Natural Language Processing Mini-Project ReportOn

Regional Language Text Summarisation and Keyword Extraction using Gensim

By

- 1. Arnav Sankhe (BE B848)
- 2. Aman Sarawgi (BE B849)
- 3. Amit Kumar Sharma (BE B855)



Department of Computer Engineering
University of Mumbai
2019-2020



Juhu-Versova Link Road Versova, Andheri(W), Mumbai-53

CERTIFICATE

Department of Computer Engineering

This is to certify that

- 1. Arnav Sankhe
- 2. Aman Sarawgi
- 3.Amit Kumar Sharma

Have satisfactory completed this project entitled

Regional Language Text Summarisation and Keywords Extraction using Gensim

Towards the partial fulfilment of the

FOURTH YEAR BACHELOR OF ENGINEERING IN (COMPUTER ENGINEERING)

as laid by University of Mumbai.

Guide H.O.D.

Prof. Suresh Mestry Dr. Satish Y. Ket

Declaration

We wish to state that the work embodied in this synopsis titled "Regional Language Text Summarisation and Keyword Extraction using Gensim" forms our own contribution to the work carried out under the guidance of "Prof. Suresh Mestry" at the Rajiv Gandhi Institute of Technology.

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Arnav Sankhe (B-848)

Aman Sarawgi (B-849)

Amit Kumar Sharma (B-855)

Abstract

In this new era, where tremendous information is available in various languages on the internet, it is most important to provide the improved mechanism to extract the information quickly and most efficiently. It is very difficult for human beings to manually extract the summary of a large documents of text. There are plenty of text materials in regional language available on the internet. So there is a problem of searching for relevant documents from the number of documents available, and absorbing relevant information from it. In order to solve the above two problems, we have created a text summarisation and keyword extraction system using Gensim. This system identifies the most important meaningful information and keywords in the given documents.

Contents

List of Figures	6
List of Algorithms	7
Introduction	
Text Summarisation	
Gensim	
Tkinter	8
Proposed Solution	9
Text summarisation with genism	9
TextRank model	9
PageRank algorithm	10
TextRank	10
Software requirements	11
Implemenation	12
• Code	
• Input	13
• Output	14
Results	16
Original Text	16
Summarised Text	17
Conclusion	18
References	19

List of Figures

1.	Text summarisation and keywords extraction code (part 1)	. 12
2.	Text summarisation and keywords extraction code (part 2)	. 12
3.	Input Text	13
4.	GUI of the system	. 14
5.	Output generated	. 15
6.	Input text for comparison	.16
7.	Summarised Text	17

List of Algorithms

1. Gensim

Introduction:

Text Summaristion:

Before going to the Text summarization, first, we have to know what a summary is. A summary is a text that is produced from one or more texts, that conveys important information in the original text, and it is of a shorter form. The goal of automatic text summarization is to present the source text into a shorter version with semantics. The most important advantage of using a summary is, it reduces the reading time. Text Summarization methods can be classified into extractive and abstractive summarization. An extractive summarization method consists of selecting important sentences, paragraphs, etc. from the original document and concatenating them into a shorter form. Abstractive summarization is an understanding of the main concepts in a document and then express those concepts in clear natural language.

There are two different groups of text summarization: indicative and informative. Inductive summarization only represents the main idea of the text to the user. The typical length of this type of summarization is 5 to 10 percent of the main text. On the other hand, the informative summarization systems give concise information of the main text. The length of informative summary is 20 to 30 percent of the main text.

Gensim:

We are using Gensim python library. Gensim is a free Python library designed to automatically extract semantic topics from documents. The gensim implementation is based on the popular TextRank algorithm. It is an open-source vector space modelling and topic modelling toolkit, implemented in the Python programming language, using NumPy, SciPy and optionally Cython for performance.

Tkinter:

We are using Tkinter module for designing the GUI of our system. The Tkinter module ("Tk interface") is the standard Python interface to the Tk GUI toolkit from Scriptics (formerly developed by Sun Labs). Both Tk and Tkinter are available on most Unix platforms, as well as on Windows and Macintosh systems. Starting with the 8.0 release, Tk offers native look and feel on all platforms. Tkinter consists of a number of modules. The Tk interface is provided by a binary extension module named "tkinter". This module contains the low-level interface to Tk, and should never be used directly by application programmers. It is usually a shared library (or DLL), but might in some cases be statically linked with the Python interpreter. The public interface is provided through a number of Python modules. The most important interface module is the Tkinter module itself. To use Tkinter, all you need to do is to import the Tkinter module:

Proposed Solution:

Gensim is a free Python library designed to automatically extract semantic topics from documents. The gensim implementation is based on the popular TextRank algorithm. It is an open-source vector space modelling and topic modelling toolkit, implemented in the Python programming language, using NumPy, SciPy and optionally Cython for performance.

Text Summarisation with Gensim (TextRank algorithm)-

We use the summarization.summarizer from gensim. This summarising is based on ranks of text sentences using a variation of the TextRank algorithm. TextRank is a general purpose, graph based ranking algorithm for NLP. TextRank is an automatic summarisation technique. Graph-based ranking algorithms are a way for deciding the importance of a vertex within a graph, based on global information recursively drawn from the entire graph.

TextRank Model -

The basic idea implemented by a graph-based ranking model is that of voting or recommendation.

When one vertex links to another one, it is basically casting a vote for that vertex. The higher the number of votes cast for a vertex, the higher the importance of that vertex.

Text as a graph -

We have to build a graph that represents the text, interconnects words or other text entities with meaningful relations.

TextRank includes two NLP tasks-

- 1. Keyword extraction task
- 2. Sentence extraction task

Keyword Extraction -

The task of keyword extraction algorithm is to automatically identify in a text a set of terms that best describe the document.

The simplest possible approach is to use a frequency criterion.

HOWEVER, this leads to poor results.

The TextRank keyword extraction algorithm is fully unsupervised. No training is necessary.

Sentence Extraction –

TextRank is very well suited for applications involving entire sentences, since it allows for a ranking over text units that is recursively computed based on information drawn from the entire text.

To apply TextRank, we first build a graph associated with the text, where the graph vertices are representative for the units to be ranked. The goal is to rank entire sentences, therefore, a vertex is added to the graph for each sentence in the text.

PageRank Algorithm -

It is the foundation of TextRank.

PageRank used by Google search.

Used to compute the rank of web pages. It is not named after its use (ranking pages) but after its creator Larry Page.

Fundamentals -

Important pages are linked by important pages.

The PageRank value of a page is the probability of a user visiting that page.

In TextRank, the only difference is that we consider sentences instead of pages.

TextRank-

TextRank is an unsupervised algorithm for the automated summarization of texts that can also be used to obtain the most important keywords in a document. It was introduced by Rada Mihalcea and Paul Tarau in [1].

The algorithm applies a variation of PageRank in [2] over a graph constructed specifically for the task of summarization. This produces a ranking of the elements in the graph: the most important elements are the ones that better describe the text. This approach allows TextRank to build summaries without the need of a training corpus or labeling and allows the use of the algorithm with different languages.

Text as a Graph:

For the task of automated summarization, TextRank models any document as a graph using sentences as nodes [3]. A function to compute the similarity of sentences is needed to build edges in between. This function is used to weight the graph edges, the higher the similarity between sentences the more important the edge between them will be in the graph. In the

domain of a Random Walker, as used frequently in PageRank in [2], we can say that we are more likely to go from one sentence to another if they are very similar.

TextRank determines the relation of similarity between two sentences based on the content that both share. This overlap is calculated simply as the number of common lexical tokens between them, divided by the length of each to avoid promoting long sentences.

The function featured in the original algorithm can be formalized as:

Definition 1. Given S_i , S_j two sentences represented by a set of n words that in S_i are represented as $S_i = w_1^i, w_2^i, ..., w_n^i$. The similarity function for S_i , S_j can be defined as:

$$Sim(S_i, S_j) = \frac{|\{w_k | w_k \in S_i \& w_k \in S_j\}|}{log(|S_i|) + log(|S_j|)}$$
 (1)

The result of this process is a dense graph representing the document. From this graph, PageRank is used to compute the importance of each vertex. The most significate sentences are selected and presented in the same order as they appear in the document as the summary.

Software Requirements:

- Python
- Tkinter
- Gensim

Implementation

Code:

```
from tkinter import *
from gensim.summarization import summarize, keywords
root1= Tk()
root = Frame(root1)
root.pack()

label1 = Label(root, text='TEXT SUMMARISATION')
label1.config(font=('helvetica', 14))
label1.pack(pady=10)

label2 = Label(root, text='ENTER YOUR TEXT:')
label2.config(font=('helvetica', 10))
label1.pack(pady=10)

text = Text(root, width=50, height=10, padx=10, pady=10, bd=5, selectbackground="blue")
text.pack(pady=10)

##root.geometry("300x300+120+120")
```

Fig 1: Text summarisation and keywords extraction code (part 1)

```
def txtsummarise():
    txt = text.get('1.0', END)
    text1.delete('1.0', END)
    text1.insert(END, summarize(txt, ratio=0.3))

text2.delete('1.0', END)
    text2.insert(END, keywords(txt, ratio=0.3))

button1 = Button(root, text='SUMMARIZE', command=txtsummarise, bg='brown', fg='white', font=('helvetica', 9, 'bold'))
button1.pack(pady=10)

text1 = Text(root, width=33, height=10, padx=10, pady=10, bd=5, selectbackground="blue")
text1 = Text(root, width=14, height=10, padx=10, pady=10, bd=5, selectbackground="blue")
text2 = Text(root, width=14, height=10, padx=10, pady=10, bd=5, selectbackground="blue")
text2.pack(side='left', pady=10)

##root.resizable(width=False, height=False)
root1.mainloop()
```

Fig 2: Text summarisation and keywords extraction code (part 2)

Input:

ठाणे - सा कोरोना िवषाणूशी लढ़ाई आपण लढ़त आहे. सव यु पातळीवर कोरोनाशी लढ़ाची सता करात आली आहे. सवनागरकांनी शासना। सुचनांचेपालन कन शासनास सहकायकरावे. िजा शासन नागरकां सुरिततेसाठी कटीब आहे. ठाणेिजात एकही ी उपाशी राहणार नाही याचेिनयोजन शासनानेकेलेआहे. कोरोना संकटािव सवानी िमळून, एकजुटीने, संघटीतपणेलढा देऊ या असेआवाहन पालकमंी एकनाथ िशंदे यांनी केले. ठाणेिजामेकोरोनाबाधीत ांची संा वाढत आहे, हेमाण रोखा। ीकोनातून व सथीतीचा आढावा घोसाठी पालकमंी एकनाथ िशंदे यांा अतेखाली ठाणेमहापािलका कै. नर बाळ सभागृहात आढावा बैठक घोत आली. या बैठकीस ठााचेखासदार राजन िवचारे, आमदार संजय केळकर, महापौर नरेश े,महापािलका आयु िवजय िसंघल, िजािधकारी राजेश नावकर, पोलीस आयु िववेक फणसळकर, तसेच वैकीय आरो अधकारी, कोरोनाचा ितबंध करासाठी महापािलकेस सहकायकरीत असलेलेसवडाॅसउपथत होते.

या बैठकीमेसंपूणपरथतीचा आढावा घेात आला. कोरोना ायरसचा फैलाव हा ामानेएकमेकांा संपकातून होत ं असानेसोशल िडीीगला ं सवानी ाधा िदलेपािहजेव याबाबत नागरकांचेबोधन करणेगरजेचेआहे. ा ीना ं कोरोनाची लणेिदसून येत असतील अशा सवानी त: पुढेयेवून तपासणी कन घेणेआवक आहेतर आिण तरच आपण या आजाराची साखळी तोडून यातून म् होवृशकतो. ा ीना ं कोरोनाची लागण झााचेिन झालेआहे, अशा ीचं ेनातेवाईक व ांा संपकात आले। सवाची तपासणी करात यावी, तसेच कोरोनाबाधीत ांवर उपचार करासाठी महापािलका कायेात ालयांची संा वाढवून यामेआवक असलेलेडॉस, नसस व इतर कमचारी उपल कन ांना आवक असलेली सववैकीय उपकरणेउपल कन ांची िनवास व भोजन वथा यो कारे राहीत याकडेकटाानेल राहील अशा सूचना पालकमंी एकनाथ िशंदेयांनी शासनाला िदा. तसेच आजाराचेवाढतेमाण लात घेवून आपाला अिधकचेडॉर िनयु करणेआवक आहे, यासाठी मुंबईा धतवर डॉरांना यो तेमानधन देवून याबाबत तातडीनेकायवाही कराबाबता सूचना सामा िजा ालय तसेच महापािलका शासनाला िदा. तसेच काही िठकाणी भाजीपाला व अधा खरेदीसाठी नागरकांची गद होत असाची बाब िनदशनास येत आहेयासाठी नागरकांमेजागृती करासाठी आवकतेनुसार िचफलक लावात यावेत व पोलीसांनी िदले। सूचनांचेनागरकांनी पालन करावेअसेही आवाहन पालकमंयांनी या बैठकीत केले. राामेलागुकेला संचारबंदी। कालावधीत अनेक नागरक अडकलेअसून तसेच हातमजुरी कन उदरिनवाह करणाया अनेक कुटुंबावर उपासमारीची वेळ आली आहे, अशा नागरकांना अनेक संथांा मामातून िनयिमत जेवण पुरिवात येत आहे. त:न पुढेयेवून ा संथांनी अितशय मोलाचेयोगदान िदलेआहे।बल पालकमंी एकनाथ िशंदे यांनी ांच आभार कन अिभनंदन केले. परंतुया सामािजक संथा गे। 25 िदवसापासून अधााचेवाटप करीत आहेंाना आता आवक अधा व भाजीपाला उपल कन देणेगरजेचेआहे, या संथा त: अ िशजवून ाचे िवतरण कराचेकाम करीत आहेतरी या संथांना तांदूळ, डाळ, तेल, मसाले, भाजीपाला आवकतेनुसार भागसिमतीिनहाय उपल कन देा। सूचना यावेळी िजा शासनाला िदा आहे. कोरोना संकट हेमोठेआहे, याला ितबंध करासाठी सवच यंणा वेगानेकाम करीत आहे, परंतुइतर आजारांनी असले। ांना उपचार घेणेआवक आहे, याकडेदेखील दुल चालणार नाही तरी अशा ांना यो उपचार वेळेत िमळतील या ीनेसवखाजगी ांलयेस राहतील या ीनेसवसंबंिधतांना आदेश ावेत अशी सूचना या बैठकीत पालकमंांनी केली. कोरोनाचा सामना करासाठीच सवच िवभाग भावीपणेकाम करीत आहे, यामेकोणतीही अडचण िनमाण झाास पालकमंी या नाानेमला वैयीक िनदशनास आणून ावी याबाबत आपाला तातडीनेयो ती मदत व

Fig 3: Input Text

Output:								
∅ tk				<u> </u>		×		
TEXT SUMMARISATION								
ठाणे - सा कोरोना िवषाणूशी लढाई आपण लढत आहे. सव यु पातळीव र कोरोनाशी लढाची सता करात आली आहे. सवनागरकांनी शासनाा सुचनांचेपालन कन शासनास सहकायकरावे. िजा शासन नागरकांा सुरिततेसाठी कटीब आहे. ठाणेिजात एकही ी उपाशी राहणार नाही याचेिनयोजन शा सनानेकेलेआहे. कोरोना संकटािव सवानी िमळून, एकजुटीने, संघटीतपणेलढा देऊ या असेआवाहन पालकमं ी एकनाथ िशंदे यांनी केले. ठाणेिजामेकोरोनाबाधीत ांची संा वाढत आहे, हेमाण रोखाा ीकोनातून व सथीतीचा आढावा								
		SUMMA	RIZE					

Fig 4: GUI of the system

TEXT SUMMARISATION

ठाणे - सा कोरोना िवषाणूशी लढाई आपण लढत आहे. सव यु पातळीव र कोरोनाशी लढाची सता करात आली

आहे. सवनागरकांनी शासनाा सुचनांचेपालन कन शासनास सहकायकरावे. िजा शासन नागरकांा सुरिततेसाठी

कटीब आहे. ठाणेिजात एकही ी उपाशी राहणार नाही याचेिनयोजन शा सनानेकेलेआहे. कोरोना संकटािव

सवानी िमळून, एकजुटीने, संघटीतपणेलढा देऊ या असेआवाहन पालकमं

ठाणेिजामेकोरोनाबाधीत ांची संा वाढत आहे, हेमाण रोखाा ीकोनातून व संथीतीचा आढावा

SUMMARIZE

ठाणे - सा कोरोना िवषाणूशी लढाई आपण लढ त आहे.

सवानी िमळून, एकजुटीने, संघटीतपणेलढा देऊ या असेआवाहन पालकमंी एकनाथ िशंदे यांनी केले.

ठाणेिजामेकोरोनाबाधीत ांची संा वाढत आहे, हेमाण रोखाा ीकोनातून व सधीतीचा आढावा कोरोनाची लणेिदसून येत असतील अशा सवानी त: पुढेयेवून तपासणी कन घेणेआवक आहेतर आण तरच आपण सहक आपण असलल फणसळकर असच तसच अडकलअसन दलप धतवर कटब

Fig 5: Output generation

Results

Original Text:

ठाणे - सा कोरोना िवषाणूशी लढाई आपण लढत आहे. सव यु पातळीवर कोरोनाशी लढाची सता करात आली आहे. सवनागरकांनी शासना। सुचनांचेपालन कन शासनास सहकायकरावे. िजा शासन नागरकां सुरिततेसाठी कटीब आहे. ठाणेिजात एकही ी उपाशी राहणार नाही याचेिनयोजन शासनानेकेलेआहे. कोरोना संकटािव सवानी िमळून, एकजुटीने, संघटीतपणेलढा देऊ या असेआवाहन पालकमंी एकनाथ िशंदे यांनी केले. ठाणेिजामेकोरोनाबाधीत ांची संा वाढत आहे, हेमाण रोखा। ीकोनातून व सथीतीचा आढावा घोसाठी पालकमंी एकनाथ िशंदे यां अतेखाली ठाणेमहापािलका कै. नर बाळ सभागृहात आढावा बैठक घोत आली. या बैठकीस ठााचेखासदार राजन िवचारे, आमदार संजय केळकर, महापौर नरेश े,महापािलका आयु िवजय िसंघल, िजािधकारी राजेश नावकर, पोलीस आयु िववेक फणसळकर, तसेच वैकीय आरो अधकारी, कोरोनाचा ितबंध करासाठी महापािलकेस सहकायकरीत असलेलेसवडाॅसउपथत होते.

या बैठकीमेसंपूणपरथतीचा आढावा घेात आला. कोरोना ायरसचा फैलाव हा ामानेएकमेकांा संपकातून होत ं असानेसोशल िडीागला ं सवानी ाधा िदलेपािहजेव याबाबत नागरकांचेबोधन करणेगरजेचेआहे. ा ीना ं कोरोनाची लणेिदसून येत असतील अशा सवानी त: पुढेयेवून तपासणी कन घेणेआवक आहेतर आिण तरच आपण या आजाराची साखळी तोडून यातून म् होवूशकतो. ा ीना ं कोरोनाची लागण झााचेिन झालेआहे, अशा ीचं ेनातेवाईक व ांा संपकात आला सवाची तपासणी करात यावी, तसेच कोरोनाबाधीत ांवर उपचार करासाठी महापािलका कायेात ालयांची संा वाढवून यामेआवक असलेलेडॉस, नसस व इतर कमचारी उपल कन ांना आवक असलेली सववैकीय उपकरणेउपल कन ांची िनवास व भोजन वथा यो कारे राहीत याकडेकटाानेल राहील अशा सूचना पालकमंी एकनाथ िशंदेयांनी शासनाला िदा. तसेच आजाराचेवाढतेमाण लात घेवून आपाला अिधकचेडॉर िनयु करणेआवक आहे, यासाठी मुंबईा धतवर डॉरांना यो तेमानधन देवून याबाबत तातडीनेकायवाही कराबाबता सुचना सामा िजा ालय तसेच महापािलका शासनाला िदा. तसेच काही िठकाणी भाजीपाला व अधा खरेदीसाठी नागरकांची गद होत असाची बाब िनदशनास येत आहेयासाठी नागरकांमेजागृती करासाठी आवकतेनुसार िचफलक लावात यावेत व पोलीसांनी िदला सूचनांचेनागरकांनी पालन करावेअसेही आवाहन पालकमंयांनी या बैठकीत केले. राामेलागुकेला संचारबंदी। कालावधीत अनेक नागरक अडकलेअसन तसेच हातमजुरी कन उदरिनवाह करणाया अनेक कुटुंबावर उपासमारीची वेळ आली आहे, अशा नागरकांना अनेक संथांा मामातून िनयिमत जेवण पुरिवात येत आहे. त:न पुढेयेवून ा संथांनी अितशय मोलाचेयोगदान िदलेआहे।बल पालकमंी एकनाथ िशंदे यांनी ांच आभार कन अिभनंदन केले. परंतुया सामािजक संथा गे। 25 िदवसापासून अधााचेवाटप करीत आहेंाना आता आवक अधा व भाजीपाला उपल कन देणेगरजेचेआहे, या संथा त: अ िशजवून ाचे िवतरण कराचेकाम करीत आहेतरी या संथांना तांदुळ, डाळ, तेल, मसाले, भाजीपाला आवकतेनुसार भागसिमतीिनहाय उपल कन दे।। सूचना यावेळी िजा शासनाला िदा आहे. कोरोना संकट हेमोठेआहे, याला ितबंध करासाठी सवच यंणा वेगानेकाम करीत आहे, परंत्इतर आजारांनी असले। ांना उपचार घेणेआवक आहे, याकडेदेखील दुल चालणार नाही तरी अशा ांना यो उपचार वेळेत िमळतील या ीनेसवखाजगी ांलयेस् राहतील या ीनेसवसंबंिधतांना आदेश ावेत अशी सूचना या बैठकीत पालकमंांनी केली. कोरोनाचा सामना करासाठीच सवच िवभाग भावीपणेकाम करीत आहे, यामेकोणतीही अडचण िनमाण झासि पालकमंी या नाानेमला वैयीक िनदशनास आणन ावी याबाबत आपाला तातडीनेयो ती मदत व

Fig 6: Input text for comparision

Summarised Text:

ठाणे - सा कोरोना िवषाणूशी लढाई आपण लढत आहे.
सवानी िमळून, एकजुटीने, संघटीतपणेलढा देऊ या असेआवाहन पालकमंी एकनाथ िशंदे यांनी केले.
ठाणेिजामेकोरोनाबाधीत ांची संा वाढत आहे, हेमाण रोखाा ीकोनातून व सथीतीचा आढावा
कोरोनाची लणेिदसून येत असतील अशा सवानी तः पुढेयेवून तपासणी कन घेणेआवक आहेतर आिण तरच आपण
राहीत याकडेकटाानेल राहील अशा सूचना पालकमंी एकनाथ िशंदेयांनी शासनाला िदा.
तसेच आजाराचेवाढतेमाण लात घेवून आपाला अिधकचेडार िनयु करणेआवक आहे, यासाठी मुंबईा
धतवर डॉरांना यो तेमानधन देवून याबाबत तातडीनेकायवाही कराबाबता सूचना सामा िजा ालय तसेच
करणाया अनेक कुटुंबावर उपासमारीची वेळ आली आहे, अशा नागरकांना अनेक संथांा मामातून िनयिमत जेवण
पुरिवात येत आहे.
तःन पुढेयेवून ा संथांनी अितशय मोलाचेयोगदान िदलेआहे।बल पालकमंी एकनाथ िशंदे
भागसिमतीिनहाय उपल कन दे। सूचना यावेळी िजा शासनाला िदा आहे.
कोरोना संकट हेमोठेआहे, याला ितबंध करासाठी सवच यंणा वेगानेकाम करीत आहे, परंतुइतर आजारांनी
असला ांना उपचार घेणेआवक आहे, याकडेदेखील दुल चालणार नाही तरी अशा ांना यो उपचार वेळेत
कोरोनाचा सामना करासाठीच सवच िवभाग भावीपणेकाम करीत आहे, यामेकोणतीही अडचण
सहकायकेलेजाईल अशी ाही पालकमंी एकनाथ िशंदेयांनी यावेळी िदली.

Fig 7: Summarised text

Conclusion:

We covered some interesting areas in NLP and text analytics with regard to information extraction, document summarization and keywords extraction. We implemented this project using Gensim library from python which helps us in summarisation for text provided in various regional languages without using and training model.

References

- 1. Mihalcea, R., Tarau, P.: Textrank: Bringing order into texts. In: Lin, D., Wu, D. (eds.) Proceedings of EMNLP 2004. pp. 404–411. Association for Computational Linguistics, Barcelona, Spain (July 2004).
- 2. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. In: Proceedings of the 7th International World Wide Web Conference. pp. 161–172. Brisbane, Australia (1998).
- 3. Christopher D. Manning, Prabhakar Raghavan, H.S.: Introduction to Information Retrieval. Cambridge University Press (2008).
- 4. Horacio Saggion and Thierry Poibeau. 2013. Automatic text summarization: Past, present and future. In Multi-source, Multilingual Information Extraction and Summarization. Springer, 3-21.
- 5. Yogesh Sankarasubramaniam, Krishnan Ramanathan, and Subhankar Ghosh. 2014. Text summarization using Wikipedia. Information Processing & Management 50, 3 (2014), 443–461.