

Work Plan:

In order to achieve the above mentioned objectives the following methodology will be followed:

- **Systematic Review of Literature and Research Work:** A systematic review of Information Retrieval models and ongoing research work in information retrieval in websites will be taken, by keeping track of latest research. This would be done by reading the related articles published in last 10 years journals with impact factor.
- **Analysis of Role of information retrieval techniques:** The significance of Information retrieval techniques would be analyzed in web applications.
- **Identification of the areas of Improvement:** The issues related to information access which needs to be tackled, and which need further research would be explored.
- **Use of information retrieval methods:** The algorithms may be refined and would be applied to improve upon the effectiveness of web search.
- **Assessment of the proposed methods:** The proposed methods would be evaluated upon various criteria.
- **Testing and Evaluation of model:** Experiments would be performed to check the performance of the solution, using test cases.
- **Conclusion:** On the basis of the result, the conclusions shall be drawn.
- **Report Generation:** A thesis report shall be prepared on the basis of the research work undergone. The work being proposed & the results of the research shall be published.

Time Plan:

The work is planned to be carried out in following phases within approximately two years of time period:

Activity / Months	2	4	6	8	10	12	14	16	18	20	22	24
Literature Survey & Develop Questionnaire												
Review of Models of diff. Search Engine and Analysis of Data												
Identification of areas of Improvements												
Enhancement of model of Search Engine												
Evaluation and Comparison												
Testing												
Thesis Writing												

Research Methodology:

The above-mentioned objectives can be achieved by carrying a proper and planned research involving different types and methods. The data collected will lay the foundations for the study and will give a platform for the analysis and findings, and then will be updating in existing technology which lead to the fulfillment of the objectives.

The types of research design will be used in this study are descriptive, exploratory in nature.

Descriptive research includes surveys and fact-finding enquiries of different kinds. Primary data will be collected by observation, interviews and questionnaires through different available technologies, and knowledge seekers from different stratum like researchers, students and developers. Observations will be carried through existing search engines available on internet. The questionnaire will be used to know about the knowledge seekers perception about existing search engines. Data will be also collected by direct interaction with the users ie through interviews to know their opinion about the information retrieval through search engines.

While Exploratory research includes detail study of existing data (technologies available). The basic objective of the study is to explore and obtain clarity about the problem situation. It involves qualitative investigation. This data will be collected through the websites, journals, articles and research papers published in journals and different books on the unstructured data mining information retrieval techniques. And be used to know about the different techniques used to retrieve information through search engines and their limitations and future scope by considering different parameters like time response, accuracy of outcomes, importance of the results and relevancy of results.

The data collection and analysis paves way for the recommendation and conclusion of the study that reveals some important findings regarding the strategy.

Type of Study: Exploratory

Research Method: Experimental

Nature of Study: Qualitative

Type of Analysis: Statistical

Sample Design: Random Sampling

Sample Size: 30 % of population

Source of Data Collection: Primary as well as Secondary

Scope of the Research:

The concept of Information Retrieval is very vast and too many models of search engines are available in the market. In this research all the information retrieval techniques will not be considered. Even for Comparative study not all models of search engines will be studies or compared. That is this research is limited with comparison of very few models and also testing will be done with small data not with big data.

Tools and Techniques:

To carry out purposed research, a few techniques and tools shall be required for performing different tasks. A brief summary of these tools and techniques is given below. This is not an exhaustive list. During research, if a new technique is found, it may be integrated into the work.

A. Statistical Tools:

For performing comparative data analysis any one of the statistical analysis tools may be used:

- **MS-Excel**
- **MATLAB**
- **SPSS**

B. Graphical Tools:

For diagrams and flowcharts, any one of the tools may be used:

- **MS Visio-** *Microsoft Visio* (formerly Microsoft Office Visio) is a diagramming and vector graphics application
- **Gliffy-** *Gliffy.com* is a web-based diagram editor. Create and share flowcharts, network diagrams, floor plans, user interface designs and other drawings online.
- **LucidChart-** *Lucidchart* is your solution for visual communication. Flow charts, diagrams, UML sketches, and ER models have never been easier.
- **DrawAnywhere-** *Draw Anywhere, Anything, Any time* - complete web-based diagramming solutions. Create and share flow charts, block diagrams, process diagrams, Organizational charts and more.

C. Design/Modeling Techniques:

For modeling purpose, any of the following approaches may be used:

- Probabilistic or Statistical Model
- **Filtering algorithms-** Filtering algorithms filter out portions of the text that cannot possibly contain a match, and, at the same time, find positions that can possibly match.

- **Ranking algorithms-** PageRank is an [algorithm](#) used by [Google Search](#) to rank websites in their search engine results.
- **Indexing algorithms-** Search engine indexing collects, parses, and stores [data](#) to facilitate fast and accurate [information retrieval](#).
- **Language models-** A statistical language model assigns a [probability](#) to a sequence of words. Language modeling is used in many [natural language processing](#) applications such as [speech recognition](#), [machine translation](#), [part-of-speech tagging](#), [parsing](#) and [information retrieval](#).
- **Semantic Web-** The Semantic Web is the extension of the World Wide Web that enables people to share *content* beyond the boundaries of applications and websites.
- **Machine Learning-** Machine learning, a branch of [artificial intelligence](#), concerns the construction and study of systems that can [learn](#) from data.

D. Testing/Evaluation Techniques:

For evaluating, any one or combination of the following techniques shall be used:

- **Weka Mining Tool-** Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.
- **Test Collections-** Used for text categorization (Test collections are standard data sets used to measure the effectiveness of information retrieval systems. Most were originally developed to support research on IR, but practitioners often find them useful as well. The most widely used text categorization test collection.)

E. Implementation Techniques:

- **Dot Net**
- **MySQL**