

# **KEYWORD EXTRACTION FROM SHORT TEXT MESSAGE USING IBM KNOWLEDGE STUDIO**

A MINI PROJECT REPORT Submitted to

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY,  
HYDERABAD**

*In partial fulfillment of the requirements for the award of the degree of*

**BACHELOR OF TECHNOLOGY**

In

**COMPUTER SCIENCE AND ENGINEERING**

BY

E. YASHWANTH (18UK1A05M8)

SB. SRIVASTAVA (18UK1A05H2)

M.PRASHANTH (18UK1A05K9)

P. SAI CHARAN (18UK1A05L2)

Under the esteemed guidance of

**Mr. P. ILANNA**

Assistant Professor



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**VAAGDEVI ENGINEERING COLLEGE**

Affiliated to JNTU, HYDERABAD

BOLLIKUNTA, WARANGAL - 506005.

2018-2022

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**  
**VAAGDEVI ENGINEERING COLLEGE**

Affiliated to JNTU, HYDERABAD  
BOLLIKUNTA, WARANGAL-506005.

2018-2022



**CERTIFICATE**

This is to certify that the Mini Project Report entitled “**KEYWORD EXTRACTION FROM SHORT TEXT MESSAGES USING IBM KNOWLEDGE STUDIO**” is being submitted by E.YASHWANTH (18UK1A05M8), SB.SRIVASTAVA (18UK1A05H2), M.PRASHANTH (18UK1A05K9), P. SAI CHARAN (18UK1A05L2) in partial fulfillment of the requirements for the award of the Degree in Bachelor of Technology in computer science and engineering during the academic year 2018-2022

**Mr. P.JLANNA**

Project Guide

**Dr. R. NAVEEN KUMAR**

Head of the Department

**EXTERNAL EXAMINER**

## ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. Prasad Rao**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this mini project in the institute.

We extend our heartfelt thanks to **Dr. R. Naveen Kumar**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the mini project.

We express heartfelt thanks to the Mini Project Coordinator, **Ms. G. Aruna Kranthi**, Assistant Professor, Department of CSE for her constant support and giving necessary guidance for completion of this mini project.

We express heartfelt thanks to the guide, **Ms. P. ILANNA**, Assistant Professor, Department of CSE for her constant support and giving necessary guidance for completion of this mini project.

We express our sincere thanks and gratitude to **The Smart Bridge**, for providing internship.

Finally, we express our sincere thanks and gratitude to our family members, friends for their encouragement and outpouring their knowledge and experiencing throughout thesis.

## **ABSTRACT**

Due to the considerable growth of the volume of text documents on the Internet and in digital libraries, manual analysis of these documents is no longer feasible. Having efficient approaches to keyword extraction in order to retrieve the ‘key’ elements of the studied documents is now a necessity. Keyword extraction has been an active research field for many years, covering various applications in Text Mining, Information Retrieval, and Natural Language Processing, and meeting different requirements.

However, it is not a unified domain of research. In spite of the existence of many approaches in the field, there is no single approach that effectively extracts keywords from different data sources. This shows the importance of having a comprehensive review, which discusses the complexity of the task and categorizes the main approaches of the field based on the features and methods of extraction that they use. This paper presents a general introduction to the field of keyword extraction. This review is intended to help readers find their way around all the works related to keyword extraction and guide them in choosing or designing a method that is appropriate for the application they are targeting.

.

# **CONTENTS**

## **LIST OF CHAPTERS**

### **1.INTRODUCTION**

1.1 OVERVIEW OF THE PROJECT

1.2 PURPOSE

### **2. LITERATURE SURVEY**

### **3. BLOCK DIAGRAM**

3.1 Hardware/Software Designing

### **4. EXPERIMENTAL INVESTIGATION**

### **5. FLOW CHART**

### **6. RESULT**

### **7. ADVANTAGES AND DISADVANTAGES**

### **8. APPLICATIONS**

### **9. CONCLUSION**

### **10. FUTURE SCOPE**

# 1. INTRODUCTION

Now a day due to the considerable growth of the volume of text documents on the Internet and in digital libraries, manual analysis of these documents is no longer feasible. We need to quickly go through large amounts of textual information to find out documents related to our interests and this document space is growing on a daily basis at an overwhelming rate. Now days it is common to store several million web-pages and hundreds of thousands of text files. Analyzing such huge quantities of data can be made easier if we can have a subset of words (Keywords) which can provide us with the main features, concept, theme etc of the document. Appropriate keywords can serve as a highly concise summary of a document and help us in easily organize documents and retrieve them based on their content. Keywords are used in academic articles to give an idea to the reader about the content of the article. In a textbook they are useful for the readers to identify and retain the main points in their mind about a particular section. As keywords represent the main theme of a text, they can be used as a measure of similarity for text clustering.

## 1.1. OVER VIEW

Now a day due to the considerable growth of the volume of text documents on the Internet and in digital libraries, manual analysis of these documents is no longer feasible. We need to quickly go through large amounts of textual information to find out documents related to our interests and this document space is growing on a daily basis at an overwhelming rate.

Text analysis is a machine learning technique that allows organizations to automatically extract and classify text data, such as tweets, emails, support tickets, product reviews, and survey responses. Organizations might want to extract specific information, like keywords, names, or company information from a bunch of emails, or categorize survey responses by sentiment and topic. Current natural language processing techniques cannot extract or interpret data that is domain or industry-specific because entities have different meanings in different domains. The best answer to such a problem is IBM's Watson Knowledge Studio.

Creating Necessary Services

Upload the entity types in knowledge studio

Uploading the training samples

Annotate (labeling) the training data with entity types

Deploy the model in NLU (Natural Language Understanding ) Service

Building UI(user interface) for analysis for SMS

The Natural Language Understanding service is a collection of text analysis functions that derive semantic information from your content. You can input text, HTML, or a public URL, and leverage sophisticated natural language processing techniques to get a quick high-level understanding of your content and obtain detailed insights.

With Natural Language Understanding, developers can analyze semantic features of input text and extract metadata from content, such as categories, concepts, emotion, entities, keywords, metadata, relations, semantic roles, and sentiment. With custom annotation models developed using IBM Watson Knowledge Studio, you can further customize the service to identify domain-specific entities and relations in your content. Natural Language Understanding can be useful in many scenarios that demand rapid analysis of unstructured text without requiring in-depth natural language processing expertise. For example, you can monitor sentiment and emotion in customer support chat transcripts, or you can quickly categorize blog posts and sort them based on general concepts, keywords, and entities.

## **1.2 PURPOSE**

Keyword extraction 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. 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, to buy an item from market or not, by just looking at offers and reviews given by the customers.

The Extraction Of keyword not only reduces the time of doing work but also saves lot of time. This process increases the effectiveness of the doing work.

## 2. LITERATURE SURVEY

In this paper, keyword extraction from short text message using IBM Watson studio. In this designated area. The main task of extracting keyword is to know about the overview of the topic. In this topic some amount of data trained before a extracting a keyword. Here text is given to the trained model and it returns a keyword. For creating a services different services are used. The services we created are to solve the problem

Creating knowledge studio services

Build Custom Models

Creating Web-Application

Extracting keyword from the designated area is not an important task, but the important task is to extract the without any mistake. To perform this task knowledge studio created services some of them are natural language understanding and node red services and in building custom models some the services that are created is workspace services. This important task is done by this model that is to train the annotations that we were created. The overall aim is to extract the accurate keyword.

### 2.1 Existing Problem

Text analysis is a machine learning technique that allows organizations to automatically extract and classify text data, such as tweets, emails, support tickets, product reviews, and survey responses. Organizations might want to extract specific information, like keywords, names, or company information from a bunch of emails, or categorize survey responses by sentiment and topic. Current natural language processing techniques cannot extract or interpret data that is domain or industry-specific because entities have different meanings in different domains. The best answer to such a problem is IBM's Watson Knowledge Studio. Consider a case where we need to extract entities present in a commercial SMS. For example:

Get 1+1 on booking Cinepolis Cinemas tickets on Paytm.

Use code: LOVEFEB to avail the offer.

Max. Cashback\* Rs.500. Click <http://m.p-y.tm/mcn>. \*T&C apply.



The example above has a few interesting entities which could not be extracted with conventional NLP techniques, but by using Watson services we can find out the following:

What is the offer?

Who is the merchant?

What is the offer name?

What is the offer's validity period?

What is the merchant's phone number?

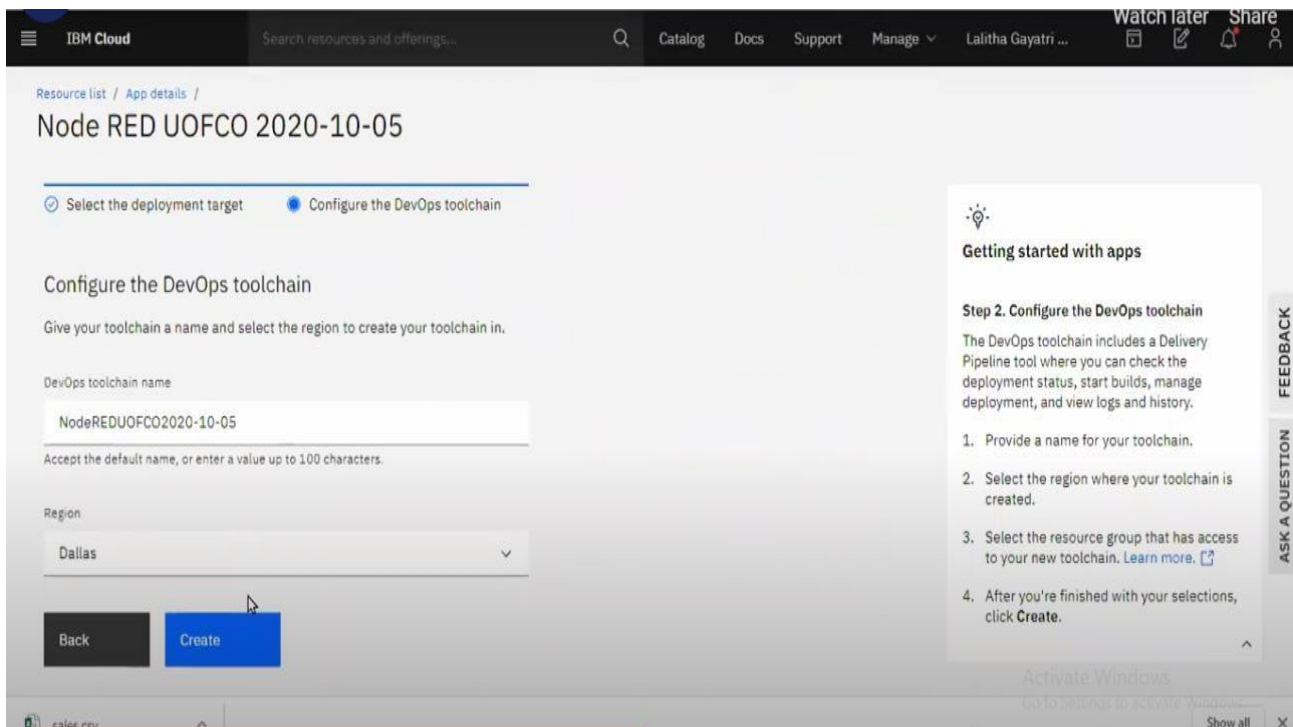
What is the merchant's website?

## **2.2 PROPOSED SOLUTION**

In this project, you will be building a custom model to analyze SMS messages with Watson Knowledge Studio (WKS) and Watson's Natural Language Understanding (NLU) capability to extract entities in the data. The project flow is followed as shown

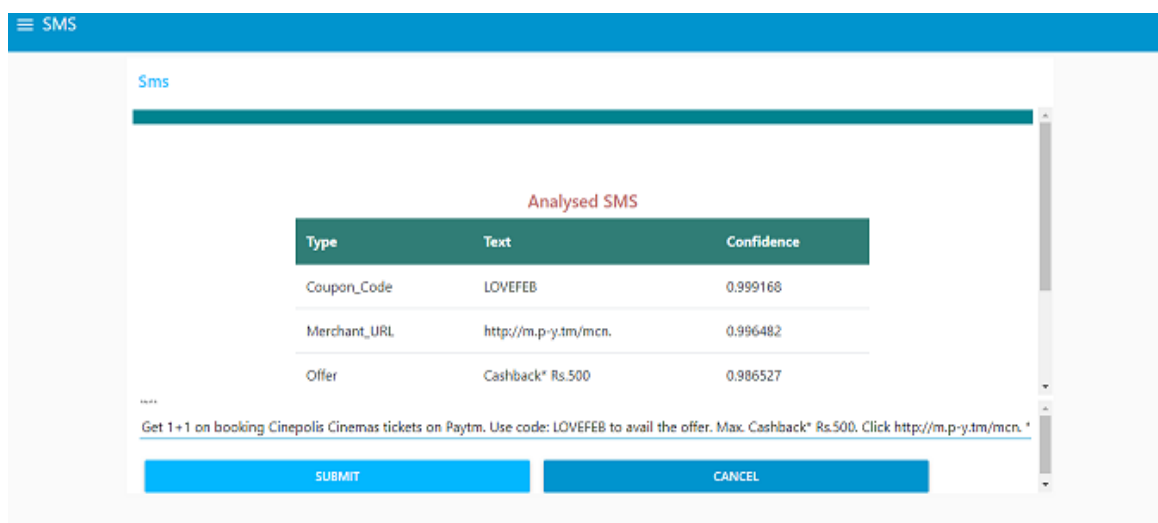
- Create IBM account
- Log in to IBM Account
- Create Watson Knowledge Studio and Watson Natural language Understanding Services
- Load entity type system and Training data files into Watson Knowledge Studio.
- Generate model by training and evaluating data.
- The WKS model is deployed to Watson NLU.
- Build Node-RED application (UI)
- A user provides an SMS message to the app for analysis.
- The SMS message is analyzed by Watson NLU for processing and returns extracted domain-specific entities

First we need to create Watson knowledge studio service. The below steps are followed to create Watson studio services. Create Knowledge Studio Service and next create natural language understanding and at last create node red services. Now click on create. In the next step we build custom models. IN custom model we need to create workspace and upload

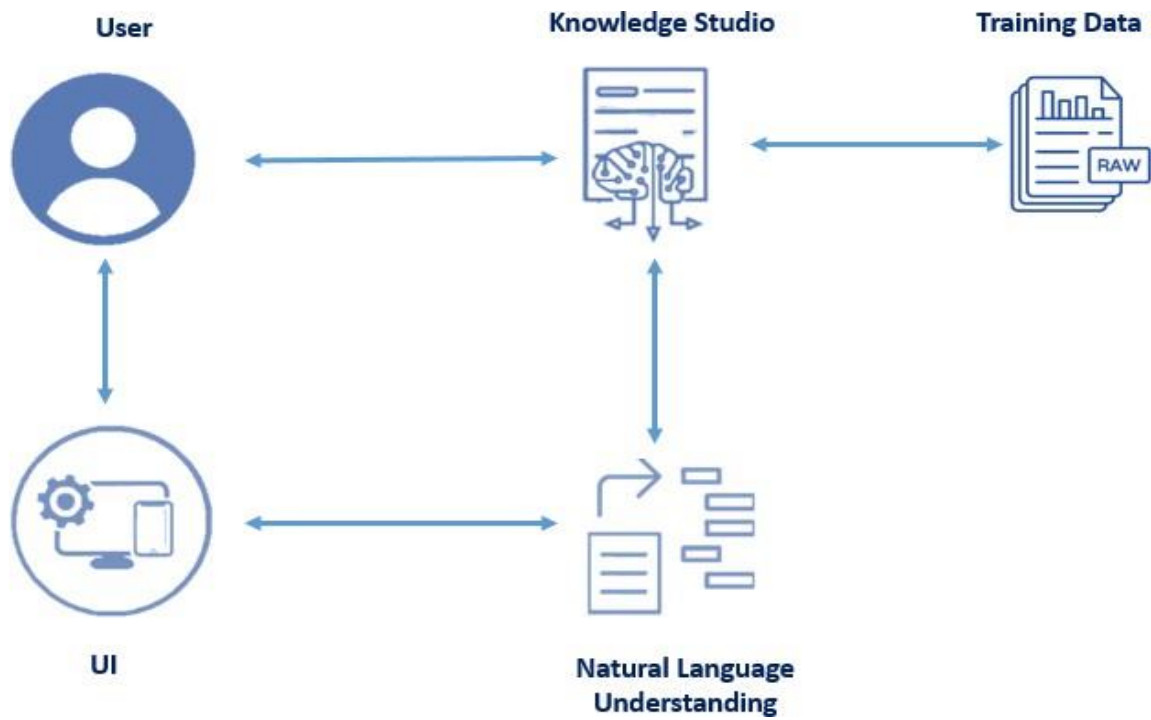


entity types after uploading the entity types. Now we need to create annotations and after annotating the document set, we need to train the model nothing but training the annotation.

In the next step we need create we application. In this activity, you will be building a web application where the user enters an SMS and the entities in the message are analyzed and returned on to the user-interface as shown below.



### 3. BLOCK DIAGRAM



#### 3.1 SOFTWARE DESINING

- Knowledge Studio
- Natural Language Understanding &
- Node-Red Services

## 4. EXPERIMENTAL INVESTIGATION

SMS

Sms

Analysed SMS

Type	Text	Confidence
Coupon_Code	LOVEFEB	0.999168
Merchant_URL	<a href="http://m.p-y.tm/mcn">http://m.p-y.tm/mcn</a>	0.996482
Offer	Cashback* Rs.500	0.986527

code

Get 1+1 on booking Cinepolis Cinemas tickets on Paytm. Use code: LOVEFEB to avail the offer. Max. Cashback\* Rs.500. Click <http://m.p-y.tm/mcn>. \*

SUBMIT

CANCEL

## 5. FLOW CHART



## 6. RESULT

When a text or message is entered through a Web-based application it returns the keyword as shown in the figure.

SMS

Sms

Analysed SMS

Type	Text	Confidence
Coupon_Code	LOVEFEB	0.999168
Merchant_URL	<a href="http://m.p-y.tm/mcn">http://m.p-y.tm/mcn</a>	0.996482
Offer	Cashback* Rs.500	0.986527

Get 1+1 on booking Cinepolis Cinemas tickets on Paytm. Use code: LOVEFEB to avail the offer. Max. Cashback\* Rs.500. Click <http://m.p-y.tm/mcn> \*

SUBMIT

CANCEL

## **7. ADVANTAGES & DISADVANTAGES**

### **ADVANTAGES:**

- Easy to understand and launch the models.
- No additional code is required to build a model.
- All the services are predefined.
- Any of the domain problem can be solved with the help of the services provide by IBM Watson Studio.
- Once created services can be used any number of times.
- Large amount of data can be analyzed easily.
- No additional services are required to send an SMS.
- No need to differentiate data in the dataset.
- Large amount trained models are already present.

### **DISADVANTAGES**

- To use services, need to pay money
- Tough to select most appropriate key words
- Training is required to create services
- Lots of data is needed to train the model.
- Difficult to recognize problem.
- Training of data takes lot of time when compared to other techniques

## **8. APPLICATIONS**

- Expressed by consumers in reviews, feedback, and suggestions, and gain valuable insights and product development keys.
- By making it simple to find, organize, and access relevant stuff, you may learn new things.
- By automating the ticket tagging process, your customer support personnel can focus on really fixing issues.
- Examining how public opinion changes over time on any particular product line or service.
- Finding Keywords to maintain track on social media trends and trending topics as well as keeping an eye on your competitors' online presence.



## **9. CONCLUSIONS**

Online comments from students can provide an effective way of teaching and receiving feedback and extracting the keyword from text messages will easy to identify the overall meaning. How to automatically extract insightful information from these comments are important. In this work, we rely on IBM WATSON STUDIO to extract topic keywords from the given topic. By doing so, we can summarize the information of online subject posts for improving the quality of learning and teaching in higher education and user can differentiate wether message is fake or real looking at the extracted message.

In this paper we have discussed the steps required to create different IBM services which are useful for extracting a keyword from short text message. The selected methods ranged from traditional approaches to deep learning algorithms.

## **10. FUTURE SCOPE**

- Can implement to recognize either the tweets are real or fake
- Can implement to extract offers from the message
- Can implement to extract keywords from news paper
- Can implement when large amount of data is present to evaluate.