Chapter 40
Enhanced Artificial Social Cockroaches (EASC) for Modern Information Retrieval

Hadj Ahmed Bouarara
Tahar Moulay University of Saida Algeria, Algeria

Reda Mohamed Hamou
Tahar Moulay University of Saida Algeria, Algeria

Amine Abdelmalek
Tahar Moulay University of Saida Algeria, Algeria

ABSTRACT

This article deals on an improved version of the recently developed Artificial Social Cockroaches (ASC) algorithm based on several modifications. The EASC has as input a set of artificial cockroaches and N selected shelters. It is based on a random displacement step and a set of operators (selection cockroaches, shelter attraction, congener's attraction, shelter permutation). Each cockroach must be hidden in the shelter where it feels safer (evaluation function). In the recent years with the coming of the world wide web, the amount of unstructured documents available in the digital society increases and becomes easily accessible, all this has led that satisfy the needs of users in terms of relevant information has become a substantial problem in the scientific community. The second component of the authors' study is to apply the algorithm (EASC) as an information retrieval system using multilingual pre-processing and thesaurus to solve the problems of multilingual query and searching with synonymy. The relevant documents will be rendered as a list of ranked and classified documents from the most relevant to the least relevant. Lastly the authors apply the benchmark Medline and a series of valuation measures (precision, recall, f-measure, entropy, error, accuracy, specificity, TCR, ROC) for the experimentation, also they have compared their results with the outcomes of set of existed systems (social worker bees, taboo search, genetic algorithm, simulating annealing, naïve method). The third component of the authors' system is the visualization step that ensures the presentation of the result in the form of a cobweb with some realism to be understandable by users.

DOI: 10.4018/978-1-5225-5191-1.ch040

1. INTRODUCTION AND PROBLEMATIC

The current scientific world was considerably built up with the first appearance of novel concepts and paradigm such as bio-inspired techniques. The artificial social cockroaches (ASC) was proposed by Bouarara et al in (Bouarara[1] && al, 2015) is based on the phenomenon that the cockroaches of the same colony must be hiding under the same shelter with less luminosity and the displacement of each cockroach is guided by the displacement of its congeners. Similar to other optimization algorithms, the goal is to find the ideal solution of an optimization problem with constraints in the form bound min x $\in\Omega$ f (x), where f: R N $\rightarrow$ R is a nonlinear function, and Ω is the search space.

The ASC is based on the principal that each cockroach seeks the most attractive shelter. Subsequently, the algorithm will evolve from time t to time t + 1 and each cockroach will move from one position to another until reaching the shelter where it feel safer. The first part of our work is to develop a new version of ASC by improved its limits.

The appearance of the Internet and the incredibly rapid development of telecommunication technology have made the world a global village. Nowadays, the information has become a pillar of our civilization, no one can escape it and we find it everywhere. Depending on the statistic given in 2014 more than 90% of the information presented in the web was in unstructured format (textual documents). Nevertheless, find the desired document is not an obvious thing. For example, we are confronted with a set of documents and asking a human to obtain documents which hold the query "artificial intelligence". Human will read all the documents one by one and classify them into two categories (relevant documents and irrelevant) this is generally a real boring process. In the last decade several information retrieval systems have been produced to allow better access to data, it is at this stage where is positioned the context of the second portion of our work.

An information retrieval system (IRS) is a set of model and process, allowing the selection of information and permits to find a subset of pertinent documents from a collection of documents, to meet the needs of a user. The actual search engines based on classical methods are not satisfactory concerning the quality of results returned that suffers from several limitations and they are face of several troubles in terms of:

- **Choice of Parameters:** The majority of information retrieval systems are based on the parameters (distance measures and texts representation methods). A poor choice of those parameters may cause degradation in the research performance as the works presented in (Bouarara[1] & al, 2014).
- The quality of results is the major limit, where the classical methods based on a simple functioning do not perform well against this challenge. They return a lot of:
  a. **Relevant Misclassification:** The document really relevant and the system classified them as irrelevant.
  b. **Irrelevant Misclassification:** The documents really irrelevant and the system classified them as relevant.
- **Multilingual Research:** The conventional information retrieval system selects only the documents with language similar to the query. For E.g. we put the query "mining" in Google. Mining means "fouiller" in French. In this case Google will look only for English documents that contain the word mining and ignores the documents with other languages despite that mining and "fouiller" have the same meaning. To address this problem, we'll use the Google translation API and added a service that asks the user if he wants to make a multilingual search or not and what are