

# Representation of Partial Knowledge and Query Answering in Locally Complete Databases

*Álvaro Cortés Calabuig*

*Marc Denecker*

*Ofer Arieli*

*Maurice Bruynooghe*

*Report CW457, August 2006*



Katholieke Universiteit Leuven  
Department of Computer Science

Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

# Representation of Partial Knowledge and Query Answering in Locally Complete Databases

*Álvaro Cortés Calabuig*

*Marc Denecker*

*Ofer Arieli*

*Maurice Bruynooghe*

*Report CW457, August 2006*

Department of Computer Science, K.U.Leuven

## **Abstract**

The *Local Closed-World Assumption* (LCWA) is a generalization of Reiter's Closed-World Assumption (CWA) for relational databases that may be incomplete. Two basic questions that are related to this assumption are: (1) how to *represent* the fact that only part of the information is known to be complete, and (2) how to properly *reason* with this information, that is: how to determine whether an answer to a database query is complete even though the database information is incomplete. In this paper we concentrate on the second issue. For this we consider a fixpoint semantics for declarative theories that represent locally complete databases. This semantics is based on 3-valued interpretations that allow to distinguish between the certain and possible consequences of the database's theory.

**Keywords :** Databases, Knowledge Representation, Incomplete Knowledge.

# Representation of Partial Knowledge and Query Answering in Locally Complete Databases

Álvaro Cortés-Calabuig<sup>1</sup>, Marc Denecker<sup>1</sup>, Ofer Arieli<sup>2</sup>, and Maurice Bruynooghe<sup>1</sup>

<sup>1</sup> Department of Computer Science, Katholieke Universiteit Leuven, Belgium  
{Alvaro, Marc.Denecker, Maurice.Bruynooghe}@cs.kuleuven.be

<sup>2</sup> Department of Computer Science, The Academic College of Tel-Aviv, Israel  
oarieli@mta.ac.il

**Abstract.** The *Local Closed-World Assumption* (LCWA) is a generalization of Reiter’s Closed-World Assumption (CWA) for relational databases that may be incomplete. Two basic questions that are related to this assumption are: (1) how to *represent* the fact that only part of the information is known to be complete, and (2) how to properly *reason* with this information, that is: how to determine whether an answer to a database query is complete even though the database information is incomplete. In this paper we concentrate on the second issue based on a treatment of the first issue developed in earlier work of the authors. For this we consider a fixpoint semantics for declarative theories that represent locally complete databases. This semantics is based on 3-valued interpretations that allow to distinguish between the certain and possible consequences of the database’s theory.

## 1 Introduction

In database theory it is common to falsify any atomic fact that does not appear in the database instance. This approach follows Reiter’s Closed-World Assumption (CWA) [16], that presupposes a complete knowledge about the database’s domain of discourse.

Databases, however, are not always complete<sup>1</sup>. There are many reasons for this fact, including ignorance about the domain, lack of proper maintenance, incomplete migration, accidental deletion of tuples, the intrinsic nature of database mediator-based systems (see [10]), and so forth. Unless properly handled, partial information in database systems might lead to erroneous conclusions, as illustrated in the following example:

*Example 1.* Consider a database of a computer science (CS) department which stores information about the telephone numbers of the department’s members and collaborators. A fragment of the database is represented in Figure 1. A reasonable assumption in this case is that this database is complete with respect to all CS department members, but possibly incomplete regarding its external collaborators. Thus, appropriate answers for the queries `Telephone(Bart Delvaux,3962836)` and `Telephone(Leen Desmet,3212445)` are “no” and “unknown”, respectively.

---

<sup>1</sup> Nor they are always correct, but we do not address this problem here.

Telephone		Department	
Name	Telephone	Name	Department
Leen Desmet	6531421	Bart Delvaux	Computer Science
Bart Delvaux	5985625	Leen Desmet	Philosophy
Tom Demans	5845213	Tom Demans	Computer Science

**Fig. 1.** A database of contact phone numbers for a CS department

Example 1 illustrates a situation in which database information is *locally* complete, and so applying the CWA is not realistic, and might even lead to wrong conclusions. The other extreme approach, known as *Open-World Assumption* (OWA) [8, 9, 13], is often used for maintaining distributed knowledge, e.g. for information integration systems. In this approach a relational database is considered as a correct but possibly incomplete representation of the domain of discourse. The main weakness of the OWA is that it does not allow to express locally complete information, and so in the example above, for instance, one cannot state a full knowledge regarding the phone numbers of the CS department members.

In order to overcome the drawbacks of the CWA and the OWA in representing partial knowledge in reality, Etzioni [5] and Motro [14] introduced the notion of *Local Closed-World Assumption* (LCWA) that, intuitively, is “a specification of the areas in the real world in which a database contains all true tuples” [2]. In Example 1, for instance, such an assumption would state a meta-knowledge that the information in the Telephone relation is complete for the members of the CS department.

At the practical level, the LCWA poses some important challenges. First, a proper way of *representing* the fact that only part of the information is known to be complete is required. In the literature there are several proposals for this, using e.g. theories in a logic programming style [6] or second-order circumscriptive formulae [4]. Here we follow the first-order representation considered in [2].<sup>2</sup> Another challenge, which is addressed in this paper, is the problem of query answering in the presence of LCWA. This involves not only the query computation itself, but also a determination whether the query answer is complete even though the database information is incomplete. Our approach is based on a three-valued fixpoint semantics and corresponding algorithms for constructing a 3-valued interpretation that evaluates queries under certain and possible semantics. More specifically, the following issues are addressed:

- **Fixpoint theory for the LCWA.** A sound fixpoint operator for the LCWA is introduced, and conditions for assuring its completeness are defined. This yields a mechanism for computing a 3-valued interpretation that approximates all the 2-valued interpretations of the database’s theory and so allows informative query answering.

---

<sup>2</sup> It is shown there that this representation and its second-order derivative capture and generalize both Reiter’s CWA and the OWA.

- **Query answering algorithm.** A simple yet general algorithm for query answering under the LCWA is defined. It distinguishes between certain and possible answers and can be easily implemented by standard relational databases engines.
- **Reconciliation of paradigms.** Alternative approaches to the LCWA are considered. In particular, LCWA handling in the context of database systems [2, 14, 11] is related to LCWA formalizations for modelling logic-based agents [4, 5].

The rest of this paper is organized as follows. In Section 2 we recall some preliminary definitions and facts about the LCWA. In Section 3 we introduce a fixpoint semantics for locally complete databases, and in Section 4 we define corresponding query answering formalisms. Some related works are discussed in Section 5 and future research is sketched in Section 6.

## 2 Preliminaries

In what follows we denote by  $\Sigma$  a first-order vocabulary consisting of predicate symbols  $\mathcal{R}(\Sigma)$  (a relational schema in database terminology) and a finite set  $\mathcal{C}(\Sigma)$  of constants representing the elements of the domain of discourse. For a formula  $\Psi$  in  $\Sigma$  we denote by  $\Psi[\bar{x}]$  that the free variables of  $\Psi$  are a subset of  $\bar{x}$ . The Herbrand base of  $\Sigma$  is the set  $HB(\Sigma)$  of atomic formulas formed using  $\mathcal{C}(\Sigma)$  and the predicate symbols in  $\mathcal{R}(\Sigma)$ . A *database* is a finite set of ground atoms in  $\Sigma$ .

**Definition 1.** [2] A *local closed-world assumption* (LCWA), is an expression of the form

$$\mathcal{LCWA}(P(\bar{x}), \Psi[\bar{x}]),$$

where  $P \in \Sigma$  is a predicate symbol, called the LCWA's *object*; and  $\Psi[\bar{x}]$ , called a *window of expertise* for  $P$ , is a first-order formula over  $\Sigma$ .

*Example 2.* The expression  $\mathcal{LCWA}(\text{Tel}(x, y), \text{Dept}(x, \text{CS}))$  is a local closed-world assumption stating that the telephone numbers of all the members of the computer science department are known. That is, for all elements  $x$  of  $\{x \mid \text{Dept}(x, \text{CS})\}$  (the window of expertise for  $\text{Tel}$ ), all atoms of the form  $\text{Tel}(x, y)$  are in the database.

**Definition 2.** [2] Let  $\theta = \mathcal{LCWA}(P(\bar{x}), \Psi[\bar{x}])$  be a local closed-world assumption and  $D$  a database under vocabulary  $\Sigma$ . Denote by  $P^D$  the set of tuples corresponding to the atoms of  $P$  in  $D$ . We abbreviate the formula  $\bigvee_{\bar{a} \in P^D} (\bar{t} = \bar{a})$  by  $P(\bar{t}) \in P^D$ , where  $\bar{t}$  is a tuple of terms. The *meaning* of  $\theta$  under  $D$  is the formula

$$\mathcal{M}_D(\theta) = \forall \bar{x} \left( \Psi[\bar{x}] \supset (P(\bar{x}) \supset (P(\bar{x}) \in P^D)) \right).$$

The intuition behind this formula is simple: for all tuples  $\bar{x}$  of domain elements such that  $\Psi(\bar{x})$  holds in the *real world*, if  $P(\bar{x})$  is true (again, in *reality*), then  $\bar{x}$  must be a tuple in the table of  $P$  in  $D$ . An alternative but semantically equivalent meaning is expressed by the formula

$$\mathcal{M}_D(\theta) = \forall \bar{x} \left( \Psi[\bar{x}] \supset ((P(\bar{x}) \notin P^D) \supset \neg P(\bar{x})) \right),$$

which can be read as follows: if a tuple  $\bar{x}$  that belongs to the window of expertise  $\Psi$  is not in the database of  $P$ , then  $P(\bar{x})$  must be false.

*Example 3.* The meaning of  $\theta = \mathcal{LCWA}(\text{Tel}(x, y), \text{Dept}(x, \text{CS}))$  in the database  $D$  of Example 1 is given by

$$\begin{aligned} \mathcal{M}_D(\theta) &= \forall x \forall y (\text{Dept}(x, \text{CS}) \supset \text{Tel}(x, y) \supset \\ &\quad ((x = \text{Leen Desmet} \wedge y = 6531421) \vee (x = \text{Bart Delvaux} \wedge y = 5985625) \\ &\quad \vee (x = \text{Tom Demans} \wedge y = 5845213))) \end{aligned}$$

In some cases, we may want to express within the same expression local closed-world assumption on different predicates. In doing so, we need to extend the basic notion of an LCWA expression to allow for set of objects:

$$\theta = \mathcal{LCWA}(\{P_1(\bar{x}_1), \dots, P_n(\bar{x}_n)\}, \Psi[\bar{x}]),$$

where the  $P_i \in \Sigma$  are predicate symbols (the LCWA's objects) and  $\Psi[\bar{x}]$  is a first-order formula over  $\Sigma$  with free variables  $\bar{y}$  s.t.  $\bar{y} \subseteq \bigcup_{i=1}^n \bar{x}_i = \bar{x}$ . When an LCWA expression takes this form, the meaning of  $\theta$  under a database  $D$  is extended as follows:

$$\mathcal{M}_D(\theta) = \forall \bar{x} (\Psi[\bar{x}] \supset (\bigwedge_{i=1}^n (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D))).$$

In this paper we assume only one predicate object for each LCWA expression. As the following proposition shows, this assumption does not harm generality:

**Proposition 1.** Given a formula  $\Psi$ , denote by  $\exists|\bar{x}\Psi$  the existential quantification of all free variables in  $\Psi$ , except those in  $\bar{x}$ . Let  $\theta = \mathcal{LCWA}(\{P_1(\bar{x}_1), \dots, P_n(\bar{x}_n)\}, \Psi)$  and  $\theta_i = \mathcal{LCWA}(P_i(\bar{x}_i), \exists|\bar{x}_i\Psi)$ ,  $i = 1, \dots, n$ . Then  $\mathcal{M}_D(\theta) \equiv \bigwedge_{i=1}^n \mathcal{M}_D(\theta_i)$ .

*Proof.* The equivalence is obtained by applying some simple rewriting rules on the relevant formulae. Indeed, denote  $\bar{x} = \bigcup_{i=1}^n \bar{x}_i$  and  $\bar{z} = \bar{x} \setminus \bar{y}$ . Then:

$$\begin{aligned} \mathcal{M}_D(\theta) &\equiv \forall \bar{x} (\Psi[\bar{x}] \supset (\bigwedge_{i=1}^n (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D)))) \\ &\equiv \forall \bar{y} (\Psi(\bar{y}) \supset (\forall \bar{z} (\bigwedge_{i=1}^n (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D)))))) \\ &\equiv \forall \bar{y} (\Psi(\bar{y}) \supset (\bigwedge_{i=1}^n \forall \bar{z} (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D)))) \\ &\equiv \forall \bar{y} (\bigwedge_{i=1}^n (\Psi(\bar{y}) \supset \forall (\bar{x}_i \setminus \bar{y}) (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D)))) \\ &\equiv \bigwedge_{i=1}^n \forall \bar{y} (\Psi(\bar{y}) \supset \forall (\bar{x}_i \setminus \bar{y}) (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D))) \\ &\equiv \bigwedge_{i=1}^n \forall (\bar{y} \cap \bar{x}_i) (\exists|\bar{x}_i\Psi(\bar{y}) \supset \forall (\bar{x}_i \setminus \bar{y}) (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D))) \\ &\equiv \bigwedge_{i=1}^n \mathcal{M}_D(\theta_i). \end{aligned}$$

Thus the equivalence is obtained.  $\square$

Similarly, one may split a disjunctive window of expertise to its disjuncts and still preserve the original LCWA, and any collection of LCWAs on the same predicate may be combined to one (disjunctive) LCWA.

**Proposition 2.** Let  $\theta = \mathcal{LCWA}(P(\bar{x}), \bigvee_{i=1}^n \Psi_i[\bar{x}_i])$  and  $\theta_i = \mathcal{LCWA}(P(\bar{x}), \Psi_i[\bar{x}_i])$ ,  $i=1, \dots, n$ . Then  $\mathcal{M}_D(\theta) \equiv \bigwedge_{i=1}^n \mathcal{M}_D(\theta_i)$ .

*Proof.* Indeed,

$$\begin{aligned} \bigwedge_{j=1}^n \mathcal{M}_D(\theta_j) &\equiv \bigwedge_{j=1}^n \forall \bar{x} \left( \Psi_j[\bar{x}] \supset (P(\bar{x}) \supset (P(\bar{x}) \in P^D)) \right) \\ &\equiv \forall \bar{x} \left( \bigvee_{j=1}^n \Psi_j[\bar{x}] \supset (P(\bar{x}) \supset (P(\bar{x}) \in P^D)) \right) \equiv \mathcal{M}_D(\theta). \quad \square \end{aligned}$$

In the sequel we assume, without loss of generality, one LCWA expression per predicate of  $\mathcal{R}(\Sigma)$ . The predicates in  $\mathcal{R}(\Sigma)$  that do not appear as objects in any LCWA expression are considered as objects of LCWAs in which the windows of expertise is false. In other words, there is no context in which those predicates are complete.

The meaning of a database is now defined by the conjunction of its atoms augmented with the meaning of the given local closed-word assumptions, and the following two general assumptions:

- *Domain Closure Axiom:*  $\text{DCA}(\Sigma) = \forall x (\bigvee_{i=1}^n x = C_i)$ ,
- *Unique Name Axiom:*  $\text{UNA}(\Sigma) = \bigwedge_{1 \leq i < j \leq n} C_i \neq C_j$ ,

where  $C_1, \dots, C_n$  are the constant symbols of  $\Sigma$ .

**Definition 3.** Let  $D$  be a database and let  $\mathcal{L}$  be a set of LCWA expressions  $\theta_j = \mathcal{LCWA}(P_j, \Psi_j)$ ,  $j=1, \dots, n$  applied on  $D$ . The *meaning* of  $D$  and  $\mathcal{L}$  is given by:

$$\mathcal{M}(D, \mathcal{L}) = \bigwedge_{A \in D} A \wedge \bigwedge_{j=1}^n \mathcal{M}_D(\theta_j) \wedge \text{UNA}(\Sigma) \wedge \text{DCA}(\Sigma).$$

The theory consisting of the meaning of  $D$  and  $\mathcal{L}$  is consistent and decidable<sup>3</sup>. Consistency is shown below and decidability follows from the fact that the language does not contain function symbols and that UNA and DCA are imposed, ensuring a fixed and finite domain. Note also that each model of this theory is isomorphic to a Herbrand model.

**Proposition 3.** For every database  $D$  and every set  $\mathcal{L}$  of LCWA expressions, the meaning of  $D$  and  $\mathcal{L}$  is always consistent.

*Proof.* Let  $\theta_j = \mathcal{LCWA}(P_j, \Psi_j)$ ,  $j=1, \dots, n$  be all the local closed-world assumptions in  $\mathcal{L}$ . We construct an interpretation  $I$  for  $\Sigma$  and show that it is a model of  $\mathcal{M}(D, \mathcal{L})$ . Let  $I$  be the Herbrand interpretation associated with the database of  $D$  (taken as a 2-valued structure): the domain is  $\mathcal{C}(\Sigma)$  and  $P^I = P^D$ . By construction of  $I$ ,  $I \models P_i(d_{i_1}, \dots, d_{i_k})$  for every ground atom in  $D$ . When  $\Psi_j[\bar{x}]$  is false in  $I$ , then trivially  $I \models \mathcal{M}(D, \mathcal{L})$ . If  $\Psi_j[\bar{x}]$  is true in  $I$ , then by construction, whenever  $I \models P_j$ , also  $I \models P_j(\bar{x}) \in P^D$ .  $\square$

<sup>3</sup> There is an effective way of deciding whether any given sentence is a theorem of the theory.

The next proposition implies that for our semantics of the local closed-world assumption is a conservative extension of Reiter's closed-world assumption for relational databases; the present approach allows to express in such cases that a single database has complete knowledge about the world.

**Proposition 4.** *Let  $D$  a database and  $\mathcal{L} = \{\theta_i = \mathcal{LCWA}(P_i(\bar{x}), \mathbf{t})\}$ ,  $i = 1 \dots n$ , where  $\bigcup_{i=1}^n P_i(\bar{x}) = \mathcal{R}(\Sigma)$ . Then  $\mathcal{M}(D, \mathcal{L})$  coincides with Reiter's axiomatization of the closed-world assumption (see [16]) of  $D$ .*

*Proof.* Reiter's axiomatization of closed-world assumption of  $D$  is a first-order theory  $\Gamma$ , consisting of the following formulae: (1)  $\text{DCA}(\Sigma)$ , (2)  $\text{UNA}(\Sigma)$ , (3) the ground atomic facts in  $D$ , and (4) completion axioms for each predicate  $P_i$  of  $\Sigma$ :

$$\forall \bar{x} (P_i(\bar{x}_i) \supset P(\bar{x}_i) \in P_i^D), \quad i = 1, \dots, n.$$

By Definition 3,  $\mathcal{M}(D, \mathcal{L})$ , includes (1), (2) and (3), so it remains to show that  $\mathcal{LCWA}(\bar{P}_i, \mathbf{t})$  is equivalent to (4). Indeed, the formula that is induced from this assumption is

$$\forall \bar{x} (P_i(\bar{x}_i) \supset (P_i(\bar{x}_i) \in P_i^D)),$$

when the windows of expertise is a tautology, which is equivalent to Reiter's closure axiom.  $\square$

Our goal is to evaluate queries with respect to the meaning  $\mathcal{M}(D, \mathcal{L})$  of a database  $D$  and a set  $\mathcal{L}$  of LCWA expressions. In this context, particularly interesting queries are those formulas that are either entailed by  $\mathcal{M}(D, \mathcal{L})$  or are necessarily falsified by it. Such queries induce definitive answers. This is formalized in different ways in [2, 5, 11]. In what follows we adopt the definition of [5].

**Definition 4.** A first-order theory  $\Gamma$  determines *complete world information (CWI)* on a query  $Q[\bar{x}]$  iff for every ground tuple  $\bar{d}$ , either  $\Gamma \models Q[\bar{d}]$  or  $\Gamma \models \neg Q[\bar{d}]$  holds.

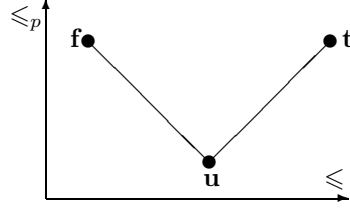
Observe that the LCWA and CWI are related concepts that capture different phenomena. The LCWA expresses completeness of a set of atoms in a relational database and the CWI identifies completeness of queries posed to the databases. Frequently, one or more LCWAs determine CWI on a query with respect to a given database. In Section 3 we consider sufficient conditions for assuring this.

As the meaning  $\mathcal{M}(D, \mathcal{L})$  of a database is a first-order formula, one way of evaluating queries is by using off-the-shelf theorem provers. However, this requires a new derivation for every ground instance of the query, which makes the whole process time consuming. An alternative approach is to generate a 3-valued Herbrand interpretation approximating all models of  $\mathcal{M}(D, \mathcal{L})$  and then to evaluate different queries with respect to this interpretation. The advantage of this approach is twofold. From a theoretical point of view, it is a good tool to distinguish the complete consequences of the theory from the incomplete ones, and in particular CWI can be easily determined. From a more practical perspective, the 3-valued Herbrand interpretation can be used to compute approximating answers to queries. In the next sections we consider this approach.

### 3 Three-valued fixpoint theory for LCWA

#### 3.1 Three-valued semantics

The truth values  $\mathcal{THRE}\mathcal{E} = \{\mathbf{t}, \mathbf{u}, \mathbf{f}\}$  of three-valued semantics are usually arranged in two orders: the truth order,  $\leq$ , which is linear order on  $\mathcal{THRE}\mathcal{E}$  such that  $\mathbf{t}$  is the maximal element,  $\mathbf{f}$  is the minimal element and  $\mathbf{u}$  is a middle element, and the precision order,  $\leq_p$ , which is a partial order on  $\mathcal{THRE}\mathcal{E}$  in which  $\mathbf{u}$  is the least element, and  $\mathbf{t}$  and  $\mathbf{f}$  are incomparable maximal elements. The structure of  $\mathcal{THRE}\mathcal{E}$  is drawn in the following diagram.



The conjunction  $\wedge$  in  $\mathcal{THRE}\mathcal{E}$  is defined by the  $\leq$ -glb of this structure, the disjunction  $\vee$  is defined by the  $\leq$ -lub, and the negation operator  $\neg$  is associated with the  $\leq$ -involution, that is:  $\neg\mathbf{t} = \mathbf{f}$ ,  $\neg\mathbf{f} = \mathbf{t}$ , and  $\neg\mathbf{u} = \mathbf{u}$ .

In this paper we focus on Herbrand interpretations. Two-valued (respectively, three-valued) (Herbrand) interpretations of  $\Sigma$  are total functions from the set  $HB(\Sigma)$  of all ground atoms of  $\Sigma$  to the set of truth values  $\{\mathbf{t}, \mathbf{f}\}$  (respectively, to the elements in  $\mathcal{THRE}\mathcal{E}$ ). Equivalently, two-valued interpretations are sometimes represented as sets of (true) atoms. Extensions to complex formulas are as usual. An interpretation  $I$  satisfies a formula  $\psi$  if  $(\psi)^I = \mathbf{t}$ ;  $I$  satisfies a set  $\Gamma$  of formulas if it satisfies every formula in  $\Gamma$ . In this case  $I$  is called a *model* of  $\Gamma$ . The orders  $\leq$  and  $\leq_p$  between truth values are extended to three-valued Herbrand interpretations by pointwise extension. For a fixed  $\Sigma$ , the set  $\mathcal{L}$  of 2-valued Herbrand  $\Sigma$ -interpretations forms a complete lattice under  $\leq$ . The set  $\mathcal{L}^c$  of 3-valued Herbrand  $\Sigma$ -interpretations is a chain complete poset<sup>4</sup> under  $\leq_p$ .

There is an interesting lattice theoretic way to construct the three-valued interpretations from  $\mathcal{L}$ . Given a lattice  $\mathcal{L}$ , we can define  $\mathcal{L}_c$  as the set of *consistent* pairs  $(x, y) \in \mathcal{L}$ , i.e., pairs such that  $x \leq y$ . On  $\mathcal{L}_c$ , we can define two orders:

1.  $(x, y) \leq (x', y')$  if  $x \leq x'$  and  $y \leq y'$
2.  $(x, y) \leq_p (x', y')$  if  $x \leq x'$  and  $y' \leq y$

In general,  $\leq$  is a lattice order and  $\leq_p$  is a chain-complete order. The following mapping from three-valued interpretations  $\mathcal{K}$  to consistent pairs  $(I, J)$  of two-valued interpretations is a one-to-one correspondence from  $\mathcal{L}^c$  to  $\mathcal{L}_c$ , preserving both  $\leq$  and  $\leq_p$ :

$$I = \{P(\bar{a}) \in HB(\Sigma) \mid P(\bar{a})^{\mathcal{K}} = \mathbf{t}\}$$

$$J = \{P(\bar{a}) \in HB(\Sigma) \mid P(\bar{a})^{\mathcal{K}} = \mathbf{t} \text{ or } P(\bar{a})^{\mathcal{K}} = \mathbf{u}\}$$

<sup>4</sup> A poset  $P$  is chain-complete if every totally ordered subset  $C \subseteq P$  has a least upper bound.

Conversely,  $\mathcal{K}$  can be constructed from  $(I, J)$  by defining for every atom  $P(\bar{a})$ :

$$P(\bar{a})^{\mathcal{K}} = \begin{cases} \mathbf{t} & \text{if } P(\bar{a})^I = \mathbf{t} \\ \mathbf{f} & \text{if } P(\bar{a})^J = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases}$$

### 3.2 Fixpoint operators for LCWAs

Below, we focus on theories  $\Gamma$  that include UNA and DCA, i.e., every model is isomorphic with a Herbrand model.

**Definition 5.** Let  $\Gamma$  be a consistent theory based on  $\Sigma$ . We say that a 3-valued Herbrand  $\Sigma$ -interpretation  $\mathcal{K}$  *approximates*  $\Gamma$  (from below) iff for every 2-valued Herbrand model  $M$  of  $\Gamma$ ,  $\mathcal{K} \leq_p M$ . The *optimal approximation* for  $\Gamma$  is a 3-valued  $\Sigma$ -interpretation, defined by  $\mathcal{K}_{opt}(\Gamma) = glb_{\leq_p}(\{M \mid M \models \Gamma\})$ , where  $M$  ranges over 2-valued Herbrand models of  $\Gamma$ .

$\mathcal{K}_{opt}(\Gamma)$  is the most precise of all 3-valued Herbrand  $\Sigma$ -interpretations approximating  $\Gamma$  and is well-defined since every nonempty set  $S \subseteq \mathcal{L}^c$  has a greatest lower bound and  $\Gamma$  is consistent. Observe that  $\mathcal{M}(D, \mathcal{L})$  satisfies the consistency condition, hence  $\mathcal{K}_{opt}(\mathcal{M}(D, \mathcal{L}))$  is well-defined.

In order to construct a 3-valued approximation for  $\mathcal{M}(D, \mathcal{L})$ , we first introduce a fixpoint operator on the chain complete poset of 3-valued Herbrand  $\Sigma$ -interpretations.

**Definition 6.** Define an operator  $\mathcal{O}_{LCWA} : \mathcal{L}^c \rightarrow \mathcal{L}^c$  as follows: for every  $\mathcal{K} \in \mathcal{L}^c$ , the interpretation  $\mathcal{K}' = \mathcal{O}_{LCWA}(\mathcal{K})$  is defined for each ground atom  $P(\bar{a})$  by:

$$P(\bar{a})^{\mathcal{K}'} = \begin{cases} \mathbf{t} & \text{if } P(\bar{a}) \in D \\ \mathbf{f} & \text{if there is } LCWA(P(\bar{x}), \Psi[\bar{x}]) \in \mathcal{L} \text{ such that} \\ & \Psi[\bar{a}]^{\mathcal{K}} = \mathbf{t} \text{ and } P(\bar{a}) \notin D \\ \mathbf{u} & \text{otherwise} \end{cases}$$

The idea is to iterate  $\mathcal{O}_{LCWA}$  starting with total ignorance (a valuation that assigns  $\mathbf{u}$  to every ground atom), and gradually extend the definite knowledge according to the database and its LCWAs.

**Proposition 5.**  $\mathcal{O}_{LCWA}$  is a  $\leq_p$ -monotone operator on the chain complete poset of 3-valued  $\Sigma$ -interpretations, thus it has a least fixpoint. Moreover, the least fixpoint can be computed in polynomial time in the size of the database.

*Proof.* Monotonicity follows from  $\leq_p$ -monotonicity of the truth assignment. By an extension of the well-known Knaster-Tarski theorem, every monotone operator in a chain complete poset has a fixpoint. Polynomial complexity follows from the fact that per application of the operator, the number of queries to be solved is polynomial in the size of the database and each query can be solved in polynomial time, while the number of iterations is at most polynomial in the size of the database.  $\square$

**Definition 7.** Denote by  $\mathcal{O}_{LCWA}^\uparrow$  the  $\leq_p$ -least fixpoint of  $\mathcal{O}_{LCWA}$ .

*Example 4.*

1. If  $D = \emptyset$  and  $\theta_1 = \mathcal{LCWA}(P, R)$  then  $R^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{u}$  and  $P^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{u}$ .
2. Suppose that  $D = \emptyset$ ,  $\theta_1 = \mathcal{LCWA}(Q, \mathbf{t})$ , and  $\theta_2 = \mathcal{LCWA}(P, \neg Q)$ . In this case  $Q^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{f}$  and  $P^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{f}$ .
3. Consider the database  $D$  of Example 1 and  $\mathcal{LCWA}(\text{Tel}(x, y), \text{Dept}(x, \text{CS}))$ , that is discussed in Example 2. Here  $\text{Tel}(\text{Bart Delvaux}, 1234567)^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{f}$ , while  $\text{Tel}(\text{Leen Desmet}, 1234567)^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{u}$ .

The following theorem shows that  $\mathcal{O}_{\mathcal{LCWA}}^\dagger$  is a sound approximation of  $\mathcal{M}(D, \mathcal{L})$ .

**Theorem 1.** (*Soundness*)  $\mathcal{O}_{\mathcal{LCWA}}^\dagger$  approximates  $\mathcal{M}(D, \mathcal{L})$ .

*Proof.* Denote by  $\perp$  the interpretation that assigns  $\mathbf{u}$  to every ground atom, and let  $M$  be a model of  $\mathcal{M}(D, \mathcal{L})$ . We show by induction that for every  $i$ ,  $\mathcal{O}_{\mathcal{LCWA}}(\perp)^i \leq_p M$ .<sup>5</sup> For  $i = 1$  this is obvious since the only atoms to which  $\mathcal{O}_{\mathcal{LCWA}}(\perp)^1$  does not assign  $\mathbf{u}$  are the database facts, and to those atoms both  $\mathcal{O}_{\mathcal{LCWA}}(\perp)^1$  and  $M$  assign  $\mathbf{t}$ . For  $i > 1$ , if  $P^{\mathcal{O}_{\mathcal{LCWA}}(\perp)^i} = \mathbf{t}$  then  $P \in D$  and so  $P^M = \mathbf{t}$  as well. If  $P(\bar{a})^{\mathcal{O}_{\mathcal{LCWA}}(\perp)^i} = \mathbf{f}$  then  $P(\bar{a}) \notin D$  and there is  $\mathcal{LCWA}(P, \Psi) \in \mathcal{L}$  s.t.  $\Psi^{\mathcal{O}_{\mathcal{LCWA}}(\perp)^{i-1}} = \mathbf{t}$ . By induction hypothesis,  $\mathcal{O}_{\mathcal{LCWA}}(\perp)^{i-1} \leq_p M$ . Since the operators in the language are  $\leq_p$ -monotonic, we have that for every formula  $\Phi$ ,  $\Phi^{\mathcal{O}_{\mathcal{LCWA}}(\perp)^{i-1}} \leq_p \Phi^M$ . In particular,  $\Psi^M = \mathbf{t}$ . By the LCWA,  $P(\bar{a})^M = \mathbf{f}$  as well. Thus, we have shown that whenever  $P(\bar{a})^{\mathcal{O}_{\mathcal{LCWA}}(\perp)^i} \in \{\mathbf{t}, \mathbf{f}\}$  necessarily  $P(\bar{a})^{\mathcal{O}_{\mathcal{LCWA}}(\perp)^i} = P(\bar{a})^M$ . It follows, therefore, that  $\mathcal{O}_{\mathcal{LCWA}}(\perp)^i \leq_p M$ .  $\square$

*Example 5.* In the following two cases  $\mathcal{O}_{\mathcal{LCWA}}^\dagger$  is strictly less than  $\mathcal{K}_{opt}(\mathcal{M}(D, \mathcal{L}))$ :

1.  $D = \emptyset$ ,  $\theta = \mathcal{LCWA}(Q, P \vee \neg P)$ .  
All the models of  $\mathcal{M}(D, \mathcal{L})$  assign  $\mathbf{f}$  to  $Q$ , thus  $Q^{\mathcal{K}_{opt}(\mathcal{M}(D, \mathcal{L}))} = \mathbf{f}$ . However, as in the standard 3-valued Kripke-Kleene semantics  $P \vee \neg P$  is unknown whenever  $P$  is unknown, we have that  $Q^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{u}$ .
2.  $D = \emptyset$ ,  $\theta_1 = \mathcal{LCWA}(P, R)$ ,  $\theta_2 = \mathcal{LCWA}(Q, R \supset \neg P)$ .  
Here again  $Q^{\mathcal{O}_{\mathcal{LCWA}}^\dagger} = \mathbf{u}$ , while  $Q^{\mathcal{K}_{opt}(\mathcal{M}(D, \mathcal{L}))} = \mathbf{f}$ .

One way to address the phenomenon in item (1) is to extend the Kripke-Kleene semantics to supervaluations [3]. Under this semantics, two-valued and three-valued tautologies/contradictions coincide. In what follows we avoid this problem by representing tautologies and contradictions only with the standard  $\mathbf{t}$  and  $\mathbf{f}$  symbols (respectively).

The difference between  $\mathcal{K}_{opt}(\mathcal{M}(D, \mathcal{L}))$  and  $\mathcal{O}_{\mathcal{LCWA}}^\dagger$  in item 2 is more subtle. In this case  $\mathcal{M}(D, \mathcal{L})$  is the formula  $(R \supset \neg P) \wedge ((R \supset \neg P) \supset \neg Q)$ , which obviously entails  $\neg Q$ . The intuitive reason for the difference is that the window of expertise in  $\theta_2$  is exactly the meaning of  $\theta_1$ , and this link is not captured by  $\mathcal{O}_{\mathcal{LCWA}}$ . To gain completeness, some restrictions need to be imposed to the windows of expertise of the LCWAs. In the following section we study such conditions.

<sup>5</sup> Where  $\mathcal{O}_{\mathcal{LCWA}}(\perp)^i$  denotes the  $i$ -th iteration of  $\mathcal{O}_{\mathcal{LCWA}}$ , starting with  $\perp$ .

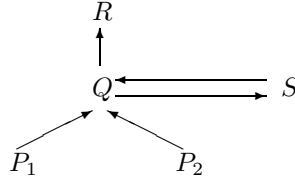
### 3.3 A hierarchy of LCWAs

**Definition 8.** A LCWA dependency graph that is determined by a set of LCWAs  $\mathcal{L} = \{\mathcal{LCWA}(P_i(\bar{x}), \Psi_i[\bar{x}])\}$ , is a directed graph whose nodes correspond to  $R(\Sigma)$ , and for each predicate  $Q$  occurring in  $\Psi_i$  there is a directed edge from  $Q$  to  $P_i$ .

*Example 6.* Consider the following set of local-closed world assumptions:

$$\mathcal{L} = \left\{ \begin{array}{lll} \mathcal{LCWA}(P_1(x), \mathbf{t}) & \mathcal{LCWA}(P_2(x), \mathbf{t}) & \mathcal{LCWA}(Q(x), P_1(x) \wedge P_2(x)) \\ \mathcal{LCWA}(Q(x), S(x)) & \mathcal{LCWA}(S(x), Q(x)) & \mathcal{LCWA}(R(x), Q(x)) \end{array} \right\}$$

The corresponding (cyclic) dependency graph is shown below:



**Definition 9.** A local closed-world assumption expression  $\mathcal{LCWA}(P(\bar{x}), \Psi[\bar{x}])$  is *primitive* iff  $\Psi[\bar{x}]$  is  $\mathbf{t}$ ,  $\mathbf{f}$ , contains only the equality predicate or any Boolean combination of those.

Primitive LCWA expressions induce CWI on the object predicates. Consider for instance  $\theta = \mathcal{LCWA}(P(x), x = a)$ . This is a primitive LCWA and it specifies that CWI on  $P$  is obtained for the domain constant  $a$ . Thus  $\mathcal{M}(D, \mathcal{L}) \models P(a)$  or  $\mathcal{M}(D, \mathcal{L}) \models \neg P(a)$ , for any database  $D$ . The following proposition formalizes this property and establishes the relationship between primitive LCWA and CWI.

**Proposition 6.** Let  $D$  be a database and let  $\theta = \mathcal{LCWA}(P(\bar{x}), \Psi) \in \mathcal{L}$  be a primitive LCWA on  $D$ . Then for any  $\bar{c}$  s.t.  $\Psi[\bar{c}]$  is true,  $\mathcal{M}(D, \mathcal{L}) \models P(\bar{c})$  iff  $P(\bar{c}) \in D$  and  $\mathcal{M}(D, \mathcal{L}) \models \neg P(\bar{c})$  iff  $P(\bar{c}) \notin D$ .

*Proof.* Let  $HU$  be the Herbrand universe of  $\Sigma$ . As  $\theta$  is a primitive LCWA, for every  $\bar{c} \in HU^n$  either  $\Psi[\bar{c}]$  holds in every Herbrand interpretation of  $\mathcal{M}(D, \mathcal{L})$  or else  $\Psi[\bar{c}]$  is falsified by every Herbrand interpretation. Assume that  $\Psi[\bar{c}]$  is true. If  $P(\bar{c}) \in D$ , then  $\mathcal{M}(D, \mathcal{L})$  contains the conjunct  $P(\bar{c})$ , hence every model of  $\mathcal{M}(D, \mathcal{L})$  satisfies  $P(\bar{c})$ . If  $P(\bar{c}) \notin D$ , then every model  $M$  of  $\mathcal{M}(D, \mathcal{L})$  satisfies:

$$\Psi[\bar{c}] \supset (P(\bar{c}) \supset P(\bar{c}) \in D)$$

It follows that  $P(\bar{c})$  is false in  $M$ . □

**Corollary 1.** Let  $D$  be a database and let  $\theta = \mathcal{LCWA}(P(\bar{x}), \Psi) \in \mathcal{L}$  be a primitive LCWA on  $D$ . For any  $\bar{c}$  such that  $\Psi[\bar{c}]$  holds,  $\mathcal{M}(D, \mathcal{L})$  determines CWI on  $P(\bar{c})$ .

*Proof.* This is an immediate consequence of Proposition 6, together with the fact that for each ground tuple  $\bar{c}$ , either  $P(\bar{c}) \in P^D$  or  $P(\bar{c}) \notin P^D$ . □ By Proposition 6 and since for every ground tuple  $\bar{c}$ , either  $P(\bar{c}) \in P^D$  or  $P(\bar{c}) \notin P^D$ . □

Another interesting relation between LCWA and CWI is the following:

**Proposition 7.** *Let  $\theta = \mathcal{LCWA}(P(\bar{x}), \Psi[\bar{x}]) \in \mathcal{L}$ . If  $\mathcal{M}(D, \mathcal{L})$  determines CWI on  $\Psi[\bar{x}]$ , then it also determines CWI on  $P(\bar{x}) \wedge \Psi[\bar{x}]$  and on  $\neg P(\bar{x}) \wedge \Psi[\bar{x}]$ .*

*Proof.* Again, denote by  $HU$  the Herbrand universe of  $\Sigma$ . By assumption, for every  $\bar{c} \in HU^n$  either  $\mathcal{M}(D, \mathcal{L}) \models \Psi[\bar{c}]$  or  $\mathcal{M}(D, \mathcal{L}) \models \neg \Psi[\bar{c}]$ . In the second case,  $\mathcal{M}(D, \mathcal{L}) \models \neg(P(\bar{c}) \wedge \Psi[\bar{c}])$  and  $\mathcal{M}(D, \mathcal{L}) \models \neg(\neg P(\bar{c}) \wedge \Psi[\bar{c}])$ . In the first case, the values of  $P(\bar{c}) \wedge \Psi[\bar{c}]$  and  $\neg P(\bar{c}) \wedge \Psi[\bar{c}]$  are determined by the question whether  $P(\bar{c})$  and  $\neg P(\bar{c})$  are true. As in the proof of Proposition 6, this reduces to whether  $P(\bar{c}) \in P^D$ .  $\square$

*Example 7.* Consider again Example 6. By the last proposition, some of the formulas to which  $\mathcal{M}(D, \mathcal{L})$  determines CWI can be inductively defined by the following stages:

1.  $P_1(x), \neg P_1(x), P_2(x), \neg P_2(x)$
2.  $P_1(x) \wedge P_2(x), \neg P_1(x) \wedge P_2(x), P_1(x) \wedge \neg P_2(x), \neg P_1(x) \wedge \neg P_2(x)$
3.  $Q \wedge P_1(x) \wedge P_2(x), \neg Q \wedge P_1(x) \wedge P_2(x)$ , and so forth.

**Definition 10.** A *hierarchically closed* database  $\mathfrak{D}$  based on vocabulary  $\Sigma$  is a tuple  $(D, \mathcal{L})$ , where  $D$  is a database and  $\mathcal{L}$  is a set of LCWAs inducing a cycle-free dependency graph.

The transitive closure of a cycle-free LCWA dependency graph is a well-founded strict order on  $\mathcal{R}(\Sigma)$ , denoted by  $<_{\mathcal{L}}$ . The minimal predicates in this order are those that are the object of a primitive LCWA (recall that every predicate is the object of exactly one LCWA in  $\mathcal{L}$ ). This property together with the definition of  $\mathcal{O}_{\mathcal{LCWA}}$  are the corner stones for the following constructive definition of an approximation of  $\mathcal{M}(D, \mathcal{L})$ .

**Definition 11.** Let  $\mathfrak{D} = (D, \mathcal{L})$  be a hierarchically closed database. The interpretation  $\mathcal{K}_{\mathcal{L}}$  that is *induced* by  $\mathfrak{D}$  is defined by induction on  $<_{\mathcal{L}}$  as follows: for each predicate  $P$  of  $\Sigma$  and every tuple  $\bar{a}$ ,

$$P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \begin{cases} \mathbf{t} & \text{if } P(\bar{a}) \in D \\ \mathbf{f} & \text{if there exists } \mathcal{LCWA}(P(\bar{x}), \Psi[\bar{x}]) \in \mathcal{L} \text{ s.t. } \Psi[\bar{a}]^{\mathcal{K}_{\mathcal{L}}} = \mathbf{t} \\ & \text{and } P(\bar{a}) \notin D \\ \mathbf{u} & \text{otherwise} \end{cases}$$

*Note 1.* In spite of their similar forms, Definitions 6 and 11 define 3-valued interpretations in different ways. In Definition 6 a given operator is iterated so that several 3-valued interpretations are constructed until a fixpoint is reached. Definition 11, on the other hand, induces a gradual construction of a *single* interpretation, starting from bottom elements of the underlying LCWA dependency graph. This construction only works for cycle-free graphs.

It is easy to see how an algorithm to compute  $\mathcal{K}_{\mathcal{L}}$  could look like: first, a primitive predicate  $P$  from  $\Sigma$  is (non-deterministically) selected. For every tuple  $\bar{a}$  of the domain, if  $P(\bar{a})$  is in the database, then  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \mathbf{t}$ . Otherwise, if the corresponding window of expertise  $\Psi_P(\bar{a})$  holds in  $\mathcal{K}_{\mathcal{L}}$ , then  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \mathbf{f}$ . Otherwise,  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}}$  is  $\mathbf{u}$ . The same steps can be repeated for any predicate  $Q$  once all the predicates on which  $Q$  depends have been evaluated.

**Proposition 8.**  $\mathcal{K}_{\mathcal{L}}$  can be computed in polynomial time in the size of the database  $D$ .

**Proposition 9.**  $\mathcal{K}_{\mathcal{L}} \equiv \mathcal{O}_{\mathcal{L}CW\mathcal{A}}^{\dagger}$  for every hierarchically closed database  $\mathfrak{D} = (D, \mathcal{L})$ .

*Proof.* Let  $rank(P) = \max(\{1 + rank(Q) \mid Q <_{\mathcal{L}} P\})$  and  $rank(Q) = 1$  iff  $Q$  is primitive. The proof is by induction on these ranks: If  $rank(P) = 1$  then as  $P$  is a primitive predicate, by the definitions of  $\mathcal{O}_{\mathcal{L}CW\mathcal{A}}^{\dagger}$  and  $\mathcal{K}_{\mathcal{L}}$ ,  $P^{\mathcal{O}_{\mathcal{L}CW\mathcal{A}}^{\dagger}} = P^{\mathcal{K}_{\mathcal{L}}}$ . If  $rank(P) = n$  for some  $n > 1$ , then  $P^{\mathcal{O}_{\mathcal{L}CW\mathcal{A}}^{\dagger}}$  is computed using the elements in  $\{Q \mid Q <_{\mathcal{L}} P\}$ . As the rank of each element  $Q$  in this set is strictly smaller than  $n$ , by induction hypothesis  $Q^{\mathcal{O}_{\mathcal{L}CW\mathcal{A}}^{\dagger}} = Q^{\mathcal{K}_{\mathcal{L}}}$ . It follows then that  $\mathcal{O}_{\mathcal{L}CW\mathcal{A}}^{\dagger}$  and  $\mathcal{K}_{\mathcal{L}}$  must assign the same truth value for  $P$ , so again  $P^{\mathcal{O}_{\mathcal{L}CW\mathcal{A}}^{\dagger}} = P^{\mathcal{K}_{\mathcal{L}}}$ .  $\square$

Given Proposition 9, the soundness of  $\mathcal{K}_{\mathcal{L}}$  is obtained as a corollary of Theorem 1.

**Corollary 2.** (*Soundness*) Let  $(D, \mathcal{L})$  be a hierarchically closed database. Then  $\mathcal{K}_{\mathcal{L}}$  approximates  $\mathcal{M}(D, \mathcal{L})$ .

The next theorem states conditions under which *optimal* approximation for  $\mathcal{M}(D, \mathcal{L})$  can be effectively constructed. Below,  $\leq_{\mathcal{L}}$  denotes the reflexive closure of  $<_{\mathcal{L}}$ .

**Theorem 2.** (*Completeness*) Let  $(D, \mathcal{L})$  be a hierarchically closed database. If for every local closed-world assumption  $\theta$  in  $\mathcal{L}$   $\Psi_{\theta}$  is a conjunction of literals, then  $\mathcal{K}_{\mathcal{L}} = \mathcal{K}_{opt}(\mathcal{M}(D, \mathcal{L}))$ .

*Proof.* Let  $S \subseteq \Sigma$ . Define:

- $\mathcal{M}_S(D, \mathcal{L})$  as the meaning of  $D$  and  $\mathcal{L}$  restricted to the predicates of  $S$  as follows:

$$\mathcal{M}_S(D, \mathcal{L}) = \bigwedge_{(Q \in S \wedge Q(\bar{a}) \in D)} Q(\bar{a}) \wedge \bigwedge_{(Q \in S)} (\forall x (\Psi_Q(\bar{x}) \supset (Q(\bar{x}) \supset Q(\bar{x}) \in D)),$$

where  $\Psi_Q$  is the windows of expertise of the unique LCWA in which  $Q$  is the object predicate.

- $<_{\mathcal{L}}^P = \{Q \mid Q <_{\mathcal{L}} P\}$ .
- $\leq_{\mathcal{L}}^P = \{Q \mid Q \leq_{\mathcal{L}} P\}$ .
- A *pre-interpretation*  $J$  of a relational first-order vocabulary  $\Sigma$  consists of a non-empty set  $U$ , called the domain of  $J$ , and an assignment  $a^J : \mathcal{C}(\Sigma) \supset U$  for each constant symbol  $a$  in  $\Sigma$ .
- A set  $S$  of predicates is *downward closed* w.r.t. the dependency order  $\leq_{\mathcal{L}}$  if  $Q \leq_{\mathcal{L}} P \in S$  implies  $Q \in S$ .
- $I|_S$  is the (2 or 3-valued) Herbrand  $\Sigma$ -interpretation  $I$  restricted to the predicates in  $S \subseteq \Sigma$ .

**Lemma 1.** *If  $S \subseteq \Sigma$  is downward closed and  $I$  is a Herbrand  $S$ -structure,  $I \models \mathcal{M}_S(D, \mathcal{L})$  iff there exists a  $\Sigma$ -model  $J$  of  $\mathcal{M}(D, \mathcal{L})$  s.t.  $I = J|_S$ .*

**Proof of lemma 1.** We prove  $(\Rightarrow)$  (the other direction is trivial). Let  $D|_{\Sigma \setminus S}$  denote the set of ground atoms in  $D$  restricted to vocabulary  $\Sigma \setminus S$ . Select  $J$  a Herbrand  $\Sigma$ -structure extending  $I$  with the tuples in  $D|_{\Sigma \setminus S}$ . We prove that  $J \models \mathcal{M}(D, \mathcal{L})$ . By construction,  $J \models \bigwedge_{A \in D} A$ . Indeed, select any atom  $A = P(\bar{a})$  from  $D$ . If  $P \in S$ , since  $J$  coincides with  $I$  in  $S$ ,  $J \models A$ . If  $P \in \Sigma \setminus S$  then, by construction of  $J$ ,  $J \models A$ . We prove now that for each  $Q \in \Sigma$ ,  $J \models \forall \bar{x} (\Psi_Q(\bar{x}) \supset (Q(\bar{x}) \supset Q(\bar{x}) \in D))$  (\*). Again, if  $Q \in S$ , since  $J$  coincides with  $I$  for the predicates in  $S$ , and  $I$  satisfies (\*),  $J \models \forall \bar{x} (\Psi_Q(\bar{x}) \supset (Q(\bar{x}) \supset Q(\bar{x}) \in D))$ . Suppose now that  $Q \in \Sigma \setminus S$ . In this case, for all tuples of domain values  $\bar{a}$ , it clearly holds that  $Q(\bar{a})^J = \mathbf{t}$  implies that  $Q(\bar{a}) \in D$ . Hence, the inner implication of (\*) is satisfied and also (\*).  $\square$

**Lemma 2.** *Let  $\mathcal{S}$  be a collection of sets. If  $S = \cup_{S' \in \mathcal{S}} S'$  and if for all  $S' \in \mathcal{S}$ ,  $\mathcal{K}|_{S'} = \text{glb}_{\leq_p}(\{I|_{S'} \mid I \models \mathcal{M}(D, \mathcal{L})\})$ , then  $\mathcal{K}|_S = \text{glb}_{\leq_p}(\{I|_S \mid I \models \mathcal{M}(D, \mathcal{L})\})$ .*

**Proof of lemma 2.** Take an arbitrary  $P \in \cup_{S \in \mathcal{S}} S$ , and a tuple  $\bar{a} \in \text{Dom}^n$ . Assume  $P \in S'$ .

$$\begin{aligned} P(\bar{a})^{\mathcal{K}|_{\cup_{S \in \mathcal{S}} S}} &= P(\bar{a})^{\mathcal{K}|_{S'}} \\ &= P(\bar{a})^{\text{glb}_{\leq_p}(\{I|_{S'} \mid I \models \mathcal{M}(D, \mathcal{L})\})} \text{ (by the hypothesis)} \\ &= P(\bar{a})^{\text{glb}_{\leq_p}(\{I \mid I \models \mathcal{M}(D, \mathcal{L})\})} \\ &= P(\bar{a})^{\text{glb}_{\leq_p}(\{I|_{\cup_{S \in \mathcal{S}} S} \mid I \models \mathcal{M}(D, \mathcal{L})\})} \end{aligned}$$

$\square$

**Lemma 3.** *Let  $\mathfrak{D} = (D, \mathcal{L})$  be a hierarchically closed database and  $\mathcal{K}_{\mathcal{L}}$  a 3-valued interpretation constructed using definition 11. For every  $P \in \Sigma$ ,*

$$\mathcal{K}_{\mathcal{L}}|_{\leq_{\mathcal{L}}^P} = \text{glb}(\{I|_{\leq_{\mathcal{L}}^P} \mid I \models \mathcal{M}(D, \mathcal{L})\}),$$

where the interpretations  $I$  are 2-valued.

**Proof of Lemma 3.** By lemma 1, it holds for each  $P$  that  $\text{glb}_{\leq_P}(\{I|_{\leq_{\mathcal{L}}^P} \mid I \models \mathcal{M}(D, \mathcal{L})\}) = \text{glb}_{\leq_P}(\{I' \mid I' \models \mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})\})$ , where  $I$  ranges over  $\Sigma$ -interpretations and  $I'$  ranges over  $\leq_{\mathcal{L}}^P$ -interpretations extending  $J$ . Therefore, it suffices to prove that  $\mathcal{K}_{\mathcal{L}}|_{\leq_{\mathcal{L}}^P} = \text{glb}_{\leq_P}(\{I|_{\leq_{\mathcal{L}}^P} \mid I \models \mathcal{M}(D, \mathcal{L})\})$ . The proof is by induction over the well-founded order of the dependency graph of  $\mathcal{L}$ .

- Base case. Let  $P$  depend only on primitive predicates. Since  $\Psi_P$  is primitive, then  $\leq_{\mathcal{L}}^P = \leq_{\mathcal{L}}^{\{P\}}$  and  $<_{\mathcal{L}}^P = \emptyset$ , and so  $\mathcal{K}|_{<_{\mathcal{L}}^P}$  is the Herbrand  $\Sigma$ -pre-interpretation  $J$ . Let  $\mathcal{M}_D(P) = \forall \bar{x} (\Psi_P(\bar{x}) \supset (P(\bar{x}) \supset (P(\bar{x}) \in P^D)))$  be the meaning of the LCWA of  $P$ . Consider an arbitrary tuple  $\bar{a}$  of domain elements. We must prove that  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = P(\bar{a})^{\text{glb}_{\leq_P}(\{I \mid I \models \mathcal{M}_{\{P\}}(D, \mathcal{L})\})}$ . It is proven in corollary 2 (soundness result) that if  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \mathbf{t}$  (or  $\mathbf{f}$ ), then  $P(\bar{a})^{\text{glb}_{\leq_P}(\{I \mid I \models \mathcal{M}_{\{P\}}(D, \mathcal{L})\})} = \mathbf{t}$  (or

f). Then, we only need to prove the property for the case in which  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \mathbf{u}$ . Consider such case. By construction of  $\mathcal{K}$ ,  $P(\bar{a}) \notin D$  and  $\Psi_P(\bar{a})^J = \mathbf{f}$ . We prove that there exist  $I_1$  and  $I_2$   $\{P\}$ -interpretations extending  $J$  s.t.  $P(\bar{a})^{I_1} = \mathbf{f}$  and  $P(\bar{a})^{I_2} = \mathbf{t}$ , and they both are models of  $\mathcal{M}_{\{P\}}(D, \mathcal{L})$ . Construct  $I_1$  as the pre-interpretation  $J$  extended with  $P^D$ . Clearly,  $I_1$  satisfies all atoms of  $P$  in the database and  $I_1 \models \mathcal{M}_D(P)$  since for all tuples  $\bar{b}$ , if  $P(\bar{b})^{I_1} = \mathbf{t}$ , then  $P(\bar{b}) \in D$ . Construct now  $I_2$  identical to  $I_1$  except that  $P(\bar{a})^{I_2} = \mathbf{t}$ .  $I_2$  still satisfies all atoms of  $P$  in  $D$ . We show that  $I_2 \models \mathcal{M}_D(P)$ . Select an arbitrary tuple  $\bar{d}$  in the domain. We need to show that  $\Psi_P(\bar{d}) \supset (P(\bar{d}) \supset (P(\bar{d}) \in P^D))$  (\*). If  $\bar{d} \neq \bar{a}$ , then the inner implication of (\*) is satisfied, just as for  $I_1$ . If  $\bar{d} = \bar{a}$  then the condition  $\Psi_P(\bar{d})$  of (\*) is false in  $J$  and hence in  $I_2$ . Since the choice of  $\bar{d}$  was arbitrary, the whole formula  $\mathcal{M}_D(P)$  is true in  $I_2$ . This concludes the proof for the base case.

– Induction step. The Induction Hypothesis (IH) is that for all  $Q <_{\mathcal{L}} P$ ,

$$\mathcal{K}_{\mathcal{L}}|_{\leq_{\mathcal{L}}^Q} = \text{glb}_{\leq_P}(\{I|_{\leq_{\mathcal{L}}^Q} \mid I \models \mathcal{M}_{\leq_{\mathcal{L}}^Q}(D, \mathcal{L})\}).$$

It follows from Lemma 2 that

$$\mathcal{K}_{\mathcal{L}}|_{<_{\mathcal{L}}^P} = \text{glb}_{\leq_P}(\{I|_{<_{\mathcal{L}}^P} \mid I \models \mathcal{M}_{\leq_{\mathcal{L}}^Q}(D, \mathcal{L})\}).$$

To prove that  $\mathcal{K}_{\mathcal{L}}|_{\leq_{\mathcal{L}}^P} = \text{glb}_{\leq_P}(\{I|_{\leq_{\mathcal{L}}^P} \mid I \models \mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})\})$  it suffices to show that for every tuple  $\bar{a}$  of domain elements:  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = P(\bar{a})^{\text{glb}_{\leq_P}(\{I|_{\leq_{\mathcal{L}}^P} \mid I \models \mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})\})}$ . As in the base case, we use the soundness result of corollary 2: if  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \mathbf{t}$  (or  $\mathbf{f}$ ), then  $P(\bar{a})^{\text{glb}_{\leq_P}(\{I|_{\leq_{\mathcal{L}}^P} \mid I \models \mathcal{M}_{\{P\}}(D, \mathcal{L})\})} = \mathbf{t}$  (or  $\mathbf{f}$ ). We need to prove the property for the case in which  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \mathbf{u}$ .  $P(\bar{a})^{\mathcal{K}_{\mathcal{L}}} = \mathbf{u}$  implies that  $P(\bar{a}) \notin P^D$  and  $\mathcal{K}|_{<_{\mathcal{L}}^P} \not\models \Psi_P(\bar{a})$ . We must prove that there exist  $\leq_{\mathcal{L}}^P$ -structures  $I_1, I_2$  s.t.  $I_1 \models \mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})$ , and  $P(\bar{a})^{I_1} = \mathbf{f}$ ; and  $I_2 \models \mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})$  and  $P(\bar{a})^{I_2} = \mathbf{t}$ . Since  $\mathcal{K}|_{<_{\mathcal{L}}^P} \not\models \Psi_P(\bar{a})$ , there is a literal  $L_i$  in  $\Psi[\bar{a}]$  s.t.  $\mathcal{K}|_{<_{\mathcal{L}}^P} \not\models L_i$ . Because of IH, either  $\mathcal{K}|_{<_{\mathcal{L}}^P} \models \neg L_i$  or  $L_i^{\mathcal{K}|_{<_{\mathcal{L}}^P}} = \mathbf{u}$ . In the first case, select  $I'$  as any model of  $\mathcal{M}_{<_{\mathcal{L}}^P}(D, \mathcal{L})$ . In the second case, select  $I'$  such that  $I' \models \neg L_i$  and  $I' \models \mathcal{M}_{<_{\mathcal{L}}^P}(D, \mathcal{L})$  (by IH  $I'$  can always be constructed). Construct  $I_1$  extending  $I'$  with  $P^D$ . Since  $P(\bar{a})^{\mathcal{K}} = \mathbf{u}$ ,  $P(\bar{a}) \notin P^D$ . We prove that  $I_1 \models \mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})$ . Since  $I_1$  is an extension of  $I'$  with the atoms of the database of  $P$ ,  $I_1$  satisfies all database atoms  $P(\bar{a})$  in  $\mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})$ . We only need to prove that for every tuple  $\bar{d}$  in the domain  $I_1 \models \Psi_P(\bar{d}) \supset (P(\bar{d}) \supset (P(\bar{d}) \in D))$ . This is easy since  $I_1 \models P(\bar{d})$  implies  $P(\bar{d}) \in D$ . Construct now  $I_2$  identical to  $I_1$  except that  $P(\bar{a})^{I_2} = \mathbf{t}$ . Since by construction  $I_2$  satisfies the database atoms, we just need to prove that  $I_2 \models \Psi_P(\bar{d}) \supset (P(\bar{d}) \supset (P(\bar{d}) \in P^D))$ , for every tuple  $\bar{d}$  of the domain. The proof for the cases where  $\bar{d} \neq \bar{a}$  is analog to the case of  $I_1$ , so we focus on the case  $\bar{d} = \bar{a}$ . We selected  $I'$  such that  $I' \models \neg L_i$ , hence  $I_2 \models \neg L_i$ , and therefore  $I_2 \models \neg \Psi_P(\bar{a})$ . It follows that  $I_2 \models \Psi_P(\bar{a}) \supset (P(\bar{a}) \supset (P(\bar{a}) \in P^D))$ . As a result,  $I_2 \models \mathcal{M}_{\leq_{\mathcal{L}}^P}(D, \mathcal{L})$ , and this concludes the proof of the lemma.  $\square$

The theorem is a direct consequence of lemmas 1 and 3, and the fact that  $\Sigma = \bigcup_{P \in \Sigma} \leq_{\mathcal{L}}^P$ .  $\square$

*Note 2.* As Example 5 shows, the requirement in Theorem 2 that  $\Psi_\theta$  is a conjunction of literals is indeed necessary.

*Example 8.* Consider a database in which  $D = \{P_1(a), P_1(b), P_2(a), Q(c)\}$  and  $\mathcal{L}$  is the set of LCWA of Example 6 without  $\mathcal{LCWA}(Q(x), S(x))$ . That is,

$$\mathcal{L} = \left\{ \begin{array}{lll} \mathcal{LCWA}(P_1(x), \mathbf{t}) & \mathcal{LCWA}(P_2(x), \mathbf{t}) & \mathcal{LCWA}(Q(x), P_1(x) \wedge P_2(x)) \\ \mathcal{LCWA}(S(x), Q(x)) & \mathcal{LCWA}(R(x), Q(x)) & \end{array} \right\}$$

This database is hierarchically closed as the dependency graph induced by  $\mathcal{L}$  is acyclic. Also, as each window of expertise is a conjunction of literals, the conditions of Theorem 2 are satisfied in this case. Clearly,  $Q(a)^{\mathcal{K}_\mathcal{L}} = \mathbf{f}$ ,  $Q(b)^{\mathcal{K}_\mathcal{L}} = \mathbf{u}$  and  $Q(c)^{\mathcal{K}_\mathcal{L}} = \mathbf{t}$ . By Corollary 2, then,  $\mathcal{M}(D, \mathcal{L}) \models \neg Q(a)$  and  $\mathcal{M}(D, \mathcal{L}) \models Q(c)$ . Moreover, by Theorem 2,  $\mathcal{K}_\mathcal{L}$  is an optimal approximation of  $\mathcal{M}(D, \mathcal{L})$ .

## 4 Query Answering

In the previous section we presented conditions to compute a (optimal) 3-valued interpretation for the meaning of a database and its set of LCWAs. In this section we show how these interpretations can be used for query answering in incomplete databases.

**Definition 12.** Given a 3-valued  $\Sigma$ -interpretation  $\mathcal{K}$  and a query  $Q[\bar{x}]$  in  $\Sigma$ , define:

- *Certain answers:*  $Cert(Q[\bar{x}]) = \{\bar{x} \mid Q[\bar{x}]^{\mathcal{K}} = \mathbf{t}\}$ .
- *Possible answers:*  $Poss(Q[\bar{x}]) = \{\bar{x} \mid Q[\bar{x}]^{\mathcal{K}} \geq \mathbf{u}\}$ .

*Note 3.* The notions of certain and possible answers proposed here differ from definitions considered in some domains of incomplete databases (see for instance [1]), where the certain and possible answers depend on whether an open or closed-world is assumed. Our definition is based on 3-valued semantics and it does not rely on the assumption that is adopted for the database.

Given a 3-valued Herbrand  $\Sigma$ -interpretation  $\mathcal{K}$  and a query  $Q[\bar{x}]$  in  $\Sigma$ , we compute certain/possible answers by algorithm 1. Steps 1. and 2. are computed once (unless  $\mathcal{K}$  is changed), while step 3. is executed for each new query. Observe that in step 3.a we under-approximate positive occurrences and over-approximate negative occurrences of predicates, while in 3.b we do the converse.

---

**Algorithm 1** : Computing Certain/Possible Answers

---

- 1: For each predicate  $P$ , extend  $\Sigma$  with new predicate symbols  $P_t$  and  $P_u$ .
- 2: Define  $I^{\mathcal{K}}$  as the 2-valued Herbrand  $\Sigma$ -interpretation such that:

- $R^{I^{\mathcal{K}}} = R^{\mathcal{K}}$  if  $R^{\mathcal{K}}$  is 2-valued in  $\mathcal{K}$ .
- $P_t^{I^{\mathcal{K}}} = \{\bar{d} \mid P(\bar{d})^{\mathcal{K}} = \mathbf{t}\}$ .
- $P_u^{I^{\mathcal{K}}} = \{\bar{d} \mid P(\bar{d})^{\mathcal{K}} \geq \mathbf{u}\}$ .

- 3: Consider the query  $Q[\bar{x}]$ .

3.a To obtain **certain** answers for  $Q[\bar{x}]$  compute  $Q^c[\bar{x}]$  as follows:

- Replace positive occurrences of 3-valued predicates  $P$  by  $P_t$ .
- Replace negative occurrences of 3-valued predicates  $P$  by  $P_u$ .

3.b To obtain **possible** answers for  $Q[\bar{x}]$  compute  $Q^p[\bar{x}]$  as follows:

- Replace positive occurrences of 3-valued predicates  $P$  by  $P_u$ .
  - Replace negative occurrences of 3-valued predicates  $P$  by  $P_t$ .
- 

**Theorem 3.** *The ground tuple  $\bar{d}$  is a certain (alternatively possible) answer for  $Q[\bar{x}]$  in  $\mathcal{K}$  iff  $\bar{d}$  is an answer to  $Q^c$  (alternatively  $Q^p$ ) in  $I^{\mathcal{K}}$ .*

*Proof.* Consider the (non-standard representation of) truth assignment interpreting a formula  $\varphi$  in a pair of structures  $(I, J)$  with the same domain, such that positively occurring atoms of  $\varphi$  are interpreted by  $I$ , and negatively occurring ones by  $J$ . A satisfaction relation  $\models$  between these valuations and formulas in  $\Sigma$  is inductively defined as follows:

- $(I, J) \models P(\bar{d})$  iff  $I \models P(\bar{d})$ , i.e.,  $\bar{d}^I \in P^I$ ;
- $(I, J) \models \neg\varphi$  iff  $(J, I) \not\models \varphi$ ;
- $(I, J) \models \varphi \vee \psi$  iff  $(I, J) \models \varphi$  or  $(I, J) \models \psi$ ;
- $(I, J) \models \exists \bar{x} \varphi(\bar{x})$  iff there is a  $\bar{d} \in \text{dom}(I)$ , such that  $(I[\bar{x}/\bar{d}], J[\bar{x}/\bar{d}]) \models \varphi(\bar{x})$ .

There is a strong link with 3-valued logic. Indeed, when  $(I, J)$  is the result of splitting a 3-valued interpretation  $\mathcal{K}$  in a 2-valued underestimation  $I$  and 2-valued overestimation  $J$  (see section 3.1), then it is well-known that :

- $\varphi^{\mathcal{K}} = \mathbf{f}$  iff  $(J, I) \not\models \varphi$ ;
- $\varphi^{\mathcal{K}} = \mathbf{u}$  iff  $(J, I) \models \varphi$  and  $(I, J) \not\models \varphi$ ;
- $\varphi^{\mathcal{K}} = \mathbf{t}$  iff  $(I, J) \models \varphi$

Let  $(I, J)$  be the pair of 2-valued Herbrand  $\Sigma$ -interpretations associated with  $\mathcal{K}$ . By construction of  $I^{\mathcal{K}}$  (see step 2 above), for every 3-valued predicate  $P$ ,  $(P_t)^{I^{\mathcal{K}}} = P^I$  and  $(P_u)^{I^{\mathcal{K}}} = P^J$ . It is easy to see that for any tuple of domain elements  $\bar{d}$ ,  $(I, J) \models Q[\bar{d}]$  iff  $I^{\mathcal{K}} \models Q[\bar{d}]$  and  $(J, I) \models Q[\bar{d}]$  iff  $I^{\mathcal{K}} \models Q^p[\bar{d}]$ . It follows that if  $I^{\mathcal{K}} \models Q^c[\bar{d}]$ , then  $Q[\bar{d}]^{\mathcal{K}} = \mathbf{t}$ , i.e.,  $\bar{d}$  is a certain answer for  $Q$ , and if  $I^{\mathcal{K}} \models Q[\bar{d}]$  then  $Q[\bar{d}]^{\mathcal{K}} \geq \mathbf{u}$ , i.e.,  $\bar{d}$  is a possible answer for  $Q$ .  $\square$

**Corollary 3.** *Let  $\mathfrak{D} = (D, \mathcal{L})$  be a hierarchically closed database. If the ground tuple  $\bar{d}$  is a certain answer for  $\mathcal{Q}$  in  $\mathcal{K}_{\mathcal{L}}$ , then  $\mathcal{M}(D, \mathcal{L}) \models \mathcal{Q}[\bar{d}]$ . satisfiable).*

*Example 9.* Consider the database  $\mathfrak{D}$  of Example 8. Let us check the query  $\mathcal{Q}[x] = P_1(x) \wedge R(x)$  for  $x = b$ . Let  $\mathcal{K}_{\mathcal{L}}$  be the 3-valued interpretation that is induced by  $\mathfrak{D}$  (obtained by the inductive construction of Definition 11), and let  $I$  be the 2-valued interpretation derived from  $\mathcal{K}_{\mathcal{L}}$  by step 2 above. Since  $P_1$  is 2-valued in  $\mathcal{K}_{\mathcal{L}}$ , for every constant  $a$ ,  $P_1(a)^{\mathcal{K}_{\mathcal{L}}} = P_1(a)^I$ . The predicate  $R$  is 3-valued in  $\mathcal{K}_{\mathcal{L}}$ , and following step 2 we replace it by  $R_t$  and  $R_u$  in  $I$ . It follows that  $R_t(b)^I = \mathbf{f}$  and  $R_u(b)^I = \mathbf{t}$ . Rewriting  $\mathcal{Q}[x]$  using step 3, we obtain  $\mathcal{Q}^c[x] = P_1(x) \wedge R_t(x)$  and  $\mathcal{Q}^p[b] = P_1(b) \wedge R_u(b)$ . It follows that  $x = b$  is a possible but not a certain answer for  $\mathcal{Q}[x]$  (that is,  $b \in \text{Poss}(\mathcal{Q}) - \text{Cert}(\mathcal{Q})$ ).

The last example demonstrates query answering in hierarchically closed databases. The whole process can be summarized as follows:

- 1) Given a database  $\mathfrak{D} = (D, \mathcal{L})$ , check whether it is hierarchically closed (this is simple, as both the construction of the dependency graph and cycle checking for it are polynomial in the size of  $D$ ).
- 2) Apply the procedure that is sketched in the paragraph above the proof of Proposition 8 for constructing the 3-valued interpretation  $\mathcal{K}_{\mathcal{L}}$  that is induced by  $\mathfrak{D}$ .
- 3) Given a query  $\mathcal{Q}$ , apply Algorithm 1 with  $\mathcal{K}_{\mathcal{L}}$  and  $\mathcal{Q}$ .

## 5 Related Works

In [5] the authors introduced the concept of closed-world information on a formula  $\mathcal{Q}$  in the context of logical agents. The idea amounts to the specification of which parts of a logical theory are complete. More formally, an agent has CWI relative to a formula  $\mathcal{Q}(\bar{x})$  if every ground instance  $\mathcal{Q}(\bar{a})$  is either entailed by the first-order knowledge base or necessarily false. As argued in this paper (see the discussion after Definition 4), the notions of CWI and LCWA (as presented here) are closely related, but they capture different properties of different entities.

Doherty et al. [4] propose a semantics for the CWI of Etzioni et al. in terms of (second-order) circumscription. This approach generalizes the one in [5] by allowing limited use of negation and disjunction in the representation of CWI, while retaining tractability of reasoning. The query processing in [4] is based on fixpoint semantics. One of the robust characteristics of this approach is the possibility to express inductive definitions by means of CWI formulas. The correspondence between this approach and our notion of LCWA is considered in detail in [2].

The database community has made important contributions to the study of locally complete databases. In [14], Motro formalizes the concept of partial completeness in relational databases by means of *completeness constraints*, which are concerned with “true information that must be part of a complete database”. Such constraints are represented in by conjunctive relational calculus. The role of Motro’s completeness constraints is taken here by the windows of expertise, which are represented by a more expressive language (function-free first-order). In [11] Levy investigates how to determine whether the answer to a given query is complete even if the database is incomplete.

The notion of completeness is semantically characterised in terms of *virtual relations* (those that are true in the real world) and *available relations* (the actual database). Neither Motro nor Levy address the problem of obtaining possible answers to queries, as we do in this paper.

In the context of answer-set programming, Gelfond et al. [6] introduce the possibility to *partially* define a predicate by combining in one rule classical negation and negation as failure. Consider, for instance, the following program:<sup>6</sup>  $\{\neg p \leftarrow \text{not } p, \varphi\}$ . Under stable-models semantics [7], the truth of  $\neg p$  can be established if no evidence is found about  $p$  and  $\varphi$  is provable. The relationship with the LCWA in this context is clear:  $\varphi$  is a window of expertise of  $p$ . Observe, however, the different setting: The knowledge encoded in a logic program is specified by its formal semantics, and the answers to queries are retrieved according this semantics. On the other hand, an (incomplete) relational database is an entity without an a priori formal meaning, and the interpretation of the answers is frequently left to the user. In this work we tackle this informality by providing a semantics and a technique to answer queries based on the formal specification.

## 6 Conclusions and Future Work

We have presented a general algebraic fixpoint theory for the local closed-world assumption of relational databases. The framework proposed here re-conciliates the notions of CWI and LCWA used in the contexts of knowledge-base agents and relational databases. More specifically, (i) important cases in which LCWA induces CWI are identified, and (ii) a simple query answering mechanism that allows to distinguish between possible and certain answers is introduced. Future lines of research include the following:

- **Extension from relational to deductive databases.** In deductive databases (sometimes referred as Datalog), so called intensional predicates are defined in terms of extensional relations. Given a relational database and set of LCWAs, it would be informative to identify where locally complete extensional relations induce complete knowledge on intensional ones. The immaterial nature of intensional predicates presumably requires some extension of the notion of LCWA considered in this paper.
- **Integration with mediator-based systems [10].** In [2] it was shown how to represent LCWA information over a number of different data-sources. Informally, the idea amounts to represent that a set of data-sources, taken together, may store complete knowledge about certain predicate. An open question is how to explore this additional knowledge to retrieve more informative answers from queries to the mediator. An approach based on the semantic considerations presented in this paper could provide a well-founded solution to the elusive problem of answering negative queries from mediator systems.

---

<sup>6</sup> Following standard conventions, the symbols  $\neg$  and  $\text{not}$  represent classical negation and negation as failure, respectively.

- **Efficient query answering techniques for the LCWA.** The methods we presented here for query answering require the computation of a 3-valued interpretation. This approach allows a straightforward identification of the exact knowledge endorsed by a database, but if the set of the LCWA expressions is updated the 3-valued interpretation must be re-computed. Incremental methods for updating 3-valued interpretations may be incorporated for reducing the computational complexity of the revision process.

## References

1. S. Abiteboul and O.M. Duschka. Complexity of answering queries using materialized views. In *PODS*, pages 254–263, 1998.
2. A. Cortés-Calabuig, M. Denecker, O. Arieli, B. Van Nuffelen, and M. Bruynooghe. On the local closed-world assumption of data-sources. In *Proc. 8th LPNMR, LNCS 3662*, pages 145–157, 2005.
3. M. Denecker, V.W. Marek, and M. Truszczynski. Uniform semantic treatment of default and autoepistemic logics. *Artif. Intell.*, 143(1):79–122, 2003.
4. P. Doherty, W. Łukaszewicz, and A. Szalas. Efficient reasoning using the local closed-world assumption. In *Proc. 9th AIMSA, LNCS 2407*, pages 49–58, 2000.
5. O. Etzioni, K. Golden, and D. Weld. Sound and efficient closed-world reasoning for planning. *Artificial Intelligence*, 89(1-2):113–148, 1997.
6. M. Gelfond and V. Lifschitz. Logic programs with classical negation. In *ICLP*, pages 579–597, 1990.
7. M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Comput.*, 9(3/4):365–386, 1991.
8. G. Grahne. Information integration and incomplete information. *IEEE Data Engineering Bulletin*, 25(3):46–52, 2002.
9. G. Grahne and A. Mendelzon. Tableau techniques for querying information sources through global schemas. In *Proc. 7th ICDT, LNCS 1540*, pages 332–347, 1999.
10. M. Lenzerini. Data integration: A theoretical perspective. In *Proc. 21st PODS*, pages 233–246, 2002.
11. Alon Y. Levy. Obtaining complete answers from incomplete databases. In *VLDB*, pages 402–412, 1996.
12. A. O. Mendelzon and G. A. Mihaila. Querying partially sound and complete data sources. In *Proc. 20th PODS*, pages 162–170, 2001.
13. T. Millstein, A. Levy, and M. Friedman. Query containment for data integration systems. In *Proc. 21st PODS*, pages 67–75, 2002.
14. A. Motro. Integrity = validity + completeness. *ACM Trans. Database Syst.*, 14(4):480–502, 1989.
15. F. Naumann, J.C. Freytag, and U. Leser. Completeness of integrated information sources. *Information Systems*, 29(7):583–615, 2004.
16. R. Reiter. Towards a logical reconstruction of relational database theory. In *On Conceptual Modelling, Perspectives from Artificial Intelligence, Databases, and Programming Languages.*, pages 191–233, 1982.