iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38105

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Contextual Image Illustrator: A Review

Kanika Sharma[1], Anita Choudhary[2]
[1]*Student, Department of Computer Science Engineering, Dronacharya College of Engineering, Khentawas, Haryana, India*
[2]*Assistant Professor, Department of Computer Science Engineering, Dronacharya College of Engineering, Khentawas, Haryana, India*

Abstract: ***Any story or any other literary content is best understood and advertised with the help of pictures. Images are used to arouse reader's interest and comprehension in the content. The contextual image illustrator will take any content description and will output the ranked images related to that content. The text can be any blog, newspaper article, any story or any other content. The image retrieval process that has been used for this purpose is Text based Image Retrieval, i.e., TBIR. Semantic keywords are extricated from the story; images are looked through an annotated database. Thereafter, an image ranking scheme will determine the relevance of each image. Then the user can choose among the images displayed. A score along with each image will also be displayed representing its relevance to the query. The keywords stemming and stop word removal has been explained in the document. Also, the algorithm that has been designed to determine the score and hence the image's significance has been calculated. Testing consisting of both unit testing and module testing of the project are explained.***

Keywords: Keyword Extraction, Image Search, Stemming, Stop word Removal, URL Score, URL Ranking

I. INTRODUCTION

Now-a-days, the textual content is growing exponentially each and every day. To make the comprehension of this large amount of textual data, it is advised to be accompanied with images. The reader can gain from the content more smoothly and more easily if it contains images and hence can understand it better and faster. Keywords are commonly used for search engines and document databases to locate information and determine if two pieces of test are associated with one another. Reading and summarizing the contents of huge entries of text into a little set of topics is difficult and time consuming for a person's , such a lot in order that it becomes nearly impossible to accomplish with limited manpower because the size of the knowledge grows. As a result, automated systems are being more commonly used to do this task. This problem is challenging due to the intricate complexities of natural language, as well as the inherent difficulty in determining if a word or set of words accurately represent topics present within the text. Keyword extraction from text data may be a common tool employed by search engines and indexes alike to quickly categorize and locate specific data supported explicitly or implicitly supplied keywords. Contextual Image Illustrator offers a system that is capable of generating ranked image results for the content writers. The image results will be displayed in descending order on the basis of their relevancy score. The content can wary from a newspaper article to a blog, or even a comic strip kind of story-telling, that tells a story with the help of sequenced images.

II. METHODOLOGY

Contextual Image Illustrator is an implementation of a system to find a set of images based on a sample text. Search should be carried based on the keywords of the images. Different types of techniques can be used for extracting the keywords of the images and conducting the search.

Major functionalities that are required in the product are,

1) Extracting image context

2) Conducting a search for the input text in the search space

The system should first be able to extract the image context. This step involves some complex keyword extraction methodologies and algorithms. These data will be used to make an index for that particular context which is later used when conducting the search. Then the system should be able to carry down the search for the given context in a given search space. In case when an exact match is not available, the searching algorithm should be able to find a set of best matching solutions. This step will involve making some approximations based on the search index of each image. And finally, the system should give the best matching solutions as the output.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

A. Keyword Extraction

Extricating Keywords are perhaps the main task when working with text. While reading, the advantage to readers are the Keywords since they can judge all the more rapidly whether the text merits perusing. Website creators advantage of keywords as they can group comparable substances by their subjects. Algorithm programmers advantage from keywords since they decrease the dimensionality of text to the main provisions.

Selection of Keywords will be on the basis of typical heuristics:

1) Stop word removal using a stop word list
2) Candidate Keywords can be the words with certain part of speech tags (e.g., nouns, adjectives, verbs)
3) n-grams can be allowed, which are found in to be Wikipedia article title
4) Extracting noun phrases that satisfy pre-defined lexicon-syntactic pattern(s)

B. Stop Word Removal

Stop words are essentially a bunch of normally utilized words in any language, not simply English. The motivation behind why stop words are basic to numerous applications is that, in the event that we eliminate the words that are regularly utilized in a given language, we can zero in on the significant words all things being equal.

1) Instances of insignificant stop word records that you can utilize
a) *Determiners* - Determiners will in general stamp things where a determiner normally will be trailed by a thing Models: the, a, an, another
b) *Coordinating Conjunctions* – Coordinating conjunctions interface words, expressions, and conditions Models: for, an, nor, however, or, yet, so
c) *Prepositions* - Prepositions express fleeting or spatial relations Models: in, under, towards, previously

2) Some of the published stop words lists, which can be referred to and used for the benefit
a) *Snowball Stop Word List* – Published under Snowball Stemmer
b) *Terrier Stop Word List* – A Comprehensive stop word list published under Terrier package.
c) *Minimal Stop Word List* – Stop word list that was compiled consisting of determiners, coordinating conjunctions and prepositions
d) *Rainbow Stop Word List*- It is a Statistical Text Classification list. It is based on the Bow library. I am using Rainbow list in this project.

C. Stemming

Stemming is a suffix stripping algorithm that fetches the root word for us to capture the exact Keyword. This helps in improving the context-based image search.

1) *Some Examples Of The Rules Include*
a) Words ending with 'ed', remove the 'ed'
b) Words ending with 'ing', remove the 'ing'
c) Words ending with 'ly', remove the 'ly'

Suffix stripping algorithms might vary in outcomes for an assortment of reasons. One such explanation is whether the algorithm compels whether the yield word should be a genuine word in the given language. A few methodologies don't need the word to really exist in the language vocabulary (the arrangement of all words in the language). Then again, some addition stripping approaches keep an information base (a huge rundown) of all referred to morphological word roots that exist as genuine words. These methodologies really take a look at the rundown for the presence of the term before settling on a choice. Normally, if the term doesn't exist, substitute move is made. This substitute activity might include a few different measures. The non-presence of a yield term might serve to make the calculation attempt a substitute postfix stripping rules. Reality can eventually show that at least two postfix stripping rules apply to a similar information term, which makes vagueness with regards to which rule to apply. The calculation might dole out (by human hand or stochastically) a need to some standard. Or then again the calculation might dismiss one guideline application since it brings about a non-existent term while the other covering rule doesn't.

©IJRASET: All Rights are Reserved

D. *Image Search*

Google Images is an inquiry administration possessed by Google and presented in July 2001 that permits users to scan the Web for picture content. The keywords for the image search depend on the filename of the image, the connection text highlighting the image, and the text nearby the image.

E. *Ranking Score*

The ranking is a connection between a bunch of things with the end goal that, for any two things, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equivalent to' the second. In math, this is known as a frail request or complete preorder of items. Ranking in our venture is an approach to show the request for importance to the setting of the text. Here in our applied way to deal with the Ranking of images, we will probably discover an image that best suites a given text. Numerous URLs from a bunch of watchwords are recovered and afterwards ranking calculation is applied to them. For ranking, numerous calculations could be utilized yet we made our own which fulfils our need.

F. *Ranking Images*

After estimation of scores, we need to rank images. So we recover the scores of every URL and submit the URLs in dropping request of their score. The better the score is, the more closely the image will be to go with the text. Here, we are utilizing input from the client to pick among the images.

III. MODELING AND ANALYSIS

Document key expressions have empowered quick and accurate searching for a given archive from a huge text assortment, and have displayed their potential in further developing numerous natural language processing (NLP) and information retrieval (IR) tasks, like text synopsis, text classification, assessment mining, and record ordering.

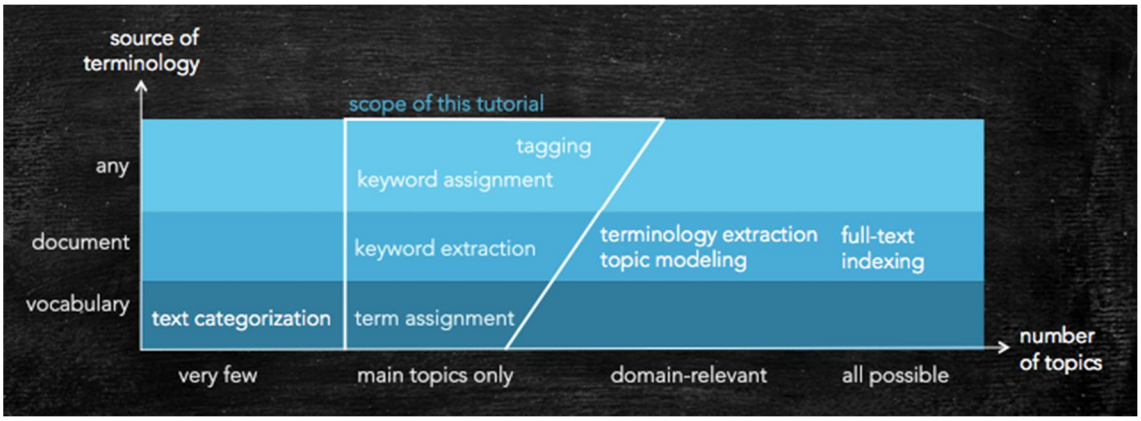


Figure1: 3D Comparison of related tasks in terms of the source of terminology and number of topics selected per document.

The stop words used in this project belong to the Rainbow Stop Word list. Rainbow usage pattern can be achieved in two steps (1) Give your document as input to Rainbow and write to disk a "model" containing their statistics,

(2) Classification or diagnostics is performed by Rainbow using the Model

Some of the Stop Words from the list based on Rainbow statistical text:

a's, able, about, above, anyways, around, as, aside, ask, asking, associated, at, available, away, awfully, be, can't, cannot, cant, cause, causes, certain, didn't, different, do, does, doesn't, doing, especially, et, etc, even, ever, every, everybody, everyone, everything, formerly, given, gives, go, goes, going, gone, got, gotten, greetings, had, hadn't, happens, hardly, has, hasn't, i'd, i'll, i'm, i've, ie, if, me, mean, you, you'd, you'll, you're, you've, your, yours, yourself, yourselves, zero.

A. *External Interface Requirements*

Initially the user is given an interface to input a keyword. Then the search space is searched according to that keyword that is extracted after stemming and removal of stop words. The search results are then ranked as per the algorithm and similarity score is generated for each URL. The image URLs are then ranked on the basis of this score and user can choose any image among them.

©IJRASET: All Rights are Reserved

B. Non Functional Requirements

1) Performance Requirements: Performance is a major non functional requirement in this system. As the system is a search engine, it should have a very short response time.

2) Security Requirements: Dataset should be well secured. It was highly recommended that end user must not be able to do any changes to the system and obtain any valuable system data from the system.

3) Software Quality Attributes: A friendly and simplistic system will go a long way. Average time must be taken for results to be out. Flexible system is more reliable as it will be easier to handle the updates in future.

IV. RESULTS AND DISCUSSION

The results deduced clearly states that the images are more relevant when searched with keywords.

A. Search Results Comparison

1) Example: Query Data: Once upon a time in a fairy tale land a cat and a dog were friends. One night, the cat invited the dog for a party at his house. The cat played the fiddle. The dog happily clapped his hands. Suddenly, they saw a cow flying in the sky. It jumped over the moon. The dog laughed. Just then, they saw a dish and a spoon from the party running away together. And they laughed even louder. After that they became the best friends.

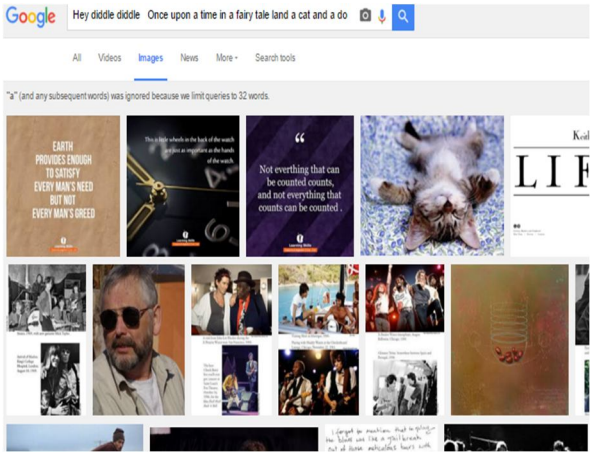


Figure2: Google image search result

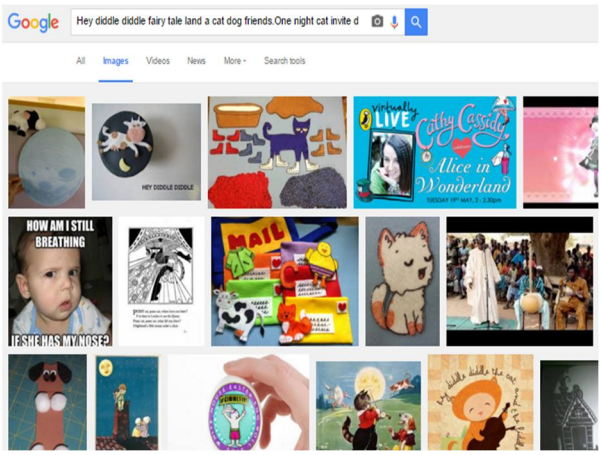


Figure3: Result after Keyword Extraction

Table 1. Test outputs of the Illustrator

S.NO.	TEST CASE	EXPECTED OUTPUT	ACTUAL OUTPUT	RESULT
1.	I'm kanika sharma.	kanika sharma.	kanika sharma.	PASS
2.	Fiercely	Fierce	fierce	PASS
3.	Positive	Positive	positive	PASS
4.	Right now I'm testing this program.	Testing program.	Testing program	PASS
5.	We were walking up the Avenue des Champs-Élysées with Venus, trying to read the story of the siege of Paris in the shell-scarred walls and the sidewalks plowed up by grape-shot.	walk avenue des champ élysées venus read story siege paris shell scar wall sidewalk plow grape shot .	walk avenue des champ élysées venus read story siege paris shell scar wall sidewalk plow grape shot .	PASS

©IJRASET: All Rights are Reserved

V. CONCLUSION

The system successfully concludes that image search results are more relevant when done using keyword extraction. The ranking score successfully ranks the image URLs by matching it to the extracted keywords. The system is able to fulfill its need and purpose.

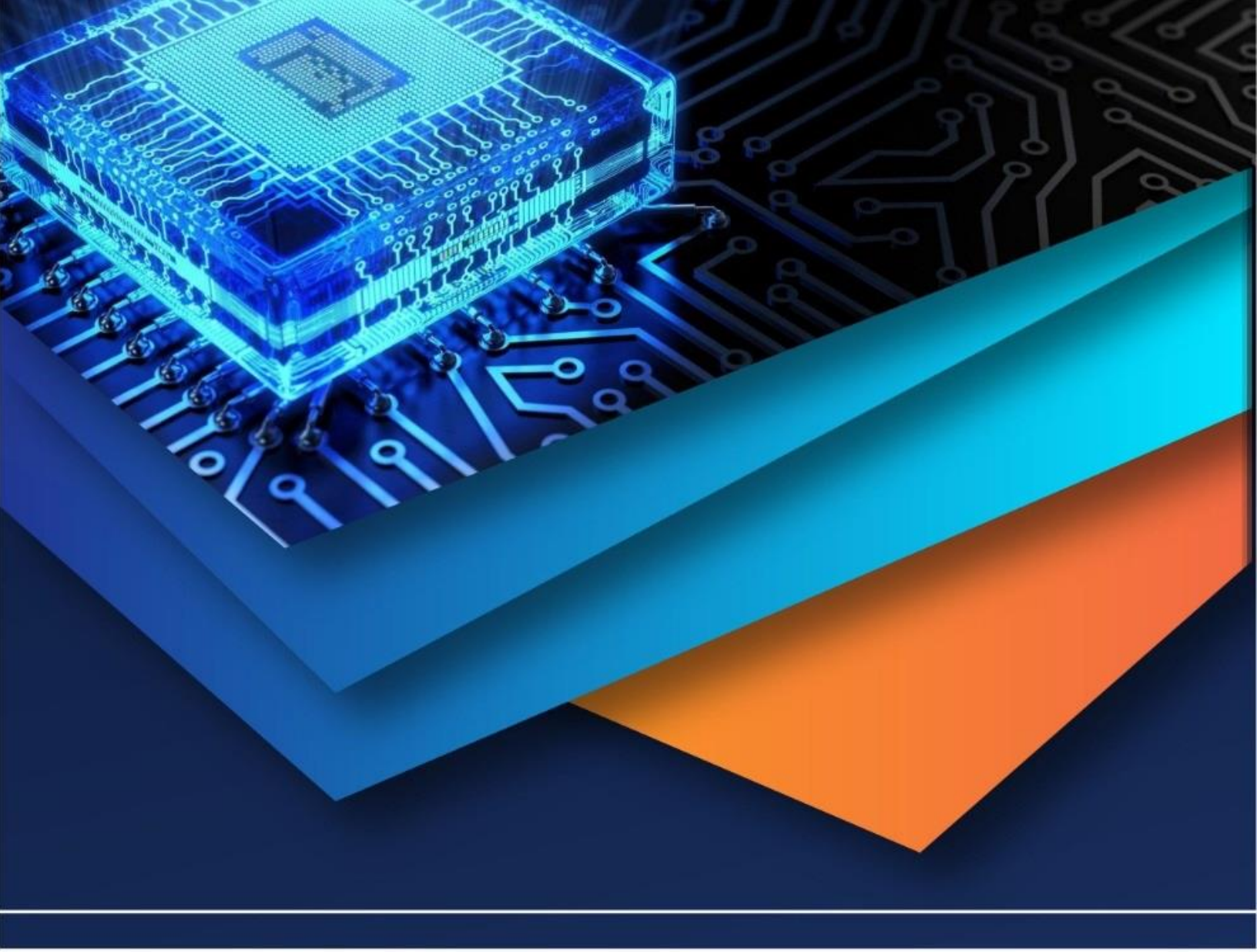
VI. ACKNOWLEDGEMENTS

In performing this project, I needed to take the assistance and direction of some regarded people, who merit my most prominent appreciation. The consummation of this task gives me much delight. Many individuals, particularly my schoolmates, have offered significant remark ideas on this proposition which gave me a motivation to work on my task. I thank every individual for their assistance straightforwardly and in a roundabout way to finish my task.

REFERENCES

[1] Omprakash, Dr Khanna Samrat Vivekanand. "Concept of Search Engine Optimization in Web Search Engine." International Journal of Advanced Engineering Research and Studies 1, no. 1 (2011): 235-237.
[2] Moral, Cristian, Angélica de Antonio, Ricardo Imbert, and Jaime Ramírez. "A survey of stemming algorithms in information retrieval." Information Research: An International Electronic Journal 19, no. 1 (2014): n1.
[3] Huu, Quynh Nguyen, Ha Nguyen Thi Thu, and Tao Ngo Quoc. "An efficient content based image retrieval method for retrieving images." International Journal Of Innovative Computing Information And Control 8, no. 4 (2012): 2823-2836.
[4] Lott, Brian. "Survey of Keyword Extraction Techniques." UNM Education(2012).
[5] Datta, Ritendra, Dhiraj Joshi, Jia Li, and James Z. Wang. "Image retrieval: Ideas, influences, and trends of the new age." ACM Computing Surveys (CSUR) 40, no. 2 (2008).
[6] Jeh, Glen, and Jennifer Widom. "SimRank: a measure of structural-context similarity." In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 538-543. ACM, 2002.
[7] Datta, Ritendra, Jia Li, and James Z. Wang. "Content-based image retrieval: approaches and trends of the new age." In Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, pp. 253-262. ACM, 2005.
[8] "Getting Started with Keyword Extraction | Text Mining Online | Text Analysis Online | Text Processing Online", Textminingonline.com, 2015. [Online]. Internet: http://textminingonline.com/getting-started-with-keyword-extraction. [Accessed: 13- Mar- 2016].
[9] NLP keyword extraction tutorial with RAKE and Maui", Airpair.com, 2016. [Online]. Internet: https://www.airpair.com/nlp/keyword-extraction-tutorial. [Accessed: 04- Feb- 2016].
[10] K. Ganesan, "Text Mining, Analytics & More: All About Stop Words for Text Mining and Information Retrieval", Text-analytics101.com, 2016. [Online]. Internet: http://www.text-analytics101.com/2014/10/all-about-stop-words-for-text-mining.html. [Accessed: 06- Feb- 2016].
[11] "Rainbow", Weka.sourceforge.net, 2015. [Online]. Internet: http://weka.sourceforge.net/doc.dev/weka/core/stopwords/Rainbow.html. [Accessed: 09- Oct- 2015].
[12] "Rainbow", Cs.cmu.edu, 2015. [Online]. Internet: http://www.cs.cmu.edu/~mccallum/bow/rainbow/. [Accessed: 10- Oct- 2015].
[13] "Google Images", Wikipedia, 2015. [Online]. Internet: https://en.wikipedia.org/wiki/Google_Images. [Accessed: 04- Apr- 2016].

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)