ABSTRACT

Recent research directions in Computer Science is closely related to two areas – machine learning and data mining that have developed a method for constructing statistical models of network data. Such models include social networks, networks of web pages, complex relational databases and data about inter-relation people, places, things and events extracted from text documents. Research in a number of academic fields has shown that social networks operates on many levels from families up to the level of nations and play a critical role in determining the way in which problems are solved, organizations are run and the degree to which individuals succeed in achieving their goals.

Online social network involves large quantity of information transmission that creates increased impact on it. Social networks can be defined as the network of interactions or instance of relationship between multiple numbers of nodes connected to it. In this social network connection, the nodes are commonly referred as actors, where the edges indicate relationship among the actors and interaction between them.

Information retrieval from the available websites becomes a challenging issue, which involves complexities. Ranking WebPages is an important mission as it assists the user to look for highly ranked pages that are relevant to the query.

Information filtering is a popular research area which has received great attention from the Social Network. Since the last decade, there has been an explosive growth of information on the Social Network. More than one trillion web pages in Social network
are indexed by search engines and hence finding the desired information is not an easy task. The proposed Page Access Coefficient (PAC) method has the potential to solve this problem and influences for further development in Social Network. It aims to discover useful information from the Web hyperlink structure, page contents and data usage. The Page Access Coefficient is used for information filtering and also provides an intelligent access to retrieve information available in the Social Network.

Social Media is a popular platform to share user’s thoughts and likes or dislikes from different graphical regions, especially with the phenomenal growth of Social Media sites and Blogs which generate vast amounts of user contents. This data can be analyzed to retrieve useful and specific information, like grouping of people with similarity in likes or dislikes. Social Media Analysis Grouping Algorithm (SMAGA) is a novel data mining technique which is suggested for grouping like-minded people. This technique is used to retrieve valuable knowledge from huge amount of data and help in informed decisions.

The proposed methods improve the quality of Web search by filtering irrelevant information. The methods are designed, implemented and tested in various application domains like Text filtering, Image filtering and Electronic mail.

The PAC was implemented in C# language, under the platform Visual Studio 2010. The SMAGA was implemented in VB using Oracle and ODBC. Existing algorithms have also been studied and their pitfalls have been eliminated. The proposed methods are evaluated by the measures like precision and time taken for calculation.