|  |
| --- |
| University of Baghdad  |
| Economic and administration  | College Name |
| Computer center | Department |
| Bushra Khireibut Jassim Al-Saidi | Full Name as written in Passport |
| **Drbushra144@yahoo.com** | e-mail |
|  **Professor**  |  **Assistant Professor** | √ **Lecturer** |  **Assistant Lecturer**  | Career  |
|  PhD√  |  Master  |  |
| A Proposed Genetic Algorithm for Clustering Web Search Engine Results | Thesis Title  |
| 2006 | Year |
| With the wide use of the Internet, and the exponential growth of the size of World Wide Web, information retrieval and resource discovery from the Web is becoming more challenging .As the size of the Web continues to grow, searching it for useful information has become increasingly difficult. Web users have been mainly relying on Web search engines to find information of interest on the Web. However, one key issue remain with Web search engines: the browsability of searching results. The long ranked list presentation of search results, which is widely adopted adds a layer of confusion to users, especially when the number of matches returned from search engines can easily exceed thousand level. This research, focuses on clustering Web search results as a solution to the browsability problem in order to help users find relevant Web information more easily and quickly. Web search results organized into topics and subtopics facilitates browsing the collection and locating results of interest.This research proposes a new genetic frequent term sets based clustering algorithm to build clusters for a collection of search results retrieved in response to a query. It also developed a user interface that enables the user cluster Web search results in two mode of clustering, overlapping and non-overlapping version. The clusters and its associated frequent terms present to the users with the ability to cluster any of the generated clusters. Genetic algorithms has been used in this research to generate frequent term sets that used to cluster the web search engine results. Some new genetic algorithms operations have been suggested that increase the performance of the algorithm in clustering web snippets. The contribution of this algorithm is the little preprocessing step needed, working independently from user and not need any prior estimation of parameter. Another contribution is the wide range of freedom the interface provided to the user in dealing with generated clusters. The research identify some requirements for clustering web search engine results and the algorithm is evaluated in term of this requirements and showed a promising results in the area of web search engine results mining . |  Abstract  |