


Chapter 15

A New Insight on the Morphology of Web Mining

Joshua Ojo Nehinbe

 <https://orcid.org/0000-0002-0098-7437>
Federal University, Oye, Nigeria

ABSTRACT

Recent surveys have revealed that about 199 million of active and over 1.2 Billion of inactive websites exist across the globe. The categories of websites have also increased beyond espionage networks of spies, computer networks for corporate organizations, networks for governments' agencies, networks for social interactions, search engines and networks for religious bodies, etc. These diversities have generated complex issues regarding the morphology and classification of webs and web mining. Thus, the validity of the generic web classification, web mining taxonomy, and contemporary studies on the regularities of web usage, web content, web semantic, web structures, and the process of extracting useful information and interesting patterns from the intricate of the Internet are frequently questionable. The existing web mining taxonomy can also lead to misinformation, misclassification, and crisscrossed issues such that numerous webs' patterns could be marked with crossing and inexplicable lines. By using qualitative virtual interviews of 26 skilled web-designers and a focus group-conference of 7 experts in web-usage to brainstorm on the above issues, this chapter comprehensively discusses the above concepts and how they relate to web classification and web mining taxonomy. The themes obtained elucidate the techniques that commonly underpin basic web mining taxonomy. New concepts like existence of esoteric web data, exoteric web data; mysterious, inexplicable, and mystifying patterns; and cryptic vocabularies are discussed to assist web analytics. Finally, the author suggests eight classification attributes for web mining patterns (illustrative, expositive, educative, advisory, interpretative, demonstrative, revealing, and informative) and proposes a new web mining taxonomy to minimize the impacts of the above concerns on global settings.

DOI: 10.4018/978-1-7998-9426-1.ch015

INTRODUCTION

The Art and science of websites' designs require web designers to possess creative and imaginative skills and capability to combine some standard technologies such as Hypertext Markup Language (HTML), Cascading Style Sheets (CSS); Extensible Markup Language (XML), Scalable Vector Graphics (SVG) to build and synthesize images and Application Programming Interfaces (APIs) or the intermediary software that enable two web applications to link and communicate with each other (W3C, 2021). Visual studies suggest that the morphology of web classification hints that modern websites now combine various branches of creative activities like music, painting and literary composition to typically produce visual works on websites that are primarily appreciated for their beauty, innovativeness and quality. In other words, the ontology and structure of the words that are frequently published on different websites and parts of such words could be classified on the basis of root, stem, prefix and suffix. For these reasons, websites regularly publish and keep the records of countless web data in diverse morphological components and formats.

Web data is a combination of piece of information and semantic of facts on websites. Fundamentally, web data subsumes diversity of web users, variety of web structure and array of web content regarding websites that are hosted on the Internet (Busetto et al., 2020; Singh et al., 2014). Web users are computer services and end-users such as the consumers, customers and clients that access websites. Web structure includes the arrangement, organization, composition, configuration, framework and makeup of websites. Similarly, web content is the variety of information that is published on the websites for the audience or web users (end-users). Information retrieval is the central part of the concept of web mining. Logically, the information that is retrieved from the web is indirectly extracted from web servers. The web servers usually log rich on the above web data. Such information may include remote hosts, successful and unsuccessful responses; parameters required to identify web users, authentications, status codes such as resource requested and the HTTP protocol, etc in standard formats. Nonetheless, web data or information that is retrieved from the webs might require high level of pre-processing due to their uniqueness and diversities in their sources, kinds, purposes and meanings. The pre-processing of web data required and the changeable meaning of the linguistics together with the logical combinations of the above groups of web data begin to pose serious challenges to web data analytics in two different ways on a daily basis (Chawan & Pamnani, 2010). Modern web data is perceived in terms of the logical semantics and lexical semantics in the above context. The logical semantics of web data are concerned with the common sense, reference, preconception and the conclusion that can be implicitly drawn from web data while the lexical semantics of web data are

18 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/a-new-insight-on-the-morphology-of-web-mining/300225

Related Content

Utilizing Past Web for Knowledge Discovery

Adam Jatowt, Yukiko Kawai and Katsumi Tanaka (2010). *Web Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 2544-2562).

www.irma-international.org/chapter/utilizing-past-web-knowledge-discovery/37752

Towards Ubiquitous and Adaptive Web-Based Multimedia Communications via the Cloud

Spyros Panagiotakis, Ioannis Vakintis, Haroula Andrioti, Andreas Stamoulias, Kostas Kapetanakis and Athanasios Malamos (2016). *Web Design and Development: Concepts, Methodologies, Tools, and Applications* (pp. 925-974).

www.irma-international.org/chapter/towards-ubiquitous-and-adaptive-web-based-multimedia-communications-via-the-cloud/137382

Collaborative Service Clouds

Liliana Ardissono, Gianni Bosio, Anna Goy, Giovanna Petrone, Marino Segnan and Fabrizio Torretta (2010). *International Journal of Information Technology and Web Engineering* (pp. 23-39).

www.irma-international.org/article/collaborative-service-clouds/49198

ThespiSTRX: Causally-Consistent Read Transactions

Carl Camilleri, Joseph G. Vella and Vitezslav Nezval (2020). *International Journal of Information Technology and Web Engineering* (pp. 1-16).

www.irma-international.org/article/thespistrx/241773

Improved Algorithm for Error Correction

Wael Toghuj and Ghazi I. Alkhatib (2011). *International Journal of Information Technology and Web Engineering* (pp. 1-12).

www.irma-international.org/article/improved-algorithm-error-correction/52802