IMPROVED CLUSTER BASED SEARCH ENGINE

GAURI S. BHAGAT[1], MRUNAL S. BEWOOR[2] AND DR. SUHAS H. PATIL[3]

[1] Research Scholar,   [2]Associate Professor and Guide ,   [3]Professor and Head
Department of Computer Engineering, Bharati Vidyapeeth Deemed University,
College of Engineering, Katraj, Pune, India **.**

Abstract

With an enormous growth of the Internet it has become very difficult for the users to find relevant documents. In response to the user's query, currently available search engines return a ranked list of documents along with their partial content (snippets). If the query is general, it is extremely difficult to identify the specific document which the user is interested in. The users are forced to sift through a long list of off-topic documents. Web page clustering identifies semantically meaningful groups of web pages and presents these to the user as clusters. Readable and unambiguous descriptions of the thematic groups are an overview of topics covered in the search results and help them to identify the specific group of documents they were looking for important factor of the overall quality of clustering. Pre-process input data applying an ontology-based heuristics for feature selection and feature aggregation. The inexperienced users, who may have difficulties in formulating a precise query, can be helped in identifying the actual information of interest. In this paper we briefly present our algorithm, which we believe is able to discover groups of related documents and describe the subject of these groups in a way meaningful to a human.