Algorithms and Framework for Comparison of Bee-Intelligence Based Peer-to-Peer Lookup

Vesna Šešum-Čavić and Eva Kühn

Technical University Vienna, Institute of Computer Languages,
Argentinierstrasse 8, 1040 Wien, Austria
{vesna,eva}@complang.tuwien.ac.at

Abstract. Peer-to-peer has proven to be a scalable technology for retrieval of information that is widely spread among distributed sites and that is subject to dynamic changes. However, selection of a right search algorithm depends on many factors related to actual data content and application problem at hand. A comparison of different algorithms is difficult, especially if many different approaches (intelligent or unintelligent ones) shall be evaluated fairly and possibly also in combinations. In this paper, we describe a generic architectural pattern that serves as an overlay network based on autonomous agents and decentralized control. It supports plugging of different algorithms for searching and retrieving data, and thus eases comparison of algorithms in various topology configurations. A further novelty is to use bee intelligence for the lookup problem, spot optimal parameters' settings, and evaluate the bee algorithm by using the architectural pattern to benchmark it with other algorithms.

Keywords: information retrieval, lookup mechanism, bee intelligence, distributed coordination patterns.

1 Introduction

Bio-inspired algorithms play an important role in the design of self-organizing software for distributed systems. Such software is typically characterized by a huge problem size concerning number of computers, clients, requests and size of queries, autonomy and heterogeneity of participating organizations, and dynamic changes of the environment. In such a setting, the common approach of one central coordinator often reaches its technical and conceptual limits. On the technical side, it represents a single point of failure with the risk of becoming a performance and availability bottleneck. On the conceptual site, it is hard to design as it must be aware of the entire business logic, possessing the complete picture of all participants, and being able to cope with all possible dynamics in the environment. Therefore, to cope with the described dynamics and vast number of unpredictable dependencies on participating components, other approaches are demanded like autonomously acting components who are inspired by nature and whose behaviors implement, e.g., bio-inspired algorithms. These components act in a dynamic, ad-hoc way and adapt quickly and self-subsistent to both changing requirements and dynamically evolving system states caused through the interplay and contribution of the many components towards a global goal. On type of

Y. Tan, Y. Shi, and H. Mo (Eds.): ICSI 2013, Part I, LNCS 7928, pp. 404–413, 2013.

bio-inspired algorithms are bee algorithms that have been already applied to several computer science problems, e.g., [13], [16], [18], [19], [21] and [23].

The novelty of this paper is to evaluate the usefulness of bee-intelligence for the problem of information placement and retrieval in distributed systems and to prove its correctness. The use case employed for evaluation is to search information in the Internet. It assumes a highly dynamic setting where new URLs are being added every day. Also, there is a certain amount of invalid links that correspond to old, discarded pages [10], [12], [14], [25].

The *contribution* of this paper is: 1) The adaptation and implementation of an intelligent lookup mechanism based on bee intelligence that is able to cope with complex queries even if the given query is incomplete. 2) The definition and implementation of a generic architectural pattern for searching and retrieving data. This test-bed allows the simple exchange of different algorithms (intelligent and non-intelligent ones) simply through plugging.

The paper is organized as follows: Section 2 gives an overview of already existing approaches of solving the considered problem - location and retrieval of information in the Internet - taking into account P2P computing paradigm with lookup realized by using bee intelligence. Section 3 describes the proposed architectural pattern. Section 4 explains the bee algorithm for searching and retrieving of information. Section 5 presents the best parameters' settings and benchmarks. Section 6 summarizes the results.

2 State-of-the-Art

This section summarizes related work in the context of location and retrieval of information in the Internet that supports "intelligent" lookup mechanisms and that integrates these with the peer-to-peer (P2P) computing paradigm [2]. Table 1 gives an overview of systems that support adaptation and/or provision of a generic software framework pattern. In [4], both adaptation and a framework are offered. Namely, it is proposed a framework that supports an approach for building P2P applications based on the MAS paradigm. It inherits the free search capability of Gnutella, without relying on inefficient broadcasting techniques. However, these systems neither support complex queries (by means of incomplete information[1]), nor flexible plug-ability of different search algorithms, nor the application of different algorithms at the same time. Table 2 depicts P2P systems that support intelligent search, according to their distribution structure. Although unstructured P2P network does not scale well, it supports dynamics very well, and therefore it fits better to our problem.

[1] *Example* (the scenario of a crime investigating process): There is some information about the person that the police search for, but unfortunately the complete description and/or the material evidence are not available. According to the police simplified pattern of information would look like (*first_name, last_name, birth_date, ID_number, hair_color, height*). In a real case, the complete information might be missing (e.g., the person under investigation was seen by a witness at a crime scene only shortly) and only 3 entities of data are known (*John, …, …, …, brown, 172*).