Computers and Electrical Engineering 000 (2016) 1–11

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

An intelligent web search framework for performing efficient retrieval of data

B. Bazeer Ahamed [a,*], T. Ramkumar [b]

[a] Faculty of Computer Science & Engineering, Sathyabama University, Chennai, India
[b] School of Information Technology & Engineering, VIT University, Vellore, India

ARTICLE INFO

Article history:
Received 7 March 2016
Revised 27 September 2016
Accepted 28 September 2016
Available online xxx

Key words:
Web content
Search engines
Web data
Multiple data source
Information retrieval

ABSTRACT

There are numerous search engines available in today's world to search and retrieve the required information. However retrieval of meaningful and appropriate formation as per the user requirement is always a challenging task. The foremost intention of any search engine is to provide the information with in a quick span of time. Since the nature of data available in World-Wide-Web shows heterogeneity in common and the sources of data are also distinct with each other, issues pertaining to schema structure and data representational are also there. In such circumstances, to eliminate inconsistencies and for enabling seamless integration of multiple data sources while retrieving web data, an efficient web search mechanism that fulfils the customer requirement is always needed. To enable the integration of multiple data sources while performing efficient retrieval of web data, an intelligent web search framework has been proposed in this paper.

1. Introduction

World-Wide-Web has been emerged as an internet-repository which collects the information from multiple heterogeneous resources and offer different collection of services to the user [1]. The numbers of users have been increased day-by-day and the information found in the web has also been increased enormously which results number web pages that are available publicly to the user [2]. The structure, representation and novelty of the data found in the web also offers more challenging issues during the process of retrieval since they have been retrieved from multiple data sources. From the perspective of World-Wide-Web, Multiple Data Sources (MDS) meant data repository which contains the required data in heterogenic format and provides information to various platforms such as web site, document achieve and others on the basis of the particular topic/keyword [3]. Though data has been retrieved from multiple data sources, the process of hiding complexity, performing optimization have been carried out by a typical search engine optimization [4] through the various steps as shown in Fig. 1. It includes development like analysis in the initial stages [5,6] followed by the keyword based search and participatory analysis through submission of directory, blogs, press release creation, and social block marking. The popularity with respect to the document is also being considered during the process of optimization.

The rest of the paper has been organized as follows; In Section 2, certain generic issues with respect to information retrieval has been discussed. The proposed intelligent framework along with the various modules is introduced in Section 3. The underlying mathematical model of our proposed framework for arriving search decision is presented in

* Corresponding author.
E-mail addresses: bazeerahamed@gmail.com, bazeermiet@gmail.com (B.B. Ahamed).

http://dx.doi.org/10.1016/j.compeleceng.2016.09.033
0045-7906/

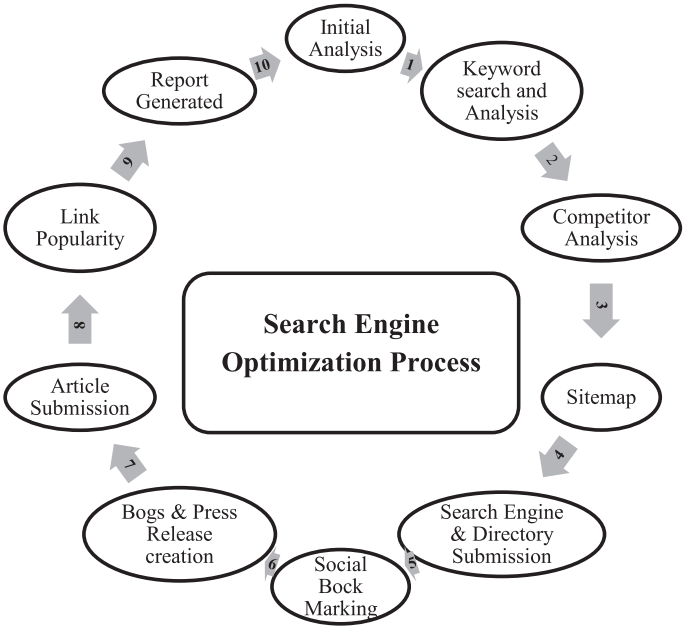

Fig. 1. Search engine optimization process.

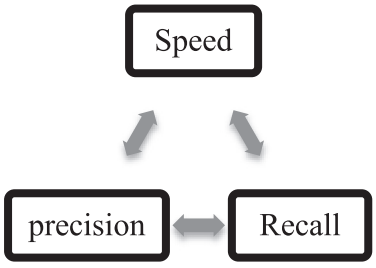

Fig. 2. Three way trade-off in search engine performance.

Section 4. The experimental investigation made for the proposed broker with respect to various accuracy measures are justified in the Section 5 along with obtained result. Section 6 provides conclusion note of our framework and also leads further scope for future work.

2. Generic issues with respect to information retrieval of web data

Since the proposed work advocates an intelligent web search framework, this section accounts the various issues with relevant to information retrieval of web data. Numerous procedures and measures [7–9] have been proposed for quantifying the accuracy of information retrieval. For an elementary model that is derived from the traditional retrieval systems distinguishes a three trade-off technique as shown in Fig. 2. The three way trade-off consists of speed, recall and precision.

The communicating response time is the most important issue that appears at the first in the list of significant issue that are present in the Web users [10]. The trade-off becomes more and more problematic as there is a tremendous increase in the number of users and their documents day by day [11]. In the background of information retrieval, exactness or the precision which is well-defined as the ratio of applicable documents to that of the total number of documents that are retrieved. If the precision calculated is 5, then the number of relevant documents is 25 and the total quantity of the documents that are retrieved is 5.

Recall is well-defined as that of the amount of the retrieving of the relevant documents. Similarly, if the recall is 5, then the amount of documents that are relevant is 25 and the entire amount of documents that are relevant is 5.The exactness can be planned by means of Eq. (1) and the recall can be planned by means of Eq. (2). These are the general equations given.

$$\text{precision} = \frac{\text{Number of revelant documents retrieved}}{\text{Total Number of retrieved documents}} \tag{1}$$

Please cite this article as: B.B. Ahamed, T. Ramkumar, An intelligent web search framework for performing efficient retrieval of data, Computers and Electrical Engineering (2016), http://dx.doi.org/10.1016/j.compeleceng.2016.09.033