



# Computing Concept Covers: a Preliminary Report

Mohand-Saïd Hacid, Alain Léger, Christophe Rey, Farouk Toumani

## ► To cite this version:

Mohand-Saïd Hacid, Alain Léger, Christophe Rey, Farouk Toumani. Computing Concept Covers: a Preliminary Report. Proceedings of the 2002 International Workshop on Description Logics (DL2002), 2002, Toulouse, France. hal-02079201

**HAL Id: hal-02079201**

**<https://hal.science/hal-02079201>**

Submitted on 5 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Computing concept covers: a preliminary report

M.S. Hacid<sup>†</sup>, A. Leger<sup>‡</sup>, C. Rey<sup>§</sup> and F.Toumani<sup>§</sup>

<sup>†</sup> LISI, Université Lyon I, France (mshacid@lisi.fr)

<sup>‡</sup> France Telecom R&D (alain.leger@rd.francetelecom.com)

<sup>§</sup> LIMOS, Université Blaise Pascal, France (rey@isima.fr, ftoumani@isima.fr)

## Abstract

Motivated by a novel application of description logics in the area of electronic commerce, this paper investigates a new instance of the problem of rewriting concepts using terminologies, namely the *best covering problem*: given a concept description  $Q$  and a terminology  $\mathcal{T}$ , the problem consists in finding a rewriting of  $Q$  that uses only concept names from  $\mathcal{T}$  and contains as much as possible of *common information* with  $Q$  and as less as possible of *extra information* with respect to  $Q$ . This problem is investigated in a restricted framework of description logics with a structural subsumption algorithm.

**Keywords:** Rewriting concepts, difference of descriptions, hypergraphs.

## 1 Introduction and Motivation

In [2], a general framework for *rewriting using terminologies* is defined as follows: given a terminology  $\mathcal{T}$  (i.e., a set of concept descriptions), a concept description  $Q$  that does not contain concept names defined in  $\mathcal{T}$  and a binary relation  $\rho$  between concept descriptions, can  $Q$  be rewritten into a description  $E$ , built using (some) of the names defined in  $\mathcal{T}$ , such that  $Q\rho E$ ?

Additionally, some optimality criterion is defined in order to select the relevant rewritings. Already investigated instances of this problem are the minimal rewriting problem [2] and rewriting queries using views [3, 7]. In the former, the goal is to rewrite a concept description  $Q$  into a shorter but *equivalent* description (hence,  $\rho$  is instantiated by equivalence modulo  $\mathcal{T}$  and the size of the rewriting is used as the optimality criterion). The interest in this case is to obtain a rewriting that is shorter and better readable than the original description. In the latter, the purpose is to rewrite a query  $Q$  into a query expression that uses only a set of views  $\mathcal{V}$  and is maximally contained in  $Q$ . Rewriting queries using views plays an important role in many applications such as query optimization, data warehousing or information integration. As highlighted in [2], since views can be seen as concept definitions and queries as concepts, this problem can be regarded as another instance of the general framework in which the relation  $\rho$  is instantiated by subsumption and the optimality criterion is the inverse subsumption.

We investigate a new instance of the problem of rewriting concepts using terminologies, namely the *best covering problem*. Informally, this problem can be stated as follows: given a terminology  $\mathcal{T}$  and a concept description  $Q$ , find a description  $E$ , built using (some) of the names defined in  $\mathcal{T}$ , such that  $E$  contains as much as possible of *common information* with  $Q$  and as less as possible of *extra information* with respect to  $Q$ . We call such a rewriting  $E$  a *best cover* of  $Q$  using  $\mathcal{T}$ .

Our goal is to rewrite a description  $Q$  into the closest description expressed as a conjunction of (some) concept names in  $\mathcal{T}$  (hence,  $\rho$  is neither equivalence nor subsumption). To formally define the notion of *best cover* we need to be able to characterize the notion of “*extra information*”, i.e., the information contained in one description and not contained in the other. For that, we use a non standard operation in description logics, the *difference or subtraction* operation. Roughly spoken, the difference of two descriptions is defined in [9] as being a description containing all information which is a part of one argument but not a part of the other one.

We formally define the *best covering* problem in a restricted framework of description logics where the difference operation is always semantically unique. Then we show that, in this framework, the problem of computing the *best covers* of a concept  $Q$  using a terminology  $\mathcal{T}$  can be reduced to the problem of computing minimal transversals in hypergraphs [5]. Therefore, one can reuse results known for computing minimal transversals for solving the best covering problem. A (possible) extension of our approach to description logics where the difference is not semantically unique is briefly discussed.

The motivation behind this work comes from an application in electronic commerce. We are interested in supporting a *dynamic discovery* of electronic services (e-services). Informally speaking, an e-service can be defined as an application made available via Internet by a service provider, and accessible by clients [1]. Examples of e-services currently available range from on-line travel reservation or banking services to entire business functions of an organization. What makes such a vision attractive is that e-services are capable of intelligent interaction by being able to discover and negotiate with each other, compose themselves into more complex services, etc [1]. This work is a part of an ongoing research project called MKBEEM<sup>1</sup> intended to provide electronic marketplaces with intelligent, knowledge-based multilingual services.

The rest of this paper is organized as follows. Section 2 introduces the difference operation in description logics and some related results useful for our work. In Section 3 we formally define the best covering problem in a restricted framework of description logics with a structural subsumption algorithm. This problem is then addressed in Section 4. We conclude in Section 5.

---

<sup>1</sup>MKBEEM stands for Multilingual Knowledge Based European Electronic Marketplace (IST-1999-10589, 1st Feb. 2000 - 1st Aug. 2002).

## 2 Preliminaries

We assume that the reader is familiar with the field of description logics (e.g., see [4] for a survey). A description logic  $\mathcal{L}$  is uniquely identified by the set of constructors (e.g.,  $\sqcap, \forall, \dots$ ) it allows. A model-theoretic semantics of constructors, subsumption (respectively equivalence) of concepts ( $C \sqsubseteq D$ ) (respectively,  $C \equiv D$ ) as well as the notion of least common subsumer ( $lcs$ ) of a set of descriptions are defined in the usual way. A TBox (or a terminology)  $\mathcal{T}$  is a (finite) set of *concept definitions* ( $B = D$ , where  $B$  is a concept name and  $D$  is a concept). The semantics of TBoxes is defined as usual. In the sequel, we assume that a terminology  $\mathcal{T}$  is *acyclic*.

Recall that acyclic TBoxes can be unfolded by replacing defined names by their definitions until no more defined names occur on the right-hand sides. Therefore, the notion of  $lcs$  of a set of descriptions can be obviously extended to concepts containing defined names. In this case we write  $lcs_{\mathcal{T}}(C, D)$  to denote the least common subsumer of the concepts  $C$  and  $D$  w.r.t. a terminology  $\mathcal{T}$  (i.e., the  $lcs$  is applied to the unfolded descriptions of  $C$  and  $D$ ).

In our application, concept definitions are used to specify e-services. So, in the following, when appropriate, we use the term e-services to understand defined concepts in our application.

The formal definition of the difference operation is given below.

**Definition 1 (difference operation)** [9] Let  $C, D$  be two concept descriptions with  $C \sqsubseteq D$ . The difference  $C - D$  of  $C$  and  $D$  is defined by  $C - D := \max_{\sqsubseteq} \{B \mid B \sqcap D \equiv C\}$

The set  $\{B \mid B \sqcap D \equiv C\}$  is called the difference candidates. Please note that this definition of difference requires that the second argument subsumes the first one. However, the difference  $C - D$  between two incomparable descriptions  $C$  and  $D$  can be given by constructing the least common subsumer of  $C$  and  $D$ , that is,  $C - D := C - lcs(C, D)$ .

In a description logic  $\mathcal{L}$ , where all the descriptions in the set of difference candidates are semantically equivalent, the difference is semantically unique. However, in some description logics, the set  $C - D$  may contain descriptions which are not semantically equivalent as illustrated in the example below.

*Example 1* Let us consider the following descriptions  $C := (\forall R.P) \sqcap (\forall R.\neg P)$  and  $D := (\forall R.P') \sqcap (\forall R(\leq 4S))$ . The following two non-equivalent descriptions  $(\forall R.\neg P')$  and  $(\forall R(\geq 5S))$  are both members of the set  $C - D$ .

Teege [9] provides sufficient conditions to characterize the logics where the difference operation is always semantically unique and can be implemented in a simple syntactical way. Some basic notions and useful results of this work are introduced below.

**Definition 2 (reduced clause form and structure equivalence)** Let  $\mathcal{L}$  be a description logic.

- A **clause** in  $\mathcal{L}$  is a description  $A$  with the following property:  $A \equiv B \sqcap A' \Rightarrow B \equiv \top \vee B \equiv A$ . Every conjunction  $A_1 \sqcap \dots \sqcap A_n$  of clauses can be represented by the clause set  $\{A_1, \dots, A_n\}$ .
- A clause set  $A = \{A_1, \dots, A_n\}$  is called **reduced** if either  $n = 1$ , or no clause subsumes the conjunction of the other clauses:  $\forall 1 \leq i \leq n : A_i \not\sqsupseteq A \setminus A_i$ . The set  $A$  is then called a **reduced clause form (RCF)** of every description  $B \equiv A_1 \sqcap \dots \sqcap A_n$ .
- Let  $A = \{A_1, \dots, A_n\}$  and  $B = \{B_1, \dots, B_m\}$  be reduced clause sets in a description logic  $\mathcal{L}$ .  $A$  and  $B$  are **structure equivalent** (denoted by  $A \cong B$ ) iff:  $n = m \wedge \forall 1 \leq i \leq n \exists 1 \leq j, k \leq m : A_i \equiv B_j \wedge B_i \equiv A_k$
- If in a description logic for every description all its RCFs are structure equivalent, we say that RCFs are **structurally unique** in that logic.

The *structural difference operation*, denoted by  $\setminus_{\equiv}$ , is defined as being the set difference of clause sets where clauses are compared on the basis of the equivalence relation. The following theorem shows that in description logics with structurally unique RCFs, the difference operation can be straightforwardly calculated using the *structural difference operation*.

**Theorem 1** [9] Let  $\mathcal{L}$  be a description logic with structurally unique RCFs. Let  $A, B \in \mathcal{L}$  descriptions given by their RCFs with  $A \sqsupseteq B$ . Then the difference  $B - A$  is semantically unique and is given by the structural difference:  $B - A = B \setminus_{\equiv} A$ .

Let us now introduce the notion of *structural subsumption* as defined in [9].

**Definition 3** The subsumption relation in a description logic  $\mathcal{L}$  is said **structural** iff for any clause  $A \in \mathcal{L}$  and any description  $B = B_1 \sqcap \dots \sqcap B_m \in \mathcal{L}$  which is given by its RCF, the following holds:  $A \sqsupseteq B \Leftrightarrow \exists 1 \leq i \leq m : A \sqsupseteq B_i$  (\*)

The following theorem provides a sufficient condition for a description logic to have structurally unique RCFs.

**Theorem 2** Let  $\mathcal{L}$  be a description logics with structural subsumption. Let  $A$  and  $B$  be two reduced clause sets in  $\mathcal{L}$ . Then the following property holds true:  $A \equiv B \Rightarrow A \cong B$ .

Note that definition 3 of structural subsumption is different from the one usually used in the literature: generally,  $A$  is not necessary a clause and the condition (\*) applies on some kind of normal form (and not for all the RCFs of  $B$ ). Unfortunately, a consequence of this remark is that many description logics for which a structural subsumption algorithm exists (e.g.,  $\mathcal{ALN}$  [8]) do not have structurally unique RCFs.

Nevertheless, the result of theorem 2 is still interesting in practice since there exists many description logics with this property. Examples of such logics include

$\mathcal{FL}_0$  or the more expressive description logic, denoted by  $\mathcal{L}_1$  in [9], which contains the following constructors:

- $\sqcap, \sqcup, \top, \perp, (\exists R.C), (\exists f.C), (\geq R)$  for concepts where  $C$  denotes a concept,  $R$  a role and  $f$  a feature (i.e., a functional role),
- $\perp, \circ, |$  for roles,
- $\perp, \circ$  for features.

In the rest of this paper we use the term *structural subsumption* in the sense of definition 3.

**Size of a description** Let  $\mathcal{L}$  be a description logic with structural subsumption. We define the size  $|C|$  of an  $\mathcal{L}$ -concept description  $C$  as being the number of clauses in its RCFs<sup>2</sup>. If necessary, a more precise measure of a size of a description can be defined by also taking into account the size of each clause (e.g., by counting the number of occurrences of concept and role names in each clause). However, in this case one must use some kind of canonical form to abstain from different descriptions of equivalent clauses. Please note that, in a description logic with structurally unique RCFs it is often possible to define a canonical form which is itself an RCF [9].

### 3 A restricted framework for the best covering problem

In this section, we investigate the *best covering* problem in the framework of description logics with structural subsumption.

Let  $\mathcal{L}$  be a description logic with structural subsumption,  $\mathcal{T}$  be an  $\mathcal{L}$ -terminology, and  $Q \not\sqsubseteq \perp$  be a coherent  $\mathcal{L}$ -concept description. The set of defined names occurring in  $\mathcal{T}$ , hereafter called e-services, is denoted by  $\mathcal{S}_{\mathcal{T}} = \{S_i, i \in [1, n]\}$  with  $S_i \not\sqsubseteq \perp, \forall i \in [1, n]$ . We assume that the query  $Q$  and the e-services  $S_i, i \in [1, n]$  are given by their RCFs.

**Definition 4 (cover)** A cover of  $Q$  using  $\mathcal{T}$  is a conjunction  $E$  of some names  $S_i$  from  $\mathcal{T}$  such that:  $Q - lcs_{\mathcal{T}}(Q, E) \not\sqsubseteq Q$ .

Hence, a cover of a concept  $Q$  using  $\mathcal{T}$  is defined as being any conjunction of concept names occurring in  $\mathcal{T}$  which shares some common information with  $Q$ . Please note that a cover  $E$  of  $Q$  is always consistent with  $Q$  (i.e.,  $Q \sqcap E \not\sqsubseteq \perp$ ) since  $\mathcal{L}$  is a description logic with structurally unique RCFs<sup>3</sup> and we have  $Q \not\sqsubseteq \perp$  and  $S_i \not\sqsubseteq \perp, \forall i \in [1, n]$ .

---

<sup>2</sup>We recall that, since  $\mathcal{L}$  have structurally unique RCFs, all the RCFs of an  $\mathcal{L}$ -description are equivalent and thus have the same number of clauses.

<sup>3</sup>If the language  $\mathcal{L}$  contains the incoherent concept  $\perp$ , then  $\perp$  must be a clause, i.e., non trivial decompositions of  $\perp$  is not possible (that means we cannot have incoherent conjunction of coherent clauses), otherwise it is easy to show that  $\mathcal{L}$  does not have structurally unique RCFs.

To define the notion of *best cover*, we first need to characterize more precisely the remaining descriptions both in the input concept description  $Q$  (hereafter called the *rest*) and in its cover  $E$  (hereafter called the *miss*).

**Definition 5 (rest and miss)** Let  $Q$  be an  $\mathcal{L}$ -concept description and  $E$  a cover of  $Q$  using  $\mathcal{T}$ . The rest of  $Q$  with respect to  $E$ , written  $Rest_E(Q)$ , is defined as follows:  $Rest_E(Q) \doteq Q - lcs_{\mathcal{T}}(Q, E)$ .

The missing information of  $Q$  with respect to  $E$ , written  $Miss_E(Q)$ , is defined as follows:  $Miss_E(Q) \doteq E - lcs_{\mathcal{T}}(Q, E)$ .

Now we can define the notion of *best cover*.

**Definition 6 (best cover)** A concept description  $E$  is called a *best cover* of  $Q$  using a terminology  $\mathcal{T}$  iff:

- $E$  is a cover of  $Q$  using  $\mathcal{T}$ , and
- there doesn't exist a cover  $E'$  of  $Q$  using  $\mathcal{T}$  such that  $(|Rest_{E'}(Q)|, |Miss_{E'}(Q)|) < (|Rest_E(Q)|, |Miss_E(Q)|)$ , where  $<$  stands for the lexicographic order.

The *best covering problem*, noted  $BCOV(\mathcal{T}, Q)$ , is then the problem of computing all the best covers of  $Q$  using  $\mathcal{T}$ .

**Theorem 3 (Complexity of  $BCOV(\mathcal{T}, Q)$ )** The best covering problem is NP-hard.

The proof of this theorem easily follows from results already shown for the minimal rewriting problem [2].

## 4 Computing best covers using hypergraphs

Let us first recall some useful definitions regarding hypergraphs.

**Definition 7 (hypergraph and transversals)** [5]

A hypergraph  $\mathcal{H}$  is a pair  $(\Sigma, \Gamma)$  of a finite set  $\Sigma = \{V_1, \dots, V_n\}$  and a set  $\Gamma$  of subsets of  $\Sigma$ . The elements of  $\Sigma$  are called vertices, and the elements of  $\Gamma$  are called edges.

A set  $T \subseteq \Sigma$  is a transversal of  $\mathcal{H}$  if for each  $\varepsilon \in \Gamma$ ,  $T \cap \varepsilon \neq \emptyset$ . A transversal  $T$  is minimal if no proper subset  $T'$  of  $T$  is a transversal. The set of the minimal transversals of an hypergraph  $\mathcal{H}$  is noted  $Tr(\mathcal{H})$ .

Now we can show that the best covering problem can be interpreted in the framework of hypergraphs as the problem of finding the minimal transversals with a minimal cost. In the sequel, a sketch of proof is given for each lemma.

**Definition 8 (hypergraph  $\mathcal{H}_{\mathcal{T}Q}$  generated from  $\mathcal{T}$  and  $Q$ )** Let  $\mathcal{L}$  be a description logic with structural subsumption,  $\mathcal{T}$  be an  $\mathcal{L}$ -terminology, and  $Q$  be an  $\mathcal{L}$ -concept description. Given an instance  $BCOV(\mathcal{T}, Q)$  of the best covering problem, we build an hypergraph  $\mathcal{H}_{\mathcal{T}Q} = (\Sigma, \Gamma)$  as follows:

- each e-service  $S_i$  in  $\mathcal{T}$  becomes a vertex  $V_{S_i}$  in the hypergraph  $\mathcal{H}_{\mathcal{T}Q}$ . Thus  $\Sigma = \{V_{S_i}, i \in [1, n]\}$ .
- each clause  $A_i \in Q$ , for  $i \in [1, k]$ , becomes an edge in  $\mathcal{H}_{\mathcal{T}Q}$ , noted  $w_{A_i}$ , with  $w_{A_i} = \{V_{S_i} \mid S_i \in \mathcal{S}_{\mathcal{T}} \text{ and } A_i \in_{\equiv} lcs_{\mathcal{T}}(Q, S_i)\}$  where  $\in_{\equiv}$  stands for the membership test modulo equivalence of clauses and  $lcs_{\mathcal{T}}(Q, S_i)$  is given by its RCF.

For the sake of clarity we introduce the following notation.

**Notation** For any set of vertices  $X = \{V_{S_i}\}$ , subset of  $\Sigma$ , we note  $E_X \doteq \bigcap_{V_{S_i} \in X} S_i$  the concept obtained from the conjunction of the e-services corresponding to the vertices in  $X$ . Mutually, for any concept  $E \doteq \bigcap_{j \in [1, m]} S_{i_j}$ , we note  $X_E = \{V_{S_{i_j}}, j \in [1, m]\}$  the set of vertices corresponding to the e-services in  $E$ .

With lemmas 1 and 2 given below, we show that computing a cover of  $Q$  using  $\mathcal{T}$  that minimizes *rest* amounts to computing a transversal of  $\mathcal{H}_{\mathcal{T}Q}$  by considering only the non empty edges.

**Lemma 1 (characterization of the minimal rest)** Let  $\mathcal{L}$  be a description logic with structural subsumption,  $\mathcal{T}$  be an  $\mathcal{L}$ -terminology, and  $Q$  be an  $\mathcal{L}$ -concept description. Let  $\mathcal{H}_{\mathcal{T}Q} = (\Sigma, \Gamma)$  be the hypergraph built from the ontology of e-services  $\mathcal{T}$  and the concept  $Q = A_1 \sqcap \dots \sqcap A_k$  provided by its RCF. Whatever the cover  $E$  of  $Q$  using  $\mathcal{T}$  we consider, the minimal rest (i.e., the rest which size is minimal) is:  $Rest_{min} \equiv A_{j_1} \sqcap \dots \sqcap A_{j_l}, \forall j_i \in [1, k] \mid w_{A_{j_i}} = \emptyset$ .

*Proof* (sketch)

First, to prove the existence of a cover  $E$  of  $Q$  using  $\mathcal{T}$  having such a rest, it is sufficient to consider  $E$  as being the combination of all the e-services in  $\mathcal{T}$ , i.e.,  $E \doteq S_1 \sqcap \dots \sqcap S_n$ .

Second, we show that  $Rest_{min}$  has the minimal size. We recall that, for any cover  $E$ , we have  $Rest_E(Q) := Q \setminus_{\equiv} lcs_{\mathcal{T}}(Q, E)$ . Assume that  $Q$  and  $lcs_{\mathcal{T}}(Q, S_i), \forall i \in [1, n]$ , are given by their RCFs. We have  $A_{j_i} \in_{\equiv} Q$  and  $A_{j_i} \notin_{\equiv} lcs_{\mathcal{T}}(Q, S_i)$  for all  $j_i \in [1, k]$  such that  $w_{A_{j_i}} = \emptyset$  (by construction of  $\mathcal{H}_{\mathcal{T}Q}$ ). Then we can prove that for all  $j_i \in [1, k]$  such that  $w_{A_{j_i}} = \emptyset$  we have  $A_{j_i} \notin_{\equiv} lcs_{\mathcal{T}}(Q, E)$  (since  $E$  is a conjunction of some e-services  $S_{i_j}$  and  $\mathcal{L}$  is a description logic with structural subsumption). This implies that for all  $j_i \in [1, k]$  such that  $w_{A_{j_i}} = \emptyset$  we have  $A_{j_i} \in Q \setminus_{\equiv} lcs_{\mathcal{T}}(Q, E)$  and thus  $|Rest_{min}| \leq |Rest_E(Q)|$  for any cover  $E$  of  $Q$ .  $\square$

**Lemma 2 (characterization of covers that minimize the rest)** Let  $\hat{\mathcal{H}}_{\mathcal{T}Q} = (\Sigma, \Gamma')$  be the hypergraph built by removing from  $\mathcal{H}_{\mathcal{T}Q}$  the empty edges. A rewriting  $E_{min} \doteq S_{i_1} \sqcap \dots \sqcap S_{i_m}$ , with  $1 \leq m \leq n$  and  $S_{i_j} \in \mathcal{S}_{\mathcal{T}}$  for  $1 \leq j \leq m$ , is a cover of  $Q$  using  $\mathcal{T}$  that minimizes the rest  $Rest_{E_{min}}(Q)$  iff  $X_{E_{min}} = \{V_{S_{i_j}}, j \in [1, m]\}$  is a transversal of  $\hat{\mathcal{H}}_{\mathcal{T}Q}$ .

*Proof* (sketch) The main steps of the proof are:



Lemma 1  $\Leftrightarrow \forall w_{A_i} \in \widehat{\mathcal{H}}_{\mathcal{T}Q}$ , the corresponding clause  $A_i \not\in_{\equiv} Q \setminus \equiv lcs_{\mathcal{T}}(Q, E_{min})$   
 $\Leftrightarrow \forall w_{A_i} \in \widehat{\mathcal{H}}_{\mathcal{T}Q}, A_i \in_{\equiv} lcs_{\mathcal{T}}(Q, E_{min})$   
 $\Leftrightarrow \forall w_{A_i} \in \widehat{\mathcal{H}}_{\mathcal{T}Q}, \exists S_{i_j}$  with  $j \in [1, m] \mid A_i \in_{\equiv} lcs_{\mathcal{T}}(Q, S_{i_j})$  (since  $\mathcal{L}$  is a  
 description logic with structural subsumption)  
 $\Leftrightarrow \forall w_{A_i} \in \widehat{\mathcal{H}}_{\mathcal{T}Q}, \exists V_{S_{i_j}} \in X_{E_{min}} \mid V_{S_{i_j}} \in w_{A_i}$   
 $\Leftrightarrow X_{E_{min}}$  is a transversal of  $\widehat{\mathcal{H}}_{\mathcal{T}Q}$  (since  $X_{E_{min}}$  intersects each edge of  $\widehat{\mathcal{H}}_{\mathcal{T}Q}$ )

□

Having covers that minimize the rest, it remains to isolate those which minimize the miss in order to have the best covers. To express miss minimization in the hypergraphs framework, we introduce the following notion of cost.

**Definition 9 (cost of a set of vertices)**

Let  $\mathcal{BCOV}(\mathcal{T}, Q)$  be an instance of the best covering problem and  $\widehat{\mathcal{H}}_{\mathcal{T}Q} = (\Sigma, \Gamma')$  its associated hypergraph. The cost of the set of vertices  $X$  is defined as follows:  $cost(X) = |Miss_{E_X}(Q)|$ .

Therefore, the  $\mathcal{BCOV}(\mathcal{T}, Q)$  problem can be reduced to the computation of the transversals with minimal cost of the hypergraph  $\widehat{\mathcal{H}}_{\mathcal{T}Q}$ . Clearly, it appears that we can only care about *minimal* transversals. To sum up, the  $\mathcal{BCOV}(\mathcal{T}, Q)$  problem can be reduced to the computation of the minimal transversals with minimal cost of the hypergraph  $\widehat{\mathcal{H}}_{\mathcal{T}Q}$ . Therefore, one can reuse results known for computing minimal transversals for solving the best covering problem.

## 5 Discussion

In this paper we have investigated the problem of the best covering problem in a restricted framework of description logics with structural subsumption. These logics ensure that the difference operation is always semantically unique. In this context, we have shown that the best covering problem can be reduced to the problem of computing the minimal transversals with minimum cost of a weighted hypergraph.

The problem of computing minimal transversals of an hypergraph is central in various fields of computer science [5]. The precise complexity of this problem is still an open problem. In [6], it is shown that the generation of the transversal hypergraph can be done in incremental subexponential time  $k^{O(\log k)}$ , where  $k$  is the combined size of the input and the output. To our knowledge, this is the best theoretical time bound for the problem of the generation of the transversal hypergraph.

In our case, since the problem is slightly different, we are working on an adaptation of an existing algorithm with a combinatorial optimization technic (branch-and-bound) to compute the transversals with a minimum cost.

Our future work will be devoted to the extension of the proposed framework to hold the definition of the best covering problem for description logics where

the difference operation is not semantically unique. In this case, the difference operation does not yield a unique result and thus the proposed definition of a best cover is no longer valid. However, we argue that in many practical applications (e.g., the dynamic discovery of e-services) it is sufficient to compute some kind of a single "*representative*" description of the difference candidates (i.e., only one description that is useful from the application point of view). This implies that, given a description logic  $\mathcal{L}$ , one must first identify the typical cases leading to a non unique difference, for example, by identifying the sources of structurally non-unique RCFs (e.g., nontrivial decompositions of  $\perp$ ). These cases can then be handled separately.

## References

- [1] The VLDB Journal: Special Issue on E-Services. 10(1), Springer-Verlag Berlin Heidelberg, 2001.
- [2] F. Baader, R. Küsters, and R. Molitor. Rewriting concepts using terminologies – revisited. Report 00-04, LTCS, RWTH Aachen, Germany, 2000. See <http://www-lti.informatik.rwth-aachen.de/Forschung/Reports.html>.
- [3] C. Beeri, A.Y. Levy, and M-C. Rousset. Rewriting Queries Using Views in Description Logics. In L. Yuan, editor, *Proc. of the ACM PODS*, New York, USA, pages 99–108, Apr. 1997.
- [4] F. Donini and A. Schaerf M. Lenzerini, D. Nardi. Reasoning in description logics. In *Gerhard Brewka, editor, Foundation of Knowledge Representation*, pages 191–236. CSLI-Publications, 1996.
- [5] T. Eiter and G. Gottlob. Identifying the minimal transversals of a hypergraph and related problems. *SIAM Journal on Computing*, 24(6):1278–1304, 1995.
- [6] M.L. Freidman and L. Khachiyan. On the complexity of dualization of monotone disjunctive normal forms. *Journal of Algorithms*, 21:618–628, 1996.
- [7] F. Goasdoué and M-C Rousset V. Lattès. The Use of CARIN Language and Algorithms for Information Integration: The PICSEL System. *IJICIS*, 9(4):383–401, 2000.
- [8] R. Molitor. Structural Subsumption for  $\mathcal{ALN}$ . LTCS-Report LTCS-98-03, LuFG Theoretical Computer Science, RWTH Aachen, Germany, 1998.
- [9] G. Teege. Making the difference: A subtraction operation for description logics. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Proc. of the 5<sup>th</sup> Int. Conf.KR*, San Francisco, CA, 1994. Morgan Kaufmann.