Intelligent Web Companion (IWC): Personalized Web Surfing tool

W.P.A. Fernando, K.A.C. Perera, A. Ranathunga, H.G.C. Wijesooriya, D.A.B Alahakoon and J.Amararachchi

Department of Information Technology, Faculty of Computing.
Sri Lanka Institute of Information Technology,
Malabe. Sri Lanka.

{pumudufernando, ashancperera, amila.sc2000, chandana.wijesuriya, daksitha} @gmail.com, jayantha.a@sliit.lk

***Keywords*— Customized web Browsing, Internet, text classification, World Wide Web**

INTRODUCTION

World Wide Web is an essential component for day to day computer users. One of the common problems that most web surfers face is searching or the same piece of information multiple times in the web. Web browsers support bookmarking important web sites for future use. This might depend on the machine and the user may need to find the same kind of information in another machine. Redundant search on the same piece of information wastes the computer surfer's time and resources.

To address the above mentioned issues the authors have developed an Intelligent Web companion (IWC), a web surfing tool to allow users to have more personalized web browsing experience independent from the location, computer and the web browser. IWC is a web based application which allows each user to create own profile and surf the web. IWC is capable of tracking the user's browsing pattern and based on tracked information IWC provides important sets of services to the web surfer.

A. Proposed System Architecture

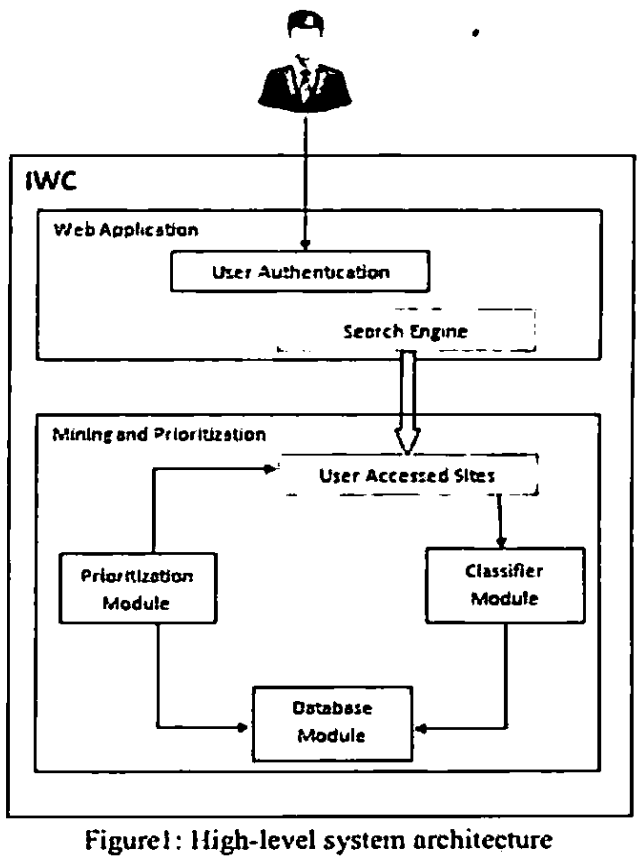

Figure1: High-level system architecture

Proposed Architecture consist of three main sub modules, where Web Application is responsible in interact with end users to display the results of the system.Mining and prioratization module is used for reading the web page content and categorize the page into one of the predefind classes. Work done by this module, can be divided into three main steps, Reading, Summarization and Categorization. Overall categorization process follows the Naïve Bayesian Classification algorithm. The output of the classifier module will be the classified category ID of the web page. To find out the probability of the popular web site for a specific user, IWC system will use an algorithm based on four main parameters, Time spent by the user on a particular web page, Total time a user spends on surfing web, number of accesses on a web site and a user's Last visited date of the web site.Database module will be used to save the related information for user visited web sites. In order to find out whether a web site has been updated or not IWC uses a hash function based calculation using the content of the web page.

B. Conclusion

During this work authors provided collection of systems created for performing similar tasks but none of those products had been able to help a user's day-today web surfing successfully because they do not have any procedure to observe user browsing pattern continuously, understand the user's interests, keep the user updated automatically with the news of his frequently accessed websites and suggesting best alternative URLs for the user's most preferred websites. IWC has overcome all these challenges by avoiding the constraints to detect a user's continuous life-time browsing pattern and introducing a systematic flow of steps to provide all the information that a user wants in day-today browsing within a time of just one click. This tool will help Internet users to make web browsing more pleasant and an interesting experience.

So far the system has been trained to categorize web pages into predefined sets of classes. This feature can be further enhanced to allow a user to come up with a set of categories according to their interest and improve the ability of Classifier module to categorize web pages more accurately. Classification module can be improved to categorize images and other multimedia content rather than textual content. So this tool will be able to fulfil desired objectives in more accurate manner.