The solutions for Bioinformatics that identify relevant background knowledge in textual documents, such as scientific publications or in database annotations, have gained enormous significance. It is obvious to the most that the focus of attention should be on the underlying problem rather than on what happens to the scientific literature since the scientists / biologists need accurate & rapid access to scientific journals. One has to keep up with an increasing number of publications. There is a growing demand to 'translate' information from text into more computable forms and to cross-link the information with relevant biological databases. One has to read so many results or records to search for scientific literature, for example, literature on cancer articles published in journals deposited in PubMed database over a period of time.

The web has greatly improved access to scientific literature. However, scientific articles on the web are largely disorganized, with research articles being spread across archive sites, institution sites, journal sites, and researcher homepages. The necessary data is widely available over internet and many kinds of data pose the current challenge in storage and retrieval. Datasets can be made more accessible, re-usable and richer in content through annotation and aggregation, as well as linking to other types of the data.
Biomedical information is growing explosively and new useful results are appearing everyday in research publications. However, getting relevant information and resources is a big challenge today in the present information world.

To overcome this problem one can focus on retrieving the data enabling us to search faster with refined records. This can be achieved by including more search terms. Hence, a new approach has been proposed in the present research. The implementation includes python program to connect a scientific database, such as, PubMed, which asks the user for a search term, start Year and end Year and cut off value to limit the number of records, and finally retrieves the records. This way the search can be faster and the user can store the records and can have access offline. Further, refining of the output can be done by selected entries to retrieve ID, Title, Authors, Abstracts and Source. On the other hand, author based search has also been written to enable specific author retrieval data from database.

This work has been implemented using Biopython through the Entrez database. The Entrez Programming Utilities (eUtils) are a set of seven server-side programs that provide a stable interface into the Entrez query and database system. The eUtils use a fixed URL syntax that translates a standard set of input parameters into the values necessary for various software components to search for and retrieve the requested data.
On the other hand, the continuous and rapidly growing volume of scientific literature and increasing diversification of inter-disciplinary fields of science and their answers to unsolved problems in medical and allied fields of science, present a major problem to scientists and librarians. Data and databases have assured prime importance with the advent of computers resulting in efficient storage and retrieval of information for the advancement of scientific research to the next level. Such an easy flow of information has become possible through various stochastic algorithms implemented in data storage and the search methodologies of mining data from databases.

It is observed that today as many as 4800 scientific journals exist in the internet of which some are online only. From researchers’ point of view, the problem is amplified when we consider today’s competition where we may not be able to spend sufficient time on experimental work owing to the fund of information already published. Therefore considering these facts on one hand and the volume of serials on the other, a study has been initiated in evaluating the scientific literature published in various journal sources. The scope of the study does not permit inclusion of all periodicals in the extensive fields of biology and hence a text mining routine was employed to extract data based on keywords. NCBI PubMed literature database was selected for this study. Initially a generalized search without any limits has been employed to retrieve articles related to bioinformatics and computational biology. The Title and
abstract search have been considered in this study because the Title field in some articles refers to the most important keywords relating to the subject.

In all the cases presented in this work, National Centre for Biotechnology Information (NCBI) PubMed database has been selected because an attempt has been made to compare the efficiency of data retrieval and search options employed in four databases, namely, NCBI, IEEE, Citeseer and ACM databases, respectively. A thorough study on the allied fields of search databases represented more fields and is more advanced in PubMed than in IEEE, Citeseer and ACM databases. From this analysis, our approach with PubMed database has been found to be more user-friendly and more advanced than the rest of the databases.