

Meeting September 11th

Describing the input, processing and output data in the algorithm

The **input data** for the algorithm is a query (Q) and a list of concrete services (C1, C2, ...). The query and concrete services are defined as follows:

$$\begin{aligned} Q(\bar{t}) &:= A_1(\bar{t}), A_2(\bar{t}), \dots, A_n(\bar{t}), Q_1, Q_2, \dots, Q_n \\ C(\bar{t}) &:= A_1(\bar{t}), A_2(\bar{t}), \dots, A_n(\bar{t}), Q_1, Q_2, \dots, Q_n \end{aligned}$$

The query and concrete services are defined as a set of abstract services (A_1, A_2, \dots, A_n) and a set of quality constraints (Q_1, Q_2, \dots, Q_n). \bar{t} is a set of input (decoration ?) or output (decoration !) variables. The first **processing step** of the algorithm creates the data structures that represent the query and concrete services. The query and concrete service data structure will be formed by its *name*, a set of head variables (V_h) and a set of abstract services (A_s).

$$\begin{aligned} Q &= \langle name, V_h, A_s \rangle \\ C &= \langle name, V_h, A_s \rangle \end{aligned}$$

For each abstract service A in the set A_s , a data structure containing its *name* and a set of variables (V) is build. Notice that the set of variables (V) can contain head variables or local variables.

$$A = \langle name, V \rangle$$

Once these structures were created, the **second step in the algorithm is to form the PCDs**. PCDs are created only for those concrete services which have abstract services that can answer that query or part of it. PCDs are defined as follows:

$$\langle S, h, \varphi, G, Def, has_opt \rangle$$

S is a concrete service. **h** are mapping between terms in the head of the service definition to terms in the right side of the service definition. **φ** are mapping between terms from the abstract composition to terms in the concrete service definition. **G** is a set of abstract services names and quality constraints covered by **S**. **Def** is a set conditions that are not covered by **S** alone. **has_opt** is a boolean flag to indicate that some abstract service in the definition of **S** has been used in **G** and has an optional parameter.

After that, all the **PCDs formed in the previous step are combined**. Once we have all the possible combinations of PCDs, **each combination is verified** to confirm if the combination is a valid rewriting of the former query or not. A valid rewriting is a list of PCDs that answers the complete query without redundancies, and without more than one mapping to the same variable with distinct values. Considering that we have the data structure for the

query (Q) and concrete services (C), the general idea of the algorithm is defined in the function *rewriting*:

```
function rewriting (Q, C)
  l := createPCDs(Q, C)
  P := combinePCD (l)
  foreach p in P
    if isRewriting(Q,p)
      R := R U {p}
  return R
```

The function *createPCDs* returns a list of PCDs **I**. **P** is a list all possible combinations of PCDs returned by the function *combinePCD*. And, for each PCD combination in **P**, the function *isRewriting* verifies the combination is a valid rewriting or not. If true, the set of rewriting **R** is updated adding the new combination. In the end, the list of rewriting **R** is returned.

The examples next show the data structure created for input data, the intermediary data generated while processing the algorithm, and the output rewriting expected.

Describing an example 1

Query:

$Q(\text{disease?}, \text{patients!}, \text{dnas!}, \text{birth!}, \text{gen!}, \text{addr!}, \text{wheatherstatistic!}) :- A1(\text{disease?}, \text{patients!}), A2(\text{patients?}, \text{dnas!}), A3(\text{patients?}, \text{birth!}, \text{gen!}, \text{addr!}), A4(\text{addr?}, \text{wheatherstatistic!})$

The data created for the query above is:

$\langle Q, [\text{disease?}, \text{patients!}, \text{dnas!}, \text{birth!}, \text{gen!}, \text{addr!}, \text{wheatherstatistic!}], [A1, A2, A3, A4] \rangle$

For each abstract service the following data is created:

$\langle A1, [\text{disease?}, \text{patients!}] \rangle$
 $\langle A2, [\text{patients?}, \text{dnas!}] \rangle$
 $\langle A3, [\text{patients?}, \text{birth!}, \text{gen!}, \text{addr!}] \rangle$
 $\langle A4, [\text{addr?}, \text{wheatherstatistic!}] \rangle$

The same procedure is done for the concrete services.

Concrete services:

$S1(d?, p!) :- A1(d?, p!)$
 $S2(d?, p!) :- A1(d?, p!)$
 $S3(p?, d!) :- A2(p?, d!)$
 $S4(p?, \text{dateofbirth!}, \text{gender!}, \text{address!}) :- A3(p?, \text{dateofbirth!}, \text{gender!}, \text{address!})$
 $S5(\text{address?}, \text{wheatherstaticalinformation!}) :- A4(\text{address?}, \text{wheatherstaticalinformation!})$
 $S6(p?, d!) :- A5(p?, d!)$
 $S7(p?, \text{vaccinateddiseases!}) :- A6(p?, \text{vaccinateddiseases!})$

Once we have all the data structure created, PCDs for abstract services that can solve the query will be created. Such as:

$\langle S1, [d? \mapsto d?, p! \mapsto p!], [d? \mapsto \text{disease?}, p! \mapsto \text{patients!}], A1, \emptyset, \text{false} \rangle$
 $\langle S2, [d? \mapsto d?, p! \mapsto p!], [d? \mapsto \text{disease?}, p! \mapsto \text{patients!}], A1, \emptyset, \text{false} \rangle$
 $\langle S3, [p? \mapsto p?, d! \mapsto d!], [p? \mapsto \text{patients?}, d! \mapsto \text{dnas!}], A2, \emptyset, \text{false} \rangle$
 $\langle S4, [p? \mapsto p?, \text{dateofbirth!} \mapsto \text{dateofbirth!}, \text{gender!} \mapsto \text{gender!}, \text{address!} \mapsto \text{address!}], [p? \mapsto \text{patients?}, \text{dateofbirth!} \mapsto \text{birth!}, \text{gender!} \mapsto \text{gen!}, \text{address!} \mapsto \text{addr!}], A3, \emptyset, \text{false} \rangle$
 $\langle S5, [\text{address?} \mapsto \text{address?}, \text{wheatherstaticalinformation!} \mapsto \text{wheatherstaticalinformation!}], [\text{address?} \mapsto \text{addr?}, \text{wheatherstaticalinformation!} \mapsto \text{wheatherstatistic!}], A4, \emptyset, \text{false} \rangle$

Based on these PCDs, all possible combination are created, and then the valid ones are returned as rewriting of the former query. The expected **rewriting results** are:

1) $Q(\text{disease}, \text{patients}, \text{dnas}, \text{birth}, \text{gen}, \text{addr}, \text{wheatherstatistic}) :-$
 $S1(\text{disease}, \text{patients}), S3(\text{patients}, \text{dnas}), S4(\text{patients}, \text{birth}, \text{gen}, \text{addr}), S5(\text{addr}, \text{wheatherstatistic})$
2) $Q(\text{disease}, \text{patients}, \text{dnas}, \text{birth}, \text{gen}, \text{addr}, \text{wheatherstatistic}) :-$
 $S2(\text{disease}, \text{patients}), S3(\text{patients}, \text{dnas}), S4(\text{patients}, \text{birth}, \text{gen}, \text{addr}), S5(\text{addr}, \text{wheatherstatistic})$

Describing an example 2

The difference between the example 1 and 2 is the concrete service S2. Now, I defined it being able to solve two parts in the query.

Query:

$Q(\text{disease?}, \text{patients!}, \text{dnas!}, \text{birth!}, \text{gen!}, \text{addr!}, \text{wheatherstatistic!}) :-$ $A1(\text{disease?}, \text{patients!}),$
 $A2(\text{patients?}, \text{dnas!}), A3(\text{patients?}, \text{birth!}, \text{gen!}, \text{addr!}), A4(\text{addr?}, \text{wheatherstatistic!})$

Concrete services:

$S1(d?, p!) :- A1(d?, p!)$
 $S2(d?, p!, dna!) :- A1(d?, p!), A2(p?, dna!)$
 $S3(p?, d!) :- A2(p?, d!)$
 $S4(p?, \text{dateofbirth!}, \text{gender!}, \text{address!}) :- A3(p?, \text{dateofbirth!}, \text{gender!}, \text{address!})$
 $S5(\text{address?}, \text{wheatherstaticalinformation!}) :- A4(\text{address?}, \text{wheatherstaticalinformation!})$
 $S6(p?, d!) :- A5(p?, d!)$
 $S7(p?, \text{vaccinateddiseases!}) :- A6(p?, \text{vaccinateddiseases!})$

Note the difference in the PCD for S2 in the following lines. Now the set **G** has indications for A1 and A2.

$\langle S1, [d? \mapsto d?, p! \mapsto p!], [d? \mapsto \text{disease?}, p! \mapsto \text{patients!}], A1, \emptyset, false \rangle$
 $\langle S2, [d? \mapsto d?, p! \mapsto p!, dna! \mapsto dna!], [d? \mapsto \text{disease?}, p! \mapsto \text{patients!}, dna! \mapsto \text{dnas!}], \{A1, A2\}, \emptyset, false \rangle$
 $\langle S3, [p? \mapsto p?, d! \mapsto d!], [p? \mapsto \text{patients?}, d! \mapsto \text{diseases!}], A2, \emptyset, false \rangle$
 $\langle S4, [p? \mapsto p?, \text{dateofbirth!} \mapsto \text{dateofbirth!}, \text{gender!} \mapsto \text{gender!}, \text{address!} \mapsto \text{address!}], [p? \mapsto \text{patients?}, \text{dateofbirth!} \mapsto \text{birth!}, \text{gender!} \mapsto \text{gen!}, \text{address!} \mapsto \text{addr!}], A3, \emptyset, false \rangle$
 $\langle S5, [\text{address?} \mapsto \text{address?}, \text{wheatherstaticalinformation!} \mapsto \text{wheatherstaticalinformation!}], [\text{address?} \mapsto \text{addr?}, \text{wheatherstaticalinformation!} \mapsto \text{wheatherstatistic!}], A4, \emptyset, false \rangle$

Based on these PCDs, the expected **rewriting results** are:

1) $Q(\text{disease}, \text{patients}, \text{dnas}, \text{birth}, \text{gen}, \text{addr}, \text{wheatherstatistic}) :-$
 $S1(\text{disease}, \text{patients}), S3(\text{patients}, \text{dnas}), S4(\text{patients}, \text{birth}, \text{gen}, \text{addr}), S5(\text{addr}, \text{wheatherstatistic})$
2) $Q(\text{disease}, \text{patients}, \text{dnas}, \text{birth}, \text{gen}, \text{addr}, \text{wheatherstatistic}) :-$
 $S2(\text{disease}, \text{patients}, \text{dnas}), S4(\text{patients}, \text{birth}, \text{gen}, \text{addr}), S5(\text{addr}, \text{wheatherstatistic})$