A Scalable Relational Database Approach for WebService Matchmaking

Deepak Chenthati
Teradata(R&D)India Pvt Ltd
Hyderabad, India
chvcdeepak@gmail.com

Hrushikesha Mohanty
Department of CIS
University of Hyderabad
Hyderabad, India
hmcs_hcu@yahoo.com

Avula Damodaram
Department of CSE
JNTU Hyderabad
Hyderabad, India
damodarama@gmail.com

Abstract — Web services have a potential to enhance Business to Business collaboration by exposing compatible services to a service of higher dimension. Composition of web services essentially requires matching of their conversation protocols in addition to service requirements. For modeling and matching of conversations, the usability of formal structure like Finite State Machine (FSM) is well studied. However we are in opinion that for storing FSMs in graphs and retrieving them is not scalable for high cost in terms of memory and time. As in near future, as millions of web services will be available this scalability problem is a challenge to research communities. In order to cope up with this problem this paper proposes a relational model to store FSM models and retrieve for service composition.

Keywords - webservice, RDBMS, matchmaking

I. INTRODUCTION

A confluence of Internet and computing technologies have given rise to a new computing paradigm that host several services for use by remote users over World Wide Web(WWW). These services use SOA: Services Oriented Architecture that not only lists the services offered on web but also provides a mechanism for their interactions. Interactions among services are necessary for making of a service of higher granularity by making use of services of lower granularity. This phenomena of service composition on synthesizing services largely depends compatibility of component services and this leads to B2B: Business to Business collaboration. As in business domain collaboration is largely welcome for greater benefits to collaborating business houses as well as users, match making of services is a problem of prime interest.

Service match making studies the composability of component services and one major aspect of studying this includes matching of conversational protocols of these services. Ideally, a conversation is feasible if for a send (of a message) by a service there is a receive at another service. Then we say both the services are in compatible for the conversation. The study of this conversation compatibility in literature is termed as service matchmaking. There has been considerable volume of research on this problem. The various techniques include ontology based [2][3][4], Fuzzy Logic based[5], Rough Set based[6] and Model based matchmaking approaches.

Model based approach is found interesting for its mathematical elegance as well as understanding. Models tried for analysis, orchestration and choreography of web services include EPC [14][15], Pertinets[16][17], and FSM[7][12]. On reviewing recent works on different modeling approaches we observe that FSM is being studied actively for modeling web services. Particularly for choreography to ensure successful communication among collaboration services. In [1] communication protocol of a web service has been extracted as a FSM and each edge is annotated with a message and its sender and receiver names. In order to ensure the correctness of a protocol between two web services, there must be well-defined FSM obtained due to interconnection of two FSMs that model communication of individual web services. A search of a compatible service for a given service essentially looks for a such well-defined FSMs from the searching and searched web services. Well-defined means the correctness in sequence of message transactions between two services i.e. for each send of a message there has to be a receive and the matching of sequence of sends with the sequence of receive of messages. An implementation of the scheme needs storage of FSMs in UDDI a repository of web services.

Because of wide spread usages of Internet there has been increasing trends in offering services, business applications on web giving rise to numerous web services. Maintaining storage of FSMs as graphs is tedious and difficult to scale up. Considering rise in number of web services available on web, managing storage at UDDI and their retrieval have become a challenge that needs an urgent attention. This paper takes up the issue and proposes a relational model to store FSMs and querying a system that matches to a given service and leads to a composed service.
II. SURVEY

FSM [7][12] has been used to model and match conversation protocols. However, we are in an opinion that storing, matchmaking and retrieving FSMs using graph structure is not scalable for representing conversation protocol for thousands of web services that will soon be available on web. In order to cope up with the challenge we propose relational model to store FSMs and a query system to retrieve services that are matching to a given service.

Another work [11] deals with the problem we address here. The authors view composability of services in terms of matching their input and output parameters. Two services match when one’s output parameters match with input parameters of the other. While matching, they have used semantic similarity based in WordNet. Service input output parameters are stored in databases. Being soft in semantic matching, we may get several services matching with a given service but at different degrees. The matching result is represented in a directed graph. Where an edge connects two matching services and the label of the edge notes the degree of matching. Then the service composition problem is seen as shortest path finding problem. Other than that QoS is also considered in deciding weight of an edge. Shortest path finding algorithm is modified to operate on relational repository of service data.

Our work is distinct from [11] by taking message transactions into consideration. We argue, for work-ability of a composed service matching of their message communications is important. In-spite of input output matching, a composition will fail to match in case of their mismatch in communication. While matching we consider the sequence as well as importance of messages (defining types like compulsory and optional). Alike [11] we have used relational database to store communication details in order to make the technique scalable. Message communications of each service are recorded in databases and matching of two services is performed by joining of two tables. The join operation is relaxed to provide results of kinds of matching Full Match, Partial Match and No Match. We have also discussed possible implementation of the proposed technique. We view our technique along with the technique proposed in [11] will give precisely workable service composition. In [13] authors proposed a method in which, web services composition are computed in advance and stored in tables. For web services composition searches, they look up the pre-computed tables rather than actual web services.

Rest of the paper is organized as follows. Section 2 illustrates FSM based approach for modeling service communication and matching of compatible services for service composition. Section 3 describes RDBMS schema to store services and its communication protocol. Also this section presents a new matching algorithm and explains it with an example. Section 4 describes the way RDBMS is embedded into the currently available jUDDI architecture and how it is used for matching of services to compose a larger service. Section 5 concludes the paper.

III. SERVICE COMPOSITION AND MATCHMAKING

A service has business logic driven by business communication. In [7] we have shown how both can be modeled by FSM. Here, we are only interested in modeling business communications by FSM as shown in Figure 1. Each node indicates a state and an edge for communication. A label on edge follows a format sender#receiver#message details. Figure 1 specifies two services name Book store and Shipping and the chain of events follow from ordering of book to shipping it.

Let us consider the example of Ordering a Book process. The chain of events will be as follows:

1) Customer(c) Requests for a Book in e-Book store (bs).
2) Book store sends availability confirmation to customer if book is available else it says Unavailable to customer.
3) Customer places an order for the book.
4) Book Store responds with make payment message.
5) Customer makes payment.
6) Book Store sends the payment confirmation.
7) bs interacts with shipping service provider (s) with a request message.
8) The shipment replies to bs with availability information if shipping is available to that particular location.
9) Then bs puts a message to Deliver the book.
10) On receiving Delivered message bs chooses any one of the modes of payment (Credit Card) CCPayment and (Debit Card) DCPayment.
11) Shipper sends a payment confirmation message on receiving the payment.
12) Book Store sends Book Delivered message to customer.

The aim of this paper is to identify service partners whose message exchange sequence matches that of the requesting service. From the figure we see that the Book Store service communicates with consumer and shipper. Here shipper is the service collaborator with whom Book Store service collaborates. The states that are darken involve communication with that of shipper. Now the aim is to find the shipper service whose message flow matches with that of the Book Store service.

Figure 1 shows the AFSM representation of Ordering a Book process. Here bs#c#Avail Conf, c#bs#Order Book, bs# c#MakePayment, bs#c#PayConf and c#bs#BookDelivered and c#bs#Payment are mandatory messages bs#s #CCPayment and bs#s#DCPayment are optional messages.
A match between two FSMs is said to have occurred if and only if there exists a non-empty intersection between the two corresponding FSMs. For non-empty intersection, first we will do intersection operation on two FSMs, then we will perform emptiness test on the resulted intersected FSM. If emptiness test fails (i.e., FSM is accepting some string of messages), then it indicates that there is a path from start state to final state in that intersected FSM. So, we can observe that two FSMs are matched. Drawback of FSM based matchmaking is that storing, matching and retrieving FSMs using graph structure is not scalable when there are huge number of services that are available on the web. Hence, we look for relational repository for this purpose.

IV. A PROPOSED SCALABLE APPROACH

Relational databases have been successful in maintaining and accessing large repository of information. For this problem, we resort to the same technology to store FSMs for each service and to perform matching required for service composition.

A. Service Repository

In the literature, a way of storing web service details and its ontology’s using Relational Database was proposed [2] [3]. But it had not taken the message flows into concern. So, we have proposed an RDBMS schema for storing not only messages but also the sequence of communications. Figure 2 shows the E-R diagram of the proposed schema. From the figure, we can observe that a single BusinessEntity can provide many BusinessServices and for every BusinessServices there is a corresponding description about its ServiceMessageFlow. We have also proposed the RequestorMessageFlow table which specifies the format a customer has to send his/her request.

BusinessEntity: A given instance of the BusinessEntity is uniquely identified by a BusinessKey. Simple textual information about the BusinessEntity is given by its name and contacts like e-mail, WebURL.

BusinessServices: A given BusinessServices entity is uniquely identified by its ServiceKey. The BusinessKey attribute uniquely identifies the BusinessEntity which is the provider of the BusinessService. Simple textual information about the BusinessServices is given by its name and short service description and its Category.

ServiceMessageFlow: A given instance of ServiceMessageFlow is uniquely identified by its ServiceKey, msg and followedByMsg. This table stores the message flow of services. Type of message, optional or mandatory, is specified by its type field. snd rcv field specifies whether the message is a sent by customer or received by customer.

RequestorMessageFlow: A given instance of RequestorMessageFlow structure is uniquely identified by its msg and followedByMsg. The table has the message flow of a requested service, which can be given externally or can be generated from the ServiceMessageFlow table. Now, we can
BusinessEntity (BusinessKey, BusinessName, Phone, Fax, e-mail, URL)
Provides (BusinessKey, ServiceKey)
BusinessServices (ServiceKey, BusinessKey, ServiceName, Description, Category)
Describes (ServiceKey, msg, followedByMsg)
ServiceMessageFlow (ServiceKey, msg, followedByMsg, snd, rcv, type, CPartner)
RequestorMessageFlow (msg, followedByMsg, snd, rcv, type)

TABLE I. SERVICE MESSAGE FLOW

<table>
<thead>
<tr>
<th>sk</th>
<th>msg</th>
<th>followedByMsg</th>
<th>snd_rcv</th>
<th>type</th>
<th>Cpart</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Req_Book</td>
<td>NULL</td>
<td>receive</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>1</td>
<td>Avail_Conf</td>
<td>Req_Book</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>1</td>
<td>Order_Book</td>
<td>Avail_Conf</td>
<td>receive</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>1</td>
<td>Make_Payment</td>
<td>Order_Book</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>1</td>
<td>Payment</td>
<td>Make_Payment</td>
<td>receive</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>1</td>
<td>CPartner</td>
<td>Payment</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>1</td>
<td>Req_Shipping</td>
<td>Pay_Conf</td>
<td>send</td>
<td>man</td>
<td>shipper</td>
</tr>
<tr>
<td>1</td>
<td>Ship_Avail</td>
<td>Req_Shipping</td>
<td>receive</td>
<td>man</td>
<td>shipper</td>
</tr>
<tr>
<td>1</td>
<td>Delivered</td>
<td>Ship_Avail</td>
<td>send</td>
<td>man</td>
<td>shipper</td>
</tr>
<tr>
<td>1</td>
<td>CCPayment</td>
<td>Delivered</td>
<td>send</td>
<td>opt</td>
<td>shipper</td>
</tr>
<tr>
<td>1</td>
<td>DCPartner</td>
<td>Delivered</td>
<td>send</td>
<td>opt</td>
<td>shipper</td>
</tr>
<tr>
<td>1</td>
<td>Ship_Pay_Conf</td>
<td>CCPayment</td>
<td>receive</td>
<td>opt</td>
<td>shipper</td>
</tr>
<tr>
<td>1</td>
<td>CPartner</td>
<td>Ship_Pay_Conf</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>2</td>
<td>Req_Shipping</td>
<td>NULL</td>
<td>receive</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>2</td>
<td>Ship_Avail_Req</td>
<td>Req_Shipping</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>2</td>
<td>Delivered</td>
<td>Ship_Avail</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>2</td>
<td>DCPartner</td>
<td>DCPartner</td>
<td>send</td>
<td>opt</td>
<td>cust</td>
</tr>
<tr>
<td>2</td>
<td>Ship_Pay_Conf</td>
<td>DCPartner</td>
<td>send</td>
<td>opt</td>
<td>cust</td>
</tr>
<tr>
<td>3</td>
<td>Req_Shipping</td>
<td>NULL</td>
<td>receive</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>3</td>
<td>Ship_Avail</td>
<td>Req_Shipping</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>3</td>
<td>Delivered</td>
<td>Ship_Avail</td>
<td>receive</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>3</td>
<td>DCPartner</td>
<td>Delivered</td>
<td>send</td>
<td>man</td>
<td>cust</td>
</tr>
<tr>
<td>3</td>
<td>Ship_Pay_Conf</td>
<td>DCPartner</td>
<td>send</td>
<td>opt</td>
<td>cust</td>
</tr>
</tbody>
</table>

summarize the tables in the proposed RDBMS schema as described in the Figure 3.

Considering the same example Ordering the book from section 2. Fig. 1 shows the FSM with annotations and Table I show the schema representation of the corresponding FSM for Book Store with ServiceKey (SK) as 1. We propose 6 column table, where msg is the message for communication, followedByMsg is the previous message which was received, snd rcv gives the detail if message is sent or received, there are two types of messages 'mandatory' (man) and 'optional' (op), Communication Partner (CPartner) gives the partner details with whom the service collaborates.

Let us take a look at the first row. Here, followedByMsg is NULL because there is no preceding message to Book message. And rcv value is receive because BookStore(bs) is receiving that message from customer(c) type is set to mandatory. Similarly all the remaining rows in that table are inserted. Number of rows in a table is equal to the number of messages in FSM. Messages CCPayment and DCPayment are entered into table as optional because, by default if annotated FSM state has two or more messages emerging from it, those messages are treated as optional.

B. Service Matching Algorithm and Analysis

By looking at the example given in section 3, we have observed that a single services message flow spans over multiple rows in a relational table. So, it is not possible for any type of query in RDBMS to perform match operation between two such multiple row spanned message flow sequences. So, we have proposed an algorithm for matchmaking of web services based on above defined tables. This algorithm is also implemented as a stored procedure in MySQL. After matching process is completed, a service falls into one of the three categories namely Exact match, Partial match and No match.
In case of **exact match**, all the messages of both service provider and requestor message flow sequence matches totally (or) some optional messages of requestor message flow sequence may not be present in message flow sequence of service provider. In case of **Partial match**, some messages of service provider message flow sequence don’t have corresponding messages in requestor message flow sequence. In case of **no match**, no messages of service provider and requestor match (or) some mandatory messages of requestor message flow sequence don’t have corresponding messages in service provider message flow sequence.

Let us take a look at the steps involved in the proposed matching algorithm. Line numbers at the end of each step refers to the part of the algorithm that is performing that particular step.

- RequestMessageFlow table shown in Table II is created from ServiceMessageFlow table which has the same ServiceKey as given input and whose Cpartner is same as req_Category.
- Temporary table is created for a service from ServiceMessageFlow that has matched Category with requestor’s category. (Line 7,8)
- Another temporary table is created by doing a full outer join operation on the above obtained table and RequestMessageFlow table based on their msg and followedByMsg.
- For each message (row or tuple), if for each send type of message there is receive type of message with the provider service or vice-versa then if the attributes msg, followedByMsg of both (service provider and requestor) in the above full outer join table are equal and if that is the first message the followed by messages would be null this check is performed and full is set to 1. (Lines 13-15)
- For the messages whose previous messages match different conditions are checked. (Line 19)
- If both message types are equal (or) if service provider has mandatory message corresponding to optional message of requestor then set full to 1. (Lines 20-26)
- If there is no message in requestor for the corresponding message in service provider then for
mandatory msg of provider partial is incremented else full is set to 1 (Lines 20-25)

- If there is no message in provider for a corresponding message in requestor if the type of requestor is mandatory partial is incremented and full is set to 0. (Lines 27-29) It is given partial as there is scope for negotiation. If type is optional then full is set to 1.

- If there is mismatch in messages then the partial value is incremented and full is set to 0.

- Now, if full is 1 and partial is 0 then the services are 'Exactly Matched' else if full is 0 and partial > 0 then they are 'Partially Matched' else they are 'Not Matched'. (Lines 53-59)

- Repeat the above steps for all services that are matched with requestor category. (Lines 5-60)

V. SIMULATION AND ANALYSIS

Let us again consider the same example of Ordering a Book service. The following Table 1 and Table 2 shows two tables: one is ServiceMessageFlow table containing three services (Assuming that only these three services are available and all of them belongs to same category), and other one is RequestorMessageFlow. Let us analyze the matching process of service 2 message flow with requestor message flow. Full outer join table of service 2 and RequestorMessageFlow tables is shown in Table III

<table>
<thead>
<tr>
<th>TABLE II. REQUESTOR MESSAGE FLOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>msg</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Req_Shipping</td>
</tr>
<tr>
<td>Ship_Avail</td>
</tr>
<tr>
<td>Deliver_Item</td>
</tr>
<tr>
<td>Delivered</td>
</tr>
<tr>
<td>DCPayment</td>
</tr>
<tr>
<td>Ship_Pay_Conf</td>
</tr>
</tbody>
</table>

Initially, full and partial are set to 0. First three messages (rows) of service 2 are exactly matched with that of requestor messages. line 14-18 of algorithm handles it. From the join table shown in Table III, There is a mismatch in the fourth message of service 2 and from the join table we can observe that the followedByMsg’s are different where as the messages are same. Since there is a skip in the message exchange both the services can negotiate hence lines 41-42 will set the value of full to 0 and increment the partial value. Again there is match in the fifth message hence full is set to 1 but partial value is 2 hence its a Partial Match. Few messages that are specified in the provider service may not be present at the requester’s end, and both the requester and provider can negotiate and agree upon the missing message.

In this way our matching algorithm performs matching of each service in ServiceMessageFlow table with the RequestorMessageFlow table. Similarly, according to the above given algorithm, service 3 matches exactly with requester’s one and service 2 has partial match with requester’s one. Due to the space constraint we have not represented all the cases. As we told earlier that we have implemented the above algorithm as a stored procedure in MySQL, the way of calling it is as follows: call match();

VI. A NEW FRAMEWORK

UDDI [8] uses WSDL [9] to describe interfaces to web services. Some organizations have already implemented UDDI depending on their requirements. One of them is jUDDI[10], which is developed by Apache group and it is open source. We have extended this jUDDI by means of adding Service communication message flow Database and Matching algorithm based on RDBMS as highlighted in Figure 3. As we told earlier, Service communication message flow Database stores the message exchange sequences of all services that are registered in UDDI. This database is used by our newly proposed matching algorithm.

A. jUDDI Extension

Now, the overall architecture of jUDDI can have the following components.

Publish Service: Service Provider who is interested in providing service will publish their service through this module. They publishes their business information in Service Database and service communication message flow details in Service Communication Message flow Database as shown in Figure 3

Query Service: Service Requestor who is interested in consuming service sends his/her request through this module. This request is processed by jUDDI by means of adding Service communication message flow Database and Matching algorithm based on RDBMS component. After the matched services are returned, requester will contact that particular service provider by getting information from Service Database.

Matching Algorithm based on RDBMS: This component receives request from requestor and performs matching with each service in Service Database by means of taking message flow of corresponding service from Service Communication Message flow Database.

Service Database: It contains the database schema required for storing information related to business service like Name, email, URL, etc Service Communication Message flow Database: It contains the database schema required for storing the communication message flow sequence of each service.

Three functionality’s are primarily provided by jUDDI namely a) Publish Service b) Query Service and c) Compose service. Service Providers who want to provide their service publishes their details into the Service Database and their message flow sequences into Service Communication
-message flow Database. The organization which requires a service places a query by specifying the type of service it wants. This query is handled by matching algorithm component. The organization which requires service composition sends its required message exchange sequence. This is handled by matching algorithm component and returns the resulted set of matched services by means of interacting with both Service Database and Service communication message flow Database components.

**TABLE III. FULL OUTER JOIN TABLE FOR SERVICE 2**

<table>
<thead>
<tr>
<th>Skey</th>
<th>msg</th>
<th>followedByMsg</th>
<th>snd_recv</th>
<th>type</th>
<th>msg</th>
<th>followedByMsg</th>
<th>snd_recv</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Req_Shipping</td>
<td>NULL</td>
<td>receive</td>
<td>man</td>
<td>Req_Shipping</td>
<td>Pay_Conf</td>
<td>send</td>
<td>man</td>
</tr>
<tr>
<td>2</td>
<td>Ship_Avail</td>
<td>Req_Shipping</td>
<td>receive</td>
<td>man</td>
<td>Ship_Avail</td>
<td>Req_Shipping</td>
<td>receive</td>
<td>man</td>
</tr>
<tr>
<td>2</td>
<td>Deliver_Item</td>
<td>Ship_Avail</td>
<td>receive</td>
<td>man</td>
<td>Deliver_Item</td>
<td>Ship_Avail</td>
<td>send</td>
<td>man</td>
</tr>
<tr>
<td>2</td>
<td>Delivered</td>
<td>Deliver_Item</td>
<td>send</td>
<td>man</td>
<td>Delivered</td>
<td>Deliver_Item</td>
<td>receive</td>
<td>man</td>
</tr>
<tr>
<td>2</td>
<td>DCPayment</td>
<td>Delivered</td>
<td>receive</td>
<td>man</td>
<td>DCPayment</td>
<td>Delivered</td>
<td>send</td>
<td>man</td>
</tr>
<tr>
<td>2</td>
<td>Ship_Pay_Conf</td>
<td>DCPayment</td>
<td>send</td>
<td>man</td>
<td>Ship_Pay_Conf</td>
<td>DCPayment</td>
<td>receive</td>
<td>opt</td>
</tr>
<tr>
<td>2</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>Delivered</td>
<td>send</td>
<td>opt</td>
</tr>
<tr>
<td>2</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>Ship_Pay_Conf</td>
<td>receive</td>
<td>opt</td>
</tr>
</tbody>
</table>

This is handled by matching algorithm component and returns the resulted set of matched services by means of interacting with both Service Database and Service communication message flow Database components.

**REFERENCES**


**VI. CONCLUSION**

This paper has proposed a scalable approach for matchmaking of web services for composition of services of higher granularity from given atomic services. FSM models for communication for each service is extracted and stored in a relational database. This provides a mechanism that makes the proposed approach scalable to store a large number of web services and retrieve a service as per the requirement of matching to a given service. The proposed idea is implemented in jUDDI open source software. The software is extended with the proposed algorithm and tested to establish the practicality of the proposed technique. This work also proposes usages of established technology in developing web services and making this available on Internet to potential users.


