Chapter X

AntWeb—Web Search Based on Ant Behavior: Approach and Implementation in Case of Interlegis

Li Weigang
University of Brasilia, Brazil

Wu Man Qi
Federal Court of Accounts, Brazil

ABSTRACT

This chapter presents a study of ant colony optimization (ACO) to Interlegis Web portal, Brazilian legislation Web site. The approach of AntWeb is inspired by ant colonies foraging behavior to adaptively mark the most significant link by means of the shortest route to arrive at the target pages. The system considers the users in the Web portal as artificial ants and the links among the pages of the Web as the researching network. To identify the group of the visitors, Web mining is applied to extract knowledge based on preprocessing Web log files. The chapter describes the theory, model, main utilities, and implementation of AntWeb prototype in Interlegis Web portal. The case study shows off-line Web mining, simulations without and with the use of AntWeb, and testing by modification of the parameters. The result demonstrates the sensibility and accessibility of AntWeb and the benefits for the Interlegis Web users.

INTRODUCTION

The World Wide Web today is considered the biggest information and knowledge source in every field. These Web pages are accessed daily by a great quantity of people with different interests. However, they are heterogeneously constructed both in available information and quality and majorities are presented in HTML language. This manner limits and makes it difficult to extract useful information and knowledge using simple search by Internet.

The process of surfing the Web is similar to ant colonies foraging (Beckers, Deneubourg &
Ants’ physical characteristics do not allow them to have a global vision of the environment, but their important and interesting foraging behavior help them to find shortest paths between food sources and their nest (Dorigo, Caro & Gambardella, 1999; Dorigo, Maniezzo & Coloni, 1996). With this insight, it is observed that the users visit the Web as a metaphor to ant colonies’ foraging process. Sometimes, they may lose their way in the immense cyberspace without knowing where the information sources are.

Ants are efficient at foraging and finding the shortest route. Also, any one of them deposits a substance called pheromone on the ground, while walking from food sources to the nest and vice versa, forming a pheromone trail. Ants can smell the pheromone left in pheromone trails and when choosing their way they tend to choose paths marked by strong pheromone concentrations (Dorigo et al., 1996).

Different from ants, visitors to the Web do not have any way to communicate among them. Each one obtains their own route to find the objective page without having the association and assistance from other users that may have previously passed through the same path. Suppose that one implements an extended Web server within the Web site, this server can record the visited path on that Web site and put information on the visited links as ants put the pheromone. In such a way, the visitors can count on teamwork to guide them establishing an indirect communication. Thus, using ant searching mechanisms, laterally, a single user can find the objective page easier with this communication and increase the possibility of surfing faster by using the best path.

The application of collective intelligence on Web search has been studied by Heylighen (1999). His work is considered an initial research on the topic. A system inspired by the ants’ foraging behavior to Web, AntWeb was proposed in the researches of Weigang, Dib, Teles, de Andrade, de Melo, and Cariolano (2002), Teles, Weigant, and Ralha (2003), and Weigang and Wu (2005). In their study, AntWeb was developed as an adaptive Web system working as a metaphor of ants’ foraging behavior in the following way: when a user visits a Web site, the system records some information of his route as the pheromone trail left by ants; other users with common objectives may be attracted by the pheromone trail using techniques from adaptive hypermedia technology and Web mining. AntWeb has been modified in two generations with the following three main utilities: the first one is to use it to evaluate the efficiency of the structure of Website (Weigang et al., 2002); the second one is to arrange a manner to assistant the new user to search his objective page with the help from the experience from the other users on the same Website (Teles et al., 2003); and the third one is to extend the Web site to have an adaptive capacity, which extracts information from the user’s sequential path to change link strengths or to create new links between the entrance point to the target page or to create a significant mark in the entrance page (Teles et al., 2003; Weigang & Wu, 2005). Other manners to use artificial life behavior in Web search have also been reported in the research of Revel (2003). Chen, Park, and Yu (1996) explored data mining capability which involves mining path traversal patterns in a distributed information providing environment like WWW. To identify the group of visitors and target pages of their visit, Web mining is introduced. There are a lot of researches in Web mining. We adapted the methods from the references of Gamma, Helm, Vlissides, and Johnson (1994), Srivastava, Cooley, Deshpande, and Tan (2000), Srikant and Yang (2001), Yao, Hamilton, and Wang (2002), Han and Chang (2002), Mobasher (2004), and so forth.

The chapter presents the second generation model, an extension version of AntWeb, with the technology of the Web mining and the implementation in Interlegis Web portal, the Brazilian legislation Web site (Weigang & Wu, 2005). The case studies show some examples such as: (1)
Related Content

A Cached-Based Stream-Relation Join Operator for Semi-Stream Data Processing

Efficient Compression and Storage of XML OLAP Cubes
[www.igi-global.com/article/efficient-compression-and-storage-of-xml-olap-cubes/129522?camid=4v1a](www.igi-global.com/article/efficient-compression-and-storage-of-xml-olap-cubes/129522?camid=4v1a)

SOM-Based Clustering of Textual Documents Using WordNet
[www.igi-global.com/chapter/som-based-clustering-textual-documents/21725?camid=4v1a](www.igi-global.com/chapter/som-based-clustering-textual-documents/21725?camid=4v1a)

Data Mining Using Fuzzy Decision Trees: An Exposition from a Study of Public Services Strategy in the USA
[www.igi-global.com/chapter/data-mining-using-fuzzy-decision/44282?camid=4v1a](www.igi-global.com/chapter/data-mining-using-fuzzy-decision/44282?camid=4v1a)