

A Semantic-Based Approach for Data Management in a P2P System

Damires Souza¹, Carlos Eduardo Pires², Zoubida Kedad³,
Patricia Tedesco⁴, and Ana Carolina Salgado⁴

¹ Federal Institute of Education, Science and Technology of Paraíba (IFPB), Brazil
damires@ifpb.edu.br

² Federal University of Campina Grande (UFCG), Computer Science Department
Av. Aprígio Veloso, 882, Bodocongó - 58109-970 - Campina Grande, PB, Brazil
cesp@dsc.ufcg.edu.br

³ Université de Versailles et Saint-Quentin-en-Yvelines (UVSQ)
45 Avenue des Etats-Unis, 78035 Versailles, France
zoubida.kedad@prism.uvsq.fr

⁴ Federal University of Pernambuco (UFPE), Center for Informatics
P.O. Box 7851, 50.732-970, Recife, PE, Brazil
{pcart,acs}@cin.ufpe.br

Abstract. Data management in P2P Systems is a challenging problem, due to the high number of autonomous and heterogeneous peers. In some Peer Data Management Systems (PDMSs), peers are semantically clustered in the overlay network. A peer joining the system is assigned to an appropriate cluster, and a query issued by a user at a given peer is routed to semantic neighbor clusters which can provide relevant answers. To help matters, semantic knowledge in the form of ontologies and contextual information has been used successfully to support the techniques used to manage data in such systems. Ontologies can be used to solve the heterogeneities between the peers, while contextual information allows a PDMS to deal with information that is acquired dynamically during the execution of a given query. The goal of this paper is to point out how the semantics provided by ontologies and contextual information can be used to enhance the results of two important data management issues in PDMSs, namely, peer clustering and query reformulation. We present a semantic-based approach to support these processes and we report some experimental results which show how semantics can improve them.

Keywords: PDMS, Semantics, Data Management, Ontologies, Context, Clustering.

1 Introduction

Peer-to-Peer (P2P) systems are massively distributed and highly volatile systems for sharing large amounts of resources. Peers communicate through a self-organizing and fault tolerant network topology which runs as an overlay on top of the physical network. Peer Data Management Systems (PDMSs) [1, 2, 3, 4, 5, 6] came into the focus