



A review on the application of evolutionary computation to information retrieval ☆

O. Cordon^{a,*}, E. Herrera-Viedma^a, C. López-Pujalte^b,
M. Luque^a, C. Zarco^c

^a *Department of Computer Science and Artificial Intelligence, University of Granada, E.T.S.I. Informatica, C/Daniel Saucedo Aranda, s/n, Granada 18071, Spain*

^b *Library Science Studies Faculty, University of Extremadura, Badajoz 06071, Spain*

^c *PULEVA Food S.A. Camino de Purchil, 66, Granada 18004, Spain*

Received 1 January 2003; accepted 1 July 2003

Abstract

In this contribution, different proposals found in the specialized literature for the application of evolutionary computation to the field of information retrieval will be reviewed. To do so, different kinds of IR problems that have been solved by evolutionary algorithms are analyzed. Some of the specific existing approaches will be specifically described for some of these problems and the obtained results will be critically evaluated in order to give a clear view of the topic to the reader.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Information retrieval; Evolutionary algorithms; Automatic indexing; Document clustering; Query definition; User profiles; Internet search agents; Image retrieval; Similarity functions; Web pages

☆ This research has been supported by CICYT under project TIC2002-03276 and by the University of Granada under project “Mejora de Metaheurísticas mediante Hibridación y sus Aplicaciones”.

* Corresponding author. Tel.: +34-958-246143; fax: +34-958-243317.
E-mail address: ocordon@decsai.ugr.es (O. Cordon).

1. Introduction

We are actually living in the information age. This leads to the fact that any media organization has a computer system endowed with data bases, which can structure different information kinds in a correct way, and with hardware which allows it to efficiently store and access to this information. Unfortunately, the large size of these data bases has made the required effort to retrieve useful information increase significantly in the last few years.

Information retrieval (IR) tries to make a suitable use of these data bases, allowing the users to access to the information which is really relevant in an appropriate time interval [47]. Unfortunately, commercial IRs, usually based on the Boolean IR model [53], have provided unsatisfactory results. Vector space, probabilistic and fuzzy models, which have been developed to extend the Boolean model [2], as well as the application of knowledge-based techniques, have solved some of these problems, but there are still some lacks [11]. In the last few years, an increasing interest on the application of artificial intelligence (AI)-based techniques to IR has been shown with the aim of solving some of those lacks.

One of the AI (or, more specifically, computational intelligence) areas with a considerable growth in the last decades is *evolutionary computation (EC)* [1], based on the use of models of evolutionary process for the design and implementation of computer-based problem solving systems. The different models which have been proposed within this philosophy are named in a generic way as *evolutionary algorithms (EAs)* [1].

In this paper, we review the application of EC to IR, analyzing the different kinds of IR problems that have been solved by EAs, describing some of the specific approaches proposed and critically evaluating the outcomes obtained.

To do so, the paper is structured as follows. In Section 2, some preliminaries are introduced by reviewing the basis of IR and EAs. Then, the different applications of EAs to IR are classified into different groups in Section 3. Sections 4–11 respectively show a review of the different proposals made in each of these groups, namely, automatic document indexing, document and term clustering, query definition, similarity function learning, image retrieval, user profile design, web page classification and Internet search agent design. Finally, Section 12 summarizes several concluding remarks.

2. Preliminaries

2.1. *Information retrieval*

IR may be defined, in general, as the problem of the selection of documentary information from storage in response to search questions provided by

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات