

## EXTENDMATCH: A NEW SCHEMA MATCHING ALGORITHM IN MODEL MANAGEMENT

DONGMING ZHANG<sup>1,2</sup>, WEI CHEN<sup>3</sup> AND GUOHUA LIU<sup>1</sup>

<sup>1</sup>School of Computer Science and Technology  
Yanshan University  
Qinhuangdao, Hebei 066004, P. R. China  
zdmqzf@yahoo.com.cn; ghliu@ysu.edu.cn

<sup>2</sup>Qinhuangdao Institute of Technology  
Beidaihe, Hebei 066100, P. R. China

<sup>3</sup>Department of Information Engineering  
Environmental Management College of China  
Qinhuangdao, Hebei 066000, P. R. China  
fallafall@163.com

Received December 2006; revised August 2007

**ABSTRACT.** *Schema matching, which works for finding semantic correspondences between elements of two schemas, plays a key role in many applications, such as data warehouse, heterogeneous data sources integration and semantic web. Currently, when a similarity flooding (SF for short) algorithm is used to do schema matching directly in the model management, there are some problems. Firstly, when column names are different, but the types of data are the same, there would be many uncertain match candidates. Secondly, when column names and the types of data are both different, the match candidates which would be matched logically can not be obtained. In this paper, we propose a new schema matching algorithm named ExtendMatch based on these shortcomings, which can not only be applied to the model management well, but also reduce the manual manipulation greatly, enhance the grade of automatization, and implement the semi-automatization or automatization of maintenance and management after publishing XML views.*

**Keywords:** Model management, Schema matching, Similarity, Mutual information, Entropy

**1. Introduction.** Publishing views on the Internet such as XML documents has become the primary content of Web Service. XML has become standard language exchanging data on the Internet. However, when the maintenance and management after publishing XML views have to operate, the manual operations are used to find out the changes from the changed source schema. And then the experts decide how to deal with the changes. Hence, the update and maintenance of publishing XML views is one of the important issues in model management.

Model management [1-3], which is a new method of metadata management, can deal with the maintenance and management after publishing XML views well. Compared with current technologies, it can provide more advanced programming design interface. Models and mappings between them are two abstract concepts. Model is a complex structure that represents a design artifact, such as relational schema, object-oriented interface, UML model, XML DTD, web-site schema, or semantic network. Schema matching is the task of finding semantic correspondences between elements of two schemas. It is a key issue in many domains, such as data warehouse, heterogeneous data sources integration