



PAPER ID: 17-05-0011	<p style="text-align: center;">Application of NLP on Big Data using Hadoop: Case Study Using Trouble Tickets</p> <p style="text-align: center;">Fauzy Che Yayah, Khairil Imran Ghauth , Choo-Yee Ting</p> <p style="text-align: center;">Faculty of Computing & Informatics, Multimedia University, Cyberjaya, Malaysia</p> <p>Telecommunication company trouble tickets system contains reported incidents tickets that related to networks service interruption or problems. Trouble tickets system is the example of the application that deals with a large amount of textual data. In order for the team that is handling the trouble tickets system, they spend a lot of time to analyze the data manually. In this paper, Natural Language Processing (NLP) approach has been introduced to solve the problem. By applying this technique, manual of activities can be automated and reduce the time and effort to find the classification of the closing resolution code. It also helps the trouble tickets system to collect incidents trending and resource utilization. New data processing method with NLP and sublanguage introduced via big data platform to deliver faster classification computation. The outcome of this study is transformation method of the original data set into the analytics series, and identification the characteristics of the trouble tickets data set to enable the classification of the resolution code. The data processing workflow shows that the linguistics of the trouble tickets fit the sublanguage theoretical framework thus enabling to tap into the unrealized value inside it. The data processing and data transformation workflow describe the linguistics of the trouble tickets fit the sublanguage theoretical framework, therefore, supporting the research to tap into the unexposed content inside it.</p> <p>Keywords: Data Processing, Incident Management, Natural Language Processing, Classification</p>
PAPER ID: 17-05-0135	<p style="text-align: center;">Distributed Join Query Processing for Big RDF Data</p> <p style="text-align: center;">Nahla Mohammed Elzein¹, Mazlina Abdul Majid¹, Mohammed Fakheridin² Ibrahim Abaker Targio Hashem³,</p> <p style="text-align: center;">¹Faculty of Computer System & Software Engineering, University Malaysia Pahang, 26600 Pekan, Pahang, Malaysia</p> <p style="text-align: center;">², Faculty of Computer Science and Information Systems, Saudi Arabia, Jazan University</p> <p style="text-align: center;">³Asia Pacific University of Technology and Innovation Technology Park Malaysia, Kuala Lumpur, Malaysia</p> <p>The expansion of the services of the Semantic Web and the evolution of cloud computing technologies have significantly enhanced the capability of preserving and publishing information in standard open web formats, such that data can be both human-readable and machine-processable. This situation meets the challenge in the current big data era to effectively store, retrieve, and analyze resource description framework (RDF) data in swarms. Moreover, efficient data storage and retrieval that can scale to large amounts of possibly schema-less data have proven to be quite difficult to achieve, specifically, RDF data storage with complex and large graph patterns for representing semantic data, and SPARQL query languages. In this paper, we provide comprehensive discussion about the proposed algorithms of Join.Query processing of RDF data by considering MapReduce Framework in a distributed environment. Moreover, we introduced a framework for RDF query processing and the benchmark that is used for the performance evaluation. Finally, we offer an evaluation discussion on distributed join query processing for big RDF data.</p> <p>Keywords: Semantic Web, Big data, Query processing.</p>