Chapter 4
Desktop Search Engines:
A Review From User Perspectives

Shaukat Ali
University of Peshawar, Pakistan

Shah Khusro
University of Peshawar, Pakistan

Mumtaz Khan
University of Peshawar, Pakistan

ABSTRACT

The advancements in desktop computing technologies and widespread use of digital devices have enabled users to generate and store tons of contents on their computers, which creates an information overload problem. Therefore, users need applications to search and retrieve required stored data instantly and accurately. This attracted the software industry to develop desktop search engines using ideas of web search engines. However, selecting an effective desktop search engine is difficult for users. In this chapter, the authors have analyzed and compared available desktop search engines using a set of parameters including users' privacy and security. A generalized architecture for desktop search engines is presented for improving understanding and unification of development efforts. The new emerging trends, supporting users' privacy and security, and several open issues and challenges in the desktop search engines domain are also highlighted. The authors hope that this chapter will be helpful for researchers to find new research problems and users to select appropriate desktop search engines.

INTRODUCTION

The history of information retrieval (IR) has footprints long before the proliferation of internet. IR systems were needed when collection of data reached to a size beyond the strength of traditional cataloguing techniques. An IR system can search a vast collection of semi-structured and unstructured data (e.g., documents, web pages, images, videos, audios, etc.) to answer users' queries. IR systems were witnessed

in commercial and intelligence applications since 1960's. The integration of search into the fabric of desktop and mobile operating systems can be attributed to the success and high-level adoptability of the Web search engines (Sanderson & Croft, 2012).

Web technology has seen an incredible growth in the history of mankind since its inception in the mid 1990's. By September 2022, Google had indexed about 50 billion web pages. Success of the web gave rise to the problem of information overload, where finding required information is relatively a time consuming and cumbersome task. The increasing growth of information on the Web highlighted the importance of knowledge management. Several technologies have been investigated for this purpose including web search engines, and Semantic Web etc.

The evolution of computer hard drive technology has shown tremendous miniaturization and increase in the average size of storage capacity with the expense of decrease in the prices (Markscheffel, Büttner, & Fischer, 2011). The increasing number of varied application programs and the availability of new information sources has emerged new data formats, which significantly increased users' capability of storing different types of information and files on their PCs with an increase in their volume day by day. Therefore, PCs are suffering from the same information overload problem which was once experienced on the Web. Ultimately, finding/searching a required file on a PC is as hard as finding a needle in a haystack. To reduce search efforts, a painstaking user might organize his information in an advanced directory structure (Markscheffel, et al., 2011). However, studies have shown that traditional systems using desktop metaphor and folder hierarchy suffer from their limitations and might not provide adequate information management features specifically information retrieval (Jones & Dumais, 1986).

Most of the computer users normally act as "virtual pack rats" having the tendencies of hoarding their files in whatever drives space they may find (Farina, 2005). Thus, make it increasingly difficult for users to find quickly and accurately the exact locations of files on their PCs. To elevate the problem, ideas from the Web are leveraged into PCs, resulting into the development of desktop search engines (DSEs). DSEs are a new generation of desktop applications which enables users for quickly searching and retrieving required files simultaneously in multiple data sources (i.e., PCs, computers across the enterprise network, etc.) using a single query in the same way as web search engines are used for locating information on the Web (Noda & Helwig, 2005). A DSE works similarly to web search engines involving crawling, indexing, ranking, and searching activities (Narasimhan & Lowe, 2010).

Desktop search is an essential part of an operating system, but the search tools incorporated in operating systems are having certain limitations. To overcome these limitations, organizations especially the web giants such as Google, Yahoo, AOL, Microsoft, etc. have introduced their own DSEs during the last few years (Khusro, et al., 2017). However, the tools vary in cost, performance, user friendliness, and users' privacy and security (Khusro, et al., 2017). The web giants offered free of cost DSEs to convince users to use their web portals regularly. However, DSEs require new approaches because desktop files are normally different in structure than those on the Web. An average user selects a DSE that offers most attractive functionalities, which is easy-to-use, covers a variety of file types and support users' privacy and security (Cole, 2005; Khusro, et al., 2017; Narasimhan & Lowe, 2010). Similarly, due to different nature and complexities in desktop structure, desktop searching poses a number of challenges including (Cole, 2005): (1) limited capabilities of operating systems' built-in desktop search features due to keyword searching over a few file types, (2) structured files are easier to search as compared to unstructured files, (3) files of various desktop applications mostly differ structurally making search difficult and challenging, (4) IR approaches used in the Web search engines need to be incorporated for dealing with unstructured files, (5) metadata associated with files by authors requires to be extracted

Related Content

DiALog: A Distributed Model for Capturing Provenance and Auditing Information
Christoph Ringelsteinand Steffen Staab (2010). *International Journal of Web Services Research (pp. 1-20).*
www.irma-international.org/article/dialog-distributed-model-capturing-provenance/42107

An Energy-Aware and Under-SLA-Constraints VM Consolidation Strategy Based on the Optimal
Matching Method
WeiLing Li, Yongbo Wang, Yuandou Wang, YunNi Xia, Xin Luoand Quanwang Wu (2017). *International
Journal of Web Services Research (pp. 75-89).*
www.irma-international.org/article/an-energy-aware-and-under-sla-constraints-vm-consolidation-strategy-based-on-the-
optimal-matching-method/188458

An Approach of Role Updating in Context-Aware Role Mining
Jian Wang, Zejin Zhu, Junju Liu, Chong Wangand Youwei Xu (2017). *International Journal of Web Services
Research (pp. 24-44).*
www.irma-international.org/article/an-approach-of-role-updating-in-context-aware-role-mining/181298

A Similarity Measure Across Ontologies for Web Services Discovery
Aissa Fellah, Mimoun Malkiand Atilla Elci (2019). *Web Services: Concepts, Methodologies, Tools, and
Applications (pp. 859-881).*
www.irma-international.org/chapter/a-similarity-measure-across-ontologies-for-web-services-discovery/217867

Mobile Web Services for P2P Computing
Fahad Aijaz (2010). *Developing Advanced Web Services through P2P Computing and Autonomous
Agents: Trends and Innovations (pp. 133-157).*
www.irma-international.org/chapter/mobile-web-services-p2p-computing/43651