ABSTRACT

To access a unified way to different information sources while hiding the user autonomy, heterogeneity, distribution and evolution of these sources, the authors thought to integrate its different sources of information, one of the existing integration approaches is mediation or virtual approach. The integration in mediation approach is done with a schema, called schema mediation, and a set of mapping (links) associating schemas of sources to integrate with the global schema. The problem arises in creating the set of correspondences between the elements of the schema global and all the elements of local schemas, in this paper the authors present their approach FMAMS: Fuzzy Mapping Approach for Mediation Systems, a new approach based on fundamental principles of the theory of fuzzy sets. The authors’ approach is to define and associate to each link between two elements a weight that reflects the degree of its existence, it is present in an analysis using two components: syntactic and semantic.

Keywords: Data integration, Fuzzy logic, Global As View (GAV), Mapping, Similarity

1. INTRODUCTION

Considering the diversity of applications and information sources commonly used, together with user requirements that vary often, several problems have been encountered with the availability of sources with uniform access.

For a better use of these information sources, systems and applications must be able to retrieve, integrate and share the content of these sources while hiding the autonomy, heterogeneity, distribution and evolution of these sources.

The integration of information sources provides a conceptual, architecturally, functional and technology solution to hide the peculiarities of each source. It provides to users an interface that allows access in a unified manner to different sources, disregarding their specific (P. Brien, 2003). In this context, many types of integration exist: physical integration based on data warehouse, virtual integration based on mediation, service's integration based on SOA, etc.

In this section, the mediation is distinguish among integration solutions. It consist to create
an interface between users and the set of sources interviewed while hiding their particularities. Thus, it can give the user the illusion that he interrogates a single information source (Rousset, 2002; Widom, 1996).

The integration in the virtual approach (also known as mediation) is done with a schema, called schema mediation, and a set of mappings associating data sources to schema (Wiederhold, 1992; Widom, 1996).

The data are stored only in the sources, the mediator contains only abstract views on these sources, it can offer the user a unified global view of the information sources system.

Each mediator has a global schema, called integration, mediation or virtual schema that describes conceptually the content of sources to integrate. In the opposite, each source has a local schema that describes conceptually its content. The user emit his request (query) to the mediation system, in cons, the query must be executed in sources. Than we must have the correspondance between the global schema and the source’s schemas in order to rewrite the user query into sub-query, each sub-query will be formulated in term of a local schema of one source that contains a part of the final result. These correspondances represent the link, relationship or mappings associate the global schema with sources schemas to integrate (C.F. Figure 1).

In the mediation field, two main approaches of mapping have been proposed, Global As View (GAV) and Local As View (LAV). These two approaches are defined as a set of links between all the attributes of the mediator schema and all attributes of local’s schemas of sources.

To deal the heterogeneity and solve problems related to source’s interoperability requirement, it is necessary to adapt and to link data sources in different formats. This operation is done by two main tasks: matching and mapping.

The matching is the operation that takes two input schemas and produces a relationship between all the attributes of this schema, one by one (Rahm & Bernstein, 2001). Contrary to the matching, the mapping specifies how the elements of two schemas are connected, it is defined as a set of expression that describes how the data in a specific format is related with another data (Rahm & Bernstein, 2001). In other words, the matching expresses the WHAT, and the mapping expresses the HOW (Shvaiko, 2005). In this paper, we present a new approach called fuzzy mapping approach for mediation systems, which is based on the theory of fuzzy sets.

Our paper is structured in seven parts, after the introduction, the second section is devoted to definitions and application domains, the classic mapping is introduced in the third section, with mapping types given in the fourth section, the fifth section details the problematic, our contribution is the subject of the sixth section, followed by the experimental results in seventh section. Finally, the article ends with a conclusion and future plans.

2. APPLICATION DOMAINS

The mapping is used in several application areas, we cite the following domains:

Figure 1. Mediation architecture