International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

ANNOTATION OF VARIOUS WEB SEARCH RESULTS

B Ramadasu[1] K Kiran Prakash[2] Dr. G.N.R Prasad[3]

[1]*Assistant Professor, Department of CSE,Chaitanya Bharathi Institute of Chaitanya Bharathi Institute of Hyderabad, India.*
[2]*Assistant Professor, Department of CSE,Chaitanya Bharathi Institute of Chaitanya Bharathi Institute of Hyderabad, India.*
[3]*Sr. Assistant Professor, Department of MCA, Chaitanya Bharathi Institute of Chaitanya Bharathi Institute of Hyderabad, India.*

ABSTRACT : The use of Internet for searching the information to cooking in kitchen is increasing. Slowly, it is becoming the essential item in our daily life. One of the most active areas of Artificial Intelligence (AI) research is natural language processing (NLP). Numerous NLP technologies, including chatbots, and sentiment analysis software, increase productivity and efficiency in innumerable enterprises all over the world. Recent advances in NLP have even suggested a potential for assisting the speech-impaired in freely communicating with ASR systems and those around them. However, without text annotation and the businesses that offer these annotation services, none of these incredible innovations would be conceivable. An explanation or clarification is added to a text or diagram using annotation. Additionally known as marginal notes. For many search engines, the data encoded in the returned result pages comes from the underlying structured databases because a substantial percentage of the deep web is database-based. Such search engines are frequently referred to as Web Databases (WDB). Multiple search result records can be found on a typical result page generated by a WDB (SRRs). Multiple data units, one for each aspect of a real-world thing, are included in each SRR. The SRRs each represent a single book with several data components. A text fragment that semantically represents one idea of an entity is referred to as a data unit. It is equivalent to a record's value for an attribute. It differs from a text node, which is a group of text enclosed in two Hypertext Markup Language (HTML) tags. In this study, we annotate data at the data unit level. There is a lot of need for gathering relevant data from various WDBs. For instance, a book comparison shopping system must evaluate whether any two SRRs pertain to the same book after gathering several result records from various book sites. This may be accomplished by comparing the International Standard Book Number (ISBN). If ISBNs aren't accessible, you might compare the authors and titles instead. The system must also display a price comparison between each site's prices. As a result, the system has to be aware of each data unit's semantic. Unfortunately, result pages sometimes do not include the semantic labels of data units. Semantic labels for data units are crucial for the record linkage operation mentioned above as well as for saving gathered SRRs into a database table for future analysis. Early programmes' capacity to scale is significantly constrained by the enormous human labour required to manually label data units. In this study, we explore the possibility of automatically labelling the data units contained in the SRRs returned by WDBs.
Keywords : Search, Web Databases, HTML, ISBN

I INTRODUCTION

An existing system defines a data unit as a textual representation of one notion of an entity. It is equivalent to a record's value for an attribute. It differs from a text node, which denotes a group of text enclosed between

DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

two HTML elements. It thoroughly explains the connections between text nodes and data units. In this study, we annotate data at the data unit level. There is a lot of need for gathering relevant data from various WDBs. For instance, a book comparison shopping system must evaluate whether any two SRRs pertain to the same book after gathering several result records from various book sites. The primary drawbacks of the current approach are that authors and titles can be compared in the absence of ISBNs. The system must also display a price comparison between each site's prices. As a result, the system has to be aware of each data unit's semantic. Unfortunately, result pages sometimes do not include the semantic labels of data units. As an illustration, no semantic labels are provided for the values of title, author, publisher, etc. Semantic labels for data units are necessary for both the aforementioned record linkage operation and for the storage of gathered SRRs into a database table. In the proposed system, we think about how to automatically identify the data units in the SRRs that WDBs provide. Given a set of SRRs that have been extracted from a result page returned from a WDB, Our automatic annotation solution consists of three steps given a collection of SRRs that have been taken out of a result page received from a WDB. The benefits are We carefully examine the connections between HTML text nodes and data units, unlike most existing techniques that just give labels to each HTML text node. We annotate data at the unit level. To align data units into various groups so that the data units inside a group have the same semantic, we suggest a clustering-based shifting approach Our approach takes into account additional significant features shared among data units, such as their data types (DT), data contents (DC), presentation styles (PS), and adjacency (AD) information, in addition to the DOM tree or other HTML tag tree structures of the SRRs (as most current methods do). To improve data unit annotation, we make use of the integrated interface schema (IIS) across several WDBs in the same domain. We are the first to use IIS for SRR annotation, to the best of our knowledge. We use six fundamental annotators, each of which has the ability to independently label data units depending on certain characteristics of the data units. In order to aggregate the output from several annotators into a single label, we also use a probabilistic model. This paradigm is quite adaptable, allowing for easy modification of the current basic annotators and the addition of new ones without impairing the functionality of the present annotators. For every WDB, we create an annotation wrapper. The wrapper may be used to quickly add additional queries to the SRRs that were obtained from the same WDB.

II ARCHITECTURE

The architecture diagram primarily illustrates the request flow from users to databases via servers. The presentation layer, business layer, and data link layer are the three layers that make up the entire system in this situation. Three-tier architecture was used in the creation of this project.

A 3-Tier Architecture

In order to address the shortcomings of the two-tier architecture, the three-tier software architecture (also known as a three layer architecture) was developed in the 1990s. Between the user interface (client) and data management (server) components lies the third tier, sometimes known as the middle tier server. By offering services like queuing, application execution, and database staging, this middle tier offers process management where business logic and rules are put into action and can support hundreds of users (as opposed to only 100 users with the two layer design). When a distributed client/server design is required that offers (in comparison to the two tier) better performance, flexibility, maintainability, reusability, and scalability, while concealing the complexities of distributed processing from the user, the three tier

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

architecture is utilised. Three layer architectures are a common option for Internet applications and net-centric information systems because of these qualities.

These characteristics have made three layer architectures a popular choice for Internet applications and net-centric information systems. The main utilities of Three-Tier architecture are it **s**eparates functionality from presentation , it has clear separation , better understanding of the system is possible. The changes limited to well define components, it can be running on World Wide Web (WWW) but effective network performance**.**

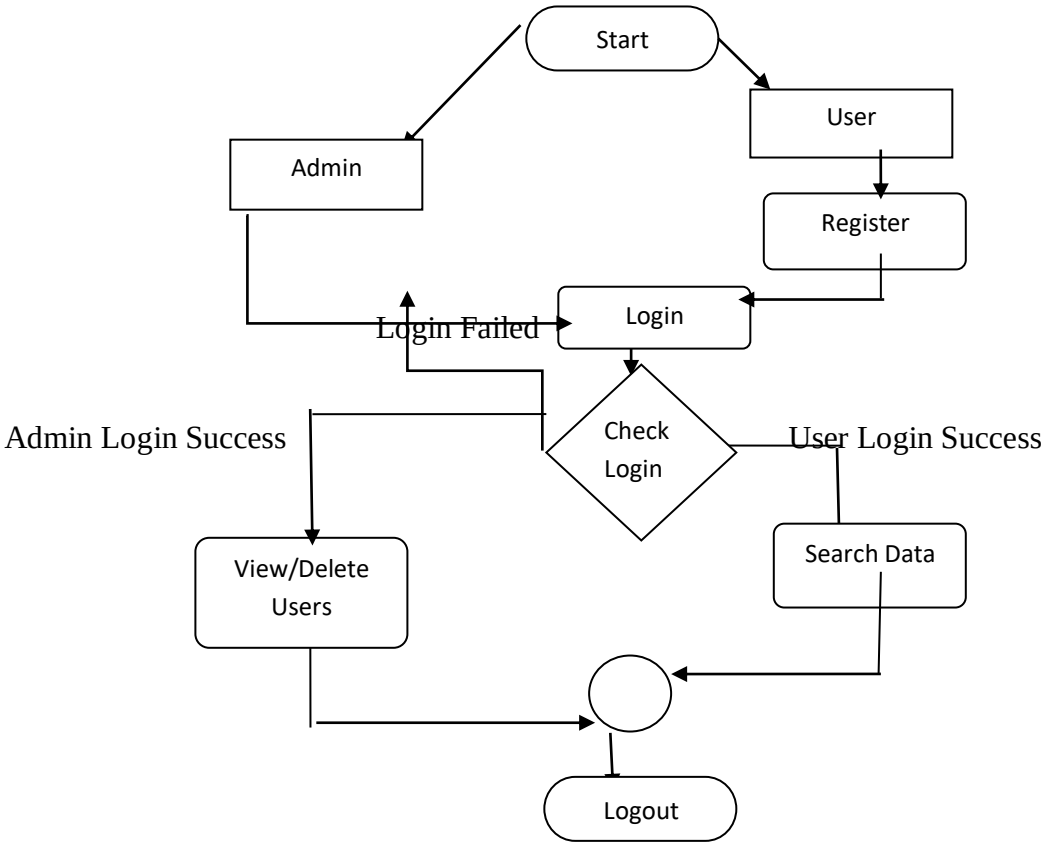

Figure : Architecture Flow

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

Column Name	Data Type	Allow Nulls
Id	int	☑
U_Name	nvarchar(50)	☑
U_Password	nvarchar(50)	☑
U_Mail	nvarchar(50)	☐
U_Mobile	nvarchar(50)	☑
U_Dob	nvarchar(50)	☑
U_Address	nvarchar(50)	☑
		☐

User_Details
Table : User Details

Two modules are proposed in this system. The User Module and Admin Module are involved to make this system to work efficiently.

User Module: Here in the user module user need to login with some login details which is get through a small registration. After user login user can search the results though some key words.
Admin Module: Here in the Admin module admin can check and view the user details who are using our application. And the admin have right to delete user whom is use our application for bad purpose.

IV RESULTS

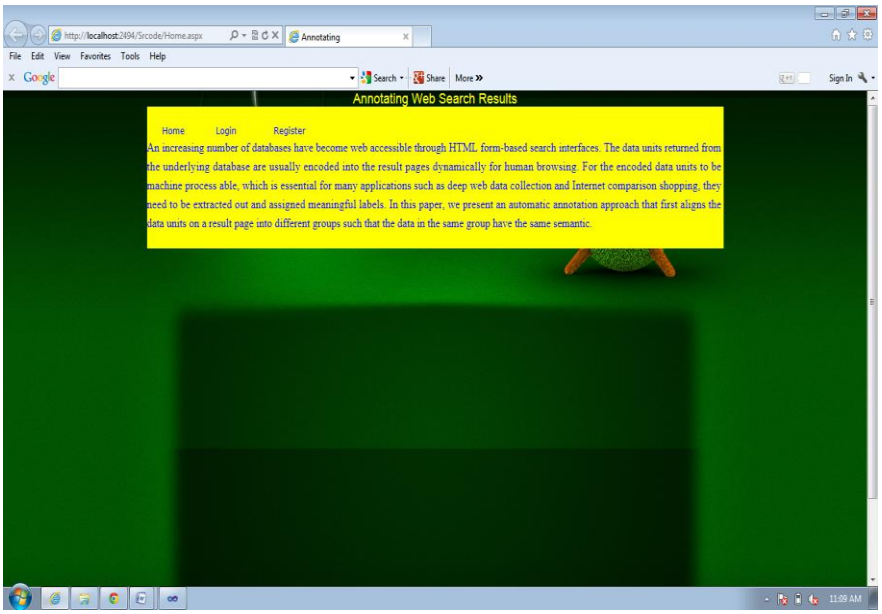

Figure : Home Page

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

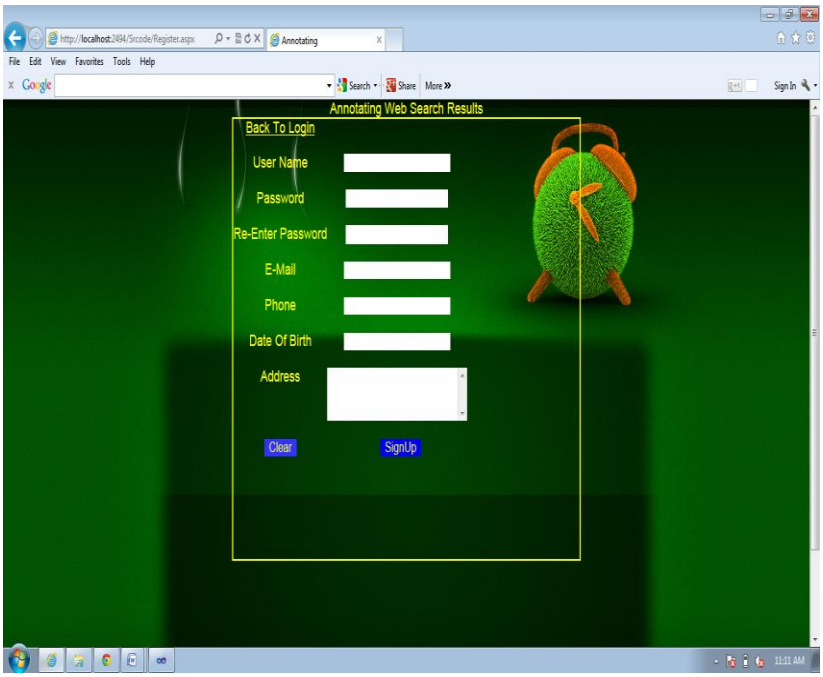

Figure : Registration Page

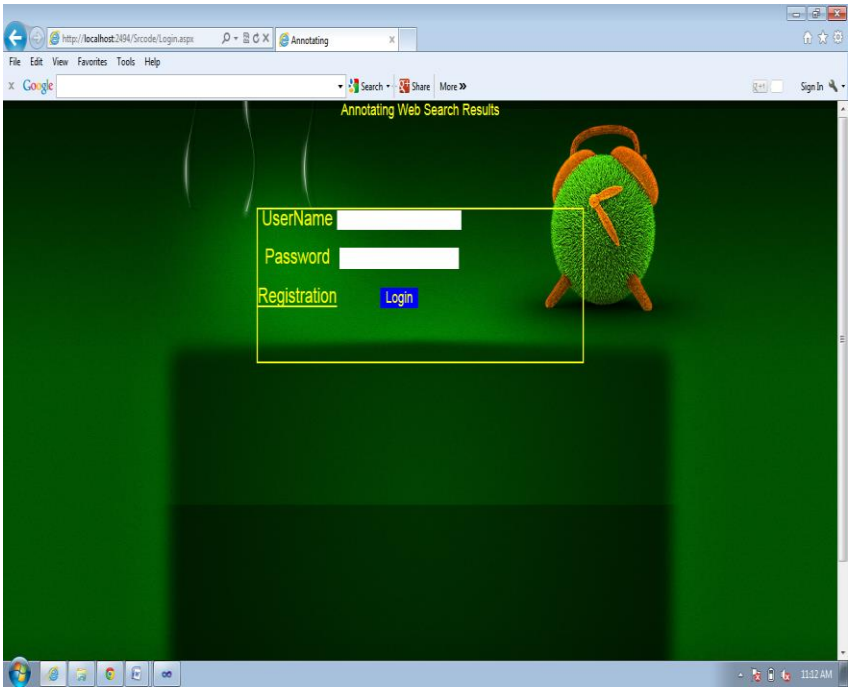

Figure : Login Page for Admin and User

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

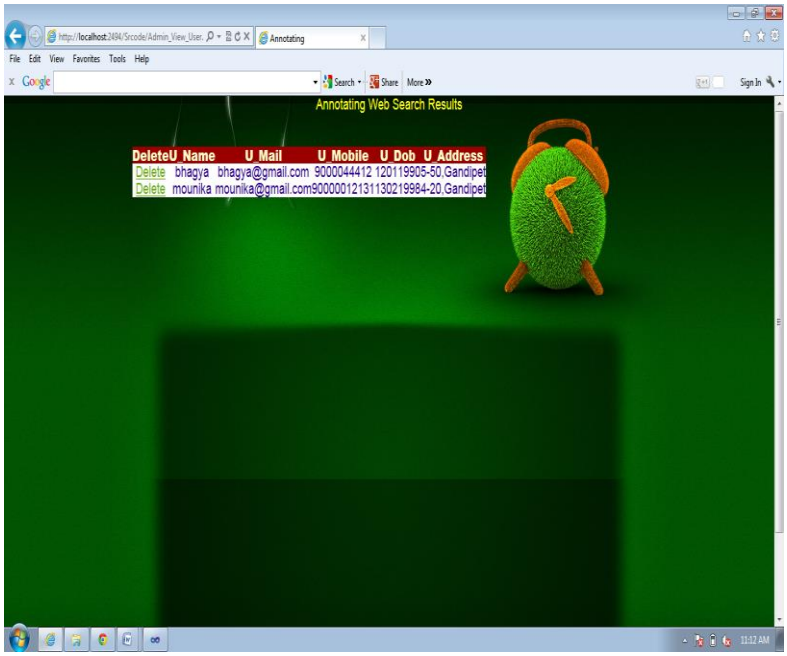

Figure : Displaying the list of Users for Admin

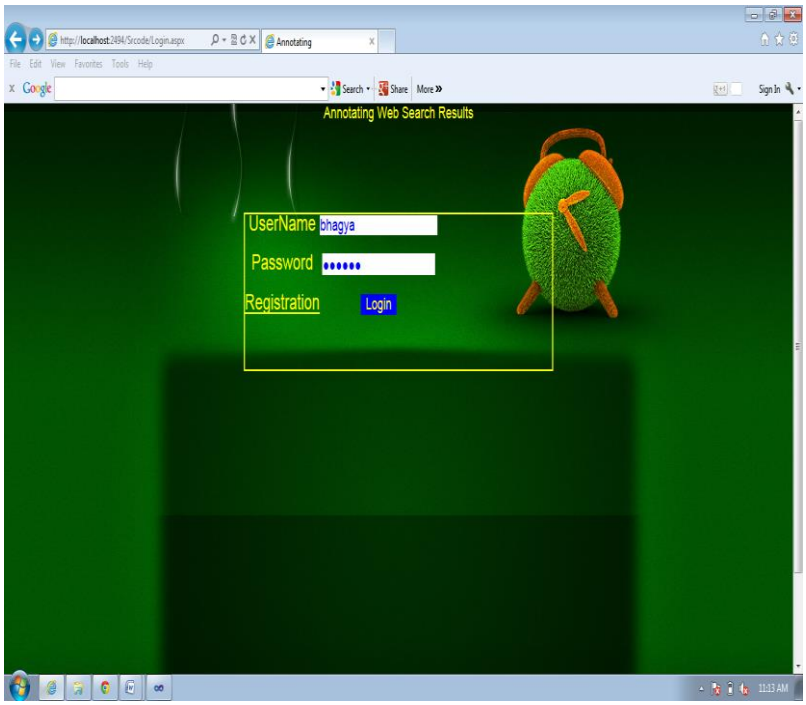

Figure : Login to the system with user credentials

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

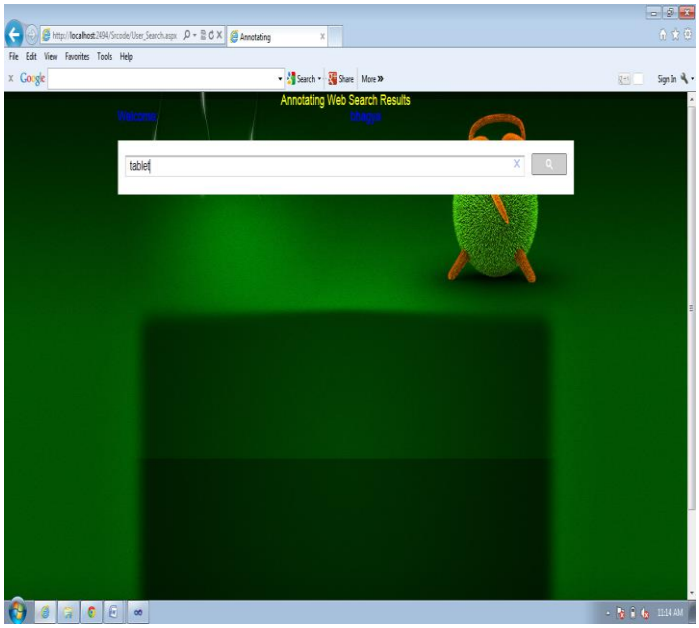

Figure : GUI to enter Search Query

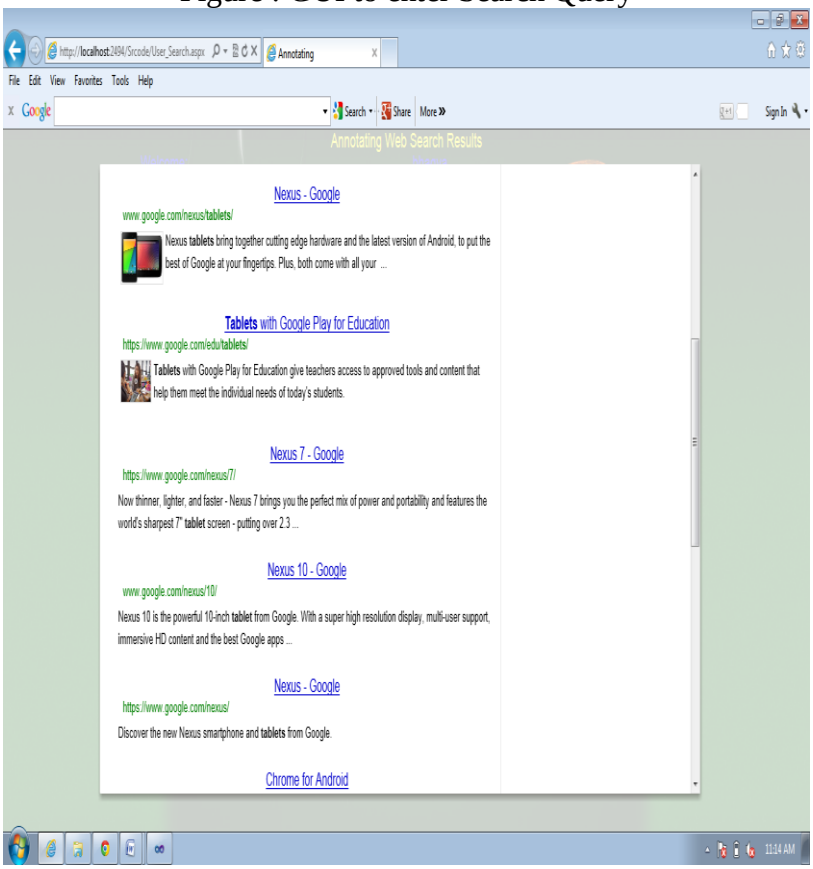

Figure : Search Results

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

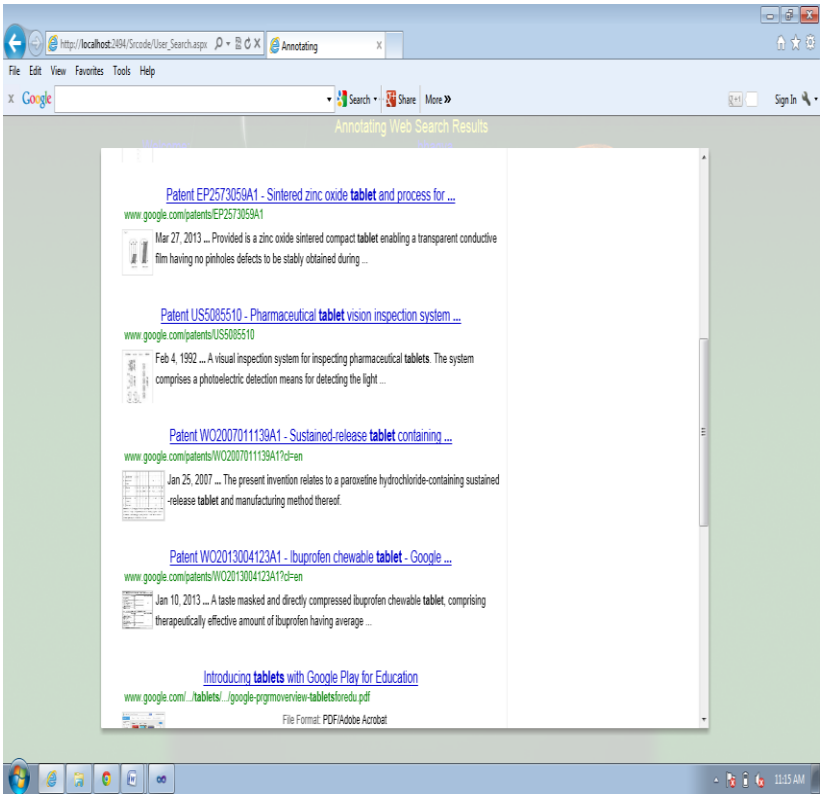

Figure : Search Results

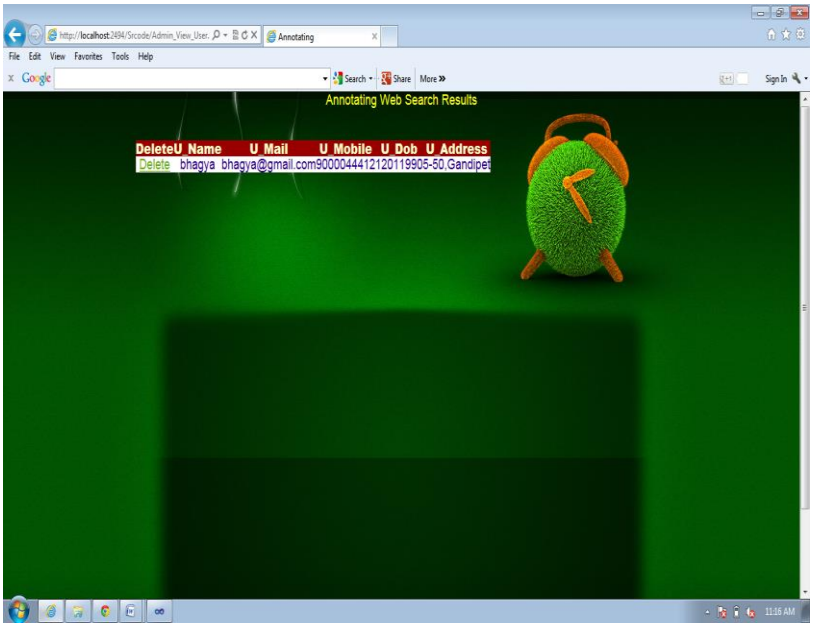

Figure : After deleting the user

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 4 Volume No.6 July – 2022
DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

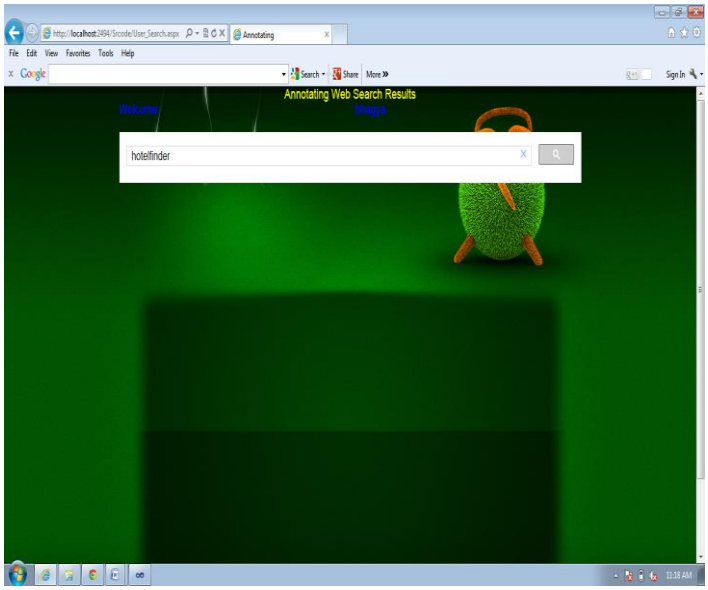

Figure : User Search Page

V CONCLUSION

In this project, we investigated the issue of data annotation and put forth a multi-annotator strategy for creating an annotation wrapper that would automatically annotate the search result records collected from any given online database. The six fundamental annotators in this methodology are combined using a probabilistic mechanism. Our experimental findings demonstrate the usefulness of each of these annotators and their ability to provide high-quality annotation when used collectively. Each of these annotators uses a certain sort of feature for annotation. . Our solution is unique in that it makes use of both the LIS and IIS of several online databases within the same domain when annotating the results returned from a web database. We also discussed how the inconsistent label and inadequate local interface schema issues may be resolved with the usage of IIS. We looked at the issue of automatic data alignment in this project. To get a comprehensive and accurate annotation, proper alignment is essential. For instance, when there are no explicit separators, we need to improve our mechanism for splitting composite text nodes.

REFERENCES

[1] A. Arasu and H. Garcia-Molina, "Extracting Structured Data from Web Pages," Proc. SIGMOD Int'l Conf. Management of Data, 2003.
[2] L. Arlotta, V. Crescenzi, G. Mecca, and P. Merialdo, "Automatic Annotation of Data Extracted from Large Web Sites," Proc. Sixth Int'l Workshop the Web and Databases (WebDB), 2003.

DOI:10.46647/ijetms.2022.v06i04.0074 ISSN: 2581-4621

[3] P. Chan and S. Stolfo, "Experiments on Multistrategy Learning by Meta-Learning," Proc. Second Int'l Conf. Information and Knowledge Management (CIKM), 1993.
[4] W. Bruce Croft, "Combining Approaches for Information Retrieval," Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval, Kluwer Academic, 2000.
[5] V. Crescenzi, G. Mecca, and P. Merialdo, "RoadRUNNER: Towards Automatic Data Extraction from Large Web Sites," Proc. Very Large Data Bases (VLDB) Conf., 2001.
[6] S. Dill et al., "SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation," Proc. 12th Int'l Conf. World Wide Web (WWW) Conf., 2003.
[7] H. Elmeleegy, J. Madhavan, and A. Halevy, "Harvesting Relational Tables from Lists on the Web," Proc. Very Large Databases (VLDB) Conf., 2009.
[8] D. Embley, D. Campbell, Y. Jiang, S. Liddle, D. Lonsdale, Y. Ng, and R. Smith, "Conceptual-Model-Based Data Extraction from Multiple-Record Web Pages," Data and Knowledge Eng., vol. 31, no. 3, pp. 227-251, 1999.
[9] D. Freitag, "Multistrategy Learning for Information Extraction," Proc. 15th Int'l Conf. Machine Learning (ICML), 1998.
[10] D. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, 1989.
[11] Prasad GNR, "Intelligent-based course material production, distribution and tracking system", International journal of progressive research in engineering management and science (IJPREMS) Vol. 02, Issue 05.