International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Survey of Data Visualization and Transformation Tools

Rahil A. Khera[1], P. S. Game[2]

[1, 2] Pune Institute of Computer Technology, Affiliated to SPPU, Pune, Maharashtra, India

Abstract: ***Data visualization and translation tools are one of the prime requirements of business processes. These tools are also known as mappers. It is a tool, which enables a visual representation of data and allows users to generate a relation between two forms of data. Generally, these two forms are known as source and target schemas and the relation between these schemas is known as mapping. The output of a mapper is a code representing the relationship (mappings) between the data. A deep research has been done on such mapping tools and also there are some mapping tools available commercially. Hence the basic motive of this paper is to understand the research carried out in this field and the existing tools, to know their benefits and limitations.***

Keywords: Mapper, Data Visualization, Data Transformation, XSLT mapper, Business Process

1. Introduction

A mapper is a graphical user interface (GUI) based tool which can be used for generation of mapping (transformation/relationship) between two different forms of data also known as schema. These schemas are represented in the form of trees and are known as source schema and the target schema [1] [3].

The mappings between these trees are represented by lines, and each line has its own meaning which plays a specific role in the transformations [1]. On the basis of these representative lines (mappings), an equivalent code would be generated which signifies either transformation of source into target or relation between source and target. This code can be used for the transformation of data before it is given as an input to another process. The representative code can be generated in any of the available programming languages. For example, when a student fills a registration form containing his/her roll number, then that roll number has to be given as input to the insert query. Such type of relationship is known as simple mapping.

But, there could be the cases, where before giving data as an input to the target schema, some types of operations are to be performed. Such type of operations results into complex mappings [1] [5] [6] [7].

For example, when an employee fills the registration form of a company, the form contains name and surname field. Now in company, let's suppose the format of the official email – id is the concatenation of the first character of the name and surname of the employee. Now, this information and email id are required for inserting the data into employee database of the company. In such a case, employee form details will be sourced schema while the data which will be used for insert query would be targeted schema and there would be lines representing a mapping from name node and surname nodes of the source schema to the email node of the target schema in such a way that the first character of name would be fetched i.e. a substring type of operation and then it will be concatenated with surname. Hence, in this case, two operations are applied on two different data of source schema and then the result of later operation is given as input to the target schema. This whole scenario is an example of complex mappings.

Normally, either this mapper would be working in conjunction with other systems or would be part of some systems [5] [6] [7]. Hence it should be light in weight [2].

2. Related Work

A data visualization and transformation tool also known as mapper which requires following objectives to be taken into the consideration.

2.1 Visualization of Data

There should be two schemas available at any given point of instance, i.e. source schema and target schema, between which the mapping has to be generated. These schemas have to be graphically represented. So the user can visualize the schemas and hence can generate the mapping i.e. relationship between the corresponding elements as per the requirement. These schemas are graphically represented in the form of trees [1] [5] [6] [7].

2.2 Visualization of Mappings

The mapping between any two elements (nodes) of the tree can be represented by lines. The mappings would be used for generation of representative code. The line representation between a source node and the target node implies that either there exists a direct mapping between both the nodes or child node(s) of either one or both are involved in mapping. The representation of lines depends on whether the parent node of the node involved in mapping is open or close (expanded or collapsed) [1] [5] [6] [7].

2.3 Types of Mappings

The mappings can be classified into two categories: a) simple mapping b) complex mapping. In simple mapping, there is

one to one correspondence between the elements of source schema and target schema. Complex mappings can be further classified into two types: a) some type of operations would be carried out on the elements of the source schema, and then the output of those operations would be mapped to an element of the target schema. For example, if employee id of an employee has to be generated from the elements of an html form, in the following format name. Surname, then name and surname can be given as input from source schema with "." to the string concatenation function to generate the employee id in required format. b) Complex mappings also constitute of many to one mapping, i.e. from two or more different source nodes to single target node, merging two schemas which are related with respect to some property or attribute and converting a schema in a grouped schema, for example suppose there is a schema containing various customers, so there is a possibility that a single address has more than one customers, Hence it may be the requirement that the customers with same addressed should be grouped together[1][5][6][7].

2.4 Operations

The operations such as the ones discussed in section 2.3 can be applied using pre – defined functions or using user – defined functions. Almost all the mappers, [5] [6] [7] [9]10] discussed in the next section, support pre – defined functions, while some [7] [9] [10] also support user defined functions. The user defined functions should be able to process the input from the source schema or from the outputs of other functions and generate the output, which can be given as input to target schema with the minimal or no efforts.

2.5 Languages of Code

After governing mappings from user, representative code should be generated. The code can be of any language varying from C++, C##, J#, XSLT, JAVA [1] [3] [4] [5] [6] [7].

2.6 Automated Mapping

In case of large schemas, it is required that the user should be provided recommendations for the creations of the mappings or should initially generate the mappings. For this, heuristic approach can be applied [3] [5] [6] [7].

3. Comparative Analysis

A comparative analysis is presented in table 1, where different research works and existing mappers are stated with their contributions and limitations.

Table 1: Different Mappers, their contribution and limitations

Paper References	Work	Contribution	Limitations
Clip: a Visual Language for Explicit Schema Mappings [1]	• Created XML mapping tool • Represented the mapping in the form of lines.	• More expressive combinations of lines for representing mappings • Complex mapping from source to target. • XQuery generated as an output.	• Automated Mapping of the complex schemas is not available. • A code representing the mapping is not generated.
Implementation of XSLT-based Schema Mapper using RCP [2]	• XSLT mapper based on Rich Client Platform. • Transformation on basis of WYSIWYG.	• Supports multiple types of document format. • Schema editor to edit the generated schemas.	• Complex mappings not supported. • User – defined functions are not supported.
Automated Data Mapping Specification via Schema Heuristics and User Interaction [3]	• Heuristics based mappings for data transformations	• Automated mapping. • Interactive based, user generated mapping.	• Complex mappings are not supported. • User – defined functions not supported.
Translating Web Data [4]	• XML schema and conversion tool	• User generated mappings are converted into queries. • Constraints can apply to the mappings.	• Cannot apply complex queries.
BizTalk Mapper [5] [8]	• XML mapper for Visual Studio .NET framework	• User defined functions supported. • Supports Visual Basic, C++, C# and J# for user defined functions.	• Multiple sources to single destination (target) mapping not available. • Has to be integrated in Visual Studio. • Depending on .NET frame – work. • Works only with Windows based operating system.
Altova MapForce [6] [8]		• User Defined functions on XSLT, C++, C# and JAVA. • Can work standalone. • Can be integrated into eclipse and visual studio. • Supports various data formats.	• Multiple mapper instances cannot be generated. • Difficult to reuse.

TIBCO Business Events Mapper [7]	• XSLT mapper as an integrated tool in Eclipse	• Supports complex mappings based on various XSLT statements. • Generates code in XSLT. • XSLT code of each node can be obtained. • Supports XSLT code to Mapping generation.	• Multiple mapper instances cannot be generated. • It comes as a part of TIBCO studio, which is developed on Eclipse.
Stylus Studio [9]	• Multi – purpose XML editor for transformation and visualization.	• Supports various types of transformations. • Supports both XQuery and XSLT.	• Cross platform support not available. • Works only for windows based operating system.
Transforming Data Using XQuery [10]	• XQuery Mapper as an integrated tool in Eclipse. • Supports XQuery Transformation.	• Supported Schemas: - XSD (XML Schema Definition) • WSDL (Web Service Definition Language. • MFL (Message Format Language. • Mapping between elements to attribute, attribute to element and attribute to attribute. • Some of the complex mappings are supported.	• Cannot be used as a standalone application. It has to be used as an integrated tool for Eclipse. • Cannot generate the mappings from previously generated code. • Hence cannot give output of one mapper to another mapper.

4. Conclusion

In this paper, we have completed a deep survey of research works which have been carried out in relation with mapping tools and the existing commercial mapper tools. Hence this paper can be used as a base for selection of a mapper tool as per the requirement.

References

[1] Alessandro Raffio, Daniele Braga, "Clip: a Visual Language for Explicit Schema Mappings," in IEEE' ICDE 2008, 978-1-4244-1837-4
[2] Tae-Jin Ha, Hye-Ja Bang, "Implementation of XSLT-based Schema Mapper using RCP", in IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.5, May 2009
[3] Sebastian Bossung, Hermann Stoeckle "Automated Data Mapping Specification via Schema Heuristics and User Interaction " in Proceedings of the 19th International Conference on Automated Software Engineering 2004.
[4] Lucian Popa, Yannis Velegrakis, "Translating Web Data" in Proc. 2002 Very Large Databases Conference (VLDB'02), pp. 598-609.
[5] BizTalk Mapper, "Using BizTalk Mapper" https://msdn.microsoft.com/en-us/library/aa547076.aspx., Dec 25, 2014.
[6] ALTOVA Map Force, "MapForce – Graphical Data Mapping, Conversion, and Integration Tool" Available: http://www.altova.com/mapforce.html, Jan, 05,2015
[7] TIBCOMapper,https://docs.tibco.com/pub/businesseventsts_process_orchestration/1.1.2/doc/pdf/tib_be_process_developers_guide.pdf, Oct 18,2014
[8] David Handlos, "BizTalk Mapper 2004 V/s Map Force 2004", http://www.osnews.com/story/6809, Jan, 10,2015.
[9] "XML Editor, XML Tools and XQuery – Stylus Studio " http://www.stylusstudio.com,,, Jan, 30, 2015
[10] "Transforming Data Using XQuery Mapper", http://docs.oracle.com/cd/E13171_01/alsb/docs21/dtguide/dtguideMapper.html, Jan, 31, 2015

Author Profile

Rahil Khera is pursuing a Master of Engineering in Computer Engineering from Computer Dept. of Pune Institute of Computer Technology, under the affiliation of Savitribai Phule, Pune University. He received his B.E degree in Computer Engineering in 2010.

Prof. P.S. Game is an Assistant Professor in the computer department of Pune Institute of Computer Technology.