Chapter 7
Web Mining and Web Usage Mining for Various Human-Driven Applications

Avinash Kumar
School of Engineering and Technology, Sharda University, India

Bharat Bhushan
https://orcid.org/0000-0002-9345-4786
School of Engineering and Technology, Sharda University, India.

Nandita Pokhriya
School of Engineering and Technology, Sharda University

Raj Chaganti
Department of Computer Science, University of Texas at San Antonio, USA

Parma Nand
School of Engineering and Technology, Sharda University, India

ABSTRACT

The current century has seen a huge amount of data being produced every day by various computational platforms. These data are very large in amount when compared to the past. These data have emerged in form of big data (BD), and it's vital to manage these data for processing as well as extracting. The big data are emerging from Facebook, Instagram, and various other social as well as private platforms. Web mining has the capabilities to mine or extract data from a very huge set of data. The mining of useful information is essential for optimal functioning of various systems. In this chapter, the usage of web mining and one of its sub categories, web usage mining, for various human-driven applications has been discussed. This chapter focuses on the background that leads to the emergence of web mining. The chapter also deeply explores various subdomains that are essential for tackling data extraction using web mining.

DOI: 10.4018/978-1-7998-9426-1.ch007

INTRODUCTION

The commencement of a huge amount of data due to various new technologies and social media platforms has resulted in a new term called big data. This huge data couldn't be worked on (processed and analysed) with the help of existing technologies, theories, and methodologies (Quinto, 2018). Some typical properties of big data are volume, value, velocity, and variety (Somoza Sánchez, 2020). Studies show that the total amount of big data at the global level has reached about $58.9 billion in the year 2017, with an increment of about 29.1% (Alam et al., 2020). Big data has innate to it the capability to improve the governance, promote research, and make production more efficient (Chen et al., 2019; Emrouznejad & Marra, 2016). Cloud computing as of now has emerged as one of the novel paradigms that is being used to implement a shared pool of computing resources (storage, applications, and services), it has become increasingly easy to configure cloud computing tasks (Gill & Buyya, 2019; Miyachi, 2018). The establishment of these services based on big data makes for a great performing data cloud platform. Big data often suffers from repeated changes in size and scope, making powerful techniques the need of the day in order to cope with these changes. Therefore, it is important that the components of big data processing architecture are designed keeping in mind not only its volatile nature but also factors like cost, speed, and system scalability (Arostegi et al., 2018). Big data finds its applications in almost all walks of life, from finance to education to healthcare to defense and governance. In the Internet of Things (IoT) big data can be employed to realize applications like smart vehicles, indoor localization, innovative computer architectures (Chen et al., 2021).

Data mining has emerged as a vital step in the discovery of useful patterns and information in databases (Howard & Rayward-Smith, 1999). Many different routines from other domains such as statistics, databases, machine learning, pattern recognition, algorithms, high-performance computing, visualization, information retrieval, are included in data mining (Robardet, 2013; Yie et al., 2021). The first three methods being the entities that fundamentally contribute to data mining (Shi et al., 2018). Summarization, clustering, analysis of outliers, association, classification, regression, and analysis of trends, characterization, are some of the most common application areas of data mining methodologies (Ye et al., 2019). With the advancement in technology, even small-scale businesses have come to be a part of the community that finds itself inclined to use computer-based decision systems to get a better and clearer picture. With the usage of these technologies, the organizations more often than not, find themselves making profits leading to an overall benefit to both the stakeholders and the owners (Kumar, 2020). These systems over time have been evolved so as to reduce the burdens on humans when