

Graph-Oriented Summary for Optimized Resource Description Framework Graphs Streams Processing

Authors : Amadou Fall Dia, Maurras Ulbricht Togbe, Aliou Boly, Zakia Kazi Aoul, Elisabeth Metais

Abstract : Existing RDF (Resource Description Framework) Stream Processing (RSP) systems allow continuous processing of RDF data issued from different application domains such as weather station measuring phenomena, geolocation, IoT applications, drinking water distribution management, and so on. However, processing window phase often expires before finishing the entire session and RSP systems immediately delete data streams after each processed window. Such mechanism does not allow optimized exploitation of the RDF data streams as the most relevant and pertinent information of the data is often not used in a due time and almost impossible to be exploited for further analyzes. It should be better to keep the most informative part of data within streams while minimizing the memory storage space. In this work, we propose an RDF graph summarization system based on an explicit and implicit expressed needs through three main approaches: (1) an approach for user queries (SPARQL) in order to extract their needs and group them into a more global query, (2) an extension of the closeness centrality measure issued from Social Network Analysis (SNA) to determine the most informative parts of the graph and (3) an RDF graph summarization technique combining extracted user query needs and the extended centrality measure. Experiments and evaluations show efficient results in terms of memory space storage and the most expected approximate query results on summarized graphs compared to the source ones.

Keywords : centrality measures, RDF graphs summary, RDF graphs stream, SPARQL query

Conference Title : ICDM 2018 : International Conference on Data Mining

Conference Location : Dublin, Ireland

Conference Dates : July 23-24, 2018