International Journal of Computer Architecture and Mobility
(ISSN 2319-9229) Volume 1-Issue 4,February 2013

ANALYSIS OF SCHEMA MATCHING TECHNIQUE USING XML DATA FOR PERFORMANCE ENHANCEMENT

Swati joshi (SDBCE), Indore

ABSTRACT:

The Web becomes a large source of information and data, and knowledge extraction and pattern mining in web data is known as web mining. Web mining categorized into three popular domain of knowledge web uses mining, structure mining and content mining. Mappings between different representations of data are the essential building blocks for many information integration tasks. A schema mapping is a high-level specification of the relationship between two schemas, and represents a useful abstraction that specifies, how the data from a source format can be transformed into a target format. In this project we are going to propose the data integration technique by using web semantics and similarity techniques.

Keywords: web mining, web content mining, similarity analysis, schema mapping.

INTRODUCTION

Web mining can be categorized into three different classes based on which part of the Web is to be mined.

These three categories are

(i) Web content mining,

(ii) Web structure mining and

(iii) Web usage mining.

Web content mining [10, 9] is the task of discovering useful information available on-line. There are different kinds of Web content which can provide useful information to users, for example

multimedia data, structured (i.e. XML documents), semi-structured (i.e. HTML documents) and unstructured data (i.e. plain text). The aim of Web content mining is to provide an efficient mechanism to help the users to find the information they seek. Web content mining includes the task of organizing and clustering the documents and providing search engines for accessing the different documents by keywords, categories, contents etc.

Schemas and schema mappings are two fundamental metadata components that are at the core of heterogeneous data management. Schemas describe the structure of the various databases, while schema mappings describe the relationships between them. Schema mappings can be used either to transform data between two different schemas (a process typically called data

exchange (Fagin et al, 2005a) or data translation (Shu et al, 1977)) or to support processing of queries formulated over one schema when the data is physically stored under some other schemas (a process typically encountered in data integration (Lenzerini, 2002) and also in schema evolution (Curino et al, 2008)).

A schema mapping is typically formalized as a triple (S,T,Σ) where S is a source schema, T is a target schema, and Σ is a set of dependencies (or constraints) that specify the relationship between the source schema and the target schema. Schema mappings are necessarily dependent on the schemas they relate. Once schemas change (and this inevitably happens overtime), the mappings become invalid. A typical solution is to regenerate the mappings; however, this process can be expensive in terms of human effort and expertise, especially for complex schemas. Moreover, there is no guarantee that the regenerated mappings will reflect the original semantics of the mappings. A better solution is to provide principled solutions that reuse the original mappings and adapt them to the new schemas, while still incorporating the original semantics. This general process was first described in (Velegrakis et al, 2003), which called it mapping adaptation and also provided a solution that applied when schemas evolve in small, incremental changes. In this paper, we describe a more general formalization of the mapping adaptation problem where schema evolution can be specified by an arbitrary schema mapping. Under this formalization, which is in the spirit of model management (Bernstein, 2003), the new, adapted mapping is obtained from the original mapping through the use of schema mapping operators.

LITERATURE SURVEY:

Schema Mapping Evolution through Composition and Inversion: Mappings between different representations of data are the essential building blocks for many information integration tasks. A schema mapping is a high-level specification of the relationship between two schemas, and represents a useful abstraction that specifies how the data from a source format can be transformed into a target format. The development of schema mappings is laborious and time-consuming, even in the presence of tools that facilitate this development. At the same time, schema evolution inevitably causes the invalidation of the existing schema mappings (since their schemas change). Providing tools and methods that can facilitate the adaptation and reuse of the existing schema mappings

in the context of the new schemas is an important research problem. In this chapter, we show how two fundamental operators on schema mappings, namely composition and inversion, can be used to address the mapping adaptation problem in the context of schema evolution. We illustrate the applicability of the two operators in various concrete schema evolution scenarios, and we survey the most important developments on the semantics, algorithms and implementation of composition and inversion. We also discuss the main research questions that still remain to be addressed.

SAXM : Semi-automatic XML Schema Mapping: Data integration technology becomes essential in processing many types of data. This paper proposes a new method called Semi-Automatic XML schema Mapping (SAXM). Semantic similarity is firstly used to separate unmatchable nodes. Structural similarity and data type compatibility are then applied to ascertain the matching of two nodes. The output of SAXM is a set of similarity relationship values between two XML schemas elements. Moreover, a user can verify the mapping to provide a more accurate result. Experimental results show that our proposed method provides the highest recall and comparable precision among well-known approaches.

++Spicy: an Open-Source Tool for Second-Generation Schema Mapping and Data Exchange: Recent results in schema-mapping and data-exchange research may be considered the starting point for a new generation of systems, capable of dealing with a significantly larger class of applications. In this paper we demonstrate the first of these second-generation systems, called ++Spicy. We introduce a number of scenarios from a variety of data management tasks, such as data fusion, data cleaning, and ETL, and show how, based on the system, schema mappings and data exchange techniques can be very effectively applied to these contexts. We compare++Spicy to the previous generations of tools, to show that this is much-needed advancement in the field.

Discovery and Correctness of Schema Mapping Transformations: Schema mapping is becoming pervasive in all data transformation, exchange and integration tasks. It brings to the surface the problem of differences and mismatches between heterogeneous formats and models, respectively used in source and target databases to be mapped one to another. In this chapter, we start by describing the

problem of schema mapping, its background and technical implications. Then, we outline the early schema mapping systems, along with the new generation of schema mapping tools. Moving from the former to the latter entailed a dramatic change in the performance of mapping generation algorithms. Finally, we conclude the chapter by revisiting the query answering techniques allowed by the mappings, and by discussing useful applications and future and current developments of schema mapping tools.

PROPOSED WORK:

There are various applications are available in web for matching schema and compare them additionally arrange them. The main problem is to achieve high accurate schema mapping technique in order to provide the way where the resource consumption is minimum. All the tools that are work for matching schema is work with high performance resources because of there are a need to process the data and match them. In this project we provide semantics based schema matching algorithm or architecture which consumes less time and memory space and produces high accurate results.

RESULT:

1. Resolving the overhead and unwanted resource consumption problem.

2. Resolve the problem to find accurate results.

3. Propose design and implement a new technique of web XML based schema matching.

CONCLUSION:

1. Study of different schema matching techniques

2. derive formulas for semantically analysis of schema mapping.

3. Tool implementation for data analysis and arrangement.

REFERENCES:

[1] Discovery and Correctness of Schema Mapping Transformations Angela Bonifati, Giansalvatore Mecca, Paolo Papotti, and YannisVelegrakis

[2] ++Spicy: an Open-Source Tool for Second-Generation Schema Mapping and Data Exchange, ProceedingsoftheVLDBEndowment, Vol. 4, No. 12 Copyright 2011 VLDB Endowment 2150-8097/11/08...

[3] SAXM : Semi-automatic XML Schema Mapping, National Center for Genetic Engineering and Biotechnology, Thailand , Khon khea university, copyright 2009.

[4] Schema Mapping Evolution through Composition and Inversion, To appear in Z. Bellahsene, A. Bonifati, and E. Rahm,

editors, Schema Matching and Mapping, Springer, 2011

[5]] Li Qian ,Michael J. Cafarella, H.V .jagadish , "Sample driven schema matching", SIGMOD '12, May 20–24, 2012, Scottsdale, Arizona, USA. Copyright 2012 ACM 978-1-4503-1247-9/12/05 ...$10.00

[6]Arenas M, P´erez J, Riveros C (2008) The Recovery of a Schema Mapping: Bringing Exchanged Data Back. In: PODS, pp 13–22

[7]Bernstein PA (2003) Applying Model Management to Classical Meta-Data Problems. In: Conference on Innovative Data Systems Research (CIDR), pp 209–220

[8]Bernstein PA, Green TJ, Melnik S, Nash A (2008) Implementing Mapping Composition. VLDB Journal 17(2):333–353

[9] .Hong-Hai Do and Erhard Rahm, "COMA: a system for flexible combination of schema matching aproaches", the VLDB Journal, P. 610-621, 2002.

[10]Guangming Xing, "Fast approximate Matching between XML Documents and Schemata", Lecture Notes in Computer Science, P. 425-436, 2006.