

# Intelligent Semantic Web Search

Jayaty Jayaty  
North Carolina State University  
jjayaty

Aishwarya Seth  
North Carolina State University  
aseth

## I. INTRODUCTION

A vast majority of information has shifted online, web search has become one of the most important tools these days. However, just the possibility of having access to a humongous amount of data present online is not enough, if you are not able to extract relevant information. The ability to search across multitudes of web pages has advanced a lot but the ability to search intelligently and semantically is still relatively new. Users may not necessarily be well aware of the accurate keywords required to be used for giving them correct results. Therefore, a more lenient approach is required to provide relevant results to the users instead of a strict keyword search approach. In this project, we will try to obtain the results of a web search query in a way that is not restricted to a strict keyword search. Thus, we title this type of search as an *Intelligent Semantic Web Search*.



Fig. 1. Keyword based web search

## II. USAGE SCENARIO

The intelligent and semantic web search is the next step towards making the user experience more seamless. This application can be integrated wherever a feature for user search is needed. For example, search engines, word meaning lookup on a website. This application can also be modified to integrate intelligence to semantic code search, for finding a specific piece of code in a huge project or code base or even online.

## III. SPECIAL FEATURES OF APPLICATION

The most special feature of the application will be the semantic search. The semantic search feature will empower the user to search what they want without exactly remembering the words/keywords for it. The feature will help in displaying the relevant results even in the absence of precise keywords as well. The feature is relatively recent and unorthodox from traditional tools. Search becomes easier by identifying entities and mapping unstructured data.

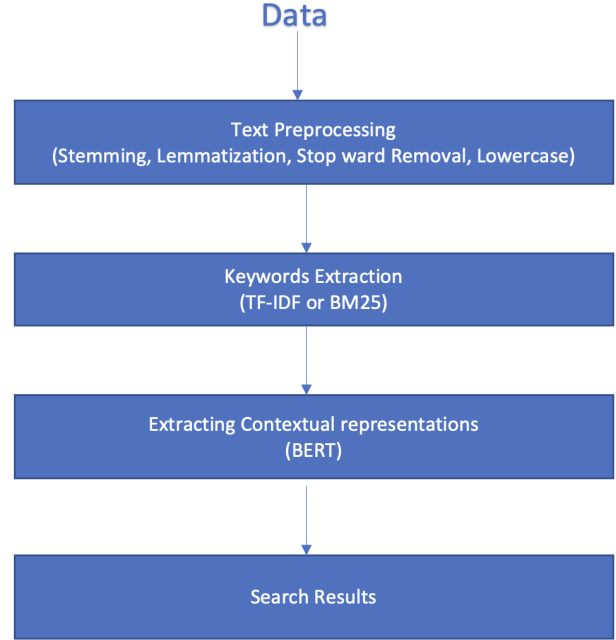


Fig. 2. Intelligent semantic web search

## IV. METHODOLOGY

Our intelligent semantic web search consists of three phases: Text Pre-processing, Keywords extraction and Contextual Representation Extraction.

### A. Text Pre-processing

The user query needs to be pre-processed first before proceeding with the task of searching. Our text-pre-processing techniques will include converting the query text into lowercase, stemming, lemmatization and finally stop words and punctuations removal(using the list of most common stop words available).

### B. Keywords Extraction

The next step is using keywords extraction. Although we argue we need something more advanced than simple keyword extraction, we still agree that keyword extraction helps in development of initial relationships and contexts. We will be using either Tf-IDF or BM25 for keywords extraction.

### C. Contextual Representation Extraction

We will be extracting contextual representation in order to get the core meaning of the query which user intends to search for. Getting the core meaning helps in getting similar words which might mean the same thing as the query and lead to actual results which the user is searching for. Contextual representation extraction will be achieved by obtaining embeddings and training them using advanced models like BERT.

#### REFERENCES

- [1] Jui-Ting Huang, Ashish Sharma, Shuying Sun, Li Xia, David Zhang, Philip Pronin, Janani Padmanabhan, Giuseppe Ottaviano, Linjun Yan, "Embedding-based Retrieval in Facebook Search", CoRR abs/2006.11632, 2020 <https://arxiv.org/abs/2006.11632>
- [2] Pascal Hitzler, "A Review of the Semantic Web Field", Communications of the ACM Volume 64Issue 2February 2021 pp 76–83 <https://doi.org/10.1145/3397512>
- [3] Hamel Husain, "How To Create Natural Language Semantic Search For Arbitrary Objects With Deep Learning", <https://towardsdatascience.com/semantic-code-search-3cd6d244a39c>
- [4] Yusuf Sermet, Ibrahim Demir, "A Semantic Web Framework for Automated Smart Assistants: COVID-19 Case Study", arXiv:2007.00747, <https://arxiv.org/abs/2007.00747>
- [5] "A decade of Semantic Web research through the lenses of a mixed methods approach", Semantic Web, vol. 11, no. 6, pp. 979-1005, 2020, <https://content.iospress.com/articles/semantic-web/sw200371>
- [6] "The Semantic Web identity crisis: In search of the trivialities that never were", Semantic Web, vol. 11, no. 1, pp. 19-27, 2020, <https://content.iospress.com/articles/semantic-web/sw190372>
- [7] "Information extraction meets the Semantic Web: A survey", Semantic Web, vol. 11, no. 2, pp. 255-335, 2020, <https://content.iospress.com/articles/semantic-web/sw180333>