A review of data quality research in achieving high data quality within organization

ABSTRACT

The aim of this review is to highlight issues in data quality research and to discuss potential research opportunity to achieve high data quality within an organization. The review adopted systematic literature review method based on research articles published in journals and conference proceedings. We developed a review strategy based on specific themes such as current research area in data quality, critical dimensions in data quality, data quality management model and methodologies and data quality assessment methods. Based on the review strategy, we select relevant research articles, extract and synthesis the information to answer our research questions. The review highlights the advancement of data quality research to resemble its real world application and discuss the available gap for future research. Research area such as organizations management, data quality impact towards the organization and database related technical solutions for data quality dominated the early years of data quality research. However, since the Internet is now taking place as the new information source, the emerging of new research areas such as data quality assessment for web and big data is inevitable. This review also identifies and discusses critical data quality dimensions in organization such as data completeness, consistency, accuracy and timeliness. We also compare and highlight gaps in data quality management model and methodologies. Existing model and methodologies capabilities are restricted to the structured data type and limit its ability to assess data quality in web and big data. Finally, we uncover available methods in data quality assessment and highlight its limitation for future research. This review is important to highlight and analyse limitation of existing data quality research related to the recent needs in data quality such as unstructured data type and big data.