IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007028 |

Keyword Extraction from Short Text Messages Using IBM Knowledge Studio

Prof. Gangamma H[1], Anusha T P[2], Megha S Naik[3],

[1]Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India

[2]B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India

[3]B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India

ABSTRACT: Keyword extraction, also known as keyword detection or keyword analysis, is a text-analysis technique that automatically identifies and extracts the most prevalent and significant terms and phrases. Unstructured test data is used in text analysis (TA), a sort of machine learning, to automatically extract meaningful information. Text analysis tools help companies quickly digest online data and documents and turn them into insights that can be used. We can use text analysis to pull out specific information, like keywords, names, or information about a company, from thousands of emails. We can also use it to sort survey responses by how they made people feel and what they were talking about. Text analysis delivers qualitative results. When a machine does text analysis, it looks at the text itself to find important information.

KEYWORDS: Keyword extraction, Textanalysis, Qualitative Results.

I. INTRODUCTION

The technique of extracting the most relevant information from a vast body of data is called "keyword extraction." In various disciplines, such as text categorization, text clustering, tracking, topic recognition, summarization, and so on, keyword extraction is an essential job. Keyword extraction is the process of obtaining information from social media sites like Twitter and Facebook and then analysing that information to make judgments. Social media analysis is an essential aspect of this process. Twitter is a widely used social media platform for exchanging and broadcasting information. The 140-character "tweets" that comprise Twitter data are referred to as "tweets." As a result, when a large amount of data is streamed in at once, such as on Twitter, reading becomes difficult. In this instance, it's critical to extract the valuable information from the remainder. This is where the concept of keyword extraction comes in.

II. LITERATURE REVIEW

Zhao, together with colleagues .[1] They used the network model, the semantic space, and the position of words to get the keyword from the Chinese microblog. How did they implement the method? To begin, we need to get the micro blog API for a certain user. Pre-processing is the second phase, and it involves cleaning up the data, dividing words, marking POS words, and removing stop words. Third, we build a graph model to locate the keywords based on how often they occur together. For each word, we apply the Score formula to find out how much weight it has. Using topic detection and TFIDF, we may construct a semantic space, and then utilise that information to calculate the statistical weight. The location of the words is taken into account in the fifth step, which determines the rank value. Because of this, it is clear to us that a lower number will be at the top of the list.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007028 |

Hromic and others .[2] The approach to structure and the development of graphs are the primary concerns of this study. This article employs a structure-based approach. We create a network model and search for popular themes and events. There are two graphs in the clustering of themes for tweets: those with the same subjects and those with distinct topics. The OSLOM technique is used to determine how people interact in a homogenous network. A ranking algorithm called rankclus is used to create a collection of numbered tweets that are arranged according to significance. In the end, we utilise Python to interpret the findings of both graphs. The tweets with the same name are then added to the group. It will be possible in the future to provide a technique for defining events using a variety of graph models.

Marujo and others .[3] As part of their research, they devised a system for extracting keywords and explored possible solutions to issues such as wide variance and lexical discrepancies They also made comparisons between the new approach and others that had previously been used. Due to lexical variation, when two words have the same meaning but vary in spelling, we utilise a different way to figure it out. For this reason, we used two strategies to get the term "keyword." Brown clumps: It is possible to determine the characteristic for each cluster by first grouping together words that have a similar meaning, such as the terms "no" and "noo."

Kim1 and others .[4] In this work, we devise a method for determining if a keyword is "popular" or "bursty." An abbreviation checker is able to identify typos and grammatical errors. First, you'll need to gather a list of potential candidates for trend and burst keywords. Any term that begins with a capital letter or is enclosed in quotation marks is a potential keyword. The next step is to merge the keywords. Acronyms, typographical errors and space are all examined in order to determine how often each phrase is used. Find and choose popular keywords from the combined candidate terms in the following stage, then. It is the "burst ratio" that is utilised to choose the "bursty" keyword phrases. You may create a prototype system that can detect keywords with a shorter shelf life.

SystemDesign

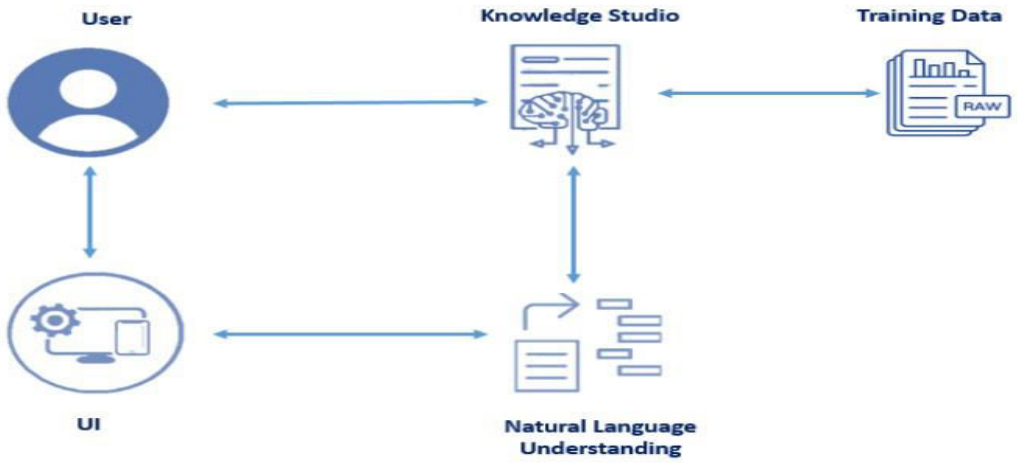


Fig: System Architecture

The system architecture of this project is shown in the figure above.
The above picture shows how the model works and how the user and the system talk to each other. It explains how the different stages work, which includes.

- Add files for the type system and the corpus to Watson Knowledge Studio.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007028 |

- A user trains and evaluates the data to make a model.
- The WKS model has been put into use by Watson NLU.
- The user gives the app an SMS message to look at.
- Watson NLU looks at the SMS message to figure out how to process it, and it sends back extracted domain-specific entities based on the WKS model.

III. RESULTS AND DISCUSSION

Snapshots

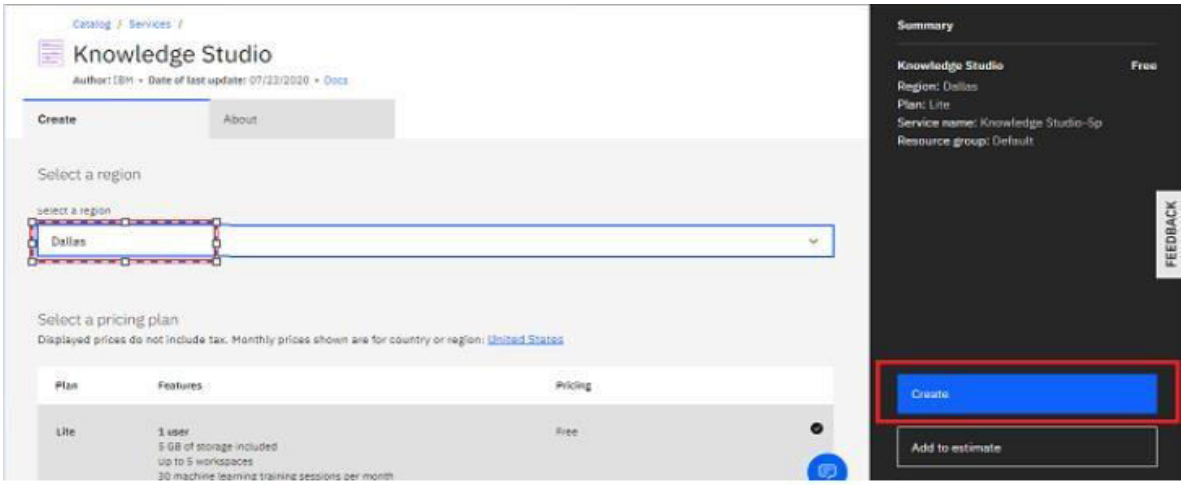
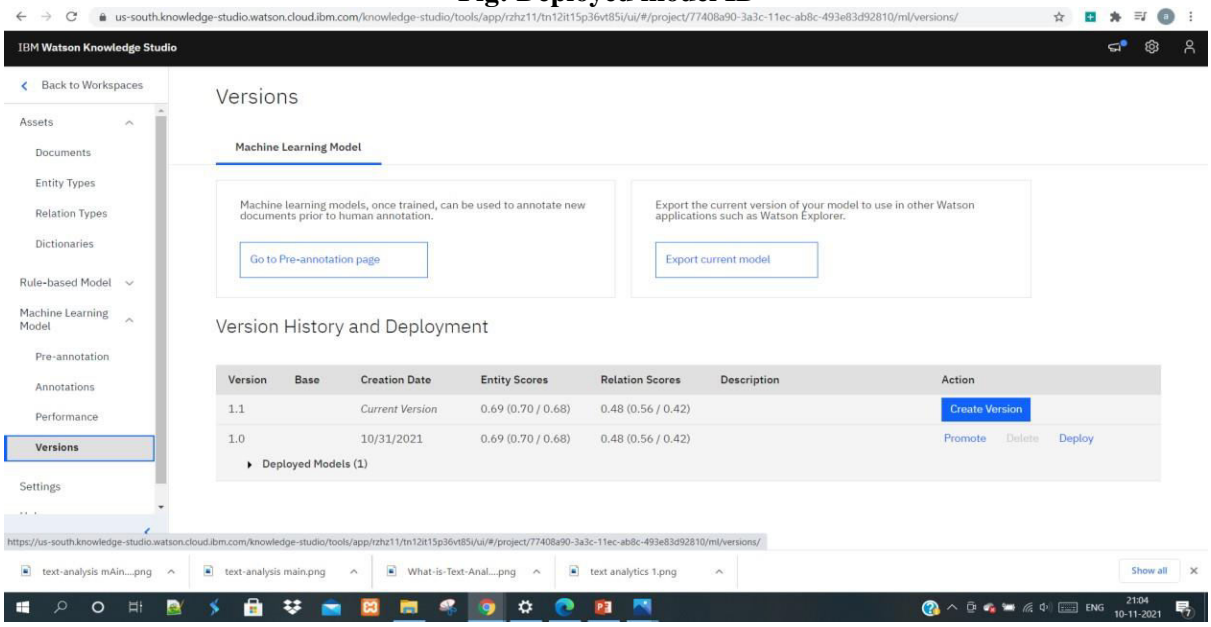


Fig: Snapshot of Creating IBM knowledge Studio Service

Fig: Deployed model ID



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007028 |

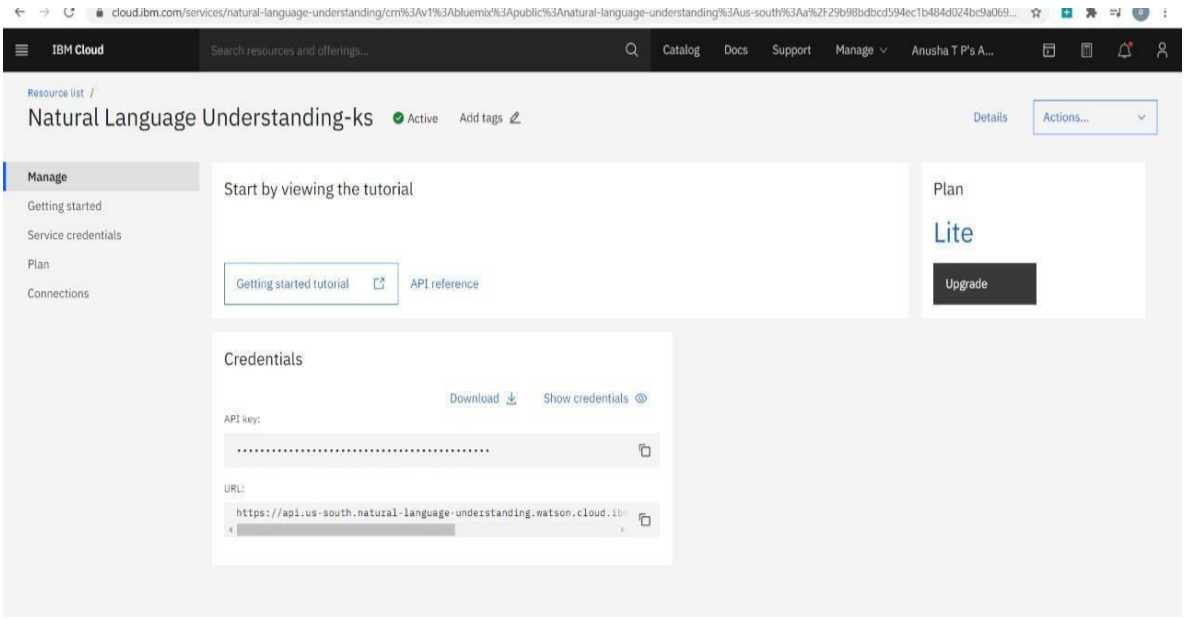


Fig: Natural language understanding service launch page

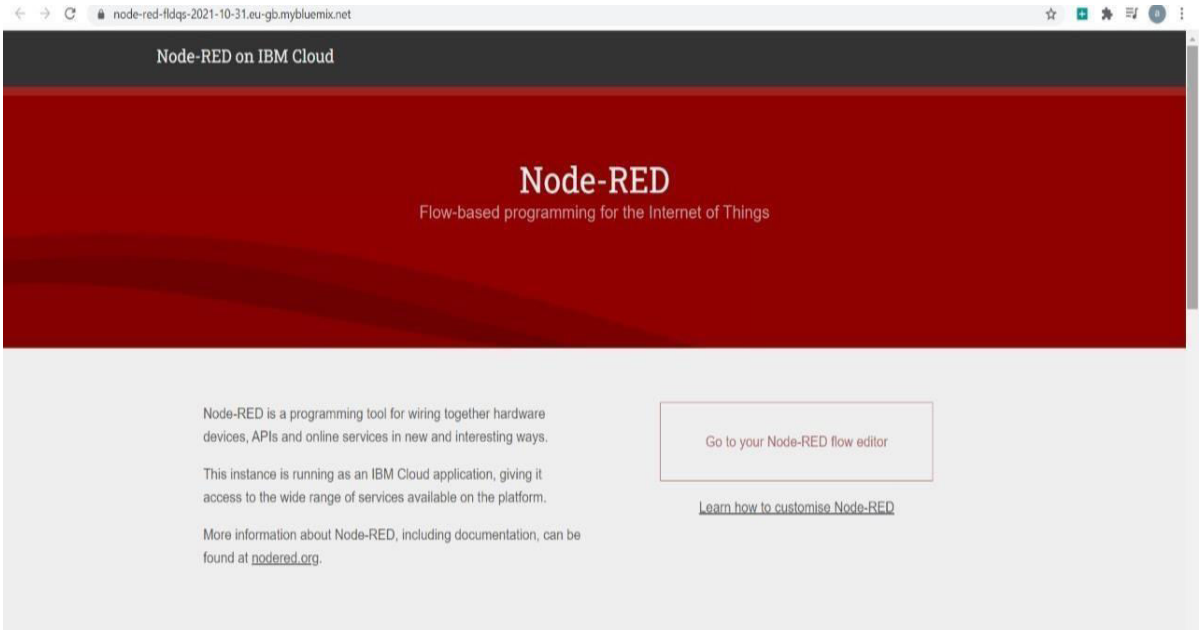


Fig: Node-Red software

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007028 |

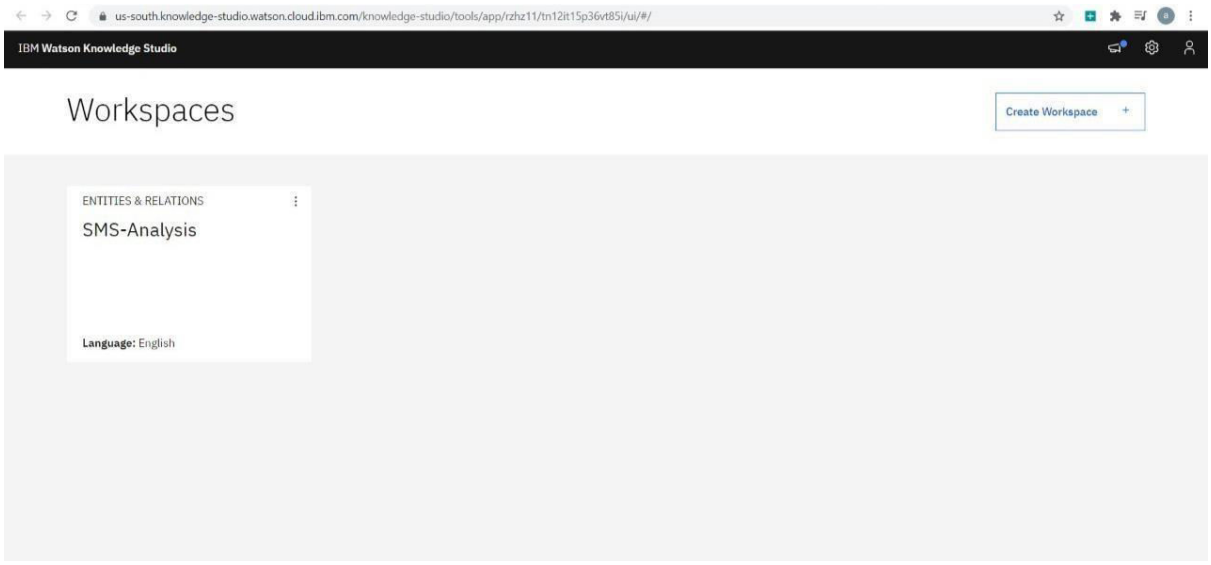


Fig: Workspace of SMS analysis

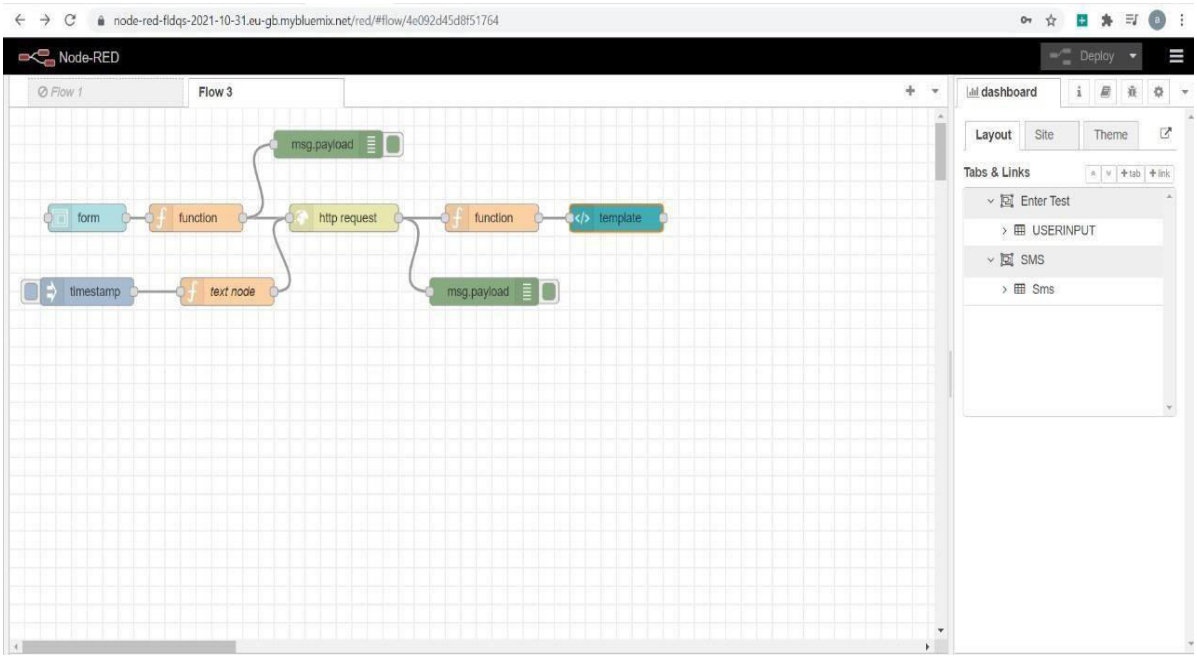


Fig: Node-Red Flows

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007028 |

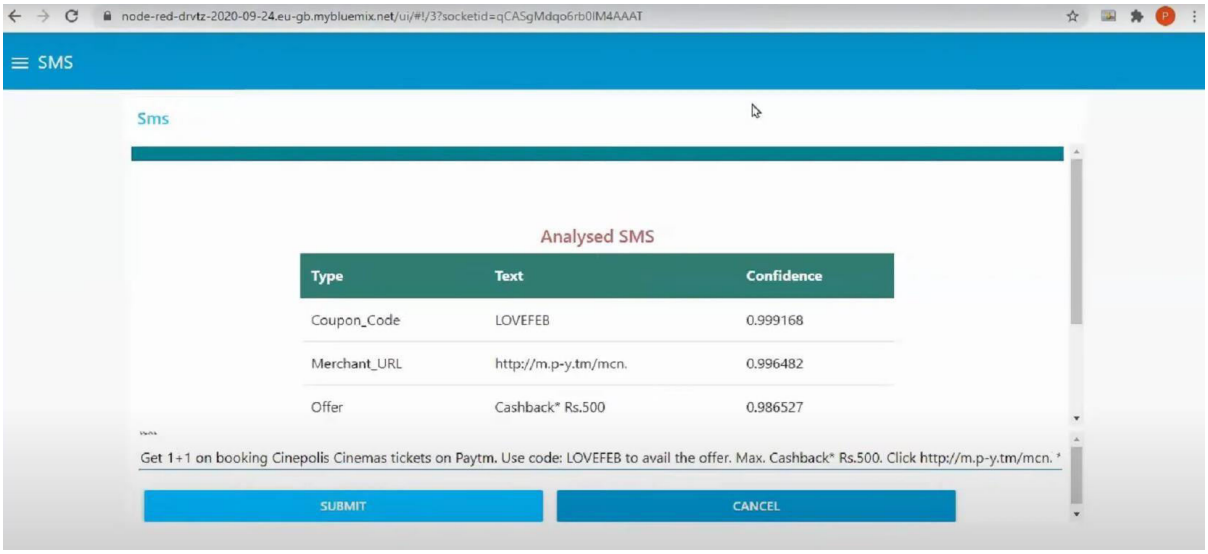


Fig: Output Page

Conclusion

The process of extracting keywords facilitates the discovery of relevant terms and phrases in unstructured material. Emails, social media postings, chat chats, and other unstructured data are all included in this category. Automating chores like labelling survey replies or responding to urgent consumer inquiries may be made much easier using keyword extraction. As a result, it may help you make better business choices based on data. Because of this, keyword extraction models are among the most straightforward to set up and operate.

REFERENCES

1. Lee, L. H., Isa, D., Choo, W. O., & Chue, W. Y. (2012). "High Relevance Keyword Extraction facility for Bayesian text classification on different domains of varying characteristic". Expert Systems with Applications, 39(1), 1147-1155.
2. Carretero-Campos, C., Bernaola-Galván, P., Coronado, A. V., & Carpena, P. (2013). "Improving statistical keyword detection in short texts: Entropic and clustering approaches". Physica A: Statistical Mechanics and its Applications, 392(6), 1481-1492.
3. J. Kaur and V. Gupta, "Effective approaches for extraction of keywords," International Journal of Computer Science Issues (IJCSI), vol. 7, p. 144, 2010.
4. K. Zhang, H Xu, J Tang, and J Li, "Keyword extraction using support vector machine," in Proceedings of the international conference on web-age information management, Springer, Berlin, Heidelberg, June 2006.
5. G. K. Palshikar, "Keyword extraction from a single document using centrality measures," in Proceedings of the International conference on pattern recognition and machine intelligence, Berlin, Heidelberg, December 2007.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details