A Systematic Review on the Profiling of Digital News Portal for Big Data Veracity

Normala Binti Che Eembi @ Jamil[a], Iskandar Bin Ishak[a], Fatimah Sidi[a], Lilly Suriani Affendey[a], Ali Mamat[a]

[a] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

Abstract

Currently, digital news portals have been one of the most important news sources for Internet users. However, the way it is written depends on the direction of the content. One approach to news reporting is through manipulative writing. Such method of writing has created a number of adverse outcomes such as political unrest, slander and negative perception towards the particular organization, personnel, and country. It is important for readers to choose and select news portal that is reporting positively and to neglect portals which practices manipulative writing approach for their own gains or causing negative impact towards the community. The aim of this study is to structure and analyzed the literature related to data veracity research that can be used to the profile of digital news portal. The method that has been used in this paper is to classify and define data veracity; a systematic literature review is a conduct. It includes journal and conference proceedings. The results come out with objectives in data veracity, the structure of research topics, research trends with publications and framework veracity model validated. This paper provides a complete review of literature related to profiling digital news portal in data veracity.

1. Introduction

One of the characteristics of big data is data veracity. Data veracity refers to principles of truth-telling, and it is grounding in respect for persons and the concept of autonomy [1]. He or she must have the information relevant to his or her decision if the individual wants to make entirely rational choices. Moreover, this information must be as transparent and comprehensible as possible. At least two ways disrupted in truth-telling. The first is by the act of untruthful, or the deliberate conversation of wrong info. But, the value of veracity is also despoiled by omission, the cautious of helpings of the truth. Finally, the principles of veracity can also be violated by the considered hiding of evidence in jargon or

linguistic. It fails to express info in a way that can be unwritten by the receiver or that intentionally misleads the recipient.

Through data veracity, four components of trustworthiness, availability, accountability and authenticity can be measured. Incorrect data can cause a lot of problems for news reporting. Therefore, we need to ensure that the data is correct as well as the analyses performed on the data are correct. We need to be sure that both the data and the analyses are correct especially in automated decision-making where no human is involved anymore.

Most of the previous studies mentioned about manipulative writing and hiding information from authorship of unknown documents [2][3][4]. Mostly paper only profile a user or social media [5][6][7][8]. To identify truth, the researcher used forensic techniques. They also highlighted about the adversarial attack which include imitation attack and obfuscation attacks. Imitation attack known as "subject attempts to frame another topic by imitating their writing styles" and obfuscation attack is "subject hide their identity." Researcher [9] reported to determine the truthfulness is related with deception detection. It can support law enforcement decision makers to identify deceptive incident report statement. They focus on different types of websites. Same with previous researcher also notify about deception detection in order to define truth such as [2][10][11][12][13] [14], [15], [16],[9],[17]. The advantages using big data veracity is it can solve many problem that relate to manipulative writing and define truthfulness in future.

In this context, a profiling digital news portal in big data veracity has gained increased attention in current research and practices [18]. The articles only focus on rating in social media. For example determining a tabloid rating for a news article from paper "increasing the veracity of event detection on social media networks through user trust modeling" [19]. Because of the manipulative writing we want to know which portal are biased or neutral. By using data veracity method, we wish to identify the profile of digital news portal. In summary, there is a lack of approach in profiling digital news portal.

The purpose of this article is to review about the research direction in big data veracity that can be applied in digital news portal using the tertiary study to review articles related to truthfulness. In particular, we concentrate on articles describing systematic literature reviews (SLRs)[20]. We describe our methodology in section 2 and present our results in section 3. We present our conclusion in section 4.

2. Method

This study has been undertaken as a systematic literature review based on the original guidelines as proposed by [21][22]. A systematic review is a synthesis of primary search studies. It is serving a particular topic with the formulation of specific questions and clear, the method explicit and reproducible search, involving critical examination in the selection process study, and communicate the results and implication [23]. The steps in the systematic literature review method are documented in figure 1.
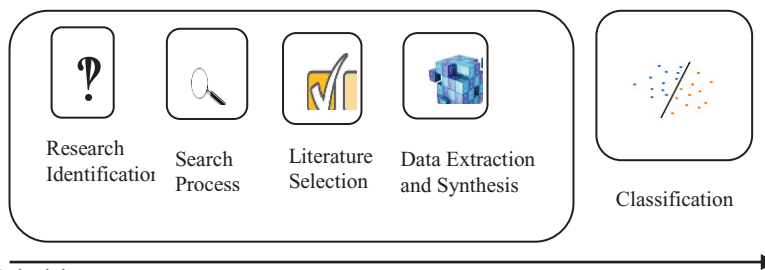Systematic Literature Review

Research Identification Search Process Literature Selection Data Extraction and Synthesis Classification

Figure 1: Research Methodology