

Chapter 52

A Particle Swarm Optimization Algorithm for Web Information Retrieval: A Novel Approach

Tarek Alloui

Constantine 2 University, Algeria

Imane Boussebough

Constantine 2 University, Algeria

Allaoua Chaoui

Constantine 2 University, Algeria

ABSTRACT

The Web has become the largest source of information worldwide and the information, in its various forms, is growing exponentially. So obtaining relevant and up-to-date information has become hard and tedious. This situation led to the emergence of search engines which index today billions of pages. However, they are generic services and they try to aim the largest number of users without considering their information needs in the search process. Moreover, users use generally few words to formulate their queries giving incomplete specifications of their information needs. So dealing this problem within Web context using traditional approaches is vain. This paper presents a novel particle swarm optimization approach for Web information retrieval. It uses relevance feedback to reformulate user query and thus improve the number of relevant results. In the authors' experimental results, they obtained a significant improvement of relevant results using their proposed approach comparing to what is obtained using only the user query into a search engine.

DOI: 10.4018/978-1-5225-5191-1.ch052

INTRODUCTION

Information Retrieval (IR) is a computer science field that focuses on acquisition, organization, storage, search and selection of information. The user can access to information through an Information Retrieval System (IRS) by submitting his information needs in the form of a keywords' set well-known as query.

But, since the late 20th century, new challenges have been posed to IR. The wide dissemination of personal computers, the development of electronic media with more and more important capacities, the advent of Internet, which is considered as the largest source of information in the world and the pervasiveness of mobile networks allow users worldwide to access a vast amount of information. The expansion and the growing popularity of The Web, basically in the field of communications and information access, have changed our way of looking to information access.

Therefore, information access has become crucial, and tracking relevant information, suitable to user needs, becomes more and more important and difficult in an environment such as the Web. This need has generated many sources of information. Moreover, the amount of information available at any time on the Web is huge and increases exponentially. However, accessing such volume of information is vain without search engines. So, the most famous search engines index billions of pages. But, they are generic information services, they have been designed to aim the largest number of users and thus, they don't take into account the user and its context in the search process.

Furthermore, this problem is accentuated by the fact that users formulate their needs using a short and ambiguous query, giving an incomplete and imprecise specification of their real information needs. Indeed, if two users with different needs submit the same query to a search engine, this will return the same results.

So, user playing an essential part in the IR process, it is necessary to consider him in the search process by personalizing the IR system according to his interests and his information needs in order to increase the relevance of returned results. But, representing user needs by a model or a structure isn't easy and few standards are established. However, it influences significantly the performance of IR systems.

In addition, the complexity of the IR problem and the constant evolution of the Web cannot be tackled with conventional approaches. So, we need powerful tools like artificial intelligence techniques to cope with these problems. In this work, we have chosen meta-heuristics and more precisely Particle Swarm Optimization (PSO) in the design and implementation of our approach. Indeed, meta-heuristics has proved their effectiveness, strength and adaptability to IR problems through several research studies (Al Mashagba et al., 2011; Bindal and Sanyal, 2012; Drias, 2011).

In this study, the information retrieval problem to solve is to find the suitable query that best matches user information needs and improves the number of relevant documents returned by the search engine. This issue is part of query reformulation which is one of the main IR research areas. Over the past four decades, it has been the subject of a lot of publications in the IR community, and until today, there is many issues remaining that must be solved before it becomes an operational standard used in information retrieval systems.

So we have adapted a PSO algorithm to the query reformulation problem in order to optimize user query and thus retrieve more relevant results to user information needs.

This paper is organized as follows. First, we briefly present query reformulation techniques and related work. Then we review the PSO algorithm principle. After that, we explain the proposed approach and the adaptation of PSO algorithm. And then, we present the experimental results obtained after testing the prototype of our approach on the Web. Finally, we end this paper with conclusions and future work.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/a-particle-swarm-optimization-algorithm-for-web-information-retrieval/198595

Related Content

Ensemble of Support Vector Machine and Ontological Structures to Generate Abstractive Text Summarization

Amita Arora (2022). *International Journal of Information Retrieval Research* (pp. 1-24).

www.irma-international.org/article/ensemble-of-support-vector-machine-and-ontological-structures-to-generate-abstractive-text-summarization/300294

Evolutionary Optimization for Prioritized Materialized View Selection: An Exploratory Analysis

Heena Madaanand Anjana Gosain (2022). *International Journal of Information Retrieval Research* (pp. 1-18).

www.irma-international.org/article/evolutionary-optimization-for-prioritized-materialized-view-selection/300295

Multi-View Meets Average Linkage: Exploring the Role of Metadata in Document Clustering

Divya Teja Ravooriand Zhengxin Chen (2015). *International Journal of Information Retrieval Research* (pp. 26-42).

www.irma-international.org/article/multi-view-meets-average-linkage/130006

An Online Measure of Discernment

Hazel C. V. Traüffer, Corné L. Bekker, Mihai C. Bocarneaand Bruce E. Winston (2013). *Online Instruments, Data Collection, and Electronic Measurements: Organizational Advancements* (pp. 254-270).

www.irma-international.org/chapter/online-measure-discernment/69745

Innovative Shot Boundary Detection for Video Indexing

Shu-Ching Chen, Mei-Ling Shyuand Chengcui Zhang (2005). *Video Data Management and Information Retrieval* (pp. 217-236).

www.irma-international.org/chapter/innovative-shot-boundary-detection-video/30767