ABSTRACT

Information Access is concerned with finding of documents that satisfy user’s information needs. It encompasses Information Retrieval as one of its tasks. Information Retrieval aims at finding more of relevant documents while ignoring the irrelevant ones from the document collection corresponding to a user query. Information access and retrieval being integral components of Information Architecture, it is necessary to improve the techniques related to the structure and organization of data as well as develop algorithms for effective retrieval to further strengthen the effectiveness of information architecture of any organization. The IR techniques for IAs need to be investigated, as the growing size of data and applications along with the rising information needs of the users of the system is further leading to complexities.

The information access and retrieval technologies for a typical IA face severe precision and recall problems when using conventional IR techniques. GA is a robust and powerful search mechanism characterized by higher probability of finding good solutions for large and complex problems and has proved to be effective in Information Retrieval. The GA has therefore been applied to IR problem in context of Information Architecture.

The thesis focuses on enhancing information access and retrieval techniques by developing a GA-based algorithm IIIARGA (Improved Information Access and Retrieval through Genetic Algorithm) to retrieve higher number of relevant documents at higher ranks. The goal is to employ evolutionary algorithm to optimise the relevance of the documents thereby improving the access and retrieval process in a typical Information Architecture and to check its applicability in a larger context. The main phases of IIIARGA are indexing phase, evolutionary phase and ranking phase.

The performance of IIIARGA is evaluated using the speed of convergence of GA unit, average execution time per query, recall, 11-point average precision, recall, mean average precision and precision at rank K. The customised fitness function is developed to yield higher performance in terms of precision and recall. The performance of IIIARGA is compared with that of classical Information Retrieval System to test the effectiveness of GA in retrieving better results. Additionally, the effectiveness of IIIARGA is tested beyond the scope of the typical IA selected for the study, on two other standard test collections, to test its practical feasibility outside the environment and consistency in its performance. In general, IIIARGA has proved to be an effective IR technique for Information Architecture that provides high quality results.