



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 9      Issue: VII      Month of publication: July 2021**

**DOI: <https://doi.org/10.22214/ijraset.2021.36258>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Review on Real-Time Search Result Minimization using NLP

Prof. (Dr) Pawan Bhaladhare<sup>1</sup>, Ankit Chaudhari<sup>2</sup>, Nikhil Chaudhari<sup>3</sup>, Dhananjay Patil<sup>4</sup>, Pratik Patil<sup>5</sup>

<sup>1, 2, 3, 4, 5</sup>Department of Information Technology, Sandip Institute of Technology and Research Centre, Sandip Foundation, Nashik, Maharashtra.

**Abstract:** *At any given time thousands of people are searching about a particular thing and only about a fraction of those people might get the answer that they wanted. Whenever we do a quick search the possibility of getting the right answer is good but when we look into the time required to reach the right answer is not always fast as for every single search query we get about hundreds of search results which is good but also is a bit confusing for the user. The user might have to try many links only to reach his desired answer. So, in our proposed system the user just has to enter the search query and has to select his desired website for the answer and in a few seconds the answer will be displayed to him. When the user inputs the search query and a particular website he data is scrapped from that particular website and is then fed to a NLP sys-tem which is responsible to minimize the size of the answer keeping in mind not to change or lose any valuable data.*

**Keywords:** *Search query, Scrapped, NLP, Text Summarization*

## I. INTRODUCTION

Use italics for emphasis; do not underline. We have usually faced a situation when we are doing a particular work and we often get stuck by not knowing how it's done. Our go to solution for this is to head on to the Internet and just search our issue hoping that it gets resolved easily. On the contrary the results retrieved are in terms of thousands of linked pages of which many of them might not be useful to us. In fact, we never know which pages exactly match our query and which do not, until we check the pages individually which is quite a tedious and time consuming task. We are creating a program that gets rid of all the hassle that one has to undergo while trying to search for the answer. Our program takes in a query from the user which is fed to a NLP sys-tem and extracts the main Points i.e. Content of the search query which is then used to get all the data that is given to the user via a general search and this search result is then minimized using a technique which summarizes the content from the result. While summarizing, the context of the content is preserved. Based on the formatting of the answer, the context is determined and then it's summarized so that if some part of the content needs to be kept intact, such as a block of code or anything that requires not to be messed with. The user also has the option not only to search normally on the web but he can also specify a particular website from where the result needs to be taken. This in-turn increases the chances of getting the right answer. To further increase the speed of the program, a database will help to retrieve the answer if it was already asked or entered by any other user.

## II. LITERATURE SURVEY

- 1) Everyday large volumes of data are gathered from different sources and are stored since they contain valuable pieces of information. The storage of data must be done in an efficient manner since it leads to difficulty during retrieval. Text data are available in the form of large documents. Understanding large text documents and extracting meaningful information out of it is time-consuming. To overcome these challenges, information in the form of text is summarized with an objective to get relevant knowledge with the help of text mining tools. Summarized text will have reduced size as compared to original one. In this paper, we have tried to highlight major techniques for extracting important information from a given text with the help of topic modeling, key phrase extraction and summary generation. For topic modeling LSI and NMF method is used, weighted TF-IDF method is used for key phrase extraction while text summary is generated by using LSA and Text Rank method.
- 2) The World Wide Web has immense resources for all kinds of people for their specific needs. Searching on the Web using search engines such as Google, Bing, Ask have become an extremely common way of locating information. Searches are factorized by using either term or keyword sequentially or through short sentences. The challenge for the user is to come up with a set of search terms/keywords/sentence which is neither too large (making the search too specific and resulting in many false negatives) nor too small (making the search too general and resulting in many false positives) to get the desired result. No matter how the user specifies the search query, the results retrieved, organized and presented by the search engines are in terms of millions of linked pages of which many of them might not be useful to the user fully. In fact, the end user never knows which pages are exactly matching the query and which the pages are not, till one checks individually. This task is quite tedious and a kind of drudgery. This is because of lack of refinement and any meaningful classification of search results. Providing the

accurate and precise result to the end users has become Holy Grail for search engines like Google, Bing, Ask etc. There are a number of implementations that have arrived on the web in order to provide better results to the users in the form of DuckDuckGo, Yippy, Dogpile etc. This research proposes development of a Meta search engine, called SEReLeC that will provide an interface for refining and classifying the search engines' results so as to narrow down the search results in a sequentially linked manner resulting in drastic reduction of number of pages.

- 3) Using a Text Mining Tool to Support Text Summarization paper presents a mining tool that is able to extract graphs from texts, and proposes their use in helping students to write summaries. The text summarization method is based on the use of the graphs as graphic organizers, leading students to further reflect about the main ideas of the text before getting to the actual task of writing. An experiment carried out demonstrated that the tool helped students reflect about the main ideas of the text and supported the writing of the summaries.
- 4) Text Summarization, is a topic that is related to the fields of philosophy and linguistics are also included in the social sciences, so often sought by researchers in computer science who prefer something in the field of exact sciences. Interestingly, the Text Summarization application is a system that will make a summary or conclusion of tens or hundreds of texts with similar themes in an overview that produces new knowledge. This system becomes very important if you have a problem to take a series of conclusions from existing text. The purpose of Application of knowledge graph for making Text Summarization (Analyzing a text of educational issues) research is to develop a knowledge-graph method to do case studies to determine whether the knowledge graph can be used as an efficient instrument in analyzing a text with the same theme in large numbers, then the results become much simpler in the form of graphs that can be used as new knowledge.

### III. SYSTEM ARCHITECTURE

The main objective of this system is to help and reduce the time required for a person to find a quick solution when they have a time restraint. Search Engines gives us solutions but it comes with lots of options and false positive results. Here this system will be working on a website which takes in the search query and preferred source from the user. This search query is fed to a program which collects all the data about our search query from the website the user gave us. This data is then given to a NLP program which is responsible to minimize the content size which helps the user understand it faster.

The search query which the user gives is used as a search parameter for the preferred website. Each website has its own structure of displaying the information on the web. The website is made up of HTML, CSS and various frameworks, but the common property of each site is that all sites have data under a particular tag which might have a class or an id. The system uses these classes or id to scrape the data from the site which is passed on to the NLP component of the system is responsible for minimizing the data. NLP has a large variety of algorithms which can be used to do this but here the system uses Knowledge Graphs which allows users to create a pipeline in which they can select steps such as tokenization, stop-words removal and named-entity recognition (NER), among others. After all the steps are complete a short version of the given data is produced to the user.

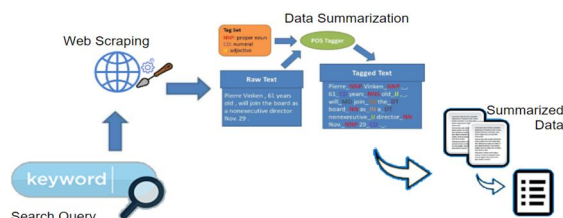


Fig. Proposed system

### IV. CONCLUSION

So finally by implementing this project we are helping those users who are looking for quick and accurate answers to their queries. This system will provide an interface for refining and classifying the search engines' results so as to narrow down the search results. The Expected Outcome of our system will surely help save valuable time for people of all ages. For the sake of future modifications in case of more enhancements, not only textual data but also other types of data can be shortened and provided to the user bound by a time constraint. Further changes will be made to make this system even faster without sacrificing accuracy.



### REFERENCES

- [1] Asha Rani Mishra, V.K Panchal, Pawan Kumar, "Extractive Text Summarization - An effective approach to extract information from Text", 2019 International Conference on contemporary Computing and Informatics (IC3I).
- [2] Vishwas Raval, Padam Kumar, "SEReleC (Search Engine Result Refinement and Classification) - a Meta search engine based on combinatorial search and search keyword based link classification", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012).
- [3] Eliseo Reategui, Miriam Klemann, Mateus David Finco, "Using a Text Mining Tool to Support Text Summarization", 2012 IEEE 12th International Conference on Advanced Learning Technologies.
- [4] Khodijah Hulliyah, Husni Teja Kusuma, "Application of knowledge graph for making Text Summarization (Analyzing a text of educational issues)", Proceeding of the 3rd International Conference on Information and Communication Technology for the Moslem World (ICT4M) 2010.



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)