

Selección de Servicios Web De Máxima Utilidad Usando Circuitos

Daniel Izquierdo

Universidad Simón Bolívar
Caracas, Venezuela
idaniel@ldc.usb.ve

Abstract. En las arquitecturas orientadas a servicios (SOA), los servicios concretos se describen en términos de parámetros de funcionalidad y calidad de servicio, mientras que las aplicaciones de *software* se especifican como flujos de trabajo de servicios abstractos y criterios no-funcionales que establecen la funcionalidad deseada y la calidad esperada respectivamente. En tiempo de ejecución de un flujo de trabajo, los servicios abstractos se reescriben con los servicios concretos que cumplen con las restricciones funcionales y no funcionales del flujo de trabajo. Dado que la selección de los servicios es hecha al momento y el espacio de servicios concretos puede ser muy grande, se requiere eficiencia y escalabilidad. En este trabajo se propone la infraestructura *McdSat^c* para resolver eficientemente el problema de selección de servicios en presencia de un número grande de servicios concretos. En primer lugar se utiliza la aproximación *Local As View* para describir la funcionalidad de los servicios concretos en términos de vistas de servicios abstractos, la calidad como una función de utilidad general que combina los distintos valores de calidad que describen cada servicio, y los flujos de trabajo como consultas conjuntivas sobre servicios abstractos. Basándose en esta representación, se modela el problema de selección de servicios como un problema conocido en el área de integración de sistemas conocido como reescritura de consultas usando vistas. Codificamos el problema de selección de servicios como una teoría proposicional cuyos modelos corresponden a la composición de los servicios que implementan el flujo abstracto; se explotan relaciones conocidas entre teorías d-DNNF y circuitos aritméticos para proveer una solución eficiente y escalable. Se reporta sobre el rendimiento de la aproximación propuesta y se observa que *McdSat^c* escala a instancias grandes del problema de selección de servicios.

1 Introducción

Bajo la sombra de la Web Semántica y con el apoyo de arquitecturas orientadas a servicios, el número de fuentes de datos y servicios Web ha explotado en los últimos años. Por ejemplo, actualmente la colección de bases de datos de biología molecular incluye 1.170 bases de datos [?], lo cual es 95 más que el año pasado [?] y 110 más que hace dos años [?]. Tanto las herramientas y servicios como el número de instancias publicadas por estos recursos siguen una progresión similar [?]. Gracias a este tesoro de datos, los usuarios se apoyan más en varias tareas digitales como obtención de datos de fuentes

públicas y análisis de datos con herramientas o servicios Web organizados en flujos de trabajo complejos.

Para diseñar una aplicación, los usuarios frecuentemente especifican la funcionalidad en términos de un flujo de trabajo de servicios abstractos, mientras que la calidad deseada se define restringiendo los valores de los parámetros QoS (calidad de servicio) que describen los servicios concretos disponibles. Para evaluar el flujo propuesto, se instancian los servicios abstractos con concretos de manera que se satisfagan las restricciones sobre parámetros QoS. Sin embargo, el rendimiento de estas tareas es impactado por el número de fuentes disponibles. Se requieren técnicas que permitan escalar eficientemente a un número grande de servicios.

En este trabajo se intenta proponer una solución a este problema y proveer técnicas que recorran eficientemente el espacio de servicios disponibles y seleccionen combinaciones que cumplan con los requerimientos funcionales y no-funcionales que describen una aplicación diseñada. Se llama *MCDSat^c* a la solución propuesta.

En *MCDSat^c*, las funcionalidades de servicios concretos se describen como vistas sobre servicios abstractos y los flujos de trabajo se representan como consultas conjuntivas sobre servicios abstractos; esta representación de los servicios es similar a la generada de manera semiautomática para el sistema DEIMOS [?]. Adicionalmente, las reglas que definen flujos de trabajo y servicios concretos en el catálogo de *MCDSat^c* son creadas de manera que las restricciones de entrada y salida de los servicios abstractos combinados sean satisfechas. Para representar las medidas de QoS, cada descripción de servicio concreto es anotada con un número real que representa un valor general para las medidas de QoS del servicio en particular. Finalmente, *MCDSat^c* modela el problema de selección de servicios web como el problema de reescritura de consultas usando vistas [?]. Este problema es importante en el contexto de integración de datos [?,?] y de optimización de consultas y mantenimiento de datos [?,?] y se han definido varias aproximaciones para escalar eficientemente a un número grande de vistas [?,?,?]. En este trabajo se propone la infraestructura *MCDSat^c* que extiende la aproximación propuesta en [?], con la capacidad de trabajar con constantes y de identificar la composición de servicios que maximice una función de utilidad dada.

MCDSat^c traduce una instancia del problema de selección de servicios en el problema de enumerar los modelos de una teoría proposicional que codifica información sobre el flujo de trabajo, los servicios concretos y las condiciones que deben ser satisfechas para generar una composición que maximice la función de utilidad. Esta teoría se representa como una fórmula CNF que es compilada a su representación d-DNNF, a partir de la cual los modelos que maximicen la función global de utilidad se obtienen en tiempo lineal usando técnicas modernas de SAT [?]. Se estudia empíricamente la calidad de *MCDSat^c* sobre una variedad de *benchmarks* y se observa que la técnica propuesta provee una solución eficiente al problema de selección de servicios que es capaz de escalar a flujos de trabajo grandes así como a números de vistas grandes.

This paper is comprised of six additional sections. The next section motivates our approach by using an example, and we summarize existing related work in section 3. Section 4 describes the *MCDSat^c* system architecture, and the experimental study is reported in section 5. In section 6 we present the formalization of the web service

selection problem as a Propositional Logic theory, and finally, section 7 outlines our conclusions and future work.

2 Motivación

Consideremos los servicios abstractos y concretos en la tabla 2 que representan aplicaciones en el dominio de vuelos.

Servicio Abstracto	Descripción
$flight(x1, x2)$	relaciona las ciudades $x1$ y $x2$ si hay un vuelo entre ellas
$uscity(x1)$	determina si una ciudad $x1$ es una ciudad de Estados Unidos
Servicio Concreto	Descripción
$national(x1, y1)$	retorna dos ciudades de Estados Unidos que están conectadas por vuelos directos
$oneway(x1, y1)$	retorna dos ciudades que están conectadas por vuelos directos de ida
$onestop(x1, y1)$	retorna dos ciudades que están conectadas por vuelos con una escala
$flightToCCS(x1, "CCS")$	retorna una ciudad que está conectada a Caracas por un vuelo directo
$onestopToCCS(x1, y1)$	retorna dos ciudades $x1$ y $y1$ tales que hay un vuelo con una escala desde la ciudad $x1$ hasta Caracas con
$fromWAS("WAS", y1)$	retorna una ciudad $y1$ a la cual llegan vuelos desde Washington.

Table 1. Servicios Abstractos y Concretos

Using the Local As View (LAV) approach [?], concrete services in Table 1 are semantically described as views on the abstract services $flight$ and $uscity$ as follows:

$$\begin{aligned}
national(x1, y1) &:- flight(x1, y1), uscity(x1), uscity(y1). \\
oneway(x2, y2) &:- flight(x2, y2). \\
onestop(x3, z3) &:- flight(x3, y3), flight(y3, z3). \\
flightToCCS(x2, "CCS") &:- flight(x2, "CCS"). \\
onestopToCCS(x3, y3) &:- flight(x3, y3), flight(y3, "CCS"). \\
fromWAS("WAS", y3) &:- flight("WAS", x3).
\end{aligned}$$

Now suppose a user is interested in implementing a workflow able to retrieve the one-stop round trip flights from US cities to any city in the world, such that flights can stop at any city. Consider the following conjunctive query that represents the workflow that defines this request in terms of abstract services. This workflow was defined in way the input parameters of each abstract service are bound by constants in the workflow, or by attributes projected out by previous abstract services.

$$\begin{aligned}
w(x, y1, y2, y3) &:- flight(x, y1), flight(y1, y2), flight(y2, y3), flight(y3, x), \\
&uscity(x).
\end{aligned}$$

Implementations of this workflow correspond to compositions of the concrete services where each concrete service can implement a subset of abstract services of the workflow, but each abstract service can be implemented by exactly one concrete service. The following composition of concrete services corresponds to one of the XX implementations of the workflow:

$$wi1(x, y1, "CCS", y3) :- national(x, y1), flightToCCS(y1, "CCS"), oneway("CCS", y3), national(y3, x).$$

However, note that the following compositions are not valid:

$$\begin{aligned} wi2(x, y1, "CCS", y3) &:- national(x, y1), flightToCCS(y1, "CCS"), \\ &\quad fromWAS("WAS", y3), national(y3, x). \\ wi3(x, y1, y2, y3) &:- national(x1, y1), onestop(x1, y2), oneway(y2, y3), \\ &\quad national(y3, x). \end{aligned}$$

The composition *wi2* is not valid because it maps the workflow variable *y2* into constants "CCS" and "WAS" which denote different cities. On the other hand, the composition *wi3* does not implement the workflow because the concrete service *onestop* does not receive as input or produce as output, the city where the flight stops (represented by the variable *y1*); thus, it is not possible to ensure that the city where this flight stops and the end city retrieved by the service *national*, are the same.

To avoid producing not valid workflow implementations, it is required to handling constants in a way that different constants are not mapped into each other either directly or indirectly (via transitivity). In addition, all the attributes in the workflow output or needed to join other services in the workflow, have to be produced by the selected concrete service.

In addition, suppose each concrete service is annotated with a utility function that aggregates the values of different QoS that characterize the behavior of the service. Then, since the space of valid service compositions of an abstract workflow can be very large, it is required a technique able to identify the ones that maximize the utility function without having to enumerate the whole set of valid compositions. In this paper, we propose a Propositional logic-based approach that takes advantage of the power of modern SAT solvers, to efficiently enumerate the service compositions that correspond to valid implementations of a given abstract workflow, as well as the compositions that maximize a given utility function.

3 Related Work

In this section we summarize existing approaches that provide solutions to the problems of service selection, knowledge representation and query rewriting.

Service Selection Solutions

The problem of selecting the services that implement an abstract workflow and best meet the QoS-based criteria is known as the QoS-aware service selection or composition problem, which has been shown to be NP-hard [?]. This problem is a combinatorial optimization problem and several heuristics have been proposed to find a relatively good solution in a reasonably short period of time. A distance metric-based heuristic to drive a backward search algorithm is proposed [?]; this metric induces an order of the services in a way that sink nodes are unlikely to be visited. In [?, ?, ?], services and workflows are described in terms of deterministic finite state machines that are encoded as a Description Logic theory whose models correspond to solutions of the problem; although reasoning methods for Description Logics formalisms could be exploited, scalability or performance of the proposed solution has not been reported. [?] propose a constraint-based approach that encodes the non-functional permissible values as a set of constraints whose violation needs to be minimized; to traverse the space of possibly optimal solutions, a hybrid algorithm that combines the tabu search and simulating annealing meta-heuristics is implemented; experimental results show that the proposed solution is able to scale up to a large number of services and abstract processes. In [?] the QoS-aware service composition problem is encoded as a Linear Programming problem providing a scalable solution to the problem. In [?] this problem is defined as a multi-objective optimization problem where the different QoS parameters are considered equally important and there is not an aggregated function to combine all of them; a genetic-based algorithm is proposed to identify a set of non-dominated service compositions that best meet all the QoS parameters. [?] propose a two-fold solution that uses a hybrid integer programming algorithm to find the decomposition of global QoS into local constraint, and then, selects the services that best meet the local constraints. Recently, two new planning-based approaches have been proposed [?, ?]. [?] extend the SHOP2 planning algorithm to select the trustworthy composition of services that implement a given OWL-S process model, while [?] propose a HTN planning-based solution where user preference metrics and domain regulations are used to guide the planner into the space of relevant compositions. Finally, [?] proposes a genetic-based algorithm to identify the composition of services that best meet the quality criteria for a set of QoS parameters.

Although these solutions are able to efficiently solve the optimization problem and scale up to a large number of abstract processes, none of them are tailored to semantically describe services in terms of the abstract process, or use these descriptions to identify the services that implement a given workflow and best meet user non-functional criteria.

Knowledge Compilation Languages

Knowledge compilation is the area in AI concerned with the problem of mapping logical theories into suitable fragments that make certain desired operations tractable [?]. Different compilation languages have been defined, for instance, Ordered Binary Decision Diagrams (OBDDs) [?], Negation Normal Form (NNF) [?], and Decomposable Negation Normal Form (DNNF) [?]. In this work we make use of the properties of the deterministic DNNFs (d-DNNF) [?] to provide an scalable and efficient solution to the service selection problem.

A Negation Normal Form (NNF) theory is constructed from literals using only conjunctions and disjunctions [?], and it can be represented as a directed acyclic graph in which the leaves are labeled with literals and the internal nodes are labeled with \wedge and \vee ; see Fig. 2 for an example. An NNF is said to be decomposable (DNNF) [?] if for each conjunction $\bigwedge_i \phi_i$, the set of variables in each conjunct are pairwise disjoint; i.e., $Vars(\phi_i) \cap Vars(\phi_j) = \emptyset$ for $i < j$. A DNNF supports a number of operations in polynomial time in the size of its DAG. For example, we can test whether a DNNF is satisfiable by a single bottom-up pass over its DAG in linear time. A DNNF is said to be deterministic (d-DNNF) [?] if for each disjunction $\bigvee_i \phi_i$, the disjuncts are pairwise logically contradictory; i.e., $\phi_i \Rightarrow \neg \phi_j$ for $i < j$. The NNF in Fig. 2, for example, is decomposable and deterministic. A d-DNNF supports the model counting and enumeration in polynomial time in the size of its DAG. *MCDSat*^c exploits the properties of a d-DNNF theory, to efficiently provide a solution to the problem of enumerating the service compositions that correspond to models of the theory, i.e., the service compositions that implement the abstract workflow and best meet the utility function. DECIR ALGO SOBRE EL ALGORITMO QUE PRODUCE LOS MEJORES.

Fig. 1. A decomposable and deterministic NNF.

Query Rewriting Solutions

A number of algorithms have been developed to find the rewritings of a given query; the most prominent being the bucket algorithm [?], the inverse rules algorithm [?,?], the minicon algorithm [?], and the *MCDSat* [?]. Generally, query rewriting algorithms work in two phases: first, they identify the views that rewrite at least one subgoal of the query; second they combine the selected views to produce a rewriting. The main difference between existing approaches, is the criteria used to choose the relevant views and reduce the space of non-useful rewritings.

The bucket algorithm reduces the number of possibilities just considering each subgoal in the query in isolation, and selecting the views that are able to produce at least the attributes projected by the query. Since, attributes involved in query joins are not verified, a large number of rewritings comprised of Cartesian products may be generated. The Inverse Rules algorithm constructs a set of rules that invert the view definition and establish how to compute tuples for the database relations from tuples of the views. Similarly, to the bucket algorithm it can produce a large number non-useful of rewritings.

The MiniCon algorithm overcomes the limitations of the previous algorithms by identifying only views that rewrite a set of the query goals and that can be combined with the rest of the query subgoals. The key idea is to identify the mappings between the variables in each subgoal of the query to the variables in one or more subgoals of the views, in a way that, join variables in the query are mapped to join variables in the body of a view or to the distinguished variables of the view. Mappings between variables and subgoals are represented in MiniCon Descriptions (MCD's)[?].

Finally, the *MCDSat* algorithm is able to identify the query rewritings of a query by translating the problem of rewriting into the problem of enumerating the models of a propositional theory $T(Q)$ whose models are in correspondence with the rewritings of the query. The *MCDSat* algorithm exploits the properties of d-DNNFs to efficiently compute the MCDs associated with theory $T(Q)$. The *MCDSat* algorithm has demonstrated to scale better than the MiniCon algorithm over a large number of benchmarks often showing performance improvements of several orders of magnitude. However, the McdSat algorithm was not designed for rewriting problems involving explicit constants, nor to compute the best rewritings with respect to a given utility function or cost model, and in this paper we propose a new encoding that overcomes these limitations.

4 The McdSat^c Architecture

In order to do so, the theory $T(Q)$ is transformed into a d-DNNF theory $\Delta(Q)$ with a CNF to d-DNNF compiler. The compiler's algorithm is similar to the DPLL algorithm for SAT but enhanced with advanced techniques such as clause learning and caching.

Fig. 2. The McdSat^c Architecture

5 Experimental Results

We have conducted an empirical analysis on the benefits of the techniques implemented in the McdSat^c system.

Dataset and Query Benchmark: we conducted our experiments over a benchmark that includes XX queries and a set of YY views. Queries and views are starts and chains, and they have between ZZ and fTT constants; queries have QQ sub-goals.

Evaluation Metrics: we report on runtime performance which corresponds to the *user time* produced by the *time* command of the Unix operation system.

Fig. 3. Compilation Time Benchmark I

6 Formalization of the Service Selection Problem

We consider databases of the form $D = \langle P, T \rangle$ where P is a set of predicates and $T = \{T_p\}_{p \in P}$ is a collection of tables that represents the predicates in extensional form. A conjunctive query Q over P is of the form

$$Q(\mathbf{x}) \text{ :- } p_1(\mathbf{x}_1), p_1(\mathbf{x}_2), \dots, p_m(\mathbf{x}_m),$$

where $p_i \in P$, \mathbf{x} is a vector of variables, and each \mathbf{x}_i is a vector of variables and constants. The result of Q over D , denoted as $Q(D)$, is the table with $|\mathbf{x}|$ columns that result of the projection of the relational join $\bowtie \{T_{p_i}\}_{i=1}^m$ over \mathbf{x} . The atoms in the body of Q are called the (sub)goals of Q .

A view V over D is a query over P . In the context of the service selection problem, the database D is an idealized description of the output produced abstract workflow implemented by multiple concrete services described as views. Given a database D , a query Q and a collection of views $E = \langle \{V_i\}_i, \{E_i\}_i \rangle$, we are required to find all the tuples in $Q(D)$ obtainable from the views in E . That is, we need to find all the *compositions* of the form

$$R(\mathbf{x}) \text{ :- } V_{i_1}(\mathbf{x}_1), V_{i_2}(\mathbf{x}_2), \dots, V_{i_n}(\mathbf{x}_n)$$

such that $R(E) \subseteq Q(D)$. A service selection problem (SSP) is a tuple $\langle P, Q, \{V_i\} \rangle$ where P is a set of predicates that represent abstract services, Q is a query over P and $\{V_i\}$ is a collection of views that define the concrete services in terms of abstract services. We assume *safe* problems in the sense that all variables mentioned in the head of the query (resp. in the head of each view) appear in the body of the query (resp., in the body of each view); also the input and output restrictions of the abstract services used in the query are satisfied. Further, we only deal with SSPs with no arithmetic predicates inside the query or views. A composition C is *valid* if for all databases $D = \langle P, T \rangle$ and extensions $\{E_i\}$, $R(E) \subseteq Q(D)$. A collection \mathcal{R} of valid compositions is a solution if for all databases $D = \langle P, T \rangle$ and extensions $\{E_i\}$, there is no other \mathcal{R}' such that $\mathcal{R}(E) \subset \mathcal{R}'(E) \subseteq Q(D)$. We are interested in finding a composition \mathcal{R} .

6.1 Logical Theories

In [?], we showed that a rewriting for a query Q with m goals can be obtained by enumerating the models of a logical theory $\mathcal{A} = \mathcal{A}_{rew} \cup \mathcal{A}_{mcd}^1 \cup \dots \mathcal{A}_{mcd}^m$ where \mathcal{A}_{rew} specified how to combine m independent copies of MCD theories \mathcal{A}_{mcd} that cover all goals in Q . Each \mathcal{A}_{mcd}^i is a copy of the theory \mathcal{A}_{mcd} in which each literal p is tagged as p^i . The theory \mathcal{A}_{mcd} consists of different groups of clauses that guarantees that its models are in correspondence with the MCDs, while the theory \mathcal{A}_{rew} contains additional clauses to guarantee a sound and complete composition of the MCDs. The reader is referred to [?] for a comprehensive description of the propositional theory.

Let us describe the difficulties that arise when constants are presents in a SSP with the following example:

$$Q(x, z) \text{ :- } p_1(x, y), p_2(y, z),$$

$$\begin{aligned}
V_1(x_1) &:- p_1(x_1, A), \\
V_2(x_2) &:- p_2(B, x_2), \\
V_3(x_3, y_3) &:- p_2(x_3, y_3)
\end{aligned}$$

where Q is the query and $\{V_1, V_2, V_3\}$ are the views. In this case, the only rewriting is $R(x, z) :- V_1(x), V_3(A, z)$ because the candidate $R(x, z) :- V_1(x), V_2(z)$ is not valid as it maps y into constants A and B that denote different objects.

The main problem when handling constants is to be sure that different constants are not mapped into each other either directly or indirectly (via transitivity).

6.2 Utility Functions

We assume a simple additive utility function in which each view V_i is associated with a utility measure $c(V_i)$, and the overall utility value of a composition is the sum of the utility values for the views in it. An optimal or best service composition is one with maximum utility value, and the optimal utility value of a SSP is the utility value of a best service composition. A SSP has always a well-defined optimal utility value (if there are no compositions, its utility is 0), but it may have multiple best service compositions. The service composition problem with utility consists in finding all the compositions of maximum utility value.

6.3 Extended Theories

The theory Δ_{mcd} makes use of propositions $t_{x,y}$ to denote that the variable/constant x in the query is mapped into the variable/constant y in the view, and propositions v_i to indicate that the MCD uses view V_i . We obtain a sound and complete theory for SSPs with constants if Δ_{mcd} is extended with the clauses:

- C1. (Inconsistent-1): $t_{x,A} \Rightarrow \neg t_{x,B}$,
- C2. (Inconsistent-2): $t_{A,x} \Rightarrow \neg t_{B,x}$,
- C3. (Inconsistent-3): $\neg t_{A,B}$,
- C4. (Transitivity-1): $v_i \wedge t_{A,y} \wedge t_{x,y} \wedge t_{x,z} \Rightarrow t_{A,z}$,
- C5. (Transitivity-2): $v_i \wedge t_{y,A} \wedge t_{y,x} \wedge t_{z,x} \Rightarrow t_{z,A}$.

Clauses C1–C3 prune candidate service compositions in which one constant is directly mapped into a different one, and the last two implement a restricted propagation of transitivity. Similarly, Δ_{rew} must be extended with the clauses:

- C6. (Inconsistent-4): $t_{x,A}^i \Rightarrow \neg t_{x,B}^j$.

[?] showed that for queries without negation or arithmetic comparisons, but with constants, and m goals and q variables, it is enough to consider service compositions of length at most m plus q subgoals [?].

Given a SSP problem $\langle Q, \{V_i\} \rangle$ possibly with constant symbols, we construct logical theories whose models are in correspondence with the service compositions. These theories are then compiled into d-DNNF from which all rewritings are extracted efficiently. Likewise, best service compositions can be computed by performing model enumeration twice: the first pass computes the optimal utility function, and the second filters out suboptimal service compositions. However, there is a better way, which we will describe in the next section.

6.4 Maximum-Utility Function

Darwiche & Marquis [?] show how to compute the rank $r^*(\Delta)$ of a propositional theory Δ efficiently when r is a literal ranking function and Δ is in d-DNNF. A literal ranking function ranks the models of the theory in terms of the rank of the literals l that are true in the model, and ranks the theory as the best rank of its models:

$$r(\omega) = \sum_{\omega \models l} r(l), \quad r^*(\Delta) = \max_{\omega \models \Delta} r(\omega).$$

Furthermore, not only $r^*(\Delta)$ can be computed efficiently from the d-DNNF but also the models that have such rank. The procedure for computing the rank converts the d-DNNF into a circuit in which the literals are replaced by their ranks, the ‘or’ nodes by ‘max’, and the ‘and’ nodes by ‘sum’. The evaluation of the circuit computes the rank of Δ .

We thus obtain a simple method for computing the best rewritings when the literal ranking function r_c induced by the cost model c is used; r_c is defined by $r_c(l) = c(V_i)$ if $l = v_i^t$ for some t , and $r_c(l) = 0$ otherwise.

7 Conclusions and Future Work

We have shown how the propositional theory used in McdSat can be extended to support constants, and how it can be used to compute the best rewritings with respect to an additive cost model. The cost model is simple yet expressive. The computation of the best models relies on an arithmetic circuit that is obtained from the d-DNNF of the propositional theory.