Chapter 6
Search Engine:
A Backbone for Information Extraction in ICT Scenario

Dilip Kumar Sharma
Shobhit University, India

A. K. Sharma
YMCA University of Science and Technology, India

ABSTRACT

ICT plays a vital role in human development through information extraction and includes computer networks and telecommunication networks. One of the important modules of ICT is computer networks, which are the backbone of the World Wide Web (WWW). Search engines are computer programs that browse and extract information from the WWW in a systematic and automatic manner. This paper examines the three main components of search engines: Extractor, a web crawler which starts with a URL; Analyzer, an indexer that processes words on the web page and stores the resulting index in a database; and Interface Generator, a query handler that understands the need and preferences of the user. This paper concentrates on the information available on the surface web through general web pages and the hidden information behind the query interface, called deep web. This paper emphasizes the Extraction of relevant information to generate the preferred content for the user as the first result of his or her search query. This paper discusses the aspect of deep web with analysis of a few existing deep web search engines.

INTRODUCTION

Information and communication technology have tremendous potential for social impact, human development and improving the lives of people they serve Through ICT peoples are able to communicate in better way and can access relevant information. It also helps in developing collaborative and research skills. People can gain confidence and avail opportunities on their potential. Information and communication technology provides appropriate hardware, software

and networking services to the search engine. To find out relevant pages instantaneously from billions of web pages available on the internet is a complex task. So, information extraction in web scenario is must to provide the relevant search to the user at the very first instant. An effective search engine is the necessity of today's information era. Search engine is a software program that searches for web sites that exist on the World Wide Web. Search engines search through its personal databases of information in order to provide the relevant information. A web crawler is an automated program that starts with a set of URLs called seeds and stores all the URL links associated with downloaded web page in a table called crawl frontier. The extractor sends all these information attached to the textual raw data to the analyzer. The analyzer then takes the entire HTML code of the downloaded web page and analyzes the code, keeping the relevant data and rejecting the rest. Some composing techniques are applied to link containing the similar types of information from the database to generate the relevant query results. Information and communication technology can be related to information extraction in web context or in search engine in a variety of ways (Anderson & Weert, 2002; Kundu & Sarangi, 2004). Traditional web crawling techniques

have been used to search the contents of the web that is reachable through the hyperlinks but they ignore the deep web contents which are hidden because there is no link is available for referring these deep web contents. The web contents which are accessible through hyperlinks are termed as surface web while the hidden contents hidden behind the html forms are termed as deep web. Deep web sources store their contents in searchable databases that produce results dynamically only in response to a direct request (Bergman, 2001) (Sharma & Sharma, 2011). Figure 1 shows the benefits of information extraction using ICT in human development in context of search engine.

ANALYSIS OF APPLICATION AREA OF ICT

Some of the area in which ICT plays a significant role in their development is analyzed below.

ICT in Education

In 1999 an analysis was done to find out the use of computer in schools. In that analysis it was found that a large number of students were sound enough to use the computers without taking help

Figure 1. Benefits of information extraction using ICT in human development in context of search engine

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/search-engine-backbone-information-extraction/68539

Related Content

Courseware and its Possible Evolution through the Use of Agent Technology
Paul Darbyshireand Glenn Lowry (2001). *Human Computer Interaction: Issues and Challenges (pp. 85-103).*
www.irma-international.org/chapter/courseware-its-possible-evolution-through/22426

Integration of Sustainability and Management Control Systems: A Challenge for Family SMEs
Paola Vola (2019). *Human Performance Technology: Concepts, Methodologies, Tools, and Applications (pp. 1780-1791).*
www.irma-international.org/chapter/integration-of-sustainability-and-management-control-systems/226644

Super Soldiers (Part 1): What Is Military Human Enhancement?
Patrick Lin, Max Mehlman, Keith Abneyand Jai Galliott (2019). *Human Performance Technology: Concepts, Methodologies, Tools, and Applications (pp. 61-81).*
www.irma-international.org/chapter/super-soldiers-part-1/226556

Systems Thinking: Dealing with Time Scales
Ian Roderick (2014). *International Journal of Systems and Society (pp. 53-54).*
www.irma-international.org/article/systems-thinking/94650

Collaborating Partnerships: A Project-Based Legitimizing Strategy amongst East African E-Learning Providers
Titus Tossyand Irwin T.J. Brown (2017). *Information Technology Integration for Socio-Economic Development (pp. 32-51).*
www.irma-international.org/chapter/collaborating-partnerships/160569