Indian Journal of Science and Technology, *Vol 9(32), DOI: 10.17485/ijst/2016/v9i32/100739, August 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Information Retrieval in the Context of Checking Semantic Similarity in Web: Vision of Future Web

Sanjib Kumar Sahu[1*], D. P. Mohapatra[2] and R. C. Balabantaray[3]

[1]Department of Computer Science, Utkal University, Bhubaneswar - 751004, Odisha, India; sahu_sanjib@rediffmail.com
[2]Department of Computer Science and Engineering, NIT, Rourkela - 769008, Odisha, India; durga@nitrkl.ac.in
[3]Department of Computer Science and Engineering, IIIT, Bhubaneswar - 751003, Odisha, India; rakesh@iiit-bh.ac.in

Abstract

Objective: This work discusses about web search based on content present in the web page. Web 3.0 in this field has been a major invention done till this date. Yet semantic web proposes to be a good choice for its enhancement towards searches based on intent of user. There has been lots of techniques proposed so far still methods are been searched based on this concept. **Methods:** Surfing the internet for relevant information as needed by the user has therefore become an issue for users. In order to overcome the problem of information deluge for relevant information retrieval, there is a need to devise a way to search data based on user's intent. This is where semantic search comes into play. **Findings:** In order to improve quality of search correctness, semantic search tries to interpret the context of data provided along with user's intent, location from where that has been looked for, variation of words etc. As a part of our work we have proposed a model using natural language processing that aims to classify data based on semantic relevance to user query. The well-known concept of NLP permits the machines to derive the purpose or the meaning meant by humans or natural language input. The aim is to provide exact needed result to the user in place of making him to look for it all over the URLs given as result. **Improvements/Applications:** Tools based on this semantic knowledge will increase relevance of search engines along with the less list of documents o go through by the user or surfer for a specific result.

Keywords: Information Retrieval (IR), Natural Language Processing (NLP), Semantic Web, Semantic Relevance, Stop Words

1. Introduction

IR a most featured concept that has been a great helper in view of gaining such information sources that are relevant to a user query from large pool of such similar sources. A one more theory over this is of information overload" i.e an automated information retrieval system with the purpose of reducing this overloaded information. IR has been a great help to large pool of educational institutes and public libraries as to give admittance for books, journals and other available documents. Search engines are the most common among these. In our world where information over sites and internet is in tremendous amount, retrieving the exact needed result for the right person and that again in a time limit manner is not a cup of tea. Techniques that we have with us for performing above stated task, have been suffering with inaccuracy and meaningless queries with irrelevance in the results. Old-fashioned IR technologies are largely based on the keyword syntax matching. Advantage of this technique is simple and quick and easy. However, lack of knowledge expression, processing and understanding leads to unsatisfactory search quality. Present IR techniques are incapable in achieving semantic knowledge from web pages which leads to incapability of giving out particular answers to a particular query. Hence, there is a need to have a method which will provide the user with the relevant information based on the semantic of the query submitted to the search engine. Semantic information retrieval has become an important topic. Semantic search seeks to improve accuracy of search by interpreting the intention of the searcher and

** Author for correspondence*

the context behind the terms entered by him. These systems are focused on various points including context, location, and intention, along with deviation of words, synonyms that are used, generalized and specialized queries, concept matching in view of providing relevant output. Rather than using ranking algorithms which are deliberately providing outputs that ware no near the actual intention with which it is searched. In order to overcome the existing anomalies in current IR systems, we are proposing a model utilizing Natural Language Processing (NLP). The focus is on delivering the information as asked for directly in place of pressurizing user to scan a set of results.WWW is different from internet in the ways of projecting subparts of internet defined in techno logical and social (techno-social) systems that comes in direct contact to the humans. This new word, techno-social platform is the model that helps humans in reasoning, communicating and cooperating with each other and the machine around them. These all parameters are essential requirement for communication and defining conditions for the network. The internet is the most transferable-information paradigm given by Tim-Burners-Lee. Progress has been made from Web 1.0, a tool based on muddle of reasoning to web 2.0 as for communication to web 3.0 incorporating the cooperation paradigm to the last web 4.0 which is famous for integrating these all paradigms from last few generations. Providing the relevant information to the user in a reasonable amount of time has become an important task of IR systems. The information returned by the search engines contains a lot of irrelevant information. The user is interested in correct, accurate and precise information. The keyword- based search engines perform keyword matching to retrieve the documents. It is most common fundamental that documents relevant to specific query are not indexed by current search engines we use. Sometimes, irrelevant documents are returned because similar terms are used but in different contexts. Because of the above stated reason present IR techniques stands incapable in giving semantic knowledge or the context of document which has led incompetency in precision of these tools. This paper provides a background of the web 3.0 evolutions and its relation with web 1.0 and web2.0 in introduction.

1.1 Web 1.0

Web 1.0, the very first stage of WWW, was entirely based on hyperlinked web pages. It was believed that at the time it was made only static web pages were built which were incapable of providing interactive content. As being the first implementation, it lasted from period of 1989 to 2005. It was a read-only type of web. It was used by different ventures to available catalogs or brochures to give their ideas through web to general people and a way to interact with other businesses. Websites built on this were of static pages and were focused on publishing information to people at any course of time and be present online. As it was the very first development of WWW, the major bottleneck faced by it is that pages designed using this were understandable to humans only. Next, the web master, a tool that was completely responsible for updating and managing pages has created dependency and finally, it lacked the dynamicity that is needed in all such tools present over web. The Figure 1 will explain the detail about Web 1.0[1].

Figure 1. Graphical View of Web 1.0.

1.2 Web 2.0

The very next innovation was Web 2.0 which was designed in 2004 and referred as second generation of WWW. This series was a technological improvement. It provides a user friendly platform that was unavailable in web 1.0. Web 2.0 was made with the specialty of using net combined with higher user interaction to provide their views and suggestions online like Wikipedia, Blogger, Digg, Technocratic, MySpace, Face book, Blogs and many more. This technique is known for its wisdom web, people's web, participative web, and read-write web. Reading and writing operation has made it bi-directional. Being an extension over web 1.0, it provides better user support. It is more flexible, provides creativity with reusability, updates, and many more features to user. One of the most significant features of web 2.0 is its paradigm of providing support to teamwork along with feature of gathering collective information. As there are pros, there exist cons too. Drawback of its is that it has met all the expectations of user but in doing that there always been chance of interference from external environment. These external forces have affected the feasibility of the output this can provide thus degrading the performance. Among these limitations there is the steady chain of iterations of change and updates, then there comes the ethical issues for constructing and using Web 2.0 and lastly, interconnectivity and knowledge sharing among different platforms were limited which are still present. The detailed explanation is shown in Figure 2.

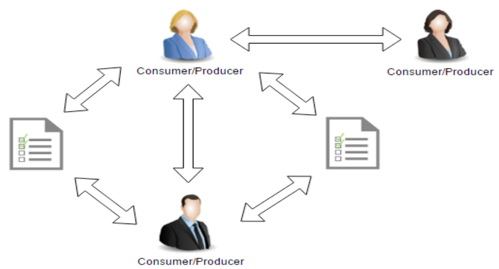


Figure 2. Graphical View of Web 2.0.

1.3 Web 3.0

The newest generation of WWW is Web 3.0 which is a joined venture of Web 2.0 and have hands on Semantic web, thus enabling it for humans and for machines too. It has enabled the Web to permit machines in accomplishing jobs assigned to them with the need of human intelligence, which leads to reduced time and efforts on internet. It is aimed in making web a better and a smarter network which is an antecedent to fully semantic Web, and the descendant of Web 2.0. However, Web 3.0 will give the Internet itself intelligence by making machine-programs that access data which understands what the data itself is. This results in searching the best information from the Web for our needs. The present internet is a pool of web pages, which is like a universal file system for most of the essential problems about web are included. The detailed explanation is shown in Figure 3.

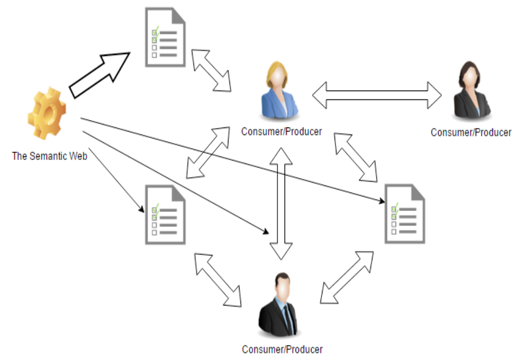


Figure 3. Graphical View of Web 3.0.

1.4 Web 4.0

Web 4.0 is an extraordinary-Intelligent Electronic Agent and symbiotic web. Symbiotic web is aimed with incorporating interaction among humans and machines. This machine is made very prevailing and powerful as human brain, growth in developing telecommunication, enhancement in nanotechnology and an inhibited interface using this tool. This has enabled machines to read the contents over web, and then to respond in the form needed. Web 4.0 model will be made to read write concurring the web. It will ensure collaboration of data globally, governing data, creating transparency participation, and finally making them available to industries, social, political and other communities.

1.5 Comparison between Web 1.0, 2.0, 3.0 and 4.0

The evolution explanation of web 1.0 to web 4.0 is described in the Table 1.The rest of the paper is organized

Table 1. Comparison among Web 1.0, Web 2.0, Web 3.0 and Web 4.0

WEB 1.0	WEB 2.0	WEB 3.0	WEB 4.0
1996-2004	2004-2010	2010-2016+	FUTURE WEB
THE INTERNET	THE SOCIAL INTERNET	THE SEMANTIC INTERNET	THE SYMBIOTIC WEB
TIM BERNER'S LEE	TIM O'REILLY	TIM BERNER'S LEE	RESEARCH GOING ON
READ INTERNET	R AND W INTERNET	EXECUTABLE INTERNET	CONCURRENCY WEB
SHARED INFO	COMMUNICATION	ENTANGLEMENT	COMMANDING INTERFACE
MILLIONS USERS	BILLIONS USERS	TRILLIONS USERS	NEVER NO
ECOSYSTEM	CONTRIBUTION	CONCERNED SELF	CONTROLLED INTERFACE
THE HTML/CGI INTERNET	THE PUBLIC INTERNET	THE SEMANTIC INTERNET	INTERACTION SOFTWARE
STATIC CONTENT	DYNAMIC CONTENT	UNDEFINED	PARTICIPATION, DISTRIBUTION

as follows: In Section 2 we discussed the problem identification and motivation of work. In Section 3 we are discussing about methodology and proposed work. In Section 4 we discussed about result and analysis. Section 5 proposes the comparison with related work. Finally we concluded the paper and defined the future scope in Section 6.

2. Problem Identification and Motivation

IR the very famous information retrieval concept states a way of gaining meaningful data from relevant resources for a specified query. IR always aimed for on providing the correct data to a specified need within time constraint from a large pool of data. The traditional IR methods often suffered from inaccuracy and incomplete results with inconsistency too. Present IR methods are incapable of using semantic knowledge within the pages and thus can never provide exact answers to exact questions. Hence, there is a need to have a method which will provide the user with the relevant information based on the semantic of the query submitted to the search engine.

2.1 Motivation

Typical search engines are Keyword-based. The problems with keyword-based Search Engine are as follows:

- Quality results set – It has been noticed that results provided for a specific query is never relevant and neither indexed.
- Time constraint- Going manually through all the results provided for a specific query is a very tedious job.
- Semantically similar queries can return different results
- Results are narrowly related with the spelling of the term but not its context
- Internet address is the basic need to find the source of specific information.
- Irrelevant results for keywords containing synonymous terms like “RESTAURANT” vs. “CAFE”, “PRC” vs. “CHINA”.

When we are entering “Lemon Tree” query in Google, the result is shown in Figure 4.

2.2 Problem Statement

The fact that Natural Language Processing (NLP) does for

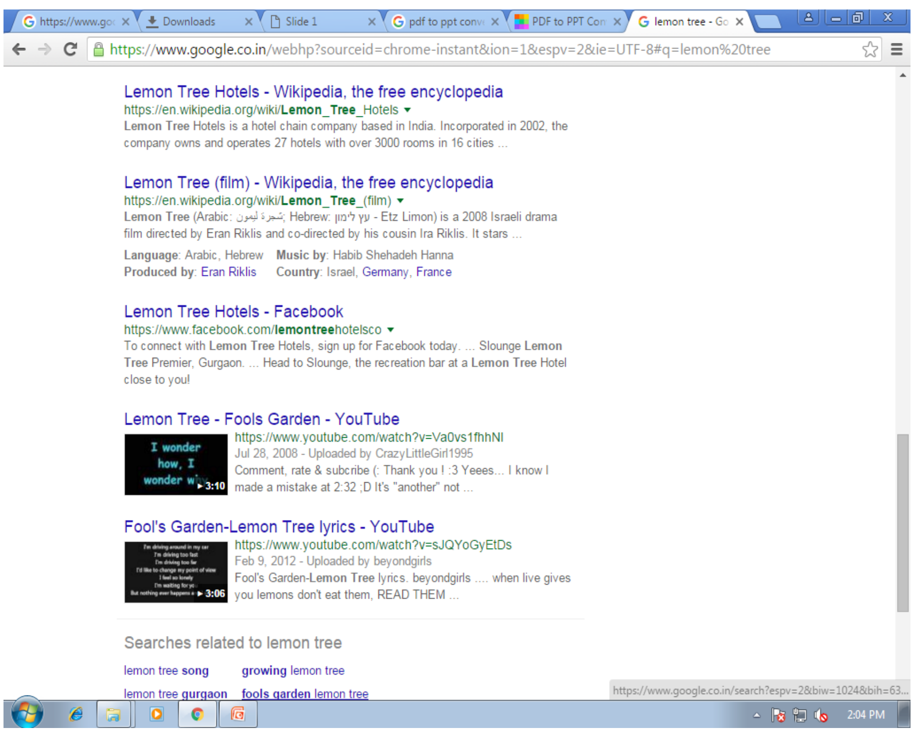


Figure 4. Output for Entering “Lemon Tree” in Google Search Engine.

accuracy is still unavailable for universal domains, which leads to plentiful efforts. This includes creating uniform languages which are semantically strong for the web we use. This would help to convert the conventional we into a semantic web which shall use conceptual representation of Web pages. Semantic search engines have always been existed for specialized areas of knowledge.

2.3 Main Intent

- We require an approach which can improve the existing web with a semantic layer that allows machines to understand it, thus giving them much better access to information resources. This would further help to improve the existing web with a semantic layer that allows machines to understand it.
- Software programs can themselves process information efficiently which helps to move from a web of "finding things" to a web of "doing things".

2.4 Scope for Solution

- Semantic search tends to find ways for improving accuracy of search using the conceptual knowledge hidden in it through guessing surfer's intent. It just don't assign ranks for predicting relevancy, it uses hidden meanings into it for giving extremely relevant outputs.
- NLP has always have its attention over syntactically, semantically, and pragmatically examination of text and discourse. Capability of analysing syntax and semantic of a phrase allows efficient results based on meaning rather than keywords.
- **Syntactic analysis** using knowledge of the grammar
- **Semantic analysis** using info. about meaning of words
- **Pragmatic analysis** using info. about context
- Above stated methods are to calculate the intent behind an ambiguous term through its context. They helps us in looking for definite parts of information within a page, further they allow us to answer solutions natural language processing queries for a corpora or structured data like FreeBase or Google's Knowledge Graph.

3. Methodology and Proposed Approach

Most of the search engines being used currently do return a lot of irrelevant information that do not meet user's requirements. Surfers now days are more interested in confined data rather than a bulk a data and over to this they want results to be precise, accurate and relevant. The success of present web engines is based on how much the keywords are suitable in framed query. With keyword matching approach it is not possible to distinguish among relevant and irrelevant document if documents use similar terms but in different context. This is one of the basic reasons in attaining high accuracy over the results. The current retrieval methods are lacking the use of semantically meaningful data and thus cannot interpret meaning of search keyword and intent behind it.

This can be explained from the example we took from online sites,

The web search engine may not retrieve relevant documents that include synonymous terms such as:

- "restaurant" vs. "café"
- "PRC" vs. "China"

The web search engine may retrieve irrelevant documents that include ambiguous terms such as:

- "bat" (baseball vs. mammal)
- "Apple" (company vs. fruit)
- "bit" (unit of data vs. act of eating)

When we enter the search query as 'Lemon Tree' on Google search engine, we get the results shown in Figure 4.

3.1 Proposed Solution

After considering the drawbacks of the traditional approach to information retrieval based on keyword matching, we have proposed a solution that will increase the accuracy of the relevant documents retrieved with respect to the user's query. The solution is based on the semantic knowledge within documents and hence will give precise answers to the user's query. It will provide the user with the relevant information based on the semantic of the query submitted to the search engine. The proposed solution uses Natural Language Processing as the base technique and it includes Latent Semantic Indexing (LSI) model for calculating the similarity score between the list of the documents and the user query submitted to the search engine.NLP has been in lime light because it has been originated from artificial intelligence which enables it to analyze, understand and generate such languages which will help humans to use tools for direct interfacing with machines in both ways. It uses Latent Semantic Indexing (LSI) uses Singular Value Decomposition (SVD) as its method to retrieve and index pages to find interrelationships between them. It is based on the concept that terms used within a page have some context that they represent and tend to give similar meanings. In our approach, we have first modeled the documents using the Tf-idf model. The Tf-idf model was then transferred into the LSI model for getting better accuracy of the results.

3.2 Assumptions

While proposing the algorithm for efficient information retrieval algorithm, we have assumed the following:

Finding information by Crawling: They help to find documents and go through the hyperlinks from one page to another. They go through these hyperlinks and give us the data about them in return to search engine's servers. The process starts from a list of internet addresses which were accessed from previous crawls and finds a site map of a specific website owner. They visit different websites, finds links present on it for different pages. Crawlers pay special care on newly added websites, any changes made to these and any dead link too. Computer programs were made in view to find which site is next to visit, how many times and number of pages it can fetch from other sites. We assume that the data the information has been collected by the software known as web crawlers. Information Indexing: Internet is ever-growing public library which contains trillions of eBooks with no main filing system. Every search engine basically collects web pages while crawling and then indexes them in respect to find which path to follow if we need it again. It is same as an index of a book, but every search engine index incorporates meaningful data of all those terms present in it with locations. We assume that the information or the web-pages has been stored in the organized way with the help of the technique known as indexing.

3.3 Proposed Algorithm

With tremendous increase in amount of information, there is an urgent need to devise a way to retrieve information relevant to users. The traditional information retrieval systems are based on keyword matching and therefore cannot understand the user's intent. As a result, the documents which are relevant to the user are often expunged from data due to expression difference. Keeping in mind the anomalies in existing IR (information retrieval) systems, we have proposed an IR system using Latent Semantic Indexing (LSI) and Natural Language Processing (NLP).

A. Dataset Collection

Data collection is the primary and foundation step in IR systems. We have used Google search results as our primary dataset. As a part of our work, we have collected 50 search results for each query. The collected data is stored in the same format which is machine and human readable.

B. Exploring and Preparing the Dataset:

a. Text extraction: It involves extracting textual data from retrieved URLs and converting them into individual documents for text processing.

b. Preprocessing: In this step stop words removal, punctuation removal and stemming are performed

C. Algorithm

a. From strings to vectors: Documents are converted into vectors, and for doing this paper will be using a bag-of-words concept for document representation. This method represents a document by one vector representing a question-answer pair.

b. tf-idf calculation: Tf –idf(term frequency-inverse doc frequency) weights each document those have been represented in the vector space. This method is composed of two terms: the first computes a normalized Term Frequency which gives a count of number of times a term is present in the page, in fraction with number of terms present in that document. The second part stands for Inverse Document Frequency, which represents a logarithm of count of documents present in the corpus in fraction with number of documents in which the term ti appears.

$$tf(t,d) = 1 + \log f_{t,d} \quad (1)$$

where

t=term in a document

d=document

$f_{t,d}$= raw frequency of term t in document d

$$Idf\,(t,D)=\log\,(N/|\{d \in D : t \in d\}|) \quad (2)$$

With

- ***N:*** total number of documents in the corpus $N=|D|$
- $|\{d \in D : t \in d\}|$**:** number of documents where the term t appears (i.e., $tf(t,d) \neq 0$). If division by zero gives whether the term is present in that set of

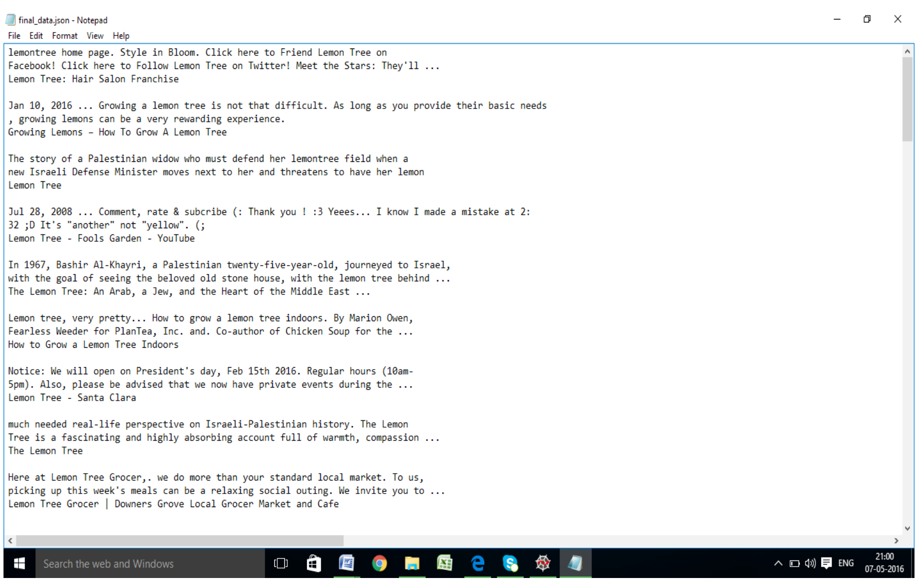
documents or not. It is therefore common to adjust the denominator to $1+|\{d\epsilon D : t\epsilon d\}|$

c. Transforming from tf-idf to LSI model: LSI is a method of indexing that uses singular value decomposition (SVD), a mathematical technique which identifies patterns giving interrelationships between the terms and concepts present in a formless group of text. It works on the principle which explains that terms used with same contexts gives away similar meanings. In this paper, Tf-idf model was finally transformed to LSI model. The goal of transformations is: To convey the concealed structure in the corpus, it finds out the relations among the words and then uses these in view to describe pages in a very new and (hopefully) a lot more semantic way. Thus document representation becomes more compact. This leads to improved efficiency (new representation consumes less resources) and efficacy (marginal data trends are ignored, noise-reduction). We transformed the trained tf-idf weighted corpus into a 2-D latent space using LSI.

Obtaining similarity measure between user query and documents: Cosine similarity is a standard measure in Vector Space Modelling using which the similarity between two documents(query and search result) is measured, the value for which lies in [-1,1]. Once the similarity is calculated, the documents can be arranged in decreasing order of similarity value.

4. Result and Analysis

For our experimentation samples of search results were taken from the most widely used web- based search engine, Google. Corresponding to each user query, a set of top 50 documents returned by Google was collected. Figure 5 and Figure 6 show the samples of extracted text from the collected documents. The documents corresponding to a particular domain are submitted as input to the above mentioned algorithm and relative semantic similarity for each is calculated. The Cosine measure returns similarities in the range *<-1, 1>* (the greater, the more similar), so that the first document has a score of 0.99809301 etc. Based on these computed values, the collected dataset is sorted so as to rank and display the documents based on their semantic similarity which helps in efficient information retrieval. Figure 7 shows no. of retrieved URL's with similarity percentage lying in a particular range. Further, a graph is plotted between similarity index and the number of documents which is shown in Figures 8 and Figure 9. We have taken another query "bolt from blue" is an idiom facing semantics problem. Only 22 URLs collected due to google server issues. We passed it through our software tool, the output is displayed in the Figure 10 and Figure 11.Finally for more comparative result we have taken another query "Nirbhaya Case Delhi". We passed it through our software tool, The output is displayed in the Figure 12 and Figure 13.


final_data.json - Notepad
File Edit Format View Help

lemontree home page. Style in Bloom. Click here to Friend Lemon Tree on
Facebook! Click here to Follow Lemon Tree on Twitter! Meet the Stars: They'll ...
Lemon Tree: Hair Salon Franchise

Jan 10, 2016 ... Growing a lemon tree is not that difficult. As long as you provide their basic needs
, growing lemons can be a very rewarding experience.
Growing Lemons – How To Grow A Lemon Tree

The story of a Palestinian widow who must defend her lemontree field when a
new Israeli Defense Minister moves next to her and threatens to have her lemon
Lemon Tree

Jul 28, 2008 ... Comment, rate & subcribe (: Thank you ! :3 Yeees... I know I made a mistake at 2:
32 ;D It's "another" not "yellow". (;
Lemon Tree - Fools Garden - YouTube

In 1967, Bashir Al-Khayri, a Palestinian twenty-five-year-old, journeyed to Israel,
with the goal of seeing the beloved old stone house, with the lemon tree behind ...
The Lemon Tree: An Arab, a Jew, and the Heart of the Middle East ...

Lemon tree, very pretty... How to grow a lemon tree indoors. By Marion Owen,
Fearless Weeder for PlanTea, Inc. and. Co-author of Chicken Soup for the ...
How to Grow a Lemon Tree Indoors

Notice: We will open on President's day, Feb 15th 2016. Regular hours (10am-
5pm). Also, please be advised that we now have private events during the ...
Lemon Tree - Santa Clara

much needed real-life perspective on Israeli-Palestinian history. The Lemon
Tree is a fascinating and highly absorbing account full of warmth, compassion ...
The Lemon Tree

Here at Lemon Tree Grocer,. we do more than your standard local market. To us,
picking up this week's meals can be a relaxing social outing. We invite you to ...
Lemon Tree Grocer | Downers Grove Local Grocer Market and Cafe

Search the web and Windows
ENG 21:00 07-05-2016

Figure 5. Samples of Extracted Text from the Collected Documents.

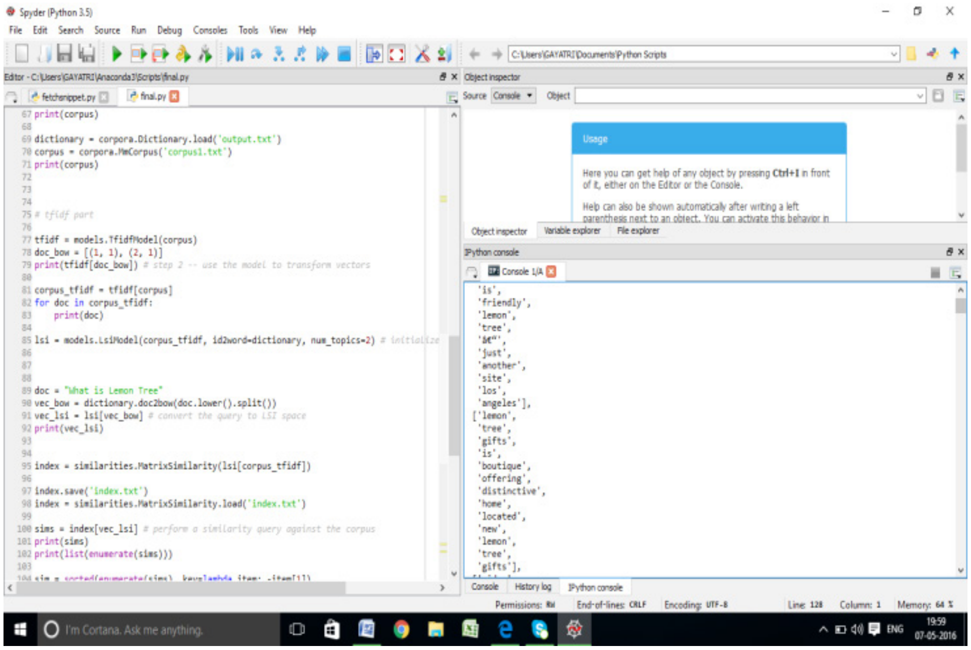


Figure 6. Tool Developed through Spider APP to Process Similarity Index (SI) using Vector Space Modeling.

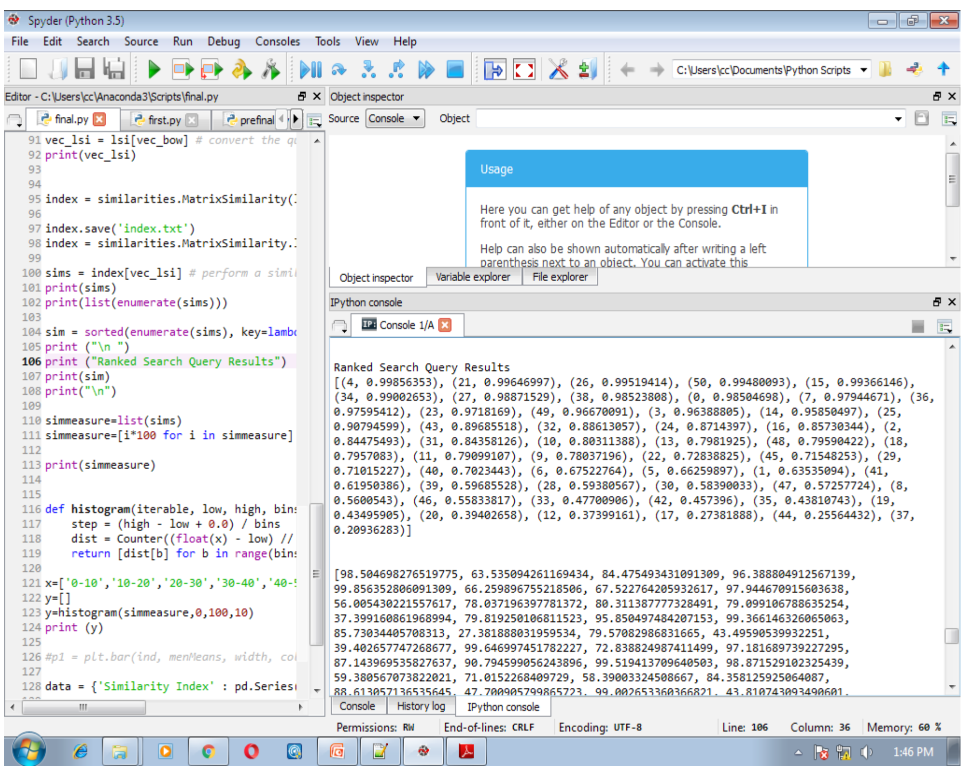


Figure 7. No. of Retrieved URL's with Similarity Percentage Lying in a Particular Range.

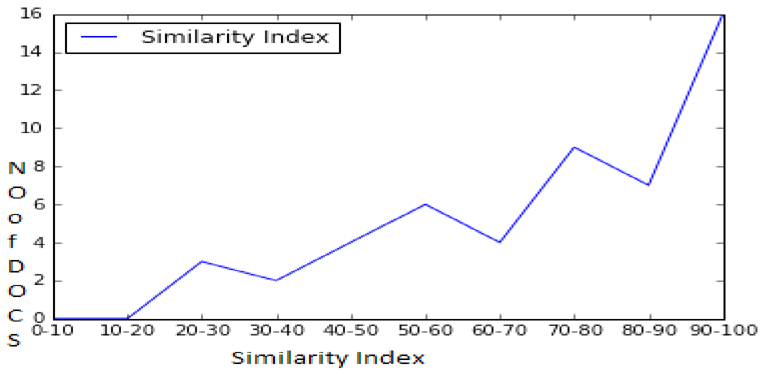


Figure 8. Graph showing the similarity index Vs number of documents.

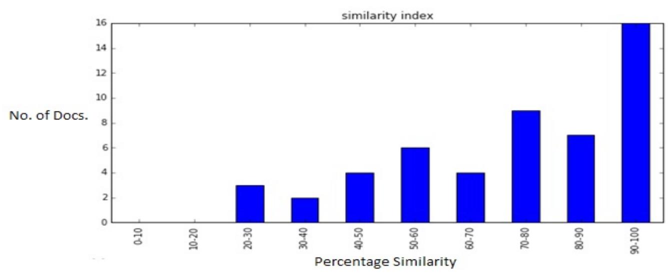


Figure 9. Another Graph Showing the Similarity Index vs the Number of Documents.

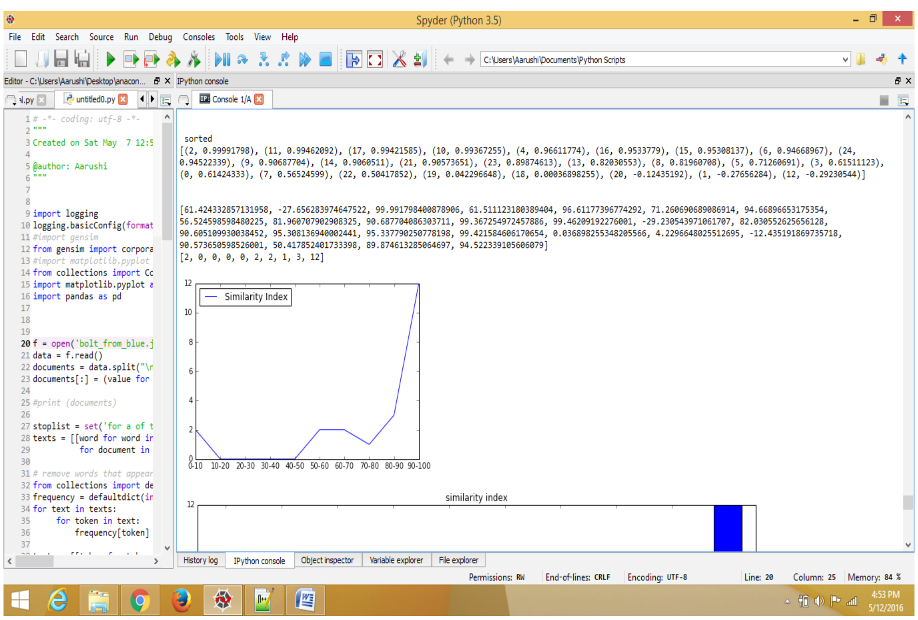


Figure 10. No. of Retrieved URL's with Similarity Percentage Lying in a Particular Range for Another Query.

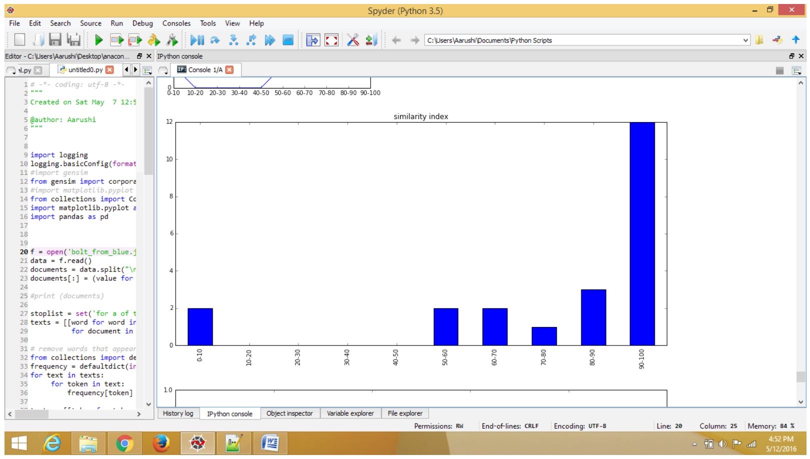


Figure 11. Graph showing the similarity index vs the number of documents for another query.

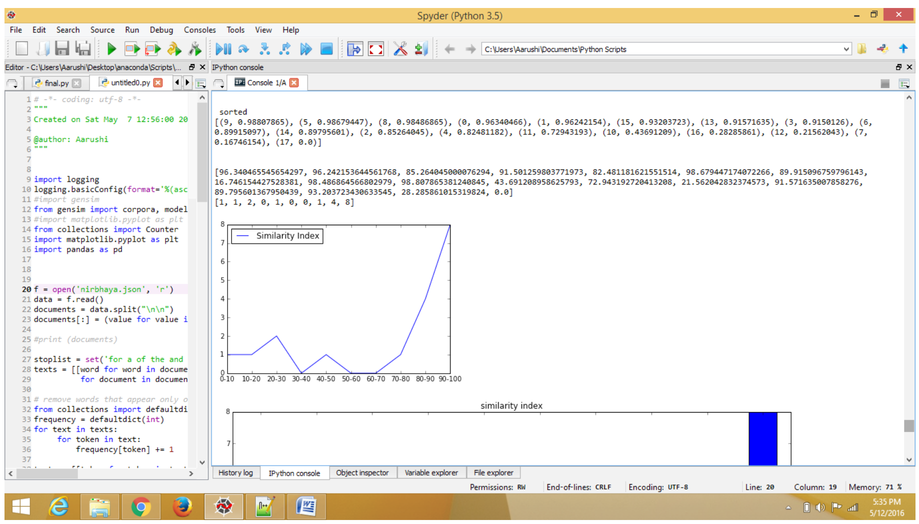


Figure 12. No. of retrieved URL's with similarity percentage lying in a particular range for another Query Nirbhaya Case Delhi.

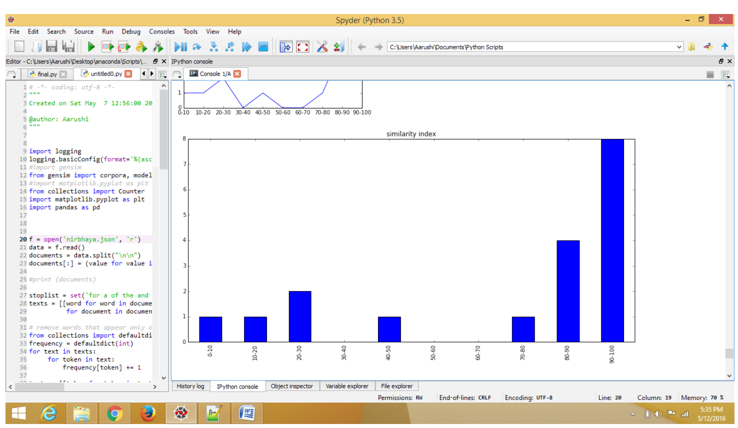


Figure 13. Graph showing the similarity index vs the number of documents for another query Nirbhaya Case Delhi.

5. Comparison with Related Work

The problem of information deluge and seeking relevant information has encouraged the researchers to work in this Area. This section briefly mentions some of the key contributions. How our work is different from their work? In [2] have proposed a personalized semantic (IR) where there's a system oriented for utilizing Semantic Web technology and Word Net ontology for supporting semantic IR capabilities. The proposed system also supports personalization through user model (UM) and text summarization. In[3] have proposed the idea of information retrieval from such documents which have both free text and semantically enriched markups. Shah in her paper has worked over a model where she worked with DAML+OIL marked up statements in queries. The idea was to improve retrieval performance by indexing text and markup together. In[4] have proposed the idea of information retrieval through semantics as with an approach of extracting information from web pages. For implementing their idea they collected documents related in some domain with the help of a crawler based on ontology and semantic content matching over search keyword entered by user. The goal behind this was to conquer such terms and queries which were semantically similar based on output given by Word net. In[5] have developed an information retrieval system based on ontology's. In their concept they have adapted the terminology specified for specific domains which were computed as to find a feature vector for every inbuilt concept. Later these vectors are used to augment user's query. The motivation of the authors was the anomaly in existing IR systems. In the existing systems users make use of ontology's to get a clear understanding of their information needs but the integration of these with traditional search is a major concern. In[6] have analyzed various ontology based IR systems. In addition, they have also performed a comparative analysis of all the available methods including vector space, probabilistic and semantic based techniques to provide developer with an appropriate choice for ontology based IR method. All the papers have done comparative analysis of different Information retrieval methods through manual methods. Nobody has given the real life example and how this type of analysis is done through software, is represented in our paper.

6. Conclusion and Future Work

To conclude, the proposed methodology helps in finding the relevance of the documents according to the user query. The similarity index obtained between the user's query and the list of large number of documents is used to separate the relevant and irrelevant documents. It is also used to rank the relevant documents in increasing order of the similarity measure to provide right, precise and accurate information to the user. The search using the semantics seeks to improve the search accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable data space. The use of NLP in our approach results in providing highly relevant search results which saves the time of the user in finding the right information.

Our proposed work can be extended in many ways. Our future work includes the followings:

1. Refining the proposed algorithm to improve the efficiency of the algorithm. It will result in more precise and accurate result according to the user's intent.
2. In this paper we had considered or studied only the Google search engine. However, the research can also be done in future on other search engines like Yahoo, Bing, Ask etc.
3. In this paper, we made two assumptions. One is finding the information using the web crawlers and the other is organizing the information using indexing in the databases. However, we can implement these two features and the algorithm for information retrieval resulting into the complete semantic search engine which the users can use to get highly relevant search results based on the query of the users.
4. We are also thinking of a new page ranking technique[7-10] in which geographical location from where the search is performed is considered as one of the factor for deciding the page rank. Currently existing techniques requires user to provide input choice option for their customised search whereas by employing our future model[11] the user gets a feel of personalised search without employing much effort.

7. References

1. Sahu SK, Saurabh JN. Comparative analysis of semantic and syntactic based search engines. Journal of Advanced Computing and Communication Technologies. 2016 Apr; 4(2):1–6.

2. Babekr ST, Fouad KM, Arshad N. Personalized semantic retrieval and summarization of web based documents. (IJACSA) International Journal of Advanced Computer Science and Applications. 2013; 4(1):177–86.
3. Shah U, Finin T, Joshi AR. Cost S, Mayfield J. Information retrieval on the semantic web. CIKM '02 Proceedings of the Eleventh International Conference on Information and Knowledge Management. 2002. p. 461–8.
4. Harb HM, Fouad KM. Semantic retrieval approach for web documents. (IJACSA) International Journal of Advanced Computer Science and Applications. 2011 Sep; 2(9):1–10.
5. Tomassen SL. Research on ontology-driven information retrieval. On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. 2006 Nov; 1460–68.
6. Sakthi Murugan RP, Aghila SBG. Ontology based information retrieval - An analysis. International Journal of Advanced Research in Computer Science and Software Engineering. 2013 Oct; 3(1):110–20.
7. Page L, Brin S, Motwani R, Winograd T. The page Rank citation ranking: bringing order to the web. Technical Report, Stanford Info Lab. 1998 Jan; 1–17.
8. Brin S, Page L. The Anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems. 1998 Apr; 30(1-7):107–17.
9. Xing W, Ghorbani A. Weighted Page Rank Algorithm. 2004 Proceedings Second Annual Conference on Communication Networks and Services Research. 2004 May. p. 305–14.
10. Ju S, Wang Z, Lv X. Improvement of page ranking algorithm based on timestamp and link. International Symposiums on Information Processing, Moscow. 2008 May; 36–40.
11. Sadhwani AA, Saxena N. A New Approach to Ranking Algorithm-Custom Personalized Searching. 2nd International Conference on Computing for Sustainable Global Development (INDIACom), Delhi. 2015. p. 130–33.