A Review of Unstructured Data Analysis

VR Weerasekara, G Gayamini, TL Weerawardane

Department of Computer Engineering, Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka

Abstract. Every day, computer programs generate enormous amounts of data in the form of your records, machine-generated papers, and other reports. This generated data constitutes the kingdom of the walking machine and contains a variety of data that can be used for machine monitoring and diagnostics. Because network monitoring systems generate a large amount of unstructured data, an automated method to obtain the required information is required. information, which now necessitates the use of many specialized analyst's Custom analyser development and testing can be time-consuming. Instead, the data can be mechanically processed and analysed in a computer-readable format, yielding a common version for standard or vendor-specific information and insights assessment, anomaly detection, intrusion detection, node failure, and a variety of apps. This paper examines a few modern approaches to mining and analyzing unstructured data, as well as the challenges associated with data extraction, the introduction of technological foundations, and the provision of a standard framework for computerized review. In this paper, I present the most recent methods for extracting unstructured text information. Relationship-based algorithms that use function template-based approaches and deep learning techniques, as well as data-driven approaches, are available to me. I also go over the analysis problems such as Values that are missing, not linguistic, No Metadata, Template of several types and tools that are available (Extraction, transformation, Modelling, Execution etc...). For each identified component, I describe the data analysis method and general techniques.

Keywords: Knowledge base, Data Mining, Information Extraction, Similarity, Natural Language Processing