET53442.2021.9659724 PMID: 34512113 PMCID: PMC8421017

Chen, Wang, Yifan Gao, Jiani Zhang, Irwin King, and Michael R. Lyu., 2019. Title-Guided Encoding for Keyphrase Generation. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, No. 01, pp. 6268-6275. DOI: 10.1609/aaai.v33i01.33016268

Chan, Hou Pong, Wang Chen, Lu Wang, and Irwin King., 2019. Neural keyphrase generation via reinforcement learning with adaptive rewards. arXiv preprint arXiv, pp. 1906.04106. DOI: 10.18653/v1/P19-1208

DeNart, Dario, and Carlo Tasso., 2014. A Domain Independent Double Layered Approach to Keyphrase Generation, In WEBIST (2), pp. 305-312. DOI: 10.5220/0004855303050312

Dong, Chenhe, Yinghui Li, Haifan Gong, Miaoxin Chen, Junxin Li, Ying Shen, and Min Yang., 2021. A Survey of Natural Language Generation. arXiv preprint arXiv:2112.11739.

El-Kassas, Wafaa S., Cherif R. Salama, Ahmed A. Rafea, and Hoda K. Mohamed., 2021. Automatic text summarization: A comprehensive survey. Expert Systems with Applications. Vol. 165,pp.113679. DOI: 10.1016/j.eswa.2020.113679

Gatt, Albert, and Emiel Krahmer., 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. Journal of Artificial Intelligence Research, Vol. 61, pp. 65-170. DOI: 10.1613/jair.5477

Hariri, Reihaneh H., Erik M. Fredericks, and Kate M. Bowers., 2019. Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data, Vol.6, No. 1, pp. 1-16. DOI: 10.1186/s40537-019-0206-3

Ibrahim, R., S. Zeebaree, and K. Jacksi., 2019. Survey on Semantic Similarity Based on Document Clustering. Adv. sci. technol. eng. syst. No. 5, pp.115-122. DOI: 10.25046/aj040515

Kalyanathaya, Krishna Prakash, D. Akila, and P. Rajesh., 2019. Advances in natural language processing-a survey of current research trends, development tools and Industry applications. International Journal of Recent Technology and Engineering, Vol. 7, No. 5C, pp. 199-202.

Khurana, Diksha, Aditya Koli, Kiran Khatter, and Sukhdev Singh., 2022. Natural language processing: State of the art, current trends and challenges. Multimedia Tools and Applications, pp. 1-32. DOI: 10.1007/s11042-022-13428-4 PMID: 35855771 PMCID: PMC9281254

Lauriola, Ivano, Alberto Lavelli, and Fabio Aiolli., 2022. An introduction to deep learning in natural language processing: Models, techniques, and tools. Neurocomputing, Vol.470, pp. 443-456. DOI: 10.1016/j.neucom.2021.05.103

Landolt, Severin, Thiemo Wambsganss, and Matthias Söllner., 2021. A taxonomy for deep learning in natural language processing. Hawaii International Conference on System Sciences. DOI: 10.24251/HICSS.2021.129

Le, Quoc, and Tomas Mikolov., 2014. Distributed representations of sentences and documents. In International conference on machine learning, pp. 1188-1196.

Li, Junyi, Tianyi Tang, Wayne Xin Zhao, and Ji-Rong Wen., 2021. Pretrained language models for text generation: A survey", arXiv preprint arXiv:2105.10311. DOI: 10.24963/ijcai.2021/612

Lourdusamy, Ravi, and Stanislaus Abraham., 2018. A survey on text preprocessing techniques and tools. International Journal of Computer Sciences and Engineering, Vol. 6, No. 03, pp. 148-157. DOI: 10.26438/ijcse/v6si3.148157

Lu, Sidi, Yaoming Zhu, Weinan Zhang, Jun Wang, and Yong Yu., 2018. Neural text generation: Past, present and beyond. arXiv preprint arXiv:1803.07133.

Luo, Yichao, Zhengyan Li, Bingning Wang, Xiaoyu Xing, Qi Zhang, and Xuanjing Huang., 2020. SenSeNet: Neural Keyphrase Generation with Document Structure.arXiv preprint arXiv, pp. 2012.06754.

Meng, Rui, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, and Yu Chi., 2017. Deep keyphrase generation.arXiv preprint arXiv, pp. 1704.06879. DOI: 10.18653/v1/P17-1054

Meng, Rui, Xingdi Yuan, Tong Wang, Sanqiang Zhao, Adam Trischler, and Daqing He., 2020. An empirical study on neural keyphrase generation. arXiv preprint arXiv:2009.10229. DOI: 10.18653/v1/2021.naacl-main.396 PMID: 34494393

Michelucci, Umberto, 2022. A Brief Introduction to Recurrent Neural Networks. In Applied Deep Learning with TensorFlow, Apress, Berkeley, CA, Vol. 2, pp. 245-255. DOI: 10.1007/978-1-4842-8020-1_8

Nair, Srikesh Rajesh, G. Gokul, AkshayAntoVadakkan, Aditya G. Pillai, and M. G. Thushara., 2021. Clustering of Research Documents-A Survey on Semantic Analysis and Keyword Extraction. In 2021 6th International Conference for Convergence in Technology (I2CT), pp. 1-6. DOI: 10.1109/I2CT51068.2021.9418197

Nasar, Zara, Syed Waqar Jaffry, and Muhammad Kamran Malik., 2019. Textual keyword extraction and summarization: State-of-the-art. Information Processing&Management,Vol.56,No.6,pp.102088. DOI: 10.1016/j.ipm.2019.102088

Otter, Daniel W., Julian R. Medina, and Jugal K. Kalita., 2020. A survey of the usages of deep learning for natural language processing. IEEE transactions on neural networks and learning systems, Vol. 32, No. 2, pp. 604-624. DOI: 10.1109/TNNLS.2020.2979670 PMID: 32324570

Parida, Upasana, Mamata Nayak, and Ajit Ku Nayak., 2021. Insight into diverse keyphrase extraction techniques from text documents. Intelligent and cloud computing, pp. 405-413. DOI: 10.1007/978-981-15-5971-6_44

Patel, Sanskruti., 2020. A comprehensive analysis of Convolutional Neural Network models. International Journal of Advanced Science and Technology, Vol. 29, No. 4, pp.771-777.

Sebastião, Pais, Cordeiro João, and Jamil M. Luqman., 2022. NLP-based platform as a service: a brief review. Journal of Big Data, Vol. 9, No. 1. DOI: 10.1186/s40537-022-00603-5

Ye, Hai, and Lu Wang., 2018. Semi-supervised learning for neural keyphrase generation. arXiv preprint arXiv, pp. 1808.06773. DOI: 10.18653/v1/D18-1447

Zhao, Jing, and YuxiangZhang., 2019. Incorporating linguistic constraints into keyphrase generation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 5224-5233. DOI: 10.18653/v1/P19-1515 PMID: 31854592

Zhang, Qi, Yang Wang, Yeyun Gong, and Xuan-Jing Huang., 2016. Keyphrase extraction using deep recurrent neural networks on Twitter. In Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 836-845. DOI: 10.18653/v1/D16-1080

Zhang, Y. and Xiao, W., 2018. Keyphrase generation based on deep seq2seq model. IEEE Access, Vol.6, pp.46047-46057. DOI: 10.1109/ACCESS.2018.2865589

Zhu, Xun, Chen Lyu, and Donghong Ji., 2020. Keyphrase Generation WithCopyNet and Semantic Web. IEEE Access, Vol. 8, pp. 44202-44210. DOI: 10.1109/ACCESS.2020.2977508