Abstract. Peer-to-Peer (P2P) is an attractive paradigm for building distributed network applications. A particularly intriguing class of distributed applications consists in Information Retrieval (IR) systems. The issue of Peer-to-Peer Information Retrieval (P2PIR) is being tackled by researchers attempting to provide valuable insights and to propose solutions to use it successfully. Nearly, all published studies have been evaluated by simulation means, using well-known document collections (usually acquired from TREC). This practice leads to two problems: First, there is little justification in favour of the document distributions used by relevant studies and second, since different studies use different experimental benchmarks, there is no common ground for comparing the solutions proposed. In this paper, we propose Peer-to-Peer Information Retrieval Benchmarking (P2PIRB), a benchmarking framework for P2PIR. P2PIRB allows to distribute documents and queries according to various ways. This work marks the start of an effort to provide more realistic evaluation environments for P2PIR systems as well as to create a common ground to compare the current and future architectures.

Keywords: P2P systems, Information Retrieval, Benchmarking.

1 Introduction

Peer-to-Peer (P2P) networks are a fast developing branch of Computer Science for which many researchers are developing new algorithms. In P2P system, each peer has a symmetric role of both a client and a server and where networks are built on software. Therefore, P2P networking has become a fast-developing research field, since it can, potentially, provide cost-effective, efficient and robust solutions. In such a networking, like any distributed system, location and retrieval of relevant information and resources are of paramount importance. In P2P network, each node possesses a files collection, and peers want to perform content-based searches over the documents collection of other nodes.

On the IR side, in a P2P network, the distribution of documents is, to a significant scale, a result of the previous location and retrieval. However, this also depends on the application specification and/or on other non-functional requirements that may be imposed (such as copyright considerations, etc.). Defining
and simulating user behaviour, especially in a very large distributed system, is a complex task. Indeed, most published P2PIR solutions have dealt with this problem indirectly. Instead of simulating user behaviour, people have attempted to reflect it in benchmarks, they have used for their evaluation. The problem with such approaches is a twofold. Firstly, there are cases where the documents distribution does not successfully reflect the application scenario and therefore such evaluation results are hardly conclusive. Secondly, each individual considers a different benchmark for experimental evaluation, the mutual comparison and the quantification of performance improvements become an impossible task. Addressing these issues, we propose a number of realistic benchmarks, suitable for the evaluation of P2PIR systems.

Emphasising the fact that there may be many, diverse potential P2PIR applications, we identify a number of possible scenarios and we propose methodologies that can be used for the creation of realistic information-sharing benchmarks.

The rest of this paper is organized as follows: In Section 2, we define what we mean by benchmark. We review, in Section 3, related work about benchmarking in P2P retrieval. Section 4 presents a number of P2PIR scenarios, their properties as well as a number of appropriate benchmarks that could address them. In Section 5, we describe our benchmarking framework. In Section 6, we present experimental study. Finally, in Section 7, we present our conclusion regarding the current work and how this may relate to future P2PIR systems.

## 2 Benchmark Definition

A benchmark is a standard by which something can be measured or judged. Dekhtyar, in [1], define IR benchmark by the following formula:

\[
\text{Benchmark} = \text{DataSet} + \text{Tasks} + \text{Answers} + \text{Evaluation measure} + \text{Data Formats}
\]

Indeed, a benchmark must provide the documents and the queries to be raised on these documents. The answers to the queries are often data provided by experts, together with the relevance judgements. Evaluation measures are the tools which the benchmark uses in order to test the relevance of the IR algorithms. Data Formats, relates to the existence of benchmark under various formats of possible data.

The benchmark formula proposed in [1] relates to a centralized benchmark. In a distributed context, new information must be defined; how to distribute the data on the various nodes of a network and which replication law to apply? In addition, we define the elements which a distributed benchmark must provide:

- **Documents collection.**
- **Queries collection.**
- **A definition of a distribution method of documents and queries among peers.**
- **A definition of a replication method of documents and queries among peers.**
- **Evaluation measures.**

In the following Section, we review various work on the benchmarking in a centralized and distributed context.