

Heterogeneous data source integration for smart grid ecosystems based on metadata mining

Abstract— The arrival of new technologies related to smart grids and the resulting ecosystem of applications and management systems pose many new problems. The databases of the traditional grid and the various initiatives related to new technologies have given rise to many different management systems with several formats and different architectures. A heterogeneous data source integration system is necessary to update these systems for the new smart grid reality. Additionally, it is necessary to take advantage of the information smart grids provide. In this paper, the authors propose a heterogeneous data source integration based on IEC standards and metadata mining. Additionally, an automatic data mining framework is applied to model the integrated information.

Index Terms—Smart grids; large-scale integration; data mining; standards; metadata mining; big data.

I. INTRODUCTION

The traditional systems in power distribution grids usually have databases with different data structure. The new technologies related to Smart Grids have provided the opportunity of new and advanced functions. Although these new systems are based on the usage of sensor networks and information systems, the systems need the information from older systems, integrating information from heterogeneous data sources. In this sense, there are several problems which need to be solved:

- Information integration. The new systems need to take advantage of old and new data sources. Thus, the integration of these heterogeneous data sources is very difficult, because each database has their own structure. This data source should be translated to a common format. In this way, the information standards provide a good source for a Common Information Models (CIM).