

Research of Data Cleaning Methods Based on Dependency Rules

Authors : Yang Bao, Shi Wei Deng, WangQun Lin

Abstract : This paper introduces the concept and principle of data cleaning, analyzes the types and causes of dirty data, and proposes several key steps of typical cleaning process, puts forward a well scalability and versatility data cleaning framework, in view of data with attribute dependency relation, designs several of violation data discovery algorithms by formal formula, which can obtain inconsistent data to all target columns with condition attribute dependent no matter data is structured (SQL) or unstructured (NoSQL), and gives 6 data cleaning methods based on these algorithms.

Keywords : data cleaning, dependency rules, violation data discovery, data repair

Conference Title : ICCCISE 2015 : International Conference on Computer, Communication and Information Sciences, and Engineering

Conference Location : Barcelona, Spain

Conference Dates : October 26-27, 2015