Semantic Data Schema Recognition

Authors : Aïcha Ben Salem, Faouzi Boufares, Sebastiao Correia

Abstract : The subject covered in this paper aims at assisting the user in its quality approach. The goal is to better extract, mix, interpret and reuse data. It deals with the semantic schema recognition of a data source. This enables the extraction of data semantics from all the available information, inculding the data and the metadata. Firstly, it consists of categorizing the data by assigning it to a category and possibly a sub-category, and secondly, of establishing relations between columns and possibly discovering the semantics of the manipulated data source. These links detected between columns offer a better understanding of the source and the alternatives for correcting data. This approach allows automatic detection of a large number of syntactic and semantic anomalies.