Information Retrieval System: 
An Overview, Issues, and Challenges

Ram Kumar, Computer Science and Engineering Discipline, DPT, IIT Roorkee, Roorkee, India
S. C. Sharma, Computer Science and Engineering Discipline, DPT, IIT Roorkee, Roorkee, India

ABSTRACT

Information Retrieval Systems (IRS) have dramatically changed the ways how people acquire information for their need. Information Retrieval (IR) enables user to find relevant document from collection of countless resources. This article presents an overview of IRS. Objectives of this article is to answer all the basic and specific questions related to IRS. In contrast to other review papers, the authors provide a complete understanding of IR in single paper. Starting from definition and importance it covers retrieval process, performance issues, and comparison among various approaches. This article also includes description of different models along with analysis of their merits and demerits. This article proposes a list of challenges, still unanswered by existing systems. Before offering a conclusion, the major applications of IR are also listed.

KEYWORDS

Indexing, Information Retrieval, Query, Ranking, Relevance, Search Engine

INTRODUCTION

In the age of information revolution, electronic documents have become main source of information. Millions of Web pages are being created daily. With increasing information availability, the need for its retrieval has become a challenge. To solve this challenge Information Retrieval Systems (IRS) was introduced (Manning, 2008). Research on information retrieval was started with advent of computers, but profound evolution in this field occurred after the commencement of Web. Since last two decade the information collections has grown exponentially because computing capacity made it easier to produces large amounts of data (Sanderson, 2012). This large scale of search space made it almost impossible to retrieve desired information without IRS. Today computer user can’t imagine to satisfy his day to day information need, unless using any information retrieval tool. The most common information retrieval tool is Web search engine that is used to find information resources of user interest. Search engine indexes useful documents of the collection and searches those indexes for relevant resources (Croft, 2010).

Information Retrieval(IR) is defined as an act of finding materials, be it text, audio or video, within large collections to satisfy an information need (Baeza-Yates, 1999). In this paper our focus will be on text information retrieval. Text based IR assume that information resources are collection of text documents only. In text based IRS user’s information need defined by text query has to be satisfied. An IRS accomplishes its job by retrieving relevant document from document collection. This collection can be a readymade database like library data or may be collection of resources.
collected by independent modules like Web crawler. After collecting a set of information, IRS index that information by indexer module. During indexing it brake large documents into meaningful terms and index them (Buttcher, 2010). A user retrieves information on the basis of matching between query keywords and index terms.

This paper presents a critical analysis of information retrieval systems. The objective of this paper is to answer basic and specific questions related to IRS. The remaining paper is divided into seven sections. Second section provide architecture of IR with brief description of each components. Next section classifies different IR models by devising the taxonomy along with critical discussion. In fourth section the authors address the core issues related to IRS. Section fifth discusses main challenges that need to be answered by every IRS. Section sixth lists some of the major applications of IR. Finally, last section offers conclusions.

INFORMATION RETRIEVAL SYSTEM ARCHITECTURE

Primary objective of IR is “finding relevant information or a document that fulfil user needs”. To achieve this, IR system employ four components named indexing, query processing, searching and ranking. The architecture of IRS consisting these modules are described in Figure 1.

Indexing

This is an essential part of IR process in which documents are represented in restate content form. It constructs an index of terms which points to stored documents. In order to find terms, it parses the text of documents, and stores text units known as terms to enable accurate and fast retrieval. The purpose of building an index is to increase retrieval speed. It also improves storage space performance

Figure 1. Architecture of Information Retrieval

![Architecture of Information Retrieval](image-url)
Related Content

Evaluating Mobile Application Development Firms: Comparing Views of Advertisers and Advertising Agencies
Pi-Fang Hsu, Tien-Chun Lu and Chia-Wen Tsai (2014). International Journal of E-Adoption (pp. 53-66).
www.igi-global.com/article/evaluating-mobile-application-development-firms/113763?camid=4v1a

Digital Economy in Egypt: The Path to Achieve It
www.igi-global.com/article/digital-economy-in-egypt/223433?camid=4v1a
Enhancing Trust in E-Commerce in Developing IT Environments: A Feedback-Based Perspective
www.igi-global.com/chapter/enhancing-trust-commerce-developing-environments/65933?camid=4v1a

A Household Internet Adoption Model Based on Integration of Technology Acceptance Model, Theory of Planned Behavior, and Uses and Gratifications Theory: An Empirical Study on Iranian Households
www.igi-global.com/article/household-internet-adoption-model-based/64342?camid=4v1a