Chapter II

Computational Intelligence Techniques Driven Intelligent Agents for Web Data Mining and Information Retrieval

Masoud Mohammadian, University of Canberra, Australia
Ric Jentzsch, University of Canberra, Australia

ABSTRACT

The World Wide Web has added an abundance of data and information to the complexity of information for disseminators and users alike. With this complexity has come the problem of finding useful and relevant information. There is a need for improved and intelligent search and retrieval engines. Current search engines are primarily passive tools. To improve the results returned by searches, intelligent agents and other technology have the potential, when used with existing search and retrieval engines, to provide a more comprehensive search with an improved performance. This research provides the building blocks for
integrating intelligent agents with current search engines. It shows how an intelligent system can be constructed to assist in better information filtering, gathering and retrieval. The research is unique in the way the intelligent agents are directed and in how computational intelligence techniques (such as evolutionary computing and fuzzy logic) and intelligent agents are combined to improve information filtering and retrieval. Fuzzy logic is used to access the performance of the system and provide evolutionary computing with the necessary information to carry out its search.

INTRODUCTION

The amount of information that is potentially available from the World Wide Web (WWW), including such areas as web pages, page links, accessible documents, and databases, continues to increase. Research has focused on investigating traditional business concerns that are now being applied to the WWW and the world of electronic business (e-business). Beyond the traditional concerns, research has moved to include those concerns that are particular to the WWW and its use. Two of the concerns are: (1) the ability to accurately extract and filter user (business and individuals) information requests from what is available; and (2) finding ways that businesses and individuals can more efficiently utilize their limited resources in this dynamic e-business world.

The first concern is, and continues to be, discussed by researchers and practitioners. Users are always looking for better and more efficient ways of finding and filtering information to satisfy their particular needs. Existing search and retrieval engines provide more capabilities today than ever before, but the information that is potentially available continues to grow exponentially. Web page designers have become familiar with ways to ensure that existing search engines find their material first, or at least in the top 10 to 20 hits. This information may or may not be what the users really want. Thus, the search engines, even though they have now become sophisticated, cannot and do not provide sufficient assistance to the users in locating and filtering out the relevant information that they need (see Jensen, 2002; Lawrence & Giles, 1999). The second area, efficient use of resources, especially labor, continues to be researched by both practitioners and researchers (Jentzsch & Gobbin, 2002).

Current statistics indicate that, by the end of 2002, there will be 320 million web users (http://www.why-not.com/company/stats.htm). The Web is said to contain more than 800 million pages. Statistics on how many databases and
The Context of IST for Solid Information Retrieval and Infrastructure Building: Study of Developing Country
www.igi-global.com/article/the-context-of-ist-for-solid-information-retrieval-and-infrastructure-building/193251?camid=4v1a

Data Mining Approach to Decision Support in Social Welfare
www.igi-global.com/chapter/data-mining-approach-to-decision-support-in-social-welfare/198625?camid=4v1a