

# Search Engine Based on Semantic Web

Harsh Vora<sup>1</sup>, Zil Shah<sup>2</sup>, Shruti Gada<sup>3</sup>, Yash Bhatia<sup>4</sup>, Moushmee Kuri<sup>5</sup>

<sup>1,2,3,4</sup>Student, Watumull Institute of Electronics Engineering and Computer Technology, (Mumbai University)  
Plot No.157, C.H.M. Campus, Ulhasnagar, Mumbai, India 421 003

<sup>5</sup>Professor and Guide, Watumull Institute of Electronics Engineering and Computer Technology, (Mumbai University)  
Plot No.157, C.H.M. Campus, Ulhasnagar, Mumbai, India 421 003

**Abstract:** *With the rapid development of the World Wide Web, one of the main tools for people to get network information is search engine. However, the search results are widely condemned due to the lack of accuracy and redundancy disadvantages. The semantic web is a technology to save data in a machine-readable format that makes it possible for the machines to intelligently match that data with related data based on its semantics. This paper starts from the traditional search engine, and firstly introduces its classification, popular technology, advantages, disadvantages, and deep Knowledge on semantic-technology, thus leads to the semantic search engine model.*

**Keywords:** Semantic Search Engine, SPARQL, Search Engine, DBpedia, Ontology, LUIS.

## 1. Introduction

The idea of a search engine and information retrieval from the search engine is not a new concept. The interesting thing about the traditional search engine is that different search engines will provide a different result for the same query. While the information was available on the web, we have some fields of problem in search engines. It has become increasingly difficult to locate meaningful results from the list of returns typical of returned search queries. Keywords, often times, alone cannot capture the intended concept with high precision. An alternative approach is to constitute Web content in a manner such that it is easily machine-processable and to use quick-witted techniques to take an edge of these representations. Hence we come to a concept of SEARCH ENGINE BASED ON SEMANTIC WEB.

The motive of this work is to scrutinize some relevant issues on querying the semantic web in a context of semantic search engines and initiate a framework that opens the door for an effective search over the proposed search engine. Searching by keyword and not understanding the coexistence of many possible meanings for a word or phrase and synonymy are some reasons why the traditional search engine is not satisfactory anymore. Then we research the semantic search technology in depth, which can be divided into enhanced semantic search based on traditional search, knowledge semantic search based on ontology inference and other semantic search types. Based on this research we provide a framework. We call this framework as ARTH. ARTH is a Dynamic Web Application, which is a semantic search engine. Here, we push towards increasing the search accuracy by comprehending the searcher's purpose and the contextual meaning of terms as they appear in the searchable space.

## 2. Literature Review

The World Wide Web (WWW) allows people to share information or a certain data from the large database store house globally. We need to search the information with

specialized tools known generically as search engines. There are many search engines available today, where retrieving meaningful information is difficult. However to overcome this problem of retrieving meaningful information intelligently in common search engines, semantic web technologies are playing a major role. Semantic web is slowly gaining power and collaborating with other areas of research like bioinformatics, e-Commerce, e-Government and social web. The need for combining information in a meaningful way creates the potential and demand for research in Semantic web. Semantic Evaluation is done for all the words in every web page. As the traditional search engine only checks for the frequency of the word; this Search Engine looks for the relation that is the relatedness of the words in a web page. It has also upgraded the efficiency due to the changes in the keyword extraction algorithm.

## 3. Methodology

The current research required the following technologies for Development: Eclipse IDE, LUIS.AI, SPARQL query Language and DBpedia Ontology.

### 3.1 Technologies Used

#### 3.1.1 Eclipse

In computing, Eclipse is an integrated development environment (IDE) which is used for developing applications using the Java programming language. One can also use other programming languages such as C/C++, Python, PERL, Ruby etc. to develop applications. The Eclipse platform which provides the base for the Eclipse IDE contains plug-ins and is outlined to be expandable using additional plug-ins. The Eclipse platform is developed using JAVA that can be used to develop rich client applications, integrated development environments and other tools. Eclipse is a platform that can be used as an IDE for any programming language for which a plug-in is available. In this project, we make the use of Eclipse to make a Dynamic Web Application which works as our search engine, ARTH.

### 3.1.2 LUIS

LUIS stands for Language Understanding Intelligent Service. It is a Cognitive Service provided by Microsoft (under Azure Cognitive Services). Using this service we can use REST APIs to extract necessary details namely 'intent', 'entity', 'phrases' etc. from any statement. We can use it for building intelligent apps that can converse and understand what a statement is trying to convey. For LUIS every statement is an "Utterance". Based on all its experience, LUIS tries to pry for the "Intent" from the sentence as in what speaker is trying to 'do'. Moreover, it tries to decipher the 'entity' from the statement as to what the 'intent' is trying to refer to. In this project, LUIS is trained for such utterances to figure out the intents and entities.

### 3.1.3 SPARQL

SPARQL is the standard query language and protocol for Linked Open Data on the web or for semantic graph databases (also called RDF triplestores). SPARQL, short for "SPARQL Protocol and RDF Query Language", helps the users to query information from databases or any data source that can be mapped to RDF. The SPARQL standard is designed and endorsed by the W3C and helps users and developers focus on what they would like to know instead of how a database is organized. The SPARQL Query Language is a Declarative Query Language (like SQL) for performing Data Manipulation and Data Definition operations on Data that is represented as a collection of RDF Language statements.

### 3.1.4 DBpedia Ontology

The DBpedia Ontology is a cross-domain ontology, which has been created manually based on the most commonly used information platform -Wikipedia. The ontology currently covers 685 classes which form a subsumption hierarchy and are described by 2,795 different properties. This ontology is generated from the manually created specifications in the "DBpedia Mappings Wiki". Each release of ontology corresponds to a new release of the DBpedia data set which consists of instance data drawn out from the different language versions of Wikipedia.

### 3.1.5 JENA

Jena is a Java framework which is used for building Semantic Web applications. It provides a large amount of Java libraries which serves developers to develop code that handles RDF, RDFS, RDFa, OWL and SPARQL in line with published W3C recommendations. Jena incorporates a rule-based inference engine which helps to perform reasoning based on OWL and RDFS ontologies.

## 3.2 Working Methodology of ARTH

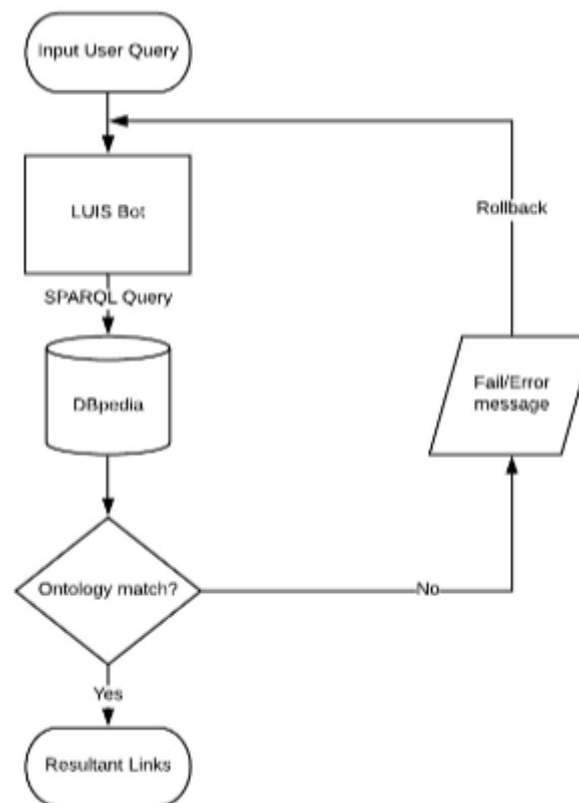


Figure 3.1: Flowchart of working of ARTH

ARTH is a semantic search engine which takes natural language query (which is the user query for search) as input. This query, with the help of Eclipse IDE, is received by LUIS. LUIS, which is trained to extract the intents and entities from the natural language query (utterances), sends these intents and entities to the IDE where a SPARQL query is generated as output. This SPARQL query is used to fetch the data from the DBpedia Ontology, where the SPARQL query is used to check if there is any ontology match. If any ontology matches are found, then the respective links, along with the film genre and its release date, appear on the result page. If no matches are found then the user is notified and is asked to enter a valid query.

Table 1: Examples of Valid Queries

Sr.No.	Examples of Queries that a User can Search For.
1	Films from 2010 (Search according to the Release Date)
2	Comedy Films (Search according to the Movie Genre)
3	30 Films (Search according to the number of Movies)
4	10 Comedy Films from 2014

As shown in the table 1, when a user wants to search about any information of a particular film, he has the facility to search by entering the number of films he wants, in general, or by entering the Genre of films that he wishes to look for or by entering the Release date of a film. With the help of this, the user can get a list of films which come under the user's desired Genre of films. The user can also get a list of films which were released in his desired year.

When a user enters any required query, the films that appear on the list are been programmed to be shown in a descending

order of the released date of the film. Which means, that for a particular search, the list of films will be displayed according to their released date, starting from December and going upto January.

#### 4. Results

The screenshots below show how ARTH gives more accurate and reliable results than the traditional search engine (here we take Google as an example). When the user enters a query which says “10 films from 2010”, Google provides with links which do not necessarily show only 10 films from 2010 whereas, ARTH provides with 10 results which are links to 10 films from 2010 based on the descending order of their released dates (that is it starts from December). Also Google does not give direct results which provide the names of the films, whereas it provides with links that are linked to some other intermediate pages, hence increasing the number of steps for the user. ARTH enables the user to reach the required results within less number of steps, which makes this framework more efficient.

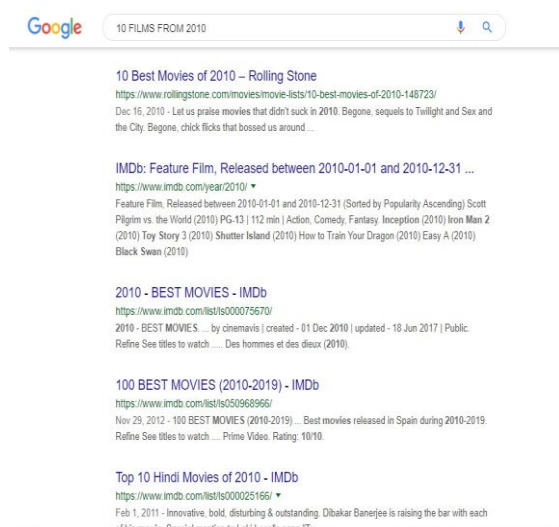


Figure 4.1: Screenshot of results on GOOGLE

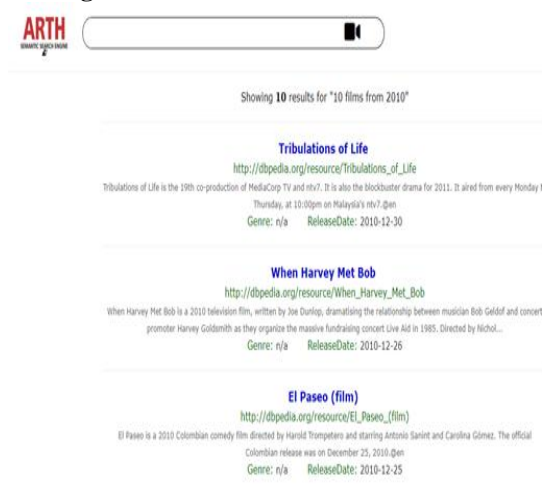


Figure 4.2: Screenshot of results on ARTH

#### 5. Conclusions

In this paper, we have discussed how semantic search engines can be more reliable platforms, for information

retrieval, as compared to the traditional search engines. We also provide a framework called ARTH, which is based on semantic search. Further, we show results which prove how semantic search engines show more reliable and accurate results as compared to traditional search engines. Hence, making Semantic Search Engines more efficient.

#### 6. Future Scope

As a part of our future work, we plan to add more detailed source code query analysis to our Eclipse IDE. As ARTH is only an experiment that helps analyze and prove how semantic web technology can prove to be an asset over the traditional web, we focused our search by choosing “FILMS” as our domain. For future, we plan on expanding this area of search based on the availability of resources. We also plan on making this framework as open source search engine which can be available to the community.

#### References

- [1] Eclipse, “Eclipse-Overview” [Online] Available: [www.tutorialspoint.com/eclipse/eclipse\\_overview.htm](http://www.tutorialspoint.com/eclipse/eclipse_overview.htm). [Accessed: Mar. 26, 2019].
- [2] Luis, “What is LUIS.AI?” [Online] Available: <https://medium.com/ai-for-developers/what-is-luis-ai-8ef7f972b7f7>. [Accessed: Mar. 26, 2019].
- [3] SPARQL, “What is SPARQL?” [Online] Available: <https://www.ontotext.com/knowledgehub/fundamentals/what-is-sparql/>. [Accessed: Mar. 27, 2019].
- [4] DBpedia, “Ontology” [Online] Available: <https://wiki.dbpedia.org/services-resources/ontology>. [Accessed: Mar. 25, 2019].
- [5] Liyang Yu, “Introduction to the Semantic Web and Semantic Web Services”, Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, London, New York, 14th June 2007.
- [6] Frank van Harmelen and Grigoris Antoniou, “A Semantic Web Primer”, The MIT Press, Cambridge, Massachusetts, London, England, 2nd Edition 2004.
- [7] K.K. Breitman, M.A. Casanova and W. Truszkowski “NASA Monographs in Systems and Software Engineering”, Springer-Verlag London Limited, 2007.
- [8] T. Burners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific American 284 (5).
- [9] Noryusliza Abdullah, Rosziati Ibrahim “Semantic Web Search Engine Using Ontology, Clustering and Personalization Techniques”, Murgante B. et al. (eds) Computational Science and Its Applications – ICCSA 2012. Lecture Notes in Computer Science Springer, Berlin, Heidelberg. Vol 7336 Part IV, pp. 364–378, 2012.
- [10] Sharmela Shaik, Prathyusha Kanakam, S Mahaboob Hussain, D. Suryanarayana “Transforming Natural Language Query to SPARQL for Semantic Information Retrieval”, International Journal of Engineering Trends and Technology (IJETT), Volume-41 No. 7, pp. 347-350, November 2016.
- [11] Nikhil Chitre “Semantic Web Search Engine”, International Journal of Advance Research in Computer

Science and Management Studies (IJARCSMS),  
Volume 4 Issue 7, pp. 47-52, July 2016.

- [12] Arooj Fatima, Cristina Luca, George Wilson “New Framework for Semantic Search Engine”, UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 446-451, 2014.
- [13] W. Wei, P. M. Barnaghi, and A. Bargiela, “The Anatomy and Design of A Semantic Search Engine,” UNMC-CS-200712, School of Computer Science, University of Nottingham Malaysia Campus, Tech. Rep., 2007.

## Author Profile



**Harsh Vora** is pursuing B.E. Computer Engineering from Watumull Institute of Electronics Engineering & Computer Technology, Mumbai University. He is interested in Semantic Web, Machine Learning, Artificial Intelligence and Developing Intelligent

Systems.



**Zil Shah** is pursuing B.E. Computer Engineering from Watumull Institute of Electronics Engineering & Computer Technology, Mumbai University. She is interested to gain the understanding of Semantic Web, Machine Learning and Artificial Intelligence and work ahead in the said fields.



**Shruti Gada** is pursuing B.E. Computer Engineering from Watumull Institute of Electronics Engineering & Computer Technology, Mumbai University. She is interested in Semantic Web, Semantic Search Engine, Database Structure Analysis and Software Quality Assurance.



**Yash Bhatia** is pursuing B.E. Computer Engineering from Watumull Institute of Electronics Engineering & Computer Technology, Mumbai University. He is interested in Data Science, Cyber Security, Machine Learning and Artificial Intelligence.



**Prof. Moushmee Kuri** has done B.E. in Computer Engineering from Goa Engineering College & M.E. in Computers from Thadomal Shahani College of Engineering. Her fields of Interest are Data Structures, Semantic Web, Compilers, and Cryptography &

Security.