

## **Exploring instances for matching heterogeneous database schemas utilizing Google similarity and regular expression**

### **ABSTRACT**

Instance based schema matching aims to identify correspondences between different schema attributes. Several approaches have been proposed to discover these correspondences in which instances including those with numeric values are treated as strings. This prevents discovering common patterns or performing statistical computation between numeric instances. Consequently, this causes unidentified matches for numeric instances which further effect the results. In this paper, we propose an approach for addressing the problem of finding matches between schemas of semantically and syntactically related attributes. Since we only fully exploit the instances of the schemas, we rely on strategies that combine the strength of Google as a web semantic and regular expression as pattern recognition. To demonstrate the accuracy of our approach, we have conducted an experimental evaluation using real world datasets. The results show that our approach is able to find 1-1 matches with high accuracy in the range of 93% - 99%. Furthermore, our proposed approach outperformed the previous approaches using a sample of instances.

**Keyword:** Schema matching; Instance based schema matching; Google similarity; Regular expression