

Abstract

System logs are essential assets in today's experimental research supporting trends like AI driven analytics, automation and reproducible science. By providing preparation techniques, the review has provided a unified framework that is able to enhance data quality, improve experimentation and increase system complexity. This later assists in developing more accurate, more accurate, scalable and futuristic models, especially in big data, machine learning and intelligent systems. This review aims to systematically find out the importances of system logs in research and to find out how logs are prepared before they are used for research. Research papers dated 2021 to 2025 were used in the review. Barbara Kitchenham's guidelines were used to review the journal papers and to write the entire paper. Several databases were used to get the literature such as IEEE, Elsevier, MDPI, Springer, SpringerOpen, Wiley, Frontiers, PLOS, PeerJ, Taylor & Francis, arXiv (preprint). The results indicated that importance of logs included doing objective event capture by use of time stamps, it's aground of truth for security or anomaly experiments, it's an input in ML experiments, acts as a provenience reproducibility artifact, it's an experimental variable, its used in behavioural and interaction analysis, it used in performance monitoring and system optimization. It was also noted that logs preparation is done in stages that is, define scope and objectives, collect and centralize, normalize timestamps, parse logs, clean and canonicalize, feature extraction, sampling or instance selection, privacy or anonymization, validation and benchmarking, document and archive. For future works, researchers can create models advancing automated log analysis, standardization of log format and metadata, reproducible research frameworks