**KEYWORD EXTRACTION FROM**

**THE NEWS FEED OR BLOGSPOTS WITH IBM CLOUD**

**1.INTRODUCTION**

**1.1 overview**

Keyword extraction (also known as keyword detection or keyword analysis) is a text analysis technique that automatically extracts the most used and most important words and expressions from a text. It helps summarize the content of texts and recognize the main topics discussed.

Keyword extraction simplifies the task of finding relevant words and phrases within unstructured text. From simple statistical approaches that detect keywords by counting word frequency, to more advanced machine learning approaches that create even more complex models by learning from previous examples.

Keyword extraction is tasked with the automatic identification of terms that best describe the subject of a document. Key phrases, key terms, key segments or just keywords are the terminology which is used for defining the terms that represent the most relevant information contained in the document.

**1.2 Purpose:**

The keyword extraction service is used to extract key words and phrases from text, such as an email or chat. The algorithm parses the text into sentences and removes the most frequent but least useful words for determining meaning.

In a nutshell, keyword extraction is a methodology to automatically detect important words that can be used to represent the text and can be used for topic modeling. This is a very efficient way to get insights from a huge amount of unstructured text data.

The keyword extraction process helps us in identifying the important words. It also effective in topic modeling tasks. You can know a lot about your text data by only a few keywords. These keywords will help you to determine whether you want to read an article or not.

**2.LITERAURE SURVEY**

**2.1 Existing problem:**

Considering that more than 80% of the data we generate every day is unstructured ― meaning it’s not organized in a predefined way, making it extremely difficult to analyze and process. You can read texts and identify key terms manually, but it would be extremely time-consuming. Automating this task gives you the freedom to concentrate on other parts of your job.

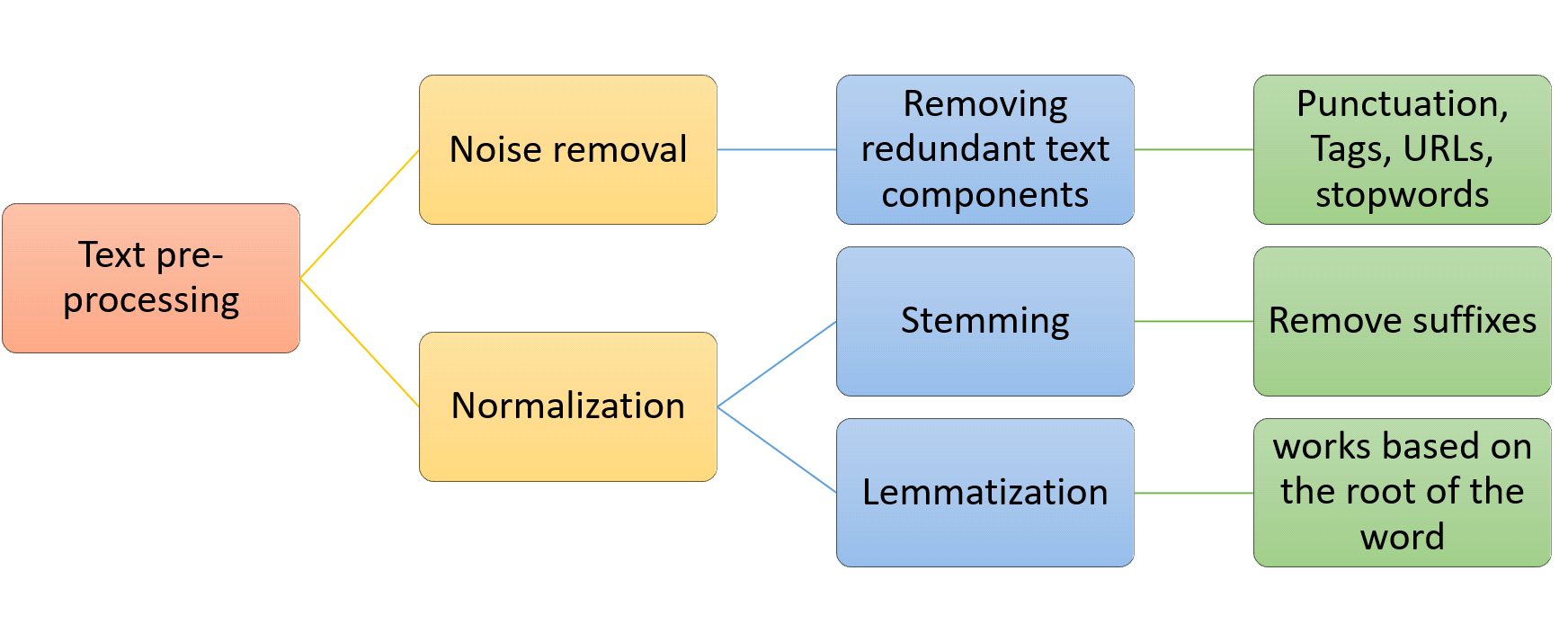
**2.2 Proposed solution**

Automated keyword extraction allows you to analyze as much data as you want Keyword extraction acts based on rules and predefined parameters. You don’t have to deal with inconsistencies, which are common in manual text analysis

keyword extraction may be the key to finding relevant keywords within massive sets of data (like new articles, papers, or journals) without having to actually read the entire content.

**3)THEORITICAL ANALYSIS**

**3.1 Block Diagram**



**3.2 Hardware/Software Designing**

System Structure Design and Software Architecture Introduction of System Structure and Functions. The primary functions based on the system requirements analysis are registration/sign-in, task management and inquire, data import/export, keyword extraction and keyword management.

1. Registration/sign-in module: the module requires personal information of users to record how users operate the system.

2. Task management and inquire module: every extraction of keyword has to be implemented Proceedings of the 2nd International Conference On Systems Engineering and Modeling (ICSEM-13) Published by Atlantis Press, Paris, France. © the authors, 2013 0720 under task management. Users can select a task or open task. The module is also featured with historical task query. A task window records created task name, address and creation time of the user. Figure 1 Flowchart of keyword extraction and management system.

3. Word segmentation module: it refers to corpus segment, which is the foundation work for keyword extraction. The module is designed to corpus segment in domain text and prepare for keyword extraction. 4

4. Keyword extraction module: it is keyword extraction which offers TF-IDF method, C-value method and combination method.

5. Keyword management module: the module offers such functions as keyword upload, query, addition, deletion and modification extraction and takes full advantage of word

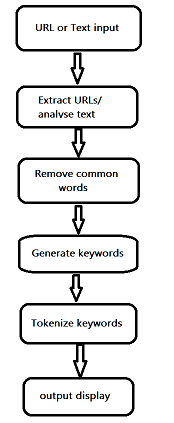
distribution information in the overall text

**4)EXPERIMENTAL INVESTIGATIONS**

**Analysis or the Investigation made while working on the Project**

As an initial and necessary step, data preprocessing cleans the noises to ensure the high quality of data to enhance the performance of models. Data cleaning, the first critical task for any data relational project, also called data cleansing [66], removes the format tags and errors for the next step in data analysis. The specific preprocessing in this work includes data cleaning, and selection of the word features for NB, LogR, and SVM. For our two datasets, the text paragraphs are collected from websites with some HTML tags and other format problems. We use the regular expression and NLTK to clean the format tags and other data noises, followed by removing the stop words and punctuation, converting all capital letters, and restoring abbreviations.

**5.FLOWCHART**

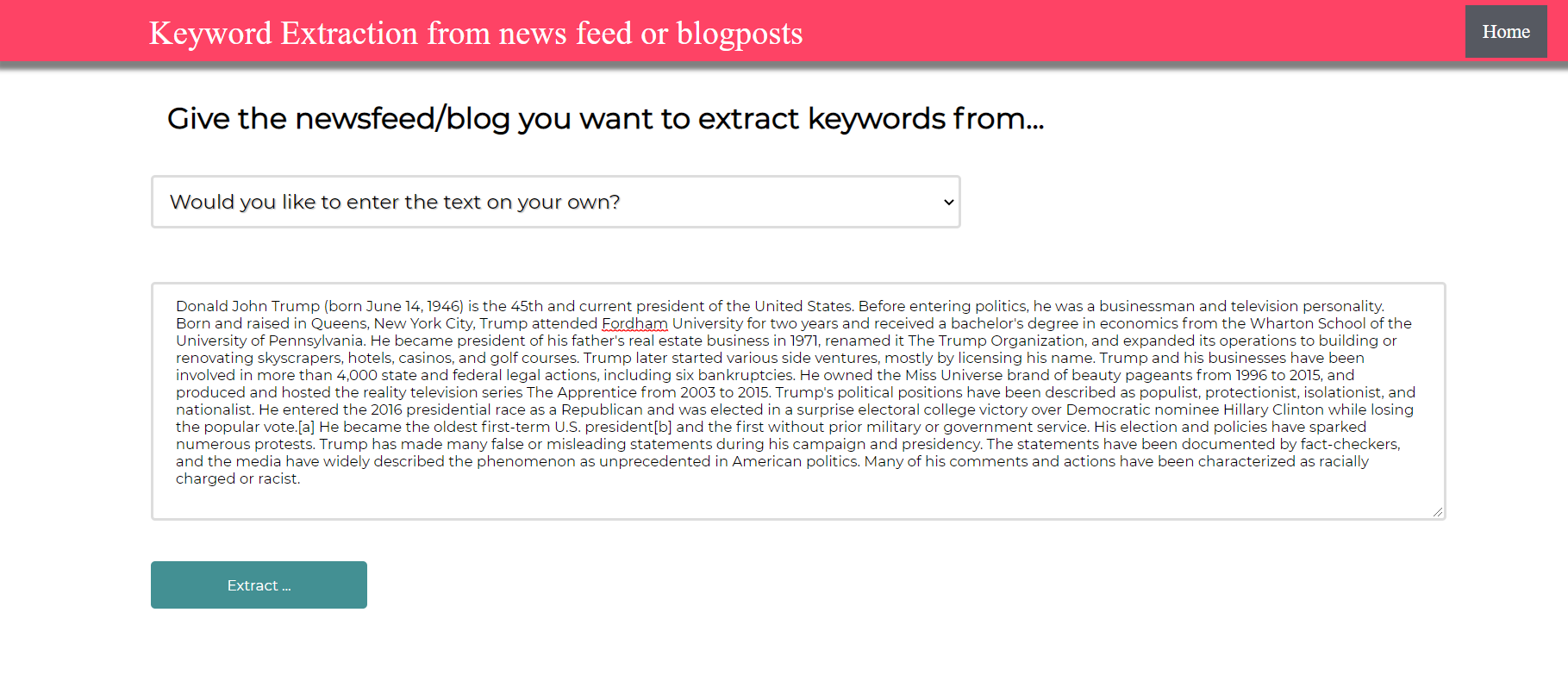


**RESULT**

We have buit our flask application and deployed it into ibm cloud.

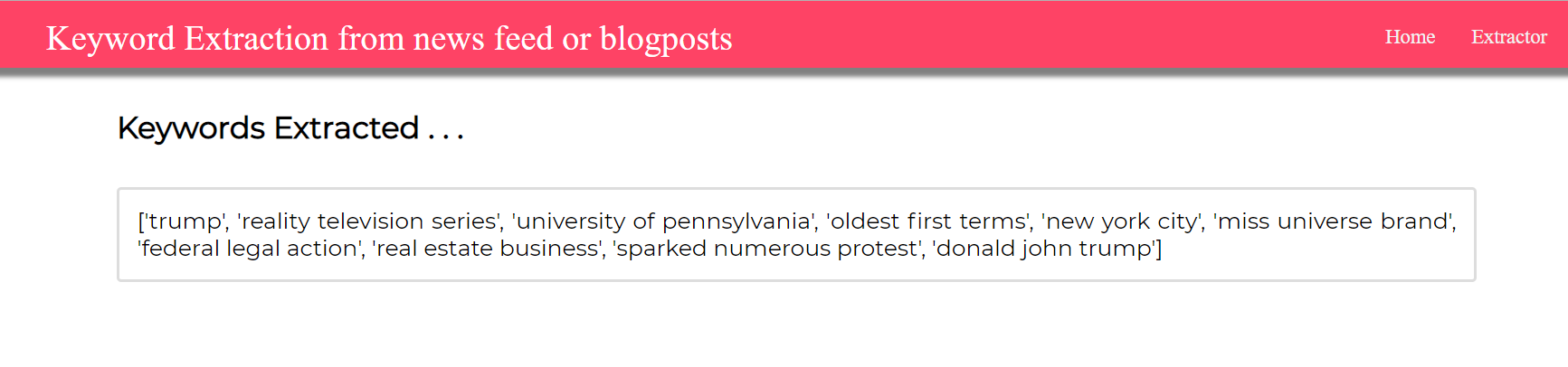
After the deployment we have given a text input to test the application.

**INPUT:**



After clicking on extract we have successfully extracted keywords out of the test input.

**OUTPUT:**



**ADVANTAGES:**

* Scalability: Automated keyword extraction allows you to analyze as much data as you want.

#### Consistent criteria:Keyword extraction acts based on rules and predefined parameters.

#### Real-time analysis:You can perform keyword extraction on social media posts, customer reviews, surveys, or customer support tickets in real-time, and get insights about what’s being said about your product as they happen and follow them over time.

**DISADVANTAGES:**

* Not accurate enough. Result is similar,so there are many words that are not keyword had listed.
* N-grams are tolerant of textual errors.
* Repetitive words are often considered as keywords.

**8.APPLICATIONS**

Every day, [internet users create 2.5 quintillion bytes of data](https://techjury.net/stats-about/big-data-statistics/). Social media comments, product reviews, emails, blog posts, search queries, chats, and so on. We have all sorts of unstructured text data at our disposal. The question is, how do we sort the chaos to find what’s relevant?

Keyword extraction can help you obtain the most important keywords or key phrases from a given text without having to actually read a single line.

Whether you are a product manager trying to analyse a pile of product reviews, a customer service manager analysing customer interactions, or a researcher that has to go through hundreds of online papers about a specific topic, you can put keyword extraction to use to easily understand what a text is about.

Thanks to keyword extraction, teams can be more efficient and take full advantage of the power of data. You can say goodbye to manual and repetitive tasks (saving countless human hours) and get access to interesting insights that will help you transform unstructured data into valuable knowledge.

Wondering what you can analyse keyword extraction? Here are some common use cases and applications:

1. [Social media monitoring](https://monkeylearn.com/keyword-extraction/#social-media-monitoring)
2. [Brand monitoring](https://monkeylearn.com/keyword-extraction/#brand-monitoring)
3. [Customer service](https://monkeylearn.com/keyword-extraction/#customer-service)
4. [Customer feedback](https://monkeylearn.com/keyword-extraction/#customer-feedback)
5. [Business intelligence](https://monkeylearn.com/keyword-extraction/#business-intelligence)
6. [Search engine optimization (SEO)](https://monkeylearn.com/keyword-extraction/#seo)
7. [Product analytics](https://monkeylearn.com/keyword-extraction/#product-analytics)
8. [Knowledge management](https://monkeylearn.com/keyword-extraction/#knowledge-management)

**9.CONCLUSION**

Keyword extraction is a text analysis technique that automatically extracts the most used and most important words and expressions from a text. It helps summarize the content of texts and recognize the main topics discussed. More than 80% of the data we generate every day is unstructured meaning it's not organized in a predefined way.

System Structure Design and Software Architecture. Introduction of System Structure and Functions. The primary functions based on the system requirements analysis are registration/sign-in, task management and inquire, data import/export, keyword extraction and keyword management.

Veena’s part

Keyword extraction can help you obtain the most important keywords or key phrases from a given text. Teams can be more efficient and take full advantage of the power of data with keyword extraction.

**10.Future Scope**

We can conclude that different methods and techniques could be used to extract keywords. Different approaches such as supervised and unsupervised are used in different scenarios. By keyword extraction we could easily find the words that are highly signified and of great importance. Some techniques resulted in a better precision rate while some has less results for this parameter. Our proposal is that the TFIDF(In information retrieval, TFIDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus) approach could provide a better result and can be used effectively. Further in the future other parameters such as purity, entropy could be considered of great importance and sentiments can be used to extract the keywords from the twitter. More in future the sentiments and emotions of the tweet can also be considered as well as some heuristic approach can be used that eliminate the noisy data which could further reduce the computational time of the algorithm.

**REFERENCES**

1.Zhao, H., & Zeng, Q. (2013). Micro-blog keyword extraction method based on graph model and semantic space.

2. .Marujo, Ling, Trancoso, Dyer, C., Black, Alan,Gershman, Carbonell & J. Automatic Keyword Extraction on Twitter. Volume 2: Short Papers, 637.

3. Kim, Rho,& Hwang, E. J. (2013). Detecting trend and bursty keywords using characteristics of Twitter stream data. International Journal of Smart Home, 7(1), 209-220.

4. Beliga, S. (2014). Keyword extraction: a review of methods and approaches. University of Rijeka, Department of Informatics, Rijeka.

5. Chen, Lin, S. J. (2010). Automatic keyword prediction using Google similarity distance. Expert Systems with Applications, 37(3), 1928-1938.

6. [NLP keyword extraction tutorial with RAKE and Maui (airpair.com)](https://www.airpair.com/nlp/keyword-extraction-tutorial)

7. Li, C., Sun, A., Weng, J., & He, Q. (2015). Tweet segmentation and its application to named entity recognition. Knowledge and Data Engineering, IEEE Transactions on, 27(2), 558-570.

8. .Abilhoa, W. D., & de Castro, L. N. (2014). A keyword extraction method from twitter messages represented as graphs. Applied Mathematics and Computation, 240, 308-325.

9. .Hong, B., & Zhen, D. (2012). An extended keyword extraction method.Physics Procedia, 24, 1120-1127.

10. 1.Y. Matsuo andM. Ishizuka(2004). Keyword extraction from a single document using word coccurrence statistical information. International Journal on Artificial Intelligence Tools,13:200